

Inference for heavy tailed stationary time series based on sliding blocks

Axel Bücher

*Ruhr-Universität Bochum, Fakultät für Mathematik
Universitätsstr. 150, 44780 Bochum, Germany
e-mail: axel.buecher@rub.de*

and

Johan Segers

*Université catholique de Louvain, ISBA
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
e-mail: johan.segers@uclouvain.be*

Abstract: The block maxima method in extreme value theory consists of fitting an extreme value distribution to a sample of block maxima extracted from a time series. Traditionally, the maxima are taken over disjoint blocks of observations. Alternatively, the blocks can be chosen to slide through the observation period, yielding a larger number of overlapping blocks. Inference based on sliding blocks is found to be more efficient than inference based on disjoint blocks. The asymptotic variance of the maximum likelihood estimator of the Fréchet shape parameter is reduced by more than 18%. Interestingly, the amount of the efficiency gain is the same whatever the serial dependence of the underlying time series: as for disjoint blocks, the asymptotic distribution depends on the serial dependence only through the sequence of scaling constants. The findings are illustrated by simulation experiments and are applied to the estimation of high return levels of the daily log-returns of the Standard & Poor's 500 stock market index.

MSC 2010 subject classifications: Primary 62G32; secondary 62M10.

Keywords and phrases: Apéry's constant, block maxima, Fréchet distribution, maximum likelihood estimator, Marshall–Olkin distribution, Pickands dependence function, return level.

Received June 2017.

Contents

1	Introduction	1099
2	Inference based on sliding blocks	1100
2.1	Maximum likelihood estimator based on sliding block maxima . .	1100
2.2	Asymptotic normality	1101
2.3	Sliding block maxima extracted from an iid sequence	1104
3	Application to return level estimation	1105
3.1	Estimator	1105
3.2	Case study	1106
4	Simulation Study	1107
5	Proofs and auxiliary results	1110

A Covariance calculations for min-stable distributions	1119
Acknowledgments	1123
References	1124

1. Introduction

Two major paradigms in extreme value theory are the block maxima method and the peaks-over-threshold method. The former, more traditional one consists of fitting an extreme value distribution to a sample of block maxima extracted from a (perhaps latent) underlying sample. The latter method consists of fitting a generalized Pareto distribution to the excesses in a sample over a high threshold.

Although the peaks-over-threshold method has become the standard one, there has been a renewed interest recently in the block maxima method, and more specifically in its asymptotic properties. The set-up is that of a triangular array of block maxima extracted from a stationary time series. The block size, r , tends to infinity as the sample size, n , tends to infinity, in such a way that the number of (disjoint) blocks, approximately n/r , tends to infinity as well. The underlying sequence of random variables can be independent [6, 7, 8] or can exhibit serial dependence [3, 4].

Usually, maxima are taken over disjoint blocks of observations. For instance, for a sequence of daily observations, one may extract weakly, monthly, quarterly or yearly maxima; see, e.g., [14, 12] or [10] for applications in finance or hydrology, respectively. Alternatively, one can slide a block or window of a given size through the sample and consider the corresponding maxima. Obviously, the blocks will be overlapping and thus dependent, even if the underlying sequence of random variables is independent. Still, as soon the underlying sequence is stationary, then so are the sliding block maxima. Moreover, the sample of sliding block maxima carries more information than the sample of disjoint block maxima, which suggests the possibility of more accurate inference. Robert et al. [17], Northrop [15] and Berghaus and Bücher [1] applied this idea to the estimation of the extremal index, a summary measure for the strength of serial dependence between extremes. They found that estimators based on sliding blocks were indeed more efficient than their counterparts based on disjoint blocks.

Here, we investigate the potential benefits of using maxima over sliding blocks rather than over disjoint blocks for fitting extreme value distributions. More precisely, we seek the asymptotic distribution of the maximum (quasi-)likelihood estimator for the shape and scale parameters of a Fréchet distribution. The likelihood is computed as if the sliding block maxima are independent, although they are not, as blocks may overlap.

The solution is based on Theorem 2.5 in [4], which states high-level conditions for the consistency and asymptotic normality of the maximum likelihood estimator of the Fréchet parameter vector based on a general triangular array of dependent random variables. The biggest challenge is the computation of the estimator's asymptotic covariance matrix. In the course of the computations, we find new formulas for moments of pairs of jointly max-stable random variables in terms of their Pickands dependence function. These formulas are then applied

to the bivariate Marshall–Olkin distribution, which describes the joint asymptotic distribution of the two maxima over a pair of overlapping blocks. The Marshall–Olkin parameter is a function of the proportion of overlap between the two blocks.

We find that the maximum likelihood estimator based on sliding blocks is more efficient than the maximum likelihood estimator based on disjoint blocks. For the estimator of the Fréchet shape parameter, the reduction in asymptotic variance is more than 18%. Remarkably, this number does not depend on the serial dependence of the underlying stationary time series, in accordance to the findings for disjoint blocks in [4]. Moreover, the efficiency gain carries over to the estimation of high return levels. The asymptotic results are confirmed in numerical experiments. We illustrate the method by estimating high quantiles of quarterly maxima of daily log-returns of the S&P500 index. The Monte Carlo simulations reveal another benefit of using sliding blocks: it makes the estimator more stable as a function of the block size.

The maximum likelihood estimator is defined and its asymptotic distribution is stated in Section 2. Estimation of high return levels is considered in Section 3, both theoretically and through a case study, followed by the results of a Monte Carlo simulation experiment in Section 4. The proofs are given in Section 5 while the covariance calculations are deferred to Appendix A.

2. Inference based on sliding blocks

2.1. Maximum likelihood estimator based on sliding block maxima

Let $(X_t)_{t \in \mathbb{Z}}$ be a strictly stationary time series: for any $k \in \mathbb{N}$ and for any $h, t_1, \dots, t_k \in \mathbb{Z}$, the distribution of $(X_{t_1+h}, \dots, X_{t_k+h})$ is the same as the distribution of $(X_{t_1}, \dots, X_{t_k})$. Further, let $P_{\alpha, \sigma}$ denote the Fréchet distribution with parameter $\theta = (\alpha, \sigma)' \in (0, \infty)^2$, given by its distribution function $P_{\alpha, \sigma}([0, x]) = \exp\{-(x/\sigma)^{-\alpha}\}$ for $x > 0$. We assume the following maximum domain-of-attraction condition. The arrow $\rightsquigarrow$ denotes convergence in distribution.

Condition 2.1 (Max-domain of attraction). For some $\alpha_0 \in (0, \infty)$, there exists a sequence $(\sigma_r)_{r \in \mathbb{N}}$, regularly varying at infinity with index $1/\alpha_0$, such that

$$\max(X_1, \dots, X_r)/\sigma_r \rightsquigarrow P_{\alpha_0, 1}, \quad r \rightarrow \infty.$$

Regular variation of the sequence $(\sigma_r)_r$ with index $1/\alpha_0$ means that $\sigma_{\lfloor rs \rfloor}/\sigma_r \rightarrow s^{1/\alpha_0}$ as $r \rightarrow \infty$ for every $s > 0$, where $\lfloor \cdot \rfloor$ denotes the integer part. A sufficient condition is that the common univariate distribution of the variables X_t is in the max-domain of attraction of the Fréchet distribution (see Section 2.3) and that the extremal index θ of the time series $(X_t)_t$ exists and is positive [11].

Suppose we observe a finite stretch of the time series, $X_1, \dots, X_n$. For some integer $r \in \{1, \dots, n\}$, let

$$M_{r,t} = M_{t:(t+r-1)} = \max\{X_t, \dots, X_{t+r-1}\}, \quad t = 1, \dots, n - r + 1,$$

denote the maximum over the r successive observations starting at time point t . The sequence $M_{r,1}, \dots, M_{r,k}$ with $k = n - r + 1$ is referred to as the sequence of *sliding block maxima*. In contrast, the classical block maxima method in extreme value statistics is based on the sequence of *disjoint block maxima* $M_{r,1}, M_{r,r+1}, \dots, M_{r,(m-1)r+1}$, where $m = \lfloor n/r \rfloor$. The common big blocks/small blocks heuristics suggests that the latter sequence may be regarded as asymptotically independent and Fréchet distributed. Any sensible estimator within the statistical model $\mathcal{P} = \{P_\theta^{\otimes m} : \theta = (\alpha, \sigma)' \in (0, \infty)^2\}$ is hence a sensible estimator when applied to the sample of disjoint block maxima as well.

Unfortunately, this idea cannot be directly transferred to the sample of sliding block maxima, as that sequence is certainly not asymptotically independent, not even for an underlying iid time series. Still, the sample of sliding block maxima is stationary and the asymptotic distribution of a single such block maximum is Fréchet. We can therefore estimate the Fréchet parameters by moment matching, for instance. Being based on empirical moments only, the maximum likelihood estimator for independent sampling from the Fréchet distribution (model $\mathcal{P}$) is a case in point.

Existence and uniqueness of the maximum likelihood estimator in model $\mathcal{P}$ is studied in Section 2.1 of [4]. The estimator is defined as

$$(\hat{\alpha}_n, \hat{\sigma}_n) = \arg \max_{\theta \in (0, \infty)^2} \sum_{t=1}^{n-r+1} \ell_\theta(X_{n,t}), \quad (2.1)$$

where $\ell_\theta(x) = \log \mathrm{d} \exp\{-(x/\sigma)^{-\alpha}\} / \mathrm{d} x$ is the contribution of an observation at $x > 0$ to the Fréchet log-likelihood of the parameter vector θ and where

$$X_{n,t} = M_{r,t} \vee c \quad (2.2)$$

with an arbitrary truncation constant $c > 0$. The reason for the left-truncation is that otherwise some of the block maxima could be zero or negative. Asymptotically, the truncation constant does not matter thanks to Condition 2.2 below. In practice, it should be chosen as small as possible, e.g., equal to the square root of the machine precision. By Lemma 2.1 in [4], the maximizer exists and is unique as soon as the $n - r + 1$ values $X_{n,1}, \dots, X_{n,n-r+1}$ are not all equal, which our conditions will guarantee to occur with probability tending to one. Note that the likelihood is constructed as if the sliding block maxima were independent, although they are not, not even if the underlying sequence is independent, since some blocks overlap. Therefore, the estimator may be more accurately referred to as a maximum quasi-likelihood estimator.

2.2. Asymptotic normality

For our main result, we need a couple of additional conditions. First of all, let $r = r_n$ be an integer sequence tending to infinity such that $r_n = o(n)$. The next condition is a slight adaptation of Condition 3.2 in [4].

Condition 2.2 (All block maxima of size $\lfloor r_n/2 \rfloor$ diverge). For every $c > 0$, the probability of the event that all disjoint block maxima of size $\tilde{r}_n = \lfloor r_n/2 \rfloor$ are larger than c converges to 1.

The condition in fact implies that the probability of the event that all sliding block maxima of size r_n are larger than c converges to 1 as well. It therefore guarantees that the left-truncation above does not matter asymptotically. In Section 2.3, the condition will be shown to hold for iid time series, provided the block sizes are not too small.

