On the Sliding Block Maxima Method in Extreme Value Statistics

Torben Staud
from Neuss

RUHR UNIVERSITÄT BOCHUM RUB

A Dissertation presented
to
the Faculty of Mathematics
Ruhr University Bochum

in fulfillment of the requirements
for the degree of
Dr. rer. nat.
in the subject of
Mathematics and Statistics

November 2024

Abstract

In both applications and statistics understanding of tail behavior is an important matter. Estimators for tail characteristics may be obtained by the classical disjoint block maxima method. Recently, this method has been improved by considering so-called sliding block maxima and desirable properties were shown in case-by-case analyses. Estimators based on the latter method might exhibit a smaller variance while maintaining a bias of similar magnitude. In this thesis, the block maxima methods are investigated systematically for large classes of estimators in the context of dependent data. These classes include so called U-statistics, a classical tool in non-parametric statistics. Limit theorems are provided allowing for the quantification of estimation errors. Applications include the probability weighted moment estimator in extremes and Kendall's τ of block maxima, a measure of tail dependence. In all situations the sliding block maxima method is found to be superior to its disjoint counterpart. The findings are extended to a setting of non-stationary time series and finite sample properties are investigated by means of Monte-Carlo simulation studies which confirm the theoretical results.

While the sliding block maxima method offers the advantage of reduced variance, deriving this variance explicitly can be challenging; the latter being important to obtain confidence intervals for sliding block maxima estimators. This thesis proposes a universal solution using novel bootstrapping methods tailored for general block maxima estimators, effectively eliminating the need for additional tuning parameters as often encountered in bootstraps for time series. A byproduct of this development is a new class of block maxima estimators. The asymptotic normality of these estimators and the formal consistency of the proposed bootstrap methods are proven. Further, the results are applied to estimators from extreme value statistics. The performance of the methods is evaluated by large-scale simulation studies, and their practical utility is demonstrated in a case study on precipitation data.

A key motivation for choosing the sliding block maxima method over its classic counterpart is the fact that estimators based on the former method exhibit smaller asymptotic variances. In the situation of one-dimensional observations, these variances can be represented as integrals over covariances of the well known Marshall Olkin bivariate exponential distribution. Solving an open problem from the literature, the maximal correlation coefficient of the latter distribution is derived and then applied, to provide an alternative proof of the important variance inequality between disjoint and sliding blocks based estimators.

Acknowledgements

First, I would like to express my deepest gratitude to my advisor Professor Axel Bücher for his outstanding guidance and support throughout the supervision of this thesis. Thank you for the numerous fruitful discussions, your enthusiasm, for believing in me and for always taking the time. Your way of teaching, doing statistics and the vast amount of expertise have always motivated and inspired me.

I am also deeply thankful to Professor Johan Segers for offering me the opportunity to conduct an educational and insightful research stay at the Université catholique de Louvain. For your hospitality and the wisdom I was able to profit from, I am thankful. Also, I am sincerely grateful for your agreement to review this thesis.

Furthermore, I owe thanks to the many individuals I had the pleasure to work with during this journey: the colleagues of the Chair of Mathematical Statistics and Probability Theory at Heinrich-Heine University and the whole statistics work group at Ruhr University; I am honored to be a child of both universities. It was a great working environment and thanks to you, the journey was even more enjoyable. Special thanks go to Tobias, Niklas and Christian for being great friends and to Professor Peter Kern and Professor Holger Dette for being my mentors.

Regarding financial support, I am grateful to the Deutsche Forschungsgemeinschaft and the RUB Research School.

Lastly, I would like to express my heartfelt gratitude to my family and friends for their constant encouragement. I owe special thanks to Lena for her invaluable support and love; without you, I would not be the man I am today.

"*Ach, was geht es uns gut!*" † 2019

Contents

1 Introduction

Extreme value theory and statistics are concerned with describing, modelling and estimating extreme occurrences of probabilistic phenomena. These often correspond to quantities exceeding critical levels. The effects of the latter can be catastrophic and have a significant impact on society resulting in a demand for reliable statistical methodologies in order to mitigate these risks. Areas of risk range from extreme weather events, such as hurricanes, floods, and heatwaves (Engeland et al., 2004; Shen et al., 2016), as well as extreme financial events like maximal stock losses (Longin, 2000) to pandemic modeling (Thomas et al., 2016) and size effect on material strength (Harter, 1978). For example, assessing the time in which one expects a certain high flood level to be exceeded is of great importance in flood protection. Another example is given by the capital reserves of banks which need to be sufficient to cover potential losses in case of financial crises. A comprehensive introduction to these applications and theoretical foundations can be found in Beirlant et al. (2004); Coles (2001); Embrechts et al. (1997). The statistical analysis of extreme values poses significant challenges, as can be seen from the following example, where the classical empirical quantile as an estimator for the 99%-quantile fails: Consider the real-valued sample $\mathcal{X}_n = (x_1, \ldots, x_n)$ and $\hat{q}_{n,.99} = \inf\{x \in \mathbb{R} : \hat{F}_n(x) \geq 0.99\}$, where $\hat{F}_n$ denotes the empirical cumulative distribution function of the sample $\mathcal{X}_n$. For all $n \leq 99$ the quantile estimator $\hat{q}_{n,.99}$ degenerates to the maximum, i.e. $\hat{q}_{n,.99}(\mathcal{X}_n) \equiv \max(\mathcal{X}_n)$, neglecting information contained in the sample. This fundamental example illustrates that classical statistical methodology may fail in extreme value theoretic frameworks where one needs to extrapolate from the bulk of the data to the tail. An introduction to the topic of extreme value theory and statistics is given by Beirlant et al. (2004).

Typically, modelling extremes is done by fitting distributions to the tail of the data of which there are two fundamental regimes. The first dates back to Gumbel (1958) in which the author laid the foundation for the so-called *block maxima method*: Consider real-valued observations $X_1, \ldots, X_n$ and a block size $r \in \mathbb{N}, r \leq n$. The *disjoint block maxima sample* $\mathcal{M}^{(\mathrm{db})} = \big(m_i^{(\mathrm{db})} : 1 \leq i \leq \lfloor n/r \rfloor\big)$ consists of the respective maximal observations of the ith disjoint block, that is $m_i^{(\mathrm{db})} = \max\{x_j : (i-1)r + 1 \leq j \leq ir\}$; Figure 1.1 illustrates this procedure.

The disjoint block maxima are then assumed to follow a certain three-parameter distribution allowing for using statistical methodology from the parametric toolbox. This assumption is legitimized by the *Fisher-Tippett-Gnedenko Theorem* due to Fisher and Tippett (1928); Gnedenko (1943): It states that if suitably normalized block maxima of independent and identically distributed (i.i.d.) random variables converge weakly to a non-degenerate distribution, then the limit distribution is already a *Generalized Extreme*

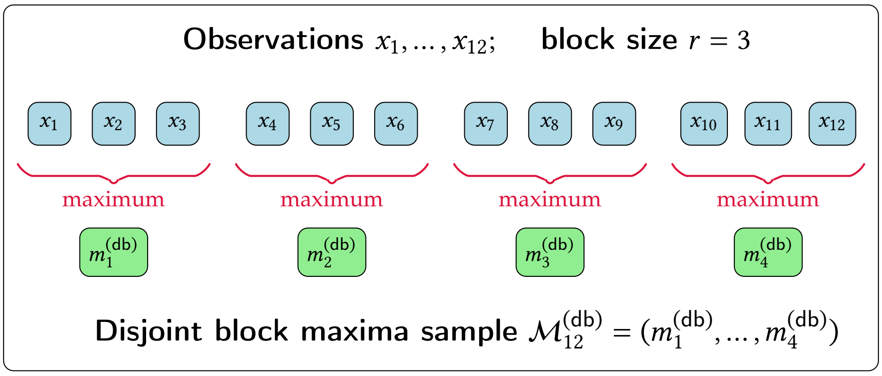

Figure 1.1: Illustration of the (disjoint) block maxima method.

Value (GEV) distribution with cumulative distribution function (cdf)

$$G(x) = G_{\mu,\sigma,\gamma}(x) = \begin{cases} \exp\left\{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\} \mathbb{1}\left(1+\gamma\frac{x-\mu}{\sigma} > 0\right), & \gamma \neq 0, \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, & \gamma = 0, \end{cases}$$

where $\mu, \gamma \in \mathbb{R}, \sigma > 0$.

The GEV-distribution has three parameters, two of which are location, scale parameters while the third parameter γ determines the shape of the tails and is thus also called *extreme value index*. It parameterizes three families: $\gamma < 0$ corresponds to the (reverse) Weibull distribution which has short (finite) tails, $\gamma = 0$ corresponds to the Gumbel distribution which has light (exponentially decreasing) tails and finally, $\gamma > 0$ leads to the Fréchet distribution which has heavy (power-law) tails. The Fisher-Tippett-Gnedenko Theorem has been generalized in different directions, one of which adresses serial dependence. The *Extremal Limit Theorem* allows for replacing the i.i.d. assumption by a short-range dependency structure, known as the $D(u_n)$ condition; (Leadbetter, 1974). This framework was further extended to include the concept of *asymptotic independence of maxima* in O'Brien (1987). These advancements have ensured that the block maxima method remains a widely used approach in applications, such as modeling temperature maxima in climatology.

Statistical theory for estimators based on disjoint block maxima can be found for example in Smith (1984); Bücher and Segers (2018b) for the maximum likelihood estimator (MLE) and in Hosking et al. (1985) where the probability weighted moments (PWM) estimator based on block maxima was considered.

Next to the block maxima method, the competing classical tool for tail inference is the *peak-over-threshold* (POT) method which is based on the works Balkema and de Haan (1974); Pickands (1975) and was fully established mathematically after the work Smith (1984), see also (Davison and Smith, 1990, Section 1). It considers observations exceeding a certain high threshold chosen beforehand; see Figure 1.2 for an illustration.

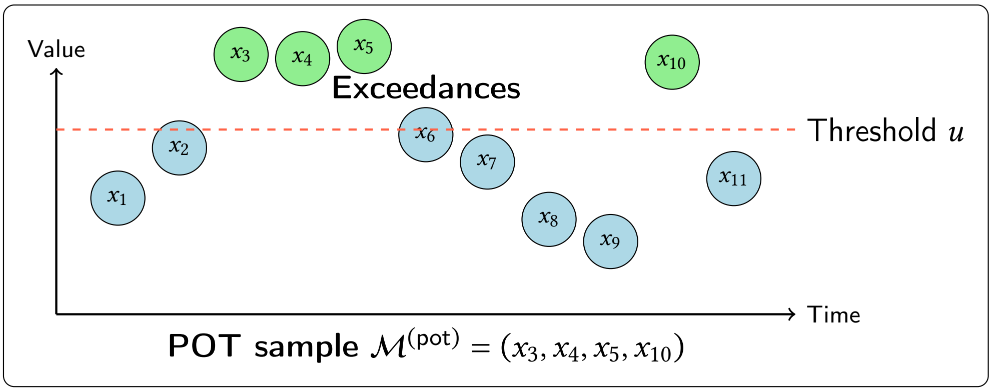

Figure 1.2: Illustration of the peak-over-threshold method.

The remaining observations are then assumed to follow a *Generalized Pareto Distribution*, which is legitimized by the *Pickands-Balkema-de Haan Theorem*; (Balkema and de Haan, 1974; Pickands, 1975). Literature on the POT method is well established, see de Haan and Ferreira (2006) among others.

One is able to synthesize both the results of Fisher-Tippett-Gnedenko and Pickands-Balkema-de Haan: under a mild regularity condition the suitably normalized sample of block maxima is asymptotically GEV-distributed if and only if the suitably normalized sample of exceedances is asymptotically GPD-distributed; (Pickands, 1975, Theorem 7). In this case, the shape parameters for the GEV and GPD coincide.
Both fundamental principles require the choice of a tuning parameter, that is the threshold and the block length for the POT method and the BM method, respectively. They are of high importance as one needs to guarantee first, the asymptotic approximation to be correct and second, the remaining sample to be large enough to mitigate increasing variance which results in a bias-variance tradeoff for a fixed size of the original sample, (Coles, 2001). For the POT principle there are well established methods at hand to do so. The *Hill plot* consisting of plotting estimates against the threshold and choosing the first occurring stability plateau (Drees et al., 2000). Also, there are formal approaches based on minimizing the asymptotic MSE with regard to the threshold u; see Danielsson et al. (2001); Drees and Kaufmann (1998). While the same principle of stability plots as for the Hill plot applies to the BM approach, often the block size is determined by nature of the problem. For instance, yearly, monthly, or weekly temperature maxima.
Intuitively the BM method might seem inefficient or "wasteful" as the size of $\mathcal{M}^{(\text{db})}$ is of order $n/r \ll n$, see e.g. Coles (2001, Section 4.1). However, Bücher and Zhou (2021) presented situations in which the block maxima method outperformed the POT method and vice-versa depending on the characteristics to estimate and the data generating processes. Nonetheless, practitioners use the block maxima method, see (Philip et al., 2020, Sections 4.2.2, 4.2.3). Reasons for that comprise the following: First, as stated before, block size choices might be more natural in climatology settings as op-

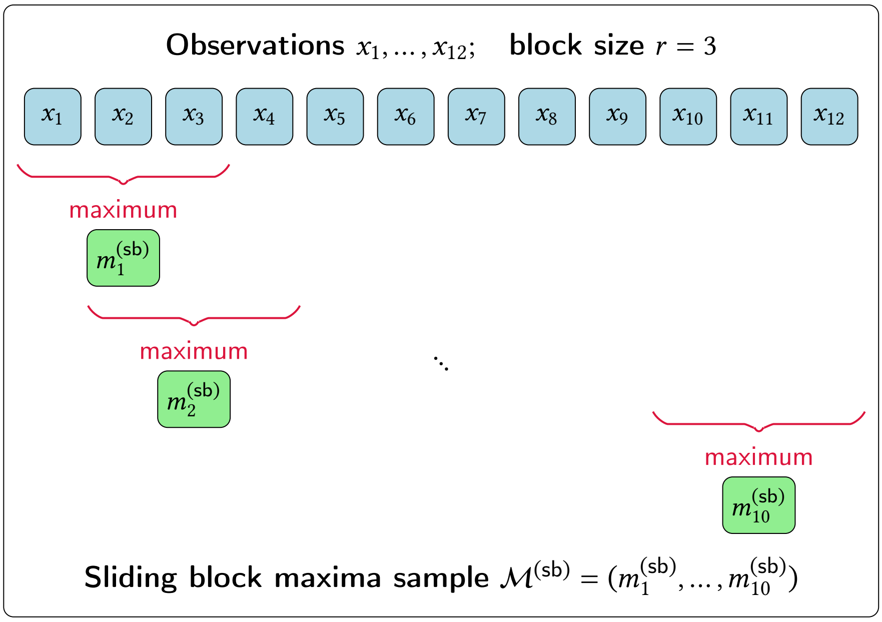

Figure 1.3: Illustration of the sliding block maxima method.

posed to setting a threshold; (Naveau et al., 2009; van den Brink et al., 2005). Secondly, observations might only be accessible as maximal values over certain blocks because underlying data is not available; (Kharin et al., 2007). Thirdly, when observations are not exactly i.i.d. block maxima retain a certain robustness as e.g. dependence in blocks is not harmful as long there is little dependence between blocks and the parameter of interest is a functional of the block maxima distribution; (Katz et al., 2002; Madsen et al., 1997). In contrast, the POT method would not allow for immediate estimation of, say, return levels but require a transformation based on an estimation of the *extremal index*. The latter parameter describes the tendency of extremes to appear in clusters; see Beirlant et al. (2004) for an introduction.

In contrast to the POT method, the block maxima method has only recently been studied under the assumption that the approximation by the GEV distribution holds asymptotically; a so-called *(Max) Domain of Attraction* (DOA) condition. In Dombry (2015) the consistency of the MLE based on disjoint block maxima was proven, while Dombry and Ferreira (2019) and Ferreira and de Haan (2015) established asymptotic normality of the GEV ML and PWM estimators based on i.i.d. block maxima, respectively. Extending these findings, a time series approach incorporating mixing conditions has been implemented in Bücher and Segers (2014), Bücher and Segers (2018b), Bücher and Segers (2018a). In line with this research, this thesis aims at establishing a wide range of statistical verification of block maxima methods assuming the more realistic DOA condition and short-range dependent time series.

Further improvement of the block maxima method has been initiated by Beirlant et al. (2004), Robert et al. (2009) in which so-called *sliding block maxima* have been proposed: Consider real-valued observations $X_1, \ldots, X_n$ and a block size $r \in \mathbb{N}, r \leq n$. The *sliding block maxima sample* $\mathcal{M}^{(\mathrm{sb})} = \left(m_i^{(\mathrm{sb})} : 1 \leq i \leq n-r+1\right)$ consists of the respective maximal observations of the ith sliding block, that is $m_i^{(\mathrm{sb})} = \max\{x_j : i \leq j \leq i+r-1\}$. Figure 1.3 illustrates this procedure. The sliding block maxima sample retains more information than its disjoint version as $\mathcal{M}^{(\mathrm{sb})}$ contains $\mathcal{M}^{(\mathrm{db})}$. It is remarkable that under the assumption of strictly stationary time series the sliding sample itself is still stationary and does not destroy temporal dependence which is the fundament of why sliding blocks-based estimators work.

Asymptotically linear estimators based on sliding block maxima typically exhibit the same bias as the respective disjoint counterpart while having a smaller variance; (Bücher and Segers, 2018a; Bücher and Zanger, 2023) consider the heavy tailed MLE for the GEV parameters or the PWM estimator, respectively. It is noteworthy that these effects have not been observed in the POT framework; (Cissokho and Kulik, 2021; Drees and Neblung, 2021). These findings point to numerous open research questions: First, is one able to prove this behavior for a large class of estimators based on sliding block maxima? Second, do the effects transfer to non-stationary situations? Third, does the form of the variance allow for the simple construction of confidence intervals or bootstrapping? Finally, is one able to introduce new block maxima methods which offer further advantages? Recently, the last question has been partially answered by the consideration of the *all block maxima method* in Oorschot and Zhou (2020) which dates back to (Segers, 2001, Chapter 5); see Figure 1.4 for an illustration. In the second reference, the author proved consistency of a U-statistics based estimator for γ. Proving asymptotic normality, in the first reference the authors showed that the (pseudo) ML estimator for the shape γ in the heavy-tailed setting based on all block maxima outperformed the disjoint and sliding block maxima based ML estimators in the sense, that the former had a smaller asymptotic variance. However, as the estimator destroys serial dependence clustering effects which might appear in time series are neglected. As a consequence scale estimation or estimation of characteristics depending on the dependence structure is not immediate; see (Oorschot and Zhou, 2020, p. 18).

The thesis presented here aims at answering the questions from the last paragraph and providing statistical methodology on block maxima methods in time series extremes. To this end we present three articles:

The first article showcased in Chapter 3 allows for answering the first question above by considering U-statistics of block maxima and estimators based on them.

Independently of each other U-statistics have been introduced in Halmos (1946) and Hoeffding (1948). They can be understood as a generalized mean: Consider i.i.d. observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and a symmetric function $h : \mathbb{R}^\rho \to \mathbb{R}$, called *kernel* of order $\rho \in \mathbb{N}$.

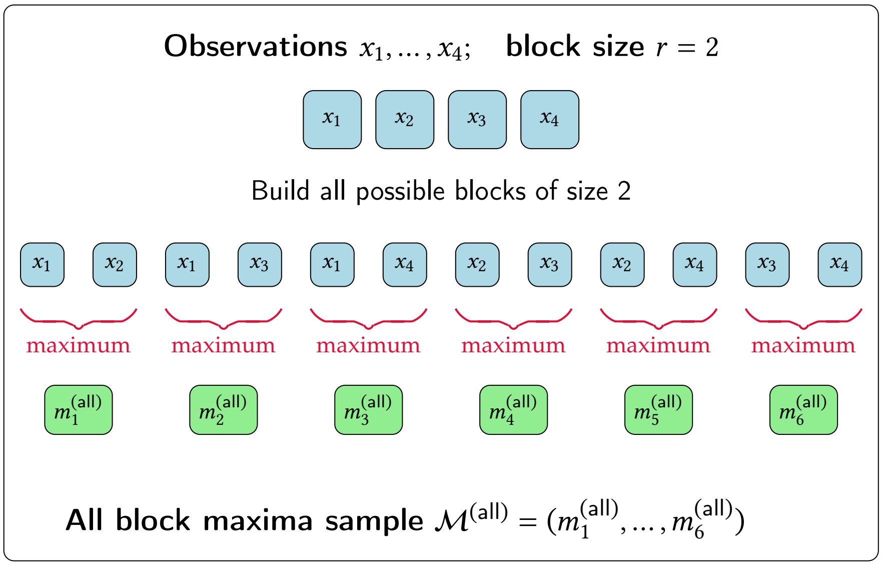

Figure 1.4: Illustration of the all block maxima method.

The characteristic of interest and assumed to exist is

$$\theta = \mathrm{E}[h(\boldsymbol{X}_1, \dots, \boldsymbol{X}_\rho)].$$

A natural estimator of this quantity is given by the expression

$$U_n(h) = \binom{n}{\rho}^{-1} \sum_{\boldsymbol{i} \in J_{n,\rho}} h(X_{i_1}, \dots, X_{i_\rho}), \tag{1.1}$$

where $J_{n,\rho} = \{\boldsymbol{i} \in \mathbb{N}^\rho : 1 \le i_1 < \dots < i_\rho \le n\}$. The prefix U stems from U-statistics being unbiased. U-statistics appear in many situations: The classical mean estimator arises from $\rho = 1, h = \mathrm{id}_\mathbb{R}$; for $k \in \mathbb{N}$ the kth moment estimator arises from $\rho = 1, h(x) = x^k$; the empirical variance arises when $\rho = 2, h(x, y) = (x - y)^2/2$. Considering multivariate observations $\boldsymbol{X}_i \in \mathbb{R}^d$ is also possible and one obtains for $d = 2, \rho = 2, h(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{1}\{(x^{(1)} - y^{(1)}) \cdot (x^{(2)} - y^{(2)}) > 0\}$ the empirical Kendall's τ to estimate the (monotone) dependence coefficient $\tau(\boldsymbol{X}) = 2\mathbb{P}((X^{(1)} - Y^{(1)}) \cdot (X^{(2)} - Y^{(2)}) > 0) - 1$, where $\boldsymbol{X}, \boldsymbol{Y} \sim \mathbb{P}_{\boldsymbol{X}_1}$. Furthermore, for $k \in \mathbb{N}_0$ the important class of kth probability weighted moment defined as

$$\beta_{\eta,k} = \mathrm{E}[M G_\eta^k(M)],$$

where $\eta = (\mu, \sigma, \gamma), M \sim \mathrm{GEV}(\eta), G_\eta := \mathbb{P}(M \le \cdot)$, for $\mu, \gamma \in \mathbb{R}, \sigma > 0$, might be estimated via a U-statistic by letting $\rho = k + 1$,

$$h(x_1, \dots, x_{k+1}) = \frac{1}{k+1} \sum_{j=1}^{k+1} \mathbb{1}\left\{ \max_{1 \le i \le k+1, i \ne j} x_i \le x_j \right\}$$

and substituting suitable block maxima for the x_i.

From now on, we will focus on the case $\rho = 2$ as for higher degrees the arguments stay the same but notation grows to be more opaque. Asymptotic normality of non-degenerate U-statistics has been established in Hoeffding (1948) by means of the so-called *Hoeffding-decomposition* in an asymptotic linear part and a degenerate part vanishing with higher order, which may be represented by

$$U_n(h) - \theta = \frac{1}{n} \sum_{i=1}^{n} \left(\mathrm{E}[h(\boldsymbol{X}_i, \boldsymbol{X}_{n+1}) \mid \boldsymbol{X}_i] - \theta \right) + o_{\mathbb{P}}\left(n^{-1/2}\right),$$

where $\boldsymbol{X}_{n+1}$ is an independent copy of $\boldsymbol{X}_1$. The line of proof consists of showing that the first part dominates the asymptotic behavior with rate $\sqrt{n}$. Since then, many refinements and improvements have been achieved; (van der Vaart, 1998, Chapter 12), (Lee, 2019; Koroljuk and Borovskich, 1994). Focusing on deviating from the independence assumption Sen (1972) proved a limit theorem where the serial dependence allows for $\star$-mixing, Yoshihara (1976) extended this to absolutely regular (also called β-mixing) time series. In Dehling and Wendler (2010a) this has been further extended to α-mixing time series. Strong mixing results typically require a faster decay of mixing-rates or bounded kernels since the important coupling lemma in Berbee (1979) fails, see Dehling (1983), and only L^1-distances may be bound in probability; (Bradley, 1983).

All together, the flexibility and asymptotic theory of U-statistics often allows for the straightforward construction of asymptotically valid tests or confidence intervals of point estimators. In Chapter 3 the theory of classic U-statistics is extended to the extremes setting. Recently, a similar effort has been done in Oorschot et al. (2023) who consider U-statistics where the kernel function has a growing order but eventually only depends on finitely many top order statistics. As opposed to the latter setting, in the presented article we consider the following class of estimators: Instead of X_i consider $M_{r,i}^{(\mathrm{mb})}$ for mb $\in \{\mathrm{db}, \mathrm{sb}\}$ and substitute them for the X_i in (1.1) to obtain

$$U_{n,r}^{(\mathrm{mb})} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n^{(\mathrm{mb})}} h\left(M_{r,i}^{(\mathrm{mb})}, M_{r,j}^{\mathrm{mb}}\right), \tag{1.2}$$

where h is a kernel of degree two, n^{mb} denotes the sample size of the respective block maxima sample $\mathcal{M}^{(\mathrm{mb})}$, i.e. $n^{\mathrm{mb}} = \lfloor n/r \rfloor$ for mb $=$ db and $n^{\mathrm{mb}} = n - r + 1$ for mb $=$ sb. The aim is to estimate the parameter $\theta_r = \mathrm{E}[h\left(M_{r,1}^{(\mathrm{db})}, \tilde{M}_{r,1}^{\mathrm{db}}\right)]$, where $\tilde{M}_{r,1}^{\mathrm{db}}$ denotes an independent copy of $M_{r,1}^{(\mathrm{db})}$. By the DOA assumption one has for fixed i that $M_{r,i}^{(\mathrm{mb})} \approx \mathrm{GEV}(\mu_r, \sigma_r, \gamma)$ for large n, legitimizing the heuristic $\theta_r \approx \mathrm{E}[h(M_{r,i}^{(\mathrm{mb})}, M_{r,j}^{(\mathrm{mb})})] \approx \mathrm{E}[U_{n,r}^{(\mathrm{mb})}]$ for well-separated indices $i \ll j$, since by mixing conditions one might look at $M_{r,i}^{(\mathrm{mb})}, M_{r,j}^{(\mathrm{mb})}$ as being independent of each other. It is noteworthy, that the considered U-statistics generally are not unbiased anymore. The unknown parameter of interest θ_r might have an asymptotic type-equivalent analogue θ which one is also able to estimate, think of the shape parameter γ. Hence, U-statistics of block maxima allow for flexible estimation of tail characteristics. From an abstract viewpoint this concerns U-statistics of weakly convergent triangular arrays; an area with ongoing research, see e.g. Janson (1988), Khashimov (1994), Tikhomirova and Chistyakov (2015), Mikhailov and Mezhennaya

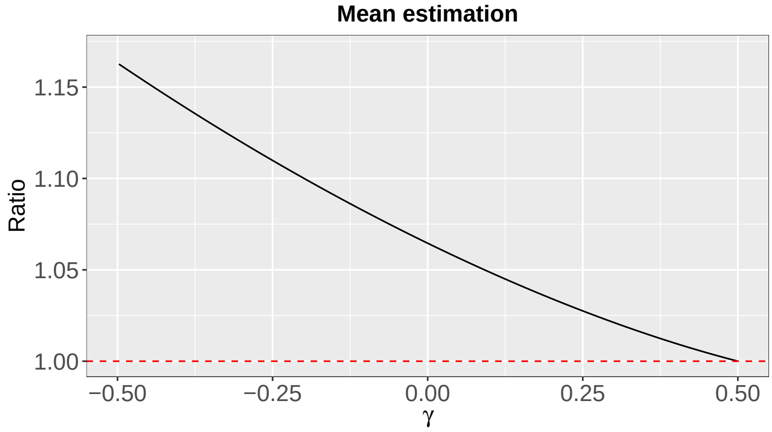

Figure 1.5: Ratio of the asymptotic variances $\sigma^2_{\mathrm{db}}/\sigma^2_{\mathrm{sb}}$ for the estimation of the block maximum mean. A benchmark red dashed line was added.

(2020), Dehling et al. (2023) among others. The analysis of such U-statistics is intricate: First, the observations might be highly dependent and second, the observations are non-stationary in n and in certain settings even in i. In the article presented we find, that U-statistics as in (1.2) are asymptotically normal under natural mild assumptions, which are standard in extreme value statistics. At the cost of stricter mixing rate assumptions we obtain these limit theorems even for $\alpha-$ instead of the more restrictive β-mixing. The results are formulated in a data adaptive way imposing a certain *kernel-transformation condition* which is satisfied in many situations. Both frameworks $\mathrm{mb} = \mathrm{db}, \mathrm{mb} = \mathrm{sb}$ allow for comparing the asymptotic variances of the respective U-statistics and in the article we show that $\mathrm{Var}(U^{(\mathrm{db})}_{n,r})/\,\mathrm{Var}(U^{(\mathrm{sb})}_{n,r}) \geq 1$ while both procedures admit the same bias, making a point to transition from disjoint block maxima to sliding block maxima usage; see e.g. Figure 1.5 for the ratio of limiting variances in the example of estimating the mean of a block maximum. These asymptotic results have been confirmed in Monte-Carlo simulation studies in the article which show that indeed sliding block based estimators perform better than their disjoint counterparts in both univariate and bivariate settings.

Analyzing deviations from the identical or stationary distribution assumptions is a highly important matter in applications as climatology; see Milly et al. (2008). Hence, we also consider a specific non-stationary situation introduced in Bücher and Zanger (2023) as *piecewise stationary* time series in which a time series is given by i.i.d. copies of time series excerpts, that is

$$(X_{n,1}, \dots X_{n,n}) = \Big(Y^{(1)}_{n,1}, \dots, Y^{(1)}_{n,r_n}, Y^{(2)}_{n,1}, \dots, Y^{(2)}_{n,r_n}, \dots, \\ Y^{(\lfloor n/r \rfloor)}_{n,1}, \dots, Y^{(\lfloor n/r \rfloor)}_{n,r_n}, Y^{(\lfloor n/r \rfloor+1)}_{n,1}, \dots, Y^{(\lfloor n/r \rfloor+1)}_{n,n-r\cdot\lfloor n/r \rfloor}\Big),$$

where $(Y^{(1)}_{n,i})_i$ is a strictly stationary time series and $(Y^{(t)}_{n,i})_i$ are i.i.d. copies of $(Y^{(1)}_{n,i})_i$ for all t. This framework is motivated by considering observation periods within each year, say summer months, concatenating the observations of each year and assuming independence between summer years as frequently done in climatology; see Philip et al. (2020) to name only one example. In Chapter 3 limiting results on U-statistics of block maxima were shown to remain valid under this new non-stationary framework showcasing that block maxima offer a versatile methodology of estimating extreme characteristics.

Significant effort has been devoted to deriving the asymptotic variance of sliding block estimators in the existing literature. Typically, these variances are of complicated form and not possible to calculate explicitly as a function of the shape γ; see e.g. Bücher and Zanger (2023). The joint asymptotic distribution of one dimensional sliding blocks can be described by a bivariate *Marshall Olkin* type extreme value distribution G_ξ with a new parameter ξ controlling the percentage of overlap. This in turn allows for expressing the rescaled asymptotic variance of linear sliding blocks estimators as

$$\sigma_{\mathrm{sb}}^2 = 2\int_0^1 \mathrm{Cov}\left(h(Z_{1,\xi}), h(Z_{2,\xi})\right) \mathrm{d}\xi,$$

where $(Z_{1,\xi}, Z_{2,\xi}) \sim G_\xi$ and $h : \mathbb{R} \to \mathbb{R}$ is a kernel. It is noteworthy that the associated extreme value copula of $(Z_{1,\xi}, Z_{2,\xi})$ is singular for $\xi \in (0,1)$ leading to explicit derivations of σ_{sb}^2 being out of reach in relevant examples, see Example 4.4 in Bücher and Jennessen (2020) among others. A universal approach avoiding the need for closed form analytic expressions of σ_{sb}^2 might be offered by resampling methods such as the *bootstrap*. Being one of the most influential statistical concepts, the bootstrap was introduced in Efron (1979). In its simplest form it works as follows: Given a sample $\boldsymbol{x} = (x_1, \dots, x_n)$ and a statistic T with $\hat{\theta} = T(\boldsymbol{x})$ to estimate θ, one draws with replacement n times from $\{x_1, \dots, x_n\}$ to obtain the bootstrap replicate $\boldsymbol{x}^* = (x_{i_1}, \dots, x_{i_n})$ which in turn yields $\hat{\theta}_n^* = T(\boldsymbol{x}^*)$. This procedure gets repeated $B \in \mathbb{N}$ times, the latter number being often determined by the computing power available to run the scheme, to obtain B bootstrap replicates of the estimator $\hat{\theta}_n^{*,[1]}, \dots, \hat{\theta}_n^{*,[B]}$. Under mild conditions these allow for inferring about the error distribution $\hat{\theta}_n - \theta$ of the estimator by analyzing the error distribution of $\hat{\theta}_n^* - \hat{\theta}_n$ instead. The latter distribution has the advantage of enabling resampling. Hence, estimator standard deviations and thus, confidence intervals –even for complicated estimators –may not be out of reach after bootstrapping. The theory about the latter is well-developed; see Efron (1982), Efron and Tibshirani (1993), Davison and Hinkley (1997) and Efron and Hastie (2016), the latter for a synthesis of modern machine learning methodology and bootstrapping.

Advantages of this method lie in both its universality but also in directly profiting from computational power which increased by a large margin over the past decades; (Moore, 1965; Schaller, 1997). Many advancements towards its usage for time series have been made; see Lahiri (2003) for an introduction. It is easy to see that modifications are necessary as otherwise temporal structure gets destroyed. A remedy lies in bootstrapping whole blocks of observations resulting in the blockwise bootstrap in Künsch (1989) or subseries approach in Carlstein (1986); see also Politis and Romano

(1994), Liu and Singh (1992), Hall et al. (1995), Bühlmann (1993). Block bootstrap methods require the choice of a well-balanced blocking parameter being essentially an unknown model parameter; (Bücher and Kojadinovic, 2016). Suboptimal choices might decrease the performance of the bootstrap significantly and several methods for optimized choices have been proposed; (Hall et al., 1995; Politis and White, 2004). Hence, a universal approach avoiding the choice of the blocking parameter is welcome and in the context of block maxima also natural.

Bootstrapping in extremes has not been well-studied. In Ferro and Segers (2003) bootstrapping functionals of extremal clusters was considered. Clusters of extreme observation appear in time series context and the inverse mean cluster size of a time series may be described by the extremal index which was introduced in Leadbetter et al. (1983), Leadbetter (1983) and originated from ideas in Loynes (1965). In Ferro and Segers (2003) an important heuristic is that intercluster exceedance times may be considered independent as clusters are nearly independent of each other under suitable mixing conditions. Nevertheless, the paper does not consider statistics of extreme observations but cluster characteristics. Most other bootstrap considerations only scratch the realm of extreme value theory: e.g. for the mean bootstrapping in the infinite-variance case (del Barrio and Matrán, 2000) where the traditional bootstrap fails, (Athreya, 1987); think of α-stable distributions or not too heavy tails. One notable exception is the recently published work de Haan and Zhou (2024) in which the authors investigate bootstrap consistency of the PWM estimators based on the POT and block maxima methods, respectively, using an asymptotic expansion of the tail quantile process which consists of a sequence of the upper order statistics. The authors were successful in proving the consistency for the POT regime, yet their results on the disjoint BM method did not imply classic desired bootstrap consistency. Despite the lack of formal validity results concerning extreme value statistical bootstrapping methods the latter are used in the literature; (Bücher and Segers, 2018a; Zanger et al., 2024) among others.

In Chapter 4 formal consistency results on bootstrapping characteristics of disjoint and sliding block maxima estimators are provided. A general framework which allows for many applications is introduced and several important examples in extreme value statistics are given. Furthermore, the results are data driven and free of oracle rates as the standardizing sequences a_r, b_r. The methods work for both multivariate block maxima and multivariate estimators. As a demonstration, the pseudo maximum likelihood estimator for the misspecified Fréchet model from Bücher and Segers (2018a) and the mean of a block maximum estimators are investigated in detail. Lastly, the methods have been applied in a case study with the result of narrower confidence bands while attaining similar coverage compared to the classic disjoint blocks based confidence intervals.

In order to establish new bootstrapping methods without the need for blocking parameters we introduce the *circular block maxima (circmax) method* which employs ideas

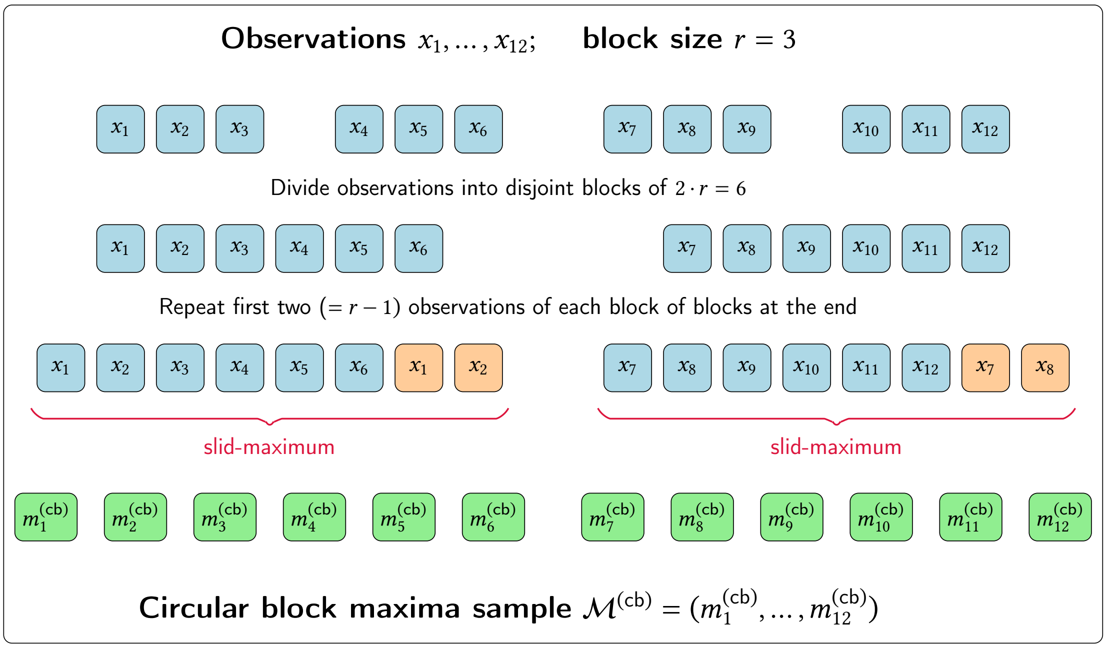

Figure 1.6: Illustration of the circular block maxima method.

from both sliding and block maxima methods as can be seen in Figure 1.6. It is possible to identify both disjoint and sliding methods as special cases of the new circmax approach. The new method offers a computationally efficient way to bootstrap sliding blocks characteristics. A new approach was necessary as we show that naive bootstrap approaches based on sliding block maxima are inconsistent. Finally, a large scale simulation and a case study support the theoretical findings by showing that confidence intervals attain comparable coverage while being more narrow than their disjoint blocks based bootstrap counterpart.

One legitimization for investigating the sliding blocks method even though the disjoint counterpart has been extensively researched lies in the fact that the asymptotic variance of estimators based on the former is less than the respective disjoint based variance. The latter fact might be heuristically explained by the fact that one considers an estimator (disjoint) summing over a subset of indices of the other (sliding). This heuristic fails, as for considering sums instead of maxima results in no advantage. The mentioned inequality can be proven by invoking Lemma A.10 from (Zou et al., 2021) with its proof relying on methods from time series analysis. Hence, in the case of one dimensional observations, it does not take into account the explicit stochastic structure of the limiting distribution of joint sliding block maxima which can be characterized by a transformed Marshall Olkin bivariate exponential distribution (BVE). In its basic form, the BVE traditionally arises from considering two random lifetimes within a two-component system where the system is determined by three independent sources of shocks, Marshall and Olkin (1967). Many characteristics of the latter distribution, such as measures of dependence like Pearson or Kendall correlation have been studied in the literature, see Lin et al. (2016) among others. Chapter 5 focuses on the *maximal correlation (coefficient)* of the BVE which is defined as

$$R = R(\mathcal{L}(X,Y)) = \sup_{f,g} \operatorname{Cor}(f(X), f(Y)),$$

where $f, g : \mathbb{R} \to \mathbb{R}$ are such that $\operatorname{Var}(f(Y_1)), \operatorname{Var}(f(Y_2)) \in (0,\infty), (X,Y)$ follow a BVE and $\mathcal{L}$ denotes the law of a random variable. Dating back to the work Gebelein (1941), the maximal correlation has been extensively studied; see (Lancaster, 1957; Sarmanov, 1958; Rényi, 1959) among others. Additionally, the coefficient has applications in numerous fields of statistics, such as optimal transport in regression (Breiman and Friedman, 1985) or Markov chains and Gibbs sampling (Liu et al., 1994). Deriving the exact value of R is an intricate problem but allows for sharp bounds of covariances. In Chapter 5 we prove a simple form of R for the two parameter BVE which formally answers an open problem, see Lin et al. (2016, Section 5, problem B). An application of this is a simple proof for $\sigma^2_{\mathrm{sb}} \leq \sigma^2_{\mathrm{db}}$, utilizing the explicit form of R, which motivates the consideration of the maximal correlation in this thesis. This creates a tangible link between sliding block maxima and the BVE and verifies the superiority of sliding blocks based estimators using the explicit limiting stochastic nature.

The rest of this thesis is structured cumulatively and organized as follows. In Chapter 2 there is a list of the included articles. Chapter 3 contains the research paper on block maxima of U-statistics. The article on bootstrapping block maxima is presented in Chapter 4. In Chapter 5 the third article on the maximal correlation of the two parameter Marshall Olkin bivariate exponential distribution is provided. Chapter 6 gives a brief outlook on potential future research extending the articles in this thesis. Finally, a statement of the authors individual contributions is deferred to the appendix.

2 Included articles

The following list contains the articles included in the thesis. They are reprinted with the permission of the respective scientific journal.

2.1 Bücher, A. and Staud, T. (2024). Limit theorems for non-degenerate U-statistics of block maxima for time series. *Electronic Journal of Statistics*, 18(2):2850-2885. (DOI: 10.1214/24-EJS2269)

2.2 Bücher, A. and Staud, T. (2024). Bootstrapping block maxima estimators for time series.
Submitted for publication and publicly available at arXiv:2409.05529.

2.3 Bücher, A. and Staud, T. (2024). On the maximal correlation coefficient for the bivariate Marshall Olkin distribution
Accepted for publication in *Statistics and Probability Letters*. Previous version publicly available at arXiv:2409.08661v2.

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

In this section we present the article Bücher and Staud (2024b) which is concerned with the asymptotic properties of U-statistics of block maxima for time series. The content of the article is reprinted with the permission of the *Electronic Journal of Statistics*. Only minor changes to improve the presentation within this thesis have been made.

Abstract

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

Keywords. Extreme Value Copula; Generalized Extreme Value Distribution; Mixing Coefficient; Sliding Block Maxima; Stationary Time Series.

MSC subject classifications. Primary 62G32, 62E20; Secondary 60G70.

A common target parameter in various domains of application is the distribution of componentwise yearly or seasonal maxima calculated from some underlying multivariate time series (Katz et al., 2002; Beirlant et al., 2004). Statistical inference on the target distribution typically involves the assumption that the block maximum distribution is an extreme value distribution. The latter is justified by probabilistic results from extreme value theory: under broad conditions on the time series, the only possible limit distribution of affinely standardized componentwise maxima, as the block size goes to infinity, are extreme value distributions; see Leadbetter (1983) for the univariate case and Hsing (1989) for multivariate extensions.

The statistical literature on estimation and testing for extreme-value distributions is abundant, ranging from univariate estimators for the parameters of the generalized extreme value distribution (Prescott and Walden, 1980; Hosking et al., 1985) to nonparametric estimators for extreme value copulas (Genest and Segers, 2009) and parametric estimators for max-stable process models (Padoan et al., 2010).

Mathematically, statistical methods are typically validated under the additional assumption that the block maxima sample is serially independent. However, heuristically, both the independence assumption as well as the assumption that block maxima genuinely follow an extreme-value distribution should only be satisfied asymptotically, for the block size tending to infinity. Dombry (2015); Ferreira and de Haan (2015); Dombry and Ferreira (2019) have shown that specific univariate estimators are consistent and asymptotically normal in a sampling scheme where the block size tends to infinity, while maintaining an i.i.d. assumption on the underlying time series. For specific univariate and multivariate estimators, Bücher and Segers (2014, 2018b) also relax the i.i.d. assumption, and allow for more general stationary time series satisfying certain mixing conditions. It has moreover been found that estimators based on block maxima may be made more efficient by considering sliding rather than disjoint block maxima, both in the univariate (Robert et al., 2009; Bücher and Segers, 2018a; Bücher and Zanger, 2023) and in the multivariate case (Zou et al., 2021). A simple but incomplete heuristic argument for this superiority consists of the fact that the sliding block maxima sample contains some observations that are not present in the disjoint block maxima sample; hence, it can be considered more informative. A formal argument for the superiority is provided in Lemma A.4 in Zou et al. (2021).

In general, the field of asymptotic statistics is based on a number of fundamental theoretical tools like the central limit theorem, the delta-method, the empirical process or the concept of U-statistics (van der Vaart, 1998). While the efficiency gain of the sliding block maxima method over its disjoint blocks counterpart mentioned in the previous paragraph has been established for classical empirical means as well as empirical (copula) processes, it has not been studied yet for the case of U-statistics. The present paper aims at filling this gap by studying non-degenerate U-statistics of disjoint and sliding block maxima samples. The topic is related but different to Oorschot et al. (2023), who study U-statistics in the univariate case where the kernel of order m is evaluated blockwise in the largest m order statistics of a (disjoint) block of observations.

In general, U-statistics comprise a number of important estimators like the empirical covariance, Wilcoxon's statistic or Kendall's tau statistic. Prominent examples from extremes are empirical probability weighted moments of order $k \geq 2$ in the univariate case (which give rise to the probability weighted moment estimator for the parameters of a Generalized Extreme Value distribution (Hosking et al., 1985)), or sample versions of Kendall's tau and Spearman's rho in the bivariate case (Beirlant et al., 2004, Page 274-275). Mathematical theory for i.i.d. random variables dates back to Hoeffding (1948); since then, several favorable statistical properties have been demonstrated (van der Vaart, 1998, Chapter 12). Asymptotic results on U-statistics have also been generalized to the time series context (Sen, 1972; Yoshihara, 1976; Dehling and Wendler, 2010a); unbiasedness then only holds asymptotically.

The main result of this paper is Theorem 3.2.5, where we establish a central limit theorem on the estimation error of U-statistics for multivariate disjoint and sliding block maxima under mild assumptions on the serial dependence and the kernel function. As in the papers mentioned before, the disjoint blocks version is found to be at most as efficient as the sliding blocks version. In selected examples, it is in fact found to be less efficient. The results are extended to a sampling scheme involving piecewise stationarities which is used to capture certain applications from environmental extremes where maxima are calculated based on, for example, summer days (Bücher and Zanger, 2023). The model is interesting mathematically, because unlike the disjoint block maxima sample the sliding block maxima sample is not stationary anymore.

The remaining parts of this paper are organized as follows: the underlying model assumptions and the definition of respective U-Statistics for disjoint and sliding block maxima are presented in Section 3.1. The main limit results are discussed in 3.2, and illustrated for three selected examples in Section 3.3. Extensions to piecewise stationary time series are presented in Section 3.4. Results from a Monte Carlo simulation study illustrate the behavior in finite-sample situations (Section 3.5). Finally, the proofs are deferred to Sections 3.6. Additional limit results under strong mixing assumptions and lengthy calculations of some asymptotic variances are postponed to a supplement.

3.1 U-statistics of block maxima

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

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

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

An extension of the classical extremal types theorem to strictly stationary time series (Leadbetter, 1983) implies that, under suitable broad conditions, affinely standardized maxima extracted from a stationary time series converge to the GEV-distribution. This was generalized to the multivariate case in Hsing (1989), where the marginals are necessarily GEV-distributed. We make this an assumption, and additionally require the scaling sequences to exhibit some common regularity inspired by the max-domain of attraction condition in the i.i.d. case (de Haan and Ferreira, 2006).

Condition 3.1.1 (Multivariate Max-domain of attraction)**.** Let $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ denote a stationary time series in $\mathbb{R}^d$ with continuous margins, where $d \in \mathbb{N} = \{1,2,\dots\}$. There exist sequences $(\boldsymbol{a}_r)_r = (a_r^{(1)},\dots,a_r^{(d)})_r \subset (0,\infty)^d$, $(\boldsymbol{b}_r)_r = (b_r^{(1)},\dots,b_r^{(d)})_r \subset \mathbb{R}^d$ and $\boldsymbol{\gamma} = (\gamma^{(1)},\dots,\gamma^{(d)}) \in \mathbb{R}$, such that, for any $s > 0$ and $j \in \{1,\dots,d\}$,

$$\lim_{r\to\infty} \frac{a_{\lfloor rs \rfloor}^{(j)}}{a_r^{(j)}} = s^{\gamma^{(j)}} \qquad \lim_{r\to\infty} \frac{b_{\lfloor rs \rfloor}^{(j)} - b_r^{(j)}}{a_r^{(j)}} = \frac{s^{\gamma^{(j)}}-1}{\gamma^{(j)}}, \tag{3.1.1}$$

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

$$\boldsymbol{Z}_r = (Z_r^{(1)},\dots,Z_r^{(d)}) \longrightarrow_d \boldsymbol{Z} \sim G \tag{3.1.2}$$

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

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

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

$$C(\boldsymbol{u}) = \exp\left(-L(-\log u^{(1)}, \dots, -\log u^{(d)})\right) \tag{3.1.3}$$

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

(L1) L is homogeneous: $L(s\boldsymbol{x}) = sL(\boldsymbol{x})$ for all $s > 0$ and all $\boldsymbol{x} \in [0, \infty]^d$;

(L2) $L(\boldsymbol{e}_j) = 1$ for $j = 1, \dots, d$, where $\boldsymbol{e}_j$ denotes the j-th unit vector in $\mathbb{R}^d$;

(L3) $\max\left(x^{(1)}, \dots, x^{(d)}\right) \leq L(\boldsymbol{x}) \leq x^{(1)} + \dots + x^{(d)}$ for all $\boldsymbol{x} \in [0, \infty]^d$;

(L4) L is convex;

see, e.g., Gudendorf and Segers (2010); Pickands (1981).

Note that (3.1.1) and (3.1.2) may for instance be deduced from Leadbetter's $D(u_n)$ condition, a domain-of-attraction condition on the associated i.i.d. sequence with stationary distribution equal to that of $\boldsymbol{X}_0$ and a weak requirement on the convergence of the c.d.f. of $\boldsymbol{Z}_r$, see Theorem 10.22 in Beirlant et al. (2004).

From now on, we assume to observe $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, an excerpt from a strictly stationary d-dimensional time series $(\boldsymbol{X}_t)_t$ satisfying Condition 3.1.1 (some generalizations will be discussed in Section 3.4). For block size parameter $r \ll n$, define componentwise block maxima of size r by

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

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

The traditional block maxima method is based on applying statistical methods to the sample of disjoint block maxima. The latter is given by $\mathcal{M}_{n,r}^{(\mathrm{db})} = \left(\boldsymbol{M}_{r,i} : i \in I_n^{\mathrm{db}}\right)$, where $I_n^{\mathrm{db}} := \{(i-1)r + 1 : 1 \leq i \leq m\}$ with $m = m_n := \lfloor n/r \rfloor$. Note that m is the number of disjoint blocks of size r that fit into the sampling period. Under Condition 3.1.1, the sample of disjoint block maxima is stationary and approximately follows the multivariate extreme value distribution G.

Instead of partitioning the observation period into disjoints blocks, one may alternatively slide the blocks through the observation period, thereby taking successive maxima of only one to the right instead of r. The resulting sliding block maxima sample is given by $\mathcal{M}_{n,r}^{(\mathrm{sb})} = \left(\boldsymbol{M}_{r,i} : i \in I_n^{\mathrm{sb}}\right)$, where $I_n^{\mathrm{sb}} := \{1, \dots, n - r + 1\}$. Under Condition 3.1.1, the sliding block maxima sample is stationary as well, with approximate c.d.f. G. Hence, statistical methods that are based on estimating unknown expectations by empirical means are meaningful.

The case of classical empirical means has been treated in varying generality in Bücher and Segers (2018a); Zou et al. (2021); Bücher and Zanger (2023). It was found that estimators based on sliding block maxima are typically more efficient than their disjoint block maxima counterparts, despite the fact that the sample $\mathcal{M}_{n,r}^{(\mathrm{sb})}$ is heavily dependent over time, even if $(\boldsymbol{X}_t)_t$ is an i.i.d. sequence. In this paper we generalize these results to U-statistics of order $p \in \mathbb{N}$, with $p = 1$ corresponding to classical empirical means.

