

TAIL INFERENCE USING EXTREME U-STATISTICS

Jochem Oorschot, Johan Segers, Chen
Zhou

REPRINT | 2023 / 06

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

Tail inference using extreme U-statistics

Dedicated to the memory of James Pickands III (1931–2022)

Jochem Oorschot

*Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam,
The Netherlands
e-mail: oorschot@ese.eur.nl*

Johan Segers

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

Chen Zhou

*Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam,
The Netherlands
e-mail: zhou@ese.eur.nl*

Abstract: Extreme U-statistics arise when the kernel of a U-statistic has a high degree but depends only on its arguments through a small number of top order statistics. As the kernel degree of the U-statistic grows to infinity with the sample size, estimators built out of such statistics form an intermediate family in between those constructed in the block maxima and peaks-over-threshold frameworks in extreme value analysis. The asymptotic normality of extreme U-statistics based on location-scale invariant kernels is established. Although the asymptotic variance coincides with the one of the Hájek projection, the proof goes beyond considering the first term in Hoeffding’s variance decomposition. We propose a kernel depending on the three highest order statistics leading to a location-scale invariant estimator of the extreme value index resembling the Pickands estimator. This *extreme Pickands U-estimator* is asymptotically normal and its finite-sample performance is competitive with that of the pseudo-maximum likelihood estimator.

Keywords and phrases: U-statistic, generalized Pareto distribution, Hájek projection, extreme value index.

Received March 2022.

Contents

1	Introduction	1114
2	Asymptotic theory: sample drawn from the GP distribution	1116
2.1	Conditions and theorem	1116
2.2	Background theory and sketch of the proof of Theorem 2.1	1117
3	Asymptotic theory: sample drawn from the DoA of a GEV distribution	1120
3.1	Conditions and general result	1120
3.2	Location-scale invariant kernels where $q = 3$	1122

4	Estimating the extreme value index	1123
5	Finite-sample performance	1125
5.1	Samples from a GP distribution	1125
5.2	Random variables in the DoA of a GEV distribution	1127
5.3	Comparison with ABM estimator	1129
6	Discussion and open problems	1131
A	Preliminary distributional representations	1131
B	Proof of Proposition 2.2	1133
B.1	Proof of sub-goal (B.3)	1135
C	Proof of Proposition 2.3	1142
D	Proof of Theorem 3.2	1146
E	Proofs for Section 4	1153
	Acknowledgments	1158
	References	1158

1. Introduction

Let $X_1, \dots, X_n$ be independent and identically distributed random variables drawn from a common distribution F . Consider the generalized average

$$U_n^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_m(X_I), \quad (1.1)$$

where $K_m : \mathbb{R}^m \rightarrow \mathbb{R}$ with $1 \leq m \leq n$ is a permutation invariant function called kernel, $[n] := \{1, \dots, n\}$ and X_I is the vector $(X_i)_{i \in I}$ indexed by the subset I . The generalized average U_n^m is called a U-statistic and is an estimator of its corresponding “kernel parameter” θ_m ,

$$\theta_m = \mathbb{E}[U_n^m] = \mathbb{E}[K_m(X_1, \dots, X_m)].$$

By efficiently exploiting the information in the sample, U-statistics corresponding to kernels that are not linked to a particular parametric model can attain uniformly minimum variance among all unbiased estimators of θ_m , see [9] for the earliest theoretical development and [12], [19, Chapter 5] or [20, Chapter 12] for an overview of U-statistics.

In this work we adapt the theory of U-statistics to the setting of extreme value analysis, which is concerned with the tail of a distribution function. If the block size m is held fixed, as in the classic setup of U-statistics, there are many blocks containing observations that cannot be considered as part of the tail of the distribution. Therefore, we require that the block size m tends to infinity as $n \rightarrow \infty$ and we let the kernel function K_m depend only on the top- q observations in a block of m observations, i.e., there is $K : \{(x_1, \dots, x_q) \in \mathbb{R}^q : x_1 \geq \dots \geq x_q\} \rightarrow \mathbb{R}$ satisfying

$$K_m(X_1, \dots, X_m) = K(X_{m:m}, \dots, X_{m-q+1:m}), \quad (1.2)$$

where $X_{m:m} \geq X_{m-1:m} \geq \dots \geq X_{1:m}$ denote the order statistics of $X_1, \dots, X_m$. Note that while m grows, the number of upper order statistics q and the kernel K on $\mathbb{R}^q$ remain fixed. U-statistics corresponding to kernel functions as in (1.2) where $m := m(n) \rightarrow \infty$ and $m = o(n)$ as $n \rightarrow \infty$ are called *extreme U-statistics*. The likelihood equation for the all-block maxima estimator in [14, eq. (2.3)] can be seen as a particular instance of an extreme U-statistic with $q = 1$.

Following usual practice in extreme value analysis, we assume that the observations are drawn from a common distribution function F in the domain of attraction (DoA) of a Generalized Extreme Value (GEV) distribution with parameter $\gamma \in \mathbb{R}$. For such F , there exist sequences a_n and b_n such that

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}) \quad (1.3)$$

for $x \in \mathbb{R}$ such that $1 + \gamma x > 0$; notation $F \in \mathcal{D}(G_\gamma)$. The parameter γ is called the extreme value index and governs the tail behaviour of F . We will show that extreme U-statistics can be valid estimators of $\lim_{m \rightarrow \infty} \theta_m = \mu(\gamma)$ for some function μ . If the function μ is invertible, we can use the extreme U-statistic to construct estimators of γ .

Extreme U-statistics corresponding to location-scale invariant kernels were first explored in [17]. It was shown there that under suitable conditions the extreme U-statistic is consistent for $\mu(\gamma)$. In this work, we study the asymptotic distribution of extreme U-statistics and we will show that, like classic U-statistics [10], they are asymptotically normally distributed. The theory is subdivided into two parts: in Section 2, to focus on the asymptotic variance, we treat the case of an independent random sample drawn from the Generalized Pareto (GP) distribution. In Section 3, our main result, Theorem 3.1, handles random samples drawn from a distribution F in the DoA of a GEV distribution, in which case an asymptotic bias term arises. In Theorem 3.2, we fix $q = 3$, the minimum number of top order statistics required to obtain a location-scale invariant kernel, and we provide easily verifiable kernel integrability conditions under which Theorem 3.1 applies. As an application of Theorem 3.2, we consider in Section 4 a location-scale invariant kernel $K_P : \mathbb{R}^3 \rightarrow \mathbb{R}$ that we call the *Pickands kernel* in view of its resemblance to the Pickands estimator [15]. The corresponding extreme U-statistic $U_{n,P}^m$ is called the *extreme U-Pickands estimator* and is shown to be an asymptotically normal estimator of γ . In a simulation study, we find that $U_{n,P}^m$ is competitive with the Maximum Likelihood (ML) estimator in the peaks-over-threshold approach based on the Generalized Pareto (GP) distribution [3].

A technical challenge for extreme-U statistics is that their asymptotic distribution cannot be obtained in the same way as for classical U-statistics. A new approach is needed, which constitutes our main theoretical contribution. Recall that, classically, one truncates the variance of the U-statistic to the first term of the Hoeffding variance decomposition [10, equation (5.12)] and exploits that the U-statistic's Hájek projection has the same variance. In contrast, in Proposition 2.2 we need to take into account more than the first term of the Hoeffding decomposition to deal with the asymptotic variance of the extreme U-statistic.

Estimators derived from extreme U-statistics are more efficient in the mean square error sense than estimators based on other block techniques in extreme value analysis. In the context of extreme value analysis, block techniques are often used to extract large observations, for example, in the classical block maxima approach, which has received renewed interest in, for instance, [5, 7]. In addition, recent studies have proposed estimators based on the sliding block technique [1, 16]. To make the comparison, note that $U_n^m = \mathbb{E}[K_m(X_1, \dots, X_m) \mid X_{1:n}, \dots, X_{n:n}]$, see for instance [20, p. 161]. This means that the conditional expectation of disjoint and sliding block statistics given the order statistics is equal to U_n^m , while, by Jensen's inequality for conditional expectations, the variance of U_n^m is not higher.

In Section 2, we develop the asymptotic theory of extreme U-statistics based on samples of the GP distribution. The extension to samples from a distribution in the DoA of some GEV distribution is treated in Section 3. The Pickands kernel leading to a novel estimator of the extreme value index is constructed in Section 4, the finite-sample performance of which is reported in Section 5. After a brief discussion in Section 6, the proofs of the results are given in a series of appendices.

2. Asymptotic theory: sample drawn from the GP distribution

2.1. Conditions and theorem

In Theorem 2.1 we state the asymptotic normality of the extreme U-statistic

$$U_{n,Z}^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_m(Z_I), \quad (2.1)$$

for K_m as in (1.2) and for independent random variables $Z_1, \dots, Z_n$ drawn from the GP(γ) distribution, i.e., $P(Z_i \leq z) = 1 - (1 + \gamma z)^{-1/\gamma}$ for $z \geq 0$ such that $1 + \gamma z > 0$, with the limiting interpretation $P(Z_i \leq z) = 1 - \exp(-z)$ if $\gamma = 0$. Usually, the GP family also includes a location and scale parameter, but these will play no role here since the kernel K will be assumed to be location-scale invariant. We start by listing the required conditions on the kernel K and the degree sequence $m := m_n$.

Condition 1 is an integrability condition on the kernel K to be checked by analysis. Condition 2 provides upper and lower bounds for the block sizes m . Both are rather weak and are satisfied for m of the order n^κ for any $0 < \kappa < 1$.

Condition 1. *The kernel K in (1.2) is location-scale invariant and not everywhere constant, implying $q \geq 3$. Moreover, for all $p > 0$ it holds that*

$$\mathbb{E}[|K((Z_{q-j:q-1})_{j=1}^q)|^p] < \infty,$$

for independent GP(γ) random variables $Z_1, \dots, Z_{q-1}$ and $Z_{0:q-1} := 0$.

Condition 2. *The degree sequence $m := m_n$ satisfies $m = o(n^{1-\epsilon})$ and $\ln(n) = o(m^{1-\delta})$ as $n \rightarrow \infty$ for some $\epsilon, \delta > 0$.*

Because of the threshold-stability property of the GP distribution, the value of $\theta_m = \mathbb{E}[K_m(Z_1, \dots, Z_m)]$ does not depend on $m \geq q$, see (A.2); we write $\theta = \theta_m$ in the remainder of this section.

Theorem 2.1. *If Conditions 1–2 are satisfied, then the extreme U-statistic $U_{n,Z}^m$ in (2.1) satisfies*

$$\sqrt{k} (U_{n,Z}^m - \theta) \xrightarrow{d} N(0, \sigma_K^2(\gamma)), \quad n \rightarrow \infty,$$

where $k = n/m$ and $\sigma_K^2(\gamma)$ depends only on K and γ and is defined in (2.11) below.

2.2. Background theory and sketch of the proof of Theorem 2.1

We start with a sketch of the derivation of the asymptotic distribution of a U-statistic in the classical case where m is fixed [19, 20], which will provide guidance in the case where $m \rightarrow \infty$. It will turn out that the same proof technique can be applied to the case where $m \rightarrow \infty$ but $m^2 = o(n)$ as $n \rightarrow \infty$, leading to the asymptotic distribution of the extreme U-statistics. By contrast, if $m^2 = o(n)$ does not hold, a novel approach is required: in the Hoeffding variance decomposition, a growing number of terms needs to be taken into account. The reader only interested in the main results can skip this subsection and jump immediately to Section 3.

Classical theory for fixed m By Hoeffding's decomposition,

$$\text{var } U_{n,Z}^m = \sum_{\ell=1}^m p_{n,m}(\ell) \zeta_{m,\ell} \quad \text{with} \quad \begin{cases} p_{n,m}(\ell) = \binom{n}{m}^{-1} \binom{m}{\ell} \binom{n-m}{m-\ell}, \\ \zeta_{m,\ell} = \text{cov}(K_m(X_S), K_m(X_{S'})), \end{cases} \quad (2.2)$$

where S and S' are sets of positive integers with $|S| = |S'| = m$ and $|S \cap S'| = \ell$. For fixed m , we get from (2.2) the asymptotic relation

$$\frac{p_{n,m}(\ell+1)}{p_{n,m}(\ell)} = O(n^{-1}), \quad \text{as } n \rightarrow \infty. \quad (2.3)$$

Provided $\zeta_{m,1} \neq 0$, it follows that the first term of Hoeffding's decomposition, $p_{n,m}(1)\zeta_{m,1}$, asymptotically dominates the value of the sum $\sum_{\ell=1}^m p_{n,m}(\ell)\zeta_{m,\ell}$. For $p_{n,m}(1)$, we find

$$p_{n,m}(1) = \frac{m!(n-m)!}{n!} \cdot m \cdot \frac{(n-m)!}{(n-2m+1)!(m-1)!} = \frac{m^2}{n} a_{m,n} \quad (2.4)$$

where

$$a_{m,n} = \frac{(n-m)!}{(n-1)!} \frac{(n-m)!}{(n-2m+1)!} = \frac{(n-m)(n-m-1) \cdots (n-2m+2)}{(n-1)(n-2) \cdots (n-m+1)}.$$

Because the kernel degree m is fixed, we have $a_{m,n} \rightarrow 1$ as $n \rightarrow \infty$. It follows that

$$\text{var } U_{n,Z}^m \sim (m^2/n) \zeta_{m,1}, \quad \text{as } n \rightarrow \infty, \quad (2.5)$$

where $x_n \sim y_n$ signifies that the ratio x_n/y_n converges to 1 as $n \rightarrow \infty$.