The following three conditions are Conditions 3.3, 3.4 and 3.5 in [4]. The alpha-mixing coefficients of the sequence $(X_t)_t$ are defined as

$$\alpha(k) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \sigma(X_j, j \leq i), B \in \sigma(X_{j+k}, j \geq i), i \in \mathbb{Z}\}, \quad k = 1, 2, \dots$$

Condition 2.3 (α -Mixing with rate). We have $\lim_{\ell \rightarrow \infty} \alpha(\ell) = 0$. Moreover, there exists $\omega > 0$ such that

$$\lim_{n \rightarrow \infty} (n/r_n)^{1+\omega} \alpha(r_n) = 0.$$

Condition 2.4 (Moments). There exists $\nu > 2/\omega$ with ω from Condition 2.3 such that

$$\limsup_{r \rightarrow \infty} \mathbb{E} [g_{\nu, \alpha_0}((M_{r,1} \vee 1)/\sigma_r)] < \infty, \quad (2.3)$$

where $g_{\nu, \alpha_0}(x) = \{x^{-\alpha_0} \mathbf{1}(x \leq e) + \log(x) \mathbf{1}(x > e)\}^{2+\nu}$.

An elementary argument shows that if Condition 2.4 holds, then (2.3) continues to hold with $M_{r,1} \vee 1$ replaced by $M_{r,1} \vee c$, for arbitrary $c > 0$.

Let $P = P_{\alpha_0, 1}$ and write $Pf = \int_0^\infty f(x) dP(x)$ for a real-valued function on $(0, \infty)$. Further, let

$$f_1(x) = x^{-\alpha_0} \log(x), \quad f_2(x) = x^{-\alpha_0}, \quad f_3(x) = \log(x) \quad (2.4)$$

and note that, by Lemma B.1 in [4],

$$\begin{aligned} Pf_1 &= -\alpha_0^{-1} \Gamma'(2) = \alpha_0^{-1}(\gamma - 1), \\ Pf_2 &= \Gamma(2) = 1, \\ Pf_3 &= -\alpha_0^{-1} \Gamma'(1) = \alpha_0^{-1} \gamma, \end{aligned}$$

where Γ and Γ' denote the gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ and its derivative, respectively, while $\gamma = 0.5772 \dots$ denotes the Euler–Mascheroni constant.

Condition 2.5 (Bias). There exists $c_0 > 0$ such that for $j = 1, 2, 3$,

$$\lim_{n \rightarrow \infty} \sqrt{\frac{n}{r_n}} \left(\mathbb{E} [f_j((M_{r,1} \vee c_0)/\sigma_r)] - Pf_j \right) = B(f_j)$$

with f_j as defined in (2.4).

As an interesting consequence of Theorem 2.6 below, the limit $B(f_j)$ can be seen not to depend on the constant c_0 under the conditions of the theorem.

Finally, we define the empirical process based on the left-truncated sliding block maxima in (2.2) as

$$\mathbb{G}_n f = \sqrt{\frac{n}{r_n}} \left(k_n^{-1} \sum_{t=1}^{k_n} f(X_{n,t}/\sigma_{r_n}) - Pf \right).$$

Recall the maximum (quasi-)likelihood estimator in (2.1). The following theorem is the main result of this paper.

Theorem 2.6. *Suppose Conditions 2.1–2.5 are met. Then, for any $c > 0$ and with probability tending to one, there exists a unique maximizer $(\hat{\alpha}_n, \hat{\sigma}_n)$ of the Fréchet log-likelihood based on the left-truncated sliding block maxima $X_{n,1}, \dots, X_{n,k_n}$, see (2.1), and we have, as $n \rightarrow \infty$,*

$$\begin{aligned} \sqrt{m_n} \begin{pmatrix} \hat{\alpha}_n - \alpha_0 \\ \hat{\sigma}_n/\sigma_{r_n} - 1 \end{pmatrix} &= M(\alpha_0) \begin{pmatrix} \mathbb{G}_n x^{-\alpha_0} \log(x) \\ \mathbb{G}_n x^{-\alpha_0} \\ \mathbb{G}_n \log(x) \end{pmatrix} + o_{\mathbb{P}}(1) \\ &\rightsquigarrow \mathcal{N}_2(M(\alpha_0)B, \Sigma(\alpha_0)). \end{aligned}$$

Here, $m_n = n/r_n$, $B = (B(f_1), B(f_2), B(f_3))'$ as in Condition 2.5 and

$$\begin{aligned} M(\alpha_0) &= \frac{6}{\pi^2} \begin{pmatrix} \alpha_0^2 & \alpha_0(1-\gamma) & -\alpha_0^2 \\ \gamma-1 & -(\Gamma''(2)+1)/\alpha_0 & 1-\gamma \end{pmatrix}, \\ \Sigma(\alpha_0) &= M(\alpha_0)\Sigma_Y M(\alpha_0)^T \approx \begin{pmatrix} 0.4946 \alpha_0^2 & -0.3236 \\ -0.3236 & 0.9578 \alpha_0^{-2} \end{pmatrix}, \end{aligned}$$

with $\Sigma_Y = (\sigma_{ij})_{i,j=1}^3$ as in Corollary 5.4 and equation (5.8).

It is interesting to note that the limiting covariance matrix is substantially smaller than for the estimator based on disjoint blocks, see Theorem 3.6 in [4]:

$$\begin{aligned} I_{(\alpha_0,1)}^{-1} &= \frac{6}{\pi^2} \begin{pmatrix} \alpha_0^2 & (\gamma-1) \\ (\gamma-1) & \alpha_0^{-2}\{(1-\gamma)^2 + \pi^2/6\} \end{pmatrix} \\ &\approx \begin{pmatrix} 0.6080 \alpha_0^2 & -0.2570 \\ -0.2570 & 1.1087 \alpha_0^{-2} \end{pmatrix}. \end{aligned}$$

The improvement is independent of the value of α_0 and of the serial dependence of the time series (e.g., of any of the characteristics like the extremal index or the cluster distribution). In particular, the quotient of the asymptotic variances for the shape and scale parameters are 0.8135 and 0.8639, respectively.

More generally, the delta method implies that the asymptotic distribution of a tail quantity that can be written as a smooth function of $(\hat{\alpha}_n, \hat{\sigma}_n)'$ will be normal with asymptotic variance equal to $\beta' I_{(\alpha_0,1)}^{-1} \beta$ (disjoint blocks) or $\beta' \Sigma(\alpha_0) \beta$ (sliding blocks), where β is a 2×1 vector of partial derivatives depending on

the estimator. The ratio of asymptotic variances is thus equal to

$$\frac{\beta' \Sigma(\alpha_0) \beta}{\beta' I_{(\alpha_0, 1)}^{-1} \beta}. \quad (2.5)$$

By the method of Lagrange multipliers, this ratio can be found to attain its minimum and maximum at the smallest and largest eigenvalues of the matrix $\Sigma(\alpha_0)(I_{(\alpha_0, 1)}^{-1})^{-1} = \Sigma(\alpha_0)I_{(\alpha_0, 1)}$. Independently of α_0 , these eigenvalues are equal to 0.9413 and 0.6448, respectively, and provide precise lower and upper bounds to the ratio in (2.5).

Should the underlying time series be positive, one might be tempted to omit the left-truncation introduced above (which amounts to setting $c = 0$). Working out the asymptotic theory is possible, but at the cost of more complicated conditions. Even if $X_t > 0$ almost surely, the lower tail of the random variable X_t in a neighbourhood of zero must not be too heavy for the moment and bias conditions to be true without truncation. In practice, left-truncation at a suitable small constant appears to be non-restrictive anyway whence we do not pursue this issue any further.

2.3. Sliding block maxima extracted from an iid sequence

If the sequence $(X_t)_t$ is iid, then all conditions can be expressed in terms of the univariate, marginal distribution function $F(x) = \mathbb{P}(X_t \leq x)$. Condition 2.1 is equivalent to regular variation of the function $-\log F$ at infinity with index $-\alpha_0$, that is,

$$\lim_{u \rightarrow \infty} \frac{-\log F(ux)}{-\log F(u)} = x^{-\alpha_0}, \quad x \in (0, \infty). \quad (2.6)$$

The scaling sequence can be chosen as $\sigma_r = \inf\{u \geq 1 : F(u) \geq e^{-1/r}\}$, for $r = 1, 2, \dots$

To control the bias (Condition 2.5), we need to reinforce regular variation in (2.6) to second-order regular variation of the function $-\log F$ together with a growth restriction on the block size sequence $(r_n)_n$. The following condition is identical to Condition 4.1 in Bücher and Segers [4, Section 4]; see also Remark 4.3 therein for additional context. For $\tau \in \mathbb{R}$, define $h_\tau : (0, \infty) \rightarrow \mathbb{R}$ by

$$h_\tau(x) = \int_1^x y^{\tau-1} dy = \begin{cases} \frac{x^\tau - 1}{\tau}, & \text{if } \tau \neq 0, \\ \log(x), & \text{if } \tau = 0. \end{cases}$$

Condition 2.7 (Second-Order Condition). There exists $\alpha_0 \in (0, \infty)$, $\rho \in (-\infty, 0]$, and a real function A on $(0, \infty)$ of constant, non-zero sign such that $\lim_{u \rightarrow \infty} A(u) = 0$ and such that, for all $x \in (0, \infty)$,

$$\lim_{u \rightarrow \infty} \frac{1}{A(u)} \left(\frac{-\log F(ux)}{-\log F(u)} - x^{-\alpha_0} \right) = x^{-\alpha_0} h_\rho(x). \quad (2.7)$$

Let $\psi = \Gamma'/\Gamma$ denote the digamma function. For $(\alpha_0, \rho) \in (0, \infty) \times (-\infty, 0]$, define the bias function

$$B(\alpha_0, \rho) = -\frac{6}{\pi^2} \begin{pmatrix} b_1(|\rho|/\alpha_0) \\ b_2(|\rho|/\alpha_0)/\alpha_0^2 \end{pmatrix}, \quad (2.8)$$

where

$$b_1(x) = \begin{cases} (1+x)\Gamma(x)\{\gamma + \psi(1+x)\}, & \text{if } x > 0, \\ \frac{\pi^2}{6}, & \text{if } x = 0, \end{cases}$$

and

$$b_2(x) = \begin{cases} -\frac{\pi^2}{6x} + (1+x)\Gamma(x)\{\Gamma''(2) + \gamma + (\gamma-1)\psi(1+x)\}, & \text{if } x > 0, \\ 0, & \text{if } x = 0. \end{cases}$$

The graphs of these functions are depicted in Figure 1 in [4].

Theorem 2.8. *Let $X_1, X_2, \dots$ be independent random variables with common distribution function F satisfying Condition 2.7. Let the block sizes r_n be such that $r_n \rightarrow \infty$ and $m_n = \lfloor n/r_n \rfloor \rightarrow \infty$ as $n \rightarrow \infty$ and assume that*

$$\lim_{n \rightarrow \infty} \sqrt{m_n} A(a_{r_n}) = \lambda \in \mathbb{R}. \quad (2.9)$$

Then, for any $c > 0$ and with probability tending to one, there exists a unique maximizer $(\hat{\alpha}_n, \hat{\sigma}_n)$ of the Fréchet log-likelihood based on the left-truncated sliding block maxima $X_{n,1}, \dots, X_{n,k_n}$, and we have

$$\sqrt{m_n} \begin{pmatrix} \hat{\alpha}_n - \alpha_0 \\ \hat{\sigma}_n/\sigma_{r_n} - 1 \end{pmatrix} \rightsquigarrow \mathcal{N}_2(\lambda B(\alpha_0, \rho), \Sigma(\alpha_0)), \quad n \rightarrow \infty,$$

with $\Sigma(\alpha_0)$ as in Theorem 2.6 and $B(\alpha_0, \rho)$ as in (2.8).