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

$$U_{n,r}^{\mathrm{mb}} := U_{n,r}^{\mathrm{mb}}(h) := \binom{n_{\mathrm{mb}}}{p}^{-1} \sum_{(i_1,\dots,i_p)\in J_n^{\mathrm{mb}}} h\big(\boldsymbol{M}_{r,i_1}, \dots, \boldsymbol{M}_{r,i_p}\big), \tag{3.1.4}$$

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

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

A standard heuristic argument suggests that, for the majority of summands in (3.1.4), the underlying block maxima can be considered as asymptotically independent. As a consequence, U_n^{mb} should be considered as an estimator for

$$\begin{aligned}\theta_r = \theta_r(h) &:= \int \cdots \int h(\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)\, \mathrm{d}\mathbb{P}_{\boldsymbol{M}_{r,1}}(x_1) \dots \mathrm{d}\mathbb{P}_{\boldsymbol{M}_{r,1}}(x_p) \\ &= \mathrm{E}[h(\tilde{\boldsymbol{M}}_{r,1}^{(1)}, \dots, \tilde{\boldsymbol{M}}_{r,1}^{(p)})],\end{aligned} \tag{3.1.5}$$

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

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

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

3.2 Limit theorems for U-statistics of block maxima

We start by introducing further conditions and notations. First, throughout the proofs we will use traditional blocking techniques relying on mixing coefficients. The latter are well-known to control the serial dependence of the underlying time series. A similar condition has been imposed in Bücher and Segers (2014), among others.

Condition 3.2.1 (Block size and serial dependence)**.** For the block size sequence $(r_n)_n$ it holds that, as $n \to \infty$,

(a) $r_n \to \infty$ and $r_n = o(n)$.

(b) There exists a sequence $(\ell_n)_n \subset \mathbb{N}$ such that $\ell_n \to \infty$, $\ell_n = o(r_n)$ and $\frac{r_n}{\ell_n}\alpha(\ell_n) = o(1)$ and $\frac{n}{r_n}\alpha(\ell_n) = o(1)$.

(c) $\left(\frac{n}{r_n}\right)^{1+\omega}\beta(r_n) = o(1)$ for some $\omega > 0$.

Here, α and β denote the alpha- and beta-mixing coefficients, see Bradley (2005) for exact definitions and basic properties. Subsequently, we often write $r = r_n$ and $\ell = \ell_n$.

The expectation and higher order moments of $h(\boldsymbol{M}_{r,i_1}, \dots, \boldsymbol{M}_{r,i_p})$ in (3.1.4) will be controlled by uniform integrablity and by relying on the convergence of rescaled block maxima from Condition 3.1.1. For that purpose, we need the kernel function h to behave well under location-scale transformations; see also Segers (2001), Chapter 5, and Oorschot et al. (2023) for a similar, slightly more restrictive assumption.

Condition 3.2.2 (Location-scale property of the kernel function)**.** There exist functions $f : (\mathbb{R}^d)^p \to (0,\infty)$, $\ell : (\mathbb{R}^d)^p \to \mathbb{R}$ such that, for all $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p, \boldsymbol{b} \in \mathbb{R}^d$ and $\boldsymbol{a} \in (0,\infty)^d$,

$$h\left(\frac{\boldsymbol{x}_1 - \boldsymbol{b}}{\boldsymbol{a}}, \dots, \frac{\boldsymbol{x}_p - \boldsymbol{b}}{\boldsymbol{a}}\right) = \frac{h(\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)}{f(\boldsymbol{a}, \boldsymbol{b})} - \ell(\boldsymbol{a}, \boldsymbol{b}), \tag{3.2.1}$$

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

Example 3.2.3. Condition 3.2.2 is met for the following kernel functions. Note that the kernels in (5) to (7) may be used to construct tests for stochastic independence; see, for instance, Leung and Drton (2018). In the current case, this corresponds to testing asymptotic independence of the coordinates of $\boldsymbol{X}_1$.

(1) The mean kernel: $h(x) = x$ with $d = 1$, $p = 1$, $f(a,b) = a$, $\ell(a,b) = b/a$.
(2) The variance kernel: $h(x,y) = (x-y)^2/2$ with $d = 1$, $p = 2$, $f(a,b) = a^2$, $\ell \equiv 0$.
(3) Gini's mean difference kernel: $h(x,y) = |x-y|/2$ with $d = 1, p = 2, f(a,b) = a, \ell \equiv 0$.
(4) The modified probability weighted moment kernel of degree $k \in \mathbb{N}$ (see also Section 3.3.2): $h_k(x_1, \dots, x_k) = \max\{x_1, \dots, x_k\}/k$ with $d = 1$, $p = k$, $f(a,b) = a$, $l(a,b) = \frac{1}{k} \cdot \frac{b}{a}$.
(5) Kendall's τ kernel: $h(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{1}\left\{(x^{(1)} - y^{(1)})(x^{(2)} - y^{(2)}) > 0\right\}$ with $d = 2$, $p = 2$, $f \equiv 1$, $\ell \equiv 0$.
(6) Spearman's ρ kernel: $h(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = 2^{-1}\sum_{\pi \in \mathrm{S}_3} \operatorname{sgn}(x_{\pi_1}^{(1)} - x_{\pi_2}^{(1)}) \operatorname{sgn}(x_{\pi_1}^{(2)} - x_{\pi_3}^{(2)})$ with $d = 2$, $p = 3$, $f \equiv 1, \ell \equiv 0$ and where S_n denotes the symmetric group of order n.
(7) Hoeffding's D kernel and Bergsma and Dassio's t^* kernel: we refer to Leung and Drton (2018) for the kernel definition, which satisfy $d = 2, f \equiv 1, \ell \equiv 0$ and $p = 4$ and $p = 5$, respectively.

From now on, for the ease of notation, we only consider the case $p = 2$ (see also Dehling and Wendler (2010a), among others). For $i \in \{1, \dots, n-r+1\}$, let

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

with $\boldsymbol{a}_r$ and $\boldsymbol{b}_r$ from Condition 3.1.1. Note that $(\boldsymbol{Z}_{r,i})_i$ is stationary with $\boldsymbol{Z}_{r,1} \rightsquigarrow G$ as $n \to \infty$. Further, under Condition 3.2.2 one has

$$h\left(\boldsymbol{M}_{r,i}, \boldsymbol{M}_{r,j}\right) = f(\boldsymbol{a}_r, \boldsymbol{b}_r)\left\{h\left(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j}\right) + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r)\right\}, \tag{3.2.2}$$

which will ultimately allow to deduce asymptotic results on $U_{n,r}^{\mathrm{mb}}$ defined in (3.1.4) from respective results on

$$U_{n,r,Z}^{\mathrm{mb}} := U_{n,r,Z}^{\mathrm{mb}}(h) := \binom{n_{\mathrm{mb}}}{2}^{-1} \sum_{(i,j)\in J_n^{\mathrm{mb}}} h\big(Z_{r,i}, Z_{r,j}\big). \tag{3.2.3}$$

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

$$\vartheta_r = \vartheta_r(h) = \int\int h(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\mathbb{P}_{Z_{r,1}}(\boldsymbol{x})\,\mathrm{d}\mathbb{P}_{Z_{r,1}}(\boldsymbol{y}) = \mathrm{E}[h(Z_{r,1}, \tilde{Z}_{r,1})] \tag{3.2.4}$$

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

$$\vartheta := \mathrm{E}[h(Z, \tilde{Z})], \tag{3.2.5}$$

under suitable integrability assumptions; here $Z, \tilde{Z} \sim G$ are independent (see Lemma 3.7.1 below). The necessary integrability condition, which will also ensure convergence of higher order moments, is as follows.

Condition 3.2.4 (Asymptotic integrability)**.** There exists a $\nu > 2/\omega$ with ω from Condition 3.2.1 such that:

(a) $\limsup_{r\to\infty} \int\int |h(\boldsymbol{x},\boldsymbol{y})|^{2+\nu}\,\mathrm{d}\mathbb{P}_{Z_{r,1}}(\boldsymbol{x})\,\mathrm{d}\mathbb{P}_{Z_{r,1}}(\boldsymbol{y}) < \infty$,

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

Note that the two moment assumptions may be understood as an asymptotic formulation of uniform moments as used in Dehling and Wendler (2010b). In many situations, the conditions are easily satisfied, see, e.g., Section 3.3. Finally, for kernels of higher order than $p = 2$, more complicated versions of this condition will be needed, see Yoshihara (1976).

Additional notation is needed to formulate the asymptotic limit results for $U_{n,r}^{\mathrm{mb}}$. Recall G from Condition 3.1.1. Let L denote the stable tail dependence function of G if $d \ge 2$, and the identity on $[0,\infty]$ if $d = 1$. For $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$ and $\xi \ge 0$, let

$$C_\xi(\boldsymbol{u},\boldsymbol{v}) = \exp\Big[-L_\xi\Big(-\log u^{(1)},\dots,-\log u^{(d)},-\log v^{(1)},\dots,-\log v^{(d)}\Big)\Big], \tag{3.2.6}$$

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

$$L_\xi(\boldsymbol{x},\boldsymbol{y}) := (\xi\wedge 1)\cdot\Big\{L(x^{(1)},\dots,x^{(d)}) + L\left(y^{(1)},\dots,y^{(d)}\right)\Big\} + (1-(\xi\wedge 1))\cdot L\left(x^{(1)}\vee y^{(1)},\dots,x^{(d)}\vee y^{(d)}\right). \tag{3.2.7}$$

As shown in Lemma 3.7.6 below, C_ξ defines a $2d$-variate extreme-value copula with stable tail dependence function L_ξ. Let G_ξ denote the $2d$-variate extreme value distribution with copula C_ξ and margins $G_{\gamma^{(1)}},\dots,G_{\gamma^{(d)}},G_{\gamma^{(1)}},\dots,G_{\gamma^{(d)}}$, i.e.,

$$G_\xi(\boldsymbol{x},\boldsymbol{y}) = C_\xi\Big\{G_{\gamma^{(1)}}(x^{(1)}),\dots,G_{\gamma^{(d)}}(x^{(d)}),G_{\gamma^{(1)}}(y^{(1)}),\dots,G_{\gamma^{(d)}}(y^{(d)})\Big\} \tag{3.2.8}$$

for $x, y \in \mathbb{R}^d$. Note that $G_\xi(x, y) = G(x)G(y)$ for $\xi > 1$. Further, G_ξ is the multivariate analogue of $G_{\alpha_0,\xi}$ in Formula (5.1) in Bücher and Segers (2018a) and, in the case $d = 1$, also appeared in Formula (13) in Bücher and Zanger (2023).

Finally, for $(Z_{1,\xi}, Z_{2,\xi}) \sim G_\xi$, let

$$\sigma^2_{\mathrm{db}} := 4\,\mathrm{Var}(h_1(Z)), \qquad \sigma^2_{\mathrm{sb}} := 8\int_0^1 \mathrm{Cov}\left(h_1(Z_{1,\xi}), h_1(Z_{2,\xi})\right)\mathrm{d}\xi. \tag{3.2.9}$$

where,

$$h_1 : \mathbb{R}^d \to \mathbb{R},\ h_1(z) := \mathrm{E}[h(z, Z)] - \vartheta \tag{3.2.10}$$

with $Z \sim G$ and ϑ from (3.2.5). The following result is the main result of this paper. Here and throughout, λ^{2d} denotes the Lebesgue measure on $\mathbb{R}^{2d}$.

Theorem 3.2.5. *Suppose Conditions 3.1.1, 3.2.1, 3.2.2 and 3.2.4 are met. Furthermore let h be λ^{2d}-a.e. continuous and bounded on compact sets. Then, for* mb $\in$ {db, sb},

$$\frac{\sqrt{m}}{f(a_r, b_r)} \cdot (U^{\mathrm{mb}}_{n,r} - \theta_r) \rightsquigarrow \mathcal{N}(0, \sigma^2_{\mathrm{mb}}),$$

with θ_r from (3.1.5) and σ^2_{mb} from (3.2.9). Moreover, $\sigma^2_{\mathrm{sb}} \le \sigma^2_{\mathrm{db}}$.

Note that, under Condition 3.2.2, $\theta_r = f(a_r, b_r)\{\vartheta_r + \ell(a_r, b_r)\}$ with ϑ_r from (3.2.4). In certain situations (in particular when $\ell = 0$ and $f \equiv$ const; see, e.g., Kendall's tau), one may be willing to regard $U^{\mathrm{mb}}_{n,r}$ as an estimator for the asymptotic analogue

$$\tilde{\vartheta}_r = f(a_r, b_r)\{\vartheta + \ell(a_r, b_r)\}. \tag{3.2.11}$$

For instance, in case of the variance kernel (see also Section 3.3.1), $\tilde{\vartheta}_r$ is the variance of the GEV(b_r, a_r, γ)-distribution, which is exactly the GEV-distribution approximating the distribution of $M_{r,1}$, see Assumption 3.1.1. Under an additional bias condition, we may deduce the following result on the estimation error.

Corollary 3.2.6. *Additionally to the assumptions made in Theorem 3.2.5, suppose that the limit $B = \lim_{n\to\infty} B_n$ exists, where*

$$B_n := \sqrt{m}(\vartheta_r - \vartheta). \tag{3.2.12}$$

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

$$\frac{\sqrt{m}}{f(a_r, b_r)} \cdot (U^{\mathrm{mb}}_{n,r} - \tilde{\vartheta}_r) \rightsquigarrow \mathcal{N}(B, \sigma^2_{\mathrm{mb}}),$$

with σ^2_{mb} from (3.2.9) and $\tilde{\vartheta}_r$ from (3.2.11).

Remark 3.2.7 (Generalizations). Using the Cramér-Wold Theorem it is possible to generalize the limit theorems to the case of joint convergence involving a finite number of kernel functions. Moreover, as mentioned before and at the cost of a more complicated notation, one might extend the results to higher kernel degrees $p \in \mathbb{N}$. Joint weak convergence then even holds for kernels of different degrees. These generalizations allow, for example, to handle the joint convergence of probability weighted moments estimators of different order, which would be needed to deduce the asymptotics of the PWM-estimator for the parameters of the GEV-distribution. Further generalizations concerning different model assumptions are worked out in Section 3.4 and Section 3.8.2 in the supplement.

Remark 3.2.8 (A bias-corrected version of the sliding blocks estimator). In view of Lemma 3.7.6 below, the block maxima $\boldsymbol{M}_{r,i}$ and $\boldsymbol{M}_{r,j}$ are asymptotically independent for $|i-j| \geq r$ and asymptotically dependent otherwise. As a consequence, the summands $h(\boldsymbol{M}_{r,i}, \boldsymbol{M}_{r,j})$ with $|i-j| < r$ induce a *dependency bias*, which suggests to replace $U_{n,r}^{\text{sb}}$ by

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

where $\tilde{J}_n^{\text{sb}} = \{(i,j) \in (I_n^{\text{sb}})^2 : j - i \geq r\}$; see also Bücher and Zanger (2023), Remark 3.1. Note that $|\tilde{J}_n^{\text{sb}}| = \binom{n_{\text{sb}}-r+1}{2} = \binom{n-2r+2}{2}$ and $|J_n^{\text{sb}} \setminus \tilde{J}_n^{\text{sb}}| = O(nr)$, which can be used to show that $\frac{\sqrt{m}}{f(\boldsymbol{a},\boldsymbol{b})}(\tilde{U}_{n,r}^{\text{sb}} - U_{n,r}^{\text{sb}}) = O_{\mathbb{P}}\big((1 + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r))m^{-1/2}\big)$ with ℓ from Condition 3.2.2. Hence, the two estimators are typically asymptotically equivalent. However, in finite-sample situations, the bias-reduction may actually be superimposed by an increase in estimation variance (see Section 3.8.4 in the supplement), whence we cannot recommend its usage in general (this is akin to the recommendation in Bücher and Zanger (2023), Section E.3).

3.3 Examples

Details are worked out for specific kernel functions of interest.

3.3.1 Variance estimation

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

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

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

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

The following result is a direct consequence of Theorem 3.2.5.

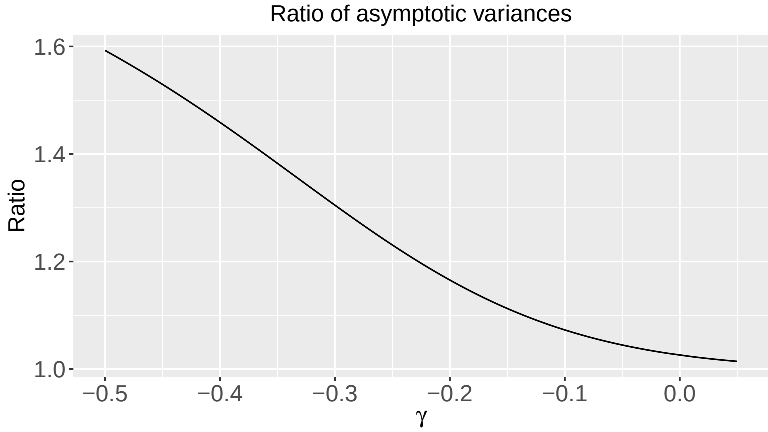

Figure 3.1: Graph of $\gamma \mapsto \sigma^2_{\mathrm{db}}/\sigma^2_{\mathrm{sb}}$ with σ^2_{mb} as in (3.8.7) and (3.8.8).

Corollary 3.3.1. *Let $d = 1$ and suppose that Conditions 3.1.1 and 3.2.1 are met with $\gamma < 1/4$. Moreover, assume that there exists a constant $\nu > 2/\omega$ such that $\limsup_r \mathbb{E}|Z_{r,1}|^{4+\nu} < \infty$. Then*

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

where σ^2_{db} and σ^2_{sb} only depend on the tail index γ. Explicit formulas are provided in (3.8.7) and (3.8.8) in the supplement, respectively. Moreover, $\sigma^2_{\mathrm{sb}} < \sigma^2_{\mathrm{db}}$.

The assumption $\gamma < 1/4$ is natural, as asymptotic normality results on empirical variances require finite fourth moments; in the case of the GEV-distribution, this exactly corresponds to $\gamma < 1/4$. Figure 3.1 shows the ratio of the asymptotic variances, $\sigma^2_{\mathrm{db}}/\sigma^2_{\mathrm{sb}}$ as a function of γ. We observe that the estimator based on sliding blocks has a significantly smaller variance for negative γ, say $\gamma < -.25$, while hardly any difference is visible for positive γ.

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

$$Y_t = \max\{\alpha Y_{t-1}, (1-\alpha)W_t\}, \quad t \in \mathbb{Z}. \tag{3.3.1}$$

The recursion has the stationary solution $Y_t := \max_{j\ge 0}(1-\alpha)\alpha^j W_{t-j}$, which has Fréchet(1) distributed marginals and extremal index $\theta = 1-\alpha$, see Example 10.5 in Beirlant et al. (2004). Define X_t as the transformed random variables $X_t := F_\gamma^{\leftarrow}(F_W(Y_t))$, where F_W is the c.d.f. of a Fréchet(1) distribution, F_γ is the c.d.f. of the Pareto family defined as

$$F_\gamma(x) := \begin{cases} \left(1-(1+\gamma x)^{-1/\gamma}\right)\mathbf{1}\{x \ge 0\}, & \gamma > 0 \\ \left(1-(1+\gamma x)^{-1/\gamma}\right)\mathbf{1}\{0 \le x \le -1/\gamma\}, & \gamma < 0\,, \\ (1-\exp(-x))\,\mathbf{1}\{x \ge 0\}, & \gamma = 0 \end{cases} \tag{3.3.2}$$

and where $F^{\leftarrow}$ is the left continuous generalized inverse of F. By Berghaus and Bücher (2018) and Bradley (2005) the untransformed time series $(Y_t)_t$ is exponentially β-mixing, which implies the same for $(X_t)_t$. This results in a large spectrum of choices for r_n and ℓ_n satisfying Condition 3.2.1, which can hence be regarded as non-restrictive. We will prove in Section 3.6.2 that, if $\gamma < 1/4$ and if $r = o(n), n = o(r^3)$, all assumptions from Corollary 3.3.1 are met, with $a_r = (r(1-\alpha))^\gamma$ and $b_r = \{(r(1-\alpha))^\gamma - 1\}/\gamma$. Hence,

$$\frac{\sqrt{m}}{(r(1-\alpha))^{2\gamma}}(\hat{\sigma}^2_{n,r,\mathrm{mb}} - \sigma_r^2) \rightsquigarrow \mathcal{N}(0, \sigma^2_{\mathrm{mb}}) \tag{3.3.3}$$

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

3.3.2 The probability weighted moment estimator

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

$$\beta_{\eta,k} := \mathrm{E}[MG_\eta^k(M)]. \tag{3.3.4}$$

It is well-known that η is a one-to-one function of the first three probability weighted moments (Hosking et al., 1985). Replacing the moments in (3.3.4) by suitable estimators and plugging those into the one-to-one function results in (the) PWM estimator for η. One version, as proposed in Landwehr et al. (1979), is given by

$$\tilde{\beta}_0 := \frac{1}{n}\sum_{i=1}^n M_i, \quad \tilde{\beta}_k := \frac{1}{n}\sum_{i=1}^n \frac{(i-1)\cdot\ldots\cdot(i-k)}{(n-1)\cdot\ldots\cdot(n-k)} M_{(i)}, \quad k \geq 1 \tag{3.3.5}$$

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

$$\hat{\beta}_{k-1} = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} h_{\mathrm{pwm},k}(M_{i_1}, \ldots, M_{i_k}) \tag{3.3.6}$$

with the permutation invariant kernel function

$$h_{\mathrm{pwm},k}(x_1, \ldots, x_k) := \frac{1}{k}\sum_{j=1}^k \mathbf{1}\Big\{\max_{1\leq i\leq k, i\neq j} x_i \leq x_j\Big\} x_j \tag{3.3.7}$$

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

In this section, we apply Theorem 3.2.5 to derive limit results for the estimator

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

with $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$. For simplicity, we restrict attention to the case $k = 2$, which yields a U-statistic of order 2. Since the function $h_{\mathrm{pwm},2}$ does not satisfy Condition 3.2.2, we will need the modified kernel function $\tilde{h}_{\mathrm{pwm},2}(x, y) := \max(x, y)/2$ from Example 3.2.3.

Proposition 3.3.2. *Let $d = 1$ and suppose that $(X_t)_{t\in\mathbb{Z}}$ satisfies Condition 3.1.1, does not contain ties with probability one, and that Condition 3.2.1 is met. If there exists $\nu > 2/\omega$ such that $\limsup_{r\to\infty} \mathbb{E}|Z_{r,1}|^{2+\nu} < \infty$, then, for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$,*

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

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

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

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

Note that similar asymptotics have also been worked out in Bücher and Zanger (2023) and Ferreira and de Haan (2015). In Bücher and Zanger (2023), who also provide explicit formulas for the asymptotic variances, the derivation was based on explicit expansions of the kernel function involving empirical cumulative distribution functions. Comparing our result with their Theorem 3.5, we observe that our result is slightly more restrictive, since we impose β-mixing rather than α-mixing. An extension to α-mixing is given in Section 3.8.2 in the supplement. Ferreira and de Haan (2015) only consider the i.i.d. case and the disjoints blocks estimator. Under this setting and using an approach based on the quantile process of the block maxima sample, they were able to provide explicit expressions for the asymptotic bias under a mild and natural second order condition; see their Theorem 2.2. They also provide an alternative representation for the asymptotic normal distribution.

3.3.3 Estimation of Kendall's tau

Kendall's tau statistic is a well-known nonparametric distribution-free measure of rank correlation that quantifies the degree of association between two variables (Kendall, 1938). The population version $\tau = \tau(\boldsymbol{X})$ for a bivariate vector $\boldsymbol{X} = (X^{(1)}, X^{(2)})$ is defined as follows: for i.i.d. copies $\boldsymbol{X}_1, \boldsymbol{X}_2$ of $\boldsymbol{X}$, we have $\tau := \pi_c - \pi_d = 2\pi_c - 1$, where $\pi_c := \mathbb{P}((X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) > 0)$ and $\pi_d := \mathbb{P}((X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) < 0)$ denote the probabilities of concordance and discordance of $\boldsymbol{X}_1, \boldsymbol{X}_2$, respectively. Applied to bivariate extreme value distributions, Kendall's tau provides a useful summary of extremal dependence; see (Beirlant et al., 2004, pp. 274-275) and the references therein.

For a bivariate sample $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ Kendall's τ-statistic can be written as $\hat{\tau}_n = \binom{n}{2}^{-1}\sum_{1\le i<j\le n}\{2h_\tau(\boldsymbol{X}_i, \boldsymbol{X}_j) - 1\}$ with h_τ as in Example 3.2.3(5). For $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$, let $\hat{\tau}_{n,r}^{\mathrm{mb}}$

denote Kendall's τ-statistic applied to the sample of disjoint or sliding block maxima. An application of Theorem 3.2.5 yields the following result.

Proposition 3.3.3. *Let $d = 2$ and suppose Conditions 3.1.1 and 3.2.1 are met. Then, with $\tau_r := \tau(M_{r,1}^{(1)}, M_{r,1}^{(2)})$, we have, for* $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$,

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

where the asymptotic variances can be represented as a function of the extreme-value copula C from (3.1.3) as follows:

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

$$\sigma_{\mathrm{sb}}^2 = 32 \int_0^1 \left(\int_{[0,1]^2 \times [0,1]^2} \big\{ C(\boldsymbol{u}) + \bar{C}(\boldsymbol{u}) \big\}\big\{ C(\boldsymbol{v}) + \bar{C}(\boldsymbol{v}) \big\} \,\mathrm{d}C_{\xi}(\boldsymbol{u}, \boldsymbol{v}) - 4\Big(\int_{[0,1]^2} C(\boldsymbol{u}) \,\mathrm{d}C(\boldsymbol{u}) \Big)^2 \right) \mathrm{d}\xi.$$

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

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

Remark 3.3.4 (Treating $\hat{\tau}_{n,r}^{\mathrm{mb}}$ as an estimator $\tau = \tau(C)$). Proposition 3.3.3 quantifies the (asymptotic) estimation error when treating $\hat{\tau}_{n,r}^{\mathrm{mb}}$ as an estimator for τ_r. In some situations one may rather be interested in treating $\hat{\tau}_{n,r}^{\mathrm{mb}}$ as an estimator for $\tau = \tau(C)$, with C the max-attractor copula from Condition 3.1.1. In that case, a bias term may show up, which can be calculated more explicitly under some suitable second-oder conditions.

It is instructive to start with the i.i.d. case: let D denote the copula of $\boldsymbol{X}_1 = (X_1^{(1)}, X_1^{(2)})$, such that the copula C_r of the block maximum distribution with block size r satisfies $C_r(u^{(1)}, u^{(2)})^{1/r} = D((u^{(1)})^{1/r}, (u^{(2)})^{1/r})$. Condition 3.1.1 implies that, for all $\boldsymbol{u} = (u^{(1)}, u^{(2)}) \in [0,1]^2$, $\lim_{r\to\infty} C_r(\boldsymbol{u}) = C(\boldsymbol{u})$. A natural second order condition (Zou et al., 2021) reads as follows: suppose there exists a function $\varphi : (0,\infty) \to (0,\infty)$ with $\lim_{r\to\infty} \varphi(r) = 0$ and a non-null-function S on $[0,1]^2$ such that

$$\lim_{r\to\infty} \frac{C_r(\boldsymbol{u}) - C(\boldsymbol{u})}{\varphi(r)} = S(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^2, \tag{3.3.8}$$

where the convergence is uniform on $[\delta, 1]^2$, for all $\delta > 0$. As shown in Zou et al. (2021), under some mild assumptions, the condition is equivalent to common second order conditions imposed on the stable tail dependence function L, and holds for selected copula families. Moreover, the convergence in (3.3.8) is necessarily uniform on $[0,1]^2$ and the function φ is regularly varying of some order $\rho \le 0$.

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

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

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

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

Overall, we obtain that

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

Hence, if the block size $r = r_n$ is chosen in such a way that the limit $\lambda_0 := \lim_{n\to\infty} \sqrt{m}\varphi(r) \geq 0$ exists, we obtain, under the conditions of Proposition 3.3.3 and in the i.i.d. case,

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

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

Treating the block size r as a tuning parameter, the previous result allows to make statements on rate-optimal choices of r. Indeed, assuming that $\varphi(r) = r^\rho$ with $\rho < 0$ for simplicity (see Zou et al. (2021) for examples), the previous equation can be restated as

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

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

Finally, note that the previous argumentation remains true in the time series case, provided we require that the convergence in (3.3.8) holds uniformly on $[0,1]^2$. However, it will typically be much harder to calculate C_r, let alone to obtain asymptotic expansions.

3.4 Extensions to piecewise stationarity

Environmental data typically involve different forms of non-stationarity. A particular source is seasonality, which may statistically be approached by restricting attention to seasons rather than years, bearing in mind that the inner-season variability should be approximately stationary. This idea may be approached mathematically by working with data satisfying the following assumption taken from Bücher and Zanger (2023).

Condition 3.4.1 (Piecewise stationary observation scheme)**.** For sample size $n \in \mathbb{N}$, we have observations $X_{n,1}, \dots, X_{n,n}$ taking values in $\mathbb{R}^d$. Moreover, for some block length sequence $(r_n)_n \subset \mathbb{N}$ diverging to infinity such that $r_n = o(n)$, we have

$$(X_{n,1}, \dots, X_{n,n}) = (Y_{1,1}, \dots, Y_{1,r_n}, Y_{2,1}, \dots, Y_{2,r_n}, \dots \\ \dots, Y_{n_{\mathrm{db}},1}, \dots, Y_{n_{\mathrm{db}},r_n}, Y_{n_{\mathrm{db}}+1,1}, \dots, Y_{n_{\mathrm{db}}+1,n-n_{\mathrm{db}}r_n}),$$

where $n_{\mathrm{db}} = \lfloor n/r_n \rfloor$ and where $(Y_{1,t})_t, (Y_{2,t})_t, \dots$ denote i.i.d. copies from a stationary time series satisfying Condition 3.1.1 with continuous marginal c.d.f. F. Note that $Y_{j,t}$ should be regarded as the t-th observation in the j-th season.

We refer to Bücher and Zanger (2023) for further discussions of Condition 3.4.1, see in particular Remark 2.3. For the rest of this section, we tacitly assume Condition 3.4.1 and write $X_j := X_{n,j}$ for simplicity. Note that the triangular array $(X_n)_n$ is r_n-dependent, which in fact simplifies the analysis of the disjoint block maxima method. For the sliding block maxima method however, mathematical challenges arise from the fact that the sliding block maxima sample is typically non-stationary. Indeed, for $x \in \mathbb{R}^d$, generally

$$\mathbb{P}\left(M_{r,1} \le x\right) \neq \mathbb{P}\left(X_2, \dots, X_r \le x\right) \cdot \mathbb{P}\left(X_{r+1} \le x\right) = \mathbb{P}\left(M_{r,2} \le x\right).$$

In Bücher and Zanger (2023), Lemma 2.4, it is shown that this non-stationarity disappears asymptotically, which suggests that statistical methodology derived under stationarity assumptions (as in Section 3.2) may also be applicable under Condition 3.4.1. For deriving respective limit results, some modifications of the previous conditions are necessary. First of all, the integrability conditions from Condition 3.2.4 take the following, slightly more involved form.

Condition 3.4.2. There exists a $\nu > 2/\omega$ with ω from Condition 3.2.1 such that

(a) $\limsup_{r\to\infty} \sup_{1\le i\le j\le r} \int\int |h(x,y)|^{2+\nu} \, \mathrm{d}\mathbb{P}_{Z_{r,i}}(x) \, \mathrm{d}\mathbb{P}_{Z_{r,j}}(y) < \infty$,

(b) $\limsup_{r\to\infty} \sup_{1\le i\le j\le r} \mathrm{E}[|h(Z_{r,i}, Z_{r,j})|^{2+\nu}] < \infty$.

It is worth noting that, if there exist monotone functions g_1, g_2 such that $|h(x,y)| \le |g_1(x)| + |g_2(y)|$, the inner supremum may be omitted; examples can be found in Section 3.3.

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

$$\vartheta_{r,i,j} := \mathrm{E}[h(Z_{r,i}, \tilde{Z}_{r,j})], \qquad \bar{\vartheta}_r := \frac{1}{r^2} \sum_{1\le i,j\le r} \vartheta_{r,i,j}, \tag{3.4.1}$$

where $(\tilde{Z}_{r,j})_{j=1,\dots,r}$ is an independent copy of $(Z_{r,j})_{j=1,\dots,r}$. Note that $\vartheta_{r,1,1} = \vartheta_r$ with ϑ_r from (3.2.4), while $\vartheta_{r,i,j} \neq \vartheta_r$ in general. We do however have $\bar{\vartheta}_r = \vartheta_r + o(1)$ under the previous conditions (see also Lemma B.5 and B.6 in Bücher and Zanger (2023) for similar results).

Lemma 3.4.3. *Suppose Conditions 3.2.2, 3.4.1 and 3.4.2(a),(b) are met and that h is λ^{2d}-a.e. continuous. Then, for $n \to \infty$,*

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

This result suggests that the non-stationarity of the sliding block maxima method under Condition 3.4.1 may show up in the asymptotic bias of the U-statistic $U_{n,r}^{\mathrm{sb}}$. The following assumption requires r to be sufficiently large to make this bias negligible.

Condition 3.4.4 (Negligibility of the bias due to non-stationarity)**.** The limit $D := \lim_{n\to\infty} D_n$ exists, where

$$D_n = \sqrt{m}\big(\bar{\vartheta}_r - \vartheta_r\big). \tag{3.4.2}$$

Theorem 3.4.5. *Within the setting of Condition 3.4.1, suppose that the block size and the underlying time series $(Y_{j,t})_t$ satisfy Condition 3.2.1(a),(b) and that the kernel satisfies Condition 3.2.2. Additionally, for* mb = db, *suppose that Condition 3.2.4(a) is met, and for* mb = sb, *suppose that Condition 3.4.2 and 3.4.4 are met. Then, if h is λ^{2d}-a.e. continuous and bounded on compact sets, we have, for $n \to \infty$,*

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

with σ_{mb}^2 from (3.2.9) satisfying $\sigma_{\mathrm{sb}}^2 \le \sigma_{\mathrm{db}}^2$. If, additionally, the limit $B = \lim_{n\to\infty} B_n$ with B_n from (3.2.12) exists, then, again for $n \to \infty$,

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

with $\tilde{\vartheta}_r$ from (3.2.11).

3.5 Simulation study

A Monte Carlo simulation study was conducted to evaluate the finite sample performance of two selected estimators based on U-statistics: the empirical variance (univariate) as well as Kendall's τ statistic (bivariate). The study mainly aims at comparing the disjoint and sliding block maxima method for various extreme value indices and time series models. The discussion is divided into two subsections, depending on the nature of the target parameter: it can either be a parameter of the block maximum distribution with some fixed r (e.g., $r = 365$ or $r = 90$), or a parameter of the max-attractor distribution from Condition 3.1.1; in that case, n should be considered fixed while r may be treated as a tuning parameter.

3.5.1 Estimating parameters of the block maximum distribution

3.5.1.1 Estimating the block maxima variance In Section 3.3.1, the empirical variance based on sliding block maxima, $\hat{\sigma}^2_{n,r,\mathrm{sb}}$, was found to be an asymptotically more efficient estimator of $\sigma_r^2 := \mathrm{Var}(M_{r,1})$ than its disjoint blocks counterpart, $\hat{\sigma}^2_{n,r,\mathrm{db}}$. We assess the performance in finite-sample situations for data-generating processes made up from the following marginal and temporal models:

Stationary distribution of X_t: We consider the generalized Pareto distribution $\mathrm{GPD}(0,1,\gamma)$ with shape parameter $\gamma \in \{-0.4, -0.2, 0, 0.1\}$, see (3.3.2). Note that the largest value of $\gamma = 0.1$ is close to the non-integrability point 0.25 for the variance estimator.

Time series models: In addition to the i.i.d. case, two time series models were considered, each with three parameter choices. The first model is the (transformed) ARMAX(1) model, see Section 3.3.1, with time series parameter $\alpha \in \{0.25, 0.5, 0.75\}$; note that the extremal index is $\theta = 1-\alpha$. As the second model we chose the Cauchy-AR model, defined as the stationary solution $(Y_t)_t$ of the Cauchy-AR recursion

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

with time series parameter $\phi \in \{0.25, 0.5, 0.75\}$. This corresponds to the extremal index $\theta = 1-\phi$, see, e.g., Problem 7.9 in Kulik and Soulier (2020). Realizations from the model were transformed to the $\mathrm{GPD}(0,1,\gamma)$ distribution by setting $X_t = F_\gamma^{\leftarrow}(F_Y(Y_t))$, where F_Y and F_γ denote the c.d.f. of the Cauchy(0,1) and the GPD(0,1,γ)-distribution, respectively.

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

The results for the i.i.d. and the ARMAX-models are illustrated in Figure 3.2, where we depict the ratio $\mathrm{MSE}(\hat{\sigma}^2_{n,r,\mathrm{db}})/\mathrm{MSE}(\hat{\sigma}^2_{n,r,\mathrm{sb}})$ as a function of the number of seasons (results for the Cauchy-AR-model are omitted because they are qualitatively similar). Across all considered numbers of seasons, tail indices and time series parameter, the sliding blocks estimator consistently outperforms its disjoint blocks counterpart. Notably, the depicted ratio is significantly larger than one for small tail indices and for small sample sizes. This particular observation is promising because obtaining large sample sizes is sometimes challenging in the area of extreme value statistics. Also, it should be noted that the serial dependence does not substantially influence the relative

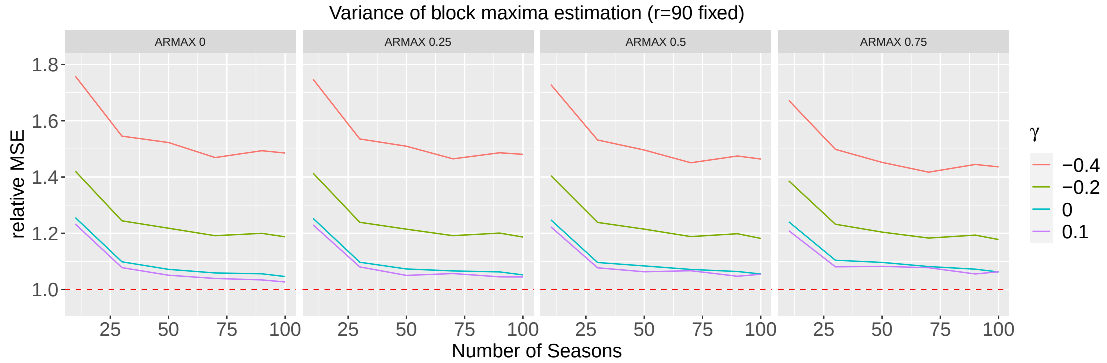

Figure 3.2: For the estimation of $\sigma_r^2 = \mathrm{Var}(M_{r,1})$, the ratio $\mathrm{MSE}(\hat\sigma^2_{n,r,\mathrm{db}})/\mathrm{MSE}(\hat\sigma^2_{n,r,\mathrm{sb}})$ is depicted as a function of number of blocks m.

performance (as was to be expected from the asymptotic results). Finally, we would like to report that the estimation variance was found to be of much larger order than the bias, whence the MSE-ratio is nearly the same as the respective variance ratio. This is different when the target variable is a parameter of the max-attractor distribution G from Condition 3.1.1, as will be discussed in Section 3.5.2 below.

3.5.1.2 Estimating Kendall's tau We investigate the finite-sample performance in the bivariate case for the estimation of Kendall's $\tau = \tau_r = \tau(M_{r,1}^{(1)}, M_{r,1}^{(2)})$ based on the estimators $\hat\tau_{n,r}^{\mathrm{db}}$ and $\hat\tau_{n,r}^{\mathrm{sb}}$ from Section 3.3.3. Note that both Kendall's τ and its estimators do not depend on the marginal distributions (in case they are continuous). The data generating processes are as follows:

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

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

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

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

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

Copula of $\boldsymbol{W}_t$: Seven different copulas were considered: the independence copula, the Gaussian copula, the t_ν-copula with $\nu = 4$ degrees of freedom, and the Gumbel-Hougard copula, where the parameter of three last-named copulas was chosen in such a way

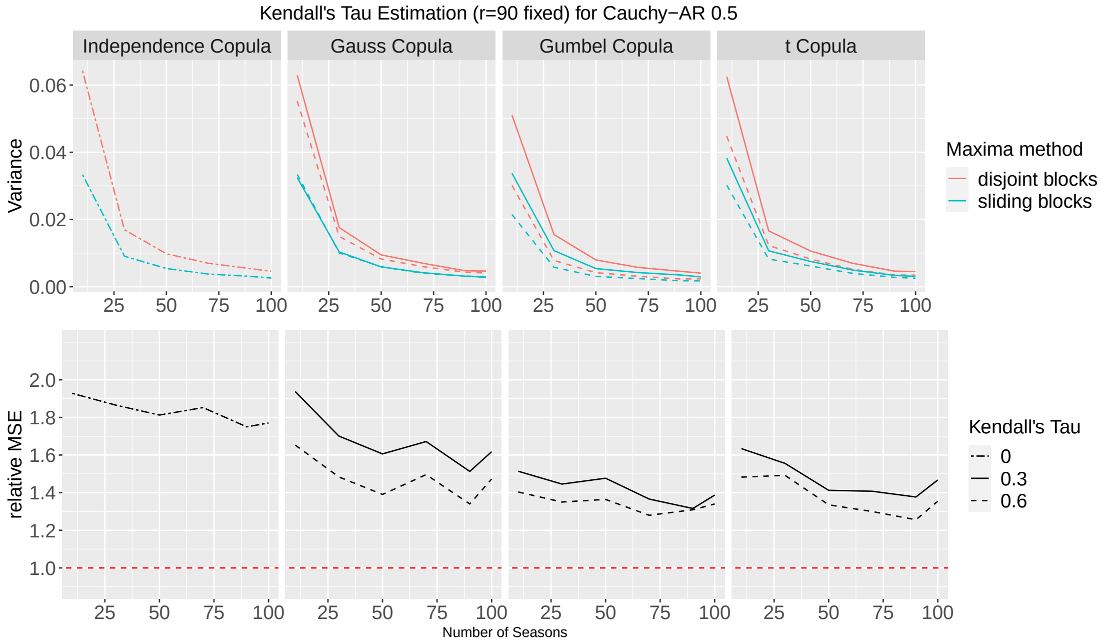

Figure 3.3: MSE of $\hat{\tau}_{n,r}^{\text{mb}}$ (upper panel) and MSE ratio $\text{MSE}(\hat{\tau}_{n,r}^{\text{db}})/\text{MSE}(\hat{\tau}_{n,r}^{\text{sb}})$ (lower panel) plotted against the number of blocks m.

that the associated value of Kendall's tau is in $\{0.3, 0.6\}$. Note that the Gaussian copula is tail independent, while the t- and Gumbel copula exhibit upper tail dependence. The upper tail dependence coefficients as a function of Kendall's tau are given by $2 \cdot t_5(-\sqrt{5(1-\sin(\pi\tau/2))/(1+\sin(\pi\tau/2)}) \in \{0.23, 0.5\}$ and $2 - 2^{1-\tau} \in \{0.375, 0.68\}$ for the t_4 and Gumbel-Hougard copula, respectively; see Embrechts et al. (2001).

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

The results are presented in Figure 3.3, where we restrict attention to the Cauchy-AR(1) model, as the performance in the other two time series is nearly identical. As in the previous section, the bias was found to be of much smaller order than the variance, whence we further restrict attention to $\text{MSE}(\hat{\tau}_{n,r}^{\text{db}})$ and to the MSE ratio $\text{MSE}(\hat{\tau}_{n,r}^{\text{db}})/\text{MSE}(\hat{\tau}_{n,r}^{\text{sb}})$. We observe that the sliding blocks estimator consistently outperforms the disjoint blocks counterpart. The level of dependence impacts the performance in that the estimation is more precise for higher dependence (for both estimators), and in that the advantage of the sliding blocks estimator over its disjoint blocks counterpart is highest for low levels of dependence/independence. Furthermore, as in the previous section, the sliding blocks estimator's advantage is slowly decreasing in the number of blocks.

3.5.2 Estimating parameters of the max-attractor distribution

If the target variable is a parameter of the max-attractor distribution, the total sample size n can be considered fixed, with the block size r serving as a tuning parameter to be chosen by the statistician. The arguments from Remark 3.3.4 then suggest that a bias-variance trade-off shows up: a large block size should correspond to small bias and large variance, and vice versa.

We restrict attention to the estimators $\hat{\tau}_{n,r}^{\mathrm{db}}$ and $\hat{\tau}_{n,r}^{\mathrm{sb}}$ from Section 3.3.3, now considered as estimators for $\tau = \tau(Z^{(1)}, Z^{(2)})$ with $(Z^{(1)}, Z^{(2)})$ from Condition 3.1.1. We chose to fix the sample size to $n = 1000$, and consider various choices for the block length between $r = 4$ and $r = 50$. For simplicity, we restrict attention to the iid case, where explicit calculations of $\tau = \tau(C)$ are possible such that we do not need to rely on preliminary Monte Carlo approximations. Regarding the copula D of $\boldsymbol{X}_i = (X_i^{(1)}, X_i^{(2)})$, we chose to work with to the outer power transform of the Clayton copula (see Bücher and Segers (2014), Formula (4.5)) with parameter θ fixed to $\theta = 1$ and with parameter β chosen in such a way that Kendall's tau $\tau = \tau(C)$ of the extreme value attractor C (which is the Gumbel-Hougaard copula with parameter β) varies in $\{0.35, 0.4, 0.45, 0.5\}$.

The results, based on $N = 1000$ simulation runs each, are summarized in Figure 3.4. As expected, the squared bias is an increasing function of m, while the variance is decreasing. In fact, the curves agree with the theoretical results from Remark 3.3.4 and Section 3.1 in Zou et al. (2021), where it is shown that φ from (3.3.8) can be chosen as r^{-1} for i.i.d. samples from the outer power Clayton copula. As a consequence, the squared bias is of the theoretical order $r^{-2} = (m/n)^2$, while the variance is of the order m^{-2}. The sum of the two rates corresponds to the MSE, which is a u-shaped function of m with minimal value attained at a point proportional to $m = n^{2/3}$, which in our case is $n = 100$. This is indeed close to the argmins of the MSE-curves depicted in Figure 3.4.

3.6 Proofs

Recall the definitions of θ_r, ϑ_r, σ_{mb}^2 from (3.1.5), (3.2.4) and (3.2.9), respectively.

3.6.1 Proofs for Section 3.2

Proof of Theorem 3.2.5. Recall the definition of $U_{n,r,Z}^{(\mathrm{mb})}$ in (3.2.3). By Condition 3.2.2 we have

$$\frac{U_{n,r}^{\mathrm{mb}} - \theta_r}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)} = U_{n,r,Z}^{(\mathrm{mb})} - \vartheta_r.$$

Hence it suffices to show that

$$\sqrt{m} \cdot (U_{n,r,Z}^{\mathrm{mb}} - \vartheta_r) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{mb}}^2), \tag{3.6.1}$$

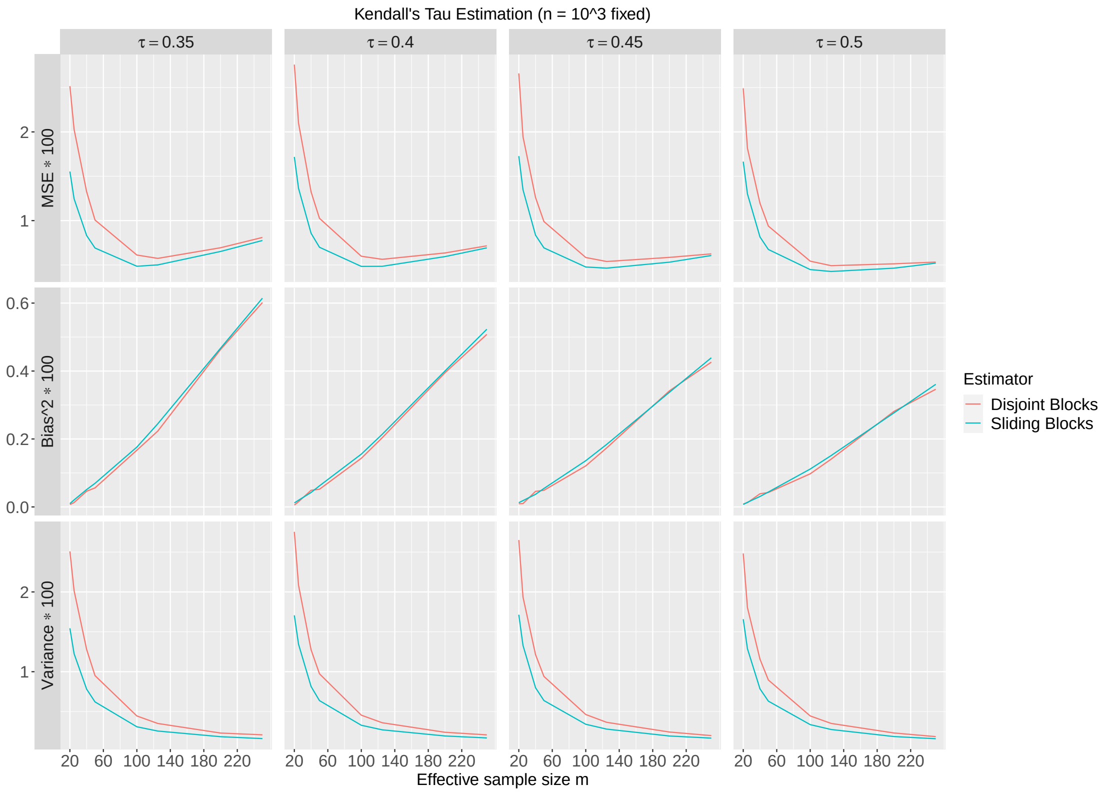

Figure 3.4: Mean squared error (top row), squared bias (middle row) and estimation variance (bottom row) when considering $\hat{\tau}_{n,r}^{\mathrm{mb}}$ as an estimator for $\tau = \tau(C)$ with fixed sample size $n = 1000$. From left to right: Kendall's tau of the attractor copula $0.35, 0.4, 0.45, 0.5$.

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

$$\sigma_{\mathrm{sb}}^2 \leq \sigma_{\mathrm{db}}^2. \tag{3.6.2}$$

For the proof of (3.6.1) we will use a Hoeffding decomposition and verify weak convergence of the linear part to the normal limit and L^2-convergence to zero of the asymptotically degenerate part. For both parts we will employ common blocking techniques to deal with the serial dependence, (see e.g., Dehling and Philipp (2002), page 31). Define

$$\begin{aligned} h_{1,r} &: \mathbb{R}^d \to \mathbb{R}, \quad \boldsymbol{x} \mapsto h_{1,r}(\boldsymbol{x}) := \mathrm{E}[h(\boldsymbol{x}, \boldsymbol{Z}_{r,1})] - \vartheta_r \qquad (3.6.3)\\ h_{2,r} &: \mathbb{R}^{2d} \to \mathbb{R} \quad (\boldsymbol{x}, \boldsymbol{y}) \mapsto h_{2,r}(\boldsymbol{x}, \boldsymbol{y}) := h(\boldsymbol{x}, \boldsymbol{y}) - h_{1,r}(\boldsymbol{x}) - h_{1,r}(\boldsymbol{y}) - \vartheta_r \end{aligned}$$

and notice the algebraic identity

$$\begin{aligned} U_{n,r}^{\mathrm{mb}} - \vartheta_r &= \frac{2}{n_{\mathrm{mb}}} \sum_{i \in I_n^{\mathrm{mb}}} h_{1,r}(\boldsymbol{Z}_{r,i}) + \frac{2}{n_{\mathrm{mb}} \cdot (n_{\mathrm{mb}} - 1)} \sum_{(i,j) \in J_n^{\mathrm{mb}}} h_{2,r}(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j}) \\ &\equiv L_{n,r}^{\mathrm{mb}} + D_{n,r}^{\mathrm{mb}}. \end{aligned} \tag{3.6.4}$$

Proof of (3.6.1) for mb = db: we start by proving

$$\sqrt{m} \cdot L_{n,r}^{\mathrm{db}} = \frac{2}{\sqrt{m}} \sum_{i \in I_n^{\mathrm{db}}} h_{1,r}(Z_{r,i}) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{db}}^2). \tag{3.6.5}$$

By Lemma 3.7.5, we may switch to i.i.d. copies of $Z_{r,i}$. The assertion then follows from Ljapunov's central limit theorem, with the Ljapunov Condition being a straightforward consequence of Lemma 3.7.2 and Condition 3.2.4.