Recall that $\theta = \mathbb{E}[K_m(Z_1, \dots, Z_m)]$. Define the Hájek projection of $U_{n,Z}^m - \theta$ as

$$\begin{aligned} \hat{U}_{n,Z}^m &:= \sum_{j=1}^n \sum_{I \subset [n], |I|=m} \binom{n}{m}^{-1} \mathbb{E}[K_m(Z_I) - \theta \mid Z_j] \\ &= \sum_{j=1}^n \binom{n-1}{m-1} \binom{n}{m}^{-1} \hat{K}_m(Z_j) = \frac{1}{k} \sum_{j=1}^n \hat{K}_m(Z_j) \end{aligned} \quad (2.6)$$

where $k = n/m$ and

$$\hat{K}_m(x) = \mathbb{E}[K_m(Z_1, \dots, Z_{m-1}, x)] - \theta.$$

Because the asymptotic variance of $U_{n,Z}^m$, i.e., $(m^2/n)\zeta_{m,1}$, is equal to the variance of $\hat{U}_{n,Z}^m$, the asymptotic distribution of $U_{n,Z}^m - \theta$ is the same as that of $\hat{U}_{n,Z}^m$, i.e., as $n \rightarrow \infty$,

$$\frac{\hat{U}_{n,Z}^m}{\sqrt{\text{var } \hat{U}_{n,Z}^m}} - \frac{U_{n,Z}^m - \theta}{\sqrt{\text{var } U_{n,Z}^m}} \xrightarrow{p} 0.$$

The Hájek projection $\hat{U}_{n,Z}^m$ is asymptotically normal by the central limit theorem, and thus

$$\frac{k}{\sqrt{n\zeta_{m,1}}} (U_{n,Z}^m - \theta) \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty,$$

provided that the kernel is square-integrable and $\zeta_{m,1} > 0$. For more details, see the proof of Theorem 12.3 in [20]. Note again that we assumed that m is fixed.

Letting $m \rightarrow \infty$: first attempt We aim to adapt the above proof to an extreme U-statistic, that is, when $m \rightarrow \infty$ in such a way that $m = o(n)$ as $n \rightarrow \infty$. Such an adaptation is non-trivial because $\zeta_{m,1}$ varies as $m \rightarrow \infty$ and the combinatorial numbers $p_{n,m}(\ell)$ no longer satisfy equation (2.3) in general, which makes it difficult to establish the key variance approximation in (2.5).

If we consider again the first term $p_{n,m}(1)$ in (2.4), we need a further restriction on the range of m to ensure $a_{m,n} \rightarrow 1$ as $n \rightarrow \infty$. Clearly,

$$\left(1 - \frac{m-1}{n-m+1}\right)^{m-1} \leq a_{m,n} \leq \left(1 - \frac{m-1}{n-1}\right)^{m-1}. \quad (2.7)$$

The lower and upper bound tend to 1 if and only if $m^2 = o(n)$ as $n \rightarrow \infty$. The condition $m^2 = o(n)$ is also necessary to ensure that $p_{n,m}(1)$ is the dominant term in the sum $\sum_{\ell=1}^m p_{n,m}(\ell)$ as $n \rightarrow \infty$. Taken together, these facts lead to the variance approximation (2.5) but only under the additional restriction $m^2 = o(n)$ as $n \rightarrow \infty$.

Case $m \rightarrow \infty$: novel approach We overcome the restriction $m^2 = o(n)$ and obtain the desired asymptotic equivalence relation in (2.5) by further analyzing the subsequent terms $\zeta_{m,\ell}$ for $\ell > 1$ in Hoeffding's decomposition (2.2). For integers a and b such that $a \leq b$ write $[a:b] = [a, b] \cap \mathbb{Z} = \{a, \dots, b\}$.

Without loss of generality, let the two blocks in (2.2) be $S = [1:m]$ and $S' = [(m - \ell + 1):(2m - \ell)]$. Their overlap $S \cap S' = [(m - \ell + 1):m]$ contains exactly ℓ elements. It follows that, for $\ell \in [1:m]$,

$$\zeta_{m,\ell} = \mathbb{E} [K_m(Z_{[1:m]}) K_m(Z_{[(m-\ell+1):(2m-\ell)]})] - \theta^2.$$

The dependence between $K_m(Z_S)$ and $K_m(Z_{S'})$ stems from observations Z_i with $i \in S \cap S'$ in the top- q of both Z_S and $Z_{S'}$. If ℓ is small compared to m , then usually there will be at most one such observation. We can thus expect that $\zeta_{m,\ell}$ can be approximated by $\ell \zeta_{m,1}$. Recognizing the term $p_{n,m}(\ell)$ as the probability mass function of a hypergeometric random variable, we use the expectation formula to find

$$\sum_{\ell=1}^m p_{n,m}(\ell) \ell \zeta_{m,1} = \frac{m^2}{n} \zeta_{m,1} = \frac{m \zeta_{m,1}}{k}.$$

We obtain the bound

$$\left| \text{var } U_{n,Z}^m - \frac{m \zeta_{m,1}}{k} \right| \leq \sum_{\ell=1}^m p_{n,m}(\ell) |\zeta_{m,\ell} - \ell \zeta_{m,1}|. \quad (2.8)$$

In order to formally derive (2.5) as well as the asymptotic distribution of the extreme U-statistic $U_{n,Z}^m$, we show three limit relations. First, we show that the bound in (2.8) is $o(1/k)$ as $n \rightarrow \infty$, that is,

$$\lim_{n \rightarrow \infty} |k \text{var } U_{n,Z}^m - m \zeta_{m,1}| = 0. \quad (2.9)$$

Second, we show that

$$\lim_{m \rightarrow \infty} m \zeta_{m,1} = \sigma^2 > 0. \quad (2.10)$$

Note then that (2.9) and (2.10) imply (2.5) as well as

$$\frac{\text{var } U_{n,Z}^m}{\text{var } \hat{U}_{n,Z}^m} = \frac{\text{var } U_{n,Z}^m}{(n/k^2) \zeta_{m,1}} = \frac{k \text{var } U_{n,Z}^m}{m \zeta_{m,1}} \rightarrow 1, \quad \text{as } n \rightarrow \infty.$$

Finally, we show the asymptotic normality of the Hájek projection $\hat{U}_{n,Z}^m$, which, in the extreme value setting, is a centered and reduced row sum of a triangular array of row-wise independent and identically distributed random variables.

Limit relation (2.9) is the content of Proposition 2.2. Limit relation (2.10) and the asymptotic normality of the Hájek projection $\hat{U}_{n,Z}^m$ are the content of Proposition 2.3. Together, Propositions 2.2 and 2.3 lead to the main result of this section in Theorem 2.1.

Proposition 2.2. *If Conditions 1 and 2 are satisfied, we have*

$$\left| \text{var } U_{n,Z}^m - \frac{m}{k} \zeta_{m,1} \right| = o(1/k), \quad n \rightarrow \infty.$$

Let $\overline{K}_m = K_m - \theta$ denote the centered kernel. Recall that for integer $q \geq 1$, the Erlang(q) distribution is the one of the sum $E_1 + \dots + E_q$ of q independent unit exponential random variables.

Proposition 2.3. *If Condition 1 holds, then*

$$\lim_{m \rightarrow \infty} m\zeta_{m,1} = \int_0^\infty (\mathbb{E} [\overline{K}_{q+1}((Z_{q-j:q-1})_{j=1}^q, h_\gamma(S_q/x))])^2 dx := \sigma_K^2(\gamma), \quad (2.11)$$

where S_q is an Erlang(q) random variable, independent of the independent GP(γ) random variables $Z_1, \dots, Z_{q-1}$, and where $Z_{0:q-1} := 0$. If, additionally, Condition 2 holds, the Hájek projection $\hat{U}_{n,Z}^m$ in (2.6) satisfies

$$\sqrt{k} \hat{U}_{n,Z}^m \xrightarrow{d} N(0, \sigma_K^2(\gamma)), \quad n \rightarrow \infty. \quad (2.12)$$

The proofs of Propositions 2.2 and 2.3 are long and are given in Sections B and C, respectively. Together, the two propositions are the required ingredients for deriving the asymptotic distribution of $U_{n,Z}^m$ in Theorem 2.1.

3. Asymptotic theory: sample drawn from the DoA of a GEV distribution

3.1. Conditions and general result

Consider again the U-statistic U_n^m in (1.1) based on a location-scale invariant kernel, but, in contrast to Section 2, let $X_1, \dots, X_n$ be a random sample with common distribution function F that satisfies the DoA condition (1.3) with shape parameter $\gamma \in \mathbb{R}$. Our goal now is to prove the asymptotic normality of U_n^m in full generality, not only for generalized Pareto distributed samples.

The technique relies on a coupling construction, leading to a GP(γ) sample $Z_1, \dots, Z_n$ associated to $X_1, \dots, X_n$. The extreme U-statistic $U_{n,Z}^m$ of the GP(γ) sample will be close to U_n^m , up to an asymptotic bias term.

For a continuous distribution function F , define the tail quantile function as

$$U(y) := \inf \{x \in \mathbb{R} : (1 - F(x))^{-1} \geq y\}, \quad y > 1. \quad (3.1)$$

For instance, the tail quantile function of GP(γ) distribution is

$$h_\gamma(y) = \int_1^y u^{\gamma-1} du = \begin{cases} (y^\gamma - 1)/\gamma & \text{if } \gamma \neq 0, \\ \ln y & \text{if } \gamma = 0. \end{cases} \quad (3.2)$$

To establish asymptotic theory, we construct the i.i.d. sample as

$$(X_1, \dots, X_n) = (U(Y_1^*), \dots, U(Y_n^*)), \quad (3.3)$$

for independent Pareto(1) random variables $Y_1^*, \dots, Y_n^*$, i.e., $P(Y_i^* \leq y) = 1 - 1/y$ for $y \geq 1$.

Since the distribution function F is in the DoA of a GEV distribution with shape parameter $\gamma \in \mathbb{R}$, there exists a measurable auxiliary function $a : (1, \infty) \rightarrow (0, \infty)$ such that

$$\lim_{x \rightarrow \infty} \frac{U(xy) - U(x)}{a(x)} = h_\gamma(y), \quad \text{locally uniformly in } y \in (0, \infty), \quad (3.4)$$

see for instance [4, Theorem 1.1.6]. Define

$$(Z_1, \dots, Z_n) = (h_\gamma(Y_1^*), \dots, h_\gamma(Y_n^*)).$$

for the same Pareto(1) random variables as in (3.3). As in Section 2, let the U-statistic with kernel K_m in (1.2) corresponding to the i.i.d. $\text{GP}(\gamma)$ random variables $Z_1, \dots, Z_n$ be denoted by

$$U_{n,Z}^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_m(Z_I).$$

We make the following assumptions on the tail quantile function U and the intermediate sequence k .

Condition 3. *The tail quantile function U in (3.1) is twice differentiable. Furthermore, the function A defined by*

$$A(t) = \frac{t U''(t)}{U'(t)} - \gamma + 1$$

is eventually positive or negative for large t and satisfies $\lim_{t \rightarrow \infty} A(t) = 0$, while the function $|A|$ is regularly varying with index $\rho < 0$.

Condition 4. *The sequence $k := n/m$ satisfies $\sqrt{k} A(m) \rightarrow \lambda \in \mathbb{R}$ as $n \rightarrow \infty$.*

Condition 3 is slightly stronger than the typical second-order condition; see [4, Theorem 2.3.12]. Condition 4 is a typical assumption in extreme value statistics leading to a finite asymptotic bias.

Recall that I is a set of m positive integers. Define

$$\xi_m^I = \frac{K_m(X_I) - K_m(Z_I)}{A(m)}$$

for A as in Condition 3 and simply write ξ_m whenever only the marginal distribution is of concern. We find

$$\begin{aligned} \sqrt{k} (U_n^m - U_{n,Z}^m) &= \sqrt{k} \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} (K_m(X_I) - K_m(Z_I)) \\ &= \sqrt{k} A(m) U_{n,\xi}^m, \end{aligned}$$

where, for fixed m ,

$$U_{n,\xi}^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} \xi_m^I$$

is another U-statistic.

The following condition assumes the weak convergence of the U-statistic $U_{n,\xi}^m$:

Condition 5. As $n \rightarrow \infty$, we have $U_{n,\xi}^m \xrightarrow{p} B_K$ for some constant $B_K \in \mathbb{R}$.

Theorem 3.1. Suppose that $X_1, \dots, X_n$ are i.i.d. random variables with a common distribution function F that satisfies the DoA condition (1.3) with shape parameter $\gamma \in \mathbb{R}$. Under Conditions 1–5, the corresponding extreme U -statistic, U_n^m , satisfies

$$\sqrt{k}(U_n^m - \theta) \xrightarrow{d} N(\lambda B_K, \sigma_K^2(\gamma)),$$

where $\sigma_K^2(\gamma)$ is as in Proposition 2.3 and $\theta = \mu(\gamma)$ is as in Theorem 2.1.

Proof of Theorem 3.1. Write

$$\begin{aligned} \sqrt{k}(U_n^m - \theta) &= \sqrt{k}(U_n^m - U_{n,Z}^m) + \sqrt{k}(U_{n,Z}^m - \theta) \\ &= \sqrt{k}A(m)U_{n,\xi}^m + \sqrt{k}(U_{n,Z}^m - \theta) \\ &=: I_1 + I_2. \end{aligned}$$

With Condition 1 and 2, we can apply Theorem 2.1 to get that $I_2 \xrightarrow{d} N(0, \sigma_K^2(\gamma))$. Condition 3–5 imply that $I_1 \xrightarrow{d} \lambda B_K$, which completes the proof. $\square$

Note that $\sigma_K^2(\gamma)$ does not depend on the distribution of X . Therefore, a practical consequence of Theorem 3.1 is that, for kernels K satisfying the required conditions, one can do a parametric bootstrap using $\text{GP}(\hat{\gamma})$ distributed random variables to obtain an estimate of $\sigma_K^2(\gamma)$, where $\hat{\gamma}$ can be any reasonable estimator of γ , provided that $\sigma_K^2(\gamma)$ is a continuous function of γ .

To make use of Theorem 3.1, one needs to verify that the kernel K and the tail quantile function U satisfy Condition 5. In Section 3.2, we deal with kernels where $q = 3$ — the minimum value required to obtain a location-scale invariant kernel that is not everywhere constant — and proceed to provide verifiable sufficient conditions under which Condition 5 holds, giving in addition an explicit expression for B_K .