Compared to the estimator based on disjoint block maxima [4, Theorem 4.2], the asymptotic bias is the same, but the asymptotic variance is smaller, as explained after Theorem 2.6.

3. Application to return level estimation

3.1. Estimator

Let $F_r(x) = \mathbb{P}(M_{1:r} \leq x)$. For $T \geq 1$, the T -return level of the sequence of disjoint block maxima is defined as the $1 - 1/T$ quantile of F_r , that is,

$$\text{RL}(T, r) = F_r^{\leftarrow}(1 - 1/T) = \inf\{x \in \mathbb{R} : F_r(x) \geq 1 - 1/T\}.$$

Since disjoint blocks are asymptotically independent, it will take on average T disjoint blocks of size r until the first such block whose maximum exceeds $\text{RL}(T, r)$.

T	50	100	200	500	1000	5000	10000
Sliding Blocks	11.01	14.40	18.26	24.07	29.02	42.35	48.87
Disjoint Blocks	12.37	16.34	20.88	27.77	33.66	49.59	57.41
Ratio	0.89	0.88	0.87	0.87	0.86	0.85	0.85

TABLE 1

Asymptotic variances for the sliding and disjoint blocks versions of $\widehat{\text{RL}}_n(T, r_n)$, alongside with their ratio, for $\alpha_0 = 1$.

By Condition 2.1, for large r , we may approximate F_r by G_{α_0, σ_r} , the cdf of the Fréchet distribution with shape parameter α_0 and scale parameter σ_r . The quantile function of the Fréchet family is given by $G_{\alpha, \sigma}^{-1}(p) = \sigma \{-\log(p)\}^{-1/\alpha_0}$. A reasonable estimator of $\text{RL}(T, r_n)$ is therefore

$$\widehat{\text{RL}}_n(T, r_n) = \hat{\sigma}_n b_T^{-1/\hat{\alpha}_n}, \quad b_T = -\log(1 - 1/T).$$

Also, let $\widetilde{\text{RL}}(T, r_n) = \sigma_{r_n} b_T^{-1/\alpha_0}$.

Corollary 3.1. *Additionally to the conditions in Theorem 2.6 assume the bias condition*

$$\Lambda_n(T) = \sqrt{m_n} \left(\frac{\widetilde{\text{RL}}(T, r_n)}{\widehat{\text{RL}}_n(T, r_n)} - 1 \right) \rightarrow \Lambda(T) \in \mathbb{R}, \quad n \rightarrow \infty.$$

Then, as $n \rightarrow \infty$,

$$\begin{aligned} \sqrt{m_n} \left(\frac{\widehat{\text{RL}}_n(T, r_n)}{\widetilde{\text{RL}}(T, r_n)} - 1 \right) &= \beta(T, \alpha_0)' \cdot \sqrt{m_n} \left(\frac{\hat{\alpha}_n - \alpha_0}{\hat{\sigma}_n / \sigma_{r_n}} - 1 \right) + \Lambda_n(T) + o_{\mathbb{P}}(1) \\ &\rightsquigarrow \mathcal{N} \left(\beta(T, \alpha_0)' M(\alpha_0) B + \Lambda(T), \beta(T, \alpha_0)' \Sigma(\alpha_0) \beta(T, \alpha_0) \right), \end{aligned}$$

where $\beta(T, \alpha_0) = (\alpha_0^{-2} \log(b_T), 1)'$.

The same result holds true for the disjoint blocks estimator, but with $\Sigma(\alpha_0)$ replaced by $I_{(\alpha_0, 1)}^{-1}$, the inverse of the Fisher information matrix for the Fréchet family. In Table 1, the asymptotic variances are given for various values of T and with $\alpha_0 = 1$.

3.2. Case study

We consider daily log-returns on the S&P500 index (data downloaded from Yahoo Finance) in the fifty-year period from 1967 to 2016, yielding $n = 12\,584$ observations in total. For such a long period, the log-returns can hardly be modelled by a stationary time series, and therefore we consider ten-year periods instead, yielding approximately $n = 2\,500$ daily observations per period. We consider a block size equal to $r = 62$, which corresponds to approximately one quarter, yielding $m = 40$ disjoint quarters in a ten-year period. In light of

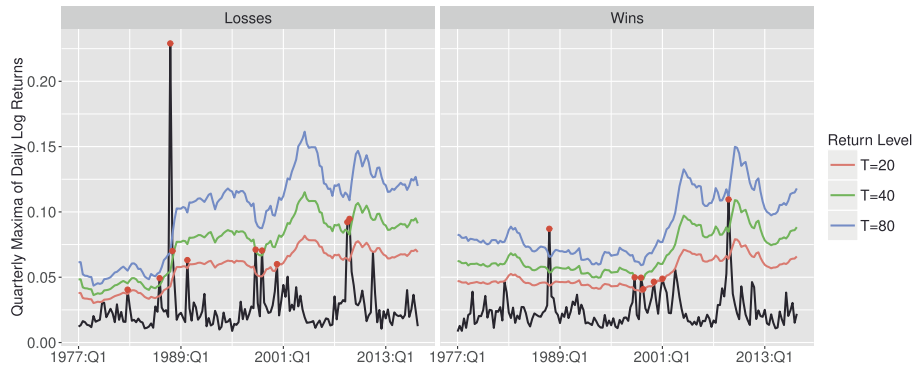


FIG 1. Quarterly block maxima of the negative (left) and positive (right) log-returns of the S&P500 index (black lines) together with estimates of the $1 - 1/T$ quantile of the block maximum distribution (coloured lines). The estimates for a given quarter are based on sliding block maxima in the ten-year period immediately preceding that quarter.

the results in McNeil [14] and Longin [12], such a block size guarantees that the block maxima are approximately Fréchet distributed, while at the same time the effective sample size of $m \approx 40$ is large enough to guarantee a reasonably small variance of the estimators. We estimate the $1 - 1/T$ quantile of the distribution of the quarterly maximum using the sliding block maxima and we check whether this level is exceeded by the quarterly maximum immediately following the ten-year training sample. We then let the ten-year window roll through the whole fifty-year period in steps of one quarter, giving $40 \times 4 = 160$ estimated return levels in total, of which, on average, $160/T$ should be exceeded. We consider $T = 20, 40, 80$, for which one would thus expect 8, 4, 2 exceedances, respectively. Doing the calculations separately for the positive and negative log-returns, we find 7, 3, 1 exceedances for the wins and 10, 7, 1 exceedances for the losses.

The results are shown in Figure 1. The black lines depict the quarters following each ten-year training period, from 1977:Q1 up to 2016:Q4. The estimated return levels are represented by colored lines; for instance, the value of the red line at 1977:Q1 corresponds to the estimated $1 - 1/20$ quantile of the cdf of the quarterly block maxima, estimated from the data in the period 1967–1976. The dots correspond to exceedances of the estimated return levels.

In Figure 2, we show estimated values and 95% confidence intervals (based on the normal approximation and ignoring the bias) for the Fréchet shape and scale parameters. The black lines are the same as in Figure 1, except for an affine transformation. The impact of the occurrence of large block maxima is clearly visible (decrease in α , increase in σ).

4. Simulation Study

We compare the performance of the disjoint and sliding blocks variations of the maximum likelihood estimator of the Fréchet shape parameter α . We also

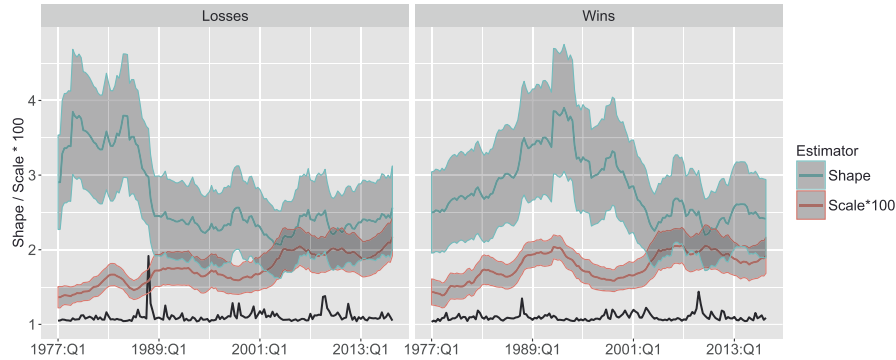


FIG 2. Estimated Fréchet shape (blue) and scale (red) parameters with pointwise 95% confidence intervals of the distribution of the quarterly block maxima of the negative (left) and positive (right) log-returns of the S&P500 index. The estimates for a given quarter are based on the sliding block maxima in the ten-year period immediately preceding that quarter and are depicted together with the actually observed block maximum (affinely transformed, black line).

compare the results with the Hill estimator [9], which is the maximum likelihood estimator for the one-parameter Pareto distribution given by $\Pr(Y > y) = (y \vee 1)^{-\alpha}$ fitted to the relative excesses over a high threshold. To put the estimators on equal footing, we set the threshold equal to the $m + 1$ largest order statistic, of which there are m excesses, where $m = \lfloor n/r \rfloor$ is the number of disjoint blocks of size r and which acts as an effective sample size.

We consider two different dependence scenarios: iid sequences and a max-autoregressive (ARMAX) model

$$X_t = \max\{\beta X_{t-1}, (1 - \beta)Z_t\}, \quad t \in \mathbb{Z},$$

with parameter $\beta \in [0, 1]$ and where $(Z_t)_t$ are nonnegative iid random variables whose distribution is in the max-domain of attraction of the Fréchet distribution with shape parameter $\alpha > 0$. The process $(X_t)_t$ is strictly stationary and admits the causal representation $X_t = \max_{j \geq 0} \{\beta^j (1 - \beta) Z_{t-j}\}$. Since $\mathbb{P}(M_{1:r} \leq x) = \mathbb{P}(X_1 \leq x) \mathbb{P}\{Z_1 \leq x/(1 - \beta)\}^{r-1}$, rescaled block maxima of X_t converge weakly to the same Fréchet distribution, but with scaling sequence depending on β . For the simulations, we fix $\beta = 1/2$, and use a burn-in period of length 200 to arrive at approximate stationarity.

We consider three different choices for the iid case and for the innovation distribution of the max-autoregressive model: the Fréchet distribution itself, the Pareto distribution with shape parameter α , and the absolute value of the Student t -distribution with α degrees of freedom.

The mean squared error, squared bias and variance for the estimators of α_0 are shown in Figure 3. We consider a fixed sample size $n = 1000$, block sizes $r = 2, 3, \dots, 50$ for the blocks estimators (i.e., $m = n/r$ ranging from 20 to 500) and numbers of upper order statistics $m = 20, 30, \dots, 500$ for the Hill estimator. The results are based on 3000 repetitions.