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

$$\frac{m}{n_{\mathrm{db}}^4} \sum_{\substack{(i_1,i_2) \in J_n^{\mathrm{db}} \\ (j_1,j_2) \in J_n^{\mathrm{db}}}} \mathrm{E}[h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})] = o(1). \tag{3.6.6}$$

The Cauchy-Schwarz inequality, standard inequalities for the expectation and Condition 3.2.4 imply that

$$\begin{aligned} \sup_{\substack{(i_1,i_2) \in J_n^{\mathrm{db}} \\ (j_1,j_2) \in J_n^{\mathrm{db}}}} \left| \mathrm{E}[h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})] \right| &\le \sup_{s \in \mathbb{N}} \mathrm{E}[h_{2,r}^2(Z_{r,1}, Z_{r,1+s})] \\ &= O(1). \end{aligned}$$

Consider a tuple $(\boldsymbol{i}, \boldsymbol{j}) = (i_1, i_2, j_1, j_2) \in \{(i_1, i_2) \in J_n^{\mathrm{db}}, (j_1, j_2) \in J_n^{\mathrm{db}}\}$ such that both the distance between the smallest, $\min(\boldsymbol{i}, \boldsymbol{j})$, and the second smallest index and the largest, $\max(\boldsymbol{i}, \boldsymbol{j})$, and the second largest index is at most $2r$. Clearly, the cardinality of the set of all those $(\boldsymbol{i}, \boldsymbol{j})$ is of the order $O(m^2)$, whence the expression in (3.6.6) with the sum restricted to those tuples is of the order $O(m^{-1})$. It is hence sufficient to consider the sum over those summands for which either the distance between the smallest index and all other indices is strictly larger than $2r$, or the distance between the largest index and all other indices is strictly larger than $2r$. We only consider the first case, as the other can be treated similarly. Without loss of generality, let i_1 be the smallest index, and let $\mathcal{J}_{n,\mathrm{db}}$ = denote the respective set of indices, that is, $\mathcal{J}_{n,\mathrm{db}} = \{(\boldsymbol{i}, \boldsymbol{j}) \in J_n^{\mathrm{db}} \times J_n^{\mathrm{db}} : i_2 - i_1 > 2r, j_1 - i_1 > 2r\}$.

For each tuple $(\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{J}_{n,\mathrm{db}}$, we may use Berbee's coupling Lemma (Berbee, 1979) to construct a random variable Z_{r,i_1}^* having the same distribution as Z_{r,i_1} that is independent of $(Z_{r,i_2}, Z_{r,j_1}, Z_{r,j_2})$ and which satisfies $\mathbb{P}(Z_{r,i_1} \neq Z_{r,i_1}^*) \le \beta(\sigma(Z_{r,i_1}), \sigma(Z_{r,i_2}, Z_{r,j_1}, Z_{r,j_2})) \le \beta(r)$ where $\sigma(X)$ denotes the initial σ-field of X. Now decompose

$$\begin{aligned} &\mathrm{E}[h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})] \\ &= \mathrm{E}[\left\{ h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) - h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2}) \right\} h_{2,r}(Z_{r,j_1}, Z_{r,j_2})] \\ &\qquad + \mathrm{E}[h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})] =: I_{1,n}^{(\boldsymbol{i},\boldsymbol{j})} + I_{2,n}^{(\boldsymbol{i},\boldsymbol{j})}. \end{aligned}$$

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

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

$$\begin{aligned}|I_{1,n}^{(\boldsymbol{i},\boldsymbol{j})}| &\le \mathrm{E}\Big[\mathbf{1}\left\{Z_{r,i_1} \neq Z_{r,i_1}^*\right\} \big|\left\{h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) - h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2})\right\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times h_{2,r}(Z_{r,j_1}, Z_{r,j_2})\big|\Big] \\ &\lesssim \beta(r)^{\nu/(2+\nu)},\end{aligned}$$

where we have used Condition 3.2.4. Overall,

$$\begin{aligned}\frac{m}{n_{\mathrm{db}}^4} \sum_{(\boldsymbol{i},\boldsymbol{j})\in\mathcal{J}_{n,\mathrm{db}}} |\mathrm{E}[h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})]| &\lesssim m \cdot \sup_{(\boldsymbol{i},\boldsymbol{j})\in I} |I_{1,n}^{(\boldsymbol{i},\boldsymbol{j})} + I_{2,n}^{(\boldsymbol{i},\boldsymbol{j})}| \\ &\lesssim \left(m^{1+2/\nu}\beta(r)\right)^{\nu/(2+\nu)}\end{aligned}$$

which converges to zero by Condition 3.2.1 (c) as $2/\nu < \omega$. This implies (3.6.6), and in combination with (3.6.4) and (3.6.5) we obtain (3.6.1).

Proof of (3.6.1) for mb = sb: In order to show that the degenerate part of the rescaled sliding blocks U-statistic converges to zero, it is sufficient to show that

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

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

It remains to show

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

For this purpose use Theorem 3.7.7 with $f_{r,s} := h_{1,r}$, $f := h_1$ and note that all conditions are satisfied, where we use Lemma 3.7.6 and an easy adaptation of Lemma B.15 in Bücher and Zanger (2023) to obtain the weak convergence condition in (3.7.3).

Proof of (3.6.2): The inequality follows from Lemma A.10 in Zou et al. (2021), where $X_{n,i} := h_{1,r}(Z_{r,i})$ and the preconditions of Lemma A.10 can be deduced from Condition 3.2.1(a), (c) and 3.2.4(a). □

Proof of Corollary 3.2.6. By Condition 3.2.2 and the assumption on B_n, we have

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

Hence, the assertion follows from Theorem 3.2.5 and Slutsky's theorem. □

3.6.2 Proofs for Section 3.3

Proof of Corollary 3.3.1. Note that $|h_{\mathrm{Var}}(x,y)|^{2+\nu/2} \le 2^{2+\nu/2}(|x|^{4+\nu}+|y|^{4+\nu})$. Hence, by assumption, Condition 3.2.4 is met for a $\nu>0$ as in the formulation of Theorem 3.2.5. The statement follows by the continuity of h_{Var} and Example 3.2.3. □

Proof of Equation (3.3.3). Fix $\gamma < 1/4$ and omit the lower index 1 everywhere; e.g., write Z_r instead of $Z_{r,1}$. We need to verify the conditions of Corollary 3.3.1.

We start by proving Condition 3.1.1, for which we restrict attention to the case $\gamma > 0$ since the other cases can be treated similarly. Using $F_W^{\leftarrow}(F_\gamma(t)) = -1/\log\{1-(1+\gamma t)^{-1/\gamma}\}$ for $t>0$ and equation (10.5) from Beirlant et al. (2004) we have

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

where we substituted $a_r = (r(1-\alpha))^\gamma, b_r = \{(r(1-\alpha))^\gamma - 1\}/\gamma$. Hence Condition 3.1.1 is satisfied.

Condition 3.2.1(a) holds by assumption. Conditions (b) and (c) hold since $r=o(n^3)$ and since there exists $c>0$ with $\alpha(k)\le\beta(k)\le\exp(-ck)$ for $k\in\mathbb{N}$ by the discussion in Section 3.3.1. Note that (c) does hold for all $\omega>0$.

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

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

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

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

$$\begin{aligned}&\mathrm{E}[|\tilde Z_r|^{4+\nu}]\\ &= \frac{1-\alpha+\alpha/r}{\gamma}\int_{1/r^\gamma}^{\infty}\left(\frac{|t-1|}{\gamma}\right)^{4+\nu} t^{-1-1/\gamma}\left(1-\frac{t^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1}\mathrm{d}t\\ &\le \frac{1-\alpha+\alpha/r}{\gamma^{5+\nu}}\left\{\int_{1/r^\gamma}^{1/2} t^{-1-1/\gamma}\left(1-\frac{t^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1}\mathrm{d}t\right.\\ &\qquad\left.+\int_{1/2}^{\infty} t^{3+\nu-1/\gamma}\left(1-\frac{t^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1}\mathrm{d}t\right\} =: I_{r1}+I_{r2}\end{aligned}$$

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

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

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

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

Proof of Proposition 3.3.2. Write $h_{\mathrm{pwm},2} = h_{\mathrm{pwm}}$ and $\tilde{h}_{\mathrm{pwm},2} = \tilde{h}_{\mathrm{pwm}}$. First of all, we have

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

where

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

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

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

Next, we have

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

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

The second statement follows from Corollary 3.2.6, applied to $U_{n,r}^{\mathrm{mb}}(\tilde{h}_{\mathrm{pwm}})$. Finally, the inequality for the asymptotic variances can be found in Bücher and Zanger (2023). □

Proof of Proposition 3.3.3. Recall Example 3.2.3(5) and apply Theorem 3.2.5. A short calculation yields the formulas for the asymptotic variances. □

3.6.3 Proofs for Section 3.4

Proof of Lemma 3.4.3. For $\xi \in (0,1)$, let $\xi_r = 1 + \lfloor r\xi \rfloor$. Then,

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

By Lemma 3.7.9, we have $Z_{r,\xi_r} \rightsquigarrow Z \sim G$ for any $\xi \ge 0$. Hence, by independence and the continuous mapping theorem, $h(Z_{r,\xi_r}, \tilde{Z}_{r,\xi_r'}) \rightsquigarrow h(Z, \tilde{Z})$. Therefore, by the previous display, dominated convergence (use Condition 3.4.2(a)) and Example 2.21 in van der Vaart (1998), we obtain $\bar{\vartheta}_r = \vartheta + o(1)$. This implies the second statement, since $\vartheta_r = \vartheta + o(1)$ by Lemma 3.7.1.

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

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

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

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

Overall,

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

□

Proof of Theorem 3.4.5. For mb = db, $(Z_{r,i})_{i\in I_n^{\mathrm{db}}}$ is an i.i.d. sample. Thus the proof essentially is an easier version of the proof of Theorem 3.2.5.

For mb = sb note that, by Lemma 3.4.3, Conditions 3.4.4 and 3.2.2, it is sufficient to show that $\sqrt{m}(U_{n,r,Z}^{\mathrm{sb}} - \mathrm{E}[U_{n,r,Z}^{\mathrm{sb}}]) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{sb}}^2)$. Note that we might replace $n-r+1$ with n, since $(n-r+1)/n = 1 + O(m^{-1})$. Unlike in the situation from Theorem 3.2.5, the sliding block maxima sample is not stationary anymore, which requires a different version of the Hoeffding decomposition. For $1 \le i, j \le n$, introduce functions $h_{1,r,i} : \mathbb{R}^d \to \mathbb{R}$ and $h_{2,r,i,j} : \mathbb{R}^{2d} \to \mathbb{R}$ by

$$\begin{aligned} h_{1,r,i}(x) &:= \mathrm{E}[h\left(x, Z_{r,i}\right)] - \vartheta_{r,i,i} \\ h_{2,r,i,j}(x,y) &:= h(x,y) - h_{1,r,i}(x) - h_{1,r,j}(y) - \vartheta_{r,i,j}, \end{aligned}$$

with $\vartheta_{r,i,j}$ from (3.4.1). Note, that by Lemma 3.4.3

$$\begin{aligned} &U_{n,r,Z}^{\mathrm{sb}} - \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \mathrm{E}[h(Z_{r,i}, Z_{r,j})] \\ &= \frac{2}{n}\sum_{i=1}^{n} h_{1,r,i}(Z_{r,i}) + \frac{2}{n(n-1)} \sum_{1\le i<j\le n} h_{2,r,i,j}(Z_{r,i}, Z_{r,j}) + O(m^{-1}) \qquad (3.6.7) \\ &\equiv L_{n,r}^{\mathrm{sb}} + D_{n,r}^{\mathrm{sb}} + O(m^{-1}) \end{aligned}$$

The asymptotic normality of $\sqrt{m}L_{n,r}^{\mathrm{sb}}$ follows from Theorem 3.7.8 with $f_{r,i} := 2h_{1,r,i}$, where the preconditions are met since the time series is piecewise stationary and by assumption; moreover, (3.7.3) is a consequence of Lemma 3.7.11. We omit the proof of $\sqrt{m}D_{n,r}^{\mathrm{sb}} = o_{\mathbb{P}}(1)$ as the proof is similar to the proof of the respective statement in the proof of Theorem 3.2.5. □

3.7 Auxiliary results

3.7.1 Disjoint blocks - stationary case

Lemma 3.7.1 (Convergence of ϑ_r). *Assume Condition 3.1.1 is met. Furthermore suppose that h is λ^{2d}-a.e. continuous and that there exists $\nu > 0$ with $\limsup_{r\to\infty} \int\int |h(x,y)|^{1+\nu} \,\mathrm{d}\mathbb{P}_{Z_{r,1}}(x)\,\mathrm{d}\mathbb{P}_{Z_{r,1}}(y) < \infty$. Then $\lim_{r\to\infty} \vartheta_r = \vartheta$ with ϑ_r and ϑ from (3.2.4) and (3.2.5), respectively.*

Proof. We have $h(Z_{r,1}, \tilde{Z}_{r,1}) \rightsquigarrow h(Z, \tilde{Z})$ by independence and the continuous mapping theorem. The assertion then follows from Example 2.21 in van der Vaart (1998) and the integrability assumption. □

Recall the definition of $h_{1,r}$ from (3.6.3).

Lemma 3.7.2 (Weak convergence of $h_{1,r}(Z_{r,1})$). *Suppose Conditions 3.1.1, 3.2.4(a) hold and that h is λ^{2d}-a.e. continuous and bounded on compact sets. Then, for $r \to \infty$,*

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

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

Proof. Since $\vartheta_r \to \vartheta$ by Lemma 3.7.1, we may assume $\vartheta_r \equiv 0$. We will use Wichura's Theorem (Billingsley, 2013, Theorem 4.2). Note that

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

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

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

In order to show weak convergence of $T_r(b)$ to $T(b)$ we use the extended continuous mapping theorem (Theorem 1.11.1 in van der Vaart and Wellner (1996)). Let $x_r \to x \in \mathbb{R}^d$ and note that the map $(x, y) \mapsto h(x, y)\mathbf{1}\left\{y \in B\right\}$ is $\mathbb{P}_Z^{\otimes 2}$-a.e. continuous. By the ordinary continuous mapping theorem we obtain weak convergence of $h(x_r, Z_{r,1})\mathbf{1}\left\{Z_{r,1} \in B\right\}$ to $h(x, Z)\mathbf{1}\{Z \in B\}$. Next, since there exists a compact set A containing $(x_r)_r$, we have

$$\limsup_{r\to\infty} \mathrm{E}[h^2(x_r, Z_{r,1}) \cdot \mathbf{1}\left\{Z_{r,1} \in B\right\}] \le \sup_{x\in A, z\in B} h^2(x, z) < \infty,$$

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

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

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

$$= o(1)$$

as $b \to \infty$.

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

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

Applying the Cauchy-Schwarz inequality and taking the limit over r results in the upper bound $C \cdot \mathbb{P}(\boldsymbol{Z} \in B(b)^c)^{1/2}$ by the Portmanteau Theorem and Condition 3.2.4 (a), for a constant C not depending on b. The bound goes to 0 for $b \to \infty$ since $B(b)^c \downarrow \emptyset$.

Wichura's theorem implies weak convergence of $h_{1,r}(\boldsymbol{Z}_{r,1})$ to $h_1(\boldsymbol{Z})$, for $r \to \infty$. The stated convergence of moments follows by Example 2.21 in van der Vaart (1998) using the Jensen inequality and Condition 3.2.4 (a). □

Let $\ell = \ell_n \in \mathbb{N}$ denote the sequence from Condition 3.2.1(b). We may assume that $1 < \ell < r$. For $j \in \mathbb{N}$, recall that

$$\boldsymbol{M}_{r-\ell,j} = \max(\boldsymbol{X}_j, \dots \boldsymbol{X}_{j+r-\ell-1}), \quad \boldsymbol{Z}_{r-\ell,j} = (\boldsymbol{M}_{r-\ell,j} - \boldsymbol{b}_{r-\ell})/\boldsymbol{a}_{r-\ell}. \tag{3.7.1}$$

Lemma 3.7.3 (Weak convergence of clipped blocks)**.** *Suppose Conditions 3.1.1 and 3.2.1(a) and (b) are met. Then, as $n \to \infty$,*

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

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

Condition 3.1.1 yields local uniform convergence, see the proof of Lemma B.15 in Bücher and Zanger (2023), hence $a_r/a_{r-\ell} = 1 + o(1)$ and $(b_{r-\ell} - b_r)/a_r = o(1)$. By Lemma B.15 from Bücher and Zanger (2023) we have, for any $\varepsilon > 0$,

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

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

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

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

For the next results, let $\Delta_{r,\ell}(j) := h_{1,r}(Z_{r,j}) - h_{1,r-\ell}(Z_{r-\ell,j})$ for $j \in I_n^{\mathrm{db}}$. Furthermore let $(\tilde{X}_j, \dots, \tilde{X}_{j+r-1})_{j\in I_n^{\mathrm{db}}}$ be i.i.d. copies of $(X_j, \dots, X_{j+r-1})_{j\in I_n^{\mathrm{db}}}$ and define $\tilde{M}_{r,j} = \max(\tilde{X}_j, \dots, \tilde{X}_{j+r-1})$ and $\tilde{Z}_{r,j} = (\tilde{M}_{r,j} - b_r)/a_r$ and $\tilde{M}_{r-\ell,j}, \tilde{Z}_{r-\ell,j}$ analogously to (3.7.1).

Lemma 3.7.4. *Suppose Conditions 3.1.1,3.2.1(a), (b) and 3.2.4(a) are met and that h is λ^{2d}-a.e. continuous. Then*

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

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

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

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

by Jensen's inequality. Since $\lim_{r\to\infty}\vartheta_r = \vartheta$, we have $r_{2,n} = o(1)$ for $n \to \infty$. By Lemma 3.7.3 and independence, the vector $(Z_{r,1}, Z_{r-\ell,1}, \tilde{Z}_{r,1}, \tilde{Z}_{r-\ell,1})$ converges weakly to $(Z, Z, \tilde{Z}, \tilde{Z})$, where $Z, \tilde{Z}$ are i.i.d. with cdf G. Therefore, the continuous mapping theorem yields $|h(Z_{r,1}, \tilde{Z}_{r,1}) - h(Z_{r-\ell,1}, \tilde{Z}_{r-\ell,1})| = o_{\mathbb{P}}(1)$. By Condition 3.2.4(a) we have asymptotic uniform integrability of $|h(Z_{r,1}, \tilde{Z}_{r,1})|^{2+\nu}$ so that $r_{1,n} = o(1)$ by Example 2.21 in van der Vaart (1998) and stationarity.

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

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

By Lemma 3.11 in Dehling and Philipp (2002), Condition 3.2.4(a) and 3.2.1 (c), there exists a constant $C > 0$ that is independent of $s \ge 2$ and n such that $|\mathrm{Cov}(\Delta_{r,\ell}(1), \Delta_{r,\ell}(1 + rs))| \le C\alpha(r)^{\nu/(2+\nu)}$. Now $\alpha(r) \le \beta(r)$ and Condition 3.2.1 (c) imply that the sum in the upper display converges to 0. Hence, an application of the second claim of this lemma implies the first claim and the proof is finished. □

Lemma 3.7.5 (Restriction to independent blocks)**.** *Suppose Conditions 3.1.1, 3.2.1(a), (b), and 3.2.4(a) are met, and that h is λ^{2d}-a.e. continuous. Then $m^{-1/2}\sum_{j\in I_n^{\mathrm{db}}} h_{1,r}(Z_{r,j})$ converges weakly if and only if $m^{-1/2}\sum_{j\in I_n^{\mathrm{db}}} h_{1,r}(\tilde{Z}_{r,j})$ converges weakly. In that case the weak limits coincide.*

Proof. The result follows from a standard argument involving characteristic functions and Lemma 3.7.4; see, for instance, the proof of Theorem 3.6 in Bücher and Segers (2018b). □

3.7.2 Sliding blocks - stationary case

Recall the definitions of G_ξ, L_ξ and C_ξ from (3.2.8), (3.2.7) and (3.2.6), respectively. Recall the convention that $L := \mathrm{id}_{[0,\infty]}$ if $d = 1$, which implies $C = \mathrm{id}_{[0,1]}$. The following is a generalization of Lemma B.3 in (the supplementary material to) Bücher and Zanger (2023) for dimensions $d \geq 1$.

Lemma 3.7.6. *Suppose that Conditions 3.1.1, 3.2.1(a) and (b) are met. Then, for any $\xi \geq 0$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,*

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

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

Proof. We only consider the case $\xi \in [0, 1]$; the case $\xi > 1$ can be treated similarly. By the same arguments as in the proof of Lemma B.3 in Bücher and Zanger (2023), we obtain that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(Z_{r,1} \leq \boldsymbol{x}, Z_{r,\xi_r} \leq \boldsymbol{y}) \\
&\qquad = G\left(\xi^{-\gamma^{(1)}} x^{(1)} + \frac{\xi^{-\gamma^{(1)}} - 1}{\gamma^{(1)}}, \dots, \xi^{-\gamma^{(d)}} x^{(d)} + \frac{\xi^{-\gamma^{(d)}} - 1}{\gamma^{(d)}} \right) \\
&\qquad \times G\left(\xi^{-\gamma^{(1)}} y^{(1)} + \frac{\xi^{-\gamma^{(1)}} - 1}{\gamma^{(1)}}, \dots, \xi^{-\gamma^{(d)}} y^{(d)} + \frac{\xi^{-\gamma^{(d)}} - 1}{\gamma^{(d)}} \right) \\
&\qquad \times G\Bigg((1-\xi)^{-\gamma^{(1)}} \left(x^{(1)} \wedge y^{(1)} \right) + \frac{(1-\xi)^{-\gamma^{(1)}} - 1}{\gamma^{(1)}}, \dots \\
&\qquad\qquad \dots, (1-\xi)^{-\gamma^{(d)}} \left(x^{(d)} \wedge y^{(d)} \right) + \frac{(1-\xi)^{-\gamma^{(d)}} - 1}{\gamma^{(d)}} \Bigg). \qquad (3.7.2)
\end{aligned}$$

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

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

A straightforward calculation then shows that the expression on the right-hand side of (3.7.2) can be written as $G_\xi(\boldsymbol{x}, \boldsymbol{y})$. In particular, C_ξ is a copula, which can easily be seen to be max-stable, i.e., $C_\xi(\boldsymbol{u}^s, \boldsymbol{v}^s) = C_\xi(\boldsymbol{u}, \boldsymbol{v})^s$ for all $s > 0$ and $\boldsymbol{u}, \boldsymbol{v} \in [0, 1]^d$. It is hence an extreme-value copula with the given stable tail-dependence function L_ξ and G_ξ is the cdf of an extreme-value distribution. □

Theorem 3.7.7 (CLT for sliding blocks). *Suppose that Conditions 3.1.1, 3.2.1(a), (b) are satisfied and that there exists an $\omega > 0$ with $m_n^{1+\omega}\alpha(r_n) \to 0$. For each $r = r_n$ let $\mathcal{F}_r = \{f_{r,i} : \mathbb{R}^d \to \mathbb{R} \mid i \in \mathbb{N}\}$ be a family of deterministic maps with the following properties:*

(i) $f_{r,r+s} = f_{r,s}$ for all $s \in \mathbb{N}$ and $r = r_n$ with $n \in \mathbb{N}$;

(ii) *The random variables $f_{r,i}(Z_{r,i})$ are centered for all $i \in \mathbb{N}$ and $r = r_n$ with $n \in \mathbb{N}$;*

(iii) *There exists a $\nu > 2/\omega$ with $\limsup_{n\to\infty} \sup_{i\in\mathbb{N}} \mathrm{E}[|f_{r,i}(Z_{r,i})|^{2+\nu}] < \infty$.*

(iv) *There exists a map $f : \mathbb{R}^d \to \mathbb{R}$ such that for, all $\xi \in [0,1], \xi' \in [0,2]$, as $n \to \infty$,*

$$\left(f_{r,\xi_r}(Z_{r,\xi_r}), f_{r,\xi_r'}(Z_{r,\xi_r'})\right) \rightsquigarrow \left(f(Z_{1,|\xi-\xi'|}), f(Z_{2,|\xi-\xi'|})\right), \tag{3.7.3}$$

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

Then, for $n \to \infty$,

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

Proof. The proof is very similar to the one of Theorem 2.6 in Bücher and Segers (2018a). For completeness, it is carried out in the supplement. □

3.7.3 Sliding blocks - non-stationary case

The following theorem is an adaptation of Theorem 3.7.7 to the non-stationary setting of Section 3.4.

Theorem 3.7.8. *Suppose that the sampling scheme from Condition 3.4.1 is met and that the underlying time-series $(Y_t)_t$ satisfies Conditions 3.2.1(a), (b) and $m_n^{1+\omega}\alpha(r_n) \to 0$ for some $\omega > 0$. For each $r = r_n$, let $\mathcal{F}_r = \{f_{r,i} : \mathbb{R}^d \to \mathbb{R} \mid i \in \mathbb{N}\}$ be a family of deterministic maps satisfying Conditions (i) - (iv) of Theorem 3.7.7. Then, for $n \to \infty$,*

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

Proof. The proof is essentially the same as for Theorem 3.7.7, with the following simple adaptation: independence of $S_{n,1}^+, S_{n,2}^+, \ldots$ is a direct consequence of the imposed sampling scheme. □

The following result is an extension of Lemma 2.4 from Bücher and Zanger (2023) to multivariate time series.

Lemma 3.7.9. *Suppose the sampling scheme from Condition 3.4.1 is met. Then, for every $\xi \geq 0$ and $\boldsymbol{x} \in \mathbb{R}^d$,*

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

with G from Condition 3.1.1 and with $\xi_r := 1 + \lfloor r\xi \rfloor$.

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

$$G\left[\left(\frac{x^{(i)}}{s^{\gamma^{(i)}}} + \frac{s^{-\gamma^{(i)}}-1}{\gamma^{(i)}}\right)_{i=1,\ldots,d}\right] = C\left[\left(G_{\gamma^{(i)}}\left(\frac{x^{(i)}}{s^{\gamma^{(i)}}} + \frac{s^{-\gamma^{(i)}}-1}{\gamma^{(i)}}\right)\right)_{i=1,\ldots,d}\right]$$

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

by (3.1.3) and (L1) from Condition 3.1.1.

By piecewise stationarity and Condition 3.1.1 it suffices to show the result for $\xi \in (0,1)$. Analogous to the proof of Lemma 2.4 from Bücher and Zanger (2023) we have

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

where the last equality follows from the identity in the previous display. □

Lemma 3.7.10. *Suppose the sampling scheme from Condition 3.4.1 is met and that the underlying time series $(Y_t)_t$ satisfies Conditions 3.2.1(a) and (b). Then, for any $\xi, \xi' \ge 0$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,*

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

Proof. This is a slight adaption of the proof of Lemma 3.7.6 using standard clipping techniques and Lemma 3.7.9. □

Lemma 3.7.11. *Suppose the sampling scheme from Condition 3.4.1 is met and that the underlying time series $(Y_t)_t$ satisfies Conditions 3.2.1(a) and (b). Furthermore, let h be λ^{2d}-a.e. continuous and bounded on compact sets and suppose that Condition 3.4.2(a) is met. Then, for $\xi, \xi' \in [0,\infty)$*

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

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

Proof. We proceed similar as in the proof of Lemma 3.7.2 and employ the Cramér-Wold Theorem and Wichura's Theorem. Fix $\boldsymbol{a} = (a^{(1)}, a^{(2)}) \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$ and let

$$\begin{aligned}
T_n &:= a^{(1)} h_{1,r,\xi_r}(\boldsymbol{Z}_{r,\xi_r}) + a^{(2)} h_{1,r,\xi'_r}(\boldsymbol{Z}_{r,\xi'_r}) \\
&= \int a^{(1)} h(\boldsymbol{Z}_{r,\xi_r}, \boldsymbol{y}_1) + a^{(2)} h(\boldsymbol{Z}_{r,\xi'_r}, \boldsymbol{y}_2) \,\mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,\xi_r}} \otimes \mathbb{P}_{\boldsymbol{Z}_{r,\xi'_r}}(\boldsymbol{y}_1, \boldsymbol{y}_2), \\
T &:= a^{(1)} h_1(\boldsymbol{Z}_{1,|\xi-\xi'|}) + a^{(2)} h_1(\boldsymbol{Z}_{2,|\xi-\xi'|}) \\
&= \int a^{(1)} h(\boldsymbol{Z}_{1,|\xi-\xi'|}, \boldsymbol{y}_1) + a^{(2)} h(\boldsymbol{Z}_{2,|\xi-\xi'|}, \boldsymbol{y}_2) \,\mathrm{d}\mathbb{P}_{\boldsymbol{Z}}^{\otimes 2}(\boldsymbol{y}_1, \boldsymbol{y}_2)
\end{aligned}$$

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

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

We may now proceed analogous to the proof of Lemma 3.7.2, where we use the extended continuous mapping theorem and the weak convergence from Lemma 3.7.10 to show that $T_n(b) \rightsquigarrow T(b)$ for $n \to \infty$. □

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), DFG project 465665892, which is gratefully acknowledged. The authors are grateful to two unknown referees and an associate editor for their constructive comments that helped to improve the presentation substantially. The authors also appreciate valuable comments by multiple participants of the Extreme Value Analysis (EVA) Conference in Milano in 2023.

3.8 Supplement

In Section 3.8.1, we provide the proof of Theorem 3.7.7 from the main paper. In Section 3.8.2, we provide an extension of Theorem 3.2.5 to strong mixing. In Section 3.8.3, we provide explicit formulas for some asymptotic variances from Section 3.3 in the main paper. In Section 3.8.4, we provide some additional simulation results on the bias-corrected sliding blocks estimator.

3.8.1 Remaining proofs

Proof of Theorem 3.7.7. The proof is similar to the one of Theorem 2.6 in Bücher and Segers (2018a). For $j \in \{1, \dots, m\}$, let $I_j := \{(j-1)r+1, \dots, jr\}$. Choose $m^* = m_n^* \in \mathbb{N}$ with $3 \le m^* \le m$ such that $m^* \to \infty$ and $m^* = o(m^{\nu/(2(1+\nu))})$. Next, define $q := q_n := m/m^*$ and assume without loss of generality that $q \in \mathbb{N}, n/r \in \mathbb{N}$. For $j \in \mathbb{N}$ define $J_j^+ := I_{(j-1)m^*+1} \cup \dots \cup I_{jm^*-2}$ as the index set making up the big blocks, and $J_j^- := I_{jm^*-1} \cup I_{jm^*}$ as the index set making up the small blocks. Note that $\#J_j^+ = (m^*-2)r$ and $\#J_j^- = 2r$. The previous definitions allow to rewrite

$$\frac{\sqrt{m}}{n} \sum_{i=1}^n f_{r,i}(Z_{r,i}) = \frac{1}{\sqrt{q}} \sum_{j=1}^q (S_{n,j}^+ + S_{n,j}^-), \tag{3.8.1}$$

where $S_{n,j}^+ := \sqrt{q/(nr)} \sum_{s \in J_j^+} f_{r,s}(Z_{r,s})$ and $S_{n,j}^- := \sqrt{q/(nr)} \sum_{s \in J_j^-} f_{r,s}(Z_{r,s})$.

Note that the random variables $(S_{n,j}^\pm)_j$ are stationary by (i). Hence, by (ii)

$$\begin{aligned} \operatorname{Var}\Big(\frac{1}{\sqrt{q}} \sum_{j=1}^q S_{n,j}^-\Big) &= \frac{1}{q} \sum_{j=1}^q \operatorname{Var}(S_{n,j}^-) + \frac{2}{q} \sum_{1 \le i < j \le q} \operatorname{Cov}(S_{n,i}^-, S_{n,j}^-) \\ &\le 3 \operatorname{Var}(S_{n,1}^-) + 2 \sum_{k=2}^{q-1} |\operatorname{Cov}(S_{n,1}^-, S_{n,1+k}^-)| =: R_{n1} + R_{n2}. \end{aligned}$$

Properties (ii), (iii) and the definition of m^* and q yield

$$\Big(\frac{R_{n1}}{3}\Big)^{1/2} \le \sqrt{\frac{q}{nr}} \sum_{s \in J_1^-} \|f_{r,s}(Z_{r,s})\|_2 = O\Big(\sqrt{\frac{1}{m^*}}\Big) = o(1).$$

For $R_{n,2}$ note that by property (iii) and Lemma 3.11 from Dehling and Philipp (2002) we have that $\sup_{k\ge 2} |\operatorname{Cov}(S_{n,1}^-, S_{n,1+k}^-)| \lesssim (m^*)^{-1} \alpha(r)^{\nu/(2+\nu)}$. Since $m^* \ge 3$ we obtain $R_{n,2} \lesssim m(m^*)^{-2} \alpha(r)^{\nu/(2+\nu)} = o(1)$ by assumption. Therefore, in view of $\mathrm{E}[S_{n,j}^-] = 0$ for all j by (ii), we obtain $q^{-1/2} \sum_{j=1}^q S_{n,j}^- = o_{\mathbb{P}}(1)$.

Concerning the sum over $S_{n,j}^+$ note that we may assume that $S_{n,1}^+, S_{n,2}^+, \dots$ are independent by arguing as in the proof of Lemma 3.7.5, since there is a lag of r between any two big blocks. Hence, we may subsequently apply Ljapunov's central limit theorem.

We will show below that

$$\lim_{n\to\infty} \sigma_n^2 = \sigma_f^2, \qquad \sigma_n^2 := \operatorname{Var}\Big(\frac{\sqrt{m}}{n} \sum_{i=1}^n f_{r,i}\left(Z_{r,i}\right)\Big). \tag{3.8.2}$$

If $\sigma_f^2 = 0$, we immediately obtain the assertion. If $\sigma_f^2 > 0$, we obtain

$$\operatorname{Var}\Big(\sum_{j=1}^{q} S_{n,j}^+\Big) = q\Big\{\operatorname{Var}\Big(\frac{\sqrt{m}}{n}\sum_{i=1}^{n} f_{r,i}(Z_{r,i})\Big) + o(1)\Big\} = q\{\sigma_f^2 + o(1)\},$$

where the o term is due to (3.8.1) and $\operatorname{Var}(m^{-1/2}\sum_{j=1}^{q} S_{n,j}^-) = o(1)$. Moreover, by property (iii), we have $\sup_{j\in\mathbb{N}} \mathrm{E}[|S_{n,j}^+|^{2+\nu}] = O((m^*)^{1+\nu/2})$. As a consequence,

$$\frac{\sum_{j=1}^{q} \mathrm{E}[|S_{n,j}^+|^{2+\nu}]}{\{\operatorname{Var}(\sum_{j=1}^{q} S_{n,j}^+)\}^{1+\nu/2}} \lesssim \frac{q(m^*)^{1+\nu/2}}{(q(\sigma_f^2 + o(1)))^{1+\nu/2}} \lesssim \frac{(m^*)^{1+\nu}}{m^{\nu/2}} = o(1)$$

by choice of m^*. Ljapunvov's central limit theorem implies the assertion.

It remains to prove (3.8.2). For $k \in \{1,\dots,m\}$, let $A_k := \sum_{s\in I_k} f_{r,s}(Z_{r,s})$ and note that $\sum_{i=1}^{n} f_{r,i}(Z_{r,i}) = \sum_{k=1}^{m} A_k$. Here, Assumption (i) implies stationarity of $(A_k)_k$, whence

$$\sigma_n^2 = \frac{1}{nr}\operatorname{Var}\Big(\sum_{k=1}^{m} A_k\Big) = \frac{m}{nr}\{\operatorname{Var}(A_1) + 2\operatorname{Cov}(A_1, A_2)\} + \frac{R_n}{nr},$$

where $R_n := -2\operatorname{Cov}(A_1, A_2) + 2\sum_{k=2}^{m-1}(m-k)\operatorname{Cov}(A_1, A_{1+k})$. Lemma 3.11 in Dehling and Philipp (2002), together with Assumptions (ii) and (iii), implies that

$$\frac{1}{nr}|R_n| \lesssim \frac{r^2}{nr} + \frac{m^2r^2}{nr}\alpha^{\frac{\nu}{2+\nu}}(r) \lesssim \frac{1}{m} + \big(m^{1+2/\nu}\alpha(r)\big)^{\nu/(2+\nu)} = o(1),$$

by assumption. Hence $\sigma_n^2 = r^{-2}\{\operatorname{Var}(A_1) + 2\operatorname{Cov}(A_1, A_2)\} + o(1)$. Define, for $\xi, \xi' \geq 0$,

$$g_n(\xi, \xi') := \operatorname{Cov}\big(f_{r,\xi_r}(Z_{r,\xi_r}), f_{r,\xi'_r}(Z_{r,\xi'_r})\big),$$

and note that, by (3.7.3), Assumption (iii) and the continuous mapping theorem,

$$\lim_{n\to\infty} g_n(\xi, \xi') = g(\xi, \xi') := \operatorname{Cov}\big(f(Z_{1,|\xi-\xi'|}), f(Z_{2,|\xi-\xi'|})\big).$$

The dominated convergence theorem implies

$$\frac{\operatorname{Var}(A_1)}{r^2} = \int_0^1\int_0^1 g_n(\xi, \xi')\,\mathrm{d}\xi\,\mathrm{d}\xi' \to \int_0^1\int_0^1 g(\xi, \xi')\,\mathrm{d}\xi\,\mathrm{d}\xi',$$

as by (ii) and (iii) $g_n(\xi, \xi')$ may be bounded uniformly in n, ξ, ξ'. Similarly, we obtain $r^{-2}\operatorname{Cov}(A_1, A_2) \to \int_0^1\int_1^2 g(\xi, \xi')\,\mathrm{d}\xi\,\mathrm{d}\xi'$. We may finally proceed as in the proof of Lemma B.9 in the supplement of Bücher and Zanger (2023) to obtain $\lim_n \sigma_n^2 = \sigma_f^2$. □

3.8.2 Limit results under strong mixing

The proof of Theorem 3.2.5 strongly relies on Berbee's coupling Lemma, which is a coupling result for beta-mixing time series (Berbee, 1979). In the case of strong mixing, there is generally no similar coupling result that yields equality between the original and the coupling variables with high probability (Dehling, 1983). To the best of our knowledge, the strongest comparable result for alpha-mixing is due to Bradley (1983), which yields a coupling with a small L^1-distance. When deriving respective asymptotic results for U-statistics, Bradley's coupling construction makes it necessary to impose additional continuity assumptions on the kernel. Subsequently, we closely follow the concept of $\mathbb{P}$-Lipschitz continuity from Borovkova et al. (2001) which has been applied to U-statistics in the strongly mixing case in Dehling and Wendler (2010a).

As common kernels with multivariate arguments do not satisfy the following regularity conditions, we will for reasons of simplicity assume $d = 1$.

Condition 3.8.1 (Regularity of the kernel function)**.** There exists a non-negative function $g : \mathbb{R}^3 \to [0, \infty)$ that is λ^3-almost everywhere continuous and a $\kappa > 1$ such that the following two conditions are met:

(a) For all $(x_1, x_2, y) \in \mathbb{R}^3$, we have

$$|h(x_1, y) - h(x_2, y)| \leq |x_1 - x_2| g(x_1, x_2, y).$$

(b) With $\mathcal{F}_r := \{(Y_1, Y_2, Y_3) : \forall j \in \{1, 2, 3\} \exists i_j \in \mathbb{N} : \mathbb{P}_{Y_j} = \mathbb{P}_{Z_{r,i_j}}\}$, where the Y_i are random variables, we have

$$\limsup_{r\to\infty} \sup_{(Y_1,Y_2,Y_3)\in\mathcal{F}_r} \mathrm{E}[g^\kappa(Y_1, Y_2, Y_3)] < \infty.$$

The uniform integrability condition from Condition 3.2.4 must be strengthened as follows.

Condition 3.8.2 (Asymptotic integrability)**.** There exist $\nu, \rho > 0$ with $\nu > 1/(\kappa - 1) - 1$ and with κ from Condition 3.8.1 such that Conditions 3.2.4(a), (b) hold and

$$\limsup_{r\to\infty} \mathrm{E}[|Z_{r,1}|^\rho] < \infty. \tag{3.8.3}$$

In specific applications, Conditions 3.2.4(a), (b) may often be deduced from moment constraints as in (3.8.3). Hence, Condition 3.8.2 is not substantially stronger than Condition 3.2.4. Next, as we weaken the mixing requirements from absolute regularity to strong mixing we need the following stronger assumptions on the mixing rates.

Condition 3.8.3 (Block size and serial dependence)**.** The block size sequence $(r_n)_n$ satisfies Conditions 3.2.1(a) and (b). Moreover, $(n/r_n)^{3/2+2/\nu+1/(2\rho)+1/(\rho\nu)}\alpha(r_n) = o(1)$, where ρ and ν are from Condition 3.8.2.

Finally, recalling the definitions of $U_{n,r}^{\mathrm{mb}}$ in (3.1.4), of θ_r in (3.1.5), of σ_{mb}^2 in (3.2.9) and of $\tilde{\vartheta}_r$ in (3.2.11), we have the following result.

Theorem 3.8.4. *Suppose Conditions 3.2.2, 3.8.1, 3.8.2 and 3.8.3 are met. Furthermore, let h be λ^{2d}-a.e. continuous and bounded on compact sets. Then, for* mb $\in$ {db, sb}*,*

$$\frac{\sqrt{m}}{f(a_r, b_r)} \cdot (U_{n,r}^{\mathrm{mb}} - \theta_r) \longrightarrow_d \mathcal{N}(0, \sigma_{\mathrm{mb}}^2).$$

If, additionally, the limit $B = \lim_{n\to\infty} B_n$ with B_n from (3.2.12) exists, then

$$\frac{\sqrt{m}}{f(a_r, b_r)} \cdot (U_{n,r}^{\mathrm{mb}} - \tilde{\vartheta}_r) \longrightarrow_d \mathcal{N}(B, \sigma_{\mathrm{mb}}^2).$$

Proof of Theorem 3.8.4. We use the same Hoeffding decompositon as in the proof of Theorem 3.2.5; see the algebraic identity (3.6.4). Since the proof of the asymptotic normality of L_n^{mb} does not make use Condition 3.2.1(c), the proof also applies in the current setting for α-mixing. Moreover, for the disjoint blocks case, it is sufficient to show (3.6.6). By the same arguments as in the mentioned proof we may restrict attention to the sum over the indices from $\mathcal{J}_{n,\mathrm{db}} = \{(\boldsymbol{i},\boldsymbol{j}) \in J_n^{\mathrm{db}} \times J_n^{\mathrm{db}} \ : \ \min(i_2 - i_1, j_1 - i_1) > 2r\}$. By Condition 3.8.2 and Theorem 3 in Bradley (1983), after enlarging the probability space if necessary, there exist, for any $(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{J}_{n,\mathrm{db}}$, random variables Z_{r,i_1}^* with the following properties:

(i) Z_{r,i_1}^* is independent of $(Z_{r,i_2}, Z_{r,j_1}, Z_{r,j_2})$,
(ii) Z_{r,i_1} and Z_{r,i_1}^* have the same distributions,
(iii) $\forall \varepsilon > 0$: $\mathbb{P}(|Z_{r,i_1} - Z_{r,i_1}^*| \ge \varepsilon) \le K\alpha^{2\rho/(2\rho+1)}(r)/\varepsilon^{\rho/(2\rho+1)}$

where the constant K does not depend on $(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{J}_{n,\mathrm{db}}$. By the same arguments as in the proof of Theorem 3.2.5 we have, for any $\varepsilon > 0$,

$$
\begin{aligned}
&\mathrm{E}\left[h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})\right] \\
&= \mathrm{E}\left[\left\{h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) - h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2})\right\} h_{2,r}(Z_{r,j_1}, Z_{r,j_2})\right] \\
&= \mathrm{E}\Big[\left\{h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) - h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2})\right\} \\
&\qquad\qquad \times h_{2,r}(Z_{r,j_1}, Z_{r,j_2}) \mathbf{1}\{|Z_{r,i_1} - Z_{r,i_1}^*| < \varepsilon\}\Big] \\
&\quad + \mathrm{E}\Big[\left\{h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) - h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2})\right\} \\
&\qquad\qquad \times h_{2,r}(Z_{r,j_1}, Z_{r,j_2}) \mathbf{1}\{|Z_{r,i_1} - Z_{r,i_1}^*| \ge \varepsilon\}\Big] \\
&\equiv R_{1,n} + R_{2,n}.
\end{aligned} \tag{3.8.4}
$$

Note that

$$
\begin{aligned}
&h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) - h_{2,r}(Z_{r,i_1}^*, Z_{r,i_2}) \\
&\qquad = h(Z_{r,i_1}, Z_{r,i_2}) - h(Z_{r,i_1}^*, Z_{r,i_2}) + \int h(Z_{r,i_1}, y) - h(Z_{r,i_1}^*, y)\, \mathrm{d}\,\mathbb{P}_{Z_{r,1}}(y),
\end{aligned}
$$

and apply Hölder's inequality with $p = 1 + 1/(1+\nu) < \kappa$, $q = 2 + \nu$ to obtain

$$
\begin{aligned}
|R_{1,n}| \le \varepsilon\, \mathrm{E}\Bigg[&\left\{g(Z_{r,i_1}, Z_{r,i_1}^*, Z_{r,i_2}) + \int g(Z_{r,i_1}, Z_{r,i_1}^*, y)\, \mathrm{d}\,\mathbb{P}_{Z_{r,1}}(y)\right\} \\
&\times \left|h_{2,r}(Z_{r,j_1}, Z_{r,j_2})\right|\Bigg] \lesssim \varepsilon,
\end{aligned} \tag{3.8.5}
$$

by Condition 3.8.1(a), (b) and Condition 3.8.2. Next, another application of Hölder's inequality with $p = (2+\nu)/2$, $q = (2+\nu)/\nu$ yields

$$
|R_{2,n}| \lesssim \mathbb{P}\left(|Z_{r,i_1} - Z_{r,i_1}^*| \ge \varepsilon\right)^{\nu/(2+\nu)} \lesssim \frac{\alpha^{2\mu}(r)}{\varepsilon^{\mu}}, \tag{3.8.6}
$$

where $\mu = \rho\nu/\{(2\rho+1)(\nu+2)\}$. Setting $\varepsilon = \alpha^{2\mu/(\mu+1)}(r)$, we have, by (3.8.4), (3.8.5) and (3.8.6),

$$
|\,\mathrm{E}[h_{2,r}(Z_{r,i_1}, Z_{r,i_2}) h_{2,r}(Z_{r,j_1}, Z_{r,j_2})]| \lesssim \alpha^{2\mu/(\mu+1)}(r) + \alpha^{2\mu - 2\mu^2/(\mu+1)}(r)
$$

$$= 2\alpha^{(2\rho\nu)/(3\rho\nu+4\rho+\nu+2)}(r),$$

uniformly in $\boldsymbol{i}, \boldsymbol{j}$. Hence, we obtain by Condition 3.8.3

$$\frac{m}{n_{\text{db}}^4} \sum_{(\boldsymbol{i},\boldsymbol{j})\in\mathcal{J}_{n,\text{db}}} \left|\mathrm{E}[h_{2,r}(Z_{r,i_1}, Z_{r,i_2})h_{2,r}(Z_{r,j_1}, Z_{r,j_2})]\right|$$
$$\lesssim m\alpha^{(2\rho\nu)/(3\rho\nu+4\rho+\nu+2)}(r)$$
$$= \left(m^{3/2+2/\nu+1/(2\rho)+1/(\rho\nu)}\alpha(r)\right)^{(2\rho\nu)/(3\rho\nu+4\rho+\nu+2)} = o(1),$$

as $\#\mathcal{J}_{n,\text{db}} = O(m^4)$.

The sliding blocks version can be treated similarly, see also the proof of Theorem 3.2.5.

Finally, the statements concerning centering at $\tilde{\vartheta}_r$ follow by the same arguments as in the proof of Corollary 3.2.6. □

3.8.3 Formulas for the asymptotic variances in Section 3.3.1

Write $g_j := \Gamma(1 - j\gamma)$ for $\gamma < 1/j$, $j \in \mathbb{N}$ and let $\zeta(3)$ denote Apery's constant.

Lemma 3.8.5. *Let $\gamma < 1/4$. The asymptotic variances in Corollary 3.3.1 can be written as*

$$\sigma_{\text{db}}^2 = \begin{cases} \frac{4}{\gamma^4}\left(g_4 - 4g_1g_3 - g_2^2 + 8g_1^2g_2 - 4g_1^4\right), & \gamma \neq 0 \\ \frac{22}{45}\pi^4, & \gamma = 0 \end{cases} \tag{3.8.7}$$

and

$$\sigma_{\text{sb}}^2 = \begin{cases} \frac{2}{3\gamma^3}\left(-3g_4I_{2,2} + 8g_1g_3I_{2,1} - 6g_1^2g_2I_{1,1}\right), & \gamma > 0 \\ \frac{8}{\gamma^2}\left(\Gamma(-4\gamma)I_{2,2} - 2g_1\Gamma(-3\gamma)I_{2,1} + g_1^2\Gamma(-2\gamma)I_{1,1}\right), & \gamma < 0\,, \\ 2\zeta(3) - 48 - \frac{8}{3}\pi^2 + \frac{32}{3}\log^3(2) - 48\log^2(2) + 96\log(2) + \frac{16}{3}\pi^2\log(2), & \gamma = 0 \end{cases} \tag{3.8.8}$$

where

$$I_{i,k} := \int_0^{1/2} \left(\alpha_{(j+k)\gamma}(w) - 1\right)\left\{w^{-j\gamma-1}(1-w)^{-k\gamma-1} + w^{-k\gamma-1}(1-w)^{-j\gamma-1}\right\}\mathrm{d}w \tag{3.8.9}$$

and

$$\alpha_\beta : (0,1) \to (0,\infty), \quad w \mapsto \alpha_\beta(w) = \begin{cases} \frac{1-(1-w)^{\beta+1}}{w(\beta+1)}, & \beta \neq -1 \\ -\frac{\log(1-w)}{w}, & \beta = -1 \end{cases}.$$

Proof. Recall the definition of h_1 from (3.2.10). We have $h_1(z) = z^2/2 - \mu_1 z + \mu_2/2$ for the variance kernel $h(x,y) = (x-y)^2/2$, where μ_j denotes the j-th moment of a GEV(γ) distribution.

Disjoint case: Using $\sigma_{\text{db}}^2 = 4\,\mathrm{Var}(h_1(Z))$ yields

$$\sigma_{\text{db}}^2 = \mu_4 - \mu_2^2 + 4\mu_1\left(-\mu_1^3 + 2\mu_1\mu_2 - \mu_3\right). \tag{3.8.10}$$

The first four moments of a GEV(γ) distributed random variable are given by

$$\mu_1 = \frac{g_1 - 1}{\gamma}, \qquad \mu_2 = \frac{g_2 - 2g_1 + 1}{\gamma^2},$$
$$\mu_3 = \frac{g_3 - 3g_2 + 3g_1 + 1}{\gamma^3}, \qquad \mu_4 = \frac{g_4 - 4g_3 + 6g_2 - 4g_1 + 1}{\gamma^4}.$$

Plugging these into (3.8.10) gives the result for $\gamma \neq 0$. The case $\gamma = 0$ is similarly easy and hence omitted.

