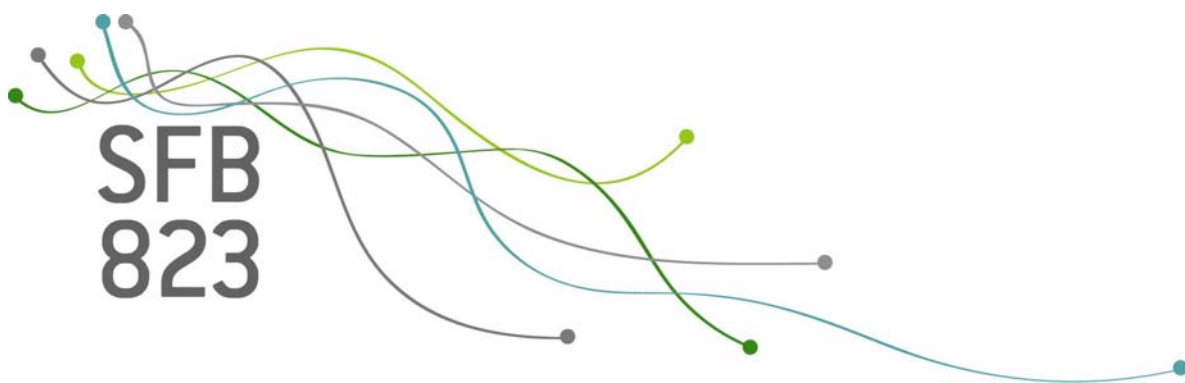# Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series

Axel Bücher, Johan Segers

Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series

Axel Bücher[*] and Johan Segers[†]

November 24, 2015

Abstract

The block maxima method in extreme-value analysis proceeds by fitting an extreme-value distribution to a sample of block maxima extracted from an observed stretch of a time series. The method is usually validated under two simplifying assumptions: the block maxima should be distributed according to an extreme-value distribution and the sample of block maxima should be independent. Both assumptions are only approximately true.

For general triangular arrays of block maxima attracted to the Fréchet distribution, consistency and asymptotic normality is established for the maximum likelihood estimator of the parameters of the limiting Fréchet distribution. The results are specialized to the setting of block maxima extracted from a strictly stationary time series. The case where the underlying random variables are independent and identically distributed is further worked out in detail. The results are illustrated by theoretical examples and Monte Carlo simulations.

Keywords and Phrases: block maxima method, maximum likelihood estimation, asymptotic normality, heavy tails, triangular arrays, stationary time series.

1 Introduction

For the analysis of extreme values, two fundamental approaches can be distinguished. First, the *peaks-over-threshold method* consists of extracting those values from the observation period which exceed a high threshold. To model such threshold excesses, asymptotic theory suggests the use of the Generalized Pareto distribution (Pickands, 1975). Second, the *block maxima method* consists of dividing the observation period into a sequence of non-overlapping intervals and restricting attention to the largest observation in each time interval. Thanks to the extremal types theorem, the probability distribution of such block maxima is approximately Generalized Extreme-Value (GEV),

---

[*]Technische Universität Dortmund, Fakultät Statistik, Vogelpothsweg 87, 44221 Dortmund, Germany. E-mail: axel.buecher@rub.de

[†]Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: johan.segers@uclouvain.be.

popularized by Gumbel (1958). The block maxima method is particularly common in environmental applications, since appropriate choices of the block size yield a simple but effective way to deal with seasonal patterns.

For both methods, honest theoretical justifications must take into account two distinct features. First, the postulated models for either threshold excesses or block maxima arise from asymptotic theory and are not necessarily accurate at sub-asymptotic thresholds or at finite block lengths. Second, if the underlying data exhibit serial dependence, then the same will likely be true for the extreme values extracted from those data.

How to deal with both issues is well-understood for the peaks-over-threshold method. The model approximation can be justified under a second-order condition (see, e.g., de Haan and Ferreira, 2006 for a vast variety of applications), while serial dependence is taken care of in Drees (2000) or Rootzén (2009), among others. Excesses over large thresholds often occur in clusters, and such serial dependence usually has an impact on the asymptotic variances of estimators based on these threshold excesses.

Surprisingly, perhaps, is that for the block maxima method, no comparable analysis has yet been done. With the exception of some recent articles, which we will discuss in the next paragraph, the commonly used assumption is that the block maxima constitute an independent random sample from a GEV distribution. The heuristic justification for assuming independence over time, even for block maxima extracted from time series data, is that for large block sizes, the occurrence times of the consecutive block maxima are likely to be well separated.

A more accurate framework is that of a triangular array of block maxima extracted from a sequence of random variables, the block size growing with the sample size. While Dombry (2013) shows consistency of the maximum likelihood estimator (Prescott and Walden, 1980) for the parameters of the GEV distribution, Ferreira and de Haan (2015) show both consistency and asymptotic normality of the probability weighted moment estimators (Hosking et al., 1985). In both papers, however, the random variables from which the block maxima are extracted are supposed to be independent and identically distributed. In many situations, this assumption is clearly violated. To the best of our knowledge, Bücher and Segers (2014) is the only reference treating both the approximation error and the time series character, providing large-sample theory of nonparametric estimators of extreme-value copulas based on samples of componentwise block maxima extracted out of multivariate stationary time series.

The aim of the paper is to show the consistency and asymptotic normality of the maximum likelihood estimator for more general sampling schemes. For technical reasons explained below, we restrict attention to the heavy-tailed case. The block maxima paradigm then suggests to use the two-parametric Fréchet distribution as a model for a sample of block maxima extracted from that time series.

The first main result, Theorem 2.5, is that for triangular arrays of random variables whose empirical measures, upon rescaling, converge in an appropriate sense to a Fréchet distribution, the maximum likelihood estimator for the Fréchet parameters based on those variables is consistent and asymptotically normal. The theorem can be applied to the set-up of block maxima extracted from an underlying time series, and the second

main result, Theorem 3.6, shows that, in this case, the asymptotic variance matrix is the inverse of the Fisher information of the Fréchet family: asymptotically, it is as if the data were an independent random sample from the Fréchet attractor. In this sense, our theorem confirms the soundness of the common simplifying assumption that the block maxima can be treated as if they were serially independent. Interestingly enough, the result also allows for time series of which the strong mixing coefficients are not summable, allowing for some long range dependence scenarios.

Restricting attention to the heavy-tailed case is done because of the non-standard nature of the three-parameter GEV distribution. The issue is that the support of a GEV distribution depends on its parameters. Even for the maximum likelihood estimator based on independent random sample from a GEV distribution, asymptotic normality has not yet been established. The article usually cited in this context is Smith (1985), although no formal result is stated therein. Even the differentiability in quadratic mean of the three-parameter GEV is still to be proven; Marohn (1994) only shows differentiability in quadratic mean for the one-parameter GEV family (shape parameter only) at the Gumbel distribution. We feel that solving all issues simultaneously (irregularity of the GEV model, finite block size approximation error and serial dependence) is a far too ample program for one paper. For that reason, we focus on the analytically simpler Fréchet family, while thoroughly treating the triangular nature of the array of block maxima and the issue of serial dependence within the underlying time series.

We will build up the theory in three stages. First, we consider general triangular arrays of observations that asymptotically follow a Fréchet distribution in Section 2. Second, we apply the theory to the set-up of block maxima extracted from a strictly stationary time series in Section 3. Third, we further specialize the results to the special case of block maxima formed from independent and identically distributed random variables in Section 4. This section can hence be regarded as a continuation of Dombry (2013) by reinforcing consistency to asymptotic normality, albeit for the Fréchet domain of attraction only. We work out an example and present finite-sample results from a simulation study in Section 5. The main proofs are deferred to Appendix A, while some auxiliary results concerning the Fréchet distribution are mentioned in Appendix B. The proofs of the less central results are postponed to a supplement.

2 Triangular arrays of block maxima

In this section, we summarize several results concerning the maximum likelihood estimator for the parameters of the Fréchet distribution. Given a sample of observations which are not all tied, the Fréchet likelihood admits a unique maximum (Subsection 2.1). If the observations are based on a triangular array which is approximately Fréchet distributed in the sense that certain functionals admit a weak law of large numbers or a central limit theorem, the maximum likelihood estimator is consistent or asymptotically normal, respectively (Subsections 2.2 and 2.3). The proofs of the results in this section are given in Subsection A.1.

2.1 Existence and uniqueness

Let P_θ denote the two-parameter Fréchet distribution on $(0, \infty)$ with parameter $\theta = (\alpha, \sigma) \in (0, \infty)^2 = \Theta$, defined through its cumulative distribution function

$$G_\theta(x) = \exp\{-(x/\sigma)^{-\alpha}\}, \quad x > 0.$$

Its probability density function is equal to

$$p_\theta(x) = \frac{\alpha}{\sigma} \exp\{-(x/\sigma)^{-\alpha}\} (x/\sigma)^{-\alpha-1}, \quad x > 0,$$

with log-likelihood function

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

and score functions $\dot{\ell}_\theta = (\dot{\ell}_{\theta,1}, \dot{\ell}_{\theta,2})^T$, with

$$\dot{\ell}_{\theta,1}(x) = \partial_\alpha \ell_\theta(x) = \alpha^{-1} + \left((x/\sigma)^{-\alpha} - 1\right)\log(x/\sigma), \tag{2.1}$$

$$\dot{\ell}_{\theta,2}(x) = \partial_\sigma \ell_\theta(x) = \left(1 - (x/\sigma)^{-\alpha}\right)\alpha/\sigma. \tag{2.2}$$

Let $\boldsymbol{x} = (x_1, \ldots, x_k) \in (0, \infty)^k$ be a sample vector to which the Fréchet distribution is to be fitted. Consider the log-likelihood function

$$L(\theta \mid \boldsymbol{x}) = \sum_{i=1}^{k} \ell_\theta(x_i). \tag{2.3}$$

Further, define

$$\Psi_k(\alpha \mid \boldsymbol{x}) = \frac{1}{\alpha} + \frac{\frac{1}{k}\sum_{i=1}^{k} x_i^{-\alpha}\log(x_i)}{\frac{1}{k}\sum_{i=1}^{k} x_i^{-\alpha}} - \frac{1}{k}\sum_{i=1}^{k}\log(x_i), \tag{2.4}$$

$$\hat{\sigma}(\alpha \mid \boldsymbol{x}) = \left(\frac{1}{k}\sum_{i=1}^{k} x_i^{-\alpha}\right)^{-1/\alpha}. \tag{2.5}$$

Lemma 2.1. (Existence and uniqueness) *If the scalars $x_1, \ldots, x_k \in (0, \infty)$ are not all equal ($k \geq 2$), then there exists a unique maximizer*

$$\hat{\theta}(\boldsymbol{x}) = \arg\max_{\theta \in \Theta} L(\theta \mid \boldsymbol{x}).$$

The maximizer is given by

$$\hat{\theta}(\boldsymbol{x}) = \left(\hat{\alpha}(\boldsymbol{x}), \hat{\sigma}(\hat{\alpha}(\boldsymbol{x}) \mid \boldsymbol{x})\right),$$

where $\hat{\alpha}(\boldsymbol{x})$ is the unique zero of the strictly decreasing function $\alpha \mapsto \Psi_k(\alpha \mid \boldsymbol{x})$:

$$\Psi\left(\hat{\alpha}(\boldsymbol{x}) \mid \boldsymbol{x}\right) = 0. \tag{2.6}$$

It is easily verified that the estimating equation for α is scale invariant: for any $c \in (0, \infty)$, we have $\Psi_k(\alpha \mid c\boldsymbol{x}) = \Psi_k(\alpha \mid \boldsymbol{x})$. As a consequence, the maximum likelihood estimator for the shape parameter is scale invariant:

$$\hat{\alpha}(c\boldsymbol{x}) = \hat{\alpha}(\boldsymbol{x}).$$

Moreover, the estimator for σ is a scale parameter in the sense that

$$\hat{\sigma}(\hat{\alpha}(c\boldsymbol{x}) \mid c\boldsymbol{x}) = c\,\hat{\sigma}(\hat{\alpha}(\boldsymbol{x}) \mid \boldsymbol{x}).$$

Until now, the maximum likelihood estimator is defined only in case not all x_i values are identical. For definiteness, if $x_1 = \ldots = x_k$, we define $\hat{\alpha}(\boldsymbol{x}) = \infty$ and $\hat{\sigma}(\infty \mid \boldsymbol{x}) = \min(x_1, \ldots, x_k) = x_1$.

2.2 Consistency

We derive a general condition under which the maximum likelihood estimator for the parameters of the Fréchet distribution is consistent. The central result, Theorem 2.3 below, shows that, apart from a not-all-tied condition, the only thing that is required for consistency is a weak law of large numbers for the functions appearing in the estimating equation (2.6) for the shape parameter.

Suppose that for each positive integer n, we are given a random vector $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,k_n})$ taking values in $(0, \infty)^{k_n}$, where $k_n \geq 2$ is a positive integer sequence such that $k_n \to \infty$ as $n \to \infty$. One may think of $X_{n,i}$ as being (approximately) Fréchet distributed with shape parameter $\alpha_0 > 0$ and scale parameter $\sigma_n > 0$. This statement is made precise in Condition 2.2 below. On the event that the k_n variables $X_{n,i}$ are not all equal, Lemma 2.1 allows us to define

$$\hat{\alpha}_n = \hat{\alpha}(\boldsymbol{X}_n), \tag{2.7}$$

the unique zero of the function $0 < \alpha \mapsto \Psi_{k_n}(\alpha \mid \boldsymbol{X}_n)$. Further, as in (2.5), put

$$\hat{\sigma}_n = \hat{\sigma}(\hat{\alpha}_n \mid \boldsymbol{X}_n) = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} X_{n,i}^{-\hat{\alpha}_n} \right)^{-1/\hat{\alpha}_n}. \tag{2.8}$$

For definiteness, put $\hat{\alpha}_n = \infty$ and $\hat{\sigma}_n = X_{n,1}$ on the event $\{X_{n,1} = \ldots = X_{n,k_n}\}$. Subsequently, we will assume that this event is asymptotically negligible:

$$\lim_{n \to \infty} \Pr(X_{n,1} = \ldots = X_{n,k_n}) = 0. \tag{2.9}$$

We refer to $(\hat{\alpha}_n, \hat{\sigma}_n)$ as the maximum likelihood estimator.

The fundamental condition guaranteeing consistency of the maximum likelihood estimator concerns the asymptotic behavior of sample averages of $f(X_{n,i}/\sigma_n)$ for certain functions f. Consider the function classes

$$\mathcal{F}_0 = \{x \mapsto x^{-\alpha} : \alpha \in (0, \infty)\} \tag{2.10}$$

$$\mathcal{F}_1 = \{x \mapsto x^{-\alpha} \log(x) : \alpha \in [0, \infty)\}, \tag{2.11}$$

5

all functions being from $(0, \infty)$ into $\mathbb{R}$. Since $\alpha = 0$ is allowed in (2.10), the logarithm belongs to $\mathcal{F}_1$ too. Let the arrow '$\rightsquigarrow$' denote weak convergence.

Condition 2.2. There exists $\alpha_0 \in (0, \infty)$ and a positive sequence σ_n such that, for all $f \in \mathcal{F}_0 \cup \mathcal{F}_1$,

$$\frac{1}{k_n} \sum_{i=1}^{k_n} f(X_{n,i}/\sigma_n) \rightsquigarrow \int_0^\infty f(x)\, p_{\alpha_0, 1}(x)\, \mathrm{d}x, \qquad n \to \infty \tag{2.12}$$

Theorem 2.3. (Consistency) *Let* $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,k_n})$ *be a sequence of random vectors in* $(0, \infty)^{k_n}$, *where* $k_n \to \infty$. *Assume that equation* (2.9) *and Condition* 2.2 *hold. On the complement of the event* $\{X_{n,1} = \ldots = X_{n,k_n}\}$, *the random vector* $(\hat{\alpha}_n, \hat{\sigma}_n)$ *is the unique maximizer of the loglikelihood* $(\alpha, \sigma) \mapsto L(\alpha, \sigma \mid X_{n,1}, \ldots, X_{n,k_n})$. *Moreover, the maximum likelihood estimator is consistent in the sense that*

$$(\hat{\alpha}_n, \hat{\sigma}_n/\sigma_n) \rightsquigarrow (\alpha_0, 1), \qquad n \to \infty.$$

2.3   Asymptotic distribution

We formulate a general condition under which the estimation error of the maximum likelihood estimator for the Fréchet parameter vector converges weakly. The central result is Theorem 2.5 below.

Recall the function classes $\mathcal{F}_0$ and $\mathcal{F}_1$ in (2.10) and (2.11) and define yet another one,

$$\mathcal{F}_2 = \{x \mapsto x^{-\alpha}(\log x)^2 : \alpha \in [0, \infty)\}. \tag{2.13}$$

Furthermore, consider the following triple of real-valued functions on $(0, \infty)$

$$\mathcal{H} = \{f_1, f_2, f_3\} = \{x \mapsto x^{-\alpha_0} \log(x),\ x \mapsto x^{-\alpha_0},\ x \mapsto \log x\}. \tag{2.14}$$

The following condition strengthens Condition 2.2.