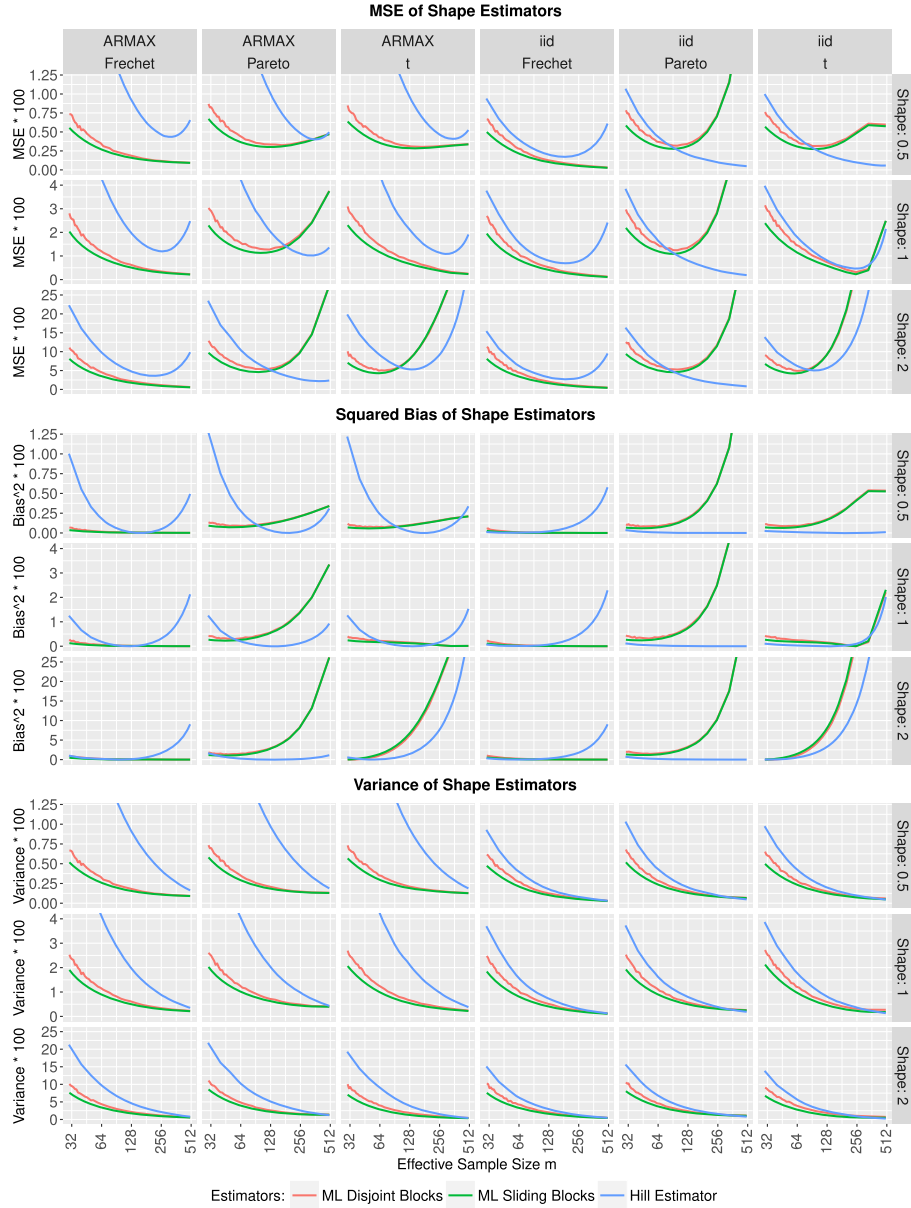


FIG 3. MSE, Squared Bias and Variance for the estimation of α_0 (multiplied by 100) as a function of the effective sample size m , where $m = n/r$ for the blocks estimators and m is equal to the number of upper order statistics used for the Hill estimator.

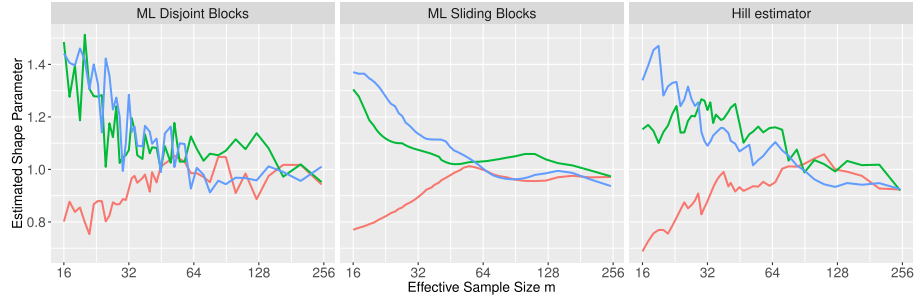


FIG 4. Three typical trajectories of three different estimators of the Fréchet shape parameter, based on iid samples of size $n = 1000$ from the t -distribution with 1 degree of freedom (i.e., $\alpha_0 = 1$). The blocks estimators are based on the 44 unique values r that are obtained by calculating the integer part of n/m_0 with m_0 ranging from 16, 17, ... up to 250. The Hill estimator is based on the respective 44 values $m = \lfloor n/r \rfloor$.

The bias-variance trade-off is clearly visible: small m (large r) yields a large variance but a small bias, while increasing m (decreasing r) decreases the variance but potentially increases the bias (with some exceptions for the Hill estimator in the ARMAX-model). The appearance of the bias curves in the iid scenarios may be explained as follows: The Hill estimator is the maximum likelihood estimator in the iid Pareto model, while the blocks estimators are maximum (quasi-)likelihood estimators in the iid Fréchet models. The absolute t -distribution is in between. The size of the bias is ordered accordingly.

In all cases, the sliding blocks estimator is more accurate than the disjoint blocks estimator due to its smaller variance.

Finally, another advantage of the sliding blocks estimator is that its trajectories fluctuate less as a function of r , as illustrated in Figure 4.

5. Proofs and auxiliary results

The proof of Theorem 2.6 is based on a sequence of auxiliary lemmas. Let $Z_{r_n,t} = (M_{r_n,t} \vee c) / \sigma_{r_n}$, with c specified in the subsequent lemmas. All convergences will be for $n \rightarrow \infty$, if not stated otherwise. Throughout the proofs, we will write $r = r_n$, $m = m_n$ etc.

Lemma 5.1 (Joint weak convergence of sliding block maxima). *Suppose that Condition 2.1 is met and that there exists an integer sequence $(\ell_n)_n$ such that $\ell_n = o(r_n)$ and $\alpha(\ell_n) = o(\ell_n/r_n)$ as $n \rightarrow \infty$. Then, for any $c \geq 0$ and any $\xi \geq 0$, we have, for $x > 0$ and $y > 0$ and as $n \rightarrow \infty$,*

$$\begin{aligned} & \lim_{r \rightarrow \infty} \mathbb{P}(Z_{r_n,1} \leq x, Z_{r_n,1+\lfloor r_n \xi \rfloor} \leq y) \\ &= G_{\alpha_0, \xi}(x, y) = \begin{cases} \exp\{-\xi x^{-\alpha_0} - (1-\xi)(x \wedge y)^{-\alpha_0} - \xi y^{-\alpha_0}\}, & \text{if } 0 \leq \xi \leq 1, \\ \exp(-x^{-\alpha_0} - y^{-\alpha_0}), & \text{if } \xi \geq 1. \end{cases} \end{aligned} \quad (5.1)$$

Surprisingly, the joint limiting law does not depend on any quantities related to the serial dependence of the time series (like the extremal index). The two margins of the limit distribution are always $P_{\alpha_0,1}$. For $\xi \geq 1$, the two block maxima are asymptotically independent. For $0 \leq \xi \leq 1$, the limiting distribution $G_{\alpha_0,\xi}$ is bivariate max-stable with $P_{\alpha_0,1}$ margins and with Pickands dependence function [16] equal to

$$A_\xi(w) = \xi + (1 - \xi)(w \wedge (1 - w)), \quad w \in [0, 1]. \quad (5.2)$$

If $(Z_1, Z_2) \sim G_{\alpha_0,\xi}$, then $(Z_1^{-\alpha_0}, Z_2^{-\alpha_0})$ is a pair of unit exponential random variables with joint survival function

$$\begin{aligned} \mathbb{P}(Z_1^{-\alpha_0} \geq s, Z_2^{-\alpha_0} \geq t) &= \exp\{-\xi s - (1 - \xi)(s \wedge t) - \xi t\} \\ &= \exp\{-(s + t) A_\xi(s/(s + t))\}. \end{aligned} \quad (5.3)$$

This is the Marshall–Olkin distribution [13] with dependence parameter ξ . The distribution depends on ξ in such a way that the dependence increases as ξ decreases, i.e., as the overlap between the two blocks increases.

Proof of Lemma 5.1. Write $r = r_n$, $\ell = \ell_n$ and $\alpha = \alpha_0$. Since $c/\sigma_r \rightarrow 0$ as $r \rightarrow \infty$, we may redefine $Z_{r,i} = M_{r,i}/\sigma_r$. By Condition 2.1,

$$\lim_{r \rightarrow \infty} \sigma_{\lfloor r\xi \rfloor} / \sigma_r = \xi^{1/\alpha}$$

for any $\xi > 0$. As a consequence,

$$\mathbb{P}(M_{1:\lfloor r\xi \rfloor} \leq \sigma_r x) = \mathbb{P}\{M_{1:\lfloor r\xi \rfloor} \leq \sigma_{\lfloor r\xi \rfloor} (\frac{\sigma_r}{\sigma_{\lfloor r\xi \rfloor}} x)\} \rightarrow \exp(-\xi x^{-\alpha}), \quad r \rightarrow \infty.$$

Consider the case $\xi \in (0, 1)$. We will show below that

$$\begin{aligned} &\mathbb{P}(Z_{r,1} \leq x, Z_{r,1+\lfloor r\xi \rfloor} \leq y) \\ &= \mathbb{P}\{M_{1:\lfloor r\xi \rfloor} \leq \sigma_r x, M_{(\lfloor r\xi \rfloor+1):r} \leq \sigma_r(x \wedge y), M_{(r+1):(r+\lfloor r\xi \rfloor)} \leq \sigma_r y\} \\ &= \mathbb{P}\{M_{1:\lfloor r\xi \rfloor} \leq \sigma_r x\} \mathbb{P}\{M_{(\lfloor r\xi \rfloor+1):r} \leq \sigma_r(x \wedge y)\} \mathbb{P}\{M_{(r+1):(r+\lfloor r\xi \rfloor)} \leq \sigma_r y\} \\ &\quad + o(1) \end{aligned} \quad (5.4)$$

as $r \rightarrow \infty$. As a consequence of the previous two displays and by strict stationarity, we obtain (5.1) for $\xi \in (0, 1)$. The case $\xi \geq 1$ can be treated similarly while the case $\xi = 0$ is trivial. It remains to show (5.4).