3.2. Location-scale invariant kernels where $q = 3$

For a location-scale invariant kernel $K : \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 > x_2 > x_3\} \rightarrow \mathbb{R}$, we may write

$$\begin{aligned} K(x_1, x_2, x_3) &= K(x_1 - x_2, 0, x_3 - x_2) \\ &= K\left(\frac{x_1 - x_2}{x_2 - x_3}, 0, -1\right) \\ &=: g\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right) \end{aligned} \tag{3.5}$$

for some function $g : \mathbb{R} \rightarrow \mathbb{R}$, with $x_1 > x_2 > x_3$. We assume that F is continuous, so that ties do not occur with probability one. Further, note that a location-scale invariant function of one or two real arguments is necessarily constant, which is why we consider $q = 3$.

Next we assume the existence of first-order partial derivatives of K , which is equivalent to the differentiability of g . Let g' denote the derivative of g . We assume the following condition on g' .

Condition 6. *The derivative g' is continuous and satisfies one of the following two properties:*

(a) *The function g' is monotone and there exists $\epsilon > 0$ such that*

$$\limsup_{m \geq 3} \mathbb{E} \left[\left| g' \left(\ln \left(\frac{X_{m:m} - X_{m-1:m}}{X_{m-1:m} - X_{m-2:m}} \right) \right) \right|^{2+\epsilon} \right] < \infty,$$

(b) *The function g' is bounded.*

In addition, we assume that the moments of the kernel function do not grow too quickly as $n \rightarrow \infty$.

Condition 7. *There exist $\epsilon > 0$ and $a > 0$ such that*

$$\mathbb{E} \left[|K_m(X_1, \dots, X_m)|^{2+\epsilon} \right] = O(m^a), \quad m \rightarrow \infty. \quad (3.6)$$

We obtain the following consequence of Theorem 3.1.

Theorem 3.2. *For a kernel $K : \mathbb{R}^3 \rightarrow \mathbb{R}$, Conditions 3, 6 and 7 imply Condition 5 with*

$$B_K = B_K(\rho, \gamma) := \mathbb{E} [S_3^{-\rho}] \mathbb{E} \left[\sum_{j=1}^2 \dot{K}_j (\{h_\gamma(Y_{3-j:2})\}_{j=1}^3) H_{\gamma,\rho}(Y_{3-j:2}) \right], \quad (3.7)$$

where S_3 is an Erlang(3) random variable and where $H_{\gamma,\rho}$ is defined in (D.3).

Under Conditions 1–4, 6 and 7, the extreme U-statistic U_n^m satisfies

$$\sqrt{k} (U_n^m - \theta) \xrightarrow{d} N(\lambda B_K(\rho, \gamma), \sigma_K^2(\gamma)), \quad n \rightarrow \infty$$

where $\sigma_K^2(\gamma)$ is the same as in Proposition 2.3 and for $\theta = \mu(\gamma)$ as defined prior to Theorem 2.1.

4. Estimating the extreme value index

We develop a kernel depending on the top-3 observations in a block producing an extreme U-statistic that is an unbiased estimator of the shape parameter $\gamma \in \mathbb{R}$ when the observations are independently sampled from a $\text{GP}(\gamma)$ distribution. For integer $m \geq 3$, define the permutation and location-scale invariant kernel $K_{P,m} : \mathbb{R}^m \rightarrow \mathbb{R}$ by

$$K_{P,m}(x_1, \dots, x_m) = K_P(x_{m:m}, x_{m-1:m}, x_{m-2:m}), \quad (4.1)$$

with $K_P : \{(y_1, y_2, y_3) \in \mathbb{R}^3 : y_1 > y_2 > y_3\} \rightarrow \mathbb{R}$ given by

$$K_P(y_1, y_2, y_3) = 2 \ln \left(\frac{y_1 - y_2}{y_2 - y_3} \right) - \ln \left(\frac{y_1 - y_3}{y_2 - y_3} \right) = \ln \left(\frac{(y_1 - y_2)^2}{(y_1 - y_3)(y_2 - y_3)} \right). \quad (4.2)$$

Because of the resemblance with the estimator of the extreme value index defined in [15], we call K_P the *Pickands kernel*. It consists of two terms: the first one keeps the so-called Galton ratios as small as possible, see [18], while the second one was considered in [17].

The extreme U-statistic based on K_P has a particularly attractive property.

Proposition 4.1. *For $\gamma \in \mathbb{R}$, integer $n \geq m \geq 3$, and independent $\text{GP}(\gamma)$ random variables $Z_1, \dots, Z_n$, we have*

$$\mathbb{E}[K_{P,m}(Z_1, \dots, Z_m)] = \gamma.$$

Consequently, the U-statistic

$$\binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_{P,m}(Z_I)$$

is an unbiased estimator of γ .

For a random sample $X_1, \dots, X_n$ from a distribution F with extreme value index γ , we need a weak additional condition on F to guarantee the asymptotic normality of $\sqrt{k}(U_{n,P}^m - \gamma)$, where

$$U_{n,P}^m := \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_{P,m}(X_I) \quad (4.3)$$

is the *extreme U-Pickands estimator*.

Condition 8. *The distribution function F admits a density function f with support $\mathcal{E} = \{x \in \mathbb{R} : f(x) > 0\}$, say. There exist $\epsilon > 0$ and $p > 2$ such that*

$$\int_{-\infty}^{-1} (\ln(-x))^p f(x) dx < \infty, \\ \sup_{x \in \mathcal{E}} \left| F(x) - F(x + e^{-r^{1/p}}) \right| = O(r^{-(1+\epsilon)}), \quad \text{as } r \rightarrow \infty. \quad (4.4)$$

Note that (4.4) holds as soon as the density f is bounded, since $F(y) - F(x) = \int_x^y f(t) dt \leq (y - x) \sup_{t \in \mathcal{E}} f(t)$ for real $x < y$.

Theorem 4.2. *If Conditions 2, 3, 4 and 8 are fulfilled, then*

$$\sqrt{k}(U_{n,P}^m - \gamma) \xrightarrow{d} N(\lambda B_{K_P}(\gamma, \rho), \sigma_{K_P}^2(\gamma)), \quad n \rightarrow \infty,$$

where the asymptotic variance $\sigma_{K_P}^2$ and the bias B_{K_P} are given in equations (2.11) and (3.7), respectively, with K to be replaced by K_P defined in (4.2).

The finite-sample performance of $U_{n,P}^m$ as an estimator of γ will be explored in Section 5 by a simulation study. For calculation purposes, it is convenient to express $U_{n,P}^m$ explicitly as a linear combination of the log-spacings.

Lemma 4.3. *The U-statistic $U_{n,P}^m$ based on a random sample $X_1, \dots, X_n$ can be written in terms of the order statistics as*

$$U_{n,P}^m = \sum_{j=2}^{n-m+3} \frac{\binom{n-j}{m-3}}{\binom{n}{m}} \left(2 \frac{n-j+1}{m-2} - j \right) \sum_{i=1}^{j-1} \ln(X_{n-i+1:n} - X_{n-j+1:n}). \quad (4.5)$$

For given n and m , the coefficients on the right-hand side of (4.5) lend themselves well for recursive calculation, without the need of invoking factorials:

$$\begin{aligned} \frac{\binom{n-2}{m-3}}{\binom{n}{m}} &= \frac{m(m-1)(m-2)}{n(n-1)(n-m+1)}, \\ \frac{\binom{n-j}{m-3}}{\binom{n}{m}} &= \frac{\binom{n-(j-1)}{m-3}}{\binom{n}{m}} \cdot \frac{n-j-m+4}{n-j+1}, \quad j \in [3:(n-m+3)]. \end{aligned}$$

5. Finite-sample performance

We compare the performance of the extreme U-Pickands estimator $U_{n,P}^m$ in (4.3) of $\gamma \in \mathbb{R}$ to the location-scale invariant GP Maximum Likelihood (ML) estimator [3] based on the k excesses over the $(k+1)$ -largest order statistic. While the latter is also considered an estimator of $\gamma \in \mathbb{R}$, the algorithm to compute the ML estimator is theoretically constrained to $\gamma > -1$ [8], see also [21]. We simulate independent random samples of size n from a number of distributions and for each sample we apply both estimators of γ over a range of block sizes m and threshold parameters k . To facilitate comparison between the two estimators we fix the number of tail observations to be used in GP ML estimation to $k = 3n/m$, where m is the block size of the kernel $K_{P,m}$. The motivation for this choice is as follows: theoretically, for estimators using the top-3 observations of a disjoint partitioning of the dataset in blocks of size m , the number of data points used is equal to $3n/m$ too; moreover, when $m = 3$, both $U_{n,P}^m$ and the GP ML estimator use all data.

In Section 5.3, we also compare the extreme U-Pickands estimator with the All Block Maxima (ABM) estimator in [14]. This estimator extracts the maximum out of every block of size m and then fits a two-parameter Fréchet distribution to the sample of all block maxima thus obtained; the two parameters are the shape parameter $\gamma > 0$ and the scale parameter $\sigma > 0$. The likelihood equation (2.3) in that paper is a scale-invariant function of the block maxima. Although the estimator of γ is not an extreme U-statistic in the sense of our paper, it is still close in spirit to it with $q = 1$.

5.1. Samples from a GP distribution

We start by comparing the MSE of the GP ML estimator to $U_{n,P}^m$ in the case that the data are sampled from the GP distribution with shape parameter

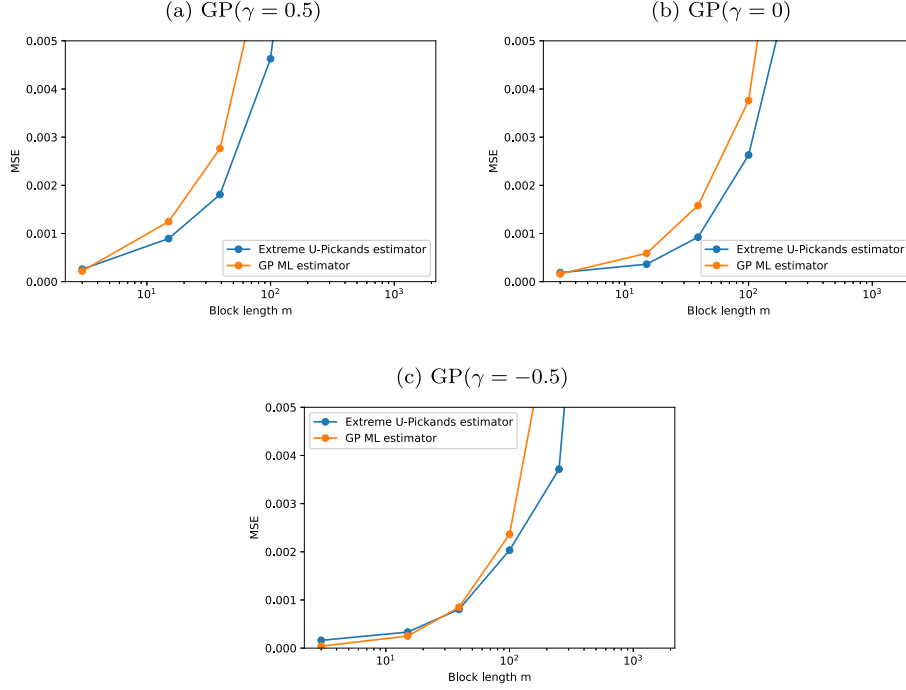


FIG 1. MSE comparison of $U_{n,P}^m$ and the GP ML estimator for 100 samples of size $n = 10\,000$ from $GP(\gamma)$

$\gamma \in \{-0.5, 0, 0.5\}$. Effectively, the comparison will be on the basis of estimator variance, since, in the GP case, $U_{n,P}^m$ is unbiased and the GP ML estimator is asymptotically unbiased.

Figure 1 shows that the MSE for both estimators is minimal when we use all available observations in estimation (block length $m = 3$). Moreover, the MSE of $U_{n,P}^m$ is significantly lower than that of the GP ML estimator over the largest range of the block length m , except when $\gamma = -0.5$. The apparent discrepancy between the two cases $\gamma \in \{0, 0.5\}$ on the one hand and the case $\gamma = -0.5$ on the other hand can be understood by a comparison of the asymptotic variance.