Condition 2.4. There exist $\alpha_0 \in (0, \infty)$ and a positive sequence $(\sigma_n)_{n \in \mathbb{N}}$ such that the following two statements hold:

(i) Equation (2.12) holds for all $f \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$.

(ii) There exists a sequence $0 < v_n \to \infty$ and a random vector $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^T$ such that, denoting

$$\mathbb{G}_n f = v_n \left(\frac{1}{k_n} \sum_{i=1}^{k_n} f(X_{n,i}/\sigma_n) - \int_0^\infty f(x)\, p_{\alpha_0, 1}(x)\, \mathrm{d}x \right), \tag{2.15}$$

we have, for f_j as in (2.14),

$$(\mathbb{G}_n f_1, \mathbb{G}_n f_2, \mathbb{G}_n f_3)^T \rightsquigarrow \boldsymbol{Y}, \qquad n \to \infty. \tag{2.16}$$

6

Let Γ be the Euler gamma function and let $\gamma = 0.5772\ldots$ be the Euler–Mascheroni constant. Recall $\Gamma''(2) = (1-\gamma)^2 + \pi^2/6 - 1$. Define the matrix

$$M(\alpha_0) = \frac{6}{\pi^2} \begin{pmatrix} \alpha_0^2 & \alpha_0(1-\gamma) & -\alpha_0^2 \\ \gamma - 1 & -(\Gamma''(2)+1)/\alpha_0 & 1-\gamma \end{pmatrix}, \qquad \alpha_0 \in (0,\infty). \qquad (2.17)$$

Theorem 2.5. (Asymptotic distribution) *Let $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,k_n})$ be a sequence of random vectors in $(0,\infty)^{k_n}$, where $k_n \to \infty$. Assume that equation (2.9) and Condition 2.4 hold. As $n \to \infty$, the maximum likelihood estimator $(\hat{\alpha}_n, \hat{\sigma}_n)$ satisfies*

$$\begin{pmatrix} v_n(\hat{\alpha}_n - \alpha_0) \\ v_n(\hat{\sigma}_n/\sigma_n - 1) \end{pmatrix} = M(\alpha_0) \begin{pmatrix} \mathbb{G}_n x^{-\alpha_0} \log(x) \\ \mathbb{G}_n x^{-\alpha_0} \\ \mathbb{G}_n \log(x) \end{pmatrix} + o_p(1) \rightsquigarrow M(\alpha_0)\boldsymbol{Y}, \qquad (2.18)$$

where $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^T$ is the limit vector in (2.16) and where the matrix $M(\alpha_0)$ is given in (2.17).

For block maxima extracted from a strongly mixing stationary time series, Condition 2.4 with $v_n = \sqrt{k_n}$, where k_n denotes the number of blocks, will be derived from the Lindeberg central limit theorem. In that case, the distribution of $\boldsymbol{Y}$ is trivariate Normal with covariance matrix

$$\Sigma_{\boldsymbol{Y}} = \frac{1}{\alpha_0^2} \begin{pmatrix} 1 - 4\gamma + \gamma^2 + \pi^2/3 & \alpha_0(\gamma - 2) & \pi^2/6 - \gamma \\ \alpha_0(\gamma - 2) & \alpha_0^2 & -\alpha_0 \\ \pi^2/6 - \gamma & -\alpha_0 & \pi^2/6 \end{pmatrix}. \qquad (2.19)$$

According to Lemma B.2 below, the right-hand side in (2.19) coincides with the covariance matrix of the random vector $\left(X^{-\alpha_0} \log(X), X^{-\alpha_0}, \log(X)\right)^T$, where X is Fréchet distributed with parameter $(\alpha_0, 1)$. From Lemma B.3, recall the inverse of the Fisher information matrix of the Fréchet family at $(\alpha, \sigma) = (\alpha_0, 1)$:

$$I_{(\alpha_0,1)}^{-1} = \frac{6}{\pi^2} \begin{pmatrix} \alpha_0^2 & (\gamma - 1) \\ (\gamma - 1) & \alpha_0^{-2}\{(1-\gamma)^2 + \pi^2/6\} \end{pmatrix}. \qquad (2.20)$$

Addendum 2.6. *If $\boldsymbol{Y}$ is Normally distributed with covariance matrix $\Sigma_{\boldsymbol{Y}}$ as in (2.19), then the limit $M(\alpha_0)\boldsymbol{Y}$ in Theorem 2.5 is also Normally distributed and its covariance matrix is equal to the inverse of the Fisher information matrix of the Fréchet family, $M(\alpha_0)\,\Sigma_{\boldsymbol{Y}}\,M(\alpha_0)^T = I_{(\alpha_0,1)}^{-1}$ in (2.20).*

3 Block maxima extracted from a stationary time series

Let $(\xi_t)_{t\in\mathbb{Z}}$ be a strictly stationary time series. For positive integer i and r, consider the block maximum

$$M_{r,i} = \max(\xi_{(i-1)r+1}, \ldots, \xi_{ir}).$$

Abbreviate $M_{r,1} = M_r$. The classical block maxima method consists of choosing a sufficiently large block size r and fitting an extreme-value distribution to a sample of block maxima $M_{r,1}, \ldots, M_{r,k}$. The likelihood is constructed under the simplifying assumption that the block maxima are independent. The present section shows consistency and asymptotic normality of this method in an appropriate asymptotic framework.

For the block maxima distribution to approach its extreme-value limit, the block sizes must increase to infinity. Moreover, consistency can only be achieved when the number of blocks grows to infinity too. Hence, we consider a positive integer sequence r_n, to be thought of as a sequence of block sizes. The number of disjoint blocks of size r_n that fit into a sample of size n is equal to $k_n = \lfloor n/r_n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of a real number x. Assume that both $r_n \to \infty$ and $k_n \to \infty$ as $n \to \infty$.

The theory will be based on an application of Theorem 2.5 to the sample of left-truncated block maxima $X_{n,i} = M_{r_n,i} \vee c$ $(i = 1, \ldots, k_n)$, for some positive constant c specified below. The estimators $\hat{\alpha}_n$ and $\hat{\sigma}_n$ are thus the ones in (2.7) and (2.8), respectively. The reason for the left truncation is that otherwise, some of the block maxima could be zero or negative. Asymptotically, such left-truncation does not matter, since all maxima will simultaneously diverge to infinity in probability (Condition 3.2 below).

In Section 4 below, we will specialize things further to the case where the random variables ξ_t are independent. In particular, we will simplify the list of conditions given in this section.

The basic assumption is that the distribution of rescaled block maxima is asymptotically Fréchet. The sequence of scaling constants should possess a minimal degree of regularity. The assumption is satisfied in case the stationary distribution of the series is in the Fréchet domain of attraction and the series possesses a positive extremal index; see Remark 3.7 below.

Condition 3.1. (Domain of attraction) The time series $(\xi_t)_{t \in \mathbb{Z}}$ is strictly stationary and there exist a positive real sequence $\sigma_n \to \infty$ and a positive real number α_0 such that

$$M_n/\sigma_n \rightsquigarrow \text{Fréchet}(\alpha_0, 1), \qquad n \to \infty. \tag{3.1}$$

Moreover, $\sigma_{m_n}/\sigma_n \to 1$ for any positive integer sequence m_n such that $m_n/n \to 1$ as $n \to \infty$.

The domain-of-attraction condition implies that, for every scalar c, we have $\Pr[M_n \leq c] = \Pr[M_n/\sigma_n \leq c/\sigma_n] \to 0$ as $n \to \infty$. In words, the block maxima become unboundedly large as the sample size grows to infinity. Still, out of a sample of k_n block maxima, the smallest of the maxima might still be small, especially when the number of blocks is large, or, equivalently, the block sizes are not large enough. The following condition prevents this from happening.

Condition 3.2. (All block maxima diverge) For every $c \in (0, \infty)$, we have

$$\lim_{n \to \infty} \Pr[\min(M_{r_n,1}, \ldots, M_{r_n,k_n}) \leq c] = 0.$$

The Fréchet distribution satisfies $\int_0^\infty x^\beta p_{\alpha_0,1}(x)\,\mathrm{d}x < \infty$ for every real β less than α_0. In order to be able to integrate (3.1) to the limit, we require an asymptotic bound on the moments of the block maxima.

Condition 3.3. (Moments) For all $c \in (0,\infty)$ and all $\beta \in (-\infty, \alpha_0)$, we have

$$\limsup_{n\to\infty} \mathbb{E}\big[\big\{(M_n \vee c)/\sigma_n\big\}^\beta\big] < \infty.$$

To control the serial dependence within the time series, we require that the Rosenblatt mixing coefficients decay sufficiently fast: for positive integer ℓ, put

$$\alpha(\ell) = \sup_{\substack{A\in\sigma(\xi_t:t\le 0)\\ B\in\sigma(\xi_t:t\ge \ell)}} |\Pr(A\cap B) - \Pr(A)\Pr(B)|.$$

Condition 3.4. (α-Mixing with rate) We have $\lim_{\ell\to\infty}\alpha(\ell) = 0$. Moreover, there exists $\omega > 0$ such that
$$k_n^{1+\omega}\,\alpha(r_n) \to 0, \qquad n \to \infty. \tag{3.2}$$

Condition 3.4 can be interpreted as requiring the block sizes r_n to be sufficiently large. For instance, if $\alpha(\ell) = O(\ell^{-a})$ for some $a > 0$, then (3.2) is satisfied as soon as r_n is of larger order than $n^{(1+\omega)/(1+\omega+a)}$. Note that the exponent a is allowed to be smaller than one, in which case the sequence of mixing coefficients is not summable.

By Condition 3.2 and Lemma A.5, the probability of the event that all block maxima $M_{r_n,1},\ldots,M_{r_n,k_n}$ are larger than some positive constant c and that they are not all equal tends to unity. On this event, we can study the maximum likelihood estimators $(\hat\alpha_n,\hat\sigma_n)$ for the parameters of the Fréchet distribution based on the sample of block maxima.

Fix $c \in (0,\infty)$ and put
$$X_{n,i} = M_{r_n,i} \vee c.$$

Let $\mathbb{G}_n$ be the empirical process associated to $X_{n,1}/\sigma_n,\ldots,X_{n,k_n}/\sigma_n$ as in (2.15) with $v_n = \sqrt{k_n}$. The empirical process is not necessarily centered, which is why we need a handle on its expectation.

Condition 3.5. (Bias) There exists $c \in (0,\infty)$ such that for every function f in $\mathcal{H}$ defined in (2.14), the following limit exists:

$$\lim_{n\to\infty} \sqrt{k_n}\left(\mathbb{E}\big[f\big((M_{r_n}\vee c)/\sigma_{r_n}\big)\big] - \int_0^\infty f(x)\,p_{\alpha_0,1}(x)\,\mathrm{d}x\right) = B(f). \tag{3.3}$$

Theorem 3.6. *Suppose that Conditions 3.1 up to 3.5 are satisfied and fix c as in Condition 3.5. Then, with probability tending to one, there exists a unique maximizer $(\hat\alpha_n,\hat\sigma_n)$ of the Fréchet log-likelihood (2.3) based on the block maxima $M_{r_n,1},\ldots,M_{r_n,k_n}$,*

9

and we have, as $n \to \infty$,

$$\begin{pmatrix} \sqrt{k_n}\,(\hat\alpha_n - \alpha_0) \\ \sqrt{k_n}\,(\hat\sigma_n/\sigma_n - 1) \end{pmatrix} = M(\alpha_0) \begin{pmatrix} \mathbb{G}_n x^{-\alpha_0} \log(x) \\ \mathbb{G}_n x^{-\alpha_0} \\ \mathbb{G}_n \log(x) \end{pmatrix} + o_p(1)$$
$$\rightsquigarrow \mathcal{N}_2\big(M(\alpha_0)\,B,\ I_{\alpha_0,1}^{-1}\big).$$

Here, $M(\alpha_0)$ and $I_{\alpha_0,1}^{-1}$ are defined in equations (2.17) and (2.20), respectively, and $B = (B(f_1), B(f_2), B(f_3))^T$, where $B(f)$ is the limit in (3.3) and where f_1, f_2, f_3 are defined in (2.14).

The proof of Theorem 3.6 is given in Subsection A.2. The conditions imposed in Theorem 3.6 are rather high-level. In the setting of a sequence of independent and identically distributed random variables, they can be brought down to analytical conditions on the tail of the stationary distribution function (Theorem 4.2). Moreover, all conditions will be worked out in a particular time series model in Section 5.1.

Interestingly, the asymptotic covariance matrix is unaffected by serial dependence and the asymptotic standard deviation of $\sqrt{k_n}(\hat\alpha_n - \alpha_0)$ is always equal to $(\sqrt{6}/\pi) \times \alpha_0 \approx 0.7797 \times \alpha_0$. The reason for this invariance is that even for time series, maxima over large disjoint blocks are asymptotically independent because of the strong mixing condition.

Remark 3.7. (Domain-of-attraction condition for positive extremal index) Let F be the cumulative distribution function of ξ_1. Assume that there exist $0 < a_n \to \infty$ and $\alpha_0 \in (0, \infty)$ such that

$$\lim_{n\to\infty} F^n(a_n x) = \exp(-x^{-\alpha_0}), \qquad x \in (0, \infty).$$

Moreover, assume that the sequence $(\xi_t)_{t\in\mathbb{Z}}$ has extremal index $\vartheta \in (0, 1]$ (Leadbetter, 1983): If $u_n \to \infty$ is such that $F^n(u_n)$ converges, then

$$\Pr(M_n \le u_n) = F^{n\vartheta}(u_n) + o(1), \qquad n \to \infty.$$

Note that we assume that $\vartheta > 0$. Put

$$\sigma_n = \vartheta^{1/\alpha_0} a_n.$$

Then Condition 3.1 is satisfied: for every $x \in (0, \infty)$, we have

$$\Pr(M_n/\sigma_n \le x) = F^{n\vartheta}(\sigma_n x) + o(1)$$
$$\to \exp\big(-\vartheta(\vartheta^{1/\alpha_0} x)^{-\alpha_0}\big) = \exp(-x^{-\alpha_0}), \qquad n \to \infty.$$

4 Block maxima extracted from an iid sample

We specialize Theorem 3.6 to the case where the random variables $\xi_1, \xi_2, \ldots$ are independent and identically distributed with common distribution function F. In this setting,

fitting extreme-value distributions to block maxima is also considered in Dombry (2013) (consistency of the maximum likelihood estimator in the GEV-family with $\gamma > -1$) and Ferreira and de Haan (2015) (asymptotic normality of the probability weighted moment estimator in the GEV-family with $\gamma < 1/2$). Assume that F is in the maximum domain of attraction of the Fréchet distribution with shape parameter $\alpha_0 \in (0, \infty)$: there exists a positive scalar sequence $(a_n)_{n \in \mathbb{N}}$ such that, for every $x \in (0, \infty)$,

$$F^n(a_n x) \to e^{-x^{-\alpha_0}}, \qquad n \to \infty. \tag{4.1}$$

Because of serial independence, the conditions in Theorem 3.6 can be simplified considerably. In addition, the mean vector of the asymptotic bivariate Normal distribution of the maximum likelihood estimator can be made explicit. Required is a second-order reinforcement of (4.1) in conjunction with a growth restriction on the number of blocks.

Equation (4.1) is equivalent to regular variation of $-\log F$ at infinity with index $-\alpha_0$ (Gnedenko, 1943): we have $F(x) < 1$ for all $x \in \mathbb{R}$ and

$$\lim_{u \to \infty} \frac{-\log F(ux)}{-\log F(u)} = x^{-\alpha_0}, \qquad x \in (0, \infty). \tag{4.2}$$

The scaling constants in (4.1) may be chosen as any sequence a_n that satisfies

$$\lim_{n \to \infty} n \left\{ -\log F(a_n) \right\} = 1. \tag{4.3}$$

For $\tau \in \mathbb{R}$, define the function $h_\tau : (0, \infty) \to \mathbb{R}$ by

$$h_\tau(x) = \int_1^x y^{\tau - 1} \, \mathrm{d}y = \begin{cases} \dfrac{x^\tau - 1}{\tau}, & \text{if } \tau \neq 0, \\ \log(x), & \text{if } \tau = 0. \end{cases} \tag{4.4}$$

The following condition reinforces (4.2) and thus (4.1).

Condition 4.1. (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 $\lim_{u \to \infty} A(u) = 0$ and such that, for all $x \in (0, \infty)$,

$$\lim_{u \to \infty} \frac{1}{A(u)} \left(\frac{-\log F(ux)}{-\log F(u)} - x^{-\alpha_0} \right) = x^{-\alpha_0} h_\rho(x). \tag{4.5}$$

The function A can be regarded as capturing the speed of convergence in (4.2). The form of the limit function in (4.5) may seem unnecessarily specific, but actually, it is not, as explained in Remark 4.3 below.