As a consequence of Lemma A.8 in [4], we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{1:(r'-\ell)} < M_{(r'-\ell+1):r'}) = 0$$

for any sequence r' such that r'/r is bounded away from 0 and infinity; note that Condition 3.1 in that paper follows from our assumption that $(\sigma_r)_r$ is regularly

varying. Applying this result four times and using the fact that $\lim_{\ell \rightarrow \infty} \alpha(\ell) = 0$, we may write the expression in the middle line of (5.4) as

$$\begin{aligned} & \mathbb{P}\{M_{1:([r\xi]-\ell)} \leq \sigma_r x, M_{([r\xi]+1):(r-\ell)} \leq \sigma_r(x \wedge y), M_{(r+1):(r+[r\xi])} \leq \sigma_r y\} \\ & \quad + o(1) \\ &= \mathbb{P}\{M_{1:([r\xi]-\ell)} \leq \sigma_r x\} \mathbb{P}\{M_{([r\xi]+1):(r-\ell)} \leq \sigma_r(x \wedge y)\} \mathbb{P}\{M_{(r+1):(r+[r\xi])} \leq \sigma_r y\} \\ & \quad + o(1) \\ &= \mathbb{P}\{M_{1:[r\xi]} \leq \sigma_r x\} \mathbb{P}\{M_{([r\xi]+1):r} \leq \sigma_r(x \wedge y)\} \mathbb{P}\{M_{(r+1):(r+[r\xi])} \leq \sigma_r y\} \\ & \quad + o(1) \end{aligned}$$

as $n \rightarrow \infty$, which proves (5.4). $\square$

Lemma 5.2 (Asymptotic covariances of functions of sliding block maxima). *Suppose Conditions 2.1 and 2.4 are met and that there exists an integer sequence $(\ell_n)_n$ such that $\ell_n = o(r_n)$ and $\alpha(\ell_n) = o(\ell_n/r_n)$ as $n \rightarrow \infty$. Then, for any $c > 0$, $\xi \in [0, 1]$ and any pair of measurable functions f, g on $(0, \infty)$ which are continuous almost everywhere and satisfy*

$$(|f| \vee |g|)^2 \leq g_{\eta, \alpha_0}(x) = \{x^{-\alpha_0} \mathbf{1}(x \leq e) + \log(x) \mathbf{1}(x > e)\}^{2+\eta}$$

for some $0 < \eta < \nu$, we have

$$\lim_{n \rightarrow \infty} \text{Cov}(f(Z_{r_n, 1}), g(Z_{r_n, 1+[r_n \xi]})) = \text{Cov}_{\alpha_0, \xi}(f(Z_1), g(Z_2))$$

where the right-hand side means that $(Z_1, Z_2) \sim G_{\alpha_0, \xi}$ as in (5.1).

Proof of Lemma 5.2. The result is a simple consequence of Lemma 5.1, the Cauchy–Schwarz inequality and Example 2.21 in [19]. $\square$

Lemma 5.3. *Suppose Conditions 2.1, 2.3 and 2.4 are met. Then, for any pair of measurable functions f, g which are continuous almost everywhere and satisfy*

$$(|f| \vee |g|)^2 \leq g_{\eta, \alpha_0}(x) = \{x^{-\alpha_0} \mathbf{1}(x \leq e) + \log(x) \mathbf{1}(x > e)\}^{2+\eta}$$

for some $0 < \eta < \nu$, we have

$$\sigma_{f, g} = \lim_{n \rightarrow \infty} \text{Cov}(\mathbb{G}_n f, \mathbb{G}_n g) = 2 \int_0^1 \text{Cov}_{\alpha_0, \xi}(f(Z_1), g(Z_2)) \, d\xi.$$

Proof of Lemma 5.3. Let $\ell_n = \max\{s_n, \lfloor r_n \sqrt{\alpha(s_n)} \rfloor\}$, where $s_n = \lfloor \sqrt{r_n} \rfloor$. Then $\ell_n \rightarrow \infty$, $\ell_n = o(r_n)$ and $\alpha(\ell_n) = o(\ell_n/r_n)$ as $n \rightarrow \infty$, so that the results of Lemma 5.1–5.2 become available.

For $h = 1, \dots, \lfloor n/r_n \rfloor$, let $I_h = \{(h-1)r_n + 1, \dots, hr_n\}$ denote the set of indices making up the h th disjoint block of size r_n . For simplicity assume that $m_n = n/r_n$ is an integer. Then, we may write

$$\frac{1}{k_n} \sum_{t=1}^{k_n} f(Z_{r, t}) = \frac{1}{k_n} \sum_{h=1}^{m_n} A_h, \quad \frac{1}{k_n} \sum_{t=1}^{k_n} g(Z_{r, t}) = \frac{1}{k_n} \sum_{h=1}^{m_n} B_h,$$

where $A_h = \sum_{s \in I_h} f(Z_{r,s})$ and $B_h = \sum_{s \in I_h} g(Z_{r,s})$. As a consequence,

$$\begin{aligned}
 & \text{Cov}(\mathbb{G}_n f, \mathbb{G}_n g) \\
 &= \frac{m_n}{k_n^2} \left(m_n \text{Cov}(A_1, B_1) + \sum_{h=1}^{m_n-1} (m_n - h) \{ \text{Cov}(A_1, B_{1+h}) \right. \\
 & \quad \left. + \text{Cov}(B_1, A_{1+h}) \} \right) \\
 &= \frac{m_n^2}{k_n^2} \text{Cov}(A_2, B_1 + B_2 + B_3) - \frac{m_n}{k_n^2} \text{Cov}(A_2, B_1 + B_3) \\
 & \quad + \frac{m_n^2}{k_n^2} \sum_{h=2}^{m_n-1} \left(1 - \frac{h}{m_n} \right) \{ \text{Cov}(A_1, B_{1+h}) + \text{Cov}(B_1, A_{1+h}) \}.
 \end{aligned} \tag{5.5}$$

Let us proceed by showing that

$$\begin{aligned}
 \lim_{n \rightarrow \infty} S_n &= 2 \int_0^1 \text{Cov}_{\alpha_0, \xi}(f(Z_1), g(Z_2)) \, d\xi, \\
 \text{where } S_n &= \frac{1}{r_n^2} \text{Cov}(A_2, B_1 + B_2 + B_3). \tag{5.6}
 \end{aligned}$$

For that purpose, define functions g_{n1} and g_{n2} on the positive real line by

$$g_{n1}(\xi) = \text{Cov}(f(Z_{r,1}), g(Z_{r,1+\lfloor r\xi \rfloor})), \quad g_{n2}(\xi) = \text{Cov}(f(Z_{r,1+\lfloor r\xi \rfloor}), g(Z_{r,1})).$$

We may then write

$$\begin{aligned}
 \frac{1}{r_n^2} \text{Cov}(A_2, B_2) &= \frac{1}{r_n^2} \sum_{s=1}^{r_n} \sum_{t=1}^{r_n} \text{Cov}(f(Z_{r,s}), g(Z_{r,t})) \\
 &= \frac{1}{r_n} g_{n1}(0) + \frac{1}{r_n} \sum_{h=1}^{r_n-1} \left(1 - \frac{h}{r_n} \right) \left\{ g_{n1}\left(\frac{h}{r_n}\right) + g_{n2}\left(\frac{h}{r_n}\right) \right\}, \\
 \frac{1}{r_n^2} \text{Cov}(A_2, B_3) &= \frac{1}{r_n^2} \sum_{s=1}^{r_n} \sum_{t=r_n+1}^{2r_n} \text{Cov}(f(Z_{r,s}), g(Z_{r,t})) \\
 &= \frac{1}{r_n} \sum_{h=1}^{r_n-1} \frac{h}{r_n} g_{n1}\left(\frac{h}{r_n}\right) + \frac{1}{r_n} \sum_{h=r_n}^{2r_n-1} \left(2 - \frac{h}{r_n} \right) g_{n1}\left(\frac{h}{r_n}\right), \\
 \frac{1}{r_n^2} \text{Cov}(A_2, B_1) &= \frac{1}{r_n^2} \sum_{s=1}^{r_n} \sum_{t=r_n+1}^{2r_n} \text{Cov}(f(Z_{r,s}), g(Z_{r,t})) \\
 &= \frac{1}{r_n} \sum_{h=1}^{r_n-1} \frac{h}{r_n} g_{n2}\left(\frac{h}{r_n}\right) + \frac{1}{r_n} \sum_{h=r_n}^{2r_n-1} \left(2 - \frac{h}{r_n} \right) g_{n2}\left(\frac{h}{r_n}\right).
 \end{aligned}$$

As a consequence of the three previous formulas,

$$S_n = \int_0^1 \{g_{n1}(\xi) + g_{n2}(\xi)\} d\xi + R_n$$

where the remainder R_n satisfies

$$|R_n| \leq \frac{1}{r_n} |g_{n1}(0)| + 2 \int_1^2 \{|g_{n1}(\xi)| + |g_{n2}(\xi)|\} d\xi.$$

Note that $\lim_{n \rightarrow \infty} g_{n\ell}(\xi) = \text{Cov}_{\alpha_0, \xi}(f(Z_1), g(Z_2))$ by Lemma 5.2, for $\ell = 1, 2$. In particular, the limit is zero for $\xi \geq 1$. Hence, we obtain both $R_n \rightarrow 0$ and (5.6) by dominated convergence.

Let us finally show that the sum on the right-hand side of (5.5) is negligible. The lemma then follows from (5.5) and (5.6). For that purpose, start by considering the sum over those summands for which $h \geq 3$. In this case, the observations making up A_1 and B_{1+h} are separated by $r_n(h-2)$ observations. As a consequence, invoking Lemma 3.11 in [5], we have

$$\text{Cov}(A_1, B_{1+h}) \leq 10 \|A_1\|_{2+\nu} \|B_1\|_{2+\nu} \{\alpha(r_n(h-2))\}^{\nu/(2+\nu)},$$

with $\nu > 0$ from Condition 2.4. Since $(m_n/k_n)^2 \|A_1\|_{2+\nu} \|B_1\|_{2+\nu} = O(1)$, we can bound the sum involving $\text{Cov}(A_1, B_{1+h})$ by a multiple of $\sum_{h=1}^{m_n-3} \{\alpha(hr_n)\}^{\nu/(2+\nu)}$, which converges to 0 by Condition 2.3. The same argument can be used to handle the sum involving $\text{Cov}(B_1, A_{1+h})$ for $h \geq 3$. It remains to consider the summand corresponding to $h = 2$. We may write

$$\begin{aligned} \frac{1}{r_n^2} \text{Cov}(A_1, B_3) &= \frac{1}{r_n^2} \sum_{s=1}^{r_n} \sum_{t=2r_n+1}^{3r_n} \text{Cov}(f(Z_{r,s}), g(Z_{r,t})) \\ &= \frac{1}{r_n} \sum_{h=r_n}^{2r_n-1} \frac{h}{r_n} g_{n1}\left(\frac{h}{r_n}\right) + \frac{1}{r_n} \sum_{h=2r_n}^{3r_n-1} \left(2 - \frac{h}{r_n}\right) g_{n1}\left(\frac{h}{r_n}\right), \end{aligned}$$

which can be bounded in absolute value by $3 \int_1^3 |g_{n1}(\xi)| d\xi$. As before, this integral converges to zero by dominated convergence and Lemma 5.2. Similarly, $|\text{Cov}(B_1, A_3)| = o(r_n^2)$, and the proof is finished. $\square$

Recall the polygamma function of order $m \geq 0$ and the Riemann zeta function:

$$\psi^{(m)}(z) = \frac{d^{m+1}}{dz^{m+1}} \log \Gamma(z), \quad (z > 0), \quad \zeta(z) = \sum_{k=1}^{\infty} k^{-z}, \quad (z > 1).$$

A special value that we will need is Apéry's constant, $\zeta(3) \approx 1.2020569$. Recall the functions f_1, f_2, f_3 in (2.4). For $i, j \in \{1, 2, 3\}$, define

$$\sigma_{ij} = 2 \int_0^1 \text{Cov}_{\alpha_0, \xi}(f_i(Z_1), f_j(Z_2)) d\xi. \quad (5.7)$$