Recall that the GP ML estimator has asymptotic variance $(1 + \gamma)^2$, provided $\gamma > -1/2$ [6]. For the asymptotic variance of $U_{n,P}^m$, we fix a grid of 50 equally spaced values of γ between -1 and 1 and approximate $\sigma_{K_P}^2(\gamma)$ by the sample variance of $U_{n,P}^m$ based on 1000 $GP(\gamma)$ samples of size $n = 10\,000$, multiplied with $k = n/m = 100$. The normalized asymptotic variance of the GP ML estimator $(1 + \gamma)^2/3$ (like before, we divide by three to facilitate comparison) together with the estimated asymptotic variance $\hat{\sigma}_{K_P}^2(\gamma)$ corresponding to $U_{n,P}^m$ are depicted in Figure 2. The asymptotic variances intersect around $\gamma \approx -0.25$, with $\sigma_{K_P}^2(\gamma)$ indeed being higher for lower values of γ but lower for all higher values of γ . In this context it is possible for the asymptotic variance $\sigma_{K_P}^2(\gamma)$ to be lower

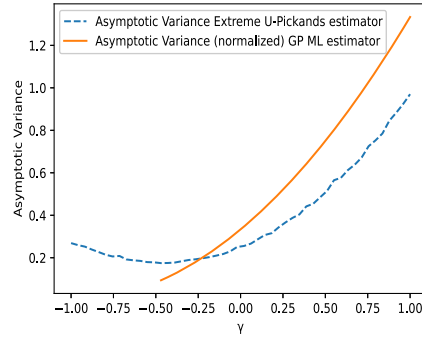


FIG 2. Normalized asymptotic variance $\sigma_{K_P}^2(\gamma)$ of $U_{n,P}^m$ and $(1+\gamma)^2/3$, corresponding to the one of the GP ML estimator

TABLE 1
Asymptotic variance $\sigma_{K_P}^2(\gamma)$ for $\gamma \in [-1, 1]$

γ	$\sigma_{K_P}^2(\gamma)$	γ	$\sigma_{K_P}^2(\gamma)$	γ	$\sigma_{K_P}^2(\gamma)$	γ	$\sigma_{K_P}^2(\gamma)$	γ	$\sigma_{K_P}^2(\gamma)$
-1.000	0.269	-0.959	0.259	-0.918	0.252	-0.878	0.239	-0.837	0.227
-0.796	0.214	-0.755	0.206	-0.714	0.208	-0.673	0.192	-0.633	0.188
-0.592	0.185	-0.551	0.179	-0.510	0.178	-0.469	0.175	-0.429	0.175
-0.388	0.177	-0.347	0.183	-0.306	0.189	-0.265	0.193	-0.224	0.198
-0.184	0.203	-0.143	0.210	-0.102	0.218	-0.061	0.230	-0.020	0.251
0.020	0.255	0.061	0.267	0.102	0.287	0.143	0.307	0.184	0.316
0.225	0.343	0.265	0.366	0.306	0.387	0.347	0.404	0.388	0.442
0.429	0.454	0.469	0.485	0.510	0.514	0.551	0.565	0.592	0.576
0.633	0.612	0.674	0.638	0.714	0.672	0.755	0.725	0.796	0.752
0.837	0.785	0.878	0.845	0.918	0.883	0.959	0.924	1.000	0.970

than the one of the GP ML estimator because $U_{n,P}^m$ uses more observations than the GP ML estimator. For completeness the approximate values of $\sigma_{K_P}^2(\gamma)$ for $\gamma \in [-1, 1]$ can be found in Table 1. These can be used for constructing asymptotic confidence intervals when using the extreme U-Pickands estimator, at least when the bias term is negligible.

5.2. Random variables in the DoA of a GEV distribution

Next we consider data drawn from a distribution function in the DoA of a GEV distribution. The optimal block length m in a MSE sense will then not be equal to three, but will in general be a higher value that better balances bias and variance.

In Figure 3 we illustrate the relative performance of $U_{n,P}^m$ in terms of MSE for a couple of well known distributions. The conclusions from the previous section remain; the minimal MSE of $U_{n,P}^m$ is competitive with that of the GP ML estimator (sometimes even lower) and the performance of $U_{n,P}^m$ is usually better over a large range of block lengths.

MSE comparisons for observations from a distribution in the DoA of a GEV distribution effectively net out the competing effects of changes in estimator vari-

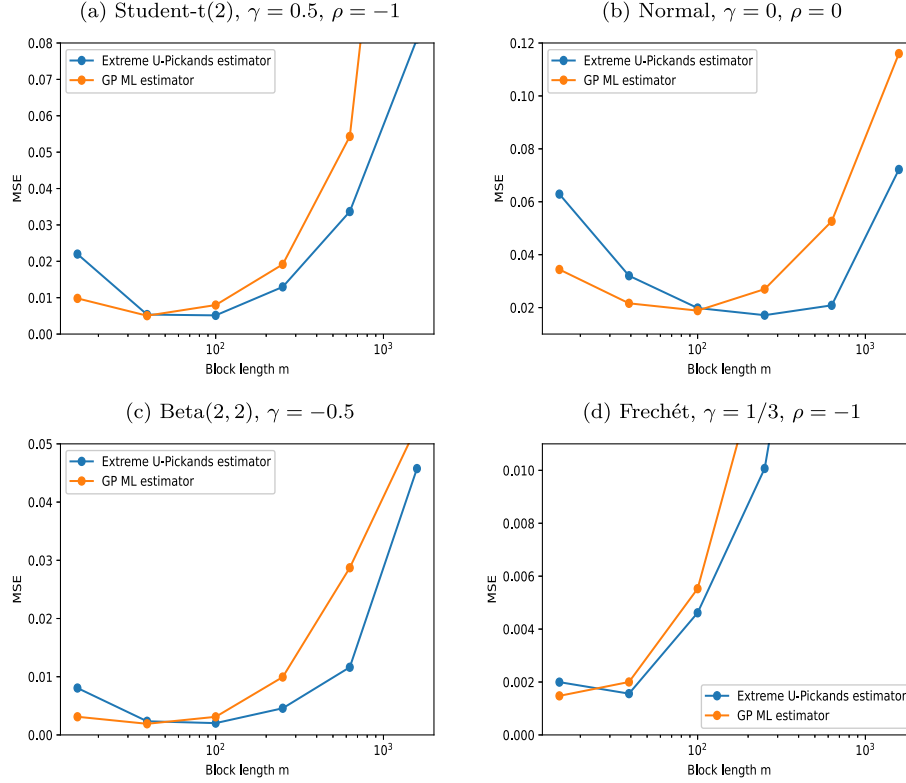


FIG 3. MSE comparison of $U_{n,P}^m$ and the GP ML estimator at $k = 3n/m$ for 100 samples of size $n = 10\,000$ from the stated distributions, with extreme value index $\gamma \in \mathbb{R}$ and second-order parameter $\rho \leq 0$ (Condition 3) as indicated

ance (depending on γ) to changes in estimator bias (depending on the second-order parameter ρ as well as on γ). For many distributions, γ and ρ cannot be controlled separately. This is why we study the Burr distribution to disentangle the effects of ρ and γ . The Burr(λ, η) distribution has distribution function $F(x) = 1 - (1/(1 + x^\eta))^\lambda$ for $\eta, \lambda, x > 0$, with extreme value index $\gamma = (\lambda\eta)^{-1}$ and second-order parameter $\rho = -\lambda^{-1}$.

Figure 4 depicts the bias for several choices of m based on 1000 simulated datasets of $n = 10\,000$ observations from a Burr distribution where the parameters are chosen such that $\gamma = 1/2$ is fixed as we vary ρ . As expected, the bias of both estimators approaches 0 as ρ decreases and increases steeply as ρ approaches 0. While subplot (a) of Figure 4 shows that the bias of $U_{n,P}^m$ is somewhat lower than that of the GP ML estimator when ρ is very negative, subplot (b) shows that when ρ approaches 0, the bias of $U_{n,P}^m$ is the higher one for all values of m .

Finally, Figure 5 shows that a single trajectory of the estimator $U_{n,P}^m$ as a function of unit increments in m is smooth, in contrast to the characteristically

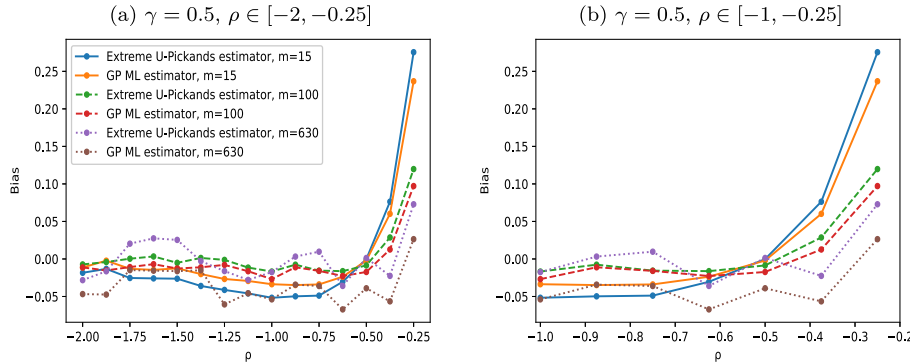


FIG 4. Bias of $U_{n,P}^m$ and the GP ML estimator based on 1000 samples of size $n = 10\,000$ from the Burr distribution with stated extreme value index γ and second-order parameter ρ . Panel (b) provides a zoom of panel (a) in the range $-1 \leq \rho \leq -0.25$.

erratic “Hill plot” of the GP ML estimator as a function of the top- k order statistics. Nonetheless, choosing the block size in practice remains a difficult task even when the single trajectory of the estimator $U_{n,P}^m$ is smooth.

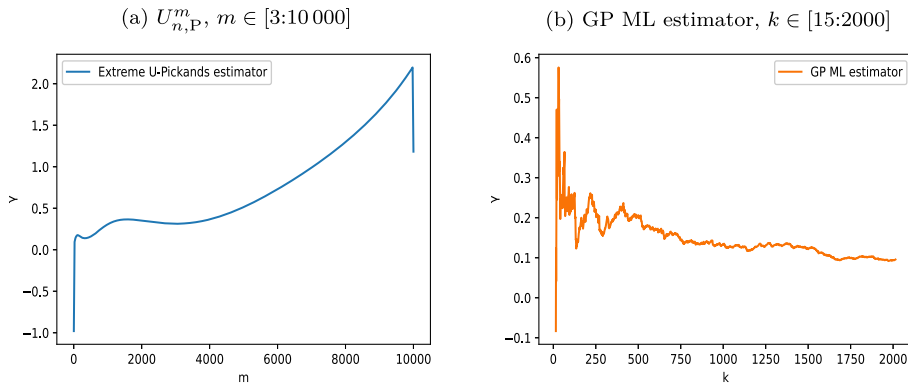


FIG 5. Single trajectories of the estimators as a function of m or k , based on a Student- $t(4)$ sample of size $n = 10\,000$

5.3. Comparison with ABM estimator

We compare the extreme U-Pickands estimator $U_{n,P}^m$ of γ to the ABM estimator in [14]. The latter estimator is restricted to the range $\gamma > 0$ and is scale invariant but not location invariant (although it can be made location invariant for instance by first centering the data at the median). Thinking in terms of disjoint blocks, the ABM estimator based on m -blocks uses n/m sample points while $U_{n,P}^m$ uses $3n/m = n/(m/3)$, three times as many. To ensure a level playing field,

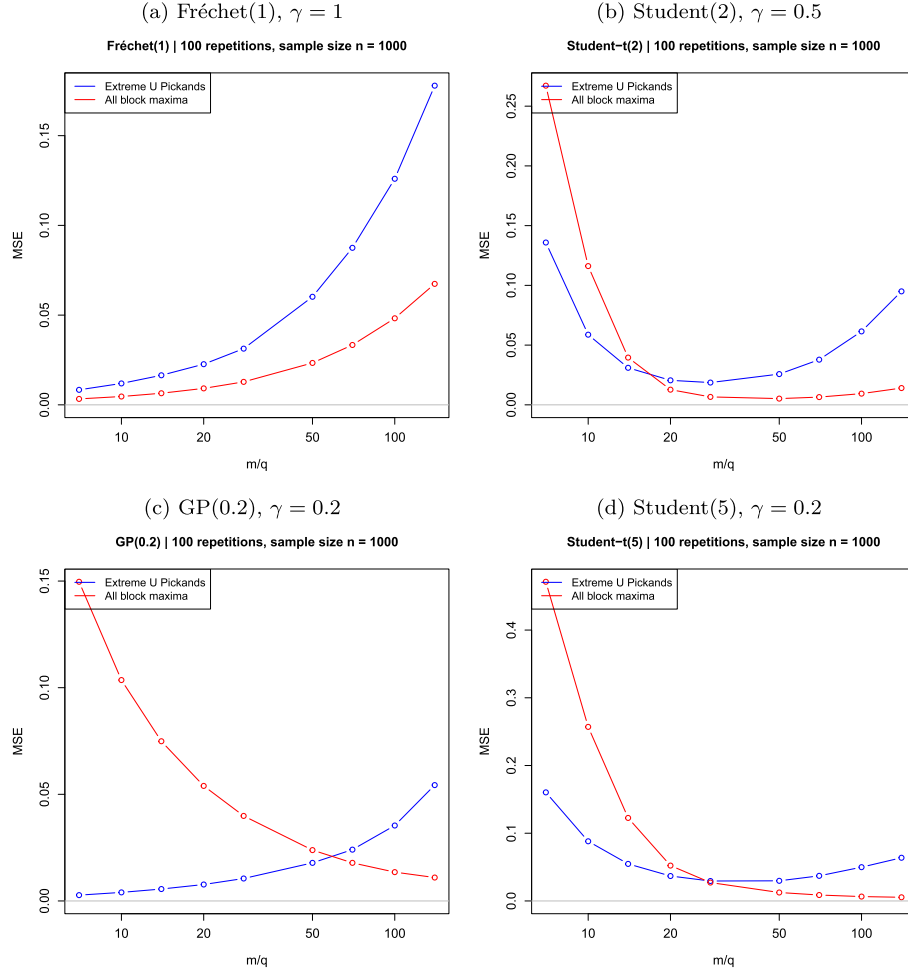


FIG 6. Comparison of MSE of ABM estimator [14] at block size m with that of $U_{n,P}^m$ at block size $3m$, based on 100 samples of size $n = 1000$ from a number of sampling distributions and for a range of m .

we compare the ABM estimator at block size m with the extreme U-Pickands estimator at block size $3m$. In other words, we compare both estimators at equal values of m/q , where m is the actual block size used for the estimator and where $q = 1$ for ABM while $q = 3$ for the extreme U-Pickands estimator.¹

For each of a range of distributions, Figure 6 shows the MSE over 100 samples of size $n = 1000$. As the ABM estimator is designed as a quasi-MLE based on the Fréchet distribution, its performance is superior for samples drawn from

¹Note that the recursive formula on page 5 of [14] should be $p_i = p_{i-1} \frac{n-i-m+2}{n-i+1}$ for $i \in [2:(n-m+1)]$.

that distribution. For samples drawn from other distributions, the extreme U-Pickands estimator is competitive, despite its simple and explicit definition.

6. Discussion and open problems

In Theorem 2.1, as well as elsewhere in the paper, we have assumed that the data are independently and identically distributed. Since γ is a parameter belonging to the marginal distribution of the observations, we expect that Theorem 2.1 can be adapted to the case where the Z_i all follow the same GP distribution but are weakly dependent. If the dependence is weak enough and the sample size is large, one may still find that within a randomly chosen block Z_I the observations are approximately independent, so that $\theta \approx \mathbb{E}[K_m(Z_I)]$ and so that the approximation $\zeta_{m,\ell} \sim \ell \zeta_{m,1}$ underlying Proposition 2.2 remains valid. The asymptotic variance will be different, however.