Sliding case: Let $C_\xi = \mathrm{Cov}(h_1(Z_{1,\xi}), h_1(Z_{2,\xi}))$. First we will consider $\gamma \neq 0$. A short calculation using the transformation $S_{i,\xi} = (1+\gamma Z_{i,\xi})^{-1/\gamma}$ gives $C_\xi = 1/4\gamma^4 \left(C_{\xi,2,2} - 4g_1 C_{\xi,2,1} + 4g_1^2 C_{\xi,1,1}\right)$, where $C_{\xi,j,k} := \mathrm{Cov}(S_{1,\xi}^{-j\gamma}, S_{2,\xi}^{-k\gamma})$ and hence

$$\sigma_{\mathrm{sb}}^2 = \frac{2}{\gamma^4}\int_0^1 (C_{\xi,2,2} - 4g_1 C_{\xi,2,1} + 4g_1^2 C_{\xi,1,1} \,\mathrm{d}\xi. \tag{3.8.11}$$

Hoeffding's Lemma will be employed to calculate $C_{\xi,j,k}$, thus we need to derive the difference of the joint and product c.d.f.s: To this end use the explicit form of G_ξ for univariate random variables, see e.g. equation (13) in Bücher and Zanger (2023),

$$G_\xi(x,y) = \exp\left[-\left\{\xi(1+\gamma x)^{-\frac{1}{\gamma}} + \xi(1+\gamma y)^{-\frac{1}{\gamma}} + (1-\xi)(1+\gamma(x\wedge y))^{-\frac{1}{\gamma}}\right\}\right]$$

for $\xi \in [0,1]$ and $G_\xi(x,y) = G_\gamma(x)G_\gamma(y)$ for $\xi > 1$ to obtain

$$\mathbb{P}(S_{1,\xi} \le s, S_{2,\xi} \le t) = 1 - e^{-s} - e^{-t} + e^{-(s+t)A_\xi(\frac{t}{t+s})}, \qquad s,t > 0,$$

where $A_\xi(w) = \xi + (1-\xi)\{w \vee (1-w)\}$. These lead to

$$\begin{aligned}&\mathbb{P}\left(S_{1,\xi}^{-j\gamma} \le s, S_{2,\xi}^{-k\gamma} \le t\right) - \mathbb{P}\left(S_{1,\xi}^{-j\gamma} \le s\right)\mathbb{P}\left(S_{2,\xi}^{-k\gamma} \le t\right)\\ &= \exp\left[(s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}})A_\xi\left(\frac{t^{-\frac{1}{k\gamma}}}{s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}}}\right)\right] - \exp\left[-\left(s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}}\right)\right],\end{aligned}$$

for $s,t > 0$, $j,k \in \{1,2\}$. Now by Hoeffding's Lemma:

$$\begin{aligned}&C_{\xi,j,k}\\ &= \int_0^\infty\int_0^\infty \exp\left[(s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}})A_\xi\left(\frac{t^{-\frac{1}{k\gamma}}}{s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}}}\right)\right] - \exp\left[-\left(s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}}\right)\right] \mathrm{d}s\,\mathrm{d}t.\end{aligned}$$

Using the substitutions $u = s^{-\frac{1}{j\gamma}} + t^{-\frac{1}{k\gamma}}$, $w = t^{-\frac{1}{k\gamma}}/(s^{-1/(j\gamma)} + t^{-1/(k\gamma)})$ we get

$$C_{\xi,j,k} = jk\gamma^2 \int_0^1\int_0^\infty \left(\mathrm{e}^{-uA_\xi(w)} - \mathrm{e}^{-u}\right) u^{-(j+k)\gamma-1}(1-w)^{-j\gamma-1}w^{-k\gamma-1}\,\mathrm{d}u\,\mathrm{d}w \tag{3.8.12}$$

Distinguish cases, first let $\gamma < 0$. Use the fact that

$$\int_0^\infty \left(\mathrm{e}^{-zt} - \mathrm{e}^{-u}\right) u^{-\beta-1}\,\mathrm{d}u = \Gamma(-\beta)\left(z^\beta - 1\right), \qquad z > 0, \beta < 0$$

to obtain

$$C_{\xi,j,k} = jk\gamma^2 \int_0^1 \Gamma(-(j+k)\gamma) \left(A_\xi^{(j+k)\gamma} - 1 \right) (1-w)^{-j\gamma-1} w^{-k\gamma-1} \, \mathrm{d}w. \tag{3.8.13}$$

Note $A_\xi(w) = A_\xi(1-w)$ for $w \in (0,1)$ and recall the definition of $\alpha_\beta(w)$. For $w \in (0,1)$ and $\beta > 0$ it then follows $\int_0^1 (\xi w + 1 - w)^\beta \, \mathrm{d}\xi = \alpha_\beta(w)$. This in conjunction with the symmetry of A_ξ and (3.8.13) yields

$$\begin{aligned} &\int_0^1 C_{\xi,j,k} \, \mathrm{d}\xi \\ &= jk\gamma^2 \Gamma(-(j+k)\gamma) \int_0^{1/2} \left(\alpha_{(j+k)\gamma}(w) - 1 \right) \left\{ w^{-j\gamma-1}(1-w)^{-k\gamma-1} \right. \\ &\qquad \left. + w^{-k\gamma-1}(1-w)^{-j\gamma-1} \right\} \mathrm{d}w. \end{aligned} \tag{3.8.14}$$

Recall the definition of $I_{j,k}$ in (3.8.9) and use (3.8.11) to obtain

$$\sigma_{\text{sb}}^2 = \frac{8}{\gamma^2} \left(\Gamma(-4\gamma) I_{2,2} - 2g_1^2 \Gamma(-3\gamma) I_{2,1} + \Gamma(-2\gamma) I_{1,1} \right), \qquad \gamma < 0.$$

Consider now $\gamma > 0$ and note

$$\int_0^\infty \left(\mathrm{e}^{-zt} - \mathrm{e}^{-u} \right) u^{-\beta-1} \, \mathrm{d}u = \frac{\Gamma(1-\beta)}{\beta} (1 - z^\beta), \qquad \beta \in (0,1).$$

to obtain via (3.8.12)

$$\begin{aligned} C_{\xi,j,k} &= -\gamma \frac{jk g_{j+k}}{j+k} \int_0^1 \left(A_\xi^{(j+k)\gamma} - 1 \right) w^{-j\gamma-1} (1-w)^{-k\gamma-1} \, \mathrm{d}w \\ &= -\gamma \frac{jk g_{j+k}}{j+k} I_{j,k}. \end{aligned}$$

Plugging this into (3.8.11) yields the in (3.8.8) stated formula for $\gamma > 0$.

Let $\gamma = 0$ and use the transformation $S_{i,\xi} = \exp(-Z_{i,\xi})$ to obtain

$$C_\xi := \operatorname{Cov}(h_1(Z_{1,\xi}), h_1(Z_{2,\xi})) = \frac{1}{4} C_{\xi,2,2} + \gamma_E C_{\xi,2,1} + \gamma_E^2 C_{\xi,1,1}, \tag{3.8.15}$$

where $C_{\xi,j,k} := \operatorname{Cov}(\log^j S_{1,\xi}, \log^k S_{2,\xi})$, $j,k = 1,2$ and γ_E denotes the Euler–Mascheroni constant. Simple but tedious calculations to derive the differences needed in Hoeffdings Lemma result in

$$\begin{aligned} &\mathbb{P}(\log S_{1,\xi} \le s, \log S_{2,\xi} \le t) - \mathbb{P}(\log S_{1,\xi} \le s) \mathbb{P}(\log S_{2,\xi} \le t) \\ &= \exp\left(-(\mathrm{e}^s + \mathrm{e}^t) A_\xi \left(\frac{\mathrm{e}^t}{\mathrm{e}^t + \mathrm{e}^s} \right) \right) - \exp\left(-(\mathrm{e}^s + \mathrm{e}^t) \right), \qquad s,t \in \mathbb{R}; \\ &\mathbb{P}(\log^2 S_{1,\xi} \le s, \log S_{2,\xi} \le t) - \mathbb{P}(\log^2 S_{1,\xi} \le s) \mathbb{P}(\log S_{2,\xi} \le t) \\ &= \exp\left(-(\mathrm{e}^{\sqrt{s}} + \mathrm{e}^t) A_\xi \left(\frac{\mathrm{e}^t}{\mathrm{e}^t + \mathrm{e}^{\sqrt{s}}} \right) \right) - \exp\left(-(\mathrm{e}^{-\sqrt{s}} + \mathrm{e}^t) A_\xi \left(\frac{\mathrm{e}^t}{\mathrm{e}^t + \mathrm{e}^{-\sqrt{s}}} \right) \right) \\ &\quad - \exp(-(\mathrm{e}^{\sqrt{s}} + \mathrm{e}^t)) + \exp(-(\mathrm{e}^{-\sqrt{s}} + \mathrm{e}^t)), \qquad s > 0, t \in \mathbb{R}; \end{aligned}$$

$$\begin{aligned}
&\mathbb{P}(\log^2 S_{1,\xi} \le s, \log^2 S_{2,\xi} \le t) - \mathbb{P}(\log^2 S_{1,\xi} \le s)\mathbb{P}(\log^2 S_{2,\xi} \le t) \\
&= \exp\left(-(\mathrm{e}^{\sqrt{s}} + \mathrm{e}^{\sqrt{t}})A_\xi\left(\frac{\mathrm{e}^{\sqrt{t}}}{\mathrm{e}^{\sqrt{t}} + \mathrm{e}^{\sqrt{s}}}\right)\right) - \exp\left(-(\mathrm{e}^{\sqrt{s}} + \mathrm{e}^{\sqrt{t}})\right) \\
&\quad + \exp\left(-(\mathrm{e}^{-\sqrt{s}} + \mathrm{e}^{-\sqrt{t}})A_\xi\left(\frac{\mathrm{e}^{-\sqrt{t}}}{\mathrm{e}^{-\sqrt{t}} + \mathrm{e}^{-\sqrt{s}}}\right)\right) - \exp\left(-(\mathrm{e}^{-\sqrt{s}} + \mathrm{e}^{-\sqrt{t}})\right) \\
&\quad - \exp\left(-(\mathrm{e}^{-\sqrt{s}} + \mathrm{e}^{\sqrt{t}})A_\xi\left(\frac{\mathrm{e}^{\sqrt{t}}}{\mathrm{e}^{\sqrt{t}} + \mathrm{e}^{-\sqrt{s}}}\right)\right) + \exp\left(-(\mathrm{e}^{-\sqrt{s}} + \mathrm{e}^{\sqrt{t}})\right) \\
&\quad - \exp\left(-(\mathrm{e}^{\sqrt{s}} + \mathrm{e}^{-\sqrt{t}})A_\xi\left(\frac{\mathrm{e}^{-\sqrt{t}}}{\mathrm{e}^{-\sqrt{t}} + \mathrm{e}^{\sqrt{s}}}\right)\right) + \exp\left(-(\mathrm{e}^{\sqrt{s}} + \mathrm{e}^{-\sqrt{t}})\right), \quad s,t > 0.
\end{aligned}$$

Hoeffding's Lemma, the upper displays and substitutions of the form $u = \mathrm{e}^{\pm\sqrt{s}} + \mathrm{e}^{\pm\sqrt{t}}$, $w = \mathrm{e}^{\pm\sqrt{t}}/(\mathrm{e}^{\pm\sqrt{s}} + \mathrm{e}^{\pm\sqrt{t}})$ yield

$$\begin{aligned}
C_{\xi,1,1} &= \int_0^1 \int_0^\infty \left(\mathrm{e}^{-uA_\xi(w)} - \mathrm{e}^{-u}\right) \frac{1}{uw(1-w)} \,\mathrm{d}u \,\mathrm{d}w, \\
C_{\xi,2,1} &= 2\int_0^1 \int_0^\infty \left(\mathrm{e}^{-uA_\xi(w)} - \mathrm{e}^{-u}\right) \frac{\log u + \log(1-w)}{uw(1-w)} \,\mathrm{d}u \,\mathrm{d}w, \\
C_{\xi,2,2} &= 4\int_0^1 \int_0^\infty \left(\mathrm{e}^{-uA_\xi(w)} - \mathrm{e}^{-u}\right) \frac{(\log u + \log(1-w))(\log u + \log w)}{uw(1-w)} \,\mathrm{d}u \,\mathrm{d}w.
\end{aligned} \tag{3.8.16}$$

Invoke the following integral identities

$$\begin{aligned}
&\int_0^\infty (\mathrm{e}^{-uz} - e^{-u}) \frac{1}{u} \,\mathrm{d}u = -\log z, \\
&\int_0^\infty (\mathrm{e}^{-uz} - e^{-u}) \frac{\log u}{u} \,\mathrm{d}u = \log z \frac{\log z + 2\gamma_E}{2}, \\
&\int_0^\infty (\mathrm{e}^{-uz} - e^{-u}) \frac{\log^2 u}{u} \,\mathrm{d}u = -\log z \frac{\pi^2 + 6\gamma_E^2 + 2\log^2 z + 6\gamma_E \log z}{6}
\end{aligned}$$

for $z > 0$ to obtain via (3.8.16)

$$\begin{aligned}
&\int_0^1 C_{\xi,1,1} \,\mathrm{d}\xi = \int_0^1 \frac{1}{w(1-w)} \int_0^1 -\log(A_\xi(w)) \,\mathrm{d}\xi \,\mathrm{d}w, \\
&\int_0^1 C_{\xi,2,1} \,\mathrm{d}\xi \\
&= \int_0^1 \frac{1}{w(1-w)} \int_0^1 \log(A_\xi(w)) \big[2\gamma_E - 2\log(1-w) \\
&\qquad + \log A_\xi(w)\big] \,\mathrm{d}\xi \,\mathrm{d}w, \\
&\int_0^1 \frac{C_{\xi,2,2}}{4} \,\mathrm{d}\xi \\
&= \int_0^1 \frac{1}{w(1-w)} \int_0^1 -\gamma_E^2 \log A_\xi(w) + \gamma_E \log A_\xi(w) \big[-\log A_\xi(w) \\
&\qquad + \log(1-w) + \log w\big] \\
&\qquad + \log A_\xi(w) \Big[-\frac{\pi^2 + 2\log^2 A_\xi(w)}{6} - \log w \log(1-w) \\
&\qquad\quad + \log A_\xi(w) \frac{\log(1-w) + \log w}{2}\Big] \,\mathrm{d}\xi \,\mathrm{d}w.
\end{aligned} \tag{3.8.17}$$

Now plug these formulas into (3.8.15), simplify and integrate to get

$$\begin{aligned}
\int_0^1 C_\xi \,\mathrm{d}\xi &= \int_0^{1/2} \frac{1}{w(1-w)} \int_0^1 \log A_\xi(w) \Big[-\frac{\pi^2 + 2\log^2 A_\xi(w)}{3} \\
&\quad + \log A_\xi(w)(\log(1-w) + \log w) - 2\log w \log(1-w) \Big] \,\mathrm{d}\xi \,\mathrm{d}w \\
&= \int_0^{1/2} \frac{1}{3w^2(1-w)} \Big[(w-1)\log^3(1-w) - 3(w-1)\log^2(1-w)\log w \\
&\quad + w(12 + \pi^2 + 6\log w) \\
&\quad + \log(1-w)(12 + \pi^2 - w(6+\pi^2) + 6\log w) \Big] \,\mathrm{d}w \\
&=: \int_0^{1/2} c_\xi \,\mathrm{d}\xi.
\end{aligned}$$

The integrand c_ξ has the antiderivative

$$\begin{aligned}
B(w) = -\frac{1}{3w} \Big[& 6w \operatorname{Li}_2(w) + 6w \operatorname{Li}_3(1-w)(\log(1-w) - 1) \\
& + w\log^3(1-w) - \log^3(1-w) - 3w \log w \log^2(1-w) \\
& + 3\log^2(1-w)\log w + 3\log^2(1-w) - 6w\log(1-w) \\
& + 6w\log(1-w)\log w + 6\log w \log(1-w) \\
& + \pi^2 \log(1-w) + 18\log(1-w) + 6w\log w \Big], \quad w \in (0, 1/2),
\end{aligned}$$

where $\operatorname{Li}_j(w) := \sum_{k=1}^\infty w^k/k^j$ denotes the Polylogarithm function for $w \in (0,1)$ and $j \in \mathbb{N}$. Take the limits to get

$$\begin{aligned}
&\lim_{w \downarrow 0} F(w) = 6 - 2\zeta(3), \\
&\lim_{w \uparrow 0} F(w) = -\frac{7}{4}\zeta(3) - \frac{\pi^2}{3} + \frac{4}{3}\log^3(2) - 6\log^2(2) + 12\log 2 + \frac{2}{3}\pi^2 \log 2,
\end{aligned}$$

which imply the formula in (3.8.8) for $\gamma = 0$. □

3.8.4 Additional simulation results on the bias-corrected estimator

By construction, the expectation of the bias-corrected sliding blocks estimator from Remark 3.2.8 should be close to the expectation of the disjoint blocks estimator. In terms of bias, no improvement 'beyond' the disjoint blocks bias should be visible. This intuition has been confirmed in simulation experiments, both where r is fixed with target parameter depending on r, and where n is fixed with target parameter depending on the limiting attractor distribution.

Regarding the former case, we chose to fix $r = 90$ and consider the estimation of the variance σ_r^2 in the ARMAX model with GPD margins. In Figure 3.5 we depict the expected difference (calculated based on averaging over 1000 simulation runs) between (1) the corrected sliding blocks estimator and the disjoint blocks and (2) the corrected sliding blocks estimator and the plain sliding blocks estimator. The results reveal that,

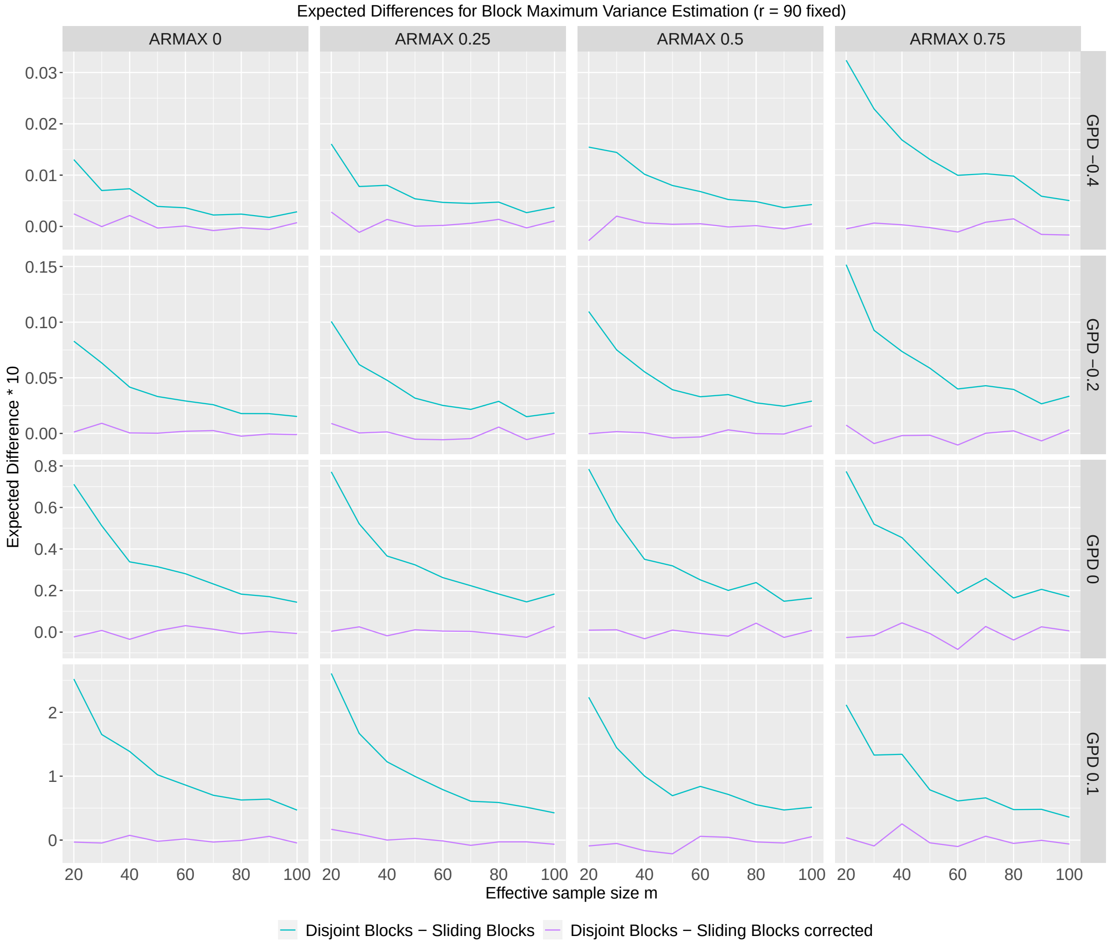

Figure 3.5: Expected differences $\mathrm{E}[\hat{\sigma}_r^{\mathrm{db}} - \hat{\sigma}_r^{\mathrm{sb}}]$ and $\mathrm{E}[\hat{\sigma}_r^{\mathrm{db}} - \hat{\sigma}_r^{\mathrm{sb-corr}}]$, with fixed block size $r = 90$ and for increasing effective sample size m (i.e., number of disjoint blocks). The four columns correspond to four time series models, and the four rows to four choices of GPD-margins for the innovations.

indeed, the expected value of the corrected sliding blocks estimator is the same as the expected value of the disjoint blocks estimator; the two hence have the same bias. The overall performance of the three estimators is compared in Figure 3.6. The results show that, once again, the bias is clearly negligible compared to the variance for all three estimators (the reason being that the target parameter is σ_r^2, which is not an 'asymptotic' parameter). In terms of estimation variance we observe that, the larger γ and the smaller m, the closer the estimation variance of the corrected estimator is to the estimation variance of the disjoint blocks counterpart (as a consequence, the plain sliding blocks estimator overall wins the race). The observed behavior in terms of m may be explained by the fact that, for fixed $r = 90$, the percentage of summands that is removed for the corrected sliding estimator is quite large for small m (e.g., $\approx 10\%$ for $m = 20$), while it is much smaller for large m (e.g., $\approx 2\%$ for $m = 100$).

Next, regarding the case where n is fixed, we chose to fix $n = 1000$ and consider esti-

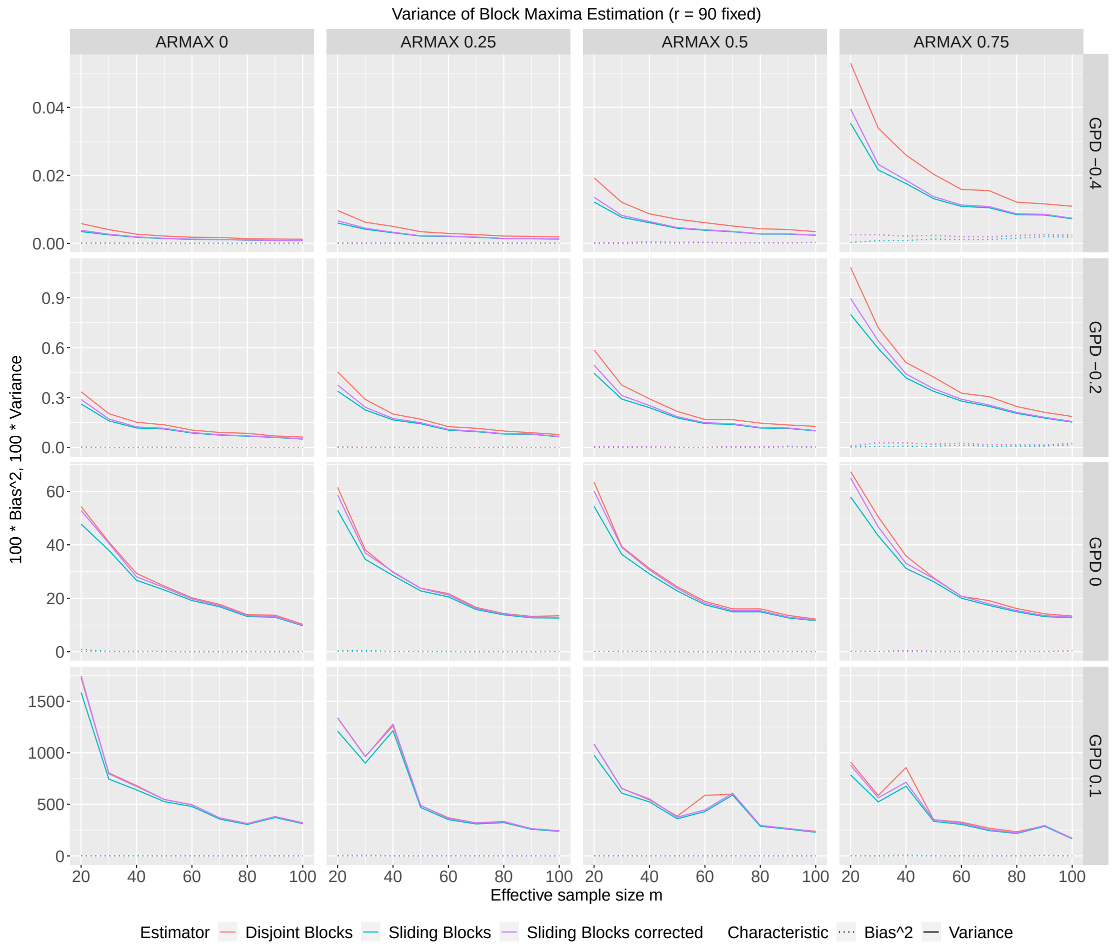

Figure 3.6: Estimation variance and squared bias of $\hat{\sigma}_r^{\mathrm{db}}, \hat{\sigma}_r^{\mathrm{sb}}$ and $\hat{\sigma}_r^{\mathrm{sb-corr}}$, with fixed block size $r = 90$ and for increasing effective sample size m (i.e., number of disjoint blocks). The four columns correspond to four time series models, and the four rows to four choices of GPD-margins for the innovations.

mation of Kendall's tau in the bivariate i.i.d. case drawn from the outer power Clayton copula (similar results were obtained for other time series models). The results are summarized in Figure 3.7. We observe that the bias of all three estimators is nearly identical (which may be explained by the fact that the bias is dominated by the difference $\tau_r - \tau$ for all three estimators), and that the variance of the bias-corrected and the plain sliding estimators are nearly identical as well (which is akin to the case $\gamma = -0.4$ in Figure 3.5).

Overall, in view of the preceding findings and the fact that the bias-corrected estimator may be computationally more expensive, we cannot recommend its use in practice in general (note that Bücher and Zanger (2023), Section E.3, came to the same conclusion).

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG, Ger-

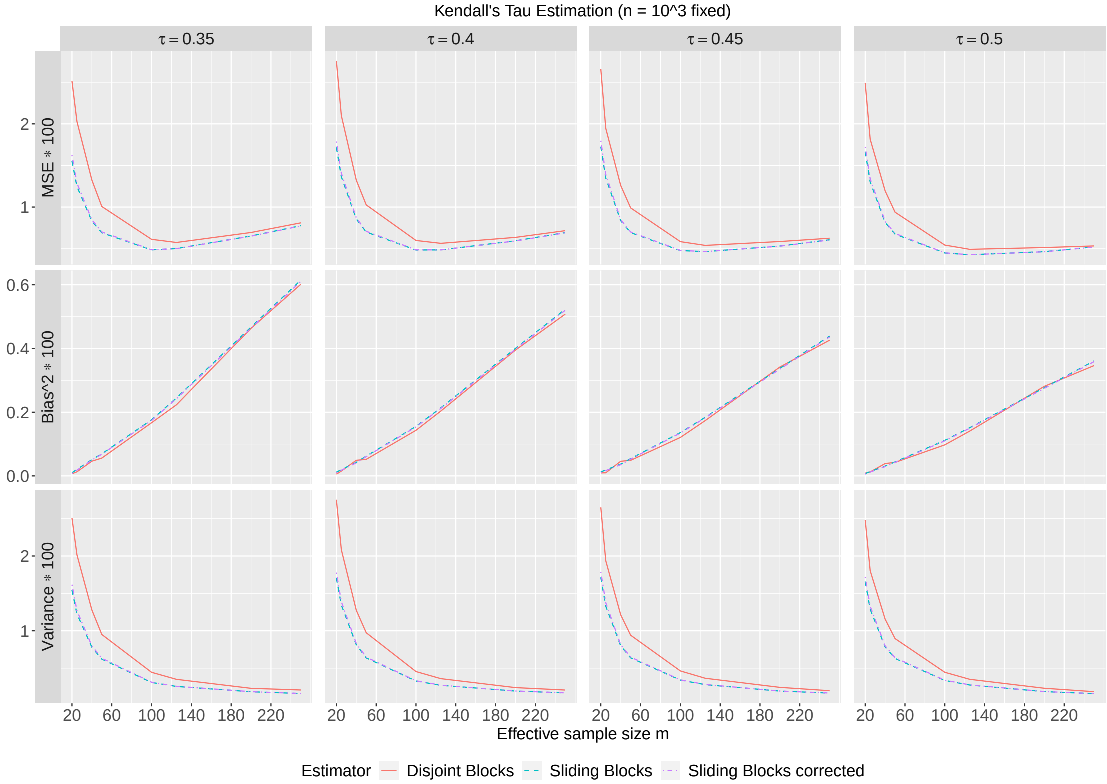

Figure 3.7: Mean squared error, squared bias and estimation variance of $\hat{\tau}^{\mathrm{db}}_{n,r}$, $\hat{\tau}^{\mathrm{sb}}_{n,r}$ and $\hat{\tau}^{\mathrm{sb-corr}}_{n,r}$, considered as estimators for Kendall's tau of the attractor copula, with fixed sample size $n = 1000$. From left to right: Kendall's tau of the attractor copula $0.35, 0.4, 0.45, 0.5$.

man Research Foundation), DFG project 465665892, which is gratefully acknowledged. The authors are grateful to two unknown referees and an associate editor for their constructive comments that helped to improve the presentation substantially. The authors also appreciate valuable comments by multiple participants of the Extreme Value Analysis (EVA) Conference in Milano in 2023.

4 Bootstrapping block maxima estimators for time series

In this section we present the preprint Bücher and Staud (2024a) which is concerned with bootstrapping within the block maxima method. Only minor changes to improve the presentation within this thesis have been made.

Abstract

The block maxima method is a standard approach for analyzing the extremal behavior of a potentially multivariate time series. It has recently been found that the classical approach based on disjoint block maxima may be universally improved by considering sliding block maxima instead. However, the asymptotic variance formula for estimators based on sliding block maxima involves an integral over the covariance of a certain family of multivariate extreme value distributions, which makes its estimation, and inference in general, an intricate problem. As an alternative, one may rely on bootstrap approximations: we show that naive block-bootstrap approaches from time series analysis are inconsistent even in i.i.d. situations, and provide a consistent alternative based on resampling circular block maxima. As a by-product, we show consistency of the classical resampling bootstrap for disjoint block maxima, and that estimators based on circular block maxima have the same asymptotic variance as their sliding block maxima counterparts. The finite sample properties are illustrated by Monte Carlo experiments, and the methods are demonstrated by a case study of precipitation extremes.

Keywords. Bootstrap Consistency; Disjoint and Sliding Block Maxima; Extreme Value Statistics; Pseudo Maximum Likelihood Estimation; Time Series Analysis.

MSC subject classifications. Primary 62F40, 62G32; Secondary 62E20.

4.1 Introduction

Let $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ with $\boldsymbol{X}_t = (X_{t1},\dots,X_{td})^\top$ denote a stationary time series in $\mathbb{R}^d$. For block size $r \in \mathbb{N}$, let $\boldsymbol{M}_r = \max_{t=1}^r \boldsymbol{X}_t = (\max_{t=1}^r X_{t1},\dots,\max_{t=1}^r X_{td})^\top$ denote the componentwise maximum over r successive observations. In view of a version of the classical extremal types theorem for multivariate time series, the distribution of $\boldsymbol{M}_r$ may be approximated by a multivariate extreme value distribution, for sufficiently large block size r (see Section 4.2.1 for details). Both the distribution of $\boldsymbol{M}_r$ and parameters related to its potential weak limit distribution are common target parameters in extreme value statistics, with applications in finance, insurance or environmental statistics among others (Beirlant et al., 2004; Katz et al., 2002; Philip et al., 2020).

Suppose $X_1, \dots, X_n$ is an observed stretch from $(X_t)_{t\in\mathbb{Z}}$. The classical block maxima method for estimating parameters associated with the law of M_r (or its limiting law for $r \to \infty$) consists of dividing the sampling period $\{1, \dots, n\}$ into m disjoint blocks of length r (for simplicity, we assume that $n = mr$), and of using the sample of (disjoint) block maxima $(M_{1,r}^{(\mathrm{db})}, \dots, M_{m,r}^{(\mathrm{db})})$ as a starting point for statistical methods; here $M_{i,r}^{(\mathrm{db})} = \max_{t=(i-1)r+1,\dots,ir} X_t$ denotes the ith disjoint block maximum. This classical approach can often be improved by instead considering the sample of sliding block maxima $(M_{1,r}^{(\mathrm{sb})}, \dots, M_{n-r+1,r}^{(\mathrm{sb})})$ as a starting point; here $M_{i,r}^{(\mathrm{sb})} = \max_{t=i,\dots,i+r-1} X_t$ denotes the block maximum of size r starting at "day" i. Note that this sample also provides a stationary, but strongly auto-correlated sample from the law of M_r. It was shown in Bücher and Segers (2018a); Zou et al. (2021) among others that estimators that explicitly or implicitly involve empirical means of the sliding block maxima sample typically have the same expectation but a slightly smaller variance than the respective counterparts based on the disjoint block maxima sample.

It seems natural that a lower estimation variance offers the possibility of constructing smaller confidence intervals for a given confidence level. The main goal of this work is to find universal practical solutions for this heuristic based on suitable bootstrap approximations. A challenge consists of the fact that the large sample asymptotics of the sliding block maxima method depend on a blocking of blocks approach involving some intermediate blocking parameter converging to infinity at a well-balanced rate, whence standard bootstrap approaches for time series like the block bootstrap (Lahiri, 2003) would depend on such an intermediate blocking parameter as well (which must be avoided due to the typically small effective sample size in extreme value analysis). However, as we will show, a novel approach based on what we call the circular (sliding) block maxima sample allows for consistent distributional approximations without relying on such an intermediate parameter sequence. At the same time, the method is computationally attractive because it avoids recalculation of any block maxima for the bootstrap sample.

In general, various variants of the bootstrap have been routinely applied in extreme value statistics (see, for instance, Eastoe and Tawn, 2008; Huser and Davison, 2014). However, to the best of our knowledge, respective statistical theory is only available for the peak-over-threshold method (Peng and Qi, 2008; Drees, 2015; Davis et al., 2018; Kulik and Soulier, 2020; Jentsch and Kulik, 2021; de Haan and Zhou, 2024). Concerning the block maxima method, within the regime where the block size is treated as a parameter sequence $r = r_n$ converging to infinity, the only reference we are aware of is de Haan and Zhou (2024), who derive an asymptotic expansion for the tail quantile process of bootstrapped disjoint block maxima. However, their results require an underlying i.i.d. sequence, they only concern the disjoint block maxima method and they do not eventually imply desirable classical consistency statements. While our main focus is on the sliding block maxima method, we establish consistency results for the disjoint block maxima method as a by-product, for serially dependent data.

Finally, we also derive limit results for empirical means and estimators based on the circular (sliding) block maxima sample, and prove that they show the same favorable asymptotic behavior as the sliding block maxima counterparts. As we will discuss in the conclusion, the result may be of independent interest when dealing with non-stationary situations involving, for example, trends; note that the latter is a typical use case for climate extremes.

The remaining parts of this paper are organized as follows: basic model assumptions and known results on the disjoint and sliding block maxima method are collected in Section 4.2. The sample of circular block maxima is introduced in Section 4.3 and supplemented by a central limit theorem. Respective bootstrap schemes are proposed in Section 4.4, where we also provide respective consistency statements. The previous high-level results are applied to a specific estimation problem for univariate heavy tailed time series in Section 4.5. Finite-sample results on a large-scale simulation study and a small case study are presented in Sections 4.6 and 4.7, respectively. Section 4.8 concludes, and all proofs and technical conditions are collected in Sections 4.9-4.11. Some additional theoretical results and simulation results are deferred to a sequence of appendices.

Throughout, weak convergence of sequences of random vectors/distributions is denoted by $\rightsquigarrow$. The code used for the simulation study is available on Github, see Staud (2024).

4.2 Mathematical preliminaries

4.2.1 Basic model assumptions

An extension of the classical extremal types theorem to strictly stationary time series (Leadbetter, 1983) implies that, under suitable broad conditions, affinely standardized maxima extracted from a stationary time series converge to the generalized extreme value distribution (GEV). This was generalized to the multivariate case in Hsing (1989), where the marginals are necessarily GEV-distributed. We make this an assumption, and additionally require the scaling sequences to exhibit some common regularity inspired by the max-domain of attraction condition in the i.i.d. case (de Haan and Ferreira, 2006).

Condition 4.2.1 (Multivariate max-domain of attraction)**.** Let $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ denote a strictly stationary time series in $\mathbb{R}^d$ with continuous margins. There exist sequences $(\boldsymbol{a}_r)_r = (a_r^{(1)},\dots,a_r^{(d)})_r \subset (0,\infty)^d, (\boldsymbol{b}_r)_r = (b_r^{(1)},\dots,b_r^{(d)})_r \subset \mathbb{R}^d$ and $\boldsymbol{\gamma} = (\gamma^{(1)},\dots,\gamma^{(d)}) \in \mathbb{R}^d$, such that, for any $s > 0$ and $j \in \{1,\dots,d\}$,

$$\lim_{r\to\infty} \frac{a_{\lfloor rs\rfloor}^{(j)}}{a_r^{(j)}} = s^{\gamma^{(j)}}, \qquad \lim_{r\to\infty} \frac{b_{\lfloor rs\rfloor}^{(j)} - b_r^{(j)}}{a_r^{(j)}} = \frac{s^{\gamma^{(j)}} - 1}{\gamma^{(j)}}, \tag{4.2.1}$$

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

$$Z_r = \frac{\boldsymbol{M}_r - \boldsymbol{b}_r}{\boldsymbol{a}_r} \rightsquigarrow Z \sim G, \tag{4.2.2}$$

where G denotes a d-variate extreme-value distribution with marginal c.d.f.s $G_{\gamma^{(1)}}, \dots, G_{\gamma^{(d)}}$ (with $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}) \mathbf{1}\{1 + \gamma x > 0\}$ the cdf of the GEV(γ)-distribution) and where division by the vector $\boldsymbol{a}_r$ is understood componentwise, that is, $\boldsymbol{Z}_r = (Z_r^{(1)}, \dots, Z_r^{(d)})$ with $Z_r^{(j)} = \{\max(X_1^{(j)}, \dots, X_r^{(j)}) - b_r^{(j)}\}/a_r^{(j)}$ for $j \in \{1, \dots, d\}$.

Note that (4.2.1) and (4.2.2) may for instance be deduced from Leadbetter's $D(u_n)$ condition, a domain-of-attraction condition on the associated i.i.d. sequence with stationary distribution equal to that of $\boldsymbol{X}_0$ and a weak requirement on the convergence of the c.d.f. of $\boldsymbol{Z}_r$, see Theorem 10.22 in Beirlant et al. (2004).

Throughout this paper, we assume to observe $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, an excerpt from a d-dimensional time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ satisfying Condition 4.2.1. Block maxima will be formed with respect to a block size parameter $r = r_n$ converging to infinity such that $r = o(n)$. For establishing the subsequent limit results, the serial dependence of $(\boldsymbol{X}_t)_t$ will additionally be controlled via the Rosenblatt mixing coefficient (Bradley, 2005). For two sigma-fields $\mathcal{F}_1, \mathcal{F}_2$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let $\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup_{A\in\mathcal{F}_1, B\in\mathcal{F}_2} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$. For positive integer p, let $\alpha(p) := \alpha\big(\sigma((\boldsymbol{X}_t)_{t\le 0}), \sigma((\boldsymbol{X}_t)_{t\ge p})\big)$, with $\sigma(\cdot)$ denoting the sigma-field generated by its argument.

Condition 4.2.2 (Block size and mixing)**.** For the block size sequence $(r_n)_n$ it holds that, as $n \to \infty$:

(a) $r_n \to \infty$ and $r_n = o(n)$.
(b) There exists a sequence $(\ell_n)_n \subset \mathbb{N}$ such that $\ell_n \to \infty$, $\ell_n = o(r_n)$, $\frac{r_n}{\ell_n}\alpha(\ell_n) = o(1)$, and $\frac{n}{r_n}\alpha(\ell_n) = o(1)$.
(c) $\left(\frac{n}{r_n}\right)^{1+\omega} \alpha(r_n) = o(1)$ for some $\omega > 0$.

4.2.2 Empirical means of rescaled disjoint or sliding block maxima

A central theoretical ingredient for establishing weak limit results on estimators based on the block maxima method is the weak convergence of either the tail quantile process or of empirical means of (unobservable) rescaled block maxima. Respective results on the former can be found in Theorem 2.1 in Ferreira and de Haan (2015), while the latter approach was taken in Theorems 2.6/B.1 in Bücher and Zanger (2023), in the proof of Theorem 2.6 in Bücher and Segers (2018a) or in Theorem 2.4 in Zou et al. (2021). Throughout this paper, we also follow the latter approach, and for completeness, we summarize the essence of the respective results in a theorem and briefly summarize and discuss potential statistical applications.

For $i \in \{1, \dots, n-r+1\}$, write $\boldsymbol{M}_{r,i} = \max(\boldsymbol{X}_i, \dots, \boldsymbol{X}_{i+r-1})$, and define $\overline{\mathcal{M}}_{n,r}^{(\mathrm{db})} = (\boldsymbol{M}_{r,i} : i \in I_n^{\mathrm{db}})$ and $\overline{\mathcal{M}}_{n,r}^{(\mathrm{sb})} = (\boldsymbol{M}_{r,i} : i \in I_n^{\mathrm{sb}})$ as the (vanilla) disjoint and sliding block maxima samples, respectively, where $I_n^{\mathrm{db}} = \{(i-1)r + 1 : 1 \le i \le n/r\}$ and $I_n^{\mathrm{sb}} = \{1, \dots, n-r+1\}$.

We are interested in the associated empirical measures $n_{\mathrm{mb}}^{-1}\sum_{i\in I_n^{\mathrm{mb}}}\delta_{\boldsymbol{M}_{r,i}}$, or their versions based on rescaled block maxima $n_{\mathrm{mb}}^{-1}\sum_{i\in I_n^{\mathrm{mb}}}\delta_{(\boldsymbol{M}_{r,i}-\boldsymbol{b}_r)/\boldsymbol{a}_r}$, with $n_{\mathrm{mb}}=|I_n^{\mathrm{mb}}|$ and $\delta_{\boldsymbol{z}}$ the Dirac-measure at $\boldsymbol{z}$. The fact that the sample size n_{mb} depends on mb is a notational nuisance which we subsequently resolve by the following asymptotically negligible modification: first, the disjoint block maxima sample may be identified with the sample $\mathcal{M}_{n,r}^{(\mathrm{db})}=(\boldsymbol{M}_{r,1},\dots,\boldsymbol{M}_{r,1},\dots,\boldsymbol{M}_{r,n/r},\dots,\boldsymbol{M}_{r,n/r})$ of size n containing each disjoint block maximum exactly r times; note that the respective empirical measures of $\overline{\mathcal{M}}_{n,r}^{(\mathrm{db})}$ and $\mathcal{M}_{n,r}^{(\mathrm{db})}$ are the same. Next, we define $\mathcal{M}_{n,r}^{(\mathrm{sb})}$ as the sliding block maxima sample of size n calculated from the extended sample $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n,\boldsymbol{X}_{n+1},\dots,\boldsymbol{X}_{n+r-1})$. In the subsequent asympototic results, this modification is asymptotically negligible since the first $n-r+1$ maxima in $\mathcal{M}_{n,r}^{(\mathrm{sb})}$ are exactly the maxima in $\overline{\mathcal{M}}_{n,r}^{(\mathrm{sb})}$. These modifications allow to define $(\boldsymbol{M}_{r,1}^{(\mathrm{mb})},\dots,\boldsymbol{M}_{r,n}^{(\mathrm{mb})}):=\mathcal{M}_{n,r}^{(\mathrm{mb})}$, and using the notation $\boldsymbol{Z}_{r,i}^{(\mathrm{mb})}:=(\boldsymbol{M}_{r,i}^{(\mathrm{mb})}-\boldsymbol{b}_r)/\boldsymbol{a}_r$, we define

$$\bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}=\sqrt{\frac{n}{r}}\big(\mathbb{P}_{n,r}^{(\mathrm{mb})}-P_r\big),\quad \mathbb{G}_{n,r}^{(\mathrm{mb})}=\sqrt{\frac{n}{r}}\big(\mathbb{P}_{n,r}^{(\mathrm{mb})}-P\big),$$

where $\mathbb{P}_{n,r}^{(\mathrm{mb})}=\frac{1}{n}\sum_{i=1}^n\delta_{\boldsymbol{Z}_{r,i}^{(\mathrm{mb})}}$, $P_r=\mathbb{P}(\boldsymbol{Z}_r\in\cdot)$ and $P=\mathbb{P}(\boldsymbol{Z}\in\cdot)$. Finally, let

$$G_\xi(\boldsymbol{x},\boldsymbol{y})=G(\boldsymbol{x})^{\xi\wedge 1}G(\boldsymbol{y})^{\xi\wedge 1}G(\boldsymbol{x}\wedge\boldsymbol{y})^{1-(\xi\wedge 1)},\qquad \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d,\tag{4.2.3}$$

and note that G_ξ is a $2d$-variate extreme-value distribution; see Formula (3.8) and its discussion in Bücher and Staud (2024b) for further details.

Theorem 4.2.3. *Under Conditions 4.2.1 and 4.2.2, for any finite set of real valued functions $h_1,\dots,h_q$ satisfying the integrability Condition 4.10.1(a) with $\nu>2/\omega$ where ω is from Condition 4.2.2, we have, writing $\boldsymbol{h}=(h_1,\dots,h_q)^\top$,*

$$\bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}\boldsymbol{h}=\big(\bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}h_1,\dots,\bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}h_q\big)^\top\rightsquigarrow\mathcal{N}_q(\boldsymbol{0},\Sigma_{\boldsymbol{h}}^{(\mathrm{mb})}),\qquad \mathrm{mb}\in\{\mathrm{db},\mathrm{sb}\},$$

where

$$\big(\Sigma_{\boldsymbol{h}}^{(\mathrm{db})}\big)_{j,j'=1}^q=\mathrm{Cov}\big(h_j(\boldsymbol{Z}),h_{j'}(\boldsymbol{Z})\big),\qquad \big(\Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}\big)_{j,j'=1}^q=2\int_0^1\mathrm{Cov}\big(h_j(\boldsymbol{Z}_{1,\xi}),h_{j'}(\boldsymbol{Z}_{2,\xi})\big)\,\mathrm{d}\xi,\tag{4.2.4}$$

with $\boldsymbol{Z}\sim G$ from Condition 4.2.1 and $(\boldsymbol{Z}_{1,\xi},\boldsymbol{Z}_{2,\xi})\sim G_\xi$ from (4.2.3). Moreover, $\Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}\le_L\Sigma_{\boldsymbol{h}}^{(\mathrm{db})}$, where $\le_L$ denotes the Loewner ordering. Finally, if the bias Condition 4.10.2 holds, then

$$\mathbb{G}_{n,r}^{(\mathrm{mb})}\boldsymbol{h}=\big(\mathbb{G}_{n,r}^{(\mathrm{mb})}h_1,\dots,\mathbb{G}_{n,r}^{(\mathrm{mb})}h_q\big)^\top\rightsquigarrow\mathcal{N}_q(\boldsymbol{B}_{\boldsymbol{h}},\Sigma_{\boldsymbol{h}}^{(\mathrm{mb})}).$$

Proof. The proof is a simplified version of our proof of Theorem 4.3.2 below, whence we omit further details. Under slightly different conditions, a proof can be found in Theorem B.1 in Bücher and Zanger (2023) or Theorem 8.7 in Bücher and Staud (2024b). □

In statistical applications, one is typically interested in distributional approximations for the estimation error of general statistics depending on the observable block maxima $\boldsymbol{M}_{r,i}$ (here and in the following, we omit the upper index 'mb'). As discussed in the following remark, such approximations can be deduced from Theorem 4.2.3.

Remark 4.2.4 (Normal approximations for statistics depending on observable block maxima). Suppose that $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\varphi}_n(\boldsymbol{M}_{r,1}, \dots, \boldsymbol{M}_{r,n})$ is some estimator of interest for a target parameter $\boldsymbol{\theta}_r \in \mathbb{R}^p$ (with $p \in \mathbb{N}$), where $\boldsymbol{\varphi}_n$ is a measurable function on $(\mathbb{R}^d)^n$ with values in $\mathbb{R}^p$. As discussed in Example 4.2.5 below, for many functions $\boldsymbol{\varphi}_n$ of practical interest there exists a positive integer q and functions $\boldsymbol{\psi}_n : (\mathbb{R}^d)^n \to \mathbb{R}^q$, $A: (0,\infty)^d \times \mathbb{R}^d \to \mathbb{R}^{p\times q}$ and $B: \mathbb{R}^d \times (0,\infty)^d \to \mathbb{R}^p$ such that

$$\boldsymbol{\varphi}_n(\boldsymbol{m}_1, \dots, \boldsymbol{m}_n) = A(\boldsymbol{a}, \boldsymbol{b})\boldsymbol{\psi}_n\Big(\frac{\boldsymbol{m}_1 - \boldsymbol{b}}{\boldsymbol{a}}, \dots, \frac{\boldsymbol{m}_n - \boldsymbol{b}}{\boldsymbol{a}}\Big) + B(\boldsymbol{a}, \boldsymbol{b}) \tag{4.2.5}$$

for all $\boldsymbol{m}_1, \dots, \boldsymbol{m}_n \in \mathbb{R}^d$ and $\boldsymbol{b} \in \mathbb{R}^d, \boldsymbol{a} \in (0,\infty)^d$. In such a case, if we further assume that there exists a sequence $\boldsymbol{\vartheta}_r \in \mathbb{R}^q$ solving the equations

$$\boldsymbol{\theta}_r = A(\boldsymbol{a}_r, \boldsymbol{b}_r)\boldsymbol{\vartheta}_r + B(\boldsymbol{a}_r, \boldsymbol{b}_r) \tag{4.2.6}$$

(again, see below for examples), we immediately obtain that the estimation error of $\hat{\boldsymbol{\theta}}_n$ can be written as

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_r = A(\boldsymbol{a}_r, \boldsymbol{b}_r)(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_r), \tag{4.2.7}$$

where $\hat{\boldsymbol{\vartheta}}_n := \boldsymbol{\psi}_n(\boldsymbol{Z}_{r,1}, \dots, \boldsymbol{Z}_{r,n})$ is a function of the rescaled block maxima. The difference $\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_r$ often allows for a linearization (for instance, by the delta-method): there exist real-valued functions $h_1, \dots, h_q$ such that

$$\sqrt{\frac{n}{r}}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_r) = \bar{\mathbb{G}}_{n,r}(h_1, \dots, h_q)^\top + R_n \tag{4.2.8}$$

for some remainder $R_n = o_{\mathbb{P}}(1)$. Loosely spoken, if the error term R_n in the previous display is sufficiently small, the previous two displays and Theorem 4.2.3 imply the normal approximation

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_r = \sqrt{\frac{r}{n}}A(\boldsymbol{a}_r, \boldsymbol{b}_r)\big\{\bar{\mathbb{G}}_{n,r}(h_1, \dots, h_q)^\top + R_n\big\} \approx_d \mathcal{N}_p\Big(0, \frac{r}{n}A(\boldsymbol{a}_r, \boldsymbol{b}_r)\Sigma_{\boldsymbol{h}}A(\boldsymbol{a}_r, \boldsymbol{b}_r)^\top\Big), \tag{4.2.9}$$

where $\Sigma_{\boldsymbol{h}}$ is the matrix from Theorem 4.2.3. More precisely, in the case where the matrix A is invertible (in particular, $p = q$), we obtain the more precise result

$$A(\boldsymbol{a}_r, \boldsymbol{b}_r)^{-1}\sqrt{\frac{n}{r}}(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r) = \bar{\mathbb{G}}_{n,r}(h_1, \dots, h_q)^\top + o_{\mathbb{P}}(1) \rightsquigarrow \mathcal{N}_p(\boldsymbol{0}, \Sigma_{\boldsymbol{h}}). \tag{4.2.10}$$

Example 4.2.5. (i) Empirical variance: consider the case $d = 1$ and the empirical variance $\hat{\theta}_n = n^{-1}\sum_{i=1}^n (M_{r,i} - \bar{M}_{r,n})^2$, where $\bar{M}_{r,n} := n^{-1}\sum_{i=1}^n M_{r,i}$, considered as an estimator for $\theta_r := \operatorname{Var}(M_{r,1})$. In that case, (4.2.5) is met with $p = 1, q = 2$, $\boldsymbol{\psi}_n(z_1, \dots, z_n) = ((\frac{1}{n}\sum_{i=1}^n z_i)^2, \frac{1}{n}\sum_{i=1}^n z_i^2)^\top$, $B(a,b) = 0$ and $A(a,b) = (-a^2, a^2)$. Moreover, $\boldsymbol{\vartheta}_r = (\mathrm{E}[Z_{r,1}]^2, \mathrm{E}[Z_{r,1}^2])^\top$ is a solution to (4.2.6) satisfying (4.2.8): indeed, with $h_1(x) = x, h_2(x) = x^2$, and using $P_r h_1 \to P h_1$, we may write

$$\sqrt{\frac{n}{r}}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_r) = \sqrt{\frac{n}{r}}\begin{pmatrix} (\mathbb{P}_{n,r}h_1)^2 - (P_r h_1)^2 \\ \mathbb{P}_{n,r}h_2 - P_r h_2 \end{pmatrix} = \sqrt{\frac{n}{r}}\begin{pmatrix} (2Ph_1 + o_{\mathbb{P}}(1))(\mathbb{P}_{n,r}h_1 - P_r h_1) \\ \mathbb{P}_{n,r}h_2 - P_r h_2 \end{pmatrix}$$

$$= \bar{\mathbb{G}}_{n,r}((2Ph_1) \cdot h_1, h_2)^\top + o_{\mathbb{P}}(1).$$

Assembling terms and solving for $\hat\theta_n - \theta_r$, we obtain $\sqrt{\frac{n}{r}} a_r^{-2}(\hat\theta_n - \theta_r) = \bar{\mathbb{G}}_{n,r}(h_2 - (2Ph_1) \cdot h_1) + o_{\mathbb{P}}(1)$. The asymptotic variance of the limiting distribution has been explicitly calculated in Bücher and Staud (2024b), see their Corollary 4.1 and their Equations (C.1) and (C.2).

(ii) Probability weighted moments: consider the case $d = 1$ and let $\boldsymbol{\varphi}_n : \mathbb{R}^n \to \mathbb{R}^p$ denote the vector containing the first p empirical probability weighted moments. In that case, Equation (4.2.5) is met with $q = p$, $\boldsymbol{\psi}_n = \boldsymbol{\varphi}_n$, $A(a,b) = \mathrm{diag}(a,\dots,a)$ and $B(a,b) = \mathrm{diag}(b, b/2, \dots, b/(p+1))$; see Formula (A.7) in Bücher and Zanger (2023). Moreover, Equation (4.2.8) can be deduced from (A.5) in that paper (for $p = 3$), with functions $h_1(x) = x$, $h_2(x) = xG_\gamma(x) + \int_x^\infty z \,\mathrm{d}G_\gamma(z)$ and $h_3(x) = xG_\gamma^2(x) + 2\int_x^\infty zG_\gamma(z)\,\mathrm{d}G_\gamma(z)$. Finally, up to the treatment of a bias term, (4.2.10) corresponds to the assertion in their Theorem 3.2.

(iii) Pseudo maximum likelihood estimation in the heavy tailed case: the main steps from Remark 4.2.4 also apply in a slightly modified setting tailored to the heavy tailed case; details are worked out in Section 4.5 below.