Corollary 5.4. For σ_{ij} as in (5.7), we have

$$\begin{aligned}\sigma_{11} &= \alpha_0^{-2} \left[4 \log(2) \left(\psi(2)^2 + \frac{\pi^2}{6} - \psi(2) \log(2) + \log^2(2)/3 \right) \right. \\ &\quad \left. + \psi(2) \frac{\pi^2}{3} - \frac{7}{8} \zeta(3) - 2\psi(2)^2 \right], \\ \sigma_{22} &= 4 \log(2) - 2, \\ \sigma_{33} &= \alpha_0^{-2} [8 \log(2) - 4], \\ \sigma_{12} &= \alpha_0^{-1} \left[2 \log^2(2) - \frac{\pi^2}{6} - (1 - \gamma)(4 \log(2) - 2) \right], \\ \sigma_{13} &= \alpha_0^{-2} \left[(1 + \psi(2)) \frac{\pi^2}{6} + 2 \log^2(2) - 4\psi(2) \log(2) + 2\psi(2) - \frac{7}{16} \zeta(3) \right], \\ \sigma_{23} &= \alpha_0^{-1} \left[4 \log(2) - 2 - \frac{\pi^2}{6} \right].\end{aligned}$$

Proof. If $(Z_1, Z_2) \sim G_{\alpha_0, \xi}$, then $(S, T) = (Z_1^{-\alpha_0}, Z_2^{-\alpha_0})$ is a pair of unit exponential random variables whose distribution is jointly min-stable with Marshall–Olkin Pickands dependence function A_ξ ; see (5.2)–(5.3). Further, we have

$$\begin{aligned}f_1(Z_1) &= Z_1^{-\alpha_0} \log(Z_1) = -\alpha_0^{-1} S \log(S), \\ f_2(Z_1) &= Z_1^{-\alpha_0} = S, \\ f_3(Z_1) &= \log(Z_1) = -\alpha_0^{-1} \log(S),\end{aligned}$$

and similarly for $f_j(Z_2)$. We obtain

$$\begin{aligned}\text{Cov}_{\alpha_0, \xi}(f_1(Z_1), f_1(Z_2)) &= \alpha_0^{-2} \text{Cov}_\xi(S \log(S), T \log(T)), \\ \text{Cov}_{\alpha_0, \xi}(f_2(Z_1), f_2(Z_2)) &= \text{Cov}_\xi(S, T), \\ \text{Cov}_{\alpha_0, \xi}(f_3(Z_1), f_3(Z_2)) &= \alpha_0^{-2} \text{Cov}_\xi(\log(S), \log(T)), \\ \text{Cov}_{\alpha_0, \xi}(f_1(Z_1), f_2(Z_2)) &= -\alpha_0^{-1} \text{Cov}_\xi(S \log(S), T), \\ \text{Cov}_{\alpha_0, \xi}(f_1(Z_1), f_3(Z_2)) &= \alpha_0^2 \text{Cov}_\xi(S \log(S), \log(T)), \\ \text{Cov}_{\alpha_0, \xi}(f_2(Z_1), f_3(Z_2)) &= -\alpha_0^{-1} \text{Cov}_\xi(S, \log(T)),\end{aligned}$$

where the index $\xi \in [0, 1]$ stresses the dependence of the covariances on the Marshall–Olkin parameter. In Appendix A, the covariances on the right-hand side are denoted by $H_{k, \ell}(a, b; \xi) = \text{Cov}_\xi(S^a (\log(S))^k, T^b (\log(T))^\ell)$, and their integrals over $\xi \in [0, 1]$ are computed in Corollary A.3. Multiplying by two, we find the stated formulas. $\square$

Approximate values of σ_{ij} in (5.7) are summarized in the following matrix:

$$\Sigma_Y = (\sigma_{ij})_{i,j=1}^3 = \frac{1}{\alpha_0^2} \begin{pmatrix} 1.5140 & -1.0107 \cdot \alpha_0 & 0.8712 \\ -1.0107 \cdot \alpha_0 & 0.7726 \cdot \alpha_0^2 & -0.8723 \cdot \alpha_0 \\ 0.8712 & -0.8723 \cdot \alpha_0 & 1.5434 \end{pmatrix}. \quad (5.8)$$

Proof of Theorem 2.6. The theorem is a consequence of Theorem 2.5 in [4]. In the notation of that paper, let $v_n = \sqrt{m_n}$ and recall that $k_n = n - r_n + 1$ and $m_n = \lfloor n/r_n \rfloor$. As before, write $k = k_n$, $m = m_n$, and so on. Subsequently, redefine $X_{n,t} = M_{r,t} \vee c_0$ with $c_0 > 0$ from Condition 2.5. Recall that, as a consequence of Condition 2.2, such a redefinition does not change the estimator on a sequence of events whose probability converges to one. Hence, the asymptotic distribution is unaffected as well and in particular the asymptotic bias (which is identifiable from the limiting distribution) does not depend on c_0 . We need to show the following three properties:

- (i) $\lim_{n \rightarrow \infty} \mathbb{P}(X_{n,1} = \dots = X_{n,k}) = 0$.
- (ii) There exist constants $0 < \alpha_- < \alpha_0 < \alpha_+ < \infty$ such that

$$\mathbb{P}_n f = \frac{1}{k} \sum_{t=1}^k f(X_{n,t}/\sigma_r) \rightsquigarrow \int_0^\infty f(x) p_{\alpha_0,1}(x) \, dx$$

for all $f \in \mathcal{F}_2(\alpha_-, \alpha_+)$, where

$$\mathcal{F}_2(\alpha_-, \alpha_+) = \{x \mapsto \log x\} \cup \{x \mapsto x^{-\alpha} (\log x)^k : k = 0, 1, 2, \alpha \in (\alpha_-, \alpha_+)\}.$$

- (iii) For f_j as in (2.4), we have

$$(\mathbb{G}_n f_1, \mathbb{G}_n f_2, \mathbb{G}_n f_3)^T \rightsquigarrow Y \sim \mathcal{N}_3(B, \Sigma_Y), \quad n \rightarrow \infty,$$

where $\mathbb{G}_n$ and B are as in Theorem 2.6 and where Σ_Y is as in Corollary 5.3, see in particular equation (5.8).

The not-all-tied property in (i) follows immediately from Lemma A.5 in [4]: note that k_n in Condition 3.3 in that paper corresponds to m_n here.

Consider (ii). Choose $\eta \in (2/\omega, \nu)$ with ω and ν from Condition 2.3 and 2.4. Further, let $0 < \alpha_- < \alpha_0 < \alpha_+$ be arbitrary (further constraints on α_+ will be imposed below). Lemma A.6 in [4] implies that $\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{P}_n f] = Pf$, as long as α_+ is chosen smaller than $2\alpha_0$ (in that case, any $f \in \mathcal{F}_2(\alpha_-, \alpha_+)$ can be bounded in absolute value by g_{0,α_0}). Further, $\mathbb{P}_n f - \mathbb{E}[\mathbb{P}_n f] = m^{-1/2} \mathbb{G}_n f = O_{\mathbb{P}}(m^{-1/2}) = o_{\mathbb{P}}(1)$, as will be shown below in the proof of (iii). These two facts imply (ii).

Consider (iii). The empirical process $\mathbb{G}_n$ can be decomposed into a stochastic term and a bias term:

$$\mathbb{G}_n = \sqrt{m}(\mathbb{P}_n - P_n) + \sqrt{m}(P_n - P) \equiv \tilde{\mathbb{G}}_n + B_n,$$

where P_n denotes the distribution of $X_{n,1}/\sigma_r$. For $j = 1, 2, 3$, we have $B_n(f_j) \rightarrow B(f_j)$ by Condition 2.5. Let us show that the finite-dimensional distributions of

$(\tilde{\mathbb{G}}_n(f))_{f \in \mathcal{F}_2(\alpha_-, \alpha_+)}$ converge weakly to those of a zero-mean Gaussian process $\mathbb{G}$ with covariance

$$\text{Cov}(\mathbb{G}f, \mathbb{G}g) = 2 \int_0^1 \text{Cov}_{Q_{\alpha_0, \xi}}(f(U_1), g(U_2)) \, d\xi, \quad f, g \in \mathcal{F}_2(\alpha_-, \alpha_+).$$

This certainly implies (iii), and is also sufficient to close the missing gap in the proof of (ii) above.

By the Cramér–Wold device, it suffices to show weak convergence $\tilde{\mathbb{G}}_n(h) \rightsquigarrow \mathbb{G}(h)$ for $h = v^T g$ where v is a column vector and where g is a column vector of functions in $\mathcal{F}_2(\alpha_-, \alpha_+)$. Here, $\tilde{\mathbb{G}}_n(h)$ and $\mathbb{G}(h)$ are defined by linearity; in particular, $\mathbb{G}(h)$ is centred Gaussian with variance $2v^T (\int_0^1 \text{Cov}_{Q_{\alpha_0, \xi}}(g(U_1), g(U_2)) \, d\xi) v$, the integral being defined entrywise. Also, note that $|h|^{2+\delta} \lesssim g_{\eta, \alpha_0}$, provided we choose $\delta \in (2/\omega, \eta)$ and $\alpha_+ > \alpha_0$ sufficiently small. Now, let

$$I_1 = \{1, \dots, r\}, \quad I_2 = \{r+1, \dots, 2r\}, \quad \dots, \quad I_m = \{(m-1)r+1, \dots, mr\}$$

denote the indices making up the m disjoint blocks of size r . Further, let $m^* = m_n^* \geq 3$, $m^* \leq m$, be an integer sequence converging to infinity such that $m^* = o(m^{\delta/(2(1+\delta))})$ as $n \rightarrow \infty$. Define

$$\begin{aligned} J_1^+ &= I_1 \cup \dots \cup I_{m^*-2}, & J_1^- &= I_{m^*-1} \cup I_{m^*}, \\ J_2^+ &= I_{m^*+1} \cup \dots \cup I_{2m^*-2}, & J_2^- &= I_{2m^*-1} \cup I_{2m^*}, \\ J_3^+ &= I_{2m^*+1} \cup \dots \cup I_{3m^*-2}, & J_3^- &= I_{3m^*-1} \cup I_{3m^*}, \quad \dots \end{aligned}$$

and so on, that is, successively merge m^*-2 of the initial disjoint blocks to a new big block J_j^+ , and then 2 of the initial disjoint blocks to a new small block J_j^- . In total, we obtain $q = q_n = \lfloor m/m^* \rfloor \rightarrow \infty$ new (disjoint) big blocks and small blocks. For simplicity, assume that $m^*q = m$, so that each time point $1, \dots, n$ is covered by exactly one of the new blocks; otherwise, a negligible remainder term arises. We may then write

$$\begin{aligned} \tilde{\mathbb{G}}_n(h) &= \sqrt{m} \left(\frac{1}{k} \sum_{t=1}^k h(X_{n,t}/\sigma_r) - \mathbb{E}[h(X_{n,t}/\sigma_r)] \right) = \frac{1}{\sqrt{q}} \sum_{j=1}^q S_{nj}^+ + \frac{1}{\sqrt{q}} \sum_{j=1}^q S_{nj}^- \\ &\equiv A_n^+ + A_n^-, \end{aligned}$$

where, for $j = 1, \dots, q$,

$$S_{nj}^\pm = \frac{\sqrt{mq}}{k} \sum_{s \in J_j^\pm} \{h(X_{n,s}/\sigma_r) - \mathbb{E}[h(X_{n,s}/\sigma_r)]\}.$$

It suffices to show that $A_n^- = o_{\mathbb{P}}(1)$ and that $A_n^+ \rightsquigarrow \mathbb{G}(h)$ as $n \rightarrow \infty$.