Furthermore, in the simulation study, when data were sampled from the $\text{GP}(\gamma)$ distribution, the estimator $U_{n,P}^3$ turned out to have the lowest variance among all unbiased estimators $U_{n,P}^m$ of γ , for $m \geq 3$. Since the estimator $U_{n,P}^3$ is also a function of the entire sample, it is an open question whether the estimator $U_{n,P}^3$ is the uniformly minimum variance unbiased estimator for γ of the $\text{GP}(\gamma)$ distribution. Because the GP distributions do not constitute an exponential family, this question is left to further research.

Finally, note that a U-statistic may also be calculated based on all exceedances over a high threshold. If the kernel is location-scale invariant, this amounts to applying a U-statistic only to a certain fraction of the largest observations in a sample. Doing so while keeping $m = q$ fixed would yield another kind of extreme U-statistic, perhaps worth studying as well.

Appendix A: Preliminary distributional representations

Let $\stackrel{d}{=}$ denote equality in distribution. For a distribution function F , recall its tail quantile function U in (3.1). The tail quantile function of the standard GP distribution with parameter $\gamma \in \mathbb{R}$ is h_γ in (3.2). Recall that a random variable Y is said to have a standard Pareto distribution if $\mathbb{P}(Y \leq y) = 1 - 1/y$ for $y \geq 1$. The function h_γ has the property that if Y follows a standard Pareto distribution, the variable $Z = h_\gamma(Y)$ follows a $\text{GP}(\gamma)$ distribution. Moreover, the function h_γ satisfies the functional equation

$$h_\gamma(xy) = h_\gamma(x) + x^\gamma h_\gamma(y), \quad x, y > 0.$$

Let $Y_1, Y_1^*, Y_2, Y_2^*, \dots$ be independent standard Pareto distributed random variables. For integer $1 \leq q \leq m$, Rényi's representation for exponential order statistics implies that the vector of top- q order statistics of $Y_1^*, \dots, Y_m^*$ admits the representation

$$(Y_{m-i+1:m}^*)_{i=1}^q \stackrel{d}{=} (Y_{m-q+1:m}^* Y_{q-i:q-1})_{i=1}^q \quad (\text{A.1})$$

where $Y_{0:q-1} = 1$. Then, we can represent the vector of top- q order statistics of an independent random sample $Z_1^*, \dots, Z_n^*$ from the $\text{GP}(\gamma)$ distribution as

$$\begin{aligned} (Z_{m-i+1:m}^*)_{i=1}^q &\stackrel{d}{=} (h_\gamma(Y_{m-i+1:m}^*))_{i=1}^q \\ &\stackrel{d}{=} (h_\gamma(Y_{m-q+1:m}^* Y_{q-i:q-1}))_{i=1}^q \\ &= (h_\gamma(Y_{m-q+1:m}^*) + (Y_{m-q+1:m}^*)^\gamma h_\gamma(Y_{q-i:q-1}))_{i=1}^q. \end{aligned}$$

We have thereby proved the following result.

Lemma A.1. *Consider the integers $1 \leq q \leq m$ and the scalar $\gamma \in \mathbb{R}$. The vector of top- q order statistics of an independent $\text{GP}(\gamma)$ random sample $Z_1^*, \dots, Z_m^*$ admits the representation*

$$(Z_{m-i+1:m}^*)_{i=1}^q \stackrel{d}{=} (Z_{m-q+1:m}^* + (1 + \gamma Z_{m-q+1:m}^*) Z_{q-i:q-1})_{i=1}^q$$

where $Z_1, \dots, Z_{q-1}$ is another independent standard $\text{GP}(\gamma)$ random sample, independent from $Z_1^*, \dots, Z_m^*$, and where $Z_{0:q-1} = 0$.

Recall that the kernel $K_m : \mathbb{R}^m \rightarrow \mathbb{R}$ of the extreme U-statistic is such that, for all integer $m \geq q$ and all $(x_1, \dots, x_m) \in \mathbb{R}^m$,

$$K_m(x_1, \dots, x_m) = K(x_{m:m}, \dots, x_{m-q+1:m}),$$

for some (symmetric) function $K : \mathbb{R}^q \rightarrow \mathbb{R}$. If the kernel is also location-scale invariant, i.e., if the kernel satisfies

$$K_m(ax_{m:m} + b, \dots, ax_{m-q+1:m} + b) = K(x_{m:m}, \dots, x_{m-q+1:m}).$$

for all $a > 0$ and $b \in \mathbb{R}$, the representation derived in Lemma A.1 has a useful consequence.

Lemma A.2. *Consider integers $3 \leq q \leq m$ and $\gamma \in \mathbb{R}$. If the kernel $K : \mathbb{R}^q \rightarrow \mathbb{R}$ is location-scale invariant and if $Z_1^*, \dots, Z_m^*$ are independent standard $\text{GP}(\gamma)$ random variables, then*

$$(Z_{m-q+1:m}^*, K_m(Z_1^*, \dots, Z_m^*)) \stackrel{d}{=} (Z_{m-q+1:m}^*, K((Z_{q-i:q-1})_{i=1}^q))$$

with $Z_1, \dots, Z_{q-1}$ as in Lemma A.1. In particular, $K_m(Z_1^*, \dots, Z_m^*)$ is independent of $Z_{m-q+1:m}^*$ and its distribution does not depend on m .

It follows from Lemma A.2 that

$$\theta := \theta_m = \mathbb{E}[K_m(Z_1^*, \dots, Z_m^*)] = \mathbb{E}[K((Z_{q-j:q-1})_{j=1}^q)] = \mu(\gamma), \quad (\text{A.2})$$

which does not depend on m , provided $m \geq q$.

Appendix B: Proof of Proposition 2.2

Recall the bound (2.8), which we state here for convenience again:

$$\left| \text{var } U_{n,Z}^m - \frac{m\zeta_{m,1}}{k} \right| \leq \sum_{\ell=1}^m p_{n,m}(\ell) |\zeta_{m,\ell} - \ell\zeta_{m,1}|. \quad (\text{B.1})$$

Further, recall that

$$p_{n,m}(\ell) = \binom{n}{m}^{-1} \binom{m}{\ell} \binom{n-m}{m-\ell}, \quad (\text{B.2})$$

for integer $1 \leq m \leq n$ and $0 \leq \ell \leq m$, represents the probability mass function of the hypergeometric distribution counting the number of successes ℓ after m draws without replacement out of an urn containing n balls of which m have the desired colour. To prove Proposition 2.2, we will decompose the sum over $\ell \in [1:m]$ in (B.1) into two pieces, $\ell \leq \ell_n$ and $\ell > \ell_n$, with $\ell_n = o(m)$ determined below.

- For $\ell \leq \ell_n$, we will show that $|\zeta_{m,\ell} - \ell\zeta_{m,1}|$ is sufficiently small.
- For $\ell > \ell_n$, we use tail bounds on the hypergeometric distribution $p_{n,m}(\ell)$.

We make these statements precise in the proof below by listing a number of properties that, in combination, imply that the bound (B.1) is $o(1/k)$ as $n \rightarrow \infty$.

Proof of Proposition 2.2. By (B.1), we have, for any $b > 0$,

$$\begin{aligned} \left| \text{var } U_n^m - \frac{m}{k} \zeta_{m,1} \right| &\leq \max_{\ell \in [1:\ell_n]} |\ell^{-1} \zeta_{m,\ell} - \zeta_{m,1}| \cdot \sum_{\ell=1}^{\ell_n} p_{n,m}(\ell) \ell \\ &\quad + \max_{\ell \in [1:m]} \ell^{-b} |\zeta_{m,\ell} - \ell\zeta_{m,1}| \cdot \sum_{\ell=\ell_n+1}^m p_{n,m}(\ell) \ell^b. \end{aligned}$$

The expectation of the hypergeometric distribution in (B.2) is $m \cdot m/n = m/k$. The previous decomposition implies the following break-down of (2.9) into three sub-goals: As $n \rightarrow \infty$,

$$\max_{\ell \in [1:\ell_n]} |\ell^{-1} \zeta_{m,\ell} - \zeta_{m,1}| = o(1/m), \quad (\text{B.3})$$

$$\max_{\ell \in [1:m]} \ell^{-b} |\zeta_{m,\ell} - \ell\zeta_{m,1}| = O(1), \quad (\text{B.4})$$

$$\sum_{\ell=\ell_n+1}^m p_{n,m}(\ell) \ell^b = o(1/k), \quad (\text{B.5})$$

where we choose b such that $1/2 \leq b \leq 1$.

If the above sub-goals are attained, we obtain that

$$\left| \text{var } U_n^m - \frac{m}{k} \zeta_{m,1} \right| \leq o(1/m) m/k + O(1) o(1/k) = o(1/k), \quad \text{as } n \rightarrow \infty.$$

The proof of the three sub-goals (B.3), (B.4) and (B.5) for the choice

$$\ell_n = \left\lfloor \left(\frac{m}{n} + \left(\frac{\ln n}{2m} \right)^{1/2} \right) m \right\rfloor, \quad (\text{B.6})$$

with $\lfloor x \rfloor$ the integer part of $x \in \mathbb{R}$, is the content of Lemmas B.1, B.2 and B.3 below. $\square$

We remark that the choice of ℓ_n is constrained by (B.3) and (B.5). On the one hand, for (B.3) to hold, we will need that $\ell_n = o(m)$ as $n \rightarrow \infty$. On the other hand, for (B.5) to hold, we will need that $\ell_n \rightarrow \infty$ sufficiently fast. Both criteria can be met for the choice of ℓ_n in (B.6), provided m is not too small or large as per Condition 2.

The proof of sub-goal (B.3) is somewhat involved and is developed in Section B.1 via a series of lemmas. First, we establish sub-goals (B.4) and (B.5).

Recall that for the independent $\text{GP}(\gamma)$ random variables $Z_1, \dots, Z_{2m-1}$,

$$\zeta_{m,\ell} = \mathbb{E} [K_m(Z_{[1:m]}) K_m(Z_{[(m-\ell+1):(2m-\ell)]})] - \theta^2,$$

for $\ell \in [1:m]$. For integer $m \geq q$ and real $a > 0$, define

$$\|K\|_a := (\mathbb{E} [|K((Z_{q-j:q-1})_{j=1}^q)|^a])^{1/a},$$

which, by Lemma A.2, does not depend on m .

Lemma B.1 (Sub-goal (B.4)). *If Condition 1 is satisfied, then, for every integer $m \geq q$, we have, for any $a > 2$,*

$$\zeta_{m,1} \leq (2q)^{1-2/a} \|K\|_a^2 \cdot m^{2/a-1}. \quad (\text{B.7})$$

As a consequence, we can take any $b \in (2/a, 1)$ to ensure that (B.4) holds.

Proof. To ensure (B.4), it is sufficient to have

$$\sup_{m \geq q} \max_{\ell \in [1:m]} \zeta_{m,\ell} < \infty, \quad (\text{B.8})$$

$$\sup_{m \geq q} m^{1-b} \zeta_{m,1} < \infty, \quad (\text{B.9})$$

provided $0 \leq b \leq 1$. By Hölder's inequality, Lemma A.2 and Condition 1,

$$\sup_{m \geq q} \max_{\ell \in [1:m]} \zeta_{m,\ell} \leq \|K\|_2^2 - \theta^2 < \infty,$$

such that (B.8) is guaranteed.

For (B.9), consider the difference

$$\Delta_m = K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]}) - K_{m-1}(Z_{[1:(m-1)]}) K_{m-1}(Z_{[(m+1):(2m-1)]}),$$

capturing the effect of omitting Z_m from $Z_1, \dots, Z_{2m-1}$. It follows that

$$\zeta_{m,1} = \mathbb{E} [K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]})] - \theta^2 = \mathbb{E}(\Delta_m).$$

Further, if Z_m does not belong to the top- q of $Z_{[1:m]}$ nor to the one of $Z_{[m:(2m-1)]}$, $K_m(Z_{[1:m]}) = K_{m-1}(Z_{[1:(m-1)]})$ and $K_m(Z_{[m:(2m-1)]}) = K_{m-1}(Z_{[(m+1):(2m-1)]})$, which implies that $\Delta_m = 0$. The probability that Z_m belongs to the top- q of $Z_{[1:m]}$ or $Z_{[m:(2m-1)]}$ is not larger than $2q/m$. We find, by applying Hölder's inequality twice, for any $a > 2$,

$$\begin{aligned} |\mathbb{E}(\Delta_m)| &\leq \mathbb{E}(|\Delta_m|) \leq \left(\mathbb{E} \left[|K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]})|^{a/2} \right] \right)^{2/a} (2q/m)^{1-2/a} \\ &\leq \|K\|_a^2 \cdot (2q/m)^{1-2/a}, \end{aligned}$$

which is the upper bound in (B.7). By multiplying the upper bound with m^{1-b} for some $b \in (2/a, 1)$, we obtain (B.4). $\square$

Lemma B.2 (Sub-goal (B.5)). *If Condition 2 holds, the sequence ℓ_n in (B.6) satisfies $\ell_n \rightarrow \infty$ and $\ell_n = o(m)$ as $n \rightarrow \infty$, and sub-goal (B.5) holds.*

Proof. The statements regarding the asymptotic order of ℓ_n follow immediately. Let H_n be a hypergeometric random variable with probability mass function $p_{n,m}$ in (B.2). We have $\mathbb{E}(H_n) = m/k$. Put $t = (\ell_n + 1)/m - 1/k$ for $t \geq 0$ to be determined. We have

$$\begin{aligned} k \sum_{\ell=\ell_n+1}^m p_{n,m}(\ell) \ell^b &\leq km^b \sum_{\ell=\ell_n+1}^m p_{n,m}(\ell) \\ &= km^b \mathbb{P}[H_n \geq (k^{-1} + t)m] \\ &\leq \exp(-2mt^2 - \ln k^{-1} + b \ln m), \end{aligned}$$

see [11] or [2]. Since $t \geq ((2m)^{-1} \ln n)^{1/2}$ and $b < 1$, we get

$$2mt^2 + \ln 1/k - b \ln(m) \geq \ln n - \ln k - b \ln m = (1 - b) \ln m \rightarrow \infty,$$

as $n \rightarrow \infty$. The lemma follows. $\square$

B.1. Proof of sub-goal (B.3)

It remains to prove the first sub-goal (B.3) for the choice of ℓ_n above. Recall that $\ell_n = o(m)$ and $\ell_n \rightarrow \infty$ as $n \rightarrow \infty$. Throughout, since we handle $\ell < \ell_n$, we assume ℓ, q, m are positive integers with $m \geq \ell + q$ and $\ell \geq 2$ (if $\ell = 1$ there is nothing to prove).