Let $\psi = \Gamma'/\Gamma$ denote the digamma function and recall the Euler–Mascheroni constant $\gamma = -\Gamma'(1) = 0.5772\ldots$. To express the asymptotic bias of the maximum likelihood estimators, we will employ the functions b_1 and b_2 defined by

$$b_1(x) = \begin{cases} (1 + x) \, \Gamma(x) \{ \gamma + \psi(1 + x) \}, & \text{if } x > 0, \\ \dfrac{\pi^2}{6}, & \text{if } x = 0, \end{cases} \tag{4.6}$$

11

and

$$b_2(x) = \begin{cases} -\dfrac{\pi^2}{6x} + (1+x)\,\Gamma(x)\{\Gamma''(2) + \gamma + (\gamma-1)\,\psi(1+x)\}, & \text{if } x > 0, \\ 0, & \text{if } x = 0. \end{cases} \tag{4.7}$$

See Figure 1 for the graphs of these two functions. For $(\alpha_0, \rho, \lambda) \in (0, \infty) \times (-\infty, 0] \times \mathbb{R}$, define the bias function

$$B(\alpha_0, \rho, \lambda) = \frac{6\lambda}{\pi^2} \begin{pmatrix} b_1(|\rho|/\alpha_0) \\ b_2(|\rho|/\alpha_0)/\alpha_0^2 \end{pmatrix}. \tag{4.8}$$

The proof of the following theorem is given in Section A.3.

Theorem 4.2. *Let $\xi_1, \xi_2, \ldots$ be independent random variables with common distribution function F satisfiying Condition 4.1. Let the block sizes r_n be such that $r_n \to \infty$ and $k_n = \lfloor n/r_n \rfloor \to \infty$ as $n \to \infty$ and assume that*

$$\lim_{n\to\infty} \frac{k_n \log(k_n)}{n} = 0, \tag{4.9}$$

$$\lim_{n\to\infty} \sqrt{k_n}\, A(a_{r_n}) = \lambda \in \mathbb{R}. \tag{4.10}$$

Then, with probability tending to one, there exists a unique maximizer $(\hat\alpha_n, \hat\sigma_n)$ of the Fréchet log-likelihood (2.3) based on the block maxima $M_{r_n,1}, \ldots, M_{r_n,k_n}$, and we have

$$\sqrt{k_n} \begin{pmatrix} \hat\alpha_n - \alpha_0 \\ \hat\sigma_n/a_{r_n} - 1 \end{pmatrix} \rightsquigarrow \mathcal{N}_2 \left(B(\alpha_0, \rho, \lambda),\ I_{(\alpha_0, 1)}^{-1} \right), \qquad n \to \infty, \tag{4.11}$$

where $I_{(\alpha_0,1)}^{-1}$ denotes the inverse of the Fisher information of the Fréchet family as in (2.20) and with $B(\alpha_0, \rho, \lambda)$ as in (4.8).

Condition (4.9) is also imposed in Dombry (2013), while a version of (4.10) is used in Ferreira and de Haan (2015) to prove asymptotic normality of probability weighted moment estimators.

Remark 4.3. (Second-order regular variation) Let F satisfy (4.2). For $x > 0$ sufficiently large such that $F(x) > 0$, define $L(x)$ by

$$-\log F(x) = x^{-\alpha_0}\, L(x). \tag{4.12}$$

In view of (4.2), the function L is slowly varying at infinity, that is,

$$\lim_{u\to\infty} \frac{L(ux)}{L(u)} = 1, \qquad x \in (0, \infty).$$

A second-order refinement of this would be that there exist $A : (0, \infty) \to (0, \infty)$ and $h : (0, \infty) \to \mathbb{R}$, the latter not identically zero, such that $\lim_{u\to\infty} A(u) = 0$ and

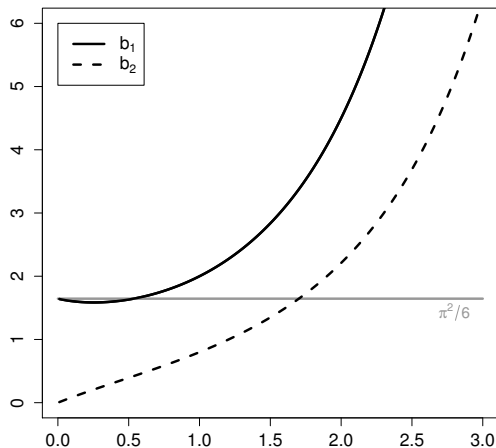$$\lim_{u\to\infty} \frac{1}{A(u)} \left(\frac{L(ux)}{L(u)} - 1 \right) = h(x), \qquad x \in (0, \infty)$$

12

Figure 1: Graphs of the functions b_1 and b_2 in (4.6) and (4.7).

Writing $g(u) = A(u) L(u)$, Theorem B.2.1 in de Haan and Ferreira (2006) (see also Bingham et al., 1987, Section 3.6) implies that there exists $\rho \in \mathbb{R}$ such that g and thus $A = g/L$ are regularly varying at infinity with index ρ. Since A vanishes at infinity, necessarily $\rho \leq 0$. Furthermore, there exists $\kappa \in \mathbb{R} \setminus \{0\}$ such that $h(x) = \kappa \, h_\rho(x)$ for $x \in (0, \infty)$, with h_ρ as in (4.4). Incorporating the constant κ into the function A, we can assume without loss of generality that $\kappa = 1$ and we arrive at Condition 4.1. The function A then possibly takes values in $(-\infty, 0)$ rather than in $(0, \infty)$.

Remark 4.4. (Mean squared error) According to (4.10) and (4.11), the distribution of the estimation error $\hat{\alpha}_n - \alpha_0$ is approximately equal to

$$\mathcal{N}\left(A(a_{r_n}) \frac{6}{\pi^2} \, b_1(|\rho|/\alpha_0), \ \frac{r_n}{n} \frac{6}{\pi^2} \, \alpha_0^2 \right).$$

The asymptotic mean squared error is therefore equal to

$$\mathrm{AMSE}(\hat{\alpha}_n) = \mathrm{ABias}^2(\hat{\alpha}_n) + \mathrm{AVar}(\hat{\alpha}_n) = |A(a_{r_n})|^2 \, \frac{36}{\pi^4} \, b_1(|\rho|/\alpha_0)^2 + \frac{r_n}{n} \frac{6}{\pi^2} \alpha_0^2.$$

The choice of the block size r_n (or, equivalently, the number of blocks k_n), thus involves a bias–variance trade-off; see Section 5. Alternatively, if ρ and $A(a_{r_n})$ could be estimated, then one could construct bias-reduced estimators, just as in the case of the Hill estimator (see, e.g., Peng, 1998, among others) or probability weighted moment estimators (Cai et al., 2013).

5 Examples and finite-sample results

5.1 Verification of conditions in a moving maximum model

Let $(Z_t)_{t \in \mathbb{Z}}$ be a sequence of independent and identically distributed random variables with common distribution function F in the domain of attraction of the Fréchet distribution with shape parameter α_0, that is, such that (4.1) is satisfied for some sequence

13

$a_n \to \infty$. Let $p \in \mathbb{N}$, $p \geq 2$, be fixed and let $b_1, \ldots, b_p$ be nonnegative constants, $b_1 \neq 0 \neq b_p$, such that $\sum_{i=1}^{p} b_i = 1$. We consider the moving maximum process ξ_t of order p, defined by

$$\xi_t = \max\{b_1 Z_t, b_2 Z_{t-1}, \ldots, b_p Z_{t-p+1}\}, \qquad t \in \mathbb{Z}.$$

A simple calculation (see also the proof of Lemma 5.1 for the stationary distribution of ξ_t) shows that the extremal index of $(\xi_t)_{t \in \mathbb{Z}}$ is equal to

$$\theta = \{\textstyle\sum_{i=1}^{p} b_i^{\alpha_0}\}^{-1} b_{(p)}^{\alpha_0},$$

where $b_{(p)} = \max_{i=1}^{p} b_i$. Let $\sigma_n = b_{(p)} a_n$. The proof of the following lemma is given in Section C in the supplementary material.

Lemma 5.1. *The stationary time series $(\xi_t)_{t \in \mathbb{Z}}$ satisfies Conditions 3.1, 3.3 and 3.4. If additionally (4.9) is met, then Condition 3.2 is satisfied as well. Finally, if F satisfies the Second-Order Condition 4.1, if (4.10) is met and if $k_n = o(n^{2/3})$ as $n \to \infty$, then Condition 3.5 is also satisfied, with $B(f)$ denoting the same constant appearing in the iid case, that is $B(f) = \boldsymbol{\beta}$ with $\boldsymbol{\beta}$ as defined in (A.24).*

As a consequence, Theorem 3.6 may be applied and the asymptotic bias of the maximum likelihood estimator is the same as specified in Theorem 4.2 for the case of independent and identically distributed random variables.

5.2 Simulation results

We report on the results of a simulation study, highlighting some interesting features regarding the finite-sample sample performance of the maximum likelihood estimator. Attention is restricted to estimation of the shape parameter, and particular emphasis is given to a comparison with the common Hill estimator. Note that the latter is based on the competing peak-over-threshold method. Its variance is of the order $O(k^{-1})$, where k is the number of upper order statistics taken into account for its calculation. Note that the asymptotic variance of the Hill estimator is given by α_0^2, which is larger than the asymptotic variance $(6/\pi^2) \times \alpha_0^2$ of the block maxima maximum likelihood estimator. We also ran numerical experiments (not shown) on the probability weighted moment estimator, but the general finding was that its asymptotic variance was much higher than the one of the maximum likelihood estimator.

We consider three time series models for $(\xi_t)_{t \in \mathbb{Z}}$: independent and identically distributed random variables, the moving maximum process from Section 5.1, and the absolute values of a GARCH(1,1) time series. In the first two models, three choices are considered for the distribution function F of either the variables ξ_t in the first model and the innovations Z_t in the second model: absolute values of a Cauchy-distribution, the standard Pareto distribution and the Fréchet(1,1) distribution itself. All three distribution functions are attracted to the Fréchet distribution with $\alpha_0 = 1$. For the moving maximum process, we fix $p = 4$ and $b_j = j/10$ for $j \in \{1, 2, 3, 4\}$. The GARCH(1,1)

14

model is based on standard Normal innovations, that is, $\xi_t = |Z_t|$, where Z_t is the stationary solution of the equations

$$\left.\begin{array}{rcl} Z_t & = & \varepsilon_t \sigma_t, \\ \sigma_t^2 & = & \lambda_0 + \lambda_1 Z_{t-1}^2 + \lambda_2 \sigma_{t-1}^2, \end{array}\right\} \tag{5.1}$$

with ε_t, $t \in \mathbb{Z}$, independent standard Normal random variables. The parameter vector $(\lambda_0, \lambda_1, \lambda_2)$ is set to either $(0.5, 0.367, 0.367)$ or $(0.5, 0.08, 0.91)$. By Mikosch and Stărică (2000), the stationary distribution associated to any of these two models is attracted to the Fréchet distribution with shape parameter being (approximately) equal to $\alpha_0 = 5$.

We generate samples from all of the afore-mentioned models for a fixed sample size of $n = 1,000$. Based on $N = 3,000$ Monte Carlo repetitions, we obtain empirical estimates of the finite sample bias, variance and mean squared error (MSE) of the competing estimators. The results are summarized in Figure 2 for the iid and the moving maxima model, and in Figure 3 for the GARCH-model.

In general, (most of) the graphs nicely reproduce the bias-variance tradeoff, its characteristic form however varying from model to model. Consider the iid scenario: since the Hill estimator is essentially the maximum likelihood estimator in the Pareto family, it is to be expected that it outperforms the block maxima estimator. On the other hand, by max-stability of the Fréchet family, the block maxima estimator should outperform the Hill estimator for that family. These expectations are confirmed by the simulation results in the left column of Figure 2. For the Cauchy distribution it turns out that the block maxima maximum likelihood estimator shows a better performance.

Now, consider the moving maxima time series scenarios (right column in Figure 2). Compared to the iid case, we observe an increase in the mean squared error (note that the scale on the axis of ordinates is row-wise identical). The block maxima method clearly outperforms the Hill estimator, except for the Pareto model. The big increase in relative performance is perhaps not too surprising, as the data points from a moving maximum process are already (weighted) maxima, which principally favors the block maxima method with small block sizes.

Finally, consider the GARCH models in Figure 3. While, as in line with the theoretical findings, the variance of the block maxima estimator is smaller than the one of the Hill estimator, the squared bias turns out to be substantially higher for a large range of values for k. The MSE-optimal point is smaller for the Hill estimator.

A   Proofs

A.1   Proofs for Section 2

Proof of Lemma 2.1. The proof extends the development in Section 2 of Balakrishnan and Kateri (2008). First, fix $\alpha > 0$ and consider the function $0 < \sigma \mapsto L(\alpha, \sigma \mid \boldsymbol{x})$. By equation (2.2), its derivative is equal to

$$\partial_\sigma L(\alpha, \sigma \mid \boldsymbol{x}) = \sum_{i=1}^k \partial_\sigma \ell_\theta(x_i) = (\alpha/\sigma)\left(k - \sigma^\alpha \sum_{i=1}^k x_i^{-\alpha}\right).$$
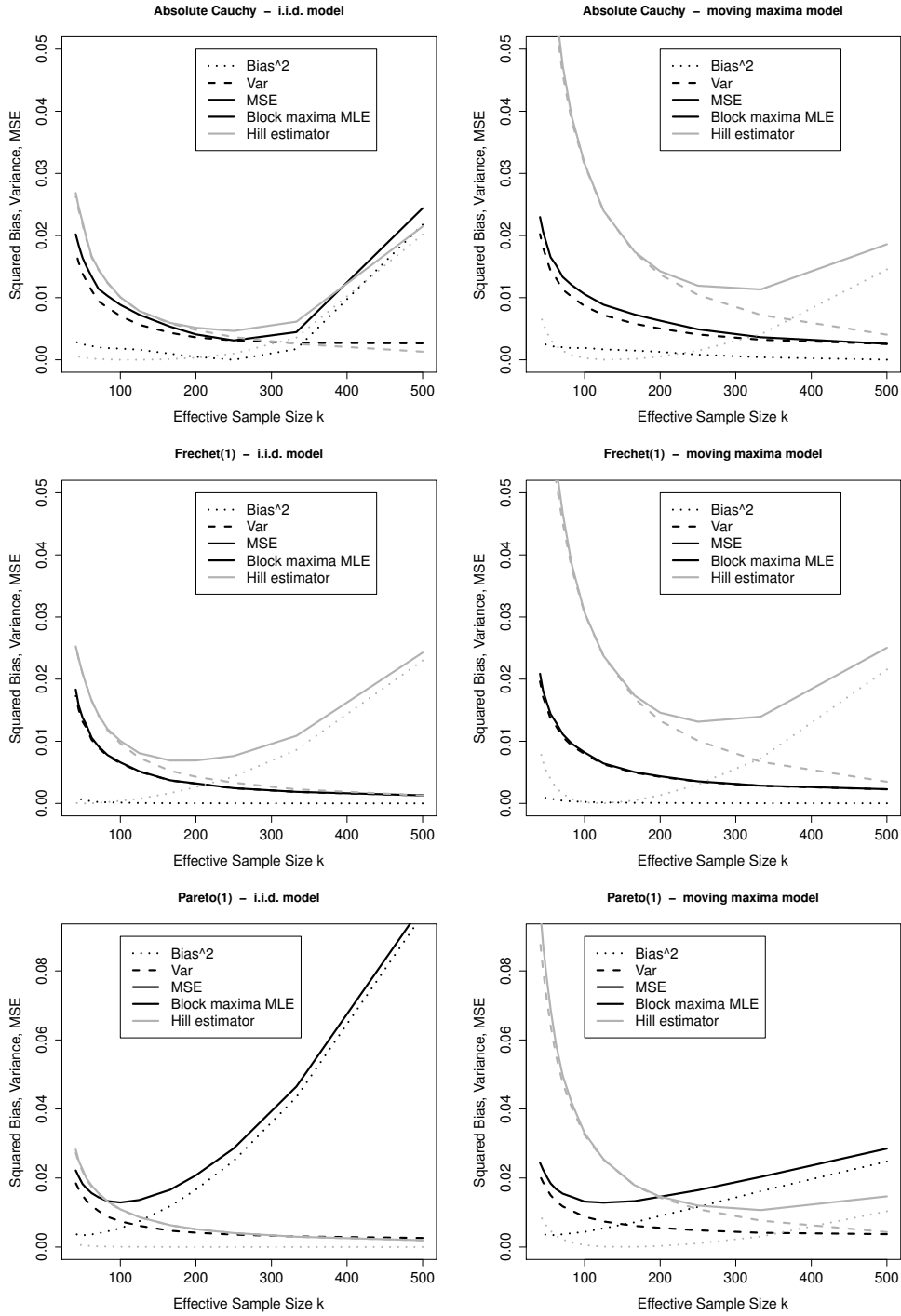
Figure 2: Simulation results (Section 5.2). Effective sample size refers to the number of blocks (block maxima MLE) or the number of upper order statistics (Hill estimator). Time series models: iid (left) and moving maximum model (right). Innovations: absolute values of Cauchy (top), unit Fréchet (middle) and unit Pareto (bottom) random variables. Block sizes $r \in \{2, 3, \ldots, 24\}$, resulting in $k \in \{500, 333, \ldots, 41\}$ blocks.
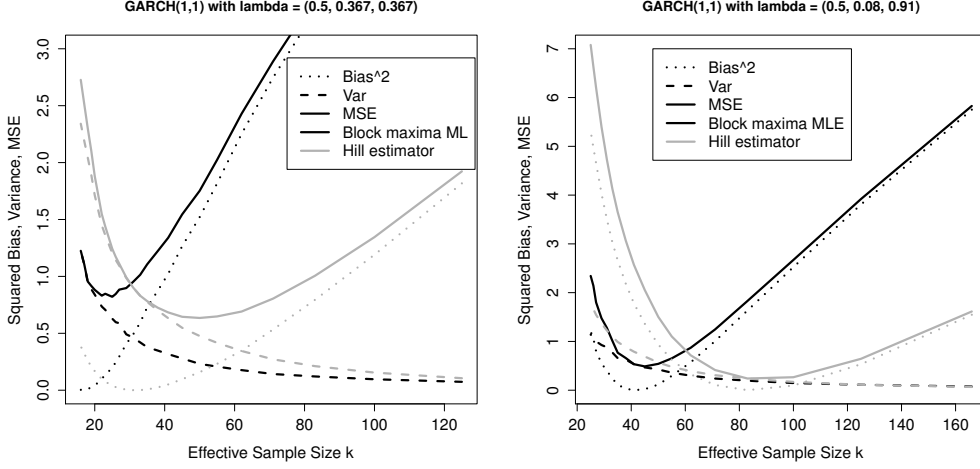
Figure 3: Simulation results (Section 5.2). Effective sample size refers to the number of blocks (block maxima MLE) or the number of upper order statistics (Hill estimator). Both panels refer to the GARCH(1,1) model in (5.1), with $(\lambda_0, \lambda_1, \lambda_2)$ equal to $(0.5, 0.367, 0.367)$ on the left and to $(0.5, 0.08, 0.91)$ on the right.