Let us prove that A_n^- is negligible, and for that purpose consider its variance, since it is already centered. We have

$$\text{Var}(A_n^-) = \text{Var}(S_{n1}^-) + \frac{2}{q} \sum_{j=1}^{q-1} (q-j) \text{Cov}(S_{n1}^-, S_{n,1+j}^-)$$

$$\leq 3 \operatorname{Var}(S_{n1}^-) + 2 \sum_{j=2}^{q-1} \operatorname{Cov}(S_{n1}^-, S_{n,1+j}^-).$$

Recall that $|h|^{2+\delta} \lesssim g_{\eta, \alpha_0}$ with $\delta \in (2/\omega, \eta)$. Hence, by Condition 2.4 and since $q = \lfloor m/m^* \rfloor$,

$$\|S_{n1}^-\|_{2+\eta} \leq \frac{\sqrt{mq}}{k} 4r \cdot \|h(X_{n,1}/\sigma_r)\|_{2+\eta} \lesssim \frac{1}{\sqrt{m^*}} \cdot \|h(X_{n,1}/\sigma_r)\|_{2+\eta} = o(1).$$

Thus $\operatorname{Var}(S_{n1}^-) = o(1)$ as well. Further, by Lemma 3.1 in [5],

$$\sum_{j=2}^q \operatorname{Cov}(S_{n1}^-, S_{n,1+j}^-) \lesssim q \|S_{n1}^-\|_{2+\delta}^2 \alpha(r)^{\delta/(2+\delta)} \lesssim \frac{m}{(m^*)^2} \alpha(r)^{\delta/(2+\delta)},$$

which is of the order $o((m^*)^{-2})$ by Condition 2.3 and the fact that $2/\delta > \omega$ by the choice of δ .

Now, consider the weak convergence of A_n^+ . By a standard argument based on characteristic functions (see, e.g., the proof of Theorem 3.6 in [4]), we may assume that the triangular array $S_{n1}^+, \dots, S_{nq}^+$ is rowwise independent. As a consequence, we may apply Lyapounov's central limit theorem (Theorem 27.3 in [2]): provided that $E[(S_{nj}^+)^2]$ converges to $\operatorname{Var}(\mathbb{G}(h))$ and that

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^q E[|S_{nj}^+|^{2+\delta}]}{(\sum_{j=1}^q E[|S_{nj}^+|^2])^{1+\delta/2}} = 0, \quad (5.9)$$

we obtain that A_n^+ converges weakly to $\mathbb{G}(h)$ and the proof of the convergence of the finite-dimensional distributions of the empirical process $(\tilde{\mathbb{G}}_n(f))_{f \in \mathcal{F}_2(\alpha_-, \alpha_+)}$ is finished. Now,

$$\begin{aligned} \|S_{nj}^+\|_{2+\delta} &\leq \frac{\sqrt{mq}}{k} 2(m^* - 2)r \cdot \|h(X_{n,1}/\sigma_r)\|_{2+\delta} \\ &\lesssim \sqrt{m^*} \cdot \|h(X_{n,1}/\sigma_r)\|_{2+\delta} = O(\sqrt{m^*}) \end{aligned}$$

by Condition 2.4 (recall that $|h|^{2+\delta} \lesssim g_{\eta, \alpha_0}$). As a consequence, provided that $E[(S_{nj}^+)^2]$ is converging, the fraction in (5.9) is of the order $O(q^{-\delta/2} (m^*)^{1+\delta/2})$. Since $q = \lfloor m/m^* \rfloor$ and $m^* = o(m^{\delta/(2(1+\delta))})$, this expression converges to 0.

It remains to show that $E[(S_{nj}^+)^2] = \operatorname{Var}(S_{nj}^+)$ converges to $\operatorname{Var}(\mathbb{G}(h))$. This follows similarly as in the proof of Lemma 5.3: since $q = \lfloor m/m^* \rfloor$ and $m = \lfloor n/r \rfloor$, we may write

$$S_{n1}^+ = \frac{\sqrt{mq}}{k} \sum_{j=1}^{m^*-2} C_j = \left(\frac{1}{r\sqrt{m^*}} \sum_{j=1}^{m^*-2} C_j \right) (1 + o(1)), \quad n \rightarrow \infty,$$

where $C_j = \sum_{s \in I_j} \{h(Z_{r,s}) - E[h(Z_{r,s})]\}$. Now,

$$\operatorname{Var} \left(\frac{1}{r\sqrt{m^*}} \sum_{j=1}^{m^*-2} C_j \right) = \frac{m^* - 2}{m^*} \frac{1}{r^2} \operatorname{Var}(C_1) + \frac{2}{r^2} \sum_{j=1}^{m^*-3} \left(1 - \frac{(2+j)}{m^*}\right) \operatorname{Cov}(C_1, C_{1+j}).$$

Exactly as in the proof of Lemma 5.3, the right-hand side can be seen to be equal to $\text{Cov}(C_2, C_1 + C_2 + C_3)/r^2 + o(1)$, which further can be seen to converge to $\text{Var}(\mathbb{G}(h)) = 2v^T \text{Cov}(g(U_1), g(U_2))v$, as asserted. $\square$

Proof of Theorem 2.8. We apply Theorem 2.6 and need to check Conditions 2.1–2.5 as well as the expression of the bias function.

The max-domain of attraction Condition 2.1 follows from first-order regular variation of $-\log F$ in (2.6), which is a consequence of second-order regular variation in Condition 2.7. Condition 2.2 on the smallest block maxima follows in the same way as the proof of Condition 3.2 in the proof of Theorem 4.2 in Bücher and Segers [4]; in particular, (2.9) implies that $\log(m_n) = o(r_n)$ as $n \rightarrow \infty$, see Remark 4.5 in Bücher and Segers [4]. Strong mixing with rate in Condition 2.3 is trivially fulfilled. Finally, Conditions 2.4 and 2.5 are the same as Conditions 3.4 and 3.5 in Bücher and Segers [4], respectively, and are shown to be implied by second-order regular variation in the proof of Theorem 4.2 in the cited paper. $\square$

Proof of Corollary 3.1. The bias condition implies that

$$\begin{aligned} \sqrt{m_n} \left(\frac{\widehat{\text{RL}}_n(T, r_n)}{\text{RL}(T, r_n)} - 1 \right) &= \sqrt{m_n} \left(\frac{\widehat{\text{RL}}_n(T, r_n)}{\widetilde{\text{RL}}(T, r_n)} - 1 \right) \frac{\widetilde{\text{RL}}(T, r_n)}{\text{RL}(T, r_n)} + \Lambda_n(T) \\ &= \sqrt{m_n} \left(\frac{\widehat{\text{RL}}_n(T, r_n)}{\widetilde{\text{RL}}(T, r_n)} - 1 \right) (1 + o(1)) + \Lambda_n(T). \end{aligned}$$

The first factor on the right-hand side of this display can be written as

$$\begin{aligned} &\sqrt{m_n} \left(\frac{\hat{\sigma}_n}{\sigma_{r_n}} b_T^{1/\alpha_0 - 1/\hat{\alpha}_n} - 1 \right) \\ &= \sqrt{m_n} \left(\frac{\hat{\sigma}_n}{\sigma_{r_n}} - 1 \right) (b_T^{1/\alpha_0 - 1/\hat{\alpha}_n} - 1) + \sqrt{m_n} \left(\frac{\hat{\sigma}_n}{\sigma_{r_n}} - 1 \right) + \sqrt{m_n} (b_T^{1/\alpha_0 - 1/\hat{\alpha}_n} - 1) \\ &= o_{\mathbb{P}}(1) + \sqrt{m_n} \left(\frac{\hat{\sigma}_n}{\sigma_{r_n}} - 1 \right) + \sqrt{m_n} (\hat{\alpha}_n - \alpha_0) \frac{\log(b_T)}{\hat{\alpha}_n \alpha_0} \\ &= \beta(T, \alpha_0)' \cdot \sqrt{m_n} \begin{pmatrix} \hat{\alpha}_n - \alpha_0 \\ \hat{\sigma}_n/\sigma_{r_n} - 1 \end{pmatrix} + o_{\mathbb{P}}(1), \end{aligned}$$

which proves the corollary. $\square$

Appendix A: Covariance calculations for min-stable distributions

Let the pair of random variables (S, T) have a min-stable distribution with unit exponential margins and Pickands dependence function $A : [0, 1] \rightarrow [1/2, 1]$, i.e.,

$$\Pr(S > x, T > y) = \exp \left\{ -(x + y) A \left(\frac{x}{x+y} \right) \right\}, \quad (x, y) \in [0, \infty)^2 \setminus \{(0, 0)\}.$$

The function A is convex and satisfies $w \vee (1 - w) \leq A(w) \leq 1$ for all $w \in [0, 1]$. [18] obtained the formula

$$\text{Cov}(\log S, \log T) = \int_0^1 \frac{-\log A(w)}{w(1-w)} dw. \quad (\text{A.1})$$

To compute the asymptotic covariance matrix of the maximum likelihood estimator based on sliding blocks, we seek to generalize (A.1) to

$$H_{k,\ell}(a, b) = \text{Cov}(S^a(\log S)^k, T^b(\log T)^\ell), \quad (\text{A.2})$$

for $a, b \in [0, \infty)$ and $k, \ell \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$. In particular, we are interested in the Marshall–Olkin distribution with parameter $\xi \in [0, 1]$, which has Pickands dependence function $A = A_\xi$ in (5.2).

For any $a \in [0, \infty)$ and $k \in \mathbb{N}_0$,

$$\mathbb{E}[S^a(\log(S))^k] = \int_0^\infty s^a(\log(s))^k e^{-s} ds = \Gamma^{(k)}(1+a),$$

where $\Gamma^{(k)}$ denotes the k -th derivative of the Euler gamma function. For $k = 1$, we may further write $\Gamma'(1+a) = \Gamma(1+a)\psi(1+a)$, where $\psi = \Gamma'/\Gamma$ is the digamma function. For notational convenience, the zero-th (partial) derivative of a function is to be interpreted as the function itself.

Lemma A.1. *Let (S, T) have a min-stable distribution with unit exponential margins and Pickands dependence function A . Then, for all $a, b > 0$ and all $k, \ell \in \mathbb{N} = \{0, 1, 2, 3, \dots\}$, we have*

$$\begin{aligned} H_{k,\ell}(a, b) &= \int_0^1 \frac{\partial^{k+\ell}}{\partial a^k \partial b^\ell} \left(ab \Gamma(a+b) \frac{w^{a-1}(1-w)^{b-1}}{(A(w))^{a+b}} \right) dw \\ &\quad - \Gamma^{(k)}(1+a) \Gamma^{(\ell)}(1+b), \\ H_{k,1}(a, 0) &= \int_0^1 \frac{\partial^k}{\partial a^k} (\Gamma(a+1)((A(w))^{-a} - 1)w^{a-1}(1-w)^{-1}) dw, \\ H_{1,1}(0, 0) &= \int_0^1 \frac{-\log A(w)}{w(1-w)} dw. \end{aligned}$$

Further cases can be obtained by symmetry: we have $H_{k,\ell}(a, b) = \tilde{H}_{\ell,k}(b, a)$, where $\tilde{H}$ is given by the above formulas, but with $A(w)$ replaced by $A(1-w)$.

Proof. In the following, let $a, b > 0$.