Lemma B.3 (Sub-goal (B.3)). *If Conditions 1 and 2 are satisfied, then sub-goal (B.3) holds, i.e., as $n \rightarrow \infty$,*

$$\max_{\ell \in [1:\ell_n]} |\ell^{-1} \zeta_{m,\ell} - \zeta_{m,1}| = o(1/m).$$

Proof of sub-goal (B.3). Consider the counting variable

$$N_{\ell,m} = \sum_{i=m-\ell+1}^m \mathbb{1}_{i,\ell,m},$$

where

$$\mathbb{1}_{i,\ell,m} = \mathbb{1}\{\{Z_i \text{ is in the top-}q \text{ of } Z_{[1:m]}\} \cup \{Z_i \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]}\}\}. \quad (\text{B.10})$$

Define

$$\Xi_{\ell,m} = K_m(Z_{[1:m]}) K_m(Z_{[(m-\ell+1):(2m-\ell)]}),$$

and write $\mathbb{E}[\Xi; A] = \mathbb{E}[\Xi \mathbb{1}_A]$ for a random variable Ξ and an event A . Then,

$$\begin{aligned} \zeta_{m,\ell} &= \mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] + \mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] + \mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \geq 2] - \theta^2 \\ \text{and } \zeta_{m,1} &= \mathbb{E}[\Xi_{1,m}; N_{1,m} = 0] + \mathbb{E}[\Xi_{1,m}; N_{1,m} = 1] - \theta^2. \end{aligned}$$

We obtain the following decomposition of the left-hand side of (B.3) into three terms:

$$\begin{aligned} |\zeta_{m,\ell} - \ell \zeta_{m,1}| &\leq |\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] - \ell \mathbb{E}[\Xi_{1,m}; N_{1,m} = 0] + (\ell - 1)\theta^2| \\ &\quad + |\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] - \ell \mathbb{E}[\Xi_{1,m}; N_{1,m} = 1]| \\ &\quad + |\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \geq 2]| \\ &:= \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3 \end{aligned} \quad (\text{B.11})$$

According to Lemmas B.4, B.5 and B.6 below, we have

$$\mathbb{I}_j \lesssim \ell (\ell/m^2)^{1-2/a} \quad (\text{B.12})$$

for $j \in \{1, 2, 3\}$ and any $a > 2$, where $a_n \lesssim b_n$ if and only if there exists a constant $c > 0$ such that $a_n \leq cb_n$ for all n . Recall the choice of ℓ_n in (B.6). Together with Condition 2, the bounds on $\mathbb{I}_j$ will then imply, as $n \rightarrow \infty$ and for sufficiently large a , that

$$\begin{aligned} \max_{\ell \in [1:\ell_n]} |\ell^{-1} \zeta_{m,\ell} - \zeta_{m,1}| &\lesssim \max_{\ell \in [1:\ell_n]} (\ell/m^2)^{1-2/a} \\ &\leq n^{2/a-1} + (2^{-1} \ln(n))^{1/2-1/a} m^{3/a-3/2} = o(1/m). \end{aligned}$$

We conclude that sub-goal (B.3) is attained. $\square$

We now list Lemmas B.4, B.5 and B.6, providing the bounds (B.12). The proofs of these lemmas rely on counting variable inequalities which are developed afterwards in Lemmas B.7 to B.11.

Lemma B.4 (Handling $\mathbb{I}_1$ in (B.11)). *If Condition 1 is satisfied, then*

$$\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] = \mathbb{P}(N_{\ell,m} = 0) \theta^2.$$

In addition,

$$\mathbb{I}_1 \leq \theta^2 \cdot \frac{q^2 \ell (\ell - 1)}{(m - 1)(m - \ell + 1)}.$$

Proof. Let $r = m - \ell$. If $N_{\ell,m} = 0$, the top- q of $Z_{[1:m]}$ is the same as the top- q of $Z_{[1:(m-\ell)]}$ and the top- q of $Z_{[(m-\ell+1):(2m-\ell)]}$ is the same as the one of $Z_{[(m+1):(2m-\ell)]}$. The event $\{N_{\ell,m} = 0\}$ is a function of the random variables $Z_{[(r+1):m]}$ and the q -largest order statistics of $Z_{[1:r]}$ and $Z_{[(m+1):(m+r)]}$. These statements taken together with a slight extension of the properties of the top order statistics of a generalised Pareto sample in Lemma A.2 imply

$$\begin{aligned}\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] &= \mathbb{E}[K_r(Z_{[1:r]})K_r(Z_{[(m+1):(m+r)]}); N_{\ell,m} = 0] \\ &= \mathbb{P}(N_{\ell,m} = 0) \theta^2.\end{aligned}$$

Thanks to that equality, we find

$$\begin{aligned}|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] - \ell \mathbb{E}[\Xi_{1,m}; N_{1,m} = 0] + (\ell - 1)\theta^2| \\ = |\mathbb{P}(N_{\ell,m} = 0) - \ell \mathbb{P}(N_{1,m} = 0) + \ell - 1| \theta^2 \\ = |\ell \mathbb{P}(N_{1,m} \geq 1) - \mathbb{P}(N_{\ell,m} \geq 1)| \theta^2.\end{aligned}$$

Apply Lemma B.10 to conclude. $\square$

In case $N_{\ell,m} = 1$, exactly one of the ℓ indicators $\mathbb{1}_{i,\ell,m}$ in (B.10) for $i \in [(m-\ell+1):m]$ is equal to one while the other $\ell-1$ indicators are zero. We find

$$\begin{aligned}\{N_{\ell,m} = 1\} &= \bigcup_{i=m-\ell+1}^m B_{i,\ell,m} \quad \text{where} \\ B_{i,\ell,m} &= \{\mathbb{1}_{i,\ell,m} = 1\} \cap \bigcap_{j \in [(m-\ell+1):m] \setminus \{i\}} \{\mathbb{1}_{j,\ell,m} = 0\},\end{aligned}\tag{B.13}$$

and the union over i being disjoint. Note that $B_{i,\ell,m}$ is the event that Z_i is in the top- q of $Z_{[1:m]}$ or $Z_{[(m-\ell+1):(2m-\ell)]}$ while the other variables Z_j for $j \in [(m-\ell+1):m] \setminus \{i\}$ are not. Further, $\{N_{1,m} = 1\} = \{\mathbb{1}_{m,1,m} = 1\}$.

Lemma B.5 (Handling $\mathbb{I}_2$ in (B.11)). *If Condition 1 is satisfied, we have*

$$\mathbb{I}_2 \leq 4(5q^2 - q)^{1-2/a} \|K\|_a^2 \ell \left(\frac{(\ell-1)}{(m-\ell+1)^2} \right)^{1-2/a},$$

for all $a > 2$, provided that m is sufficiently large.

Proof. By symmetry and since the union in (B.13) is disjoint,

$$\begin{aligned}\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] &= \sum_{i=m-\ell+1}^m \mathbb{E}[\Xi_{\ell,m}; B_{i,\ell,m}] \\ &= \ell \mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}].\end{aligned}$$

On $B_{m,\ell,m}$, the top- q order statistics of $Z_{[(m-\ell+1):(2m-\ell)]}$ are the same as those of $Z_{[m:(2m-\ell)]}$, which implies $K_m(Z_{[(m-\ell+1):(2m-\ell)]}) = K_{m-\ell+1}(Z_{[m:(2m-\ell)]})$ and thus

$$\mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}] = \mathbb{E}[K_m(Z_{[1:m]}) K_{m-\ell+1}(Z_{[m:(2m-\ell)]}); B_{m,\ell,m}].$$

On the right-hand side, we replace the indicator of the event $B_{m,\ell,m}$ first by $\mathbb{1}_{m,\ell,m}$ and then by $\mathbb{1}_{m,1,m}$, using Lemmas B.11 and B.9, respectively. We find, for $a > 2$, by the triangle inequality and Hölder's inequality,

$$\begin{aligned} & \left| \mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}] - \mathbb{E}[K_m(Z_{[1:m]}) K_{m-\ell+1}(Z_{[m:(2m-\ell)]}) \mathbb{1}_{m,1,m}] \right| \\ & \leq \|K\|_a^2 \mathbb{E}[\|\mathbb{1}_{B_{m,\ell,m}} - \mathbb{1}_{m,\ell,m}\|]^{1-2/a} + \|K\|_a^2 \mathbb{E}[\|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}\|]^{1-2/a} \\ & \leq \|K\|_a^2 \left[\left(2(2q^2 - q) \frac{\ell - 1}{(m-1)(m-\ell+1)} \right)^{1-2/a} + \left((q+1)q \frac{\ell - 1}{(m+1)m} \right)^{1-2/a} \right] \\ & \leq 2(5q^2 - q)^{1-2/a} \|K\|_a^2 \left(\frac{\ell - 1}{(m-1)(m-\ell+1)} \right)^{1-2/a}. \end{aligned}$$

The top- q order statistics of $Z_{[m:(2m-\ell)]}$ are the same as those of $Z_{[m:(2m-1)]}$ unless at least one of the variables Z_j for $j \in [(2m-\ell+1):(2m-1)]$ is larger than the q -largest order statistic of $Z_{[m:(2m-\ell)]}$. Let $C_{\ell,m}$ denote the latter event. It follows that, for $a > 2$,

$$\begin{aligned} & \left| \mathbb{E}[K_m(Z_{[1:m]}) K_{m-\ell+1}(Z_{[m:(2m-\ell)]}) \mathbb{1}_{m,1,m}] \right. \\ & \quad \left. - \mathbb{E}[K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]}) \mathbb{1}_{m,1,m}] \right| \\ & \leq \mathbb{E}[\|K_m(Z_{[1:m]})\| \|K_{m-\ell+1}(Z_{[m:(2m-\ell)]}) - K_m(Z_{[m:(2m-1)]})\| \mathbb{1}_{m,1,m}] \\ & \leq 2\|K\|_a^2 \mathbb{E}[\mathbb{1}_{m,1,m}; C_{\ell,m}]^{1-2/a}. \end{aligned}$$

Further, we have

$$\begin{aligned} & \mathbb{E}[\mathbb{1}_{m,1,m}; C_{\ell,m}] \\ & = \mathbb{P}\left(\{Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ and/or } Z_{[m:(2m-1)]}\} \right. \\ & \quad \left. \cap \bigcup_{j=2m-\ell+1}^{2m-1} \{Z_j \text{ is larger than the } q\text{-largest order statistic of } Z_{[m:(2m-\ell)]}\}\right). \end{aligned}$$

The latter probability is bounded by $\ell - 1$ times the sum

$$\begin{aligned} & \mathbb{P}(Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ and} \\ & \quad Z_{2m-\ell+1} \text{ is in the top-}q \text{ of } Z_{[(m+1):(2m-\ell+1)]}) \\ & + \mathbb{P}(Z_m \text{ and } Z_{2m-\ell+1} \text{ are both in the top-}q \text{ of } Z_{[m:(2m-\ell+1)]}) \end{aligned}$$

and thus by

$$(\ell - 1) \left(\frac{q}{m} \cdot \frac{q}{m - \ell + 1} + \frac{q(q-1)}{(m-\ell+2)(m-\ell+1)} \right) \leq (2q^2 - q) \frac{\ell - 1}{(m - \ell + 1)^2}.$$

Since,

$$\mathbb{E}[K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]}) \mathbb{1}_{m,1,m}] = \mathbb{E}[\Xi_{1,m}; N_{1,m} = 1],$$

a combination of the previous inequalities yields

$$\begin{aligned}
 & |\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] - \ell \mathbb{E}[\Xi_{1,m}; N_{1,m} = 1]| \\
 &= \ell |\mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}] - \mathbb{E}[K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]}) \mathbf{1}_{m,1,m}]| \\
 &\leq \ell \|K\|_a^2 \left[2(5q^2 - q)^{1-2/a} \left(\frac{\ell - 1}{(m-1)(m-\ell+1)} \right)^{1-2/a} \right. \\
 &\quad \left. + 2(2q^2 - q)^{1-2/a} \left(\frac{\ell - 1}{(m-\ell+1)^2} \right)^{1-2/a} \right] \\
 &\leq 4(5q^2 - q)^{1-2/a} \|K\|_a^2 \ell \left(\frac{(\ell - 1)}{(m-\ell+1)^2} \right)^{1-2/a}. \square
 \end{aligned}$$

Lemma B.6 (Handling $\mathbb{I}_3$ in (B.11)). *If Condition 1 is satisfied, then*

$$\mathbb{I}_3 \leq \|K\|_a^2 \left(\frac{\ell(\ell - 1)}{2} \right)^{1-2/a} \left(\frac{2(2q^2 - q)}{(m-1)(m-\ell+1)} \right)^{1-2/a}$$

for all $a > 2$.