In statistical applications, results like those in (4.2.9) are typically used as a starting point for inference, for instance in the form of confidence intervals for $\hat{\boldsymbol{\theta}}_n$. Routinely, such intervals would be based on normal approximations involving consistent estimators for the variances on the right-hand side of those displays, which in our case requires estimation of $\boldsymbol{a}_r, \boldsymbol{b}_r$ and $\Sigma_{\boldsymbol{h}} = \Sigma_{\boldsymbol{h}}^{(\mathrm{mb})}$. This can be a complicated task, especially for the sliding block maxima method in view of the complicated formula for the limiting variance. Alternatively, one can rely on bootstrap approximations instead of normal approximations, which can avoid the need for variance estimation. For the disjoint blocks case, the proof of Theorem 4.2.3 shows that the disjoint block maxima can essentially be considered independent (despite the serial dependence in the underlying observations). This suggests that the standard bootstrap based on resampling with replacement from the disjoint block maxima should be consistent; a conjecture that will be confirmed in Section 4.4. In the sliding block maxima case, however, the standard bootstrap cannot work since it destroys the serial dependence in the sliding block maxima sample. In classical time series analysis, a remedy consists of the block bootstrap, where blocks of successive observations are drawn with replacement. In our case, this heuristically requires the choice of a smoothing parameter sequence that is of larger order than the block size; see Remark 4.4.3 for details on the inconsistency of a naive approach to bootstrapping sliding block maxima. An alternative approach relies on the circular block maxima method as introduced in the next section.

4.3 The circular block maxima sample

In this section we introduce the circular block maxima method. We show that respective empirical means yield the same asymptotic variance as the sliding block maxima counterparts. The approach is interesting in its own (particularly for computational reasons), but most importantly for this paper it suggests a straightforward bootstrap approach for bootstrapping sliding block maxima that will be discussed in Section 4.4.

Formally, given a sample $\mathcal{X}_n = (X_1, \dots, X_n)$ as before, the circular block maxima sample

$$\mathcal{M}_{n,r}^{(\mathrm{cb})} = (M_{r,1}^{(\mathrm{cb})}, \dots, M_{r,n}^{(\mathrm{cb})})$$

is a new sample of size n containing suitable block maxima that asymptotically follow the distribution G in (4.2.2). The sample depends on two parameters: the block length parameter r and an integer-valued parameter k (for instance, $k = 2$ or $k = 3$) that determines the length of the interval over which we apply the circular maximum operation. Details for the univariate case are provided in Figure 4.1, where the sampling period $\{1, \dots, n\}$ has been decomposed into $m(k) := m_n(k) := n/(kr)$ blocks of size kr, say $I_{kr,1}, \dots, I_{kr,m(k)}$ with $I_{kr,i} = \{(i-1)kr + 1, \dots, ikr\}$, and where we assume that $m(k) \in \mathbb{N}$ for simplicity. Note that every observation X_s appears within exactly r maximum-operations, and hence has the same chance to become a block maximum (this is not the case for the plain sliding block maxima sample, where, for instance, the very first observation X_1 can appear only once as sliding block maximum). This observation is in fact the main motivation for the circularization approach within each kr-block.

More formally, the notations used in Figure 4.1 are defined as follows: for a given vector $\boldsymbol{W} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_n) \in (\mathbb{R}^d)^n$ and $i \in \{1, \dots, m(k)\}$, we let

$$\pi_{kr,i}(\boldsymbol{W}) = (\boldsymbol{w}_s)_{s \in I_{kr,i}} = \big(\boldsymbol{w}_{(i-1)kr+1}, \dots, \boldsymbol{w}_{ikr}\big)$$

denote the projection of $\boldsymbol{W}$ to its coordinates defined by the i-th kr-block. For $\boldsymbol{V} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_q) \in (\mathbb{R}^d)^q$ with $q \in \mathbb{N}_{\geq r}$, consider the sliding-maxima-operation

$$\text{slid-max}(\boldsymbol{V} \mid r) = \Big(\max_{s \in [1:r]} \boldsymbol{v}_s, \max_{s \in [2:r+1]} \boldsymbol{v}_s, \dots, \max_{s \in [q-r+1:q]} \boldsymbol{v}_s \Big) \in (\mathbb{R}^d)^{q-r+1},$$

where maxima over vectors in $\mathbb{R}^d$ are to be understood componentwise and where $[i : j] = \{i, i+1, \dots, j\}$. Next, for $\boldsymbol{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_{kr}) \in (\mathbb{R}^d)^{kr}$, the circularization function is defined as

$$\text{circ}(\boldsymbol{U} \mid r) := \big(\text{circ}_1(\boldsymbol{U} \mid r), \dots, \text{circ}_{kr+r-1}(\boldsymbol{U} \mid r)\big) := (\boldsymbol{u}_1, \dots, \boldsymbol{u}_{kr}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_{r-1}) \in (\mathbb{R}^d)^{kr+r-1}.$$

Finally, we define $\text{circ-max}(\boldsymbol{U} \mid r) := \text{slid-max}(\text{circ}(\boldsymbol{U} \mid r) \mid r)$, that is,

$$\begin{aligned}\text{circ-max}(\boldsymbol{U} \mid r) &= \big(\text{circ-max}_1(\boldsymbol{U} \mid r), \dots, \text{circ-max}_{kr}(\boldsymbol{U} \mid r)\big)\\ &= \Big(\max_{s \in [1:r]} \text{circ}_s(\boldsymbol{U} \mid r), \max_{s \in [2:r+1]} \text{circ}_s(\boldsymbol{U} \mid r), \dots, \max_{s \in [kr:kr+r-1]} \text{circ}_s(\boldsymbol{U} \mid r) \Big) \in (\mathbb{R}^d)^{kr}.\end{aligned}$$

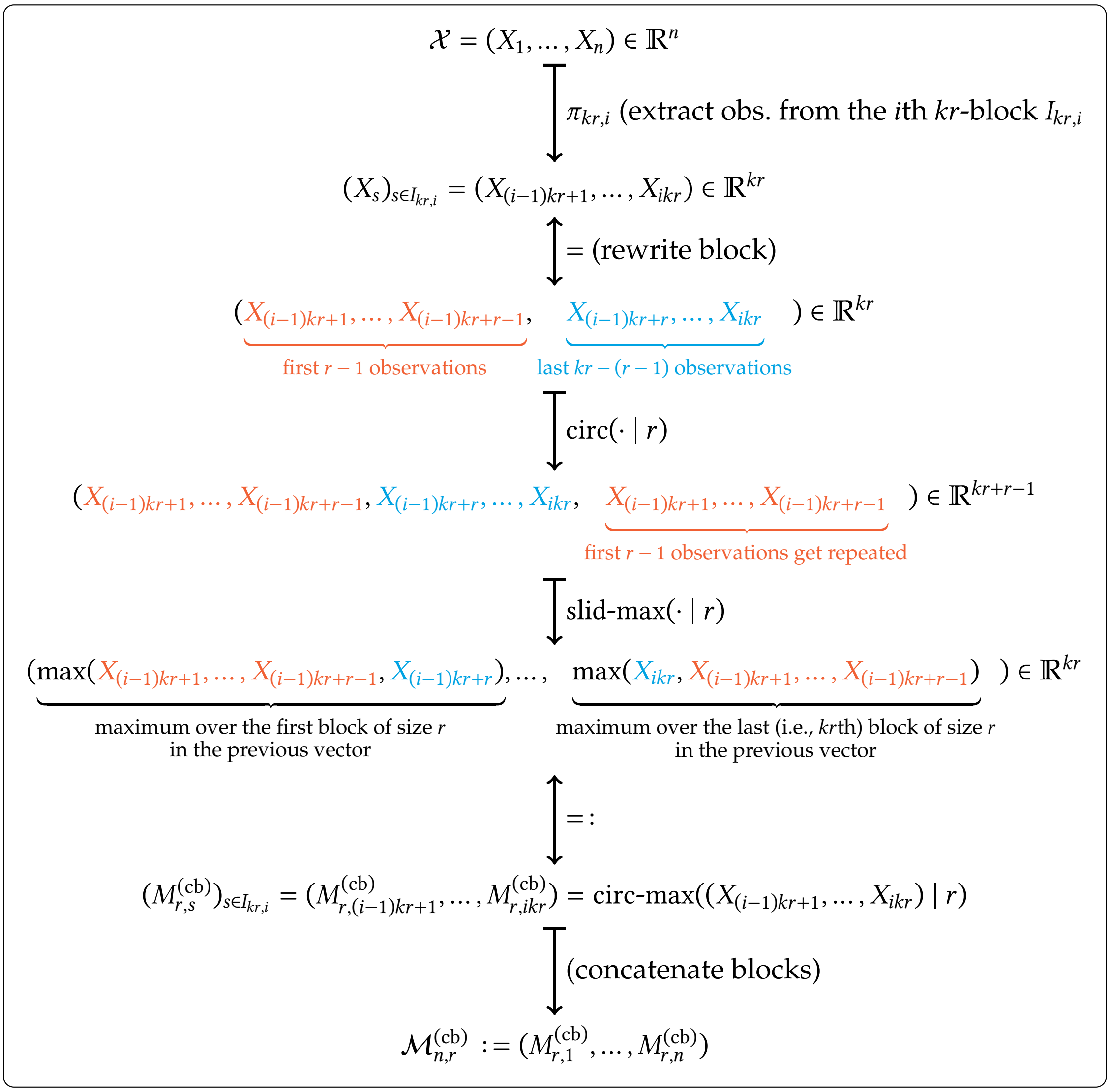

Figure 4.1: Illustration of calculating the circular block maxima sample.

It is constructive to take a closer look at the choices $k = 1$ and $k = n/r$ (the number of disjoint blocks). In the former case, the circmax-sample is the same as the disjoint block maxima sample, but with every observation repeated exactly r times. In the latter case, the first $n - r + 1$ observations of the circmax-sample coincide with the sliding block maxima sample. New samples are obtained for every other choice of k, and in the subsequent developments we will mostly be interested in choices like $k = 2$ or $k = 3$.

A key feature of the circmax-sample consists of the fact that it can be stored and evaluated efficiently. More precisely, for each kr block, the number of distinct circmax-values in that block is typically very small (typically around 10 for realistic block and sample sizes as observed in the simulation study); for instance, the largest observation in each block is necessarily appearing exactly r times. As such, the entire circmax-sample can also be regarded as a weighted sample whose (random) size corresponds to the number of distinct values in the circmax-sample.

The following central result shows that the circmax-sample can be considered as a sample from the 'correct'-limit distribution G from (4.2.2) (a similar result on joint convergence of two circmax-observations can be found in Proposition 4.11.1). For $s \in \{1, \dots, n\}$, define

$$Z_{r,s}^{(\mathrm{cb})} := \frac{\boldsymbol{M}_{r,s}^{(\mathrm{cb})} - \boldsymbol{b}_r}{\boldsymbol{a}_r}.$$

Note that we suppress the dependence on k for notational convenience.

Proposition 4.3.1 (Weak convergence of circular block maxima)**.** *Suppose that Condition 4.2.1 and 4.2.2(a) are met and that $\alpha(r_n) \to 0$. Then, for every fixed $k \in \mathbb{N}$ and $\xi \in [0, k)$, we have*

$$Z_{r,1+\lfloor \xi r \rfloor}^{(\mathrm{cb})} \rightsquigarrow Z \sim G \qquad (n \to \infty).$$

Due to Proposition 4.3.1, statistical methods based on the circular block maxima method may be expected to work, asymptotically. As discussed in Section 4.2.2, a central ingredient for studying respective methods is weak convergence of empirical means. Adopting the notation from that section, we denote the empirical processes associated with the normalized sample $Z_{r,1}^{(\mathrm{cb})}, \dots, Z_{r,n}^{(\mathrm{cb})}$ by

$$\bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})} = \sqrt{\frac{n}{r}}\big(\mathbb{P}_{n,r}^{(\mathrm{cb})} - P_{n,r}^{(\mathrm{cb})}\big), \quad \tilde{\mathbb{G}}_{n,r}^{(\mathrm{cb})} = \sqrt{\frac{n}{r}}\big(\mathbb{P}_{n,r}^{(\mathrm{cb})} - P_r\big), \quad \mathbb{G}_{n,r}^{(\mathrm{cb})} = \sqrt{\frac{n}{r}}\big(\mathbb{P}_{n,r}^{(\mathrm{cb})} - P\big), \qquad (4.3.1)$$

where $\mathbb{P}_{n,r}^{(\mathrm{cb})} = \frac{1}{n}\sum_{s=1}^{n} \delta_{Z_{r,s}^{(\mathrm{cb})}}$, $P_{n,r}^{(\mathrm{cb})} = \frac{1}{kr}\sum_{s=1}^{kr} \mathbb{P}(Z_{r,s}^{(\mathrm{cb})} \in \cdot)$, $P_r = \mathbb{P}(Z_{r,1} \in \cdot)$ and $P = \mathbb{P}(Z \in \cdot)$. The following result can be regarded as a circmax-counterpart of Theorem 4.2.3.

Theorem 4.3.2. *Suppose that Conditions 4.2.1 and 4.2.2 are met. Then, for fixed $k \in \mathbb{N}_{\ge 2}$ and any finite set of real valued functions $h_1, \dots, h_q$ satisfying the integrability Condition 4.10.1(b) with $\nu > 2/\omega$ with ω from Condition 4.2.2, we have*

$$\bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}\boldsymbol{h} \rightsquigarrow \mathcal{N}_q(\boldsymbol{0}, \Sigma_{\boldsymbol{h}}^{(\mathrm{sb})})$$

and $\lim_{n\to\infty} \mathrm{Cov}(\bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}\boldsymbol{h}) = \Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}$ *with* $\Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}$ *from (4.2.4). Moreover, if the bias Condition 4.10.3 holds, then*

$$\tilde{\mathbb{G}}_{n,r}^{(\mathrm{cb})}\boldsymbol{h} \rightsquigarrow \mathcal{N}_q\big(\tfrac{D_{\boldsymbol{h},k}+E_{\boldsymbol{h}}}{k}, \Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}\big)$$

(with $D_{\boldsymbol{h},k}$ often equal to zero, see Remark 4.3.3), and if additionally the bias Condition 4.10.2 is met, then

$$\mathbb{G}_{n,r}^{(\mathrm{cb})}\boldsymbol{h} \rightsquigarrow \mathcal{N}_q\big(B_{\boldsymbol{h}} + \tfrac{D_{\boldsymbol{h},k}+E_{\boldsymbol{h}}}{k}, \Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}\big).$$

Remark 4.3.3. The bias term $D_{\boldsymbol{h},k} + E_{\boldsymbol{h}}$ is due to the fact that the rescaled circular block maxima $\max(\boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}, \boldsymbol{X}_1, \dots, \boldsymbol{X}_t)$ with index $s = (k-1)r + 1 + t$ for some $t \in \{1, \dots, r-1\}$ are not equal in distribution to a generic block maximum variable $\boldsymbol{M}_r$. This bias has two sources: $D_{\boldsymbol{h},k}$ is a measure for the average (over t) difference induced by the approximation of $\max(\boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}, \boldsymbol{X}_1, \dots \boldsymbol{X}_t)$ by $\max(\boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_t)$ (with

$(\tilde{X}_t)_t$ an independent copy of $(\tilde{X}_t)_t$), while $E_{\boldsymbol{h}}$ is a measure for the average difference induced by the approximation of $\max(\boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_t)$ by that of a generic block maximum variable $\boldsymbol{M}_r$. In Lemma 4.10.4 we provide simple sufficient conditions for beta mixing time series that imply $D_{\boldsymbol{h},k} = B_{\boldsymbol{h}} = 0$.

The proof of Theorem 4.3.2 (see in particular Proposition 4.11.1) shows that the sample of kr-blocks containing the circular block maxima can be considered asymptotically independent. This is akin to the sample of disjoint block maxima and, as discussed after Theorem 4.2.3, it suggests to define bootstrap versions of $\mathbb{G}_{n,r}^{(\mathrm{cb})}$ by drawing $m(k)$-times from the sample of kr-blocks with replacement (which is computationally efficient: recall that all circular block maxima in a kr-block could be efficiently stored as a weighted sample with typically very few observations).

4.4 Bootstrapping block maxima estimators

Throughout this section, we discuss bootstrap approximations for the empirical processes from Section 4.3. For completeness, we also cover the disjoint block maxima case, as it can be identified with the circular block maxima sample for $k = 1$.

For fixed $k \in \mathbb{N}$, consider the circmax-sample $\mathcal{M}_{n,r}^{(\mathrm{cb})}$. Independent of the observations, let $\boldsymbol{W}_{m(k)} = (W_{m(k),1}, \dots, W_{m(k),m(k)}) = (W_1, \dots, W_{m(k)})$ be multinomially distributed with $m(k)$ trials and class probabilities $(m(k)^{-1}, \dots, m(k)^{-1})$. Given the sample $\mathcal{M}_{n,r}^{(\mathrm{cb})}$, the bootstrap sample $\mathcal{M}_{n,r}^{(\mathrm{cb}),*}$ is obtained by repeating the observations $(M_{r,s}^{(\mathrm{cb})})_{s \in I_{kr,i}}$ from the ith kr-block exactly $W_{m(k),i}$-times, for every $i = 1, \dots, m(k)$. The respective empirical measure of the rescaled bootstrap sample $Z_{r,1}^{(\mathrm{cb}),*}, \dots, Z_{r,n}^{(\mathrm{cb}),*}$ with $Z_{r,1}^{(\mathrm{cb}),*} = (\boldsymbol{M}_{r,s}^{(\mathrm{cb}),*} - \boldsymbol{b}_r)/\boldsymbol{a}_r$ can then be written as

$$\hat{\mathbb{P}}_{n,r}^{(\mathrm{cb}),*} = \frac{1}{n} \sum_{s=1}^{n} \delta_{Z_{r,s}^{(\mathrm{cb}),*}} = \frac{1}{n} \sum_{i=1}^{m(k)} W_{m(k),i} \sum_{s \in I_{kr,i}} \delta_{Z_{r,s}^{(\mathrm{cb})}},$$

and the associated (centered) empirical process is given by

$$\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*} = \sqrt{\frac{n}{r}} (\hat{\mathbb{P}}_{n,r}^{(\mathrm{cb}),*} - \mathbb{P}_{n,r}^{(\mathrm{cb})}). \tag{4.4.1}$$

Note that the conditional distribution of $\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}$ given the data $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ can in practice be approximated to an arbitrary precision based on repeated sampling of $\boldsymbol{W}_{m(k)}$ (if $\boldsymbol{a}_r, \boldsymbol{b}_r$ were known; we discuss extensions below). The following result shows that the bootstrap process $\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}$ provides a consistent distributional approximation for $\bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}$ and for $\bar{\mathbb{G}}_{n,r}^{(\mathrm{sb})}$.

Theorem 4.4.1 (Asymptotic validity of the circmax-resampling bootstrap)**.** *Suppose Conditions 4.2.1 and 4.2.2 are met. Then, for fixed $k \in \mathbb{N}_{\geq 2}$ and $\boldsymbol{h}$ satisfying the integrability Condition 4.10.1(b) with $\nu > 2/\omega$ with ω from Condition 4.2.2, we have, for* mb $\in \{\mathrm{sb}, \mathrm{cb}\}$*,*

$$d_K\Big(\mathcal{L}\big(\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}\boldsymbol{h} \mid \mathcal{X}_n\big), \mathcal{L}\big(\bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}\boldsymbol{h}\big)\Big) = o_{\mathbb{P}}(1)$$

as $n \to \infty$, where d_K denotes the Kolmogorov metric for probability measures on $\mathbb{R}^q$. As a consequence, if the bias Condition 4.10.2 is met and if r is chosen sufficiently large to guarantee that $B_{\boldsymbol{h}} = 0$, we have

$$d_K\Big(\mathcal{L}\big(\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}\boldsymbol{h} \mid \mathcal{X}_n\big), \mathcal{L}\big(\mathbb{G}_{n,r}^{(\mathrm{sb})}\boldsymbol{h}\big)\Big) = o_{\mathbb{P}}(1),$$

and if also the bias Condition 4.10.3 is met with $E_{\boldsymbol{h}} = D_{\boldsymbol{h},k} = 0$, we have

$$d_K\Big(\mathcal{L}\big(\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}\boldsymbol{h} \mid \mathcal{X}_n\big), \mathcal{L}\big(\mathbb{G}_{n,r}^{(\mathrm{cb})}\boldsymbol{h}\big)\Big) = o_{\mathbb{P}}(1).$$

Remarkably, the consistency of the circmax bootstrap for the sliding block maxima method does not require the bias Condition 4.10.3 (nor $E_{\boldsymbol{h}} = D_{\boldsymbol{h},k} = 0$ in case it is met).

Remark 4.4.2 (Bootstrapping the disjoint block maxima empirical process). Consider the disjoint block maxima sample, whose empirical means coincide with empirical means based on the circular block maxima sample with $k = 1$. The results from Theorem 4.4.1 continue to hold for $k = 1$ ($D_{\boldsymbol{h},k} = E_{\boldsymbol{h}} = 0$ is then immediate), but with 'sb' replaced by 'db' at all instances.

Remark 4.4.3 (Inconsistency of naive resampling of sliding block maxima). It seems natural to bootstrap sliding block maxima empirical means by the sliding block maxima analogue of (4.4.1), that is, by

$$\hat{\mathbb{G}}_{n,r}^{(\mathrm{sb}),*} = \sqrt{\frac{n}{r}}(\hat{\mathbb{P}}_{n,r}^{(\mathrm{sb}),*} - P_r), \qquad \hat{\mathbb{P}}_{n,r}^{(\mathrm{sb}),*} = \frac{1}{n}\sum_{s=1}^{n} \delta_{Z_{r,s}^{(\mathrm{sb}),*}} = \frac{1}{n}\sum_{i=1}^{m(k)} W_{m(k),i} \sum_{s \in I_{kr,i}} \delta_{Z_{r,s}^{(\mathrm{sb})}}.$$

Note that $\hat{\mathbb{G}}_{n,r}^{(\mathrm{sb}),*}$ depends on k, which is suppressed from the notation. However, unlike $\mathbb{G}_{n,r}^{(\mathrm{cb}),*}$, this process is inconsistent for both $\bar{\mathbb{G}}_{n,r}^{(\mathrm{sb})}$ and $\bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}$. Indeed, considering the case $d = q = 1$ for simplicity, we prove in the appendix that

$$d_K\Big(\mathcal{L}\big(\hat{\mathbb{G}}_{n,r}^{(\mathrm{sb}),*}h \mid \mathcal{X}_n\big), \mathcal{N}\big(\mathbf{0}, \Sigma_h^{(k)}\big)\Big) = o_{\mathbb{P}}(1), \tag{4.4.2}$$

where $\Sigma_h^{(k)} = \Sigma_h^{(\mathrm{sb})} - \frac{2}{k}\int_0^1 \xi \operatorname{Cov}(h(Z_{1,\xi}), h(Z_{2,\xi}))\,\mathrm{d}\xi$. If $\operatorname{Var}(h(Z)) > 0$, we have $\Sigma_h^{(k)} < \Sigma_h^{(\mathrm{sb})}$ for any $k \in \mathbb{N}$ by Lemma 4.11.4, so in view of Theorem 4.3.2 the bootstrap process has a smaller asymptotic variance than needed for approximating $\bar{\mathbb{G}}_{n,r}^{(\mathrm{sb})}$ or $\bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}$. This circumstance is one of the main motivations for working with circular block maxima.

In statistical applications, as discussed after Theorem 4.2.3, we do not want to bootstrap the estimation error of empirical means involving the (unobservable) $Z_{r,i}$, but the estimation error of general statistics depending on the block maxima $\boldsymbol{M}_{r,i}$ itself. We follow the general setting of Remark 4.2.4, that is, $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\varphi}_n(\boldsymbol{M}_{r,1}, \dots, \boldsymbol{M}_{r,n})$ is some estimator of interest for a target parameter $\boldsymbol{\theta}_r \in \mathbb{R}^p$. If similar equations/linearizations as in (4.2.7) and (4.2.8) can be shown to hold for the (unobservable) circmax-bootstrap counterpart $\hat{\boldsymbol{\vartheta}}_n^{(\mathrm{cb}),*} = \boldsymbol{\psi}_n(Z_{1,r}^{(\mathrm{cb}),*}, \dots, Z_{n,r}^{(\mathrm{cb}),*})$, that is, if

$$\hat{\boldsymbol{\theta}}_n^{(\mathrm{cb}),*} - \hat{\boldsymbol{\theta}}_n^{(\mathrm{cb})} = A(\boldsymbol{a}_r, \boldsymbol{b}_r)(\hat{\boldsymbol{\vartheta}}_n^{(\mathrm{cb}),*} - \hat{\boldsymbol{\vartheta}}_n^{(\mathrm{cb})}), \tag{4.4.3}$$

$$\sqrt{\frac{n}{r}}(\hat{\boldsymbol{\vartheta}}_n^{(\mathrm{cb}),*} - \hat{\boldsymbol{\vartheta}}_n^{(\mathrm{cb})}) = \hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}(h_1, \dots, h_q)^\top + o_{\mathbb{P}}(1) \tag{4.4.4}$$

then, loosely speaking, the distributional approximation $\mathbb{G}_{n,r}^{(\mathrm{cb})}\boldsymbol{h} \approx_d (\mathbb{G}_{n,r}^{(\mathrm{cb}),*}\boldsymbol{h} \mid \mathcal{X}_n)$ from Theorem 4.4.1 carries over to the distributional approximation $\hat{\boldsymbol{\theta}}_n^{(\mathrm{cb})} - \boldsymbol{\theta}_r \approx_d (\hat{\boldsymbol{\theta}}_n^{(\mathrm{cb}),*} - \hat{\boldsymbol{\theta}}_n^{(\mathrm{cb})} \mid \mathcal{X}_n)$. Formally, we have the following result.

Proposition 4.4.4. *Fix $k \in \mathbb{N}_{\ge 2}$ and assume that Conditions 4.2.1 and 4.2.2 are met. Suppose $\hat{\boldsymbol{\theta}}_n^{(\mathrm{mb})} = \boldsymbol{\varphi}_n(\boldsymbol{M}_{r,1}^{(\mathrm{mb})}, \dots, \boldsymbol{M}_{r,n}^{(\mathrm{mb})})$ is some estimator of interest for a target parameter $\boldsymbol{\theta}_r \in \mathbb{R}^p$ such that (4.2.5), (4.2.6) and (4.2.8) (with $\bar{\mathbb{G}}_{n,r} = \bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}$) is met for some function $\boldsymbol{h} = (h_1, \dots, h_q)^\top$ satisfying the conditions of Theorem 4.3.2. Moreover, assume that (4.4.3) and (4.4.4) is met. Then, if $\Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}$ is invertible, for* mb $\in$ {sb, cb}*,*

$$d_K\Big(\mathcal{L}\big(\hat{\boldsymbol{\theta}}_n^{(\mathrm{cb}),*} - \hat{\boldsymbol{\theta}}_n^{(\mathrm{cb})} \mid \mathcal{X}_n\big), \mathcal{L}\big(\hat{\boldsymbol{\theta}}_n^{(\mathrm{mb})} - \boldsymbol{\theta}_r\big)\Big) = o_{\mathbb{P}}(1).$$

Proof. By Equations (4.2.5), (4.2.6), (4.2.8) and (4.4.3), (4.4.4) we have

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_n^{(\mathrm{mb})} - \boldsymbol{\theta}_r &= A(\boldsymbol{a}_r, \boldsymbol{b}_r)\sqrt{\frac{r}{n}}\big(\bar{\mathbb{G}}_{n,r}^{(\mathrm{mb})}\boldsymbol{h} + o_{\mathbb{P}}(1)\big) =: \sqrt{\frac{r}{n}}A(\boldsymbol{a}_r, \boldsymbol{b}_r)S_n, \\
\hat{\boldsymbol{\theta}}_n^{(\mathrm{cb}),*} - \hat{\boldsymbol{\theta}}_n^{(\mathrm{cb})} &= A(\boldsymbol{a}_r, \boldsymbol{b}_r)\sqrt{\frac{r}{n}}\big(\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb},*)}\boldsymbol{h} + o_{\mathbb{P}}(1)\big) =: \sqrt{\frac{r}{n}}A(\boldsymbol{a}_r, \boldsymbol{b}_r)S_n^*.
\end{aligned}$$

By Theorems 4.2.3 and 4.3.2 we have $S_n \rightsquigarrow \mathcal{N}_q(\boldsymbol{0}, \Sigma_{\boldsymbol{h}}^{(\mathrm{sb})})$. Theorem 4.4.1 implies $d_K(\mathcal{L}(S_n^* \mid \mathcal{X}_n), \mathcal{L}(S_n)) = o_{\mathbb{P}}(1)$. The assertion then follows from Lemma 4.11.9. □

The results from Proposition 4.4.4 are sufficient for showing that basic bootstrap confidence intervals (Davison and Hinkley, 1997) asymptotically hold their intended level. As a proof of concept, details on a specific example are given in the next section, see Corollary 4.5.6.

4.5 Application: bootstrapping the pseudo-maximum likelihood estimator for the Fréchet distribution

In this section, we provide details how the previous methods and results can be used for a specific important estimation problem. More precisely, we consider the univariate heavy tailed case, and restrict attention to the maximum likelihood estimator for parameters associated with a suitable version of the max-domain of attraction Condition 4.2.1.

Condition 4.5.1 (Fréchet Max-Domain of Attraction)**.** Let $(X_t)_{t \in \mathbb{Z}}$ denote a strictly stationary univariate time series with continuous margins. There exists some $\alpha_0 > 0$ and some sequence $(\sigma_r)_{r \in \mathbb{N}} \subset (0, \infty)$ such that

$$\lim_{r \to \infty} \frac{\sigma_{\lfloor rs \rfloor}}{\sigma_r} = s^{1/\alpha_0}, \ (s > 0) \quad \text{and} \quad \frac{\max(X_1, \dots, X_r)}{\sigma_r} \rightsquigarrow Z \sim P_{\alpha_0,1} \qquad (r \to \infty),$$

where $P_{\alpha,\sigma}$ denotes the Fréchet-scale family on $(0, \infty)$ defined by its cdf $F_{\alpha,\sigma}(x) = \exp(-(x/\sigma)^{-\alpha})$; here $(\alpha, \sigma) \in (0, \infty)^2$.

Suppose $X_1, \dots, X_n$ is an observed sample from $(X_t)_{t\in\mathbb{Z}}$ as in Condition 4.5.1, and let $r = r_n$ denote a block length parameter. In view of the heuristics and results from Sections 4.2 and 4.3, the associated block maxima samples $\mathcal{M}_{n,r}^{(\mathrm{mb})}$ with $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}\}$ (with $k \in \mathbb{N}$ fixed) can all be considered approximate samples from P_{α_0,σ_r}. As in Bücher and Segers (2018b,a), this suggests to estimate (α_0, σ_r) by maximizing the independence Fréchet-log-likelihood, that is, we define

$$\hat{\boldsymbol{\theta}}_n^{(\mathrm{mb})} := (\hat{\alpha}_n^{(\mathrm{mb})}, \hat{\sigma}_n^{(\mathrm{mb})})^\top := \underset{\boldsymbol{\theta}=(\alpha,\sigma)\in(0,\infty)^2}{\arg\max} \sum_{M_i \in \mathcal{M}_{n,r}^{(\mathrm{mb})}} \ell_{\boldsymbol{\theta}}(M_i \vee c), \tag{4.5.1}$$

where, for $\boldsymbol{\theta} = (\alpha, \sigma)^\top \in (0,\infty)^2$,

$$\ell_{\boldsymbol{\theta}}(x) = \log(\alpha/\sigma) - (x/\sigma)^{-\alpha} - (\alpha+1)\log(x/\sigma), \quad x > 0, \tag{4.5.2}$$

denotes the log density of the Fréchet distribution $P_{(\alpha,\sigma)}$ and where $c > 0$ denotes an arbitrary truncation constant. For the case $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$, the rescaled estimation error of $\hat{\boldsymbol{\theta}}_n^{(\mathrm{mb})}$ is known to be asymptotically normal (and independent of c), that is, under suitable additional assumptions,

$$\sqrt{\frac{n}{r}} \begin{pmatrix} \hat{\alpha}_n^{(\mathrm{mb})} - \alpha_0 \\ \hat{\sigma}_n^{(\mathrm{mb})}/\sigma_{r_n} - 1 \end{pmatrix} \rightsquigarrow \mathcal{N}_2(\boldsymbol{\mu}, \Sigma^{(\mathrm{mb})}), \tag{4.5.3}$$

for some $\boldsymbol{\mu} \in \mathbb{R}^2$ and some $\Sigma^{(\mathrm{mb})} \in \mathbb{R}^{2\times 2}$ positive definite, see Bücher and Segers (2018b,a). Subsequently, we will extend these results to $\mathrm{mb} = \mathrm{cb}$, and we will derive a consistent bootstrap scheme for the estimation error.

For fixed $k \in \mathbb{N}$, consider the circmax-sample $\mathcal{M}_{n,r}^{(\mathrm{cb})}$. Independent of the observations, let $\boldsymbol{W}_{m(k)} = (W_{m(k),1}, \dots, W_{m(k),m(k)}) = (W_1, \dots, W_{m(k)})$ be multinomially distributed with $m(k)$ trials and class probabilities $(m(k)^{-1}, \dots, m(k)^{-1})$. As in Section 4.4, conditional on $\mathcal{M}_{n,r}^{(\mathrm{cb})}$, the bootstrap sample $\mathcal{M}_{n,r}^{(\mathrm{cb}),*}$ is obtained by repeating the observations $(M_{r,s}^{(\mathrm{cb})})_{s\in I_{kr,i}}$ from the ith kr-block exactly $W_{m(k),i}$-times, for every $i = 1, \dots, m(k)$. We are going to show that the independence Fréchet-log-likelihood

$$\boldsymbol{\theta} \mapsto \sum_{s=1}^{n} \ell_{\boldsymbol{\theta}}(M_{r,s}^{(\mathrm{cb}),*} \vee c) = \sum_{i=1}^{m(k)} W_{m(k),i} \sum_{s\in I_{kr,i}} \ell_{\boldsymbol{\theta}}(M_{r,s}^{(\mathrm{cb})} \vee c).$$

has a unique-maximimizer (with probability converging to one), say $(\hat{\alpha}_n^{(\mathrm{cb}),*}, \hat{\sigma}_n^{(\mathrm{cb}),*})$, and that the conditional distribution of the rescaled bootstrap-estimation error, $\sqrt{n/r}(\hat{\alpha}_n^{(\mathrm{cb}),*} - \hat{\alpha}_n^{(\mathrm{cb})}, \hat{\sigma}_n^{(\mathrm{cb}),*}/\hat{\sigma}_n - 1)$, given the observations is close to the distribution of $\sqrt{n/r}(\hat{\alpha}_n^{(\mathrm{mb})} - \alpha_0, \hat{\sigma}_n^{(\mathrm{mb})}/\sigma_{r_n} - 1)$ for both $\mathrm{mb} = \mathrm{sb}$ and $\mathrm{mb} = \mathrm{cb}$. We also show how this result can be used to derive valid asymptotic confidence intervals.

A couple of conditions akin to those imposed in Bücher and Segers (2018b,a) will be needed.

Condition 4.5.2 (All disjoint block maxima of size $\lfloor r_n/2 \rfloor$ diverge)**.** For all $c > 0$, the probability of the event that all disjoint block maxima of block size $\tilde{r}_n = \lfloor r_n/2 \rfloor$ are larger than c converges to one.

This is Condition 2.2 in Bücher and Segers (2018a), and guarantees that the probability of the event that all (blocksize r_n) block maxima in $\mathcal{M}_{n,r}^{(\mathrm{mb})}$ are larger than c converges to one as well, for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}\}$. This will guarantee that the Fréchet-log-likelihoods under consideration are well-defined, with probability converging to one. The following condition is Condition 3.4 in Bücher and Segers (2018b); recall that $M_{r,1} = \max(X_1, \dots, X_r)$.

Condition 4.5.3 (Moments)**.** There exists $\nu > 2/\omega$ with ω from Condition 4.2.2 such that

$$\limsup_{r\to\infty} \mathrm{E}[g_{\nu,\alpha_0}\big((M_{r,1} \vee 1)/\sigma_r\big)] < \infty,$$

where $g_{\nu,\alpha_0}(x) := \Big[x^{-\alpha_0}\mathbf{1}\{x \le e\} + \log(x)\mathbf{1}\{x > e\}\Big]^{2+\nu}$.

Condition 4.5.4 (Bias)**.** There exists $c_0 > 0$ such that, for $j \in \{1,2,3\}$, the limits $C(f_j) := \lim_{n\to\infty} C_n(f_j)$ and $C_k(f_j) := \lim_{n\to\infty} C_{n,k}(f_j)$ exist, where

$$C_n(f_j) := \sqrt{\frac{n}{r}}\left(\mathrm{E}\left[f_j\big((M_{r,1} \vee c_0)/\sigma_r\big)\right] - P_{\alpha_0,1} f_j\right)$$
$$C_{n,k}(f_j) := \sqrt{\frac{n}{r}}\left(\frac{1}{r}\sum_{s=(k-1)r+1}^{kr}\left\{\mathrm{E}\left[f_j\big((M_{r,s} \vee c_0)/\sigma_r\big)\right] - \mathrm{E}\left[f_j\big((M_{r,1} \vee c_0)/\sigma_r\big)\right]\right\}\right)$$

where $f_j : (0,\infty) \to \mathbb{R}$ are defined as

$$f_1(x) = x^{-\alpha_0}, \quad f_2(x) = x^{-\alpha_0}\log x, \quad f_3(x) = \log x. \tag{4.5.4}$$

The first part of this condition on C_n is Condition 3.5 in Bücher and Segers (2018b), and provides control of the bias for those block maxima which are calculated based on r successive observations. For the circmax-sample, the last r observations in a kr-block of that sample are not of that form; their contribution to the bias is controlled by the condition on $C_{n,k}$.

Subsequently, we fix a truncation constant $c > 0$, and we write

$$\mathbb{G}_{n,r}^{(\mathrm{cb})} = \sqrt{\frac{n}{r}}\big(\mathbb{P}_{n,r}^{(\mathrm{cb})} - P_{(\alpha_0,1)}\big), \quad \hat{\mathbb{G}}_n^{(\mathrm{cb}),*} = \sqrt{\frac{n}{r}}(\hat{\mathbb{P}}_{n,r}^{(\mathrm{cb}),*} - \mathbb{P}_{n,r}^{(\mathrm{cb})})$$

where

$$\mathbb{P}_{n,r}^{(\mathrm{cb})} = \frac{1}{n}\sum_{s=1}^{n}\delta_{(M_{r,s}^{(\mathrm{cb})} \vee c)/\sigma_r}, \qquad \hat{\mathbb{P}}_{n,r}^{(\mathrm{cb}),*} = \frac{1}{n}\sum_{i=1}^{m(k)} W_{m(k),i}\sum_{s\in I_{kr,i}}\delta_{(M_{r,s}^{(\mathrm{cb})} \vee c)/\sigma_r};$$

the double use of notation with (4.3.1) and (4.4.1) should not cause any confusion. Furthermore, we suppress the dependence on c in the notation, which is motivated by the fact that the limiting distribution does not depend on c, as shown in the following theorem.

Theorem 4.5.5. *Fix $k \in \mathbb{N}_{\geq 2}$ and $c > 0$, and suppose that Conditions 4.2.2 and 4.5.1–4.5.4 are satisfied. Then, with probability tending to one, there exists a unique maximizer $(\hat\alpha_n^{(\mathrm{cb})}, \hat\sigma_n^{(\mathrm{cb})})$ of the Fréchet log-likelihood $\boldsymbol\theta \mapsto \sum_{i=1}^n \ell_{\boldsymbol\theta}(M_{r,s}^{(\mathrm{cb})} \vee c)$, and this maximizer satisfies*

$$\sqrt{\frac{n}{r}} \begin{pmatrix} \hat\alpha_n^{(\mathrm{cb})} - \alpha_0 \\ \hat\sigma_n^{(\mathrm{cb})}/\sigma_{r_n} - 1 \end{pmatrix} = M(\alpha_0)\mathbb{G}_{n,r}^{(\mathrm{cb})}(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1) \rightsquigarrow \mathcal{N}_2(M(\alpha_0)(\boldsymbol C + k^{-1}\boldsymbol C_k), \Sigma^{(\mathrm{sb})}),$$

where $\boldsymbol C = (C(f_1), C(f_2), C(f_3))^\top, \boldsymbol C_k = (C_k(f_1), C_k(f_2), C_k(f_3))^\top$ and

$$M(\alpha_0) = \frac{6}{\pi^2}\begin{pmatrix} \alpha_0^2 & \alpha_0(1-\gamma) & -\alpha_0^2 \\ \gamma - 1 & -(\Gamma''(2)+1)/\alpha_0 & 1-\gamma \end{pmatrix}, \quad \Sigma^{(\mathrm{sb})} = \begin{pmatrix} 0.4946\alpha_0^2 & -0.3236 \\ -0.3236 & 0.9578\alpha_0^{-2} \end{pmatrix}$$

with $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}t$ the Euler Gamma function and $\gamma = 0.5772\ldots$ the Euler-Mascheroni constant.

Moreover, also with probability tending to one, there exists a unique maximizer $(\hat\alpha_n^{(\mathrm{cb}),*}, \hat\sigma_n^{(\mathrm{cb}),*})$ of $\boldsymbol\theta \mapsto \sum_{s=1}^n \ell_{\boldsymbol\theta}(M_{r,s}^{(\mathrm{cb}),*} \vee c)$, and this maximizer satisfies

$$\sqrt{\frac{n}{r}} \begin{pmatrix} \hat\alpha_n^{(\mathrm{cb}),*} - \hat\alpha_n^{(\mathrm{cb})} \\ \hat\sigma_n^{(\mathrm{cb}),*}/\hat\sigma_n^{(\mathrm{cb})} - 1 \end{pmatrix} = M(\alpha_0)\hat{\mathbb{G}}_{n,r}^{(\mathrm{cb}),*}(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1) \rightsquigarrow \mathcal{N}_2(0, \Sigma^{(\mathrm{sb})}).$$

Fix $\mathrm{mb} \in \{\mathrm{sb}, \mathrm{cb}\}$. If r_n is chosen sufficiently large such that, for $j \in \{1,2,3\}$, $C(f_j) = 0$ ($\mathrm{mb} = \mathrm{sb}$) or $C(f_j) = C_k(f_j) = 0$ ($\mathrm{mb} = \mathrm{cb}$) in Condition 4.5.4, we have bootstrap consistency in the following sense:

$$d_K\left[\mathcal{L}\left(\sqrt{\frac{n}{r}} \begin{pmatrix} \hat\alpha_n^{(\mathrm{cb}),*} - \hat\alpha_n^{(\mathrm{cb})} \\ \hat\sigma_n^{(\mathrm{cb}),*}/\hat\sigma_n^{(\mathrm{cb})} - 1 \end{pmatrix} \,\Big|\, \mathcal{X}_n\right), \sqrt{\frac{n}{r}} \begin{pmatrix} \hat\alpha_n^{(\mathrm{mb})} - \alpha_0 \\ \hat\sigma_n^{(\mathrm{mb})}/\sigma_{r_n} - 1 \end{pmatrix}\right] = o_{\mathbb{P}}(1).$$

The result in Theorem 4.5.5 allows for statistical inference on the parameters α_0, σ_r, for instance in the form of confidence intervals. We provide details on σ_r, using the circular block bootstrap approximation to the sliding block estimator: for $\beta \in (0,1)$, let $q_{\hat\sigma_n^*}(\beta)$ denote the β-quantile of the conditional distribution of $\hat\sigma_n^{(\mathrm{cb}),*}$ given the data, that is, $q_{\hat\sigma_n^*}(\beta) = (F_{\hat\sigma_n^*})^{-1}(\beta)$, where $F_{\hat\sigma_n^*}(x) = \mathbb{P}(\hat\sigma_n^{(\mathrm{cb}),*} \leq x \mid \mathcal{X}_n)$ for $x \in \mathbb{R}$. Note that the quantile may be approximated to an arbitrary precision by repeated bootstrap sampling. Consider the basic bootstrap confidence interval (Davison and Hinkley, 1997)

$$I_{n,\sigma}^{(\mathrm{sb,cb})}(1-\beta) = \left[2\hat\sigma_n^{(\mathrm{sb})} - q_{\hat\sigma_n^*}(1-\tfrac{\beta}{2}), 2\hat\sigma_n^{(\mathrm{sb})} - q_{\hat\sigma_n^*}(\tfrac{\beta}{2})\right].$$

Corollary 4.5.6. *Under the conditions of Theorem 4.5.5 with r_n chosen sufficiently large such that, for $j \in \{1,2,3\}$, $C(f_j) = 0$ in Condition 4.5.4, we have, for any $\beta \in (0,1)$,*

$$\lim_{n\to\infty} \mathbb{P}(\sigma_{r_n} \in I_{n,\sigma}^{(\mathrm{sb,cb})}(1-\beta)) = 1-\beta.$$

The above result only concerns the sliding block maxima estimator, but an analogous result can be derived for the disjoint block maxima estimator and under additional bias conditions for the circmax estimator. In view of the fact that the disjoint block maxima estimator exhibits a larger asymptotic estimation variance, the width of the disjoint blocks confidence interval is typically larger than the one of the circmax method, for every fixed confidence level (for an empirical illustration, see Figure 4.8 for the related problem of providing a confidence interval for the shape parameter α).

4.6 Simulation study

The finite-sample properties of the circmax estimators and of the bootstrap approaches have been investigated in a large scale Monte-Carlo simulation study. Two asymptotic regimes were considered: first, the case where the block size r is fixed and the number of disjoint blocks is increasing, and second, the case where n is fixed and the block size is treated as a tuning parameter. All performance metrics in the subsequent paragraphs have been empirically approximated based on $N = 5,000$ simulation repetitions.

4.6.1 Fixed block size

Throughout, we fix the block size r, and consider the estimation of target parameters associated with the law of $\boldsymbol{M}_r$, as well as bootstrap approximations for the respective estimation error. For illustrative purposes, we restrict attention to the one dimensional case and the most vanilla target parameter: the expected value $\mu_r = \mathrm{E}[M_r]$ of the block maximum distribution. The respective disjoint, sliding and circular block maxima estimators are $\hat{\mu}_n^{(\mathrm{mb})} := n^{-1}\sum_{i=1}^n M_{r,i}^{(\mathrm{mb})}$ with mb $\in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\}$, where $\mathrm{cb}(k)$ denotes the circmax estimator with parameter $k \in \{2, 3\}$. The finite-sample performance is assessed for the following time series model.

Model 4.6.1 (ARMAX-GPD-Model). $(X_t)_{t\in\mathbb{Z}}$ is a stationary real-valued time series whose stationary distribution is the generalized Pareto distribution $\mathrm{GPD}(0, 1, \gamma)$ with c.d.f.

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

After transformation to the Fréchet(1)-scale, the temporal dynamics correspond to the ARMAX(1)-model (Example 10.3 in Beirlant et al., 2004). More precisely, the time series $(Y_t)_t$ with $Y_t = F_W^{\leftarrow}(F_\gamma(X_t))$ satisfies the recursion

$$Y_t = \max\left(\beta Y_{t-1}, (1-\beta)W_t\right) \quad \forall t \in \mathbb{Z} \tag{4.6.1}$$

for an i.i.d. sequence $(W_t)_{t\in\mathbb{Z}}$ of Fréchet(1)-distributed random variables and some $\beta \in [0, 1)$; here, F_W denotes the c.d.f. of a Fréchet(1) distributed random variable and $F^{\leftarrow}$ denotes the generalised inverse of F. Throughout the simulation study, we consider the choices $\beta = 0$ (iid case) and $\beta = 0.5$ combined with $\gamma \in \{-0.2, 0, 0.2\}$, giving a total of six different models.

Note in passing that $(X_t)_t$ from Model 4.6.1 is exponentially beta mixing and satisfies Condition 4.2.1 with $Z \sim \mathrm{GEV}(\gamma)$, $a_r = \{r(1-\beta)\}^\gamma$ and $b_r = \{r(1-\beta)^\gamma - 1\}/\gamma$. Moreover, the mean and variance of M_r exist for $\gamma < 1$ and $\gamma < 1/2$, respectively, and in the latter case an application of Theorems 4.2.3 and 4.3.2 shows that, for $r = r_n \to \infty$ such that $r = o(n)$

and $\log(n) = o(r^{1/2})$,

$$\frac{\sqrt{n/r}}{(r(1-\beta))^\gamma}\big(\hat\mu_n^{(\mathrm{mb})} - \mu_r\big) \rightsquigarrow \mathcal{N}(0, \sigma^2_{\mathrm{mb}}), \quad \mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\},$$

where the asymptotic variance is given by

$$\sigma^2_{\mathrm{db}} = \begin{cases} \frac{g_2 - g_1^2}{\gamma^2}, & \gamma < 1/2, \gamma \neq 0, \\ \frac{\pi^2}{6}, & \gamma = 0, \end{cases} \qquad \sigma^2_{\mathrm{cb}(k)} = \sigma^2_{\mathrm{sb}} = \begin{cases} 4\Gamma(-2\gamma)I(\gamma), & \gamma < 0, \\ 4(\log 4 - 1), & \gamma = 0, \\ -\frac{g_2}{\gamma} I(\gamma) & 0 < \gamma < 1/2. \end{cases} \tag{4.6.2}$$

Here, Γ denote the gamma function, $g_j = \Gamma(1 - j\gamma)$ and $I(\gamma) := 2\int_0^{1/2} \{\alpha_{2\gamma}(w) - 1\}\{w^{-\gamma-1}(1-w)^{-\gamma-1}\}\,\mathrm{d}w$ with $\alpha_c(w) = w^{-1}\int_0^w (1-z)^c\,\mathrm{d}z$. Details are provided in Section 4.12.1.

For the simulation experiments, the block size was fixed to $r = 90$ as it roughly corresponds to the number of days within a season - a common block size in environmental applications. The effective sample size (i.e., the number of disjoint blocks of size r, abbreviated by $m = \lfloor n/r \rfloor$ hereafter) has been varied between 50 and 100 resulting in total sample sizes n between 4,500 and 9,000.

Performance of the estimators. We start by comparing the four estimators in terms of their variance, squared bias and mean squared error (MSE). For assessing the bias, the true first moment μ_r was approximated by a preliminary Monte Carlo simulation based on 10^6 independent block maxima from Model 4.6.1 with block size $r = 90$. The results are summarized in Figure 4.2. As suggested by the theory, both the sliding blocks estimator and the two circular blocks estimators perform uniformly better than the respective disjoint blocks estimator. The improvement gets smaller as the observations are getting more heavy-tailed; a phenomenon that has already been observed in the literature; see e.g. Bücher and Zanger (2023), Bücher and Staud (2024b). It is noteworthy that the additional bias effect introduced in the circular blocks estimators does not contribute in a relevant way; its contribution to the overall MSE was found to be at most 0.2% and often much smaller. Furthermore, the serial dependence does not change the qualitative results significantly.

Performance of the bootstrap. We consider each of the bootstrap estimators $\hat\mu_n^{(\mathrm{mb}),*}$ with $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\}$, with number of bootstrap replications set to $B = 1,000$. Recall that our results from the previous sections imply that the conditional bootstrap estimation errors $\hat\mu_n^{(\mathrm{cb}(2)),*} - \hat\mu_n^{(\mathrm{cb}(2))}$ and $\hat\mu_n^{(\mathrm{cb}(3)),*} - \hat\mu_n^{(\mathrm{cb}(3))}$ are consistent for the estimation error $\hat\mu_n^{(\mathrm{sb})} - \mu_r$, that $\hat\mu_n^{(\mathrm{db}),*} - \hat\mu_n^{(\mathrm{db})}$ is consistent for $\hat\mu_n^{(\mathrm{db})} - \mu_r$ and that $\hat\mu_n^{(\mathrm{sb}),*} - \hat\mu_n^{(\mathrm{sb})}$ is inconsistent for $\hat\mu_n^{(\mathrm{sb})} - \mu_r$. All four statements are illustrated in Figure 4.3 by means of QQ-plots and histograms for the parameter choice $\gamma = -0.2, \beta = 0.5$ and $m = 80$; other choices lead to similar results.