- *Case $H_{0,0}(a, b)$.* By Fubini's theorem,

$$\begin{aligned} \mathbb{E}[S^a T^b] &= \mathbb{E} \left[\int_0^S ax^{a-1} dx \int_0^T by^{b-1} dy \right] \\ &= \mathbb{E} \left[\int_{(0,\infty)^2} \mathbf{1}(S > x, T > y) ax^{a-1} by^{b-1} d(x, y) \right] \end{aligned}$$

$$\begin{aligned}
&= \int_{(0,\infty)^2} \mathbb{P}(S > x, T > y) a x^{a-1} b y^{b-1} d(x, y) \\
&= \int_{(0,\infty)^2} e^{-(x+y) A(x/(x+y))} a x^{a-1} b y^{b-1} d(x, y).
\end{aligned}$$

Substituting $x + y = s$ and $x/(x + y) = w$, so $x = sw$ and $y = s(1 - w)$, with Jacobian $|\partial(x, y)/\partial(s, w)| = s$, and following up by substituting $sA(w) = t$, we find

$$\begin{aligned}
\mathbb{E}[S^a T^b] &= \int_{w=0}^1 \int_{s=0}^{\infty} e^{-s A(w)} a (sw)^{a-1} b (s(1-w))^{b-1} s ds dw \\
&= ab \int_{w=0}^1 \int_{t=0}^{\infty} e^{-t} (tw/A(w))^{a-1} (t(1-w)/A(w))^{b-1} (A(w))^{-2} t dt dw \\
&= ab \int_{w=0}^1 \frac{w^{a-1} (1-w)^{b-1}}{(A(w))^{a+b}} \int_{t=0}^{\infty} t^{a+b-1} e^{-t} dt dw \\
&= ab \Gamma(a+b) \int_{w=0}^1 \frac{w^{a-1} (1-w)^{b-1}}{(A(w))^{a+b}} dw.
\end{aligned}$$

• *Case $H_{k,\ell}(a, b)$.* By the mean value theorem, we have, for $b > 0$, $t > 0$ and h such that $b + h > 0$, the inequality

$$\left| \frac{t^{b+h} - t^b}{h} \right| \leq \max(t^{b+h}, t^b) |\log t|.$$

Since $\mathbb{E}[T^b (\log(T))^\ell] < \infty$ for all $b > 0$ and $\ell = 0, 1, 2, \dots$, an application of the dominated convergence theorem implies that we can interchange expectation and partial derivatives to find that

$$\begin{aligned}
\mathbb{E}[S^a (\log(S))^k T^b (\log(T))^\ell] &= \mathbb{E} \left[\frac{\partial^k S^a}{\partial a^k} \frac{\partial^\ell T^b}{\partial b^\ell} \right] = \frac{\partial^{k+\ell}}{\partial a^k \partial b^\ell} \mathbb{E}[S^a T^b] \\
&= \frac{\partial^{k+\ell}}{\partial a^k \partial b^\ell} \int_{w=0}^1 ab \Gamma(a+b) \frac{w^{a-1} (1-w)^{b-1}}{(A(w))^{a+b}} dw.
\end{aligned}$$

By the same type of argument, we can interchange the partial derivatives and the integral over w : in a small neighbourhood of a fixed pair $(a, b) \in (0, \infty)^2$, the partial derivatives with respect to a and b are bounded by a constant multiple of the function $w \mapsto w^{a-h-1} |\log(w)|^k w^{b-h-1} |\log(1-w)|^\ell$, for some $0 < h < a \wedge b$, a function which is integrable over $w \in (0, 1)$.

• *Case $H_{0,1}(a, 0)$.* By Hoeffding's covariance formula,

$$\begin{aligned}
&\text{Cov}(S^a, \log(T)) \\
&= \int_{(0,\infty) \times \mathbb{R}} \{\mathbb{P}(S^a > x, \log(T) > y) - \mathbb{P}(S^a > x) \mathbb{P}(\log(T) > y)\} d(x, y),
\end{aligned}$$

where

$$\mathbb{P}(S^a > x, \log(T) > y) = \exp \left\{ -(x^{1/a} + e^y) A \left(\frac{x^{1/a}}{x^{1/a} + e^y} \right) \right\}.$$

Apply the change of variables $s = x^{1/a} + e^y$ and $w = x^{1/a}/(x^{1/a} + e^y)$, so $x = (sw)^a$ and $y = \log(s(1-w))$ with Jacobian $|\partial(x, y)/\partial(s, w)| = as^{a-1}w^{a-1}(1-w)^{-1}$. We obtain

$$\begin{aligned} \text{Cov}(S^a, \log(T)) &= \int_0^1 \left\{ \int_0^\infty (e^{-sA(w)} - e^{-s}) as^{a-1} ds \right\} w^{a-1}(1-w)^{-1} dw \\ &= \int_0^1 ((A(w))^{-a} - 1) \Gamma(a+1) w^{a-1}(1-w)^{-1} dw. \end{aligned}$$

• *Case $H_{k,1}(a, 0)$.* As in the case $H_{k,\ell}(a, b)$, we can interchange expectation (or integration) and differentiation to find

$$\begin{aligned} &\text{Cov}(S^a \log(S), \log(T)) \\ &= \text{Cov} \left(\frac{\partial^k S^a}{\partial a^k}, \log(T) \right) = \frac{\partial^k}{\partial a^k} \text{Cov}(S^a, \log(T)) \\ &= \int_0^1 \frac{\partial^k}{\partial a^k} \{ \Gamma(a+1) ((A(w))^{-a} - 1) w^{a-1}(1-w)^{-1} \} dw. \end{aligned}$$

• *Case $H_{1,1}(0, 0)$.* This is (A.1) and can be found by a similar argument as the one for $H_{0,1}(a, 0)$. $\square$

Of special interest are the cases $a = b = 1$ and $k, \ell \in \{0, 1\}$.

Corollary A.2. *Let (S, T) have a min-stable distribution with unit exponential margins and Pickands dependence function A . With $H_{k,\ell}(a, b)$ as in (A.2) we have*

$$\begin{aligned} H_{0,0}(1, 1) &= \int_0^1 \frac{1}{(A(w))^2} dw - 1, \\ H_{0,1}(1, 1) &= \int_0^1 \frac{1}{(A(w))^2} \left[1 + \log(1-w) + \psi(2) - \log(A(w)) \right] dw - \psi(2), \\ H_{1,1}(1, 1) &= \int_0^1 \frac{1}{(A(w))^2} \left[\psi(2)^2 + 2\psi(2) + \psi'(2) + 1 \right. \\ &\quad \left. + (1 + \psi(2)) \{ \log(w) + \log(1-w) - 2\log(A(w)) \} \right. \\ &\quad \left. + (\log(w) - \log(A(w))) (\log(1-w) - \log(A(w))) \right] dw - \psi(2)^2, \\ H_{0,1}(1, 0) &= \int_0^1 \frac{1 - A(w)}{(1-w)A(w)} dw, \end{aligned}$$

$$H_{1,1}(1, 0) = \int_0^1 \frac{(1 - A(w))(\log(w) + \psi(2)) - \log(A(w))}{(1 - w)A(w)} dw,$$

$$H_{1,1}(0, 0) = \int_0^1 \frac{-\log A(w)}{w(1 - w)} dw.$$

Of further special interest is the Marshall–Olkin Pickands dependence function A_ξ in (5.2). Write $H_{k,\ell}(a, b; \xi)$ for the covariance in (A.2) if $A = A_\xi$ with $\xi \in [0, 1]$. The asymptotic covariance matrix of the maximum likelihood estimator based on sliding blocks involves the integrals of $H_{k,\ell}(a, b; \xi)$ over $\xi \in [0, 1]$.

Corollary A.3. *We have*

$$\begin{aligned} \int_0^1 H_{0,0}(1, 1; \xi) d\xi &= 2\log(2) - 1, \\ \int_0^1 H_{0,1}(1, 1; \xi) d\xi &= \frac{\pi^2}{12} - \log^2(2) + (1 - \gamma)(2\log(2) - 1), \\ \int_0^1 H_{1,1}(1, 1; \xi) d\xi &= 2\log(2) \left(\psi(2)^2 + \frac{\pi^2}{6} - \psi(2)\log(2) + \log^2(2)/3 \right) \\ &\quad + \psi(2) \frac{\pi^2}{6} - \frac{7}{4}\zeta(3) - \psi(2)^2, \\ \int_0^1 H_{0,1}(1, 0; \xi) d\xi &= \frac{\pi^2}{12} + 1 - 2\log(2), \\ \int_0^1 H_{1,1}(1, 0; \xi) d\xi &= (1 + \psi(2)) \frac{\pi^2}{12} + \log^2(2) - 2\psi(2)\log(2) + \psi(2) - \frac{7}{8}\zeta(3), \\ \int_0^1 H_{1,1}(0, 0; \xi) d\xi &= 4\log(2) - 2. \end{aligned}$$

Proof. The expressions of $H_{k,\ell}(a, b; \xi)$ follow from Corollary A.2 with $A = A_\xi$. Integrating over ξ yields a double integral over ξ and w . By interchanging integration with respect to ξ and w if necessary, we can calculate all six integrals explicitly. The algebraic details are tedious and are omitted for brevity. All formulas have been checked numerically. $\square$

Acknowledgments

The authors would like to thank three anonymous referees and an Associate Editor for their constructive comments on an earlier version of this manuscript. Axel Bücher gratefully acknowledges support by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation. Johan Segers gratefully acknowledges funding by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government.

References

- [1] Berghaus, B. and A. Bücher (2016). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. ArXiv e-prints, arXiv:1608.01903.
- [2] Billingsley, P. (1979). *Probability and Measure*. John Wiley & Sons, New York-Chichester-Brisbane. [MR0534323](#)
- [3] Bücher, A. and J. Segers (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* 17(3), 495–528. [MR3252823](#)
- [4] Bücher, A. and J. Segers (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* 24(2), 1427–1462. [MR3706798](#)
- [5] Dehling, H. and W. Philipp (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pp. 3–113. Boston, MA: Birkhäuser Boston. [MR1958777](#)
- [6] 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. [MR3322325](#)
- [7] Dombry, C. and A. Ferreira (2017). Maximum likelihood estimators based on the block maxima method. ArXiv e-prints, arXiv:1705.00465.
- [8] Ferreira, A. and L. de Haan (2015, 02). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* 43(1), 276–298. [MR3285607](#)
- [9] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3, 1163–1174. [MR0378204](#)
- [10] Katz, R. W., M. B. Parlange, and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources* 25(8), 1287 – 1304.
- [11] Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* 65(2), 291–306. [MR0722133](#)
- [12] Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance* 24(7), 1097 – 1130.
- [13] Marshall, A. W. and I. Olkin (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* 62, 30–44. [MR0215400](#)
- [14] McNeil, A. (1998). Calculating quantile risk measures for financial time series using extreme value theory. Manuscript, Department of Mathematics, ETH, Swiss Federal Technical University, Zurich, Switzerland.
- [15] Northrop, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* 18(4), 585–603. [MR3418769](#)
- [16] 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, pp. 859–878, 894–902. With a discussion. [MR0820979](#)
- [17] Robert, C. Y., J. Segers, and C. A. T. Ferro (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* 3, 993–1020. [MR2540849](#)
- [18] Tiago de Oliveira, J. (1980). Bivariate extremes: foundations and statis-

- tics. In P. Krishnaiah (Ed.), *Multivariate Analysis V*, pp. 349–366. North Holland. [MR0566351](#)
- [19] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press. [MR1652247](#)