Proof. Since the kernel satisfies Condition 1, we have by Hölder's inequality

$$|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \geq 2]| \leq \mathbb{E}[|\Xi_{\ell,m}|; N_{\ell,m} \geq 2] \leq \|K\|_a^2 \mathbb{P}(N_{\ell,m} \geq 2)^{1-2/a}$$

for $a > 2$. By Lemma B.7, we have

$$\mathbb{P}(N_{\ell,m} \geq 2) \leq \frac{\ell(\ell - 1)}{2} \mathbb{E}[\mathbf{1}_{m-1,\ell,m} \mathbf{1}_{m,\ell,m}].$$

In view of Lemma B.8, we conclude

$$|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \geq 2]| \leq \|K\|_a^2 \left(\frac{\ell(\ell - 1)}{2} \right)^{1-2/a} \left(\frac{2(2q^2 - q)}{(m-1)(m-\ell+1)} \right)^{1-2/a}. \square$$

We now develop the required counting variable inequalities. In doing so, the following preliminary lemma will be useful a number of times. No originality is claimed.

Lemma B.7. *Consider a counting variable $N = \sum_{i=1}^n \mathbf{1}_i$ with indicator variables $\mathbf{1}_1, \dots, \mathbf{1}_n$ satisfying $\mathbb{E}[\mathbf{1}_i \mathbf{1}_j] \leq \varepsilon$ for all $i, j \in [1:n]$ and $i \neq j$. Then*

$$\begin{aligned}
 2\mathbb{P}(N \geq 2) &\leq \mathbb{E}(N) - \mathbb{P}(N = 1) \\
 &\leq 2(\mathbb{E}(N) - \mathbb{P}(N \geq 1)) \\
 &\leq n(n-1)\varepsilon.
 \end{aligned}$$

Proof. A case-by-case analysis reveals

$$\mathbf{1}\{N \geq 2\} \leq \frac{1}{2}(N - \mathbf{1}\{N = 1\}) \leq N - \mathbf{1}\{N \geq 1\} \leq \frac{1}{2}N(N-1)$$

and thus

$$\mathbb{P}(N \geq 2) \leq \frac{1}{2} (\mathbb{E}(N) - \mathbb{P}(N = 1)) \leq \mathbb{E}(N) - \mathbb{P}(N \geq 1) \leq \frac{1}{2} (\mathbb{E}(N^2) - \mathbb{E}(N)).$$

But since $\mathbb{1}_i^2 = \mathbb{1}_i$, we have

$$\mathbb{E}(N^2) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[\mathbb{1}_i \mathbb{1}_j] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_i] + \sum_{i,j \in [1:n], i \neq j} \mathbb{E}[\mathbb{1}_i \mathbb{1}_j] \leq \mathbb{E}(N) + n(n-1)\varepsilon,$$

and thus

$$\frac{1}{2} (\mathbb{E}(N^2) - \mathbb{E}(N)) \leq \frac{n(n-1)}{2} \varepsilon,$$

yielding the stated inequalities. $\square$

Recall the indicator variables $\mathbb{1}_{i,\ell,m}$ in (B.10).

Lemma B.8. *We have*

$$\mathbb{E}[\mathbb{1}_{m-1,\ell,m} \mathbb{1}_{m,\ell,m}] \leq \frac{2(2q^2 - q)}{(m-1)(m-\ell+1)}.$$

Proof. By symmetry,

$$\begin{aligned} & \mathbb{E}[\mathbb{1}_{m-1,\ell,m} \mathbb{1}_{m,\ell,m}] \\ &= \mathbb{P}(Z_{m-1} \text{ and } Z_m \text{ are both in the top-}q \text{ of } Z_{[1:m]} \text{ or } Z_{[(m-\ell+1):(2m-\ell)]}) \\ &\leq 2 \mathbb{P}(Z_{m-1} \text{ and } Z_m \text{ are both in the top-}q \text{ of } Z_{[1:m]}) \\ &\quad + 2 \mathbb{P}(Z_{m-1} \text{ is in the top-}q \text{ of } Z_{[1:m]} \\ &\quad \text{and } Z_m \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]}). \end{aligned}$$

The first probability on the right-hand side is equal to $q(q-1)/[m(m-1)]$ while the second one is bounded by

$$\begin{aligned} & \mathbb{P}(Z_{m-1} \text{ is in the top-}q \text{ of } Z_{[1:(m-1)]} \text{ and } Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-\ell)]}) \\ &= \frac{q}{m-1} \cdot \frac{q}{m-\ell+1}. \end{aligned}$$

Add the two bounds and note that $1/m < 1/(m-\ell+1)$ to conclude the proof. $\square$

Lemma B.9. *We have*

$$\mathbb{E}[|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}|] \leq (q+1)q \frac{\ell-1}{(m+1)m}.$$

Proof. For events A and B , we have $|\mathbb{1}_A - \mathbb{1}_B| = \mathbb{1}_{A \triangle B}$ where the symmetric difference $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the event that exactly one of A and B occurs. It follows that $|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}|$ is the indicator of the event that exactly one of the two events

$\{Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ or } Z_m \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]}\}$ and

$\{Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ or } Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-1)]}\}$

occurs. In that case, Z_m is either in the top- q of $Z_{[(m-\ell+1):(2m-\ell)]}$ but not in the top- q of $Z_{[m:(2m-1)]}$ or the other way around, and thus

$$\begin{aligned} & \mathbb{E} [|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}|] \\ & \leq \mathbb{P}(Z_m \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]} \text{ but not in that of } Z_{[m:(2m-1)]}) \\ & + \mathbb{P}(Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-1)]} \text{ but not in that of } Z_{[(m-\ell+1):(2m-\ell)]}) \\ & = 2 \mathbb{P}(Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-1)]} \text{ but not in that of } Z_{[(m-\ell+1):(2m-\ell)]}). \end{aligned}$$

For the probability of the latter event, note that there always needs to be at least one of the $\ell - 1$ random variables of $Z_{[(m-\ell+1):(m-1)]}$ larger than Z_m for Z_m not to be in the top- q of $Z_{[(m-\ell+1):(2m-\ell)]}$. Thus, by the union bound, the probability is bounded by $(\ell - 1)$ times the probability that Z_{m-1} and Z_m are both in the top- $(q+1)$ of $Z_{[(m-1):(2m-1)]}$ while Z_{m-1} is still larger than Z_m . It follows that

$$|\mathbb{E}[\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}]| \leq 2(\ell - 1) \cdot \frac{(q+1)q}{2(m+1)m} = (q+1)q \frac{\ell - 1}{(m+1)m}. \quad \square$$

Lemma B.10. *We have*

$$|\mathbb{P}(N_{\ell,m} \geq 1) - \ell \mathbb{P}(N_{1,m} = 1)| \leq 3q^2 \cdot \frac{\ell(\ell - 1)}{(m-1)(m-\ell+1)}.$$

Proof. In view of Lemma B.7, we have

$$|\mathbb{P}(N_{\ell,m} \geq 1) - \ell \mathbb{E}[\mathbb{1}_{m,\ell,m}]| \leq \frac{\ell(\ell - 1)}{2} \mathbb{E}[\mathbb{1}_{m-1,\ell,m} \mathbb{1}_{m,\ell,m}]$$

and thus, since $N_{1,m} = \mathbb{1}_{m,1,m}$,

$$\begin{aligned} |\mathbb{P}(N_{\ell,m} \geq 1) - \ell \mathbb{P}(N_{1,m} = 1)| & \leq \frac{\ell(\ell - 1)}{2} \mathbb{E}[\mathbb{1}_{m-1,\ell,m} \mathbb{1}_{m,\ell,m}] \\ & + \ell \mathbb{E} [|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}|]. \end{aligned}$$

The two expectations on the right-hand side were treated in Lemmas B.8 and B.9, respectively. Combine the bounds to obtain that

$$\begin{aligned} & |\mathbb{P}(N_{\ell,m} \geq 1) - \ell \mathbb{P}(N_{1,m} = 1)| \\ & \leq \frac{\ell(\ell - 1)}{2} \cdot \frac{2(2q^2 - q)}{(m-1)(m-\ell+1)} + \ell \cdot (q+1)q \frac{\ell - 1}{(m+1)m} \\ & \leq \frac{3q^2 \ell(\ell - 1)}{(m-1)(m-\ell+1)}. \quad \square \end{aligned}$$

Lemma B.11. *We have*

$$\mathbb{E} [|\mathbb{1}_{B_{m,\ell,m}} - \mathbb{1}_{m,\ell,m}|] \leq 2(2q^2 - q) \frac{\ell - 1}{(m-1)(m-\ell+1)}.$$

Proof. We have $\mathbb{1}_{m,\ell,m} \geq \mathbb{1}_{B_{m,\ell,m}}$ while the difference between the two indicators is equal to the indicator of the event that $\mathbb{1}_{m,\ell,m} = 1$ holds together with $\mathbb{1}_{j,\ell,m} = 1$ for some $j \in [(m - \ell + 1):(m - 1)]$. By the union bound, we obtain

$$\mathbb{E} [|\mathbb{1}_{B_{m,\ell,m}} - \mathbb{1}_{m,\ell,m}|] \leq (\ell - 1) \mathbb{E} [\mathbb{1}_{m-1,\ell,m} \mathbb{1}_{m,\ell,m}].$$

Apply Lemma B.8 to conclude. $\square$

Appendix C: Proof of Proposition 2.3

Proof of Proposition 2.3. The proof is divided in two parts. The first part will deal with the expression for the asymptotic variance in (2.11) and the second part with the asymptotic normality of the Hájek projection in (2.12).

Part 1: asymptotic variance Write $\overline{K}_m := K_m - u(\gamma)$. We start by calculating the limit

$$\lim_{r \rightarrow \infty} r \zeta_{r+1,1}, \quad \text{where} \quad \zeta_{r+1,1} = \mathbb{E} [\overline{K}_{r+1}(Z_1, \dots, Z_r, Z) \cdot \overline{K}_{r+1}(Z'_1, \dots, Z'_r, Z)]$$

and $Z, Z_1, Z'_1, Z_2, Z'_2, \dots$ are independent $\text{GP}(\gamma)$ random variables. The expectation defining $\zeta_{r+1,1}$ is finite by Condition 1 and the Cauchy–Schwarz inequality. We will prove that

$$\lim_{r \rightarrow \infty} r \zeta_{r+1,1} = \sigma_{\gamma,K}^2 = \int_0^\infty (\mathbb{E} [\overline{K}_{q+1}((Z_{q-j:q-1})_{j=1}^q, h_\gamma(S_q/x))])^2 dx < \infty, \quad (\text{C.1})$$

where $Z_{0:q-1} = 0$ and $Z_{1:q-1} \leq \dots \leq Z_{q-1:q-1}$ are the order statistics of an independent $\text{GP}(\gamma)$ random sample $Z_1, \dots, Z_{q-1}$, while $S_q = E_1 + \dots + E_q$ with $E_1, \dots, E_q$ an independent unit-exponential random sample, independent of $Z_1, \dots, Z_{q-1}$.

To prove (C.1), we will show that we can rewrite $r \zeta_{r+1,1}$ as

$$r \zeta_{r+1,1} = \int_0^\infty (\mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)) \cdot \mathbb{1}\{rU_{q:r} > x\}])^2 dx, \quad (\text{C.2})$$

where $Y_{0:q-1} := 1$, $Y_{1:q-1} \leq \dots \leq Y_{q-1:q-1}$ are the order statistics of a standard Pareto random sample of size $q - 1$ and $U_{1:r} \leq \dots \leq U_{r:r}$ are the order statistics of a $\text{uniform}(0, 1)$ random sample of size r , the two samples being independent. Since $rU_{q:r}$ converges in distribution to S_q , the remainder of the proof then consists in justifying why the limit can be interchanged with the expectation and the integral. We will do this in two steps: first we will show that the expectation inside the integral in (C.2) converges, i.e., for all $x > 0$, we have the convergence

$$\begin{aligned} \lim_{r \rightarrow \infty} \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)) \cdot \mathbb{1}\{rU_{q:r} > x\}] \\ = \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(S_q/x)) \cdot \mathbb{1}\{S_q > x\}], \quad (\text{C.3}) \end{aligned}$$

where S_q is as in (C.1). Second, we will apply Lebesgue's dominated convergence theorem to show that the integral itself converges.

We start with the proof of (C.2). Condition on Z to get

$$\begin{aligned} r \zeta_{r+1,1} &= \int_0^\infty \mathbb{E} [\overline{K}_{r+1}(Z_1, \dots, Z_r, z) \cdot \overline{K}_{r+1}(Z'_1, \dots, Z'_r, z)] \cdot r \cdot d\mathbb{P}(Z \leq z) \\ &= \int_0^\infty (\mathbb{E} [\overline{K}_{r+1}(Z_1, \dots, Z_r, z)])^2 \cdot r \cdot d\mathbb{P}(Z \leq z) \\ &= \int_0^\infty (\mathbb{E} [\overline{K}_{q+1}((Z_{r-j+1:r})_{j=1}^q, z)])^2 \cdot r \cdot d\mathbb{P}(Z \leq z), \end{aligned}$$

in terms of the ascending order statistics $Z_{1:r} \leq \dots \leq Z_{r:r}$ of $Z_1, \dots, Z_r$. Here we used the fact that the kernel K_m only depends on the top- q values of its argument. Note that if Y is a standard Pareto random variable, then the distribution of $h_\gamma(Y)$ is GP(γ). Let $Y_{1:r} \leq \dots \leq Y_{r:r}$ be the ascending order statistics of a standard Pareto random sample $Y_1, \dots, Y_r$. It follows that

$$\begin{aligned} r \zeta_{r+1,1} &= \int_1^\infty (\mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{r-j+1:r}))_{j=1}^q, h_\gamma(y))])^2 \cdot r \frac{dy}{y^2} \\ &= \int_{1/r}^\infty (\mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{r-j+1:r}))_{j=1}^q, h_\gamma(rt))])^2 \frac{dt}{t^2}, \end{aligned}$$

where we have substituted $y = rt$ in the last step. The vector of top- q standard Pareto order statistics satisfies the distributional representation

$$(Y_{r-j+1:r})_{j=1}^q \stackrel{d}{=} (Y_{q-j:q-1}/U_{q:r})_{j=1}^q$$

where $Y_{0:q-1} := 1$ and where $U_{1:r} \leq \dots \leq U_{r:r}$ denote the ascending order statistics of an independent uniform(0,1) sample $U_1, \dots, U_r$, independent of $Y_1, \dots, Y_r$; note that $1/U_{q:r}$ has the same distribution as $Y_{r-q+1:r}$. The function h_γ satisfies the functional equation

$$\forall x, y > 0, \quad h_\gamma(y/x) = \frac{h_\gamma(y) - h_\gamma(x)}{x^\gamma}.$$

By location-scale invariance of the kernel,

$$\begin{aligned} &\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}/U_{q:r}))_{j=1}^q, h_\gamma(rt)) \\ &= \overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, U_{q:r}^\gamma h_\gamma(rt) + h_\gamma(U_{q:r})) \\ &= \overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}t)). \end{aligned}$$

Following the substitution $x = 1/t$, we obtain

$$\begin{aligned} r \zeta_{r+1,1} &= \int_{1/r}^\infty (\mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}t))])^2 \frac{dt}{t^2} \\ &= \int_0^r (\mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x))])^2 dx. \end{aligned}$$

By independence and Fubini's theorem, the expectation inside the integral is equal to

$$\begin{aligned} & \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x))] \\ &= \int_0^1 \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(ru/x))] \, d\mathbb{P}(U_{q:r} \leq u). \end{aligned}$$

The expectation inside the integral on the right-hand side is zero as soon as $ru/x \leq 1$, because in that case

$$\begin{aligned} & \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(ru/x))] \\ &= \mathbb{E} [\overline{K}_q((h_\gamma(Y_{q-j:q-1}))_{j=1}^q)] = \mathbb{E} [\overline{K}_q((Z_{q-j:q-1})_{j=1}^q)] = 0. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} & \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x))] \\ &= \mathbb{E} [\overline{K}_{q+1}((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)) \cdot \mathbf{1}\{rU_{q:r} > x\}]. \end{aligned}$$

Equation (C.2) follows as the integral is equal to 0 for values of x higher than r .