In the next step, we evaluate the performance of the four bootstrap approaches in terms of their ability to provide accurate estimates of the estimation variance $\sigma^2_{\mathrm{mb}}(r) =$

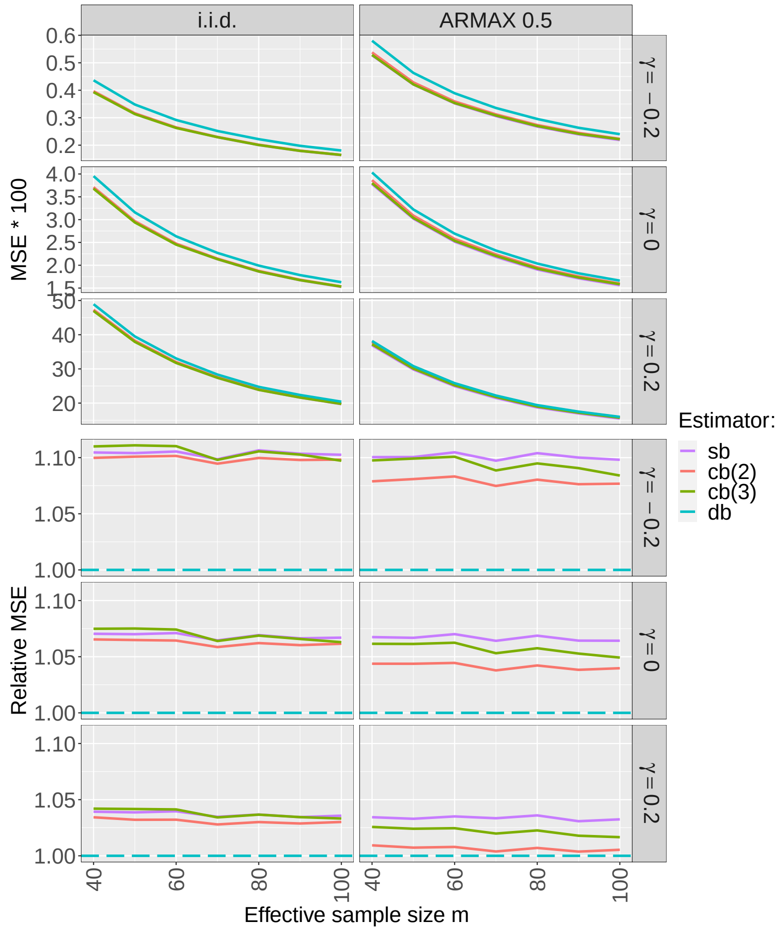

Figure 4.2: Mean estimation with fixed block size $r = 90$. Top three rows: Mean squared error $\mathrm{MSE}(\hat{\mu}_n^{(\mathrm{mb})})$. Bottom three rows: relative MSE with respect to the disjoint blocks method, i.e., $\mathrm{MSE}(\hat{\mu}_n^{(\mathrm{mb})})/\mathrm{MSE}(\hat{\mu}_n^{(\mathrm{db})})$.

$\mathrm{Var}(\hat{\mu}_n^{(\mathrm{mb})})$ with $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$. For that purpose, for each fixed sample of size n, we estimate the respective variance by the empirical variance of the sample of bootstrap estimates; recall that each such sample is of size $B = 1{,}000$, for every $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\}$. In Figure 4.4, we depict the average over the $N = 5{,}000$ bootstrap estimates, along with the true parameters determined in a presimulation based on 10^6 repetitions. We observe that the bootstrap estimates are close to the their target values for the disjoint method and for both circmax methods, while the naive sliding blocks bootstrap underestimates the true variance substantially, as mathematically demonstrated in Remark 4.4.3.

Finally, we evaluate the performance of the bootstrap approaches in terms of their ability to provide accurate confidence intervals of pre-specified coverage; clearly, smaller intervals of similar coverage would be preferable. For that purpose, we restrict ourselves to the basic bootstrap confidence interval (Davison and Hinkley, 1997) as illustrated in Corollary 4.5.6. The empirical coverage and the average widths of the respective intervals are depicted in Figure 4.5, where we omit the naive sliding method

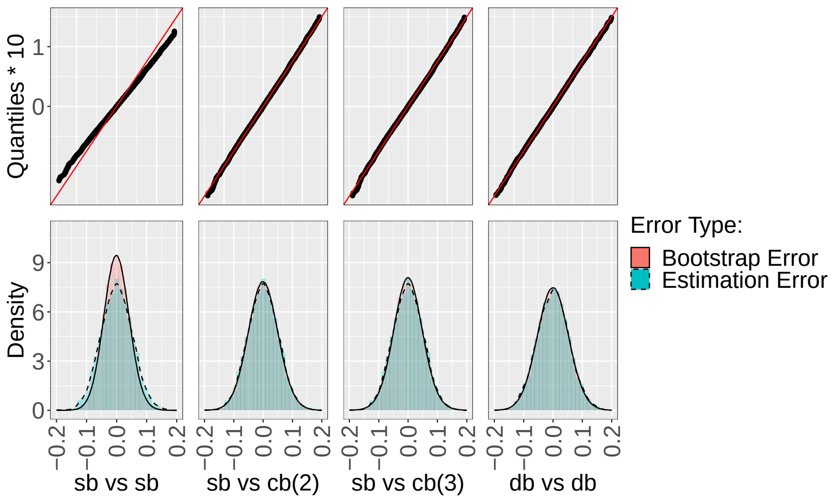

Figure 4.3: Estimation error $\hat{\mu}_n^{(\mathrm{mb})} - \mu_r$ vs. bootstrap error $\hat{\mu}_n^{(\mathrm{mb}'),*} - \hat{\mu}_n^{(\mathrm{mb}')}$ for mean estimators with fixed block size $r = 90$. Top row: qq-plots. Bottom row: histograms and kernel density estimates. Underlying data from Model 4.6.1 with $\gamma = -0.2, \beta = 0.5$ and $m = 80$.

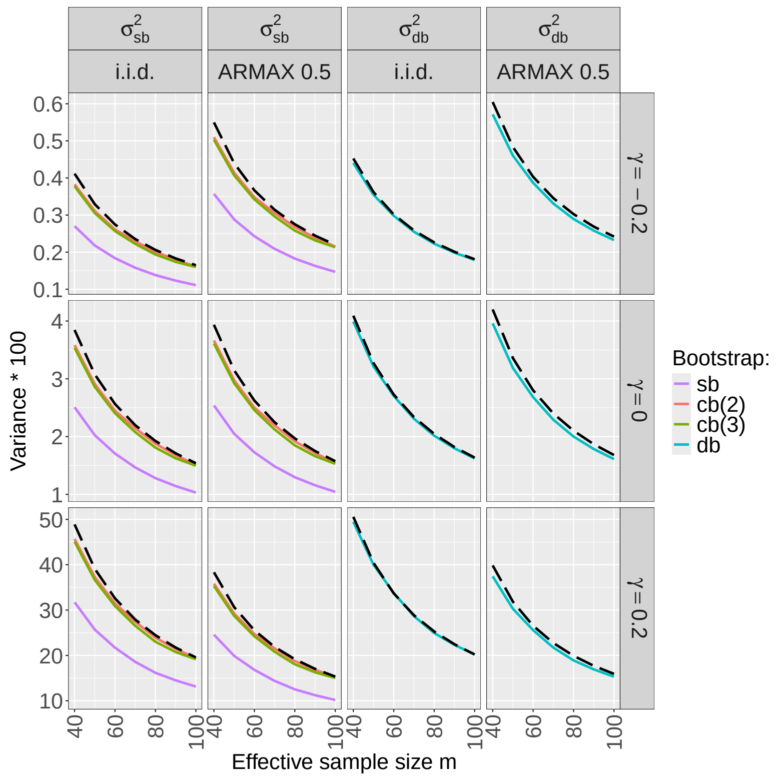

Figure 4.4: Bootstrap-based estimation of the estimation variance $\sigma^2_{\mathrm{mb}}(r) = \mathrm{Var}(\hat{\mu}_n^{(\mathrm{mb})})$ with fixed block size $r = 90$. Left two columns: mb = sb; the dashed line corresponds to $\sigma^2_{\mathrm{sb}}(r)$. Right two columns: mb = db; the dashed line corresponds to $\sigma^2_{\mathrm{db}}(r)$.

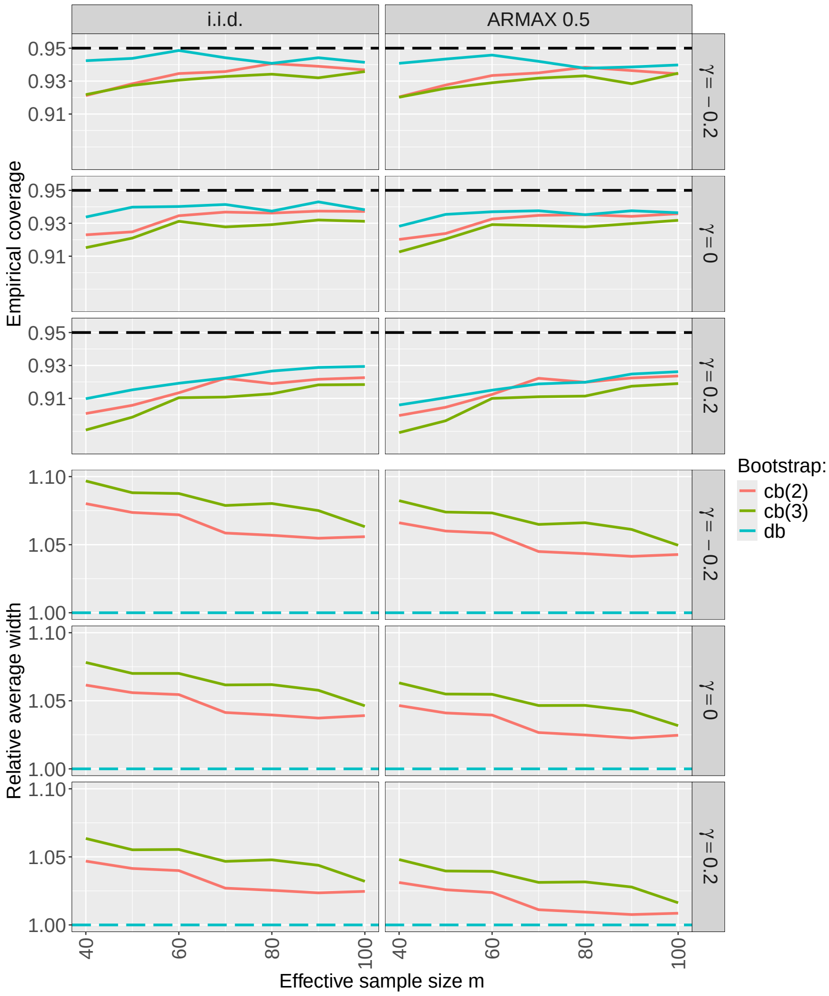

Figure 4.5: Basic bootstrap confidence intervals for μ_r with fixed block size $r = 90$. Top three rows: empirical coverage with intended coverage of 95% (dashed line). Bottom three rows: relative average width with respect to the disjoint method, i.e., width(CI(db))/width(CI(mb)).

because of its inconsistency. We find that in most scenarios the desired coverage is almost reached by any method, in particular for lighter tail behavior. The disjoint blocks approach has the best coverage overall, albeit with only minimal advantages and sometimes on a par with or even slightly worse than the cb(2)-method. However, the price for this is universally wider confidence intervals.

4.6.2 Fixed total sample size

In a second experiment, we consider the estimation of target parameters associated with the limiting law of M_r for $r \to \infty$, as well as bootstrap approximations for the respective estimation errors. For illustrative purposes, we restrict attention to the univariate, heavy tailed case and the estimation of the shape parameter α by pseudo-maximum

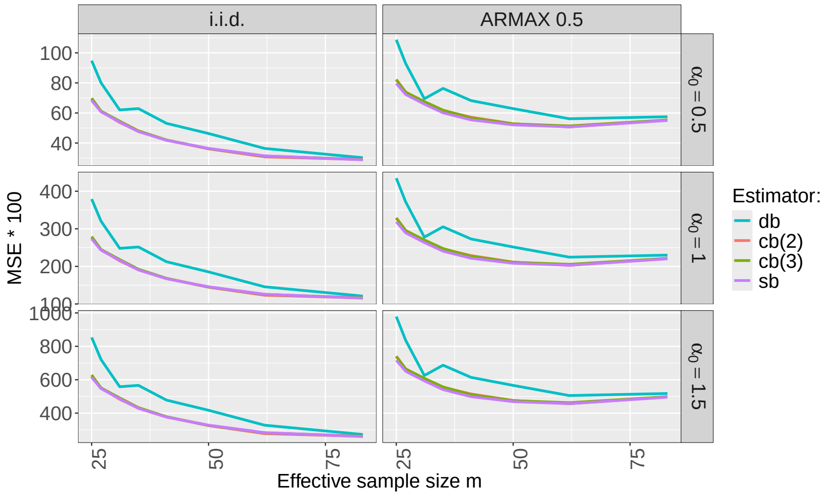

Figure 4.6: Mean squared error for the estimation of α with fixed sample size $n = 1,000$.

likelihood estimation, as extensively discussed in Section 4.5.

Model 4.6.2 (ARMAX-Pareto-Model)**.** $(X_t)_{t\in\mathbb{Z}}$ is a stationary, positive time series whose stationary distribution is the Pareto(α) distribution for some $\alpha > 0$, defined by its cdf $F_\alpha(x) = 1 - x^{-\alpha}$ for $x > 1$. After transformation to the Fréchet(1)-scale, the temporal dynamics correspond to the ARMAX(1)-model with parameter $\beta \in [0, 1)$; see Model 4.6.1 for details. Throughout, we consider the six models defined by the combinations of $\alpha \in \{0.5, 1, 1.5\}$ and $\beta \in \{0.0.5\}$.

For the simulation experiment, we fix the total sample size to $n = 1,000$, and treat the block size r as a variable tuning parameter akin to the choice of k in the peak-over-threshold method. More precisely, we consider choices of r ranging from 8 to 40, resulting in effective sample sizes $m = n/r$ ranging from 125 to 25.

Performance of the estimators. In Figure 4.6 we depict the MSE of $\hat\alpha_n^{(\mathrm{mb})}$ from Section 4.5 with mb $\in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\}$ as a function of the effective sample size $m = n/r$. For $\beta = 0.5$, we observe the typical bias-variance tradeoff resulting in a u-shaped MSE-curve: the larger the effective sample size, the smaller the variance and the larger the bias (for $\beta = 0$, the MSE is dominated by the variance even for block sizes as small as 8). Apart from that, the findings are similar to the previous results for the case where r was fixed: the sliding and circular block maxima estimator perform remarkably similar (in particular, the additional bias in the circular maxima is irrelevant), and they uniformly outperform the disjoint blocks estimator across all models under consideration.

Performance of the bootstrap. Next, we consider each of the bootstrap estimators $\hat\alpha_n^{(\mathrm{mb}),*}$ with mb $\in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\}$, with number of bootstrap replications set to $B = 1,000$. Similar as in Section 4.6.1, we start by evaluating the performance of the four bootstrap approaches in terms of their ability to provide accurate estimates of the estimation variance $\sigma^2_{\mathrm{mb}}(r) = \mathrm{Var}(\hat\alpha_n^{(\mathrm{mb})})$ with mb $\in \{\mathrm{db}, \mathrm{sb}\}$. In Figure 4.7, we depict the

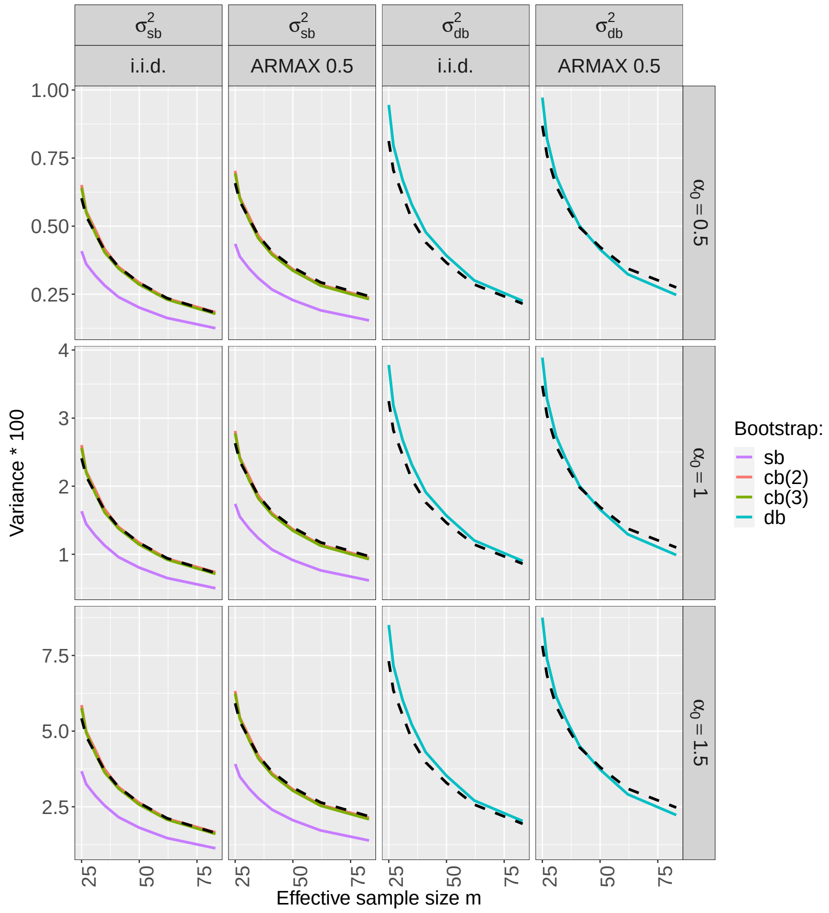

Figure 4.7: Bootstrap-based estimation of the estimation variance $\sigma^2_{\mathrm{mb}}(r) = \mathrm{Var}(\hat{\alpha}_n^{(\mathrm{mb})})$ with fixed sample size $n = 1,000$. Left two columns: mb = sb; the dashed line corresponds to $\sigma^2_{\mathrm{sb}}(r)$. Right two columns: mb = db; the dashed line corresponds to $\sigma^2_{\mathrm{db}}(r)$.

average over the $N = 5,000$ bootstrap estimates, along with the true parameters determined in a presimulation based on 10^6 repetitions. The findings are akin to those in Section 4.6.1, the only additional remarkable observation being that the disjoint blocks method tends to overestimate the variance parameter in the iid case, in particular for large block sizes.

Finally, we evaluate the performance of the bootstrap approaches in terms of their ability to provide accurate (basic bootstrap) confidence intervals of pre-specified coverage. The empirical coverage and the average widths of the respective intervals are depicted in Figure 4.8, where we omit the naive sliding method because of its inconsistency. We find that in most scenarios the desired coverage is almost reached by any method (observed coverage $\geq 92\%$), as long as the block size is not too small. Overall, the disjoint blocks method has clearly the best coverage, often reaching the intended level exactly (possibly because of the overestimation of the variance parameter ob-

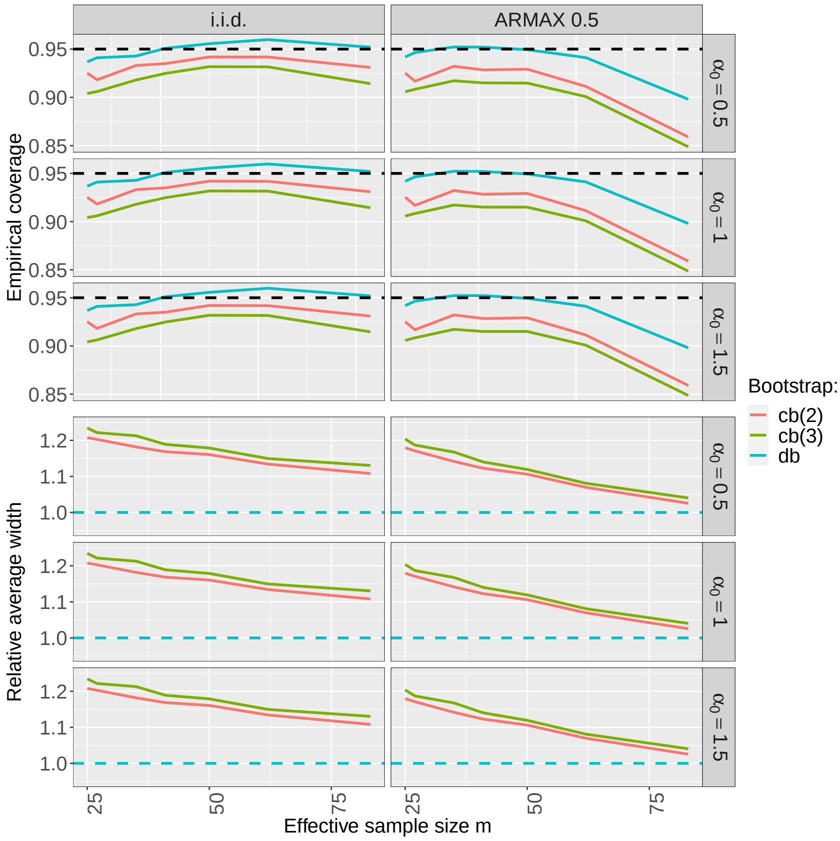

Figure 4.8: Basic bootstrap confidence intervals for α with fixed sample size $n = 1,000$. Top three rows: empirical coverage with intended coverage of 95% (dashed line). Bottom three rows: relative average width with respect to the disjoint method, i.e., width(CI(db))/width(CI(mb)).

served in Figure 4.6). The observed qualitative behavior may further be explained by the fact that the target parameter is an asymptotic parameter, such that both the sliding and the disjoint blocks method provide bias estimates (with the same asymptotic bias). Smaller confidence intervals for the circmax-methods hence necessarily imply a smaller coverage because they are concentrated around a biased estimate.

4.7 Case study

We consider daily accumulated precipitation amounts at a German weather station in Hohenpeißenberg, in the 145 year period from 1879 to 2023, resulting in $52,960$ daily observations in total.[1] As a target parameter, we consider the expected yearly maximum precipitation (Rx1day), which corresponds to a block size of $r = 365$. To account for possible non-stationarities in the target parameter over such a long observation period, we conduct the subsequent analyses on moving windows of 40 years. More precisely, for

[1]A maximum likelihood fit of the GEV-distribution to the entire sample of 145 annual maxima shows that the data are slightly heavy-tailed, with an estimated shape parameter of $\hat{\gamma} \approx 0.11$

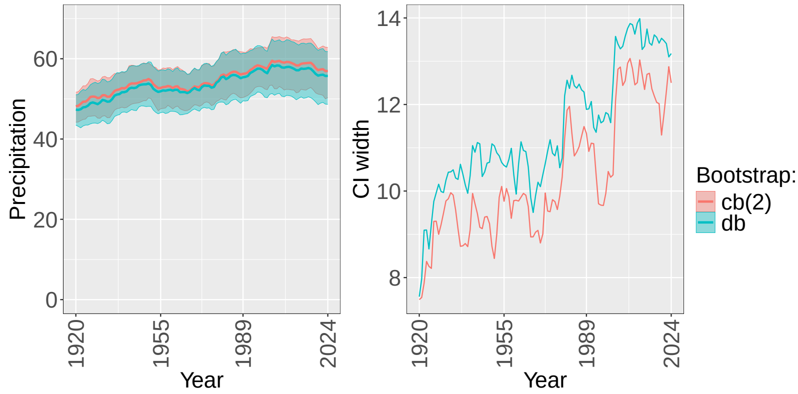

Figure 4.9: Left plot: estimated expected annual maximum precipitation based on disjoint and sliding blocks estimators $\mu_n^{(\mathrm{db})}, \mu_n^{(\mathrm{sb})}$ (solid lines), alongside with with 95% bootstrap confidence intervals based on the db- and cb(2)-bootstrap, respectively (for a moving window of 40 years). Right plot: width of the respective confidence interval. All curves have been smoothed by a moving average filter of size 2.

each fixed 40-year window w, the target parameter may be written as $\mu(w) = \mathrm{E}[M_{365}(w)]$, where $M_{365}(w)$ denotes a generic annual maximum variable corresponding to the climate over the 40 years under consideration.

Each moving window w contains approximately $n = 14,600$ daily observations, and the target parameter is estimated using the disjoint and sliding blocks estimator $\hat{\mu}_n^{(\mathrm{mb})} = \hat{\mu}_n^{(\mathrm{mb})}(w)$ with $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$. Basic bootstrap confidence intervals are constructed by either the classical approach based on resampling the disjoint block maxima (for estimator $\hat{\mu}_n^{(\mathrm{db})}$), or by the circular block bootstrap approach with $k = 2$ (for estimator $\hat{\mu}_n^{(\mathrm{sb})}$). The resulting estimates and confidence bands are presented in Figure 4.9. The years on the x-axis correspond to the endpoints of the 40-year window. As expected from the theory (see in particular Figure 4.10), we find that the confidence intervals for the sliding blocks estimator are universally smaller than those for the disjoint blocks counterpart, with the average relative width being approximately 1.07.

4.8 Conclusion

Both the block-maxima method and the bootstrap are time-honored statistical methods that have seen wide use in applied statistics for extremes. Surprisingly, bootstrap consistency has never been proven, not even for the classical block-maxima method based on disjoint blocks. In this paper, respective consistency statements were established under high-level conditions on the data-generating process. A new approach, called the circular block-maxima method, has been proposed to allow for valid and computationally efficient bootstrap inference regarding the sliding block maxima method. The approach may be of independent interest for extreme-value analysis of non-stationary extremes, for instance in the presence of a temporal trend. Indeed, a reasonable model assumption inspired by many applications of the disjoint block maxima method would consist of the assumption that the circular block-maxima within a fixed kr-block of observations, say the ith one, follow an extreme-value distribution whose distribution/parameters are depending on i. Respective methods could be studied mathematically under a suitable triangular array structure, and we conjecture that advantages over the disjoint block maxima method will eventually show up in respective estimation variances. Other possible extensions would be the inclusion of inner-seasonal non-stationarities in the data-generating process or the generalization to empirical processes indexed by non-finite function classes.

4.9 Proofs

Proof of Theorem 4.3.2. Throughout, we omit the upper index (cb). We start by showing that

$$\lim_{n\to\infty} \operatorname{Cov}(\bar{\mathbb{G}}_{n,r}\boldsymbol{h}) = \Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}. \tag{4.9.1}$$

By elementary arguments, it is sufficient to consider the case $q = 1$. For $i \in \{1, \dots, m(k)\}$ write

$$D_{n,i} = \frac{1}{kr} \sum_{s\in I_{kr,i}} \big\{ h(Z_{r,s}) - \mathrm{E}[h(Z_{r,s})] \big\}, \tag{4.9.2}$$

such that $\bar{\mathbb{G}}_{n,r}h_j = \sqrt{\frac{n}{r}} m(k)^{-1} \sum_{i=1}^{m(k)} D_{n,i}$. By stationarity and $n/r = km(k)$, we have

$$\operatorname{Var}\big(\bar{\mathbb{G}}_{n,r}h\big) = k \operatorname{Var}(D_{n,1}) + r_{n1} + r_{n2}, \tag{4.9.3}$$

where $r_{n1} = 2k(1 - \frac{1}{m(k)}) \operatorname{Cov}(D_{n,1}, D_{n,2})$ and $r_{n2} = 2k \sum_{d=2}^{m(k)-1} (1 - \frac{d}{m(k)}) \operatorname{Cov}(D_{n,1}, D_{n,1+d})$. Thus, the proof of (4.9.1) is finished once we show that

$$\lim_{n\to\infty} k \operatorname{Var}(D_{n,1}) = \Sigma_h^{(\mathrm{sb})}, \quad \lim_{n\to\infty} r_{n1} = 0, \quad \lim_{n\to\infty} r_{n2} = 0. \tag{4.9.4}$$

We start with the former, and for that purpose we define, for $\xi, \xi' \in [0, k)$,

$$\begin{aligned} f_r(\xi, \xi') &:= \operatorname{Cov}\big(h(Z_{r,1+\lfloor r\xi \rfloor}), h(Z_{r,1+\lfloor r\xi' \rfloor})\big), \\ g_{(k)}(\xi, \xi') &:= \operatorname{Cov}\big(h(Z^{(1)}_{|\xi-\xi'|,(k)}), h(Z^{(2)}_{|\xi-\xi'|,(k)})\big), \\ g(\xi, \xi') &:= \operatorname{Cov}\big(h(Z^{(1)}_{|\xi-\xi'|}), h(Z^{(2)}_{|\xi-\xi'|})\big), \end{aligned}$$

where $(Z^{(1)}_{|\xi-\xi'|,(k)}, Z^{(2)}_{|\xi-\xi'|,(k)})$ has cdf $G^{(k)}_{|\xi-\xi'|}$ from (4.11.1) and where $(Z^{(1)}_{|\xi-\xi'|}, Z^{(2)}_{|\xi-\xi'|})$ has cdf $G_{|\xi-\xi'|}$ from (4.2.3). Observe that, by Proposition 4.11.1, Condition 4.10.1 and Example 2.21 in van der Vaart (1998), $\lim_{n\to\infty} f_r(\xi, \xi') = g_{(k)}(\xi, \xi')$. Hence, by Condition 4.10.1 and Dominated Convergence

$$k \operatorname{Var}(D_{n,1}) = \frac{1}{kr^2} \sum_{s=1}^{kr} \sum_{t=1}^{kr} \operatorname{Cov}\big[h(Z_{r,s}), h(Z_{r,t})\big] = \frac{1}{k} \int_0^k \int_0^k f_r(\xi, \xi') \,\mathrm{d}\xi' \,\mathrm{d}\xi$$

$$= \frac{1}{k}\int_0^k\int_0^k g_{(k)}(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi + o(1). \qquad (4.9.5)$$

Now let $u(x) := g(x,0)$ and note that $g(\xi,\xi') = u(|\xi-\xi'|)$ and that $u(x) = 0$ for $x > 1$. By the definition of $G^{(k)}_{|\xi-\xi'|}$ in (4.11.1) we have $G^{(k)}_{|\xi-\xi'|} = G_{|\xi-\xi'|}$ if $0 \le |\xi-\xi'| \le k-1$, which also implies that $g_{(k)}(\xi,\xi') = g_{(k)}(|\xi-\xi'|,0) = g(|\xi-\xi'|,0) = u(|\xi-\xi'|)$ for $\xi \le k-1$. Therefore,

$$\begin{aligned}\int_0^k\int_\xi^k g_{(k)}(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi &= \int_0^k\int_\xi^{k\wedge(\xi+k-1)} u(\xi'-\xi)\,\mathrm{d}\xi'\,\mathrm{d}\xi + \int_0^1\int_{\xi+k-1}^k g_{(k)}(\xi'-\xi,0)\,\mathrm{d}\xi'\,\mathrm{d}\xi \\ &= \int_0^k\int_\xi^{k\wedge(\xi+1)} u(\xi'-\xi)\,\mathrm{d}\xi'\,\mathrm{d}\xi + \int_0^1\int_{k-1+\xi}^k g_{(k)}(\xi'-\xi,0)\,\mathrm{d}\xi'\,\mathrm{d}\xi, \quad (4.9.6)\end{aligned}$$

where we used that $u(x) = 0$ for $x > 1$ at the second equality. After substituting $x = \xi'-\xi$ in the inner integral and noting that $\int_0^1 u(x)\,\mathrm{d}x = \Sigma_h^{(\mathrm{sb})}/2$, the first integral on the right-hand side of (4.9.6) can be written as

$$\begin{aligned}\int_0^k\int_\xi^{k\wedge(\xi+1)} u(\xi'-\xi)\,\mathrm{d}\xi'\,\mathrm{d}\xi = \int_0^k\int_0^{(k-\xi)\wedge 1} u(x)\,\mathrm{d}x\,\mathrm{d}\xi &= \int_0^{k-1}\int_0^1 u(x)\,\mathrm{d}x\,\mathrm{d}\xi + \int_{k-1}^k\int_0^{k-\xi} u(x)\,\mathrm{d}x\,\mathrm{d}\xi \\ &= (k-1)\frac{\Sigma_h^{(\mathrm{sb})}}{2} + \int_0^1\int_0^y u(x)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

where we used the substitution $y = k-\xi$ in the last step. Moreover, substituting $x = k-\xi'+\xi$ in the inner integral, the second integral on the right-hand side of (4.9.6) can be written as

$$\int_0^1\int_{k-1+\xi}^k g_{(k)}(\xi'-\xi,0)\,\mathrm{d}\xi'\,\mathrm{d}\xi = \int_0^1\int_\xi^1 g_{(k)}(k-x,0)\,\mathrm{d}x\,\mathrm{d}\xi = \int_0^1\int_\xi^1 u(x)\,\mathrm{d}x\,\mathrm{d}\xi,$$

where we used that $g_{(k)}(k-x,0) = u(x)$ for $x \in [0,1]$ by the definition of $G^{(k)}_{k-x}$. Overall, again using that $\int_0^1 u(x)\,\mathrm{d}x = \Sigma_h^{(\mathrm{sb})}/2$, the previous three displays imply that

$$\int_0^k\int_\xi^k g_{(k)}(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi = k\frac{\Sigma_h^{(\mathrm{sb})}}{2}.$$

This implies (4.9.4) in view of the symmetry $\int_0^k\int_\xi^k g_{(k)}(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi = \int_0^k\int_0^\xi g_{(k)}(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi$ and (4.9.5).

Similar arguments as before, invoking asymptotic independence of $Z_{r,s}$ and $Z_{r,t}$ for $s \in I_{kr,1}$ and $t \in I_{kr,2}$ (see Proposition 4.11.1) and a similar identification of the covariances as integrals, imply that $r_{n1} = o(1)$. Finally, by Conditions 4.2.2(c) and 4.10.1 and Lemma 3.11 in Dehling and Philipp (2002) we obtain

$$\frac{|r_{n2}|}{k} \lesssim m(k)\alpha(kr)^{\nu/(2+\nu)} \lesssim \Big(\Big(\frac{n}{r}\Big)^{1+2/\nu}\alpha(kr)\Big)^{\nu/(2+\nu)} = o(1),$$

since $2/\omega < \nu$, where we have used that there is a lag of at least kr between the observations making up $D_{n,1}$ and $D_{n,1+d}$. Overall, we have shown all three assertions in (4.9.4), and hence the proof of (4.9.1) is finished.

To verify the asymptotic normality, by the Cramér-Wold Theorem, it is again sufficient consider the case $q = 1$. Recalling that $\bar{\mathbb{G}}_{n,r}h$ is centered, the case $\Sigma_h^{(\mathrm{sb})} = 0$ follows from (4.9.1); thus assume $\Sigma_h^{(\mathrm{sb})} > 0$. Following Bücher and Segers (2018a), we use a suitable blocking technique. Choose $m^* = m_n^* \in \mathbb{N}$ with $3 \le m^* \le m(k)$ such that $m^* \to \infty$ and $m^* = o((n/r)^{\nu/(2(1+\nu))})$ with ν from Condition 4.10.1. Next, define $p := p_n := m(k)/m^*$ and assume for simplicity that $p \in \mathbb{N}, n/r \in \mathbb{N}$. Furthermore, let $J_j^+ := \{(j-1)m^* + 1, \dots, jm^* - 1\}, J_j^- := \{jm^*\}$ and note that $|J_j^+| = (m^* - 1)$. Then,

$$\bar{\mathbb{G}}_{n,r}h = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} (S_{n,j}^+ + S_{n,j}^-), \quad \text{where} \quad S_{n,j}^{\pm} := \sqrt{\frac{np}{r}} \frac{1}{m(k)} \sum_{i \in J_j^{\pm}} D_{n,i} \tag{4.9.7}$$

with $D_{n,i}$ from (4.9.2). Noticing that the time series $(S_{n,j}^-)_j$ is stationary, we have

$$\operatorname{Var}\Big(\frac{1}{\sqrt{p}} \sum_{j=1}^{p} S_{n,j}^- \Big) \le 3 \operatorname{Var}(S_{n,1}^-) + \sum_{d=2}^{p-1} |\operatorname{Cov}(S_{n,1}^-, S_{n,1+d}^-)| =: r_{n3} + r_{n4}.$$

By Conditions 4.2.2(a) and 4.10.1 we have $r_{n3} = O(1/m^*) = o(1)$. Furthermore, by Condition 4.10.1 and Lemma 3.11 from Dehling and Philipp (2002) with $s = r = 1/(2+\nu)$ and $t = \nu/(2+\nu)$, we have $\sup_{d \ge 2} |\operatorname{Cov}(S_{n,1}^-, S_{n,1+d}^-)| \lesssim 10k(m^*)^{-1}\alpha^{\nu/(2+\nu)}(kr)$, whence $r_{n4} \lesssim r^{-1}n(m^*)^{-2}\alpha^{\nu/(2+\nu)}(kr) = o((m^*)^{-2}) = o(1)$ by Condition 4.2.2 and the fact that $2/\nu < \omega$. Put together, in view of $\mathrm{E}[S_{n,j}^-] = 0$, these convergences imply $p^{-1/2} \sum_{j=1}^{p} S_{n,j}^- = o_{\mathbb{P}}(1)$.

As there is a lag of at least kr between the observations making up S_{n,j_1}^+ and S_{n,j_2}^+ for $j_1 \neq j_2$ a standard argument involving characteristic functions, a complex-valued version of Lemma 3.9 in Dehling and Philipp (2002), Condition 4.2.2 (b) and Levy's Continuity Theorem allows for assuming that the $(S_{n,j}^+)_j$ are independent. Hence, we may subsequently apply Lyapunov's central limit theorem. To verify its conditions, note that, since $\operatorname{Var}(p^{-1/2} \sum_{j=1}^{p} S_{n,j}^-) = o(1)$ and by using (4.9.1) and (4.9.7), $\operatorname{Var}(p^{-1/2} \sum_{j=1}^{p} S_{n,j}^+) = \operatorname{Var}(\bar{\mathbb{G}}_{n,r}h) + o(1) = \Sigma_h^{(\mathrm{sb})} + o(1)$. Hence,

$$\frac{\sum_{j=1}^{p} \mathrm{E}[|S_{n,j}^+|^{2+\nu}]}{\{\operatorname{Var}(\sum_{j=1}^{p} S_{n,j}^+)\}^{1+\nu/2}} \lesssim \frac{p(np/r)^{1+\nu/2}(m^*/m(k))^{2+\nu}}{(p\sigma^2/2)^{1+\nu/2}} \lesssim \frac{(m^*)^{1+\nu}}{(n/r)^{\nu/2}} = o(1),$$

by Condition 4.10.1 and the choice of m^*. An application of Lyapunov's central limit theorem implies the assertion.

To prove the additional weak convergences note that

$$\tilde{\mathbb{G}}_{n,r}^{(\mathrm{cb})}h = \bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}h + \sqrt{\frac{n}{r}}(P_{n,r}^{(\mathrm{cb})} - P_r)h, \quad \mathbb{G}_{n,r}^{(\mathrm{cb})}h = \bar{\mathbb{G}}_{n,r}^{(\mathrm{cb})}h + \sqrt{\frac{n}{r}}(P_{n,r}^{(\mathrm{cb})} - P_r + P_r - P)h.$$

Denote by $(\tilde{X}_t)_t$ an independent copy of $(X_t)_t$. By stationarity we have

$$\begin{aligned}
(P_{n,r}^{(\mathrm{cb})} - P_r)h &= \frac{1}{m(k)} \sum_{i=1}^{m(k)} \frac{1}{kr} \sum_{s=1}^{kr} \mathrm{E}[h(Z_{r,(i-1)kr+s})] - P_r h \\
&= \frac{1}{kr} \sum_{s=(k-1)r+2}^{kr} \big(\mathrm{E}[h(Z_{r,s})] - P_r h \big)
\end{aligned}$$

$$= \frac{1}{kr}\sum_{t=1}^{r-1} \mathrm{E}\left[h\Big(\frac{\max(X_1,\dots,X_t,X_{(k-1)r+t+1},\dots,X_{kr})-b_r}{a_r}\Big) - h\Big(\frac{\boldsymbol{M}_{r,1}-b_r}{a_r}\Big)\right].$$

Condition 4.10.3 then gives $\sqrt{n/r}(P_{n,r}^{(\mathrm{cb})}-P_r)h \to k^{-1}(D_{h,k}+E_h)$. Furthermore, Condition 4.10.2 immediately implies $\sqrt{n/r}(P_r-P)h \to B_h$. Combined with the first assertion and the penultimate display, we obtain the additional two claimed weak convergences. □

Proof of Theorem 4.4.1. The second and third assertion follow immediately from the first one and the triangular inequality. For the proof of the first one, we omit the upper index cb. By Theorem 4.2.3 (for mb = sb) or Theorem 4.3.2 (for mb = cb) and Lemma 2.3 from Bücher and Kojadinovic (2019), it is sufficient to show that, unconditionally,

$$(\hat{\mathbb{G}}_{n,r}^{[1]}\boldsymbol{h}, \hat{\mathbb{G}}_{n,r}^{[2]}\boldsymbol{h}) \rightsquigarrow \mathcal{N}_q(\boldsymbol{0},\Sigma_h^{(\mathrm{sb})}) \otimes \mathcal{N}_q(\boldsymbol{0},\Sigma_h^{(\mathrm{sb})}),$$

where, for $b\in\{1,2\}$, $\hat{\mathbb{G}}_{n,r}^{[b]} = \sqrt{\frac{n}{r}}\frac{1}{n}\sum_{i=1}^{m(k)}(W_{m(k),i}^{[b]}-1)\sum_{s\in I_{kr,i}}\delta_{Z_{r,s}}$ with $\boldsymbol{W}_{m(k)}^{[b]} = (W_{m(k),1}^{[b]},\dots,W_{m(k),m(k)}^{[b]})$ two i.i.d. copies of $\boldsymbol{W}_{m(k)}$. By the Cramér-Wold device, it is sufficient to consider $q=1$, and for this in turn, yet again by the Cramér-Wold device, it is sufficient to show that

$$T_n := \sum_{b=1}^2 a_b\sqrt{\frac{n}{r}\frac{1}{n}}\sum_{i=1}^{m(k)}\big(W_{m(k),i}^{[b]}-1\big)\sum_{s\in I_{kr,i}} h(Z_{r,s}) \rightsquigarrow \mathcal{N}_2\big(0,(a_1^2+a_2^2)\Sigma_h^{(\mathrm{sb})}\big) \tag{4.9.8}$$

for all $\boldsymbol{a}=(a_1,a_2)^\top\in\mathbb{R}^2$. Note that $\sum_{i=1}^{m(k)}\big(W_{m(k),i}^{[b]}-1\big)=0$ and that $\sum_{s\in I_{kr,i}}\mathrm{E}[h(Z_{r,s})]$ does not depend on i, whence we may write

$$\begin{aligned} T_n &= \sum_{b=1}^2 a_b\sqrt{\frac{n}{r}\frac{1}{n}}\sum_{i=1}^{m(k)}\big(W_{m(k),i}^{[b]}-1\big)\Big\{\sum_{s\in I_{kr,i}} h(Z_{r,s})-\mathrm{E}[h(Z_{r,s})]\Big\} \\ &= \sum_{b=1}^2 a_b\sqrt{\frac{n}{r}}\frac{1}{m(k)}\sum_{i=1}^{m(k)}\big(W_{m(k),i}^{[b]}-1\big)D_{n,i} \end{aligned} \tag{4.9.9}$$

with $D_{n,i}$ from (4.9.2). The subsequent proof strategy, known as *Poissonization*, consists of removing the dependence of the multinomial multipliers by introducing row-wise i.i.d. multiplier sequences that approximate the multinomial multipliers. More precisely, we employ the construction from Lemma 4.11.3: for each fixed $n\in\mathbb{N}$, let $(\boldsymbol{U}_{j,m(k)}^{[b]})_{j\in\mathbb{N},b\in\{1,2\}}$ be i.i.d. multinomial vectors with 1 trial and $m(k)$ classes, with class probabilities $1/m(k)$ for each class, and independent of $(X_t)_{t\in\mathbb{Z}}$. We may then assume that

$$\boldsymbol{W}_{m(k)}^{[b]} = \sum_{j=1}^{m(k)}\boldsymbol{U}_{j,m(k)}^{[b]}.$$

Further, independent of $(\boldsymbol{U}_{j,m(k)}^{[b]})_{j\in\mathbb{N},b\in\{1,\dots,2\}}$ and of $(X_t)_{t\in\mathbb{Z}}$, let $(N_{m(k)}^{[b]})_{b=1,\dots,M}$ be i.i.d. Poisson($m(k)$) distributed random variables. Define

$$\tilde{\boldsymbol{W}}_{m(k)}^{[b]} = (\tilde W_{m(k),1}^{[b]},\dots,\tilde W_{m(k),m(k)}^{[b]}) = \sum_{j=1}^{N_{m(k)}^{[b]}}\boldsymbol{U}_{j,m(k)}^{[b]}.$$

By Lemma 4.11.3 the random vectors $\tilde{\boldsymbol{W}}^{[1]}_{m(k)}, \tilde{\boldsymbol{W}}^{[2]}_{m(k)}$ are i.i.d. Poisson$(1)^{\otimes m(k)}$ distributed. Then, in view of (4.9.9), we may write $T_n = \tilde{T}_n + a_1 R_{n1} + a_2 R_{n2}$, where

$$\tilde{T}_n = \sqrt{\frac{n}{r}} \frac{1}{m(k)} \sum_{i=1}^{m(k)} \Big\{ \sum_{b=1}^{2} a_b \big(\tilde{W}^{[b]}_{m(k),i} - 1\big) \Big\} D_{n,i},$$
$$R_{nb} = \sqrt{\frac{n}{r}} \frac{1}{m(k)} \sum_{i=1}^{m(k)} \big(W^{[b]}_{m(k),i} - \tilde{W}^{[b]}_{m(k),i}\big) D_{n,i}.$$

We may apply Lemma 4.11.2 to obtain that $\tilde{T}_n \rightsquigarrow \mathcal{N}(0, (a_1^2 + a_2^2)\Sigma_h^{(\mathrm{sb})})$. For the proof of (4.9.8), it hence remains to verify that $R_{nb} = o_{\mathbb{P}}(1)$ for $b \in \{1, 2\}$. We will suppress the upper index $[b]$ in the following. First write

$$R_{nb} = \sqrt{\frac{n}{r}} \frac{1}{m(k)} \sum_{i=1}^{m(k)} D_{n,i} \mathrm{sgn}(N_{m(k)} - m(k)) \sum_{j=1}^{\infty} \mathbf{1}\big\{ i \in I^j_{m(k)} \big\},$$

where $I^j_{m(k)} = \{i \in \{1, \dots, m(k)\} : |\tilde{W}_{m(k),i} - W_{m(k),i}| \geq j\}$.

We will start by showing that $\mathbb{P}(A_n) = o(1)$, where $A_n := \{|I^3_{m(k)}| > 0\} = \{\exists i \in \{1, \dots, m(k)\} : |\tilde{W}_{m(k),i} - W_{m(k),i}| \geq 3\}$. Fix $\delta > 0$. Invoking the central limit theorem, we may choose $C = C_\delta > 0$ sufficiently large such that $\mathbb{P}(|N_{m(k)} - m(k)| > C\sqrt{m(k)}) \leq \delta$ for all $n \in \mathbb{N}$. Next, note that, conditional on $N_{m(k)} = M$, the difference $|\tilde{W}_{m(k),i} - W_{m(k),i}|$ follows a $\mathrm{Bin}(|M - m(k)|, m(k)^{-1})$ distribution. Further, setting $c = \lceil C\sqrt{m(k)} \rceil$ and choosing M such that $|M - m(k)| \leq c$, we have that $\mathrm{Bin}(|M - m(k)|, m(k)^{-1})([3, |M - m(k)|]) \leq \mathrm{Bin}(c, m(k)^{-1})([3, c])$. As a consequence, conditioning on the event $N_{m(k)} = M$, we obtain

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{P}(|I^3_{m(k)}| > 0, |N_{m(k)} - m(k)| \leq c) + \delta \\
&\leq \sum_{i=1}^{m(k)} \mathbb{P}(|\tilde{W}_{m(k),i} - W_{m(k),i}| \geq 3, |N_{m(k)} - m(k)| \leq c) + \delta \\
&\leq m(k) \mathrm{Bin}(c, m(k)^{-1})([3, c]) + \delta \\
&\leq m(k)\{c m(k)^{-2} + \mathrm{Poisson}(c m(k)^{-1})([3, c])\} + \delta \\
&= \frac{c}{m(k)} + m(k) e^{-c/m(k)} \sum_{j=3}^{c} \frac{(c m(k)^{-1})^j}{j!} + \delta \\
&= O((m(k))^{-1/2}) + \delta,
\end{aligned}$$

where we used the approximation error of the Poisson Limit Theorem. Hence, since $\delta > 0$ was arbitrary, $\mathbb{P}(A_n) = o(1)$.

Next, note that, on A_n^c,

$$|I^1_{m(k)}| + |I^2_{m(k)}| = \sum_{j=1}^{\infty} |I^j_{m(k)}| = \sum_{i=1}^{m(k)} \sum_{j=1}^{\infty} \mathbf{1}\big\{ i \in I^j_{m(k)} \big\} = \sum_{i=1}^{m(k)} |\tilde{W}_{m(k),i} - W_{m(k),i}| = |N_{m(k)} - m(k)|,$$

using the fact that $\boldsymbol{U}_{j,m(k)}$ is multinomially distributed with 1 trial. As a consequence, letting $H_{n,i} = \mathbf{1}\big\{ N_{m(k)} \neq m(k) \big\} \big[\mathbf{1}\{i \in I^1_{m(k)}\} + \mathbf{1}\{i \in I^2_{m(k)}\}\big] / |N_{m(k)} - m(k)|$ and $\tilde{R}_{nb} = \big| \sum_{i=1}^{m(k)} H_{n,i} D_{n,i} \big|$, we obtain that

$$|R_{nb}| \le \mathbf{1}_{A_n^c}\Big|\sqrt{\frac{n}{r}}\frac{1}{m(k)}\sum_{i=1}^{m(k)} D_{n,i}\sum_{j=1}^{2}\mathbf{1}\{i\in I^j_{m(k)}\}\Big| + o_{\mathbb{P}}(1) = \mathbf{1}_{A_n^c}\sqrt{k}\times\frac{|N_{m(k)}-m(k)|}{\sqrt{m(k)}}\times\tilde R_{nb}+o_{\mathbb{P}}(1).$$

Therefore, since $|N_{m(k)}-m(k)|/\sqrt{m(k)} = O_{\mathbb{P}}(1)$ by the Central Limit Theorem, the proof of (4.9.8) is finished if we show that $\tilde R_{nb} = o_{\mathbb{P}}(1)$. For that purpose note that, for any fixed $\varepsilon>0$,

$$\mathbb{P}\Big(\frac{\mathbf{1}\{N_{m(k)}\ne m(k)\}}{|N_{m(k)}-m(k)|}\ge\varepsilon\Big)\le\mathbb{P}\Big(|N_{m(k)}-m(k)|\le\frac{1}{\varepsilon}\Big)=o(1)$$

by the Central Limit Theorem, which in turn yields

$$\max_{i=1,\dots,m(k)} H_{n,i}\le\frac{2}{|(N_{m(k)}-m(k))|}\mathbf{1}\{N_{m(k)}\ne m(k)\}=o_{\mathbb{P}}(1).$$

Now let $\sigma_r(i,j)=\mathrm{Cov}(D_{n,i},D_{n,j})$ and invoke the Conditional Tschebyscheff inequality to obtain that, for fixed $\varepsilon>0$,

$$\begin{aligned}
&\mathbb{P}\big(|\tilde R_{nb}|>\varepsilon \mid (H_{n,i})_{i=1,\dots,m(k)}\big)\\
&\lesssim \sum_{i,j=1}^{m(k)} H_{n,i}H_{n,j}|\sigma_r(i,j)|\\
&\le\Big\{\max_{j=1}^{m(k)}H_{n,j}\Big\}\times\Big\{\sum_{i=1}^{m(k)}H_{n,i}\sigma_r(1,1)+2\sum_{i=1}^{m(k)-1}H_{n,i}|\sigma_r(1,2)|+\sum_{|i-j|\ge2}H_{n,i}|\sigma_r(i,j)|\Big\}\\
&\le\Big\{\max_{j=1}^{m(k)}H_{n,j}\Big\}\times\Big\{\sigma_r(1,1)+2|\sigma_r(1,2)|+2m(k)\max_{d=2}^{m(k)-1}|\sigma_r(1,1+d)|\Big\},
\end{aligned}$$

where we used $\sum_{i=1}^{m(k)}H_{n,i}\le1$. Since $\max_{j=1}^{m(k)}H_{n,j}=o_{\mathbb{P}}(1)$ and by the (proof of the) three assertions in (4.9.4), we obtain that the expression on the right-hand side of the previous display is $o_{\mathbb{P}}(1)$. Hence, writing $\mathbb{P}(|\tilde R_{nb}|>\varepsilon)=\mathrm{E}\big[\mathbb{P}\big(|\tilde R_{nb}|>\varepsilon\mid(H_{n,i})_i\big)\big]$ and invoking the Dominated Convergence Theorem for convergence in probability, we obtain $\tilde R_{nb}=o_{\mathbb{P}}(1)$ and the proof is finished. □

Proof for Remark 4.4.3. We sketch the proof of the conditional weak convergence in (4.4.2) provided that Conditions 4.2.1, 4.2.2 and 4.10.1(a) are met. Repeating the arguments from the proof of Theorem 4.4.1, the assertion follows from unconditional weak convergence of

$$\tilde{\mathbb{G}}^{(\mathrm{sb}),*}_{n,r}h=\sqrt{\frac{n}{r}}\frac{1}{m(k)}\sum_{i=1}^{m(k)}Y_{n,i}D_{n,i},$$

where $(Y_{n,i})_i$ are iid with expectation 0 and variance 1, and where $D_{n,i}$ are defined as in (4.9.2), but with sb instead of cb. We only provide the proof for the convergence of the variance of $\tilde{\mathbb{G}}^{(\mathrm{sb}),*}_{n,r}h$ to $\Sigma_h^{(k)}$; the remaining arguments are then the same as in the proof of Lemma 4.11.2.