We find that $\partial_\sigma L(\alpha, \sigma \mid \boldsymbol{x})$ is positive, zero, or negative according to whether σ is smaller than, equal to, or larger than $\hat{\sigma}(\alpha \mid \boldsymbol{x})$, respectively. In particular, for fixed α, the expression $L(\alpha, \sigma \mid \boldsymbol{x})$ is maximal at σ equal to $\hat{\sigma}(\alpha \mid \boldsymbol{x})$. Hence we need to find the maximum of the function $0 < \alpha \mapsto L(\alpha, \hat{\sigma}(\alpha \mid \boldsymbol{x}) \mid \boldsymbol{x})$. By (2.1), its derivative is given by

$$\frac{d}{d\alpha} L(\alpha, \hat{\sigma}(\alpha \mid \boldsymbol{x}) \mid \boldsymbol{x}) = \sum_{i=1}^{k} \partial_\alpha \ell_{\alpha,\sigma}(x_i) \Big|_{\sigma=\hat{\sigma}(\alpha \mid \boldsymbol{x})} + \sum_{i=1}^{k} \partial_\sigma \ell_{\alpha,\sigma}(x_i) \Big|_{\sigma=\hat{\sigma}(\alpha \mid \boldsymbol{x})} \times \frac{d}{d\alpha} \hat{\sigma}(\alpha \mid \boldsymbol{x}).$$

The second sum is equal to zero, by definition of $\hat{\sigma}(\alpha \mid \boldsymbol{x})$. We obtain

$$\frac{d}{d\alpha} L(\alpha, \hat{\sigma}(\alpha \mid \boldsymbol{x}) \mid \boldsymbol{x}) = k\, \Psi_k(\alpha \mid \boldsymbol{x}),$$

with Ψ_k as in (2.4). This is the same expression as Eq. (2.3) in Balakrishnan and Kateri (2008), with their x_i replaced by our x_i^{-1}. Differentiating once more with respect to α, we obtain that

$$\frac{d^2}{d\alpha^2} L(\alpha, \hat{\sigma}(\alpha \mid \boldsymbol{x}) \mid \boldsymbol{x})$$

$$= -\frac{k}{\alpha^2} - k \frac{\sum_{i=1}^{k} x_i^{-\alpha} (\log(x_i))^2 \sum_{i=1}^{k} x_i^{-\alpha} - \left(\sum_{i=1}^{k} x_i^{-\alpha} \log(x_i)\right)^2}{\left(\sum_{i=1}^{k} x_i^{-\alpha}\right)^2}. \quad \text{(A.1)}$$

By the Cauchy–Schwartz inequality, the numerator of the big fraction is nonnegative, whence

$$\frac{d^2}{d\alpha^2} L(\alpha, \hat{\sigma}(\alpha \mid \boldsymbol{x}) \mid \boldsymbol{x}) \leq -\frac{k}{\alpha^2} < 0.$$

Hence, $\alpha \mapsto \Psi_k(\alpha \mid \boldsymbol{x})$ is strictly decreasing. For $\alpha \to 0$, this function diverges to ∞, whereas for $\alpha \to \infty$, this function converges to $\log(\min(x_1, \ldots, x_k)) - k^{-1} \sum_{i=1}^{k} \log(x_i)$, which is less than zero given the assumptions on $x_1, \ldots, x_k$. Hence, there exists a unique $\hat{\alpha}(\boldsymbol{x}) \in (0, \infty)$ such that this function is zero. We conclude that the function $\theta \mapsto L(\theta \mid \boldsymbol{x})$ admits a unique maximum at $\hat{\theta}(\boldsymbol{x})$. $\qquad\square$

Fix $\alpha_0 \in (0, \infty)$. Let P denote the Fréchet distribution with parameter $\theta_0 = (\alpha_0, 1)$, with support $\mathcal{X} = (0, \infty)$. The tentative limit of the functions $\alpha \mapsto \Psi_k(\alpha \mid \boldsymbol{x})$ is the function

$$\Psi(\alpha) = \frac{1}{\alpha} + \frac{\int_0^\infty x^{-\alpha} \log(x) \, dP(x)}{\int_0^\infty x^{-\alpha} \, dP(x)} - \int_0^\infty \log(x) \, dP(x),$$

Let Γ be the gamma function and let $\psi = \Gamma'/\Gamma$ be the digamma function.

Lemma A.1. *Fix $\alpha_0 \in (0, \infty)$. We have*

$$\Psi(\alpha) = \frac{1}{\alpha_0} \big(\psi(1) - \psi(\alpha/\alpha_0) \big), \qquad \alpha \in (0, \infty). \tag{A.2}$$

As a consequence, $\Psi : (0, \infty) \to \mathbb{R}$ is a decreasing bijection with a unique zero at $\alpha = \alpha_0$.

Proof of Lemma A.1. By Lemma B.1,

$$\Psi(\alpha) = \frac{1}{\alpha} + \frac{(-\alpha_0^{-1})\Gamma'(1 + \alpha/\alpha_0)}{\Gamma(1 + \alpha/\alpha_0)} - (-\alpha_0^{-1})\Gamma'(1)$$

$$= \frac{1}{\alpha_0} \big((\alpha/\alpha_0)^{-1} - \psi(1 + \alpha/\alpha_0) + \psi(1). \big)$$

The digamma function satisfies the recurrence relation $\psi(x + 1) = \psi(x) + \frac{1}{x}$. Equation (A.2) follows. The final statement follows from the fact that the digamma function $\psi : (0, \infty) \to \mathbb{R}$ is an increasing bijection. $\qquad\square$

Proof of Theorem 2.3. By Lemma 2.1, we only have to show the claimed convergence. Define a random function Ψ_n on $(0, \infty)$ by

$$\Psi_n(\alpha) = \Psi_{k_n}(\alpha \mid \boldsymbol{X}_n) = \Psi_{k_n}(\alpha \mid \boldsymbol{X}_n/\sigma_n), \tag{A.3}$$

with Ψ_k as in (2.4). Recall Ψ in (A.2). The hypotheses imply that, for each $\alpha \in (0, \infty)$,

$$\Psi_n(\alpha) \rightsquigarrow \Psi(\alpha), \qquad n \to \infty.$$

By Lemma A.1, the limit $\Psi(\alpha)$ is positive, zero, or negative according to whether α is less than, equal to, or greater than α_0. Moreover, the function Ψ_n is decreasing and

$\Psi_n(\hat{\alpha}_n) = 0$; see the proof of Lemma 2.1. Lemma 5.10 in van der Vaart (1998) then implies

$$\hat{\alpha}_n \rightsquigarrow \alpha_0, \qquad n \to \infty.$$

Second, Condition 2.2 also implies that, for each $\alpha \in (0, \infty)$ and as $n \to \infty$,

$$\frac{1}{\sigma_n} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} X_{n,i}^{-\alpha} \right)^{-1/\alpha} = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} (X_{n,i}/\sigma_n)^{-\alpha} \right)^{-1/\alpha}$$

$$\rightsquigarrow \left(\int_0^\infty x^{-\alpha} p_{\alpha_0,1}(x)\,\mathrm{d}x \right)^{-1/\alpha}$$

$$= \left(\Gamma(1 + \alpha/\alpha_0) \right)^{-1/\alpha},$$

where we used Lemma B.1 for the last identity. Both the left-hand and right-hand sides are continuous, nonincreasing functions of α. Since the right-hand side evaluates to unity at $\alpha = \alpha_0$, a standard argument then yields

$$\frac{1}{\sigma_n} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} X_{n,i}^{-\hat{\alpha}_n} \right)^{-1/\hat{\alpha}_n} \rightsquigarrow 1, \qquad n \to \infty. \qquad \square$$

The proof of Theorem 2.5 is decomposed into a sequence of lemmas. Recall Ψ_n and Ψ in (A.3) and (A.2), respectively, and define $\dot{\Psi}_n(\alpha) = (\mathrm{d}/\mathrm{d}\alpha)\Psi_n(\alpha)$ and $\dot{\Psi}(\alpha) = (\mathrm{d}/\mathrm{d}\alpha)\Psi(\alpha)$. By (A.1),

$$\dot{\Psi}_n(\alpha) = -\frac{1}{\alpha^2} - \frac{\mathbb{P}_n[x^{-\alpha}(\log x)^2]\,\mathbb{P}_n x^{-\alpha} - (\mathbb{P}_n x^{-\alpha} \log x)^2}{(\mathbb{P}_n x^{-\alpha})^2}, \qquad (A.4)$$

where $\mathbb{P}_n$ denotes the empirical distribution of the points $(X_{n,i}/\sigma_n)_{i=1}^{k_n}$ and where

$$\mathbb{P}_n f = \frac{1}{k_n} \sum_{i=1}^{k_n} f(X_{n,i}/\sigma_n).$$

The asymptotic distribution of $v_n(\hat{\alpha}_n - \alpha_0)$ can be derived from the asymptotic behavior of $\dot{\Psi}_n$ and $v_n \Psi_n$, which are investigated in the next two lemmas, respectively.

Lemma A.2. (Slope) *Let $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,k_n})$ be a sequence of random vectors in $(0, \infty)^{k_n}$, where $k_n \to \infty$. Suppose that equation (2.9) and Condition 2.4(i) are satisfied. If $\tilde{\alpha}_n$ is a random sequence in $(0, \infty)$ such that $\tilde{\alpha}_n \rightsquigarrow \alpha_0$ as $n \to \infty$, then*

$$\dot{\Psi}_n(\tilde{\alpha}_n) \rightsquigarrow \dot{\Psi}(\alpha_0) = -\frac{\pi^2}{6\alpha_0^2}, \qquad n \to \infty.$$

Proof. For $\alpha \in (0, \infty)$ and $m \in \{0, 1, 2\}$, define

$$f_{m,\alpha}(x) = x^{-\alpha}(\log x)^m, \qquad x \in (0, \infty),$$

19

with $(\log x)^0 = 1$ for all $x \in (0, \infty)$. Suppose that we could show that, for $m \in \{0, 1, 2\}$ and some $\varepsilon > 0$,

$$\sup_{\alpha:|\alpha-\alpha_0|\leq\varepsilon} \left| \mathbb{P}_n f_{m,\alpha} - \int_0^\infty f_{m,\alpha}(x) \, p_{\alpha_0}(x) \, dx \right| \rightsquigarrow 0, \qquad n \to \infty. \qquad (A.5)$$

Then from weak convergence of $\tilde{\alpha}_n$ to α_0, Slutsky's lemma and Lemma B.1 below it would follow that

$$\dot{\Psi}_n(\tilde{\alpha}_n) \rightsquigarrow -\frac{1}{\alpha_0^2} - \frac{\alpha_0^{-2}\Gamma''(2)\,\Gamma(2) - (\alpha_0^{-1}\Gamma'(2))^2}{(\Gamma(2))^2}, \qquad n \to \infty.$$

Since $\Gamma(2) = 1$, $\Gamma'(2) = 1 - \gamma$ and $\Gamma''(2) = (1 - \gamma)^2 + \pi^2/6 - 1$, the conclusion would follow.

It remains to show (A.5). We consider the three cases $m \in \{0, 1, 2\}$ separately. The value of ε does not matter, as long as $\alpha_0 - \varepsilon > 0$.

First, let $m = 0$. The maps $\alpha \mapsto (\mathbb{P}_n f_{0,\alpha})^{1/\alpha}$ and $\alpha \mapsto (\int_0^\infty f_{0,\alpha}\, p_{\alpha_0,1})^{1/\alpha}$ are monotone by Lyapounov's inequality, and the second one is also continuous by Lemma B.1. Pointwise weak convergence then implies uniform weak convergence

$$\sup_{\alpha:|\alpha-\alpha_0|\leq\varepsilon} \left| (\mathbb{P}_n f_{0,\alpha})^{1/\alpha} - \left(\int_0^\infty f_{0,\alpha}(x) \, p_{\alpha_0}(x) \, dx \right)^{1/\alpha} \right| \rightsquigarrow 0, \qquad n \to \infty.$$

Uniform continuity of the map $(y, \alpha) \mapsto y^\alpha$ on compact subsets of $(0, \infty)^2$ then yields (A.5) for $m = 0$.

Second, let $m = 1$. The maps $\alpha \mapsto \mathbb{P}_n f_{1,\alpha}$ and $\alpha \mapsto \int_0^\infty f_{1,\alpha}\, p_{\alpha_0,1}$ are continuous and nonincreasing (their derivatives are nonpositive). Pointwise weak convergence then yields (A.5) for $m = 1$.

Finally, let $m = 2$. With probability tending to one, not all variables $X_{n,i}$ are equal to σ_n, and thus $\mathbb{P}_n(\log x)^2 > 0$. On the latter event, we have

$$\mathbb{P}_n x^{-\alpha}(\log x)^2 = \mathbb{P}_n(\log x)^2 \left\{ \left(\frac{\mathbb{P}_n x^{-\alpha}(\log x)^2}{\mathbb{P}_n(\log x)^2} \right)^{1/\alpha} \right\}^\alpha.$$

By Lyapounov's inequality, the expression in curly braces is nondecreasing in α. For each $\alpha > 0$, it converges weakly to $\{\Gamma''(1 + \alpha/\alpha_0)/\Gamma''(1)\}^{1/\alpha}$, which is nondecreasing and continuous in α; see Lemma B.1. It follows that

$$\sup_{\alpha:|\alpha-\alpha_0|\leq\varepsilon} \left| \left(\frac{\mathbb{P}_n x^{-\alpha}(\log x)^2}{\mathbb{P}_n(\log x)^2} \right)^{1/\alpha} - \left(\frac{\Gamma''(1 + \alpha/\alpha_0)}{\Gamma''(1)} \right)^{1/\alpha} \right| \rightsquigarrow 0, \qquad n \to \infty.$$

Equation (A.5) for $m = 2$ follows. $\qquad \square$

Lemma A.3. *Assume Condition 2.4. Then, as $n \to \infty$,*

$$v_n \, \Psi_n(\alpha_0) = \mathbb{G}_n x^{-\alpha_0} \log(x) + \frac{1-\gamma}{\alpha_0} \mathbb{G}_n x^{-\alpha_0} - \mathbb{G}_n \log(x) + o_p(1). \qquad (A.6)$$

The expression on the right converges weakly to $Y \equiv Y_1 + \frac{1-\gamma}{\alpha_0} Y_2 - Y_3$.