Next, we show (C.3) and eventually (C.1). The well-known representation of uniform order statistics in terms of partial sums of unit-exponential random variables, i.e., $U_{q:r} \stackrel{d}{=} S_q/S_{r+1}$ where, for any integer a , $S_a = E_1 + \dots + E_a$ is the sum of independent standard exponential random variables independent of $Y_1, \dots, Y_{q-1}$, implies the weak convergence

$$rU_{q:r} \xrightarrow{d} S_q = E_1 + \dots + E_q, \quad r \rightarrow \infty. \quad (\text{C.4})$$

The question is whether in (C.3), we can switch the limit and the expectation.

Under a suitable Skorokhod construction, we can assume that (C.4) holds almost surely. By continuity of the kernel, the integrand inside the expectation on the left-hand side then converges almost surely to the integrand on the right-hand side. To show uniform integrability in (C.3) and eventually the validity of switching the limit and integral in (C.1), it is sufficient to show that the expectation of the square is bounded by an integrable function $c(x)$, i.e.,

$$\limsup_{r \rightarrow \infty} \mathbb{E} [\overline{K}_{q+1}^2((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)) \cdot \mathbf{1}\{rU_{q:r} > x\}] < c(x) < \infty,$$

where $c(x)$ is a function on $[0, +\infty)$ such that $\int_0^{+\infty} c(x)dx < \infty$. Then not only the relation (C.3) follows immediately, but also the relation (C.1) follows from the Lebesgue dominance convergence theorem.

To do so, we rely on a change-of-measure, replacing $rU_{q:r}/x$ on the event $\{rU_{q:r} > x\}$ by a standard Pareto random variable, and compensating by the likelihood ratio. For $0 < x < r$, let $p_{r,x}$ be the probability density function of $rU_{q:r}/x$ and let Y be a standard Pareto random variable, independent of

$Y_1, \dots, Y_{q-1}$. We have

$$\begin{aligned} & \mathbb{E} \left[\overline{K}_{q+1}^2 ((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)) \cdot \mathbb{1}\{rU_{q:r} > x\} \right] \\ &= \int_1^\infty \mathbb{E} \left[\overline{K}_{q+1}^2 ((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(y)) \right] \cdot p_{r,x}(y) \, dy \\ &= \int_1^\infty \mathbb{E} \left[\overline{K}_{q+1}^2 ((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(y)) \right] \cdot p_{r,x}(y) \cdot y^2 \frac{dy}{y^2} \\ &= \mathbb{E} \left[\overline{K}_{q+1}^2 ((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(Y)) \cdot p_{r,x}(Y) \cdot Y^2 \right] \\ &= \mathbb{E} \left[\overline{K}_q^2 (h_\gamma(Y_1), \dots, h_\gamma(Y_{q-1}), h_\gamma(Y)) \cdot p_{r,x}(Y) \cdot Y^{2-\epsilon} \cdot Y^\epsilon \right], \end{aligned}$$

for some $0 < \epsilon < 1/2$.

Recall that the distribution of $(h_\gamma(Y_1), \dots, h_\gamma(Y_{q-1}), h_\gamma(Y))$ is equal to the one of an independent $\text{GP}(\gamma)$ random sample of size q . In view of Condition 1, and the fact that $\mathbb{E}[Y^{2\epsilon}] < \infty$, it is sufficient to show that

$$\sup_{r > 2q} \sup_{y \geq 1} p_{r,x}(y) \cdot y^{2-\epsilon} \leq \tilde{c}(x),$$

for some finite function $\tilde{c}(x)$ such that $\int_0^{+\infty} \tilde{c}(x) dx < \infty$. Since the distribution of $U_{q:r}$ is $\text{Beta}(q, r+1-q)$, we have, for $y \in [1, \infty)$,

$$\begin{aligned} p_{r,x}(y) &= \frac{d}{dy} \mathbb{P}(rU_{q:r}/x \leq y) \\ &= \frac{d}{dy} \mathbb{P}(U_{q:r} \leq xy/r) \\ &= (x/r) \cdot \frac{\Gamma(r+1)}{\Gamma(q)\Gamma(r+1-q)} \cdot (xy/r)^{q-1} \cdot (1-xy/r)^{r-q} \cdot \mathbb{1}\{xy < r\} \\ &= \frac{x^q}{(q-1)!} \cdot \frac{r!}{(r-q)!r^q} \cdot y^{q-1} \cdot (1-xy/r)^{r-q} \cdot \mathbb{1}\{xy < r\}. \end{aligned} \quad (\text{C.5})$$

As $1-h \leq e^{-h}$ for real h , we get, for $r > 2q$, the inequalities

$$p_{r,x}(y) \leq \frac{x^q}{(q-1)!} \cdot y^{q-1} \cdot e^{-(r-q)xy/r} \leq \frac{x^q}{(q-1)!} \cdot y^{q-1} \cdot e^{-xy/2}.$$

Hence,

$$p_{r,x}(y) \cdot y^{2-\epsilon} \leq \frac{x^q}{(q-1)!} \cdot y^{q+1-\epsilon} \cdot e^{-xy/2}, \quad \text{for all } r \geq q, \text{ all } x > 0 \text{ and all } y \geq 1.$$

With straightforward calculation, we get that the upperbound is maximized at $y = 1 \vee 2(q+1-\epsilon)/x$. It follows that we can choose

$$\tilde{c}(x) = \begin{cases} \frac{(2(q+1-\epsilon))^{q+1-\epsilon}}{(q-1)!e^{q+1-\epsilon}} \cdot x^{-1+\epsilon} & \text{if } 0 < x \leq 2(q+1-\epsilon), \\ \frac{x^q}{(q-1)!} \cdot e^{-x/2} & \text{if } x > 2(q+1-\epsilon), \end{cases}$$

which satisfies $\int_0^\infty \tilde{c}(x) dx < \infty$. This concludes the proof of (C.3) and (C.1).

Part 2: asymptotic normality It remains to show that the Hájek projection $\hat{U}_{n,Z}^m$ in (2.6) of $U_{n,Z}^m - \theta$ is asymptotically normal. Note that $\hat{U}_{n,Z}^m$ is a centered row sum of a triangular array of row-wise i.i.d. random variables, with total variance $\zeta_{m,1}m^2/n$. If Lyapunov's condition is met, following the central limit theorem, we have

$$\frac{k}{\sqrt{n\zeta_{m,1}}} \hat{U}_{n,Z}^m \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty.$$

Therefore, we need to show that for some $\delta > 0$ we have

$$\lim_{m \rightarrow \infty} \frac{1}{(\sqrt{n\zeta_{m,1}})^{2+\delta}} \cdot n \mathbb{E} \left[|\hat{K}_m(Z)|^{2+\delta} \right] = 0. \quad (\text{C.6})$$

First, by Jensen's inequality,

$$|\hat{K}_m(z)|^{2+\delta} \leq \mathbb{E} \left[|\bar{K}_m(Z_1, \dots, Z_{m-1}, z)|^{2+\delta} \right].$$

Second, by Fubini's theorem, location-scale invariance of the kernel and the stability property of GP order statistics,

$$\mathbb{E} \left[|\hat{K}_m(Z)|^{2+\delta} \right] \leq \mathbb{E} \left[|\bar{K}_m(Z_1, \dots, Z_m)|^{2+\delta} \right] = \mathbb{E} \left[|\bar{K}_q(Z_1, \dots, Z_q)|^{2+\delta} \right],$$

which implies that the term $\mathbb{E}[|\hat{K}_m(Z)|^{2+\delta}]$ is uniformly bounded over all m .

Since $\lim_{m \rightarrow \infty} m\zeta_{m,1} = \sigma_K^2(\gamma) > 0$, the relation (C.6) holds provided we can choose $\delta > 0$ such that

$$\lim_{n \rightarrow \infty} \left(\frac{m}{n} \right)^{(2+\delta)/2} = 0.$$

From Condition 2, this can be achieved by choosing $\delta > 2/\epsilon$. $\square$

Appendix D: Proof of Theorem 3.2

We start by listing two preliminary lemmas.

Lemma D.1 (Potter's bounds; see [4, Proposition B.1.9]). *Let $f \in RV_\alpha$. For all $\delta_1, \delta_2 > 0$ there exists $t_0 = t_0(\delta_1, \delta_2)$ such that for any positive t and x satisfying $t \geq t_0$ and $tx \geq t_0$,*

$$(1 - \delta_1)x^\alpha \min(x^{\delta_2}, x^{-\delta_2}) \leq \frac{f(tx)}{f(t)} \leq (1 + \delta_1)x^\alpha \max(x^{\delta_2}, x^{-\delta_2}).$$

Lemma D.2. *Let $Y_{1:m}^* \leq Y_{2:m}^* \leq \dots \leq Y_{m:m}^*$ be the order statistics of an independent random sample of Pareto(1) random variables. For integer $q \geq 1$ and for a real number $\nu < q$, the sequence $\{(Y_{m-q+1:m}^*/m)^\nu : m \geq m_0\}$ is uniformly integrable for some m_0 depending on q and ν .*

Proof of Lemma D.2. Note that $Y_{m-q+1:m}^* \stackrel{d}{=} \frac{S_{m+1}}{S_q}$, where $S_l = \sum_{i=1}^l E_i$ for independent unit-exponential random variables E_i . Since it is only the distribution of $Y_{m-q+1:m}^*$ that matters, we can change probability spaces to ensure that the equality in distribution is in fact an equality between random variables.

We first deal with $(Y_{m-q+1:m}^*/m)^\nu$ for $\nu \leq 0$. We have

$$(Y_{m-q+1:m}^*/m)^\nu \leq \left(\frac{S_q}{(m+1)^{-1} \sum_{j=1}^{m+1} E_j} \right)^{-\nu}.$$

Note that $H_m = (m+1)^{-1} \sum_{j=1}^{m+1} E_j$ is distributed as a Gamma($m+1, m+1$) random variable with value $m+1$ for both the shape and rate parameter. By Cauchy's inequality,

$$\mathbb{E}[(Y_{m-q+1:m}^*/m)^{2\nu}] \leq \sqrt{\mathbb{E}[H_m^{2\nu}]} \sqrt{\mathbb{E}[S_q^{-2\nu}]}.$$

We have $\mathbb{E}[S_q^{-2\nu}] < \infty$ for all $\nu \leq 0$ while by direct calculation, provided $m > -2\nu - 1$,

$$\mathbb{E}[H_m^{2\nu}] = \frac{(m+1)^{-2\nu} \Gamma(m+2\nu+1)}{\Gamma(m+1)} \rightarrow 1 \text{ as } m \rightarrow \infty.$$

Therefore, there exists $m_0(\nu)$ such that the sequence $\{(Y_{m-q+1:m}^*/m)^\nu, m \geq m_0(\nu)\}$ is bounded in $\mathcal{L}^2$ and thus uniformly integrable.

Next we consider $(Y_{m-q+1:m}^*/m)^\nu$ for $0 < \nu < q$. Define $p_0 = (3\nu + q)/4$ and note that $\nu < p_0 < (\nu + q)/2$. Also, define $p_1 = (\nu + q)/(2p_0)$, which satisfies $p_1 > 1$, and define $z > 1$ by $1/z + 1/p_1 = 1$. For sufficiently large integer m , we find, using Hölder's inequality and the inequality $1/m \leq 2/(m+1)$, that

$$\begin{aligned} \mathbb{E}[(Y_{m-q+1:m}^*/m)^{p_0}] &\leq 2^{p_0} \mathbb{E}[(H_m/S_q)^{p_0}] \\ &\leq 2^{p_0} (\mathbb{E}[H_m^{zp_0}])^{1/z} (\mathbb{E}[S_q^{-p_1 p_0}])^{1/p_1}. \end{aligned}$$

From $p_0 p_1 = (\nu + q)/2 < q$ it follows that $\mathbb{E}[S_q^{-p_0 p_1}] < \infty$. Similar to the case where $\nu < 0$, for any $a > 0$ we have $\sup_{m \geq 1} \mathbb{E}[H_m^a] < \infty$. It follows that

$$\sup_{m \geq q} \mathbb{E}[(Y_{m-q+1:m}^*/m)^{p_0}] < \infty.$$

Since $p_0 > \nu > 0$, the sequence $\{(Y_{m-q+1:m}^*/m)^\nu, m