Since $(Y_{n,i}D_{n,i})_i$ in uncorrelated, we obtain

$$\operatorname{Var}(\tilde{\mathbb{G}}_{n,r}^{(\mathrm{sb}),*}h) = k\operatorname{Var}(D_{n,1}) = \frac{1}{kr^2}\sum_{s=1}^{kr}\sum_{t=1}^{kr}\operatorname{Cov}(h(Z_{r,s}^{(\mathrm{sb})}), h(Z_{r,t}^{(\mathrm{sb})}))$$

Using the notation after (4.9.4), we obtain

$$\lim_{n\to\infty}\operatorname{Var}(\tilde{\mathbb{G}}_{n,r}^{(\mathrm{sb}),*}h) = \frac{1}{k}\int_0^k\int_0^k g(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi.$$

We start by considering the case $k \geq 2$. Then, by the calculation in (4.9.6) but with $g_{(k)}$ replaced by g,

$$\int_0^k\int_\xi^k g(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi = I_1 + I_2$$

where

$$I_1 = (k-1)\Sigma_h^{(\mathrm{sb})}/2 + \int_0^1\int_0^y u(x)\,\mathrm{d}x\,\mathrm{d}y, \qquad I_2 = \int_0^1\int_{k-1+\xi}^k u(\xi'-\xi)\,\mathrm{d}\xi'\,\mathrm{d}\xi = 0.$$

The first integral over u can be rewritten as

$$\int_0^1\int_0^y u(x)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1\int_x^1 u(x)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1(1-x)u(x)\,\mathrm{d}x = \Sigma_h^{(\mathrm{sb})}/2 - \int_0^1 xu(x)\,\mathrm{d}x, \quad (4.9.10)$$

so that

$$I_1 = k\Sigma_h^{(\mathrm{sb})}/2 - \int_0^1 xu(x)\,\mathrm{d}x.$$

For symmetry reasons, we also have $\int_0^k\int_0^\xi g(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi = \int_0^k\int_\xi^k g(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi$ so that

$$\operatorname{Var}(\tilde{\mathbb{G}}_{n,r}^{(\mathrm{sb}),*}h) = k\operatorname{Var}(D_{n,1}) = \Sigma_h^{(\mathrm{sb})} - \frac{2}{k}\int_0^1 xu(x)\,\mathrm{d}x + o(1)$$

as asserted.

It remains to consider the case $k = 1$. A straightforward calculation then shows that

$$\int_0^1\int_\xi^1 g(\xi,\xi')\,\mathrm{d}\xi'\,\mathrm{d}\xi = \int_0^1\int_\xi^1 u(\xi'-\xi)\,\mathrm{d}\xi'\,\mathrm{d}\xi = \int_0^1\int_0^{1-\xi} u(y)\,\mathrm{d}y\,\mathrm{d}\xi = \int_0^1\int_0^x u(y)\,\mathrm{d}y\,\mathrm{d}x,$$

which implies the assertion by (4.9.10). □

Proof of Theorem 4.5.5. The results regarding $(\hat{\alpha}_n^{(\mathrm{cb})}, \hat{\sigma}_n^{(\mathrm{cb})})^\top$ and $(\hat{\alpha}_n^{(\mathrm{cb}),*}, \hat{\sigma}_n^{(\mathrm{cb}),*})^\top$ follow from an application of Theorem 4.11.6 and Theorem 4.11.7, respectively.

For the former we need to show that, with $v_n = \sqrt{n/r}, \omega_n = n$ and $X_{n,s} = M_{r,s}^{(\mathrm{cb})} \vee c$, the "no-ties-condition" in Equation (4.11.4) is met, and that the three convergences in Condition 4.11.5 are satisfied. Equation (4.11.4) follows from $\mathbb{P}(X_{n,1} = \cdots = X_{n,n}) \leq \mathbb{P}(X_{n,1} = X_{n,r+1}) = \mathbb{P}((M_{r,1} \vee c)/\sigma_r = (M_{r,r+1} \vee c)/\sigma_r)$, which converges to zero by the Portmanteau theorem, observing that $(M_{r,1} \vee c)/\sigma_r, (M_{r,r+1} \vee c)/\sigma_r)$ weakly converges to the product of two independent Fréchet$(\alpha_0, 1)$ random variables by Lemma 5.1 in Bücher and Segers (2018a). Condition 4.11.5(iii) is a consequence of Condition 4.5.4. Condition 4.11.5(ii) follows from a straightforward modification of Theorem 4.3.2 to the Fréchet domain of attraction. Finally, for α_+ chosen sufficiently close to α_0, Condition 4.11.5(ii) follows from such a modification as well, following the argumentation on page 1116 in Bücher and Segers (2018b).

Next, consider the assertions regarding $(\hat\alpha_n^{(\mathrm{cb}),*}, \hat\sigma_n^{(\mathrm{cb}),*})^\top$, which follow from Theorem 4.11.7 if we additionally show that the "no-ties-condition" in Equation (4.11.5) is met, that $\hat{\mathbb{P}}_{n,r}^{(\mathrm{cb}),*} f = P_{\alpha_0,1} f + o_{\mathbb{P}}(1)$ for all $f \in \mathcal{F}(\alpha_-, \alpha_+)$ and that (4.11.7) is met. The latter two assertions follow from similar arguments as given in the proof of Theorem 4.4.1 (adaptations to the Fréchet domain of attraction are needed), and it remains to show (4.11.5). For that purpose, let $I_1, \dots, I_{m(k)}$ be iid uniformly distributed on $\{1, \dots, m(k)\}$ independent of $(X_{n,1}, \dots, X_{n,n})$, such that

$$(X_{n,1}^*, \dots, X_{n,n}^*) \stackrel{d}{=} (X_{n,kr(I_1-1)+1}, \dots, X_{n,kr(I_1-1)+kr}, \dots, X_{n,kr(I_{m(k)}-1)+1}, \dots, X_{n,kr(I_{m(k)}-1)+kr}).$$

Then

$$\begin{aligned}
\mathbb{P}(X_{n,1}^* = \dots = X_{n,n}^*) &= \sum_{i=1}^{m(k)} \mathbb{P}(I_1 = i, X_{n,1}^* = \dots = X_{n,n}^*) \\
&\le \sum_{i=1}^{m(k)} \mathbb{P}(I_1 = i, X_{n,kr(I_1-1)+1} = \dots = X_{n,kr(I_1-1)+kr}) \\
&= \frac{1}{m(k)} \sum_{i=1}^{m(k)} \mathbb{P}(X_{n,kr(i-1)+1} = \dots = X_{n,kr(I_1-1)+kr}) = \mathbb{P}(X_{n,1} = \dots = X_{n,kr}).
\end{aligned}$$

This probability is bounded by $\mathbb{P}(X_{n,1} = X_{n,r+1})$, which converges to zero as shown above when proving Equation (4.11.4). □

Proof of Corollary 4.5.6. Throughout, we omit the upper indexes (sb) and (cb) at $\hat\sigma_n$ and $\hat\sigma_n^*$, respectively. Define $S_n = \sqrt{\frac{n}{r}}(\hat\sigma_n/\sigma_{r_n} - 1)$ and $S_n^* = \sqrt{\frac{n}{r}}(\hat\sigma_n^* - \hat\sigma_n)/\sigma_{r_n}$, and note that, with $M_2(\alpha_0)$ denoting the second row of $M(\alpha_0)$,

$$S_n = M_2(\alpha_0)\mathbb{G}_{n,r}(f_1, f_2, f_3) + o_{\mathbb{P}}(1), \qquad S_n^* = \frac{\hat\sigma_n}{\sigma_{r_n}} \sqrt{\frac{n}{r}}(\hat\sigma_n^*/\hat\sigma_n - 1) = M_2(\alpha_0)\hat{\mathbb{G}}_{n,r}^*(f_1, f_2, f_3) + o_{\mathbb{P}}(1)$$

by Theorem 4.5.5. Moreover, arguing as in the proof of that theorem, the weak limit of S_n coincides with the conditional weak limit of $\hat S_n^*$ given the observations $\mathcal{X}_n$. Therefore, observing that $F_{S_n^*}(x) = \mathbb{P}(S_n^* \le x \mid \mathcal{X}_n)$ satisfies $F_{S_n^*}(x) = F_{\hat\sigma_n^*}(\hat\sigma_n + \sqrt{\frac{r}{n}} \cdot \sigma_{r_n} x)$ and hence

$$\begin{aligned}
\mathbb{P}(\sigma_{r_n} \in I_{n,\beta}) &= \mathbb{P}\Big[\sqrt{\frac{n}{r}} \cdot \sigma_{r_n}^{-1}\{(F_{\hat\sigma_n^*})^{-1}(\tfrac{\beta}{2}) - \hat\sigma_n\} \le \sqrt{\frac{n}{r}}\Big(\frac{\hat\sigma_n}{\sigma_{r_n}} - 1\Big) \le \sqrt{\frac{n}{r}} \cdot \sigma_{r_n}^{-1}\{(F_{\hat\sigma_n^*})^{-1}(1 - \tfrac{\beta}{2}) - \hat\sigma_n\}\Big] \\
&= \mathbb{P}\Big[(F_{S_n^*})^{-1}(\tfrac{\beta}{2}) \le S_n \le (F_{S_n^*})^{-1}(1 - \tfrac{\beta}{2})\Big],
\end{aligned}$$

the assertion follows from Lemma 4.2 in Bücher and Kojadinovic (2019). □

4.10 Additional conditions

In order to integrate $\boldsymbol{h}(Z_{r,i})$ to the limit, we need mild asymptotic integrability conditions. For $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$ the condition is standard. As the circmax-sample permutates the underlying time series on a non-negligible part of the sample, we need a more involved assumption on asymptotic integrability. In most cases both conditions can be verified by similar arguments. Note that (b) implies (a) by letting $t = r$ in the supremum. Moreover, for $k = 1$ (disjoint blocks case), (a) and (b) are equivalent.

Condition 4.10.1 (Asymptotic Integrability)**.** Fix $k \in \mathbb{N}$. Let $\boldsymbol{h} = (h_1, \dots, h_q)^\top : \mathbb{R}^d \to \mathbb{R}^q$ be an a.e. continuous function with respect to the Lebesgue-measure on $\mathbb{R}^d$. There exists a $\nu > 0$ such that:

(a) $\limsup_{r\to\infty} \mathrm{E}[\|\boldsymbol{h}(Z_{r,1})\|^{2+\nu}] < \infty$.

(b) $\limsup_{r\to\infty} \sup_{t=1,\dots,r} \mathrm{E}[\|\boldsymbol{h}\big((\max\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}\} - \boldsymbol{b}_r)/\boldsymbol{a}_r\big)\|^{2+\nu}] < \infty$.

Often we additionally impose that $\nu > 2/\omega$ with ω from Condition 4.2.2.

The next two conditions specify the asymptotic bias resulting from the approximation of the distribution of the various block maxima of size r_n by the extreme value distribution G.

Condition 4.10.2 (Disjoint and sliding blocks bias)**.** Let $r = r_n \to \infty$ with $r_n = o(n)$ and let $\boldsymbol{h} = (h_1, \dots, h_q)^\top : \mathbb{R}^d \to \mathbb{R}^q$ be measurable such that $h(Z_{r,1})$ and $h(Z)$ are integrable. The following limit exists:

$$\boldsymbol{B_h} = \lim_{n\to\infty} \sqrt{\frac{n}{r}} \Big\{ \mathrm{E}\left[\boldsymbol{h}(Z_{r,1})\right] - \mathrm{E}\left[\boldsymbol{h}(Z)\right] \Big\}. \tag{4.10.1}$$

Circularization as applied within the cirmax-sample changes the order of observations used to calculate some of the cirular block maxima. The next conditions captures the resulting bias by decomposing it into two pieces, the first of which is due to a coupling with an independent copy $(\tilde{\boldsymbol{X}}_t)_t$ of $(\boldsymbol{X}_t)_t$.

Condition 4.10.3. Let $r = r_n \to \infty$ with $r_n = o(n)$ and let $\boldsymbol{h} = (h_1, \dots, h_q)^\top : \mathbb{R}^d \to \mathbb{R}^q$ be measurable and $k \in \mathbb{N}_{\geq 2}$ be fixed. The following expectations and limits exist:

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{h},k} = \lim_{n\to\infty} \sqrt{\frac{n}{r}} \cdot \frac{1}{r} \sum_{t=1}^{r-1} \mathrm{E}\Bigg[\boldsymbol{h}\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big) \\ - \boldsymbol{h}\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big)\Bigg], \end{aligned} \tag{4.10.2}$$

$$\boldsymbol{E_h} = \lim_{n\to\infty} \sqrt{\frac{n}{r}} \cdot \frac{1}{r} \sum_{t=1}^{r-1} \mathrm{E}\Bigg[\boldsymbol{h}\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big) - \boldsymbol{h}\big(Z_{r,1}\big)\Bigg].$$

Note in passing that the asymptotic bias $\boldsymbol{E_h}$ already appeared in Bücher and Zanger (2023) in a slightly modified form in the setting of observing piecewise-stationary time series.

The following result implies that in many situations the asymptotic bias terms $D_{\boldsymbol{h},k}$ and $E_{\boldsymbol{h}}$ are negligible. The result should be understood as a proof of concept in the sense that one might formulate sharper results involving stronger integrability assumptions (e.g., boundedness of h_j) or a faster decay of mixing coefficients. Recall the beta-mixing coefficient for stationary time series, $\beta(\ell) := \beta(\sigma(\dots, \boldsymbol{X}_1, \boldsymbol{X}_0), \sigma(\boldsymbol{X}_{1+\ell}, \boldsymbol{X}_{2+\ell}, \dots))$ with $\ell \in \mathbb{N}$, where $\beta(\mathcal{B}, \mathcal{F})$ denotes the beta dependence coefficient between two sigma-fields $\mathcal{B}$ and $\mathcal{F}$ (Bradley, 2005).

Lemma 4.10.4. *Fix $k \in \mathbb{N}_{\geq 2}$ and suppose that Condition 4.2.1 and 4.2.2 are met. Then $\boldsymbol{D}_{\boldsymbol{h},k} = \boldsymbol{E}_{\boldsymbol{h}} = 0$, provided the time series $(\boldsymbol{X}_t)_t$ is exponentially β-mixing (i.e., there exists $c > 0, \lambda \in (0,1)$ such that $\beta(\ell) \leq c\lambda^\ell$ for all $\ell \in \mathbb{N}$), that $\log n = o(r^{1/2})$ and that Condition 4.10.1(b) and*

$$\limsup_{n\to\infty} \sup_{t=1,\dots,r} \mathrm{E}\left[\left\| h\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big)\right\|^{2+\nu}\right] < \infty \tag{4.10.3}$$

$$\limsup_{n\to\infty} \sup_{t=\lfloor r^{1/2}\rfloor,\dots,r} \mathrm{E}\left[\left\| h\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{t-\lfloor r^{1/2}\rfloor}, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big)\right\|^{2+\nu}\right] < \infty \tag{4.10.4}$$

$$\limsup_{n\to\infty} \sup_{t=\lfloor r^{1/2}\rfloor,\dots,r} \mathrm{E}\left[\left\| h\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{t-\lfloor r^{1/2}\rfloor}, \boldsymbol{X}_{t+1}, \dots, \boldsymbol{X}_r) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big)\right\|^{2+\nu}\right] < \infty \tag{4.10.5}$$

hold for some $\nu > 2/\omega$ with ω from Condition 4.2.2, where $(\tilde{\boldsymbol{X}}_t)_t$ is an independent copy of $(\boldsymbol{X}_t)_t$.

Proof. Without loss of generality we may assume that $q = 1$. First consider $D_{h,k}$ and define, for $t \in \{1, \dots, r-1\}$,

$$D_{n,h,k,t} = h\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big) - h\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big).$$

Note that showing $\sup_{t=1,\dots,r-1}(n/r)^{1/2}|\mathrm{E}[D_{n,h,k,t}]| = o(1)$ implies $D_{h,k} = 0$. By applying Berbee's coupling Lemma (Berbee, 1979) for each fixed $t \in \{1, \dots, r-1\}$ to the vectors $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t)$ and $(\boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr})$ we may assume that the random vector $(\tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t})$ satisfies $\mathbb{P}\big((\tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) \neq (\boldsymbol{X}_{(k-1)r+t+1}, \dots, \boldsymbol{X}_{kr})\big) \leq \beta(r) \leq c\lambda^r$; where the last inequality follows by assumption. Thus, by Hölder's inequality and (4.10.3),

$$\sup_{t=1,\dots,r-1} \sqrt{\frac{n}{r}}\,\mathrm{E}[|D_{n,h,k,t}|] \lesssim \sqrt{\frac{n}{r}}\lambda^{r(1+\nu)/(2+\nu)} = \exp\left[r\Big(\log\lambda\frac{1+\nu}{2+\nu} + \frac{\log(n/r)}{2r}\Big)\right].$$

The expression on the right converges to zero since $\log n = o(r)$. As a consequence, $D_{h,k} = 0$.

It remains to show that $E_h = 0$. Writing

$$E_{n,h,t} = h\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \tilde{\boldsymbol{X}}_1, \dots, \tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r}\Big) - h(Z_{r,1}).$$

it is sufficient to show that $\sup_{t=1,\dots,r-1}(n/r)^{-1/2}|\mathrm{E}[E_{n,h,t}]| = o(1)$. For that purpose, we split the supremum according to $t > r/2$ or $t \leq r/2$; both cases can then be treated similarly, and we only provide details on the former one. For simplicity, we assume that $r/2 \in \mathbb{N}$.

Write $\ell = \lfloor r^{1/2} \rfloor$. The proof is finished once we show that

$$\sup_{t=r/2,\dots,r} \left| \sqrt{\frac{n}{r}} \operatorname{E}\left[h\Big(\frac{\max(\boldsymbol{X}_1,\dots,\boldsymbol{X}_t,\tilde{\boldsymbol{X}}_1,\dots,\tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \Big) - h\Big(\frac{\max(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{t-\ell},\tilde{\boldsymbol{X}}_1,\dots,\tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \Big) \right] \right| = o(1) \quad (4.10.6)$$

$$\sup_{t=r/2,\dots,r} \left| \sqrt{\frac{n}{r}} \operatorname{E}\left[h\Big(\frac{\max(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{t-\ell},\tilde{\boldsymbol{X}}_1,\dots,\tilde{\boldsymbol{X}}_{r-t}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \Big) - h\Big(\frac{\max(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{t-\ell},\boldsymbol{X}_{t+1},\dots,\boldsymbol{X}_r) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \Big) \right] \right| = o(1) \quad (4.10.7)$$

$$\sup_{t=r/2,\dots,r} \left| \sqrt{\frac{n}{r}} \operatorname{E}\left[h\Big(\frac{\max(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{t-\ell},\boldsymbol{X}_{t+1},\dots,\boldsymbol{X}_r) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \Big) - h(\boldsymbol{Z}_{r,1}) \right] \right| = o(1). \quad (4.10.8)$$

The proof of (4.10.7) is similar to the proof of $D_{h,k} = 0$ given above (invoking (4.10.4) instead of (4.10.3)), and is therefore omitted (the final argument requires $\log n = o(r^{1/2})$, which is then exactly met by assumption). The proof of (4.10.7) is a simplified version of the proof of (4.10.8), so we only prove the latter for the sake of brevity. In view of Condition 4.10.1(b) and (4.10.5), an application of Hölder's inequality implies

$$\sup_{t=r/2,\dots,r} \left| \sqrt{\frac{n}{r}} \operatorname{E}\left[h\Big(\frac{\max(\boldsymbol{X}_1,\dots,\boldsymbol{X}_{t-\ell},\boldsymbol{X}_{t+1},\dots,\boldsymbol{X}_r) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \Big) - h(\boldsymbol{Z}_{r,1}) \right] \right|$$
$$\lesssim \sup_{t=r/2,\dots,r} \sqrt{\frac{n}{r}} \Big\{ \mathbb{P}\big(\exists j \in \{1,\dots,d\} : \max(X_{t-\ell+1,j},\dots,X_{t,j}) > \max(X_{1,j},\dots,X_{t-\ell,j},X_{t+1,j},\dots,X_{r,j}) \big) \Big\}^{\frac{1+\nu}{2+\nu}}.$$

By the union-bound, it is sufficient to show that

$$R_n \equiv \sup_{t=r/2,\dots,r} \sqrt{\frac{n}{r}} \max_{j=1}^{d} \Big\{ \mathbb{P}\big(\max(X_{t-\ell+1,j},\dots,X_{t,j}) > \max(X_{1,j},\dots,X_{t-\ell,j},X_{t+1,j},\dots,X_{r,j}) \big) \Big\}^{\frac{1+\nu}{2+\nu}}$$
$$= o(1). \quad (4.10.9)$$

Subsequently, we write X_t instead of X_{t-j}, as all subsequent bounds are uniform in j. We then have

$$\mathbb{P}\big(\max(X_{t-\ell+1},\dots,X_t) > \max(X_1,\dots,X_{t-\ell},X_{t+1},\dots,X_r) \big)$$
$$\le \sum_{i=t-\ell+1}^{t} \mathbb{P}\big(X_i > \max(X_1,\dots,X_{t-\ell},X_{t+1},\dots,X_r) \big).$$

For fixed i in the previous sum, let $J \subset \{1,\dots,t-\ell\}$ denote the maximal set of indexes such that $|j_1 - i| \ge \ell$ and $|j_1 - j_2| \ge \ell$ for all distinct $j_1, j_2 \in J$; note that $|J| = O(r/\ell)$ since $t > r/2$. We then have

$$\mathbb{P}\big(X_i > \max(X_1,\dots,X_{t-\ell},X_{t+1},\dots,X_r) \big) \le \mathbb{P}\big(X_i > \max(X_j : j \in J) \big).$$

We may now successively apply Berbee's coupling lemma (Berbee, 1979) to construct a vector $(\check{X}_j)_{j\in J}$ with iid coordinates that is independent of X_i and satisfies $\check{X}_j =_d X_j$ for all $j \in J$ such that

$$\mathbb{P}\big(X_i > \max(X_s : s \in J) \big) \le \mathbb{P}\big(X_i > \max(\check{X}_s : s \in J) \big) + |J|\beta(\ell).$$

More precisely, writing $J = \{j_1, \dots, j_{|J|}\}$, the first application is to X_{j_1} and $(X_{j_2}, \dots, X_{j_{|J|}}, X_i)$. The second application is to $(\check{X}_{j_1}, X_{j_2})$ and $(X_{j_3}, \dots, X_{j_{|J|}}, X_i)$, where $\check{X}_{j_1}$ is the random variable constructed with the first application, and so on.

Since $\mathbb{P}\big(X_i > \max(\check{X}_s : s \in J)\big) = 2^{-|J|}$ by Fubini's theorem, we obtain from the last three displays that

$$\sup_{t=r/2,\dots,r} \mathbb{P}\big(\max(X_{t-\ell+1}, \dots, X_t) > \max(X_1, \dots, X_{t-\ell}, X_{t+1}, \dots, X_r)\big) \le \ell 2^{-|J|} + \ell |J| \beta(r) \lesssim \ell 2^{-r/\ell} + r\lambda^\ell.$$

In view of our choice of $\ell = \lfloor r^{1/2} \rfloor$, and letting $\zeta = \min(2, 1/\lambda) > 1$, the right hand side is bounded by $2r\zeta^{-\sqrt{r}}$, whence R_n from (4.10.9) can be bounded by

$$R_n \lesssim \sqrt{\frac{n}{r}} \big(r\zeta^{-\sqrt{r}}\big)^{(1+\nu)/(2+\nu)} = \exp\Big\{ -\sqrt{r}\Big(\log(\zeta)\frac{1+\nu}{2+\nu} - \frac{\log r}{\sqrt{r}}\frac{1+\nu}{2+\nu} - \frac{\log(n/r)}{\sqrt{r}} \Big) \Big\} = o(1)$$

by assumption on r. This proves (4.10.9), and the proof is finished. □

4.11 Auxiliary results

Proposition 4.11.1 (Joint weak convergence of circular block maxima). *Fix $k \in \mathbb{N}_{\ge 2}$ and suppose Conditions 4.2.1 and 4.2.2(a), (b) are met. Then, for any $\xi, \xi' \in [0, k)$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we have*

$$\lim_{n\to\infty} \mathbb{P}\Big(\boldsymbol{Z}^{(\mathrm{cb})}_{r,1+\lfloor \xi r \rfloor} \le \boldsymbol{x}, \boldsymbol{Z}^{(\mathrm{cb})}_{r,1+\lfloor \xi' r \rfloor} \le \boldsymbol{y} \Big) = G^{(k)}_{|\xi-\xi'|}(\boldsymbol{x}, \boldsymbol{y}) := \begin{cases} G_{k-|\xi-\xi'|}(\boldsymbol{x}, \boldsymbol{y}), & |\xi-\xi'| > k-1, \\ G_{|\xi-\xi'|}(\boldsymbol{x}, \boldsymbol{y}), & |\xi-\xi'| \le k-1, \end{cases} \tag{4.11.1}$$

with G_ξ from (4.2.3). Moreover, any circular block maxima taken from distinct kr-blocks are asymptotically independent, that is, for any $1 \le i \ne i' \le m(k), \xi, \xi' \in [0, k)$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\boldsymbol{Z}^{(\mathrm{cb})}_{r,(i-1)kr+1+\lfloor \xi r \rfloor} \le \boldsymbol{x}, \boldsymbol{Z}^{(\mathrm{cb})}_{r,(i'-1)kr+1+\lfloor \xi' r \rfloor} \le \boldsymbol{y} \Big) = G_1(\boldsymbol{x}, \boldsymbol{y}) = G(\boldsymbol{x})G(\boldsymbol{y}). \tag{4.11.2}$$

Proof. We start by showing marginal convergence (i.e., $\boldsymbol{y} = \infty$); the result then corresponds to Proposition 4.3.1. The case $\xi \le k-1$ is trivial. For $\xi > k-1$ and fixed $\boldsymbol{x} \in \mathbb{R}^d$, we may proceed as in the proof of Lemma 2.4 in Bücher and Zanger (2023) to obtain

$$\begin{aligned} \mathbb{P}\Big(\boldsymbol{Z}^{(\mathrm{cb})}_{r,1+\lfloor \xi r \rfloor} \le \boldsymbol{x} \Big) &= \mathbb{P}\Big(\frac{\max(\boldsymbol{X}_{1+\lfloor \xi r \rfloor}, \dots, \boldsymbol{X}_{kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{x}, \frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{r+\lfloor \xi r \rfloor - kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{x} \Big) \\ &= \mathbb{P}\Big(\frac{\max(\boldsymbol{X}_{1+\lfloor \xi r \rfloor}, \dots, \boldsymbol{X}_{kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{x} \Big) \mathbb{P}\Big(\frac{\max(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{r+\lfloor \xi r \rfloor - kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{x} \Big) + R_n, \end{aligned}$$

where $R_n = O(\alpha(r_n)) = o(1)$. Next, by stationarity, the product on the right-hand side of the previous display can be written as

$$\mathbb{P}\Big(\boldsymbol{Z}_{kr-\lfloor \xi r \rfloor} \le \frac{\boldsymbol{a}_r \boldsymbol{x} + \boldsymbol{b}_r - \boldsymbol{b}_{kr-\lfloor \xi r \rfloor}}{\boldsymbol{a}_{kr-\lfloor \xi r \rfloor}} \Big) \mathbb{P}\Big(\boldsymbol{Z}_{r+\lfloor \xi r \rfloor - kr} \le \frac{\boldsymbol{a}_r \boldsymbol{x} + \boldsymbol{b}_r - \boldsymbol{b}_{r+\lfloor \xi k \rfloor - kr}}{\boldsymbol{a}_{r+\lfloor \xi r \rfloor - kr}} \Big)$$

The convergence in Equation (4.2.1) being locally uniform we obtain, for $j \in \{1, \dots, d\}$,

$$\lim_{n\to\infty} \frac{a_r^{(j)} x^{(j)} + b_r^{(j)} - b_{kr-\lfloor \xi r\rfloor}^{(j)}}{a_{kr-\lfloor \xi r\rfloor}^{(j)}} = (k-\xi)^{-\gamma^{(j)}} x^{(j)} + \frac{(k-\xi)^{-\gamma^{(j)}} - 1}{\gamma^{(j)}},$$
$$\lim_{n\to\infty} \frac{a_r^{(j)} x^{(j)} + b_r^{(j)} - b_{r+\lfloor \xi r\rfloor - kr}^{(j)}}{a_{r+\lfloor \xi r\rfloor-kr}^{(j)}} = (1+\xi-k)^{-\gamma^{(j)}} x^{(j)} + \frac{(1+\xi-k)^{-\gamma^{(j)}} - 1}{\gamma^{(j)}}.$$

Arguing as in the proof of Lemma 8.9 from Bücher and Staud (2024b), Condition 4.2.1 and the previous three displays yield, for $n \to \infty$,

$$\mathbb{P}(\boldsymbol{Z}_{r,1+\lfloor \xi r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{x}) = G^{k-\xi}(\boldsymbol{x}) G^{1+\xi-k}(\boldsymbol{x}) + o(1) = G(\boldsymbol{x}) + o(1).$$

We proceed by proving (4.11.1), and start by considering the case $|\xi - \xi'| > k-1$. Without loss of generality let $\xi < \xi'$ such that $1 + \lfloor \xi r\rfloor \le r$ and $1 + \lfloor \xi' r\rfloor > (k-1)r$. Then, by the same arguments as for the marginal convergence,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\big(\boldsymbol{Z}_{1+\lfloor \xi r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{x}, \boldsymbol{Z}_{1+\lfloor \xi' r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{y}\big) \\
&= \lim_{n\to\infty} \mathbb{P}\Big(\frac{\max(\boldsymbol{X}_{r,1}, \dots, \boldsymbol{X}_{r,\lfloor \xi r\rfloor}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{y}, \frac{\max(\boldsymbol{X}_{r,1+\lfloor \xi r\rfloor}, \dots, \boldsymbol{X}_{r,\lfloor \xi' r\rfloor - (k-1)r}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{x} \wedge \boldsymbol{y} \\
&\qquad \frac{\max(\boldsymbol{X}_{r,\lfloor \xi' r\rfloor-(k-1)r+1}, \dots, \boldsymbol{X}_{r,r+\lfloor \xi r\rfloor}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{x}, \frac{\max(\boldsymbol{X}_{r,1+\lfloor \xi' r\rfloor}, \dots, \boldsymbol{X}_{kr}) - \boldsymbol{b}_r}{\boldsymbol{a}_r} \le \boldsymbol{y}\Big) \\
&= G^{\xi}(\boldsymbol{y}) \cdot G^{\xi'-(k-1)-\xi}(\boldsymbol{x}\wedge\boldsymbol{y}) \cdot G^{1+\xi-\xi'+(k-1)}(\boldsymbol{x}) \cdot G^{k-\xi'}(\boldsymbol{y}) \\
&= G^{k-(\xi'-\xi)}(\boldsymbol{y}) \cdot G^{1-(k-(\xi'-\xi))}(\boldsymbol{x}\wedge\boldsymbol{y}) \cdot G^{k-(\xi'-\xi)}(\boldsymbol{x}) \\
&= G_{k-(\xi'-\xi)}(\boldsymbol{x}, \boldsymbol{y}),
\end{aligned}$$

where we used (4.2.3) at the last equality.

Next, consider the case $|\xi - \xi'| \le k-1$, and again assume $\xi < \xi'$. There are two subcases to handle; first let $\xi' \in (k-1, k]$. In that case we have $\xi \ge 1$, hence there is no overlapping from the left (induced by the circ-max-operation). We may then proceed in a similar (but simpler) way as for $|\xi - \xi'| > k-1$ to obtain the stated limit. Alternatively, we have $\xi' \le k-1$. In that case, we are in the sliding block maxima case, where the result is known, see e.g. Lemma 8.6 in Bücher and Staud (2024b).

Finally, the assertion in (4.11.2) follows from similar arguments as in Lemmas A.7, A.8 from Bücher and Segers (2018b), which yield

$$\mathbb{P}\big(\boldsymbol{Z}_{(i-1)kr+1+\lfloor \xi r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{x}, \boldsymbol{Z}_{(i'-1)kr+1+\lfloor \xi' r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{y}\big) = \mathbb{P}\big(\boldsymbol{Z}_{1+\lfloor \xi r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{x}\big) \cdot \mathbb{P}\big(\boldsymbol{Z}_{1+\lfloor \xi' r\rfloor}^{(\mathrm{cb})} \le \boldsymbol{y}\big) + o(1)$$

by stationarity. The product on the right-hand side converges to $G(\boldsymbol{x})G(\boldsymbol{y})$ by the marginal convergence. □

Lemma 4.11.2 (Multiplier central limit theorem)**.** *Suppose Conditions 4.2.1 and 4.2.2 are met. Fix $k \in \mathbb{N}$ and let $\boldsymbol{h} = (h_1, \dots, h_q)^\top$ satisfy Condition 4.10.1 with $\nu > 2/\omega$ and ω from Condition 4.2.2. Let $Y_{n,1}, \dots, Y_{n,m(k)}$ be a triangular array of rowwise independent and identically distributed random variables that is independent of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \dots)$ such that* $\limsup_{n\to\infty} \mathrm{E}[|Y_{n,1}|^{2+\nu}] <$

∞ *with* ν *from Condition 4.10.1 and such that the limit* $\mu_2 := \lim_{n\to\infty} \mathrm{E}[Y_{n,1}^2] > 0$ *exists. Then, for* $n \to \infty$ *and with* $\Sigma_{\boldsymbol{h}}^{(\mathrm{mb})}$ *from (4.2.4), we have*

$$\bar{\mathbb{G}}^{\circ}_{n,r}\boldsymbol{h} := \sqrt{\frac{n}{r}}\frac{1}{n}\sum_{i=1}^{m(k)} Y_{n,i} \sum_{s\in I_{kr,i}} \left\{\boldsymbol{h}(Z_{r,s}^{(\mathrm{cb})}) - \mathrm{E}[\boldsymbol{h}(Z_{r,s}^{(\mathrm{cb})})]\right\} \rightsquigarrow \begin{cases} \mathcal{N}_q(0,\mu_2\Sigma_{\boldsymbol{h}}^{(\mathrm{db})}) & \text{if } k=1,\\ \mathcal{N}_q(0,\mu_2\Sigma_{\boldsymbol{h}}^{(\mathrm{sb})}) & \text{if } k\in\{2,3,\dots\}.\end{cases}$$

Proof. We only consider the case $k \geq 2$; the case $k = 1$ is a simple modification. By the Cramér-Wold Theorem, it is sufficient to consider the case $q = 1$. Since $m(k) = n/(kr)$, we may write $\bar{\mathbb{G}}^{\circ}_{n,r}h = \sqrt{\frac{n}{r}}\frac{1}{m(k)}\sum_{i=1}^{m(k)} Y_{n,i}D_{n,i}$, with $D_{n,i} = \frac{1}{kr}\sum_{s\in I_{kr,i}}\{h(Z_{r,s}^{(\mathrm{cb})}) - \mathrm{E}[h(Z_{r,s}^{(\mathrm{cb})})]\}$ from (4.9.2). Then, since $(Y_{n,i}D_{n,i})_i$ is uncorrelated,

$$\mathrm{Var}(\bar{\mathbb{G}}^{\circ}_{n,r}h) = k\,\mathrm{E}[Y_{n,1}^2]\,\mathrm{Var}(D_{n,1}) = \mu_2\Sigma_h^{(\mathrm{sb})} + o(1)$$

by (4.9.4). This implies the assertion for $\Sigma_h^{(\mathrm{sb})} = 0$.

If $\Sigma_h^{(\mathrm{sb})} > 0$, we may follow a similar line of reasoning as in the proof of Theorem 4.3.2: decompose

$$\bar{\mathbb{G}}^{\circ}_{n,r}h = \frac{1}{\sqrt{p}}\sum_{j=1}^{p}(T_{n,j}^{+} + T_{n,j}^{-}), \qquad \text{where} \qquad T_{n,j}^{\pm} := \sqrt{\frac{np}{r}}\frac{1}{m(k)}\sum_{i\in J_j^{\pm}} Y_{n,i}D_{n,i},$$

where $p = m(k)/m^*$ with $m^* = o(m^{\nu/(2(1+\nu))})$ and $J_j^{+} := \{(j-1)m^*+1,\dots,jm^*-1\}, J_j^{-} := \{jm^*\}$. Following the arguments from the proof of Theorem 4.3.2, we have $p^{-1/2}\sum_{j=1}^{p} T_{n,j}^{-} = o_{\mathbb{P}}(1)$. Moreover, we may assume independence of $(S_{n,j}^{+})_j$, which enables an application of Ljapunov's central limit theorem to $p^{-1/2}\sum_{j=1}^{p} T_{n,j}^{+}$ to conclude. □

The next result is possibly well-known; we skip its elementary proof (see also Section 2 in Klaassen and Wellner, 1992).

Lemma 4.11.3 (Poissonization). *For* $q \in \mathbb{N}_{\geq 2}$ *and* $(p_1,\dots,p_q) \in [0,1]^q$ *with* $p_1 + \dots + p_q = 1$, *let* $(\boldsymbol{U}_{j,q})_{j\in\mathbb{N}}$ *be i.i.d. multinomial vectors with* 1 *trial and* q *classes with class probabilities* $p_1,\dots,p_q$. *For* $m \in \mathbb{N}$ *let* $\boldsymbol{W}_{m,q} = (W_{m,q,1},\dots,W_{m,q,q}) = \sum_{j=1}^{m}\boldsymbol{U}_{j,q}$ *(which is multinomial with* m *trials and* q *classes). Further, for* $\lambda > 0$, *let* N *be a* $\mathrm{Poi}(\lambda)$*-distributed random variable that is independent of* $(\boldsymbol{U}_{j,q})_{j\in\mathbb{N}}$. *Then, the random variables* $W_{N,q,1},\dots,W_{N,q,q}$ *are independent and* $W_{N,q,j}$ *follows a* $\mathrm{Poi}(\lambda p_j)$*-distribution for* $j = 1,\dots,q$.

Lemma 4.11.4. *Let* $d = 1$ *and suppose that* h *is a real-valued function such that* $\mathrm{Var}(h(Z)) \in (0,\infty)$ *exists. Then, for all* $\xi \in [0,1]$, $\mathrm{Cov}(h(Z_{1,\xi}),h(Z_{2,\xi})) \geq 0$ *with strict inequality in a neighbourhood of* 0.

Proof. By our assumption $d = 1$, the cdf G from (4.2.2) corresponds to the GEV(γ)- distribution, which has a Lebesgue-density that we denote by g. For $\xi \in (0,1)$, we may employ the following stochastic construction for $(Z_{1,\xi},Z_{2,\xi})^{\top}$: let U,W be i.i.d. with cdf G^{ξ} and independent of V, the latter having the cdf $G^{1-\xi}$. We then have $\mathcal{L}(Z_{1,\xi},Z_{2,\xi}) = \mathcal{L}(U\vee V, W\vee V)$. Without loss of generality assume $\mathrm{E}[h(Z_{1,\xi})] = 0$ and note that

$$\mathrm{Cov}(h(Z_{1,\xi}),h(Z_{2,\xi})) = \mathrm{E}[h(U\vee V)h(W\vee V)]$$

$$= 2\,\mathrm{E}[h(U)h(V)G^{\xi}(V)\mathbf{1}\{V < U\}] \\ + \mathrm{E}[h(U)h(W)G^{1-\xi}(U \wedge W)] + \mathrm{E}[h^2(V)G^{2\xi}(V)] \equiv 2I_1 + I_2 + I_3. \tag{4.11.3}$$

Let $H : \bar{\mathbb{R}} \to \mathbb{R}, H(x) = \int_{-\infty}^{x} h(z)g(z)\,\mathrm{d}z$ and note that $H(\infty) = 0 = H(-\infty)$, where we made use of $\mathrm{E}[h(Z_{1,\xi})] = \int_{\mathbb{R}} h(z)g(z)\,\mathrm{d}z$. Using integration by parts we obtain

$$I_1 = (1-\xi)\xi \int_{-\infty}^{\infty} G^{\xi-1}(x)H'(x)H(x)\,\mathrm{d}x = \frac{\xi}{2}(1-\xi)^2 \int_{-\infty}^{\infty} G^{\xi-2}(x)g(x)H^2(x)\,\mathrm{d}x \geq 0.$$

Similarly, we have $I_2 = \xi^2(1-\xi)\int_{-\infty}^{\infty} G^{\xi-2}(x)g(x)H^2(x)\,\mathrm{d}x \geq 0$. In view of (4.11.3) and $I_3 \geq 0$ this proves the first assertion.

For the second assertion note that $I_3 = (1-\xi)\,\mathrm{E}[h^2(Z)G^{\xi}(Z)] \leq \mathrm{Var}(h(Z)) < \infty$ and $I_j \geq 0$ for $j = 1, 2, 3$. By dominated convergence and (4.11.3) we then have

$$\liminf_{\xi \downarrow 0} \mathrm{Cov}(h(\boldsymbol{W}_{1,\xi}), h(\boldsymbol{W}_{2,\xi})) \geq \liminf_{\xi \downarrow 0} I_3 = \mathrm{Var}(h(Z)) > 0,$$

which lets us conclude. □

4.11.1 A general result on bootstrapping the Fréchet independence MLE

Throughout this section, let $\mathcal{X}_n = (X_{n,1}, \dots, X_{n,\omega_n})$ denote a sequence of random vectors in $(0,\infty)^{\omega_n}$ with continuous cumulative distribution functions $F_{n,1}, \dots, F_{n,\omega_n}$, where $\omega_n \to \infty$ is a sequence of integers. Throughout, we assume that asymptotically not all observations are tied, that is,

$$\lim_{n\to\infty} \mathbb{P}(X_{n,1} = \cdots = X_{n,\omega_n}) = 0. \tag{4.11.4}$$

For some $\alpha_0 > 0$ and $\sigma_n \to \infty$, proximity of the $F_{n,1}, \dots, F_{n,\omega_n}$ to the Fréchet-distribution P_{α_0,σ_n} will be controlled by convergence conditions on empirical means of $X_{n,i}/\sigma_n$: for $0 < \alpha_- < \alpha_0 < \alpha_+ < \infty$, let $\mathcal{F}(\alpha_-, \alpha_+)$ denote the set of functions containing $x \mapsto \log x, x \mapsto x^{-\alpha}, x \mapsto x^{-\alpha}\log x$ and $x \mapsto x^{-\alpha}\log^2 x$ for all $\alpha \in (\alpha_-, \alpha_+)$. Moreover, recall $f_1(x) = x^{-\alpha_0}$, $f_2(x) = x^{-\alpha_0}\log x$ and $f_3(x) = \log x$ from (4.5.4), considered as functions on $(0,\infty)$, and let $\mathcal{H} = \{f_1, f_2, f_3\}$. For a real-valued function f defined on $(0,\infty)$ such that the following integrals/expectations exist, let

$$\mathbb{P}_n f = \frac{1}{\omega_n}\sum_{i=1}^{\omega_n} f(X_{n,i}/\sigma_n), \quad P_n f = \frac{1}{\omega_n}\sum_{i=1}^{\omega_n} \mathrm{E}[f(X_{n,i}/\sigma_n)], \quad P_{\alpha_0,1} f := \int f(x)\,\mathrm{d}P_{\alpha_0,1}(x).$$

Condition 4.11.5. (i) There exists $0 < \alpha_- < \alpha_0 < \alpha_+ < \infty$ such that $\mathbb{P}_n f \rightsquigarrow P_{\alpha_0,1} f$ for all $f \in \mathcal{F}(\alpha_-, \alpha_+)$.

(ii) There exists $0 < v_n \to \infty$ such that $\bar{\mathbb{G}}_n(f_1, f_2, f_3)^\top \rightsquigarrow N_3(\mathbf{0}, \Sigma_G)$, where $\bar{\mathbb{G}}_n f = v_n(\mathbb{P}_n f - P_n f)$ and where $\Sigma_G \in \mathbb{R}^{3\times 3}$ is positive semidefinite.

(iii) With v_n from (ii), the limit $\boldsymbol{B}_G := \lim_{n\to\infty} B_n(f_1, f_2, f_3)^\top$ exists, where $B_n f = v_n(P_n f - P_{\alpha_0,1} f)$.

Recall the log-likelihood function ℓ_θ of the Fréchet distribution from (4.5.2).

Theorem 4.11.6 (Bücher and Segers, 2018b). *(i) Suppose that (4.11.4) holds and that Condition 4.11.5(i) is met. Then, on the complement of the event $\{X_{n,1} = \cdots = X_{n,\omega_n}\}$, the independence Fréchet log-likelihood $\boldsymbol{\theta} \mapsto \sum_{i=1}^{\omega_n} \ell_{\boldsymbol{\theta}}(X_{n,i})$ has a unit maximizer $\hat{\boldsymbol{\theta}}_n = (\hat\alpha_n, \hat\sigma_n)$, and that maximizer is consistent in the sense that $(\hat\alpha_n, \hat\sigma_n/\sigma_n) = (\alpha_0, 1) + o_{\mathbb{P}}(1)$ as $n \to \infty$.*

(ii) If, additionally, Condition 4.11.5(ii)-(iii) is met, then

$$v_n \begin{pmatrix} \hat\alpha_n - \alpha_0 \\ \hat\sigma_n/\sigma_n - 1 \end{pmatrix} = M(\alpha_0)(\bar{\mathbb{G}}_n + B_n)(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1) \rightsquigarrow \mathcal{N}_2(M(\alpha_0)\boldsymbol{B}_G, M(\alpha_0)\Sigma_G M(\alpha_0)^\top)$$

as $n \to \infty$, where

$$M(\alpha_0) = \frac{6}{\pi^2}\begin{pmatrix} \alpha_0^2 & \alpha_0(1-\gamma) & -\alpha_0^2 \\ \gamma - 1 & -(\Gamma''(2)+1)/\alpha_0 & 1-\gamma \end{pmatrix}$$

with $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}t$ the Euler Gamma function and $\gamma = 0.5772\ldots$ the Euler-Mascheroni constant.

Now, conditional on $\mathcal{X}_n = (X_{n,1}, \ldots, X_{n,\omega_n})$, let $\mathcal{X}_n^* = (X_{n,1}^*, \ldots, X_{n,\omega_n}^*)$ denote a bootstrap sample of $\mathcal{X}_n$; formally, $\mathcal{X}_n^*$ is assumed to be a measurable function of both $\mathcal{X}_n$ and of some additional independent random element $\mathcal{W}_n$ taking values in some measurable space. For the subsequent consistency statements, the bootstrap sample is assumed to satisfy the not-all-tied condition

$$\lim_{n\to\infty} \mathbb{P}(X_{n,1}^* = \cdots = X_{n,\omega_n}^*) = 0 \tag{4.11.5}$$

or, equivalently, $\mathbb{P}\big(X_{n,1}^* = \cdots = X_{n,\omega_n}^* \mid \mathcal{X}_n\big) = o_{\mathbb{P}}(1)$ as $n \to \infty$. The bootstrap scheme is assumed to be regular in the sense that the conditional distribution of certain rescaled arithmetic means is close to the distribution of respective arithmetic means of the original sample. For a real-valued function f on $(0,\infty)$, define $\hat{\mathbb{P}}_n^* f = \frac{1}{\omega_n}\sum_{i=1}^{\omega_n} f(X_{n,i}^*/\sigma_n)$ and $\hat{\mathbb{G}}_n^* f = v_n(\hat{\mathbb{P}}_n^* f - \mathbb{P}_n f)$.

Theorem 4.11.7. *Suppose that (4.11.4), (4.11.5) and Condition 4.11.5 is met, that $\hat{\mathbb{P}}_n^* f = P_{\alpha_0,1} f + o_{\mathbb{P}}(1)$ for all $f \in \mathcal{F}(\alpha_-, \alpha_+)$ (or, equivalently, $\mathbb{P}(|\hat{\mathbb{P}}_n^* f - P_{\alpha_0,1}| > \varepsilon \mid \mathcal{X}_n) = o_{\mathbb{P}}(1)$ for all $\varepsilon > 0$ and all $f \in \mathcal{F}(\alpha_-, \alpha_+)$) and that*

$$d_K\big(\mathcal{L}(\hat{\mathbb{G}}_n^*(f_1, f_2, f_3)^\top \mid \mathcal{X}_n), \mathcal{N}_3(\mathbf{0}, \Sigma_G)\big) = o_{\mathbb{P}}(1). \tag{4.11.6}$$

Then, on the complement of the event $\{X_{n,1}^* = \cdots = X_{n,\omega_n}^*\}$, the independence Fréchet-log-likelihood $\boldsymbol{\theta} \mapsto \sum_{i=1}^{\omega_n} \ell_{\boldsymbol{\theta}}(X_{n,i}^*)$ has a unit maximizer $\boldsymbol{\theta}_n^* = (\hat\alpha_n^*, \hat\sigma_n^*)$, and that maximizer satisfies

$$v_n \begin{pmatrix} \hat\alpha_n^* - \hat\alpha_n \\ \hat\sigma_n^*/\hat\sigma_n - 1 \end{pmatrix} = M(\alpha_0)\hat{\mathbb{G}}_n^*(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1). \tag{4.11.7}$$

As a consequence, if additionally $\boldsymbol{B}_G = 0$, we have bootstrap consistency in the following sense

$$d_K\left[\mathcal{L}\left(v_n\begin{pmatrix} \hat\alpha_n^* - \hat\alpha_n \\ \hat\sigma_n^*/\hat\sigma_n - 1 \end{pmatrix} \,\middle|\, \mathcal{X}_n\right), \mathcal{L}\left(v_n\begin{pmatrix} \hat\alpha_n - \alpha_0 \\ \hat\sigma_n/\sigma_n - 1 \end{pmatrix}\right)\right] = o_{\mathbb{P}}(1). \tag{4.11.8}$$

Since it concerns empirical means only, the condition in (4.11.6) is typically a standard result for the bootstrap; see for instance Section 10 and 11 in Kosorok (2008).

Proof of Theorem 4.11.7. First, by Theorem 4.11.6 (which is Theorem 2.3, Theorem 2.5 and Addendum 2.6 in Bücher and Segers, 2018b applied to the sample $\mathcal{X}_n$), we have

$$v_n \begin{pmatrix} \hat\alpha_n - \alpha_0 \\ \hat\sigma_n/\sigma_n - 1 \end{pmatrix} = M(\alpha_0)(\bar{\mathbb{G}}_n + B_n)(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1) \rightsquigarrow M(\alpha_0)\mathcal{N}_3(\boldsymbol{B}_G, \Sigma_G). \tag{4.11.9}$$

Next, in view of the fact that (4.11.6) implies $\hat{\mathbb{G}}_n^*(f_1, f_2, f_3)^\top \rightsquigarrow \mathcal{N}_3(\mathbf{0}, \Sigma_G)$ (unconditionally) by Lemma 2.3 in Bücher and Kojadinovic (2019), we have

$$\mathbb{G}_n^\circ(f_1, f_2, f_3)^\top := v_n\big(\hat{\mathbb{P}}_n^*(f_1, f_2, f_3)^\top - P_{\alpha_0,1}(f_1, f_2, f_3)^\top\big) \rightsquigarrow \mathcal{N}_3(\boldsymbol{B}_G, \Sigma_G).$$

Hence, we may apply Theorem 2.3, Theorem 2.5 and Addendum 2.6 in Bücher and Segers (2018b) to the sample $\mathcal{X}_n^*$ (unconditionally) to obtain that

$$v_n \begin{pmatrix} \hat\alpha_n^* - \alpha_0 \\ \hat\sigma_n^*/\sigma_n - 1 \end{pmatrix} = M(\alpha_0)\mathbb{G}_n^\circ(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1) \tag{4.11.10}$$

As a consequence, by (4.11.9), (4.11.10) and $\hat\sigma_n/\sigma_n = 1 + o_{\mathbb{P}}(1)$, we have

$$\begin{aligned} v_n \begin{pmatrix} \hat\alpha_n^* - \hat\alpha_n \\ \hat\sigma_n^*/\hat\sigma_n - 1 \end{pmatrix} &= v_n \begin{pmatrix} \hat\alpha_n^* - \alpha_0 - (\hat\alpha_n - \alpha_0) \\ \sigma_n/\hat\sigma_n\big(\sigma_n^*/\sigma_n - 1 - (\hat\sigma_n/\sigma_n - 1)\big) \end{pmatrix} \\ &= \{1 + o_{\mathbb{P}}(1)\}\big(M(\alpha_0)(\mathbb{G}_n^\circ - \mathbb{G}_n - B_n)(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1)\big) \\ &= M(\alpha_0)\hat{\mathbb{G}}_n^*(f_1, f_2, f_3)^\top + o_{\mathbb{P}}(1), \end{aligned} \tag{4.11.11}$$

which is (4.11.7).

Finally, for the proof of (4.11.8), let $\mathcal{X}_n^\#$ denote a second bootstrap sample, generated in the same way as $\mathcal{X}_n^*$ and independent of $\mathcal{X}_n^*$, conditionally on $\mathcal{X}_n$. Denote the respective estimators and empirical measures/processes by $\hat\alpha_n^\#, \hat{\mathbb{P}}_n^\#$ etc. Then, by the expansions in (4.11.11),

$$v_n \begin{pmatrix} \hat\alpha_n^* - \hat\alpha_n \\ \hat\sigma_n^*/\hat\sigma_n - 1 \\ \hat\alpha_n^\# - \hat\alpha_n \\ \hat\sigma_n^\#/\hat\sigma_n - 1 \end{pmatrix} = \begin{pmatrix} M(\alpha_0)(\hat{\mathbb{G}}_n^*)(f_1, f_2, f_3)^\top \\ M(\alpha_0)(\hat{\mathbb{G}}_n^\#)(f_1, f_2, f_3)^\top \end{pmatrix} + o_{\mathbb{P}}(1).$$

Equation (4.11.6) and Lemma 2.3 in Bücher and Kojadinovic (2019) implies that the dominating term on the right-hand side converges weakly (unconditionally) to $\mathcal{N}_2(\mathbf{0}, M(\alpha_0)\Sigma_G M(\alpha_0)^\top)^{\otimes 2}$, which by another reverse application of that lemma implies

$$d_K\left[\mathcal{L}\left(v_n \begin{pmatrix} \hat\alpha_n^* - \hat\alpha_n \\ \hat\sigma_n^*/\hat\sigma_n - 1 \end{pmatrix} \,\Big|\, \mathcal{X}_n\right), \mathcal{N}_2(\mathbf{0}, M(\alpha_0)\Sigma_G M(\alpha_0)^\top)\right] = o_{\mathbb{P}}(1).$$

The assertion then follows from the triangular inequality and (4.11.9), noting that $\boldsymbol{B}_G = \mathbf{0}$ by assumption. □

4.11.2 Auxiliary results on bootstrap consistency

Lemma 4.11.8. *Let $S_n, T_n \in \mathbb{R}^q$ with $S_n \rightsquigarrow S, T_n \rightsquigarrow S \in \mathbb{R}^q$. Furthermore, let $\mathcal{L}(S)$ be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^q$. We then have*

$$\lim_{n\to\infty} d_K\left[\mathcal{L}(A_n S_n), \mathcal{L}(A_n T_n)\right] = 0,$$

for any sequence $(A_n)_n \subset \mathbb{R}^{p\times q}$ of matrices.