Proof. Recall that

$$\Psi_n(\alpha_0) = \Psi_{k_n}(\alpha_0 \mid \boldsymbol{X}_n/\sigma_n) = \frac{1}{\alpha_0} + \frac{\mathbb{P}_n x^{-\alpha_0} \log(x)}{\mathbb{P}_n x^{-\alpha_0}} - \mathbb{P}_n \log(x).$$

Define $\phi : \mathbb{R} \times (0, \infty) \times \mathbb{R} \to \mathbb{R}$ by

$$\phi(y_1, y_2, y_3) = \frac{1}{\alpha_0} + \frac{y_1}{y_2} - y_3.$$

The previous two displays allow to write

$$\Psi_n(\alpha_0) = \phi\big(\mathbb{P}_n x^{-\alpha_0} \log(x),\ \mathbb{P}_n x^{-\alpha_0},\ \mathbb{P}_n \log(x)\big).$$

Recall Lemma B.1 and put

$$\boldsymbol{y}_0 = \big(-\alpha_0^{-1}\Gamma'(2), \Gamma(2), -\alpha_0^{-1}\Gamma'(1)\big) = \big(\alpha_0^{-1}(\gamma - 1),\ 1,\ \alpha_0^{-1}\gamma\big).$$

As already noted in the proof of Lemma A.1, $\phi(\boldsymbol{y}_0) = \alpha_0^{-1} + \alpha_0^{-1}(\gamma - 1) - \alpha_0^{-1}\gamma = 0$. As a consequence,

$$v_n\, \Psi_n(\alpha_0) = v_n\big(\phi(\mathbb{P}_n x^{-\alpha_0} \log(x),\ \mathbb{P}_n x^{-\alpha_0},\ \mathbb{P}_n \log(x)) - \phi(\boldsymbol{y}_0)\big).$$

In view of Condition 2.4 and the delta method, as $n \to \infty$,

$$v_n\, \Psi_n(\alpha_0) = \dot{\phi}_1(\boldsymbol{y}_0)\, \mathbb{G}_n x^{-\alpha_0} \log(x) + \dot{\phi}_2(\boldsymbol{y}_0)\, \mathbb{G}_n x^{-\alpha_0} + \dot{\phi}_3(\boldsymbol{y}_0)\, \mathbb{G}_n \log(x) + o_p(1),$$

where $\dot{\phi}_j$ denotes the first-order partial derivative of ϕ with respect to y_j for $j \in \{1, 2, 3\}$. Elementary calculations yield

$$\dot{\phi}_1(\boldsymbol{y}_0) = 1, \qquad\qquad \dot{\phi}_2(\boldsymbol{y}_0) = \alpha_0^{-1}(1 - \gamma), \qquad\qquad \dot{\phi}_3(\boldsymbol{y}_0) = -1.$$

The conclusion follows by Slutsky's lemma. $\qquad\square$

Proposition A.4. (Asymptotic expansion for the shape parameter) *Assume that the conditions of Theorem 2.5 hold. Then*

$$v_n\big(\hat{\alpha}_n - \alpha_0\big) = \frac{6\alpha_0^2}{\pi^2}\, v_n\, \Psi_n(\alpha_0) + o_p(1) \rightsquigarrow \frac{6\alpha_0^2}{\pi^2}\, Y, \qquad n \to \infty, \qquad (A.7)$$

where Y is defined in Lemma A.3.

Proof. Recall that, with probability tending to one, $\hat{\alpha}_n$ is the unique zero of the random function $\alpha \mapsto \Psi_n(\alpha)$. Recall that $\dot{\Psi}_n$ in (A.4) is the derivative of Ψ_n. With probability tending to one, we have, by virtue of the mean-value theorem,

$$0 = \Psi_n(\hat{\alpha}_n) = \big(\Psi_n(\hat{\alpha}_n) - \Psi_n(\alpha_0)\big) + \Psi_n(\alpha_0)$$
$$= (\hat{\alpha}_n - \alpha_0)\, \dot{\Psi}_n(\tilde{\alpha}_n) + \Psi_n(\alpha_0);$$

21

here $\tilde{\alpha}_n$ is a convex combination of $\hat{\alpha}_n$ and α_0. Since $\dot{\Psi}_n(\alpha) \leq -1/\alpha^2 < 0$ (argument as in the proof of Lemma 2.1), we can write

$$v_n\left(\hat{\alpha}_n - \alpha_0\right) = -\frac{1}{\dot{\Psi}_n(\tilde{\alpha}_n)}\, v_n\, \Psi_n(\alpha_0).$$

By weak consistency of $\hat{\alpha}_n$, we have $\tilde{\alpha}_n \rightsquigarrow \alpha_0$ as $n \to \infty$. Lemma A.2 then gives $\dot{\Psi}_n(\tilde{\alpha}_n) \rightsquigarrow -\pi^2/(6\alpha_0^2)$ as $n \to \infty$. Apply Lemma A.3 and Slutsky's lemma to conclude. $\qquad\square$

Proof of Theorem 2.5. Combining equations (A.7) and (A.6) yields

$$v_n\left(\hat{\alpha}_n - \alpha_0\right) = \frac{6\alpha_0^2}{\pi^2}\left(\mathbb{G}_n x^{-\alpha_0} \log(x) + \frac{1-\gamma}{\alpha_0} \mathbb{G}_n x^{-\alpha_0} - \mathbb{G}_n \log(x)\right) + o_p(1)$$

as $n \to \infty$. This yields the first row in (2.18).

By definition of $\hat{\sigma}_n$, we have $(\hat{\sigma}_n/\sigma_n)^{-\hat{\alpha}_n} = \mathbb{P}_n x^{-\hat{\alpha}_n}$. Consider the decomposition

$$v_n\left((\hat{\sigma}_n/\sigma_n)^{-\hat{\alpha}_n} - 1\right) = v_n\left(\mathbb{P}_n x^{-\hat{\alpha}_n} - \mathbb{P}_n x^{-\alpha_0}\right) + v_n\left(\mathbb{P}_n x^{-\alpha_0} - 1\right). \qquad (A.8)$$

By the mean value theorem, there exists a random convex combination, $\tilde{\alpha}_n$, of $\hat{\alpha}_n$ and α_0 such that

$$\mathbb{P}_n x^{-\hat{\alpha}_n} - \mathbb{P}_n x^{-\alpha_0} = -(\hat{\alpha}_n - \alpha_0)\, \mathbb{P}_n x^{-\tilde{\alpha}_n} \log(x).$$

By the argument for the case $m = 1$ in the proof of Lemma A.2, we have

$$\mathbb{P}_n x^{-\tilde{\alpha}_n} \log(x) \rightsquigarrow -\frac{1}{\alpha_0}\Gamma'(2) = -\frac{1-\gamma}{\alpha_0}, \qquad n \to \infty.$$

By Proposition A.4 and Lemma A.3, it follows that, as $n \to \infty$,

$$\begin{aligned}
&v_n\left(\mathbb{P}_n x^{-\hat{\alpha}_n} - \mathbb{P}_n x^{-\alpha_0}\right)\\
&= v_n\left(\hat{\alpha}_n - \alpha_0\right)\frac{1-\gamma}{\alpha_0} + o_p(1)\\
&= \frac{6\alpha_0\left(1-\gamma\right)}{\pi^2}\, v_n\, \Psi_n(\alpha_0) + o_p(1)\\
&= \frac{6\alpha_0\left(1-\gamma\right)}{\pi^2}\left(\mathbb{G}_n x^{-\alpha_0} \log(x) + \frac{1-\gamma}{\alpha_0} \mathbb{G}_n x^{-\alpha_0} - \mathbb{G}_n \log(x)\right) + o_p(1).
\end{aligned}$$

This expression in combination with (A.8) yields, as $n \to \infty$,

$$\begin{aligned}
&v_n\left((\hat{\sigma}_n/\sigma_n)^{-\hat{\alpha}_n} - 1\right)\\
&= \frac{6\alpha_0\left(1-\gamma\right)}{\pi^2}\left(\mathbb{G}_n x^{-\alpha_0} \log(x) + \frac{1-\gamma}{\alpha_0} \mathbb{G}_n x^{-\alpha_0} - \mathbb{G}_n \log(x)\right) + \mathbb{G}_n x^{-\alpha_0} + o_p(1).
\end{aligned}$$
$$(A.9)$$

Write $Z_n = (\hat{\sigma}_n/\sigma_n)^{-\hat{\alpha}_n}$, which converges weakly to unity as $n \to \infty$. By the mean value theorem, we have

$$v_n\,(\hat{\sigma}_n/\sigma_n - 1) = v_n\,(Z_n^{-1/\hat{\alpha}_n} - 1) = v_n\,(Z_n - 1)\,(-1/\hat{\alpha}_n)\,\tilde{Z}_n^{-1/\hat{\alpha}_n - 1},$$

where $\tilde{Z}_n$ is a random convex combination of Z_n and 1. But then $\tilde{Z}_n \rightsquigarrow 1$ as $n \to \infty$, whence, by consistency of $\hat{\alpha}_n$ and Slutsky's lemma,

$$v_n\,(\hat{\sigma}_n/\sigma_n - 1) = (-1/\alpha_0)\,v_n\,((\hat{\sigma}_n/\sigma_n)^{-\hat{\alpha}_n} - 1) + o_p(1), \qquad n \to \infty.$$

Combinining this with (A.9), we find

$$v_n\,(\hat{\sigma}_n/\sigma_n - 1) = -\frac{6(1-\gamma)}{\pi^2}\left(\mathbb{G}_n x^{-\alpha_0}\log(x) + \frac{1-\gamma}{\alpha_0}\,\mathbb{G}_n x^{-\alpha_0} - \mathbb{G}_n \log(x)\right)$$
$$- \alpha_0^{-1}\mathbb{G}_n x^{-\alpha_0} + o_p(1)$$

as $n \to \infty$. This is the second row in (2.18). $\qquad\square$

The proof of Addendum 2.6 follows from a tedious but straightforward calculation.

A.2 Proofs for Section 3

Lemma A.5. (Block maxima rarely show ties) *Under Conditions 3.1 and 3.4, for every $c \in (0,\infty)$, we have $\Pr[M_{r_n,1} \vee c = M_{r_n,3} \vee c] \to 0$ as $n \to \infty$.*

Proof of Lemma A.5. By the domain-of-attraction condition combined with the strong mixing property, the sequence of random vectors $((M_{r_n,1} \vee c)/\sigma_n, (M_{r_n,3} \vee c)/\sigma_n)$ converges weakly to the product of two independent Fréchet$(\alpha_0, 1)$ random variables. Apply the Portmanteau lemma – the set $\{(x,y) \in \mathbb{R}^2 : x = y\}$ is closed and has zero probability in the limit. $\qquad\square$

Lemma A.6. (Moments of block maxima converge) *Under Conditions 3.1 and 3.3, we have, for every $c \in (0,\infty)$,*

$$\lim_{n\to\infty} \mathbb{E}[f\big((M_n \vee c)/\sigma_n\big)] = \int_0^\infty f(x)\,p_{\alpha_0,1}(x)\,\mathrm{d}x$$

for every measurable function $f : (0,\infty) \to \mathbb{R}$ which is continuous almost everywhere and for which there exist $\eta_+ \in [0,\alpha_0)$ and $\eta_- \in (-\infty, 0]$ such that $|f(x)| \leq \max(x^{\eta_+}, x^{\eta_-})$ whenever $x \in (0,\infty)$.

Proof of Lemma A.6. Since $c/\sigma_n \to 0$ as $n \to \infty$, the sequence $(M_n \vee c)/\sigma_n$ converges weakly to the Fréchet$(\alpha_0, 1)$ distribution in view of Condition 3.1. By Theorem 2.20 in van der Vaart (1998), we need to show that the sequence $Y_n = f((M_n \vee c)/\sigma_n)$ is asymptotically uniformly integrable, i.e.,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[|Y_n|\,\mathbb{1}_{(K,\infty)}(|Y_n|)\right] = 0. \tag{A.10}$$

23

Split the expectation into two terms, according to whether $Z_n = (M_n \vee c)/\sigma_n$ is larger than unity or not. We obtain

$$
\begin{aligned}
\mathbb{E}\left[|Y_n|\, \mathbb{1}_{(K,\infty)}(|Y_n|)\right] & \\
& \leq \mathbb{E}\left[|Z_n|^{\eta_+}\, \mathbb{1}_{(K,\infty)}(|Z_n|^{\eta_+})\, \mathbb{1}_{[1,\infty)}(Z_n)\right] + \mathbb{E}\left[|Z_n|^{\eta_-}\, \mathbb{1}_{(K,\infty)}(|Z_n|^{\eta_-})\, \mathbb{1}_{(0,1)}(Z_n)\right].
\end{aligned}
$$

Let $\varepsilon > 0$ be small enough such that $(1+\varepsilon)\eta_+ < \alpha_0$. Let $\eta \in \{\eta_-, \eta_+\}$. Since

$$
\mathbb{1}_{(K,\infty)}(|Z_n|^{\eta}) \leq \frac{|Z_n|^{\varepsilon\eta}}{K^{\varepsilon}},
$$

we find

$$
\mathbb{E}\left[|Z_n|^{\eta}\, \mathbb{1}_{(K,\infty)}(|Z_n|^{\eta})\right] \leq \frac{1}{K^{\varepsilon}}\mathbb{E}\left[|Z_n|^{(1+\varepsilon)\eta}\right].
$$

We obtain

$$
\mathbb{E}\left[|Y_n|\, \mathbb{1}_{(K,\infty)}(|Y_n|)\right] \leq \frac{1}{K^{\varepsilon}}\mathbb{E}\left[|Z_n|^{(1+\varepsilon)\eta_+} + |Z_n|^{(1+\varepsilon)\eta_-}\right].
$$

By Condition 3.3, the limsup as $n \to \infty$ is bounded by a multiple of $K^{-\varepsilon}$, which tends to zero as K tends to infinity. Equation (A.10) follows. $\qquad\square$

In order to separate maxima over consecutive blocks by a time lag of at least ℓ, we clip off the final $\ell - 1$ variables within each block:

$$
M_{r,i}^{[\ell]} = \max\{\xi_t : (i-1)r + 1 \leq t \leq ir - \ell + 1\}. \tag{A.11}
$$

Clearly, $M_{r,i} \geq M_{r,i}^{[\ell]}$. The probability that the maximum over a block of size r is attained by any of the final $\ell - 1$ variables should be small; see Lemma A.8 below.

Lemma A.7. (Short blocks are small) *Assume Condition 3.1. If $\ell_n = o(r_n)$ and if $\alpha(\ell_n) = o(\ell_n/r_n)$ as $n \to \infty$, then for all $\varepsilon > 0$,*

$$
\Pr[M_{\ell_n} \geq \varepsilon\sigma_{r_n}] = O(\ell_n/r_n), \qquad n \to \infty. \tag{A.12}
$$

Proof of Lemma A.7. Let F_r be the cumulative distribution function of M_r. By Bücher and Segers (2014, Lemma 7.1), for every $u > 0$,

$$
\Pr[F_{r_n}(M_{\ell_n}) \geq u] = O(\ell_n/r_n), \qquad n \to \infty. \tag{A.13}
$$

Fix $\varepsilon > 0$. By assumption,

$$
\lim_{n\to\infty} F_{r_n}(\varepsilon\sigma_{r_n}) = \exp(-\varepsilon^{-\alpha_0}).
$$

For sufficiently large n, we have

$$
\begin{aligned}
\Pr[M_{\ell_n} \geq \varepsilon\sigma_n] &\leq \Pr[F_{r_n}(M_{\ell_n}) \geq F_{r_n}(\varepsilon\sigma_n)] \\
&\leq \Pr[F_{r_n}(M_{\ell_n}) \geq \exp(-\varepsilon^{-\alpha_0})/2].
\end{aligned}
$$

Set $u = \exp(-\varepsilon^{-\alpha_0})/2$ in (A.13) to arrive at (A.12). $\qquad\square$

Lemma A.8. (Clipping doesn't hurt) *Assume Condition 3.1. If $\ell_n = o(r_n)$ and if $\alpha(\ell_n) = o(\ell_n/r_n)$ as $n \to \infty$, then*

$$\Pr[M_{r_n} > M_{r_n-\ell_n}] \to 0, \qquad n \to \infty. \tag{A.14}$$

Proof of Lemma A.8. Recall Lemma A.7. For every $\varepsilon > 0$ we have, by stationarity,

$$\Pr[M_{r_n} > M_{r_n-\ell_n}] \le \Pr[M_{r_n-\ell_n} \le \varepsilon\sigma_{r_n}] + \Pr[M_{\ell_n} > \varepsilon\sigma_{r_n}].$$

Since $\sigma_{r_n-\ell_n}/\sigma_{r_n} \to 1$ as a consequence of Condition 3.1 and the fact that $\ell_n = o(r_n)$ as $n \to \infty$, the first term converges to $\exp(-\varepsilon^{-\alpha_0})$ as $n \to \infty$, whereas the second one converges to 0 by Lemma A.7. Since $\varepsilon > 0$ was arbitrary, equation (A.14) follows. $\square$

Proof of Theorem 3.6. We apply Theorem 2.5 and Addendum 2.6 to the array $X_{n,i} = M_{r_n,i} \vee c$ and $v_n = \sqrt{k_n}$, where $c \in (0,\infty)$ is arbitrary and $i \in \{1,\ldots,k_n\}$. By Condition 3.2, $\lim_{n\to\infty} \Pr[\forall i = 1,\ldots,k_n : X_{n,i} = M_{r_n,i}] = 1$.