Proof. Write $A_n = (a_{n,j}^\top)_{j=1,\dots,p}$ with $a_{n,j} \in \mathbb{R}^q$. A straightforward argument shows that we may assume $a_{n,j} \neq 0$ for all $j = 1,\dots p$. Next, note that, for all diagonal matrices $D = \operatorname{diag}(d_j)_{j=1,\dots p}$ with $d_j \neq 0$ and all $\mathbb{R}^p$-valued random variables S, T, we have $d_K[\mathcal{L}(S), \mathcal{L}(T)] = d_K[\mathcal{L}(DS), \mathcal{L}(DT)]$. Hence, letting $d_j := \|a_{n,j}\|_2^{-1}$ and $\tilde{A}_n = (\tilde{a}_{n,j}^\top)_{j=1,\dots,p}$ with normed $\tilde{a}_{n,j} = a_{n,j}/\|a_{n,j}\|_2 \in \mathbb{R}^p$, we have $\tilde{A}_n = DA_n$ and therefore

$$d_K\left[\mathcal{L}(A_n S_n), \mathcal{L}(A_n T_n)\right] = d_K\left[\mathcal{L}(\tilde{A}_n S_n), \mathcal{L}(\tilde{A}_n T_n)\right]. \tag{4.11.12}$$

Since $\tilde{A}_n \in [-1,1]^{p\times q}$, the Bolzano-Weierstraß Theorem allows to find a subsequence $n' := n'(n)$ such that $E = \lim_{n\to\infty} \tilde{A}_{n'}$ exists. Slutsky's lemma then yields the weak convergences $\tilde{A}_{n'} S_{n'} \rightsquigarrow ES$ and $\tilde{A}_{n'} T_{n'} \rightsquigarrow ES$. We will show below that the cdf of ES is continuous. Since the Kolmogorov-metric d_K metrizes weak convergence to limits with continuous cdfs (van der Vaart, 1998, Lemma 2.11), an application of the triangular inequality implies

$$d_K\left[\mathcal{L}(\tilde{A}_{n'} S_{n'}), \mathcal{L}(\tilde{A}_{n'} T_{n'})\right] \le d_K\left[\mathcal{L}(\tilde{A}_{n'} S_{n'}), \mathcal{L}(ES)\right] + d_K\left[\mathcal{L}(ES), \mathcal{L}(\tilde{A}_{n'} T_{n'})\right] = o(1)$$

for $n' \to \infty$. As the previous argumentation can be repeated for subsequences n' of arbitrary subsequences n'', we then have $d_K\left[\mathcal{L}(\tilde{A}_n S_n), \mathcal{L}(\tilde{A}_n T_n)\right] = o(1)$, which implies the assertion by (4.11.12).

It remains to show that the cdf of ES is continuous. For that purpose, by Sklar's theorem and the fact that copulas are continuous, it is sufficient to show that each marginal cdf of ES is continuous. Hence, fix $j \in \{1,\dots,p\}$, and note that the jth row of E, say $e_j^\top$, has Euclidean norm 1. We may therefore construct an invertible matrix $\tilde{E} = \tilde{E}(e_j) \in \mathbb{R}^{q\times q}$ with the first row of $\tilde{E}$ being equal to $e_j^\top$. Then $\tilde{E}$ being a diffeomorphism and S being absolutely continuous with respect to the Lebesgue-measure on $\mathbb{R}^q$ implies the latter for $\tilde{E}S$. Hence, for any $y \in \mathbb{R}$, $\mathbb{P}((ES)_j = y) = \mathbb{P}\left(\tilde{E}S \in \{y\} \times \mathbb{R}^{q-1}\right) = 0$. Since $y \in \mathbb{R}$ was arbitrary, this proves the continuity of $y \mapsto \mathbb{P}((ES)_j \le y)$. □

Lemma 4.11.9. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ denote a probability space and $p, q \in \mathbb{N}$. For $n \in \mathbb{N}$, let $X_n : \Omega \to \mathcal{X}_n$, $W_n : \Omega \to \mathcal{W}_n$ denote random variables in some measurable space $\mathcal{X}_n, \mathcal{W}_n$, respectively. Let $S_n = S_n(X_n)$ and $S_n^* = S_n(X_n, W_n)$ be $\mathbb{R}^q$-valued statistics. If*

(a) $d_w\left(\mathcal{L}(S_n^* \mid X_n), \mathcal{L}(S_n)\right) \xrightarrow{\mathbb{P}} 0,$ *as $n \to \infty$;*

(b) $S_n \rightsquigarrow Q,$ *as $n \to \infty$;*

where Q is absolutely continuous with respect to the Lebesgue-measure on $\mathbb{R}^q$ and d_w denotes any metric characterizing weak convergence on $\mathbb{R}^q$; then

$$d_K\big(\mathcal{L}(A_n S_n^* \mid X_n), \mathcal{L}(A_n S_n)\big) \xrightarrow{\mathbb{P}} 0, \qquad \text{as } n \to \infty,$$

for any sequence $(A_n)_n \subset \mathbb{R}^{p\times q}$ of matrices.

Proof. Let $n' \subset \mathbb{N}$ denote a subsequence of $\mathbb{N}$. By assumption (a), we may choose a further subsequence n'' of n' and $\Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that

$$\lim_{n\to\infty} d_w\big(\mathcal{L}(S_{n''}^* \mid X_{n''}), \mathcal{L}(S_{n''})\big) = 0 \quad \text{on } \Omega_0.$$

Hence, since $S_n \rightsquigarrow Q$ by assumption, we obtain $\mathcal{L}(S_{n''}^* \mid X_{n''}) \rightsquigarrow Q$ on Ω_0. Lemma 4.11.8 then implies

$$\lim_{n\to\infty} d_K\big(\mathcal{L}(A_{n''} S_{n''}^* \mid X_{n''}), \mathcal{L}(A_{n''} S_{n''})\big) = 0 \quad \text{on } \Omega_0,$$

which lets us conclude. □

4.12 Appendix

4.12.1 Details on mean estimators in the ARMAX-GPD model

Corollary 4.12.1. *Suppose $(X_t)_{t\in\mathbb{Z}}$ is an ARMAX-GPD time series as in Model 4.6.1, for some $\beta \in [0,1)$ and some $\gamma < 1/2$. Then, if the block size parameter satisfies $r = o(n)$ and $\log n = o(r^{1/2})$,*

$$\frac{\sqrt{n/r}}{(r(1-\beta))^\gamma}\big(\hat\mu_n^{(\mathrm{mb})} - \mu_r\big) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{mb}}^2), \quad \mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}\},$$

with $\sigma_{\mathrm{cb}}^2 = \sigma_{\mathrm{sb}}^2 < \sigma_{\mathrm{db}}^2$ as provided in (4.6.2). The ratio $\gamma \mapsto \sigma_{\mathrm{db}}^2/\sigma_{\mathrm{sb}}^2(\gamma)$ is presented in Figure 4.10.

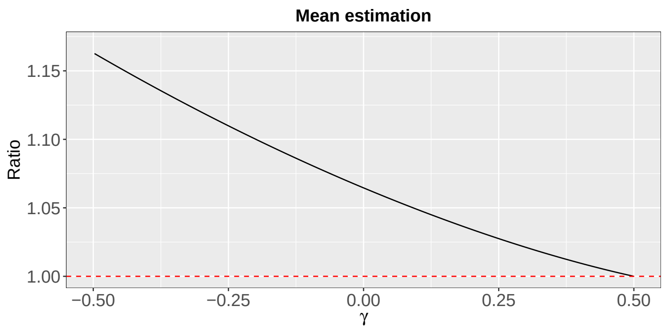

Figure 4.10: Ratio of the asymptotic variances $\sigma_{\mathrm{db}}^2/\sigma_{\mathrm{sb}}^2$ from (4.6.2).

Proof. The proof is akin to the one of Equation (4.3) in Bücher and Staud (2024b); most details are omitted for the sake of brevity and we only discuss the calculation of the asymptotic variance parameter σ_{sb}^2 (note that σ_{db}^2 simply corresponds to the variance of the GEV(γ)-distribution by definition).

We start by considering the case $\gamma \neq 0$. For $i \in \{1,2\}$, the random variables $S_{i,\xi} := (1 + \gamma Z_{i,\xi})^{1/\gamma}$ are standard exponentially distributed, and we have $\mathrm{Cov}(Z_{1,\xi}, Z_{2,\xi}) = \gamma^{-2}\,\mathrm{Cov}(S_{1,\xi}^{-\gamma}, S_{2,\xi}^{-\gamma}) =: \gamma^{-2} C_\xi$. For $\gamma < 0$, Equation (C.8) in the supplement of Bücher and Staud (2024b) implies that

$$\int_0^1 C_\xi \,\mathrm{d}\xi = 2\gamma^2\Gamma(-2\gamma)\int_0^{1/2} (\alpha_{2\gamma}(w) - 1)\big(w^{-\gamma-1}(1-w)^{-\gamma-1}\,\mathrm{d}w\big) = 2\gamma^2\Gamma(-2\gamma)I(\gamma),$$

hence $\sigma_{\mathrm{sb}}^2 = 4\Gamma(-2\gamma)I(\gamma)$. For $\gamma > 0$, the display above (C.9) in the same reference implies

$$\int_0^1 C_\xi \,\mathrm{d}\xi = -\gamma\frac{\Gamma(1-2\gamma}{2}I(\gamma).$$

Hence, $\sigma_{\mathrm{sb}}^2 = -\Gamma(1-2\gamma)/\gamma I(\gamma)$.

Finally, for $\gamma = 0$ consider the transformed random variables $S_{i,\xi} = \exp(-Z_{i,\xi})$ for $i \in \{1,2\}$, which gives $\mathrm{Cov}(Z_{1,\xi}, Z_{2,\xi}) = \mathrm{Cov}(\log S_{1,\xi}, \log S_{2,\xi}) = C_\xi$. By formula (C.11) from the same reference we have

$$\int_0^1 C_\xi \,\mathrm{d}\xi = \int_0^1 \frac{1}{w(1-w)}\int_0^1 -\log(A_\xi(w))\,\mathrm{d}\xi\,\mathrm{d}w,$$

where $A_\xi(w) = \xi + (1-\xi)\big(w \wedge (1-w)\big)$ is the Pickands-dependence function of the associated Copula C_ξ of the Marshall-Olkin distribution with dependence parameter ξ. It follows that

$$\int_0^1 C_\xi \,\mathrm{d}\xi = 2\int_{1/2}^1 \frac{1}{w(1-w)}\int_0^1 -\log(\xi + (1-\xi)w)\,\mathrm{d}\xi\,\mathrm{d}w = \log 4 - 1,$$

which implies the asserted formula. □

4.12.2 Runtime comparison

In classical situations, the computational cost of the bootstrap depends linearly on the number of bootstrap replications and is therefore high if a single evaluation of a statistic of interest is computationally intensive. Since both the sliding and the circular block maxima samples are much larger than the plain disjoint block maxima sample (sample sizes n vs. n/r, respectively), one may naively think that the former methods require substantially more computational resources. As we will argue below and prove with simulations, this naive heuristic is not correct.

For any block maxima method, the starting point for a single evaluation of a statistic of interest is the calculation of the respective block maxima samples, which requires evaluating $O(n/r)$ maxima for the disjoint block maxima method, or $O(n)$ maxima for the sliding and circular block maxima method. Subsequently, when the statistic of interest is applied to one of the samples, the fact that both the sliding and the circmax samples can be efficiently stored as a weighted sample of size $O(n/r)$ implies that the additional computational cost of the latter two methods is, approximately, only a constant multiple of the additional cost for the plain disjoint block maxima method. Next, for the bootstrap approaches proposed within this paper, no additional evaluation of maxima is ever required, whence, overall, the only major difference between the three approaches is the initial calculation of the block maxima samples. Therefore, the relative computational effort should only depend on n, and only moderately.

The above heuristic has been confirmed by Monte Carlo simulations. Exemplary results are presented in Figure 4.11, which rely on simulated data from Model 4.6.2 (parameters: $\beta = 0.5$ and $\alpha = 1.5$) with fixed block size $r = 90$ and total sample size ranging form $40 \cdot 90 = 3,600$ up to $100 \cdot 90 = 9,000$. The target parameter is the runtime for calculating $B \in \{250, 500, 750, 1,000\}$ bootstrap replicates of $\hat{\theta}_n^{(\mathrm{mb})}$ from (4.5.1) for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}, \mathrm{cb}(2), \mathrm{cb}(3)\}$, assessed by taking the median over $N = 500$ repetitions each. The disjoint blocks method has been considered as a benchmark, whence we depict relative runtimes with respect to that method. We find that, as expected, the relative runtime is mostly depending on n, with only a moderate loss in performance for the circmax-method. Similar results were obtained for other estimators and models.

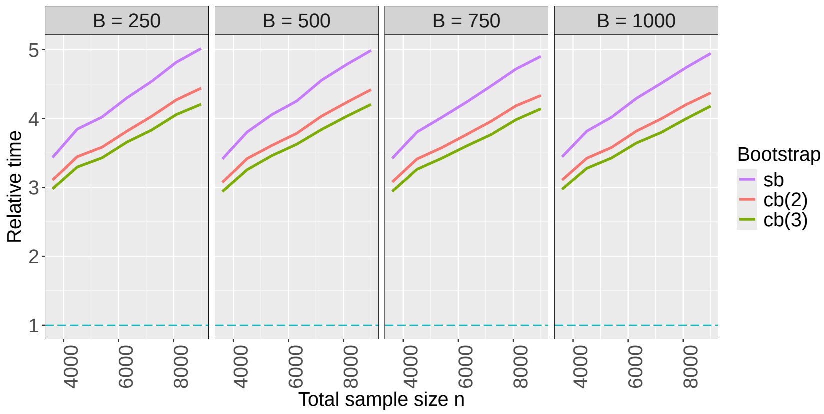

Figure 4.11: Relative median runtimes of different bootstrap algorithms for bootstrapping $\hat{\theta}_n^{(\mathrm{mb})}$ (relative to the runtime of the disjoint blocks bootstrap) for fixed blocksize $r = 90$ as a function of the effective sample size and for different numbers of bootstrap replicates B.

Acknowledgements

Financial support by the German Research Foundation (DFG grant number 465665892) and by Ruhr University Research School (funded by Germany's Excellence Initiative - DFG GSC 98/3) is gratefully acknowledged. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf. The authors are grateful to Johan Segers for fruitful discussions on the circular block maxima method, and to the participants of the Oberwolfach Workshop on "Mathematics, Statistics, and Geometry of Extreme Events in High Dimensions" for their valuable comments.

5 On the maximal correlation coefficient for the bivariate Marshall Olkin distribution

In this section we present the preprint Bücher and Staud (2024c) which is concerned with deriving the maximal correlation coefficient of the Marshall Olkin bivariate exponential distribution. This is in line with the article sharing policy of the *Statistics and Probability Letters*. Only minor changes to improve the presentation within this thesis have been made.

Abstract

We prove a formula for the maximal correlation coefficient of the bivariate Marshall Olkin distribution that was conjectured in Lin, Lai, and Govindaraju (2016, Stat. Methodol., 29:1–9). The formula is applied to obtain a new proof for a variance inequality in extreme value statistics that links the disjoint and the sliding block maxima method.

Keywords. Bivariate Exponential Distribution; Disjoint and Sliding Block Maxima; Extreme Value Statistics; Marshall Olkin Copula; Maximal Correlation Coefficient.

5.1 Introduction

The bivariate Marshall Olkin exponential distribution (Marshall and Olkin, 1967) arises from considering random lifetimes within a two-component system, say (X_1, X_2), where the components are subject to three different sources of fatal shocks. The occurrence times of the shocks are modeled by three independent exponential variables Z_1, Z_2, Z_{12} with positive parameters $\lambda_1, \lambda_2, \lambda_{12}$, respectively. The first component of the system fails as soon as any of the two shocks Z_1 or Z_{12} has occurred, that is, at time $X_1 = Z_1 \wedge Z_{12}$. Likewise, the second component fails at time $X_2 = Z_2 \wedge Z_{12}$. A straightforward calculation then shows that the joint survival function of (X_1, X_2) is

$$\begin{aligned}\bar{H}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2) &= \mathbb{P}(Z_1 > x_1, Z_2 > x_2, Z_{12} > x_1 \vee x_2) \\ &= \exp\{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12}(x_1 \vee x_2)\} \qquad (x_1, x_2 > 0),\end{aligned}$$

while the marginal survival functions satisfy $\bar{H}_j(x_j) = \mathbb{P}(X_j > x_j) = \exp\{-(\lambda_j + \lambda_{12})x_j\}$. In particular, the marginals are exponentially distributed.

The Marshall Olkin distribution has been well-studied in the literature, with precise formulas being available for its Laplace transform, its product moments, or its Pearson, Kendall or Spearman correlation. We refer to Marshall and Olkin (1967); Lin et al. (2016), among others. The present work is motivated by an open problem mentioned in Lin et al. (2016) which concerns the maximal correlation coefficient of the Marshall Olkin distribution (see their open problem B). Dating back to Gebelein (1941), the maximal correlation has been well-researched. The Gebelein-Lancaster Theorem (Gebelein, 1941; Lancaster, 1957) states the remarkable property, that for a bivariate normal distribution with Pearson correlation ρ the coefficient is given by $|\rho|$. For further properties, see Yu (2008) among others. The maximal correlation has applications in numerous fields of statistics, such as optimal transport in regression Breiman and Friedman (1985). Recall that the maximal correlation coefficient is defined as

$$R(H) := R(X_1, X_2) := \sup_{f,g} \operatorname{Corr}(f(X_1), g(X_2)), \tag{5.1.1}$$

where the supremum is taken over all functions f and g such that $\operatorname{Var}(f(X_1)), \operatorname{Var}(g(X_2)) \in (0, \infty)$ exists and where H denotes the cumulative distribution function (cdf) of (X_1, X_2). Dating back to Gebelein (1941), the maximal correlation coefficient has been extensively studied. For instance, it is known to satisfy a number of desirable properties of a dependence measure (Rényi, 1959) and it has applications in numerous fields of statistics, such as optimal transport for regression (Breiman and Friedman, 1985). We refer to Yu (2008) and the references therein for further properties and applications. In general, the calculation of maximal correlation coefficients is difficult (with the Gebelein-Lancaster theorem (Gebelein, 1941; Lancaster, 1957) being a notable exception: for a bivariate normal distribution with Pearson correlation ρ the coefficient is given by $|\rho|$), but based on extensive moment calculations, Lin et al. (2016) conjecture that

$$R(H) = R(X_1, X_2) = \frac{\lambda_{12}}{\sqrt{\lambda_1 + \lambda_{12}}\sqrt{\lambda_2 + \lambda_{12}}} \tag{5.1.2}$$

for the bivariate Marshall Olkin distribution H. The main result of this note is a proof, given in Section 5.2. Our proof is based on certain elegant arguments from Yu (2008), who derived a new proof of the Gebelein-Lancaster theorem.

Next to the proof of (5.1.2), a major contribution of this note is an application of (5.1.2) to provide a new and elegant proof for an important variance inequality in extreme value statistics. Details are provided in Section 5.3.

5.2 The maximal correlation for the Marshall Olkin distribution

In view of the continuity of the marginal survival functions, Sklar's theorem implies that the random vector (X_1, X_2) has a unique survival copula $\hat{C}$, that is, a bivariate cdf $\hat{C}$ with standard uniform margins, such that $\bar{H}(x_1, x_2) = \hat{C}(\bar{H}_1(x_1), \bar{H}_2(x_2))$ for all $x_1, x_2 \geq 0$ (Nelsen, 2006). A straightforward calculation shows that this copula is given by $\hat{C} = C_{\phi,\psi}$, where

$$C_{\phi,\psi}(u, v) = \min(u^{1-\phi}v, uv^{1-\psi}), \quad u, v \in [0, 1]^2, \tag{5.2.1}$$

with $\phi = \lambda_{12}/(\lambda_1 + \lambda_{12})$ and $\psi = \lambda_{12}/(\lambda_2 + \lambda_{12})$, see also Lin et al. (2016); Embrechts et al. (2001).

Theorem 5.2.1. *For parameters $\phi, \psi \in [0, 1]^2$, we have $R(C_{\phi,\psi}) = \sqrt{\phi\psi}$.*

Equation (5.1.2) is now an immediate corollary of the former theorem; we state it for the sake of reference.

Corollary 5.2.2 (Open problem B in Lin et al., 2016)**.** *The bivariate Marshall Olkin distribution H with parameters* $\lambda_1, \lambda_2, \lambda_{12} > 0$ *satisfies* (5.1.2).

Proof. Since the maximal correlation coefficient is invariant under (square-integrable) transformations of the margins and recalling the definitions of ϕ, ψ, we have $R(H) = R(\hat{C}) = R(C_{\phi,\psi}) = \sqrt{\phi\psi} = \lambda_{12}/\big(\sqrt{\lambda_1 + \lambda_{12}}\sqrt{\lambda_2 + \lambda_{12}}\big)$ by Theorem 5.2.1. □

The proof of Theorem 5.2.1 is based on the following two lemmas from Yu (2008), which we quote in full for the sake of readability.

Lemma 5.2.3 (Yu, 2008)**.** *If non-degenerate random variables* X *and* Y *are conditionally independent given* Z*, then* $R(X,Y) \leq R(X,Z)R(Y,Z)$*. Moreover, equality holds if* (X,Z) *and* (Y,Z) *have the same distribution.*

Lemma 5.2.4 (Yu, 2008)**.** *If non-degenerate random variables* X *and* Y *are independent and identically distributed, and* $Z = f(X,Y)$*, where* f *is a symmetric function of* x *and* y*, then* $R(X,Z) \leq 2^{-1/2}$.

Proof of Theorem 5.2.1. The assertion is trivial if either $\phi \in \{0,1\}$ or $\psi \in \{0,1\}$; so let $\phi, \psi \in (0,1)$. Furthermore, let X, Y, Z be independent standard uniform on $[0,1]$, and define

$$(U,V) = (X^{1/(1-\phi)} \vee Z^{1/\phi}, Y^{1/(1-\psi)} \vee Z^{1/\psi}),$$

which has cdf $C_{\phi,\psi}$. For $k \geq 0$, define $f_k(x) = x^{k+1}/(k+1)$ with derivative $f_k'(x) = x^k$. A straightforward calculation yields $\operatorname{Var}(f_k(U)) = \{(2k+3)(k+2)^2\}^{-1}$.

$$\begin{aligned}
\operatorname{Cov}(f_k(U), f_\ell(V)) &= \int_0^1\int_0^1 \big\{\mathbb{P}(U > u, V > v) - (1-u)(1-v)\big\} u^k v^\ell \,\mathrm{d}u\,\mathrm{d}v \\
&= \int_0^1\int_0^1 \big\{\min(u^{1-\phi}v, uv^{1-\psi}) - uv\big\} u^k v^\ell \,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{\phi\psi}{(k+2)(\ell+2)\{(\ell+2)\phi + (k+2)\psi - \phi\psi\}},
\end{aligned}$$

where the first equality is due to an extension of Hoeffding's covariance formula (Lo, 2017, Theorem 3.1), the second equality uses the fact that $\mathbb{P}(U > u, V > v) - (1-u)(1-v) = \mathbb{P}(U \leq u, V \leq v) - uv$ and the last equality follows from lengthy but elementary calculations. As a consequence

$$\operatorname{Corr}(f_k(U), f_\ell(V)) = \frac{\phi\psi\sqrt{(2k+3)(2\ell+3)}}{(\ell+2)\phi + (k+2)\psi - \phi\psi}. \tag{5.2.2}$$

Letting $k = \phi m$ and $\ell = \psi m$, this expression converges to $\sqrt{\phi\psi}$ for $m \to \infty$. As a consequence, $R(C_{\phi,\psi}) = R(U,V) \geq \sqrt{\phi\psi}$.

For the reverse inequality, note that U is conditionally independent of V given Z. As a consequence, by Lemma 5.2.3,

$$R(U,V) \le R(U,Z)R(V,Z).$$

Note that (U,Z) has joint cdf $D_\phi(u,v) = u^{1-\phi}(u^\phi \wedge v)$, and that, likewise, (V,Z) has joint cdf D_ψ. It hence remains to show that $r(\xi) := R(D_\xi) \le \sqrt{\xi}$ for all $\xi \in (0,1)$. In fact, we will show $r(\xi) = \sqrt{\xi}$, since we need '$\ge$' in the proof of '$\le$'.

We start by proving $r(\xi) \ge \sqrt{\xi}$. For that purpose, reconsider the function $f_k(x) = x^{k+1}/(k+1)$ with $k \ge 0$. A similar elementary calculation as for the proof of (5.2.2) shows that

$$\operatorname{Corr}(f_{k\xi}(S), f_k(T)) = \xi \frac{\sqrt{(2k+3)(2k\xi+3)}}{2k\xi+\xi+2}, \quad (S,T) \sim D_\xi, \tag{5.2.3}$$

which converges to $\sqrt{\xi}$ for $k \to \infty$ and hence implies $r(\xi) \ge \sqrt{\xi}$.

For the proof of $r(\xi) \le \sqrt{\xi}$, recall X, Y, Z from the beginning of the proof, and for $\xi_1, \xi_2 \in (0,1)$, let $\tilde W = Z$, $\tilde V = Y^{1/(1-\xi_2)} \vee Z^{1/\xi_2}$ and $\tilde U = X^{1/(1-\xi_1)} \vee \tilde V^{1/\xi_1}$, i.e.,

$$(\tilde U, \tilde V, \tilde W) = (X^{1/(1-\xi_1)} \vee Y^{1/\{\xi_1(1-\xi_2)\}} \vee Z^{1/(\xi_1\xi_2)}, Y^{1/(1-\xi_2)} \vee Z^{1/\xi_2}, Z).$$

A straightforward calculation shows that $(\tilde U, \tilde V)$ has cdf D_{ξ_1}, that $(\tilde V, \tilde W)$ has cdf D_{ξ_2} and that $(\tilde U, \tilde W)$ has cdf $D_{\xi_1\xi_2}$. Furthermore, $\tilde U$ and $\tilde W$ are conditionally independent given $\tilde V$, whence, by Lemma 5.2.3,

$$r(\xi_1\xi_2) = R(\tilde U, \tilde W) \le R(\tilde U, \tilde V)R(\tilde W, \tilde V) = r(\xi_1)r(\xi_2). \tag{5.2.4}$$

This implies monotonicity of $\xi \mapsto r(\xi)$. Furthermore, by setting $\xi_1 = \xi_2 = \xi$, we get equality in the previous display by Lemma 5.2.3, which yields

$$r(\xi) = r(\xi^2)^{1/2}. \tag{5.2.5}$$

Next, an application of Lemma 5.2.4 gives $r(1/2) \le 2^{-1/2}$, which, in view of the previous display, implies that $r(2^{-1/2}) = r(1/2)^{1/2} \le 2^{-1/4}$. Since we have already shown $r(\xi) \ge \sqrt{\xi}$ for all ξ, we obtain that $r(2^{-1/2}) = 2^{-1/4}$.

For $m \in \mathbb{N}$, we may apply (5.2.4) m-times to obtain that

$$2^{-m/4} \le r(2^{-m/2}) \le r(2^{-1/2})^m = 2^{-m/4},$$

whence $r(2^{-m/2}) = 2^{-m/4}$. Next, for any $n \in \mathbb{N}$, we may apply (5.2.5) $(n-1)$-times to obtain that

$$r(2^{-m/2^n}) = r(2^{-m/2^{n-1}})^{1/2} = r(2^{-m/2^{n-2}})^{1/2^2} = \cdots = r(2^{-m/2})^{1/2^{n-1}} = (2^{-m/4})^{1/2^{n-1}} = (2^{-m/2^n})^{1/2}.$$

We have hence shown that $r(x) = \sqrt{x}$ for all $x \in C := \{2^{-m/2^n} \in (0,1) : n, m \in \mathbb{N}\}$. For any fixed $\xi \in (0,1)$, we can choose sequences $(x_k)_k, (y_k)_k$ in C converging to ξ such that $x_k \le \xi \le y_k$ for all k. Hence, by monotonicity of r, $\sqrt{x_k} = r(x_k) \le r(\xi) \le r(y_k) = \sqrt{y_k}$, which implies $r(\xi) = \xi$ by taking the limit for $k \to \infty$. This finalizes the proof. □

5.3 An application in extreme value statistics

The Marshall Olkin copula from (5.2.1) is easily seen to be max-stable, that is, we have

$$C_{\phi,\psi}(u,v) = C_{\phi,\psi}(u^{1/m}, v^{1/m})^m \qquad \forall u,v \in [0,1], m \in \mathbb{N}.$$

As a consequence, it is an extreme-value copula (Gudendorf and Segers, 2010) and may hence appear as the weak limit copula of affinely standardized bivariate maxima. In fact, it happens to occur as a limit in the following simple situation: let $(X_n)_n$ denote an independent and identically distributed (iid) sequence of random variables satisfying the standard domain of attraction (DOA) condition (de Haan and Ferreira, 2006) that $(\max_{i=1}^r X_i - b_r)/a_r$ converges weakly to a non-degenerate limit distribution for $r \to \infty$, where $(b_r)_r \subset \mathbb{R}$ and $(a_r)_r \subset (0,\infty)$ are suitable scaling sequences. In that case, by the Fisher-Tippett-Gnedenko Theorem (Fisher and Tippett, 1928; Gnedenko, 1943), the limit distribution is necessarily the generalized extreme value distribution with cdf $G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}$ for x such that $1+\gamma x > 0$ and defined by continuity if $\gamma = 0$; here, $\gamma \in \mathbb{R}$ denotes the extreme value index. Now, under the DOA condition, we have, for any $\zeta \in [0,1]$ and writing $\zeta_r = \lfloor r\zeta \rfloor$,

$$\lim_{r\to\infty} \mathbb{P}\Big(\frac{\max_{i=1}^r X_i - b_r}{a_r} \le x, \frac{\max_{i=\zeta_r+1}^{\zeta_r+r} X_i - b_r}{a_r} \le y \Big) = G_{\zeta,\gamma}(x,y) := C_{1-\zeta,1-\zeta}\{G_\gamma(x), G_\gamma(y)\} \tag{5.3.1}$$

for all $x,y \in \mathbb{R}$; see, for instance, Lemma B.3 in Bücher and Zanger (2023). In fact, the result in (5.3.1) even holds if the iid sequence is replaced by a stationary time series, provided the long range dependence is suitably controlled.

The weak convergence in (5.3.1) is fundamental for the so-called sliding block maxima method in extreme value statistics. We refer to Bücher and Segers (2018a); Zou et al. (2021); Bücher and Zanger (2023) among others for details on the general approach. As it happens, even for time series data, the asymptotic behavior of respective estimators is typically driven by certain empirical means satisfying a central limit theorem with asymptotic variance formula given by

$$\sigma^2_{\mathrm{sb}}(h) := 2\int_0^1 \mathrm{Cov}(h(Y_{1,\zeta}), h(Y_{2,\zeta}))\,\mathrm{d}\zeta,$$

where $(Y_{1,\zeta}, Y_{2,\zeta}) := (Y_1, Y_2) \sim G_{\zeta,\gamma}$ with $G_{\zeta,\gamma}$ from (5.3.1) and where h is square-integrable with respect to G_γ. On the other hand, the traditional (disjoint) block maxima method satisfies respective limit theorems with asymptotic variance given by

$$\sigma^2_{\mathrm{db}}(h) := \mathrm{Var}(h(Y_1)),$$

where $Y_1 \sim G_\gamma$. The following theorem is essential for showing that the sliding block maxima method is statistically more efficient than the traditional disjoint block maxima method (again, we refer to the aforementioned references). The first part can be deduced from a technical result in Zou et al. (2021), see their Lemma A.10, but Theorem 5.2.1 above offers the possibility for an elegant and short proof.

Theorem 5.3.1. *For $h : \mathbb{R} \to \mathbb{R}$ with $\int h(x)^2 \, \mathrm{d}G_\gamma(x) < \infty$, we have $\sigma^2_{\mathrm{sb}}(h) \le \sigma^2_{\mathrm{db}}(h)$. Moreover, equality holds if and only if h is a function such that $\mathrm{Corr}(h(Y_{1,\zeta}), h(Y_{2,\zeta})) = R(G_{\gamma,\zeta}) = 1 - \zeta$ for Lebesgue almost every value of $\zeta \in [0, 1]$.*

Proof. We have

$$\begin{aligned}
\sigma^2_{\mathrm{sb}}(h) = 2\int_0^1 \mathrm{Cov}(h(Y_{1,\zeta}), h(Y_{2,\zeta}))\, \mathrm{d}\zeta &= 2\sigma^2_{\mathrm{db}}(h) \int_0^1 \mathrm{Corr}(h(Y_{1,\zeta}), h(Y_{2,\zeta}))\, \mathrm{d}\zeta \\
&\le 2\sigma^2_{\mathrm{db}}(h) \int_0^1 R(C_{1-\zeta,1-\zeta})\, \mathrm{d}\zeta \\
&= 2\sigma^2_{\mathrm{db}}(h) \int_0^1 1 - \zeta \, \mathrm{d}\zeta = \sigma^2_{\mathrm{db}}(h),
\end{aligned}$$

where we used Theorem 5.2.1 at the penultimate equality. The second statement is immediate. □

Acknowledgements

Financial support by the German Research Foundation (DFG grant number 465665892) is gratefully acknowledged. The authors are grateful to the participants of the Oberwolfach Workshop on “Mathematics, Statistics, and Geometry of Extreme Events in High Dimensions” for their valuable comments. The authors also are grateful to an unknown referee and a Co-Editor-in-Chief for their constructive comments that helped to improve the presentation substantially.

6 Outlook

In this section we will provide a brief discussion of possible future research directions based on the results obtained in the articles presented.

In Chapter 3 we have shown that U-statistics of block maxima are asymptotically normal under mild assumptions; in particular an affine-linear transformation property on the kernel function. The authors assume that this condition may be weakened to only hold asymptotically and might be embedded in a suitable framework leading to results with the same assertion but weaker conditions.
Furthermore, the limiting results are meaningful if the linear part of the U-statistic is (asymptotically) non-degenerate. Otherwise, the weak convergence is against null and thus one can expect to obtain faster rates depending on the level of degeneracy. Already for the classic i.i.d. situation the lines of argument change in a notable manner: the limit distribution is derived via a spectral decomposition of the kernel in regard to the underlying distribution. This method of proof requires restrictive assumptions on the associated eigenfunctions and the validity is often out of reach to verify. In Leucht (2012) the author presents new sufficient conditions comprising moment and kernel smoothness constraints which are easier to verify. It is an interesting open question in what way the results from Chapter 3 translate to degenerate settings comparable to Leucht (2012) and to develop Cramér von Mises type tests based on block maxima.
The classic partial sum process $1/\sqrt{n}\sum_{i=1}^{\lfloor ns \rfloor} X_i$ for $s \in (0,1)$ and suitable observations X_i whose limit distribution has been derived in Donsker (1951) can be transferred to (dependent) U-statistics, see Bücher and Kojadinovic (2016). The latter reference established consistent bootstrap procedures based on dependent multipliers to provide asymptotic confidence intervals and demonstrated applications to change point detection. The proof techniques in Chapter 3 may be synthesized with the techniques of the latter reference to allow generalizations to limit theorems for partial sum processes of U-statistics of block maxima.
Furthermore, the results might be extended to general U-processes as in Nolan and Pollard (1987) who pose abstract conditions on the function family which parameterizes the U-process.

The main result of Chapter 4 provides formal bootstrap consistency of general linear block maxima estimators comprising popular estimators in extreme value statistics as the GEV maximum likelihood estimator, pseudo/misspecified variants, probability weighted moment estimators and method of moments based estimators. A natural extension of the results would be to consider the bootstrap consistency of U-statistics of block maxima or more generally, asymptotically linear estimators based on block maxima. It is reasonable to assume that arguments from the proof of Propositions 2.5 and 3.2 in Bücher and Kojadinovic (2016) would be helpful.
A different scheme to Efron's bootstrap is given by subsampling. Subsampling per-

forms superior to bootstrapping twofold: First, fewer assumptions lead to formal consistency results as drawing without replacement leads to true subsamples, (Politis and Romano, 1994). Second, for rank based estimators, as the empirical copula, subsampling was found to outperform the bootstrap, (Kojadinovic and Stemikovskaya, 2019). In the disjoint block maxima setting this does indeed promise performance gains but in the sliding counterpart ties do appear subasymptotically regardless of the resampling method. Nevertheless, it is to be seen whether the advantages of the subsampling method prevail in the sliding regime. Yet, a general problem in extreme value statistics is the small number of effective samples which would result in a need for a sophisticated choice of the subsample size B.
Additionally, practical improvements of the resampling methods in order to obtain confidence intervals should be researched. In particular the recently introduced cheap bootstrap might allow for obtaining asymptotic niveau α confidence intervals while requiring a considerably smaller amount of resamples, c.f. Lam (2022).
Finally, the introduction of simple non-stationarity scheme as in Chapter 3 that is, piecewise stationary time series, could be considered in the context of bootstrapping block maxima estimators to improve flexibility of the methods. The techniques used in proving the consistency results invoke hope of easily transferring the arguments to this specific setting.

In Chapter 5 the maximal correlation coefficient of the two parameter Marshall Olkin distribution has been derived and the important inequality linking sliding and disjoint block maxima variances proven. The latter proof only works for one dimensional block maxima, but the result has been used for multivariate block maxima. It is of theoretical interest to derive the maximal correlation coefficient for distributions taking the form as in Chapter 4 equation (4.2.3).

The all block maxima (ABM) method investigated in Oorschot and Zhou (2020) is a permutation invariant method which differentiates it from all the other block maxima methods in this thesis. The latter paper provided theory for i.i.d. settings and simulation results for the dependent case which showed that the method performs better even in the presence of dependence. Moreover, the ABM method ignores dependence in the time series as it does not pertain ordering which is a crucial aspect in estimating functionals of the dependence structure as copulas, extremal index, tail dependence index or Kendall's τ. Yet, if the extreme value index γ is of interest these drawbacks do not apply. Thus, theory regarding formal bootstrap consistency results based on ABM estimators would be of interest.

Lastly, it is an important task for extreme value statistics to develop methodology in order to assess complex non-stationarities as appearing in climatology where the changing climate has to be taken into account, see Philip et al. (2020, Sec. 4.3), Alexander et al. (2006). First, the vanilla case of estimators based on observations M_t already following a non-stationary GEV$(\mu_t, \sigma_t, \gamma_t)$ distribution should be investigated. A prominent class of estimators would be given by conditional maximum likelihood estimators with covariates; see Zanger et al. (2024) for a proof sketch. Following, a further extension into the direction of max domain of attraction assumptions should then be incorporated to allow for a more general framework. Finally, asymptotic comparisons for different methods as disjoint, sliding and possibly all block maxima estimators in this framework would be worthwhile to conduct.

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111(D5).

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.*, 15(2):724–731.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probability*, 2:792–804.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: Theory and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.

Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam.

Berghaus, B. and Bücher, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.*, 46(5):2307–2335.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Borovkova, S., Burton, R., and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to u-statistics and dimension estimation. *Transactions of the American Mathematical Society*, 353(11):4261–4318.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.*, 30(1):69–81.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144. Update of, and a supplement to, the 1986 original.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, 80(391):580–619. With discussion and with a reply by the authors.

Bücher, A. and Jennessen, T. (2020). Method of moments estimators for the extremal index of a stationary time series. *Electronic Journal of Statistics*, 14(2):3103 – 3156.

Bücher, A. and Kojadinovic, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32(3):1145–1165.

Bücher, A. and Segers, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes*, 17(3):495–528.

Bücher, A. and Segers, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Stat.*, 12(1):1098–1125.

Bücher, A. and Segers, J. (2018b). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462.

Bücher, A. and Staud, T. (2024a). Bootstrapping estimators based on the block maxima method. *arXiv:2409.08661*.

Bücher, A. and Staud, T. (2024b). Limit theorems for non-degenerate u-statistics of block maxima for time series. *Electron. J. Statist.*, 18(2):2850–2885.

Bücher, A. and Staud, T. (2024c). On the maximal correlation coefficient for the bivariate marshall olkin distribution. *arXiv:2409.05529*.

Bücher, A. and Zanger, L. (2023). On the disjoint and sliding block maxima method for piecewise stationary time series. *Ann. Statist.*, 51(2):573–598.

Bücher, A. and Zhou, C. (2021). A horse race between the block maxima method and the peak–over–threshold approach. *Statistical Science*, 36(3):360–378.

Bühlmann, P. L. (1993). *The blockwise bootstrap in time series and empirical processes*. ProQuest LLC, Ann Arbor, MI. Thesis (Dr.Sc.Math)–Eidgenoessische Technische Hochschule Zürich (Switzerland).

Bücher, A. and Kojadinovic, I. (2016). Dependent multiplier bootstraps for non-degenerate u-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference*, 170:83–105.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, 14(3):1171–1179.

Cissokho, Y. and Kulik, R. (2021). Estimation of cluster functionals for regularly varying time series: sliding blocks estimators. *Electron. J. Stat.*, 15(1):2777–2831.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2):226–248.

Davis, R. A., Drees, H., Segers, J., and Warchoł, M. (2018). Inference on the tail process with application to financial time series modeling. *J. Econometrics*, 205(2):508–525.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B*, 52(3):393–442. With discussion and a reply by the authors.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York. An introduction.

de Haan, L. and Zhou, C. (2024). Bootstrapping extreme value estimators. *Journal of the American Statistical Association*, 119(545):382–393.

Dehling, H. (1983). A note on a theorem of Berkes and Philipp. *Z. Wahrsch. Verw. Gebiete*, 62(1):39–42.

Dehling, H., Giraudo, D., and Schmidt, S. K. (2023). U-statistics of local sample moments under weak dependence. *ALEA Lat. Am. J. Probab. Math. Stat.*, 20(2):1511–1535.

Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pages 3–113. Birkhäuser Boston, Boston, MA.

Dehling, H. and Wendler, M. (2010a). Central limit theorem and the bootstrap for u-statistics of strongly mixing data. *Journal of Multivariate Analysis*, 101(1):126–137.

Dehling, H. and Wendler, M. (2010b). Law of the iterated logarithm for U-statistics of weakly dependent observations. In *Dependence in probability, analysis and number theory*, pages 177–194. Kendrick Press, Heber City, UT.

del Barrio, E. and Matrán, C. (2000). The weighted bootstrap mean for heavy-tailed distributions. *J. Theoret. Probab.*, 13(2):547–569.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, 21(1):420–436.

Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli*, 25(3):1690–1723.

Donsker, M. D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6:12.

Drees, H. (2015). Bootstrapping empirical processes of cluster functionals with application to extremograms. *ArXiv e-prints: 1511.00420*.

Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75(2):149–172.

Drees, H. and Neblung, S. (2021). Asymptotics for sliding blocks estimators of rare events. *Bernoulli*, 27(2):1239–1269.

Drees, H., Resnick, S., and de Haan, L. (2000). How to make a Hill plot. *The Annals of Statistics*, 28(1):254 – 274.

Eastoe, E. F. and Tawn, J. A. (2008). Modelling Non-Stationary Extremes with Application to Surface Level Ozone. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 58(1):25–45.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Efron, B. and Hastie, T. (2016). *Computer age statistical inference*, volume 5 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, New York. Algorithms, evidence, and data science.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*. Springer.

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

Engeland, K. r., Hisdal, H., and Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: a case study. *Extremes*, 7(1):5–30.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.*, 43(1):276–298.

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(2):545–556.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Gebelein, H. (1941). Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Z. Angew. Math. Mech.*, 21:364–379.

Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.*, 37(5B):2990–3022.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula theory and its applications*, volume 198 of *Lect. Notes Stat. Proc.*, pages 127–145. Springer, Heidelberg.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press, New York.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Halmos, P. R. (1946). The Theory of Unbiased Estimation. *The Annals of Mathematical Statistics*, 17(1):34 – 43.

Harter, H. L. (1978). A bibliography of extreme-value theory. *Internat. Statist. Rev.*, 46(3):279–306.

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Hsing, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.*, 29(2):274–291.

Huser, R. and Davison, A. C. (2014). Space-time modelling of extreme events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 76(2):439–461.

Janson, S. (1988). Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.*, 16(1):305–312.

Jentsch, C. and Kulik, R. (2021). Bootstrapping Hill estimator and tail array sums for regularly varying time series. *Bernoulli*, 27(2):1409–1439.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8):1287–1304.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30:81–93.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C. (2007). Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations. *Journal of Climate*, 20(8):1419 – 1444.

Khashimov, S. A. (1994). The central limit theorem for generalized u-statistics for weakly dependent vectors. *Theory of Probability & Its Applications*, 38(3):456–469.

Klaassen, C. A. J. and Wellner, J. A. (1992). Kac empirical processes and the bootstrap. In *Probability in Banach spaces, 8 (Brunswick, ME, 1991)*, volume 30 of *Progr. Probab.*, pages 411–429. Birkhäuser Boston, Boston, MA.

Kojadinovic, I. and Stemikovskaya, K. (2019). Subsampling (weighted smooth) empirical copula processes. *Journal of Multivariate Analysis*, 173:704–723.

Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U-statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Kulik, R. and Soulier, P. (2020). *Heavy-tailed time series*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer Ser. Stat. New York, NY: Springer.

Lam, H. (2022). A cheap bootstrap method for fast inference. *arXiv:2202.00090*.

Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, 44(1/2):289–292.

Landwehr, J. M., Matalas, N., and Wallis, J. (1979). Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. *Water resources research*, 15(5):1055–1064.

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 28:289–303.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete*, 65(2):291–306.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin.

Lee, A. J. (2019). *U-statistics: Theory and Practice*. Routledge.

Leucht, A. (2012). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli*, 18(2):552 – 585.

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics*, 46(1):280 – 307.

Lin, G. D., Lai, C.-D., and Govindaraju, K. (2016). Correlation structure of the Marshall-Olkin bivariate exponential distribution. *Stat. Methodol.*, 29:1–9.

Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81(1):27–40.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 225–248. Wiley, New York.

Lo, A. (2017). Functional generalizations of Hoeffding's covariance lemma and a formula for Kendall's tau. *Statist. Probab. Lett.*, 122:218–226.

Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance*, 24(7):1097–1130.

Loynes, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.*, 36:993–999.

Madsen, H., Pearson, C. P., and Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. regional modeling. *Water Resources Research*, 33(4):759–769.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.*, 62:30–44.

Mikhailov, V. G. and Mezhennaya, N. M. (2020). Normal approximation for U- and V-statistics of a stationary absolutely regular sequence. *Sib. Èlektron. Mat. Izv.*, 17:672–682.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, 319(5863):573–574.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Nolan, D. and Pollard, D. (1987). U-processes: Rates of convergence. *The Annals of Statistics*, 15(2):780–799.

O'Brien, G. L. (1987). Extreme values for stationary and markov sequences. *The Annals of Probability*, 15(1):281 – 291.

Oorschot, J., Segers, J., and Zhou, C. (2023). Tail inference using extreme U-statistics. *Electron. J. Stat.*, 17(1):1113–1159.

Oorschot, J. and Zhou, C. (2020). All block maxima method for estimating the extreme value index. *ArXiv preprints: 2010.15950*.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.*, 105(489):263–277.

Peng, L. and Qi, Y. (2008). Bootstrap approximation of tail dependence function. *J. Multivariate Anal.*, 99(8):1807–1824.

Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R., et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6(2):177–203.

Pickands, III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131.

Pickands, III, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, volume 49 (2), pages 859–878, 894–902. With a discussion.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.*, 22(4):2031–2050.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.*, 23(1):53–70.

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

Rényi, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.*, 10:441–451 (unbound insert).

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.

Sarmanov, O. V. (1958). The maximum correlation coefficient (symmetric case). *Dokl. Akad. Nauk SSSR*, 120:715–718.

Schaller, R. (1997). Moore's law: past, present and future. *IEEE Spectrum*, 34(6):52–59.

Segers, J. (2001). *Extremes of a Random Sample: Limit Theorems and Applications*. Phd thesis, Katholieke Universiteit Leuven.

Sen, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 25:71–82.

Shen, L., Mickley, L. J., and Gilleland, E. (2016). Impact of increasing heat waves on u.s. ozone episodes in the 2050s: Results from a multimodel analysis using extreme value theory. *Geophysical Research Letters*, 43(8):4017–4025.

Smith, R. L. (1984). Threshold methods for sample extremes. In *Statistical extremes and applications (Vimeiro, 1983)*, volume 131 of *NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.*, pages 621–638. Reidel, Dordrecht.

Staud, T. (2024). Bootstraping Block Maxima Estimators. https://github.com/torbenstaud/Bootstrapping-block-maxima.

Thomas, M., Lemaitre, M., Wilson, M. L., Viboud, C., Yordanov, Y., Wackernagel, H., and Carrat, F. (2016). Applications of extreme value theory in public health. *PLOS ONE*, 11(7):1–7.

Tikhomirova, M. I. and Chistyakov, V. P. (2015). On the asymptotic normality of some sums of dependent random variables. *Diskret. Mat.*, 27(4):141–149.

van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., and Burgers, G. (2005). Estimating return periods of extreme events from ecmwf seasonal forecast ensembles. *International Journal of Climatology*, 25(10):1345–1354.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes - Springer Series in Statistics*. Springer, New York.

Yoshihara, K.-i. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(3):237–252.

Yu, Y. (2008). On the maximal correlation coefficient. *Statist. Probab. Lett.*, 78(9):1072–1075.

Zanger, L., Bücher, A., Kreienkamp, F., Lorenz, P., and Tradowsky, J. S. (2024). Regional pooling in extreme event attribution studies: an approach based on multiple statistical testing. *Extremes*, 27(1):1–32.

Zou, N., Volgushev, S., and Bücher, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.*, 49(1):295–320.

Author contribution statement

In the following, the individual contributions of the authors of the included articles in this thesis are presented.

2.1 Bücher, A. and Staud, T. (2024). Limit theorems for non-degenerate U-statistics of block maxima for time series. *Electronic Journal of Statistics*, 18(2):2850-2885.

    The initial idea in this article to investigate U-statistics of block maxima was developed by the first author, who also wrote the abstract and improved upon the introduction. The idea to generalize dependence to α-mixing time series was developed by the second author while the kernel transformation condition was proposed by the first author. The second author established most theoretical results and conducted the simulation studies. Both authors contributed to the final version of the article. The manuscript was mostly drafted by the second author and throughout the project the first author supervised the second author and made several improvements and corrections leading to the final version of the article. Both authors contributed equally to the revision upon first submission to the above-mentioned journal.

2.2 Bücher, A. and Staud, T. (2024). Bootstrapping block maxima estimators for time series. *Submitted for publication*.

    The idea in this article to investigate bootstrapping block maxima estimators was developed by the first author, who also wrote the abstract and improved upon the introduction. The idea of the circmax bootstrapping procedure was developed by the first author in earlier works. Investigating the bootstrap of the pseudo maximum likelihood estimation in the Fréchet model was proposed by the first author. The second author established most theoretical results and conducted both the simulation and case studies. Both authors contributed to the final version of the article. The manuscript was drafted mostly by the second author and improved upon by the first author. The GitHub repository was developed and maintained by the second author. The first author supervised the second author and made several improvements leading to the final version of the article.

2.3 Bücher, A. and Staud, T. (2024). On the maximal correlation coefficient for the bivariate Marshall Olkin distribution.
    Accepted for publication in *Statistics and Probability Letters*.

    The open problem to derive the maximal correlation coefficient for the Marshall

Olkin distribution was identified by the second author. The second section was drafted by the first author while the second author provided complementing arguments and drafted the third section, concerning application to extremes. Both authors contributed to the final version of the article. The revision upon first submission to the above-mentioned journal was mostly done by the second author.

Versicherung an Eides statt

Ich versichere an Eides statt, dass ich die eingereichte Dissertation selbstständig und ohne unzulässige fremde Hilfe verfasst, andere als die in ihr angegebene Literatur nicht benutzt und dass ich alle ganz oder annähernd übernommenen Textstellen sowie verwendete Grafiken, Tabellen und Auswertungsprogramme kenntlich gemacht habe.

Außerdem versichere ich, dass die vorgelegte elektronische mit der schriftlichen Version der Dissertation übereinstimmt und die Abhandlung in dieser oder ähnlicher Form noch nicht anderweitig als Promotionsleistung vorgelegt und bewertet wurde.

Bochum, den

- -

Ort, Datum **M. Sc. Torben Staud**