The no-tie property (2.9) has been established in Lemma A.5.

We need to check Condition 2.4, and in particular that the distribution of the random vector $\boldsymbol{Y}$ in (2.16) is $\mathcal{N}_3(B, \Sigma_{\boldsymbol{Y}})$ with B as in the statement of the theorem and $\Sigma_{\boldsymbol{Y}}$ as in (2.19). Essentially, the proof employs the Bernstein big-block-small-block method in combination with the Lindeberg central limit theorem.

Let $\ell_n = \max\{s_n, \lfloor r_n\sqrt{\alpha(s_n)}\rfloor\}$, where $s_n = \lfloor\sqrt{r_n}\rfloor$. Clearly,

$$\ell_n \to \infty, \quad \ell_n = o(r_n) \quad \text{and} \quad \alpha(\ell_n) = o(\ell_n/r_n), \text{ as } n \to \infty. \tag{A.15}$$

Consider the truncated and rescaled block maxima

$$Z_{r,i} = (M_{r_n,i} \vee c)/\sigma_r,$$
$$Z_{r,i}^{[\ell_n]} = (M_{r_n,i}^{[\ell_n]} \vee c)/\sigma_r,$$

with $M_{r,i}^{[\ell_n]}$ as in (A.11). To these random variables correspond the following empirical and population probability measures:

$$\mathbb{P}_n f = \frac{1}{k_n}\sum_{i=1}^{k_n} f(Z_{r_n,i}), \qquad P_n f = \mathbb{E}[f(Z_{r_n,i})],$$

$$\mathbb{P}_n^{[\ell_n]} f = \frac{1}{k_n}\sum_{i=1}^{k_n} f(Z_{r_n,i}^{[\ell_n]}), \qquad P_n^{[\ell_n]} f = \mathbb{E}[f(Z_{r_n,i}^{[\ell_n]})].$$

Abbreviate the tentative limit distribution by $P = \text{Fréchet}(\alpha_0, 1)$. We will also need the following empirical processes:

$$\mathbb{G}_n = \sqrt{k_n}(\mathbb{P}_n - P) \qquad\qquad \text{(uncentered)},$$
$$\tilde{\mathbb{G}}_n = \sqrt{k_n}(\mathbb{P}_n - P_n) \qquad\qquad \text{(centered)},$$
$$\tilde{\mathbb{G}}_n^{[\ell_n]} = \sqrt{k_n}(\mathbb{P}_n^{[\ell_n]} - P_n^{[\ell_n]}) \qquad \text{(centered)}.$$

Finally, the bias arising from the finite block size is quantified by the operator

$$B_n = \sqrt{k_n}(P_n - P).$$

Proof of Condition 2.4(i). Recall the function classes $\mathcal{F}_j$ ($j = 0, 1, 2$) in (2.10), (2.11) and (2.13). For $f \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_2$ not of the form $x^{-\alpha_0}$, $\log(x)$ or $x^{-\alpha_0}\log x$, we just need to show that

$$\mathbb{P}_n f = Pf + o_p(1), \qquad n \to \infty.$$

The domain-of-attraction property (Condition 3.1) and the asymptotic moment bound (Condition 3.3) yield

$$\mathbb{E}[\mathbb{P}_n f] = P_n f \to Pf, \qquad n \to \infty,$$

by uniform integrability, see Lemma A.6. Further,

$$\mathbb{P}_n f - P_n f = \frac{1}{\sqrt{k_n}}\tilde{\mathbb{G}}_n f.$$

Below, see (A.17), we will show that

$$\tilde{\mathbb{G}}_n f = \tilde{\mathbb{G}}_n^{[\ell_n]} f + o_p(1) = O_p(1) + o_p(1) = O_p(1), \qquad n \to \infty. \tag{A.16}$$

It follows that

$$\mathbb{P}_n f = (\mathbb{P}_n f - P_n f) + P_n f = o_p(1) + Pf + o(1) = Pf + o_p(1), \qquad n \to \infty,$$

as required.

Proof of Condition 2.4(ii). We can decompose the empirical process $\mathbb{G}_n$ in a stochastic term and a bias term:

$$\mathbb{G}_n = \sqrt{k_n}(\mathbb{P}_n - P_n) + \sqrt{k_n}(P_n - P)$$
$$= \tilde{\mathbb{G}}_n + B_n.$$

For $f \in \mathcal{H} = \{f_1, f_2, f_3\}$, the bias term $B_n f$ converges to $B(f)$ thanks to Condition 3.5. It remains to treat the stochastic term $\tilde{\mathbb{G}}_n f$, for all $f \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$ [in view of the proof of item (i); see (A.16) above]. We will show that the finite-dimensional distribuitons of $\tilde{\mathbb{G}}_n$ converge to the one of a P-Brownian bridge, $\mathbb{G}$, i.e., a zero-mean, Gaussian stochastic process with covariance function given by

$$\text{cov}(\mathbb{G}f, \mathbb{G}g) = P\big((f - Pf)(g - Pf)\big) = \text{cov}_P\big(f(X), g(X)\big),$$

for $f, g \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$.

Decompose the stochastic term in two parts:

$$\tilde{\mathbb{G}}_n = \tilde{\mathbb{G}}_n^{[\ell_n]} + \Delta_n. \tag{A.17}$$

We will show that Δ_n converges to zero in probability and that the finite-dimensional distributions of $\tilde{\mathbb{G}}_n^{[\ell_n]}$ converge to those of $\mathbb{G}$.

26

First, we treat the main term, $\tilde{\mathbb{G}}_n^{[\ell_n]}$. By the Cramér–Wold device, it suffices to show that $\tilde{\mathbb{G}}_n^{[\ell_n]}g \rightsquigarrow \mathbb{G}g$ as $n \to \infty$, where g is an arbitrary linear combination of functions $f \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$. Define

$$\phi_{ni}(t) = \exp\left[-\mathrm{i}tk_n^{-1/2}\{g(Z_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g\}\right],$$

with i the imaginary unit. Note that the characteristic function of $\tilde{\mathbb{G}}_n^{[\ell_n]}g$ can be written as $t \mapsto \mathbb{E}[\prod_{i=1}^{k_n} \phi_{ni}(t)]$. Successively applying Lemma 3.9 in Dehling and Philipp (2002), we obtain that

$$\left|\mathbb{E}\left[\prod_{i=1}^{k_n} \phi_{ni}(t)\right] - \prod_{i=1}^{k_n} \mathbb{E}[\phi_{ni}(t)]\right| \leq 2\pi k_n \max_{i=1}^{k_n} \alpha\left(\sigma\{\phi_{ni}(t)\}, \sigma\left\{\prod_{j=i+1}^{k_n} \phi_{nj}(t)\right\}\right),$$

where $\alpha(\mathcal{A}_1, \mathcal{A}_2)$ denotes the alpha-mixing coefficient between the sigma-fields $\mathcal{A}_1$ and $\mathcal{A}_2$. Since the maxima $Z_{r,i}^{[\ell_n]}$ over different blocks are based on observations that are at least ℓ_n observations apart, the expression on the right-hand side of the last display is of the order $O(k_n\alpha(\ell_n))$, which converges to 0 as a consequence of Equation (3.2). We can conclude that the weak limit of $\tilde{\mathbb{G}}_n^{[\ell_n]}g$ is the same as the one of

$$\tilde{\mathbb{H}}_n^{[\ell_n]}g = \sqrt{k_n}\left\{\frac{1}{k_n}\sum_{i=1}^{k_n} g(\bar{Z}_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g\right\}$$

where $\bar{Z}_{r_n,i}^{[\ell_n]}$ are independent over $i \in \mathbb{N}$ and have the same distribution as $Z_{r_n,i}^{[\ell_n]}$. By the classical central limit theorem for row wise independent triangular arrays, the weak limit of $\tilde{\mathbb{H}}_n^{[\ell]}g$ is $\mathbb{G}g$: first, its variance

$$\mathrm{Var}(\tilde{\mathbb{H}}_n^{[\ell_n]}g) = P_n^{[\ell_n]}g^2 - (P_n^{[\ell_n]}g)^2$$

converges to $\mathrm{Var}(\mathbb{G}g)$ by Lemma A.6. Note that any linear combination g of functions $f \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$ can be bounded by a multiple of $\max(x^{\eta_+}, x^{\eta_-})$ for some $\eta_+ \in [0, \alpha_0)$ and some $\eta_- \in (-\infty, 0]$. Second, the Lindeberg Condition follows from the Cauchy–Schwarz and Markov inequalities: for any $\varepsilon > 0$,

$$\frac{1}{k_n}\sum_{i=1}^{k_n} \mathbb{E}\left[|g(\bar{Z}_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g|^2 \mathbb{1}(|g(\bar{Z}_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g| > \varepsilon k_n^{1/2})\right]$$

$$\leq \frac{1}{k_n}\sum_{i=1}^{k_n} \mathbb{E}[|g(\bar{Z}_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g|^4]^{1/2} \mathbb{P}(|g(\bar{Z}_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g| > \varepsilon k_n^{1/2})^{1/2}$$

$$\leq \frac{1}{k_n}\sum_{i=1}^{k_n} \mathbb{E}[|g(\bar{Z}_{r_n,i}^{[\ell_n]}) - P_n^{[\ell_n]}g|^4](\varepsilon k_n^{1/2})^{-2},$$

which converges to 0 as $n \to \infty$ again as a consequence of Lemma A.6, as arbitrary powers of g can also be bounded by multiples of $\max(x^{\eta_+}, x^{\eta_-})$, for some $\eta_+ \in [0, \alpha_0)$ and some $\eta_- \in (-\infty, 0]$.

Second, consider the remainder term Δ_n. Since $\tilde{\mathbb{G}}_n f$ and $\tilde{\mathbb{G}}_n^{[\ell_n]} f$ are centered, so is $\Delta_n f$, and

$$\mathbb{E}[(\Delta_n f)^2] = \mathrm{var}(\Delta_n f) = \frac{1}{k_n}\,\mathrm{var}\left(\sum_{i=1}^{k_n} \Delta_{r_n,i}^{[\ell_n]} f\right)$$

where $\Delta_{r,i}^{[\ell_n]} f = f(Z_{r,i}) - f(Z_{r,i}^{[\ell_n]})$. By stationarity and the Cauchy–Schwartz inequality,

$$\mathbb{E}[(\Delta_n f)^2] = \mathrm{var}\left(\Delta_{r_n,1}^{[\ell_n]} f\right) + \frac{2}{k_n} \sum_{h=1}^{k_n-1} (k_n - h)\,\mathrm{cov}\left(\Delta_{r_n,1}^{[\ell_n]} f, \Delta_{r_n,1+h}^{[\ell_n]} f\right)$$

$$\leq 3\,\mathrm{var}\left(\Delta_{r_n,1}^{[\ell_n]} f\right) + 2 \sum_{h=2}^{k_n-1} \left|\mathrm{cov}\left(\Delta_{r_n,1}^{[\ell_n]} f, \Delta_{r_n,1+h}^{[\ell_n]} f\right)\right|. \qquad (A.18)$$

Please note that we left the term $h = 1$ out of the sum; whence the factor three in front of the variance term.

Since $\ell_n = o(r_n)$ as $n \to \infty$ by Condition 3.4, we have $\sigma_{r_n-\ell_n+1}/\sigma_{r_n} \to 1$ as $n \to \infty$ by Condition 3.1. The asymptotic moment bound in Condition 3.3 then ensures that, for every $p > 0$ and every $f \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$, we have, by Lemma A.6,

$$\limsup_{n\to\infty} \mathbb{E}\left[\left|\Delta_{r_n,1}^{[\ell_n]} f\right|^p\right] < \infty. \qquad (A.19)$$

On the event that $M_{r_n,1} = M_{r_n-\ell_n+1}$, we have $\Delta_{r_n,1}^{[\ell_n]} f = 0$. The mixing rate in (A.15) together with Lemma A.8 then imply

$$\Delta_{r_n,1}^{[\ell_n]} f = o_p(1), \qquad n \to \infty.$$

Lyapounov's inequality and the asymptotic moment bound (A.19) then ensure that

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\Delta_{r_n,1}^{[\ell_n]} f\right|^p\right] = 0, \qquad f \in \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2,\ p > 0. \qquad (A.20)$$

Recall Lemma 3.11 in Dehling et al. (2002): for random variables ξ and η and for numbers $p, q \in [1, \infty]$ such that $1/p + 1/q < 1$,

$$|\mathrm{cov}(\xi, \eta)| \leq 10\,\|\xi\|_p\,\|\eta\|_q\,\{\alpha(\sigma(\xi), \sigma(\eta))\}^{1-1/p-1/q}, \qquad (A.21)$$

where $\sigma(\,\cdot\,)$ denotes the σ-field generated by its argument and where $\alpha(\mathcal{A}_1, \mathcal{A}_2)$ denotes the strong mixing coefficient between two σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$. Use inequality (A.21) to bound the covariance terms in (A.18): for every $p > 2$,

$$\mathbb{E}[(\Delta_n f)^2] \leq 3\,\|\Delta_{r_n,1}^{[\ell_n]} f\|_2^2 + 20\,k_n\,\|\Delta_{r_n,1}^{[\ell_n]} f\|_p^2\,\{\alpha(r_n)\}^{1-2/p}.$$

Recall $\omega > 0$ in Condition 3.4. In view of (A.20), a sufficient condition for the right-hand side to converge to zero is that p is big enough: $1/(1 - 2/p)$ should be smaller than $1 + \omega$. $\qquad \square$

28

A.3 Proof of Theorem 4.2

Proof of Theorem 4.2. We apply Theorem 3.6. To this end, we verify its conditions.

Proof of Condition 3.1. The second-order regular variation condition (4.5) implies the first-order one in (4.2), which is in turn equivalent to weak convergence of partial maxima as in (4.1). Condition 3.1 follows with scaling sequence $\sigma_n = a_n$. The latter sequence is regularly varying (Resnick, 1987, Proposition 1.11) with index $1/\alpha_0$, which implies that $\lim_{n\to\infty} a_{m_n}/a_n = 1$ whenever $\lim_{n\to\infty} m_n/n = 1$.

Proof of Condition 3.2. For any real c we have, since $\log F(c) < 0$ and since $\log(k_n) = o(r_n)$ by (4.9),

$$\Pr[\min(M_{r_n,1}, \ldots, M_{r_n,k_n}) \leq c]$$
$$\leq k_n \, F^{r_n}(c) = \exp\{\log(k_n) + r_n \log F(c)\} \to 0, \qquad n \to \infty.$$

Proof of Condition 3.3. This is Lemma A.9 below.

Proof of Condition 3.4. Trivial, since $\alpha(\ell) = 0$ for integer $\ell \geq 1$.

Proof of Condition 3.5. Recall Remark 4.3 and therein the functions L and $g(u) = A(u)L(u)$. We begin by collecting some non-asymptotic bounds on the function L. Fix $\delta \in (0, \alpha_0)$. Potter's theorem (Bingham et al., 1987, Theorem 1.5.6) implies that there exists some constant $x'(\delta) > 0$ such that, for all $u \geq x'(\delta), x \geq x'(\delta)/u$,

$$\frac{L(u)}{L(ux)} \leq (1+\delta) \, \max(x^{-\delta}, x^{\delta}). \tag{A.22}$$

As a consequence of Theorem B.2.18 in de Haan and Ferreira (2006), accredited to Drees (1998), there exists some further constant $x''(\delta) > 0$ such that, for all $u \geq x''(\delta), x \geq x''(\delta)/u$,

$$\left| \frac{L(ux) - L(u)}{g(u)} \right| \leq \; c(\delta) \, \max(x^{\rho - \delta}, x^{\rho + \delta}), \tag{A.23}$$

for some constant $c(\delta) > 0$. Define $x(\delta) = \max\{x'(\delta), x''(\delta), 1\}$.

We are going to show Condition 3.5 for $c = x(\delta)$ and $\sigma_{r_n} = a_{r_n}$. For $i = 1, \ldots, k_n$, define $X_{n,i} = M_{r_n,i} \vee x(\delta)$. Let P_n denote the common distribution of the rescaled, truncated block maxima $X_{n,i}/a_{r_n}$ and let P denote the Fréchet$(\alpha_0, 1)$ distribution. Write $B_n = \sqrt{k_n}(P_n - P)$ and define the three-by-one vector $\boldsymbol{\beta}$ by

$$\boldsymbol{\beta} = \frac{\lambda}{|\rho|\,\alpha_0} \begin{pmatrix} \Gamma(2 + \frac{|\rho|}{\alpha_0}) + \Gamma'(2 + \frac{|\rho|}{\alpha_0}) - 2 + \gamma \\ \alpha_0 - \alpha_0 \Gamma(2 + \frac{|\rho|}{\alpha_0}) \\ \Gamma(1 + \frac{|\rho|}{\alpha_0}) - 1 \end{pmatrix} \tag{A.24}$$

if $\rho < 0$ and by

$$\boldsymbol{\beta} = \frac{\lambda}{\alpha_0^2} \begin{pmatrix} (1-\gamma)^2 + \pi^2/6 - \gamma \\ \alpha_0(\gamma - 1) \\ -\gamma \end{pmatrix}$$

if $\rho = 0$. We will show that

$$\lim_{n\to\infty} \left(B_n x^{-\alpha_0} \log x, \, B_n x^{-\alpha_0}, \, B_n \log x\right)^T = \boldsymbol{\beta}. \tag{A.25}$$

Elementary calculations yield that $M(\alpha_0)\,\boldsymbol{\beta} = B(\alpha_0, \rho, \lambda)$ as required in (4.8).

Equation (A.25) can be shown coordinatewise. We begin by some generalities. For any $f \in \mathcal{H}$ as in (2.14), we can write, for arbitrary $x, x_0 \in (0, \infty)$,

$$f(x) = \begin{cases} f(x_0) - \int_x^{x_0} f'(y)\,dy, & \text{if } 0 < x \le x_0, \\ f(x_0) + \int_{x_0}^x f'(y)\,dy, & \text{if } x_0 < x < \infty. \end{cases}$$

By Fubini's theorem, with G_n and G denoting the cdf-s of P_n and P, respectively,

$$\begin{aligned}
Pf &= \int_{(0,x_0]} f(x)\,dP(x) + \int_{(x_0,\infty)} f(x)\,dP(x) \\
&= f(x_0) - \int_{x\in(0,x_0]} \int_{y=x}^{x_0} f'(y)\,dy\,dP(x) + \int_{x\in(x_0,\infty)} \int_{y=x_0}^x f'(y)\,dy\,dP(x) \\
&= f(x_0) - \int_{y=0}^{x_0} \int_{x\in(0,y]} dP(x)\,f'(y)\,dy + \int_{y=x_0}^{\infty} \int_{x\in(y,\infty)} dP(x)\,f'(y)\,dy \\
&= f(x_0) - \int_0^{x_0} G(y)\,f'(y)\,dy + \int_{x_0}^{\infty} \{1 - G(y)\}\,f'(y)\,dy,
\end{aligned}$$

and the same formula holds with P and G replaced by P_n and G_n, respectively. We find that

$$B_n f = \sqrt{k_n}(P_n - P)f = -\int_0^{\infty} \sqrt{k_n}\,\{G_n(y) - G(y)\}\,f'(y)\,dy.$$

Note that

$$G(y) = \exp(-y^{-\alpha_0})\,\mathbb{1}_{(0,\infty)}(y), \qquad G_n(y) = F^{r_n}(a_{r_n}y)\,\mathbb{1}_{[x(\delta)/a_{r_n},\infty)}(y),$$

From the definition of L in (4.12), we can write, for $y \ge x(\delta)/a_{r_n}$,

$$G_n(y) = \exp\left(-y^{-\alpha_0} r_n\{-\log F(a_{r_n})\}\,\frac{L(a_{r_n}y)}{L(a_{r_n})}\right).$$

For the sake of brevity, we will only carry out the subsequent parts of the proof in the case where F is ultimately continuous, so that $r_n\{-\log F(a_{r_n})\} = 1$ for all sufficiently large n. In that case, $B_n f = J_{n1}(f) + J_{n2}(f)$ where

$$J_{n1}(f) = \sqrt{k_n} \int_0^{x(\delta)/a_{r_n}} \exp(-y^{-\alpha_0}) f'(y)\,dy,$$

$$J_{n2}(f) = -\sqrt{k_n} \int_{x(\delta)/a_{r_n}}^{\infty} \left[\exp\left(-y^{-\alpha_0}\frac{L(a_{r_n}y)}{L(a_{r_n})}\right) - \exp(y^{-\alpha_0})\right] f'(y)\,dy,$$

Let us first show that $J_{n1}(f)$ converges to 0 for any $f \in \mathcal{H}$. For that purpose, note that any $f \in \mathcal{H}$ satisfies $|f'(x)| \leq K x^{-\alpha_0 - \varepsilon - 1}$ for any $\varepsilon < 1$ and for some constant $K = K(\varepsilon) > 0$. As a consequence, by (4.10), for sufficiently large n,

$$\max_{f \in \mathcal{H}} |J_{n1}(f)| \leq \{\lambda + o(1)\} \frac{K}{A(a_{r_n})} \int_0^{x(\delta)/a_{r_n}} \exp(-y^{-\alpha_0}) y^{-\alpha_0 - \varepsilon - 1} \, dy.$$

Since $A(x)$ is bounded from below by a multiple of $x^{\rho - \varepsilon}$ for sufficiently large x (by Remark 4.3 and Potter's theorem), the expression on the right-hand side of the last display can be easily seen to converge to 0 for $n \to \infty$.

For the treatment of J_{n2}, note that

$$J(f, \rho) \equiv \int_0^\infty h_\rho(y) \exp\left(-y^{-\alpha_0}\right) y^{-\alpha_0} f'(y) \, dy$$

$$= \begin{cases} \int_0^\infty h_\rho(y) \exp\left(-y^{-\alpha_0}\right) y^{-2\alpha_0 - 1} (1 - \alpha_0 \log y) \, dy & , f(y) = y^{-\alpha_0} \log y \\ \int_0^\infty h_\rho(y) \exp\left(-y^{-\alpha_0}\right) (-\alpha_0 y^{-2\alpha_0 - 1}) \, dy & , f(y) = y^{-\alpha_0} \\ \int_0^\infty h_\rho(y) \exp\left(-y^{-\alpha_0}\right) y^{-\alpha_0 - 1} \, dy & , f(y) = \log y \end{cases}$$

$$= \begin{cases} \mathbb{E}[h_\rho(Y) Y^{-\alpha_0} (\alpha_0^{-1} - \log Y)] & , f(y) = y^{-\alpha_0} \log y \\ -\mathbb{E}[h_\rho(Y) Y^{-\alpha_0}] & , f(y) = y^{-\alpha_0} \\ \alpha_0^{-1} \mathbb{E}[h_\rho(Y)] & , f(y) = \log y, \end{cases}$$

where Y denotes a Fréchet$(\alpha_0, 1)$ random variable. By Lemma B.1 this implies

$$J(x^{-\alpha_0} \log x, \rho) = \frac{1}{\rho \alpha_0} \left\{ \Gamma(2 + \tfrac{|\rho|}{\alpha_0}) + \Gamma'(2 + \tfrac{|\rho|}{\alpha_0}) - 1 - \Gamma'(2) \right\}$$

$$= \frac{1}{|\rho| \alpha_0} \left\{ 2 - \gamma - \Gamma(2 + \tfrac{|\rho|}{\alpha_0}) - \Gamma'(2 + \tfrac{|\rho|}{\alpha_0}) \right\},$$

$$J(x^{-\alpha_0}, \rho) = \frac{1}{\rho} \left\{ \Gamma(2) - \Gamma(2 + \tfrac{|\rho|}{\alpha_0}) \right\} = \frac{1}{|\rho|} \left\{ \Gamma(2 + \tfrac{|\rho|}{\alpha_0}) - 1 \right\},$$

$$J(\log x, \rho) = \frac{1}{\rho \alpha_0} \left\{ \Gamma(1 + \tfrac{|\rho|}{\alpha_0}) - 1 \right\} = \frac{1}{|\rho| \alpha_0} \left\{ 1 - \Gamma(1 + \tfrac{|\rho|}{\alpha_0}) \right\}$$

for $\rho < 0$ and

$$J(x^{-\alpha_0} \log x, 0) = -\frac{1}{\alpha_0^2} \left\{ \Gamma'(2) + \Gamma''(2) \right\} = \frac{1}{\alpha_0^2} \left\{ \gamma - (1 - \gamma)^2 - \pi^2/6 \right\},$$

$$J(x^{-\alpha_0}, 0) = \frac{1}{\alpha_0} \Gamma'(2) = \frac{1 - \gamma}{\alpha_0},$$

$$J(\log x, 0) = -\frac{1}{\alpha_0^2} \Gamma'(1) = \frac{\gamma}{\alpha_0^2}.$$

Hence, $\boldsymbol{\beta} = -\lambda \big(J(x^{-\alpha_0} \log x, \rho), J(x^{-\alpha_0}, \rho), J(\log x, \rho) \big)^T$ and it is therefore sufficient to show that, for any $f \in \mathcal{H}$,

$$J_{n2}(f) \to -\lambda J(f, \rho) \tag{A.26}$$

31

as $n \to \infty$. By the mean value theorem, we can write $J_{n2}(f)$ as

$$J_{n2}(f) = -\sqrt{k_n} A(a_{r_n}) \int_{x(\delta)/a_{r_n}}^{\infty} \frac{L(a_{r_n}y) - L(a_{r_n})}{A(a_{r_n})L(a_{r_n})} \exp\left(-y^{-\alpha_0}\xi_n(y)\right) y^{-\alpha_0} f'(y)\, dy$$

for some $\xi_n(y)$ between $L(a_{r_n}y)/L(a_{r_n})$ and 1. For $n \to \infty$, the factor in front of this integral converges to $-\lambda$ by assumption (4.10), while the integrand in this integral converges to

$$h_\rho(y) \exp\left(-y^{-\alpha_0}\right) y^{-\alpha_0} f'(y),$$

pointwise in $y \in (0, \infty)$, by Condition 4.1. Hence, the convergence in (A.26) follows from dominated convergence if we show that

$$f_n(y) = \mathbb{1}\left(y > \frac{x(\delta)}{a_{r_n}}\right) \left| \frac{L(a_{r_n}y) - L(a_{r_n})}{A(a_{r_n})L(a_{r_n})} \right| \exp\left(-y^{-\alpha_0}\xi_n(y)\right) y^{-\alpha_0} f'(y)$$

can be bounded by an integrable function on $(0, \infty)$. We split the proof into two cases.

First, for any $1 \geq y \geq x(\delta)/a_{r_n}$,

$$\left| \frac{L(a_{r_n}y) - L(a_{r_n})}{A(a_{r_n})L(a_{r_n})} \right| \leq c(\delta)y^{\rho-\delta}$$

from (A.23) and

$$\xi_n(y) \geq \min\left(1, \frac{L(a_{r_n}y)}{L(a_{r_n})}\right) \geq (1+\delta)^{-1}y^{\delta}$$

from (A.22). Moreover, for any $f \in \mathcal{H}$, the function $f'(y)$ is bounded by a multiple of $y^{-\alpha_0-\delta-1}$ for $y \leq 1$. Therefore, for any $y \in (0,1)$,

$$f_n(y) \leq c'(\delta) \exp\{-(1+\delta)^{-1}y^{-\alpha_0+\delta}\}y^{-2\alpha_0-2\delta-1+\rho}$$

and the function on the right is integrable on $(0, 1)$ since $\delta < \alpha_0$.

Second, for $y \in [1, \infty)$, we have

$$\left| \frac{L(a_{r_n}y) - L(a_{r_n})}{A(a_{r_n})L(a_{r_n})} \right| \leq c(\delta)y^{\rho+\delta}$$

from (A.23) and

$$\xi_n(y) \geq \min\left(1, \frac{L(a_{r_n}y)}{L(a_{r_n})}\right) \geq (1+\delta)^{-1}y^{-\delta}$$

from (A.22). Moreover, $f'(y)$ is bounded by a multiple of y^{-1} for any $y \geq 1$ and any $f \in \mathcal{H}$. Therefore,

$$f_n(y) \leq c''(\delta) y^{-\alpha_0-1+\rho+\delta}$$

which is easily integrable on $[1, \infty)$. $\qquad \square$

32

Lemma A.9. *Let $\xi_1, \xi_2, \ldots$ be independent random variables with common distribution function F satisfying (4.2). Let $M_n = \max(\xi_1, \ldots, \xi_n)$. For every $\beta \in (-\infty, \alpha_0)$ and any constant $c > 0$, we have*

$$\limsup_{n \to \infty} \mathbb{E}\big[((M_n \vee c)/a_n)^\beta\big] < \infty.$$

Proof of Lemma A.9. Since the case $\beta = 0$ is trivial, there are two cases to be considered: $\beta \in (-\infty, 0)$ and $\beta \in (0, \alpha_0)$. Write $Z_n = (M_n \vee c)/a_n$ and note that

$$\Pr[Z_n < y] = \Pr[M_n \vee c < a_n y] = F^n(a_n y)\, \mathbb{1}_{(c/a_n, \infty)}(y).$$

Case $\beta \in (-\infty, 0)$. We have

$$
\begin{aligned}
\mathbb{E}[Z_n^\beta] &= \int_0^\infty \Pr[Z_n^\beta > x]\, dx \\
&= \int_0^\infty \Pr[Z_n < x^{1/\beta}]\, dx \\
&= \int_0^\infty \Pr[Z_n < y]\, |\beta|\, y^{\beta - 1}\, dy \\
&= \int_{c/a_n}^\infty F^n(a_n y)\, |\beta|\, y^{\beta - 1}\, dy.
\end{aligned}
$$

We split the integration domain in two pieces. For $y \in (1, \infty)$, the integrand is bounded by $|\beta|\, y^{\beta - 1}$, which integrates to unity. Hence we only need to consider the integral over $y \in (c/a_n, 1]$. We have

$$F^n(a_n y) = \exp\{n \log F(a_n y)\} = \exp\left(-n\{-\log F(a_n)\} \frac{-\log F(a_n y)}{-\log F(a_n)}\right).$$

Fix $\delta \in (0, \alpha_0)$. By (4.3), we have $n\{-\log F(a_n)\} \geq 1 - \delta$ for all n larger than some $n(\delta)$. By Potter's theorem (Bingham et al., 1987, Theorem 1.5.6), there exists $x(\delta) > 0$ such that, for all n such that $a_n \geq x(\delta)$ and for all $y \in (x(\delta)/a_n, 1]$,

$$\frac{-\log F(a_n)}{-\log F(a_n y)} \leq (1 + \delta)\, y^{\alpha_0 - \delta}.$$

Without loss of generality, assume $x(\delta) > c$. For all $y \in (c/a_n, x(\delta)/a_n]$, we have

$$
\begin{aligned}
\frac{-\log F(a_n)}{-\log F(a_n y)} &\leq \frac{-\log F(a_n)}{-\log F(x(\delta))} \\
&\leq (1 + \delta)\, (x(\delta)/a_n)^{\alpha_0 - \delta} \\
&\leq (1 + \delta)\, x(\delta)^{\alpha_0 - \delta} c^{\delta - \alpha_0}\, y^{\alpha_0 - \delta}.
\end{aligned}
$$

Combining the previous two displays, we see that there exists a constant $c(\delta) > 0$ such that, for all $y \in (c/a_n, 1]$,

$$\frac{-\log F(a_n y)}{-\log F(a_n)} \geq c(\delta)\, y^{-\alpha_0 + \delta}.$$

33

We conclude that, for all sufficiently large n and all $y \in (c/a_n, 1]$,

$$F^n(a_n y) \leq \exp\left(-c(\delta)\, y^{-\alpha_0+\delta}\right),$$

where $c(\delta)$ is a positive constant, possibly different from the one in the previous equation. For such n, we have

$$\int_{c/a_n}^{1} F^n(a_n y)\, |\beta|\, y^{\beta-1}\, dy \leq \int_0^1 \exp\left(-c(\delta)\, y^{-\alpha_0+\delta}\right)\, |\beta|\, y^{\beta-1}\, dy < \infty.$$

Case $\beta \in (0, \alpha_0)$. Let $\delta > 0$ be sufficiently small such that $\beta + \delta < \alpha$. Let $x(\delta) > 0$ be as in Potter's theorem. Let $n(\delta)$ be sufficiently large such that $a_n \geq x(\delta) \vee c$ for all $n \geq n(\delta)$. Put $K = \sup_{n \geq 1} n\{1 - F(a_n)\}$, which is finite by (4.3) and the fact that $-\log F(x) \sim 1 - F(x)$ for $x \to \infty$. For $n \geq n(\delta)$, we have

$$\begin{aligned}
\mathbb{E}[Z_n^\beta] &= \int_0^\infty \Pr[Z_n > x^{1/\beta}]\, dx \\
&= \int_0^\infty \Pr[M_n \vee c > a_n x^{1/\beta}]\, dx \\
&\leq 1 + \int_1^\infty \Pr[M_n > a_n x^{1/\beta}]\, dx \\
&\leq 1 + \int_1^\infty n\{1 - F(a_n x^{1/\beta})\}\, dx \\
&\leq 1 + K \int_1^\infty \frac{1 - F(a_n x^{1/\beta})}{1 - F(a_n)}\, dx.
\end{aligned}$$

By Potter's theorem, the integral on the last line is bounded by

$$(1+\delta) \int_c^\infty (x^{1/\beta})^{-\alpha_0+\delta}\, dx.$$

The latter integral is finite, since $(-\alpha_0 + \delta)/\beta < -1$. $\qquad\square$

B   Auxiliary results

Let $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$ be the gamma function and let Γ' and Γ'' be its first and second derivative, respectively. All proofs for this section are given in Section D in the supplementary material.

Lemma B.1. (Moments) *Let P denote the Fréchet distribution with parameter vector* $(\alpha_0, 1)$, *for some $\alpha_0 \in (0, \infty)$. For all $\alpha \in (-\alpha_0, \infty)$,*

$$\int_0^\infty x^{-\alpha}\, dP(x) = \Gamma(1 + \alpha/\alpha_0),$$

$$\int_0^\infty x^{-\alpha} \log(x)\, dP(x) = -\frac{1}{\alpha_0} \Gamma'(1 + \alpha/\alpha_0),$$

$$\int_0^\infty x^{-\alpha} (\log(x))^2\, dP(x) = \frac{1}{\alpha_0^2} \Gamma''(1 + \alpha/\alpha_0).$$

Lemma B.2. (Covariance matrix) *Let X be a random variable whose distribution is Fréchet with parameter vector $(\alpha_0, 1)$. The covariance matrix of the random vector*

$$\boldsymbol{Y} = (Y_1, Y_2, Y_3)^T = \left(X^{-\alpha_0} \log(X), \ X^{-\alpha_0}, \ \log(X) \right)^T$$

is equal to

$$\mathrm{cov}(\boldsymbol{Y}) = \frac{1}{\alpha_0^2} \begin{pmatrix} 1 - 4\gamma + \gamma^2 + \pi^2/3 & \alpha_0(\gamma - 2) & \pi^2/6 - \gamma \\ \alpha_0(\gamma - 2) & \alpha_0^2 & -\alpha_0 \\ \pi^2/6 - \gamma & -\alpha_0 & \pi^2/6 \end{pmatrix}.$$

Lemma B.3. (Fisher information) *Let P_θ denote the Fréchet distribution with parameter $\theta = (\alpha, \sigma) \in (0, \infty)^2$. The Fisher information $I_\theta = P_\theta(\dot{\ell}_\theta \dot{\ell}_\theta^T)$ is given by*

$$I_\theta = \begin{pmatrix} \iota_{11} & \iota_{12} \\ \iota_{21} & \iota_{22} \end{pmatrix} = \begin{pmatrix} \{(1-\gamma)^2 + \pi^2/6\}/\alpha^2 & (1-\gamma)/\sigma \\ (1-\gamma)/\sigma & \alpha^2/\sigma^2 \end{pmatrix}.$$

Its inverse is given by

$$I_\theta^{-1} = \frac{6}{\pi^2} \begin{pmatrix} \alpha^2 & (\gamma - 1)\sigma \\ (\gamma - 1)\sigma & (\sigma/\alpha)^2 \{(1-\gamma)^2 + \pi^2/6\} \end{pmatrix}.$$

Acknowledgments

References

Balakrishnan, N. and M. Kateri (2008). On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data. *Statistics & Probability Letters 78*, 2971–2975.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.

Bücher, A. and J. Segers (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes 17*, 495–528.

Cai, J.-J., L. de Haan, and C. Zhou (2013). Bias correction in extreme value statistics with index around zero. *Extremes 16*(2), 173–201.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York. An introduction.

Dehling, H., T. Mikosch, and M. Søorensen (2002). *Empirical Process Techniques for Dependent Data.* Birkhäuser.

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.

Dombry, C. (2013). Maximum likelihood estimators for the extreme value index based on the block maxima method. *Bernoulli (to appear).* arXiv:1301.5611.

Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist. 25*(1), 187–210.

Drees, H. (2000, 11). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab. 10*(4), 1274–1301.

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.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2) 44*, 423–453.

Gumbel, E. J. (1958). *Statistics of extremes.* New York: Columbia University Press.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics 27*(3), 251–261.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete 65*(2), 291–306.

Marohn, F. (1994). On testing the Exponential and Gumbel distribution. In *Extreme Value Theory and Applications*, pp. 159–174. Kluwer Academic Publishers.

Mikosch, T. and C. Stărică (2000). Limit theory for the sample autocorrelations and extremes of a GARCH $(1,1)$ process. *Ann. Statist. 28*(5), 1427–1451.

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters 38*(2), 107 – 115.

Pickands, J. (1975, 01). Statistical inference using extreme order statistics. *Ann. Statist. 3*(1), 119–131.

Prescott, P. and A. T. Walden (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika 67*(3), 723–724.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, Volume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York.

Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications 119*(2), 468 – 490.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika 72*(1), 67–90.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge: Cambridge University Press.

<center>Supplementary Material on</center>

<center># "Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series"</center>

<center>Axel Bücher and Johan Segers</center>

<center>November 24, 2015</center>

<center>**Abstract**</center>

<center>This supplementary material contains the proofs of Lemma 5.1 (Section C) and of the auxiliary lemmas from Section B (Section D) from the main paper.</center>

C Proofs for Section 5

Proof of Lemma 5.1. We only give a sketch proof for the case $p = 2$, the general case being similar, but notationally more involved. Set $b_1 = b, b_2 = 1 - b$, such that $b_{(2)} = b \vee (1 - b)$. Clearly,

$$\mathbb{P}(M_n \leq x) = \mathbb{P}\{Z_0 \leq x(1-b)^{-1}, Z_1 \leq xb_{(2)}^{-1}, \dots, Z_{n-1} \leq xb_{(2)}^{-1}, Z_n \leq xb^{-1}\}$$
$$= F\{x(1-b)^{-1}\} \cdot F\{xb^{-1}\} \cdot F^{n-1}(xb_{(2)}^{-1}).$$

As a consequence, with $b_{(1)} = b \wedge (1 - b)$,

$$H_n(x) = \mathbb{P}(M_n \leq xb_{(2)}a_n)$$
$$= F(a_n x \tfrac{b_{(2)}}{1-b}) \cdot F(a_n x \tfrac{b_{(2)}}{b}) \cdot F^{n-1}(a_n x)$$
$$= F(a_n x \tfrac{b_{(2)}}{b_{(1)}}) \cdot F^n(a_n x).$$

Since, by assumption, $F^n(xa_n) \to \exp(-x^{-\alpha_0})$, Condition 3.1 is satisfied.

The proof of Condition 3.3 can be be carried out along the lines of the proof of Lemma A.9. For $\beta < 0$, simply use that

$$\mathbb{P}\{(M_n \vee c)/\sigma_n \leq x\} = H_n(x)\mathbb{1}(x \geq c/\sigma_n) \leq F^n(xa_n) \cdot \mathbb{1}(x \geq c/\sigma_n),$$

while, for $\beta > 0$

$$\mathbb{P}(M_n > \sigma_n x^{1/\beta}) \leq 2n \cdot \mathbb{P}(Z_1 > \sigma_n x^{1/\beta}b_{(2)}) = 2n\{1 - F(a_n x^{1/\beta})\}$$

for any $x > 1$.

Condition 3.4 is trivial, since the process is p-dependent.

Since $\log k_n = o(r_n)$, Condition 3.2 follows from

$$\mathbb{P}[\min(M_{r_n,1}, \ldots, M_{r_n,k_n}) \leq c]$$
$$\leq k_n \mathbb{P}(M_{r_n} \leq c)$$
$$= \exp\{\log k_n + (r_n - 1)\log F(cb_{(2)}^{-1})\} \cdot F\{c(1-b)^{-1}\} \cdot F\{cb^{-1}\}.$$

Finally, consider Condition 3.5. As in the proof of Theorem 4.2, write

$$\sqrt{k_n}\left(\mathbb{E}\left[f\left((M_{r_n} \vee c)/\sigma_{r_n}\right)\right] - Pf\right) = -\int_0^\infty \sqrt{k_n}\{\tilde{H}_n(y) - G(y)\}f'(y)\,dy,$$

where $G(y) = \exp(-y^{-\alpha_0})$ and where

$$\tilde{H}_n(y) = \mathbb{P}\{(M_{r_n} \vee c)/\sigma_{r_n} \leq y\} = A_n(y)G_n(y)$$

with

$$A_n(y) = F(a_{r_n}y\tfrac{b_{(2)}}{b_{(1)}}), \qquad G_n(y) = F^{r_n}(ya_{r_n})\mathbb{1}(y \geq c/\sigma_{r_n})$$

Write

$$\int_0^\infty \sqrt{k_n}\{\tilde{H}_n(y) - G(y)\}f'(y)\,dy = -\int_0^{c/\sigma_{r_n}} \sqrt{k_n}G(y)f'(y)\,dy$$
$$+ \int_{c/\sigma_{r_n}}^\infty \sqrt{k_n}A_n(y)\{G_n(y) - G(y)\}f'(y)\,dy$$
$$+ \int_{c/\sigma_{r_n}}^\infty \sqrt{k_n}\{A_n(y) - 1\}G(y)f'(y)\,dy \quad \text{(C.1)}$$

The first integral converges to 0 as shown in the proof of Theorem 4.2, treatment of $J_{n1}(f)$. The integrand of the second integral converges pointwise to the same limit as in the iid case. The integrand can further be bounded by an integrable function as shown in the treatment of J_{n2} in the proof of Theorem 4.2, after splitting the integration domain at 1. Hence, the limit of that integral is the same as in the iid case by dominated convergence.

Consider the last integral in the latter display. Decompose

$$\sqrt{k_n}|A_n(y) - 1| = \frac{\sqrt{k_n}}{r_n} \cdot \frac{1 - A_n(y)}{-\log A_n(y)} \cdot \frac{-\log A_n(y)}{-\log F(a_{r_n})},$$

where we used the fact that $r_n\{-\log F(a_{r_n})\} = 1$. The second factor is bounded by 1, since $\log(x) \leq x - 1$ for all $x > 0$. Consider the third factor. With $L(x) = -\log\{F(x)\}x^{\alpha_0}$, we have

$$\frac{-\log A_n(y)}{-\log F(a_{r_n})} = (yb_{(2)}/b_{(1)})^{-\alpha_0}\frac{L(a_{r_n}yb_{(2)}/b_{(1)})}{L(a_{r_n})}.$$

The fraction on the right-hand side is bounded by a multiple of $y^\delta \vee y^{-\delta}$ by Potter's theorem, for some $0 < \delta < \alpha_0$. Further note that, up to a factor, $f'(y) \leq y^{-\alpha_0-\delta-1}$ for

38

$y \leq 1$ and $f'(y) \leq y^{-1}$ for $y > 1$. We obtain that the integrand of the third integral on the right-hand side of (C.1) is bounded by a multiple of

$$\sqrt{k_n}/r_n \cdot \exp(-y^{-\alpha_0})y^{-2\alpha_0 - 2\delta - 1}$$

for $y \leq 1$ and by a multiple of

$$\sqrt{k_n}/r_n \cdot y^{-\alpha_0 - 1 + \delta}$$

for $y > 1$. Both functions are integrable on its respective domains. Since $k_n = o(n^{2/3})$ is equivalent to $\sqrt{k_n} = o(r_n)$, the third integral converges to 0. Hence, Condition 3.5 is satisfied. $\qquad\square$

D   Proofs for Section B

Proof of Lemma B.1. If Y is a unit exponential random variable, then the law of Y^{-1/α_0} is equal to P. The integrals stated in the lemma are equal to $\mathbb{E}[Y^{\alpha/\alpha_0}]$, $-\alpha_0^{-1}\,\mathbb{E}[Y^{\alpha/\alpha_0}\log(Y)]$, and $\alpha_0^{-2}\,\mathbb{E}[Y^{\alpha/\alpha_0}(\log Y)^2]$, respectively. First,

$$\int_0^\infty x^{-\alpha}\,\mathrm{d}P(x) = \int_0^\infty y^{\alpha/\alpha_0}\,\exp(-y)\,\mathrm{d}y = \Gamma(1 + \alpha/\alpha_0).$$

Second,

$$\int_0^\infty x^{-\alpha}\log(x)\,\mathrm{d}P(x) = -\frac{1}{\alpha_0}\int_0^\infty \log(y)\,y^{\alpha/\alpha_0}\,\exp(-y)\,\mathrm{d}y = -\frac{1}{\alpha_0}\Gamma'(1 + \alpha/\alpha_0).$$

Third,

$$\int_0^\infty x^{-\alpha}(\log x)^2\,\mathrm{d}P(x) = \frac{1}{\alpha_0^2}\int_0^\infty (\log y)^2\,y^{\alpha/\alpha_0}\,\exp(-y)\,\mathrm{d}y = \frac{1}{\alpha_0^2}\Gamma''(1 + \alpha/\alpha_0). \quad\square$$

Proof of Lemma B.2. Recall a few special values of the first two derivatives of the Gamma function:

$$\Gamma'(1) = -\gamma, \qquad\qquad \Gamma''(1) = \gamma^2 + \pi^2/6,$$
$$\Gamma'(2) = 1 - \gamma, \qquad\qquad \Gamma''(2) = (1 - \gamma)^2 + \pi^2/6 - 1,$$
$$\Gamma'(3) = 3 - 2\gamma, \qquad\qquad \Gamma''(3) = 2((3/2 - \gamma)^2 + \pi^2/6 - 5/4).$$

Applying the formulas in Lemma B.1 with $\alpha \in \{0, \alpha_0, 2\alpha_0\}$, we find

$$\mathrm{var}(Y_1) = \alpha_0^{-2}\left\{\Gamma''(3) - (\Gamma'(2))^2\right\} = \alpha_0^{-2}(1 - 4\gamma + \gamma^2 + \pi^2/3),$$
$$\mathrm{var}(Y_2) = \Gamma(3) - (\Gamma(2))^2 = 1,$$
$$\mathrm{var}(Y_3) = \alpha_0^{-2}\left(\Gamma''(1) - (\Gamma'(1))^2\right) = \alpha_0^{-2}\pi^2/6,$$

as well as

$$\mathrm{cov}(Y_1, Y_2) = \alpha_0^{-1}\left((-\Gamma'(3)) - (-\Gamma'(2))\Gamma(2)\right) = \alpha_0^{-1}(\gamma - 2),$$
$$\mathrm{cov}(Y_1, Y_3) = \alpha_0^{-2}\left(\Gamma''(2) - (-\Gamma'(2))(-\Gamma'(1))\right) = \alpha_0^{-2}(\pi^2/6 - \gamma),$$
$$\mathrm{cov}(Y_2, Y_3) = \alpha_0^{-1}\left((-\Gamma'(2)) - \Gamma(2)(-\Gamma'(1))\right) = -\alpha_0^{-1}. \qquad\square$$

Proof of Lemma B.3. If $X \sim P_{(\alpha,\sigma)}$, then $Z = X/\sigma \sim P_{(\alpha,1)}$. Therefore, by (2.1) and Lemma B.1,

$$\iota_{11} = \mathbb{E}\big[\{\alpha^{-1} + (Z^{-\alpha} - 1)\log Z\}^2\big]$$
$$= \frac{1}{\alpha^2}\big[1 - 2\{\Gamma'(2) - \Gamma'(1)\} + \{\Gamma''(3) - 2\Gamma''(2) + \Gamma''(1)\}\big]$$
$$= \frac{1}{\alpha^2}\{(1 - \gamma)^2 + \pi^2/6\}$$

Similarly, by (2.1) and (2.2),

$$\iota_{12} = \frac{\alpha}{\sigma}\mathbb{E}\big[(1 - Z^{-\alpha})\{\alpha^{-1} + (Z^{-\alpha} - 1)\log Z\}\big]$$
$$= \frac{\alpha}{\sigma}\big[\alpha^{-1}\{\Gamma(1) - \Gamma(2)\} + \alpha^{-1}\{\Gamma'(1) - 2\Gamma'(2) + \Gamma'(3)\}\}\big]$$
$$= \frac{1 - \gamma}{\sigma}.$$

Finally,

$$\iota_{22} = \frac{\alpha^2}{\sigma^2}\mathbb{E}[(1 - Z^{-\alpha})^2] = \frac{\alpha^2}{\sigma^2}\{\Gamma(1) - 2\Gamma(2) + \Gamma(3)\} = \frac{\alpha^2}{\sigma^2}. \qquad \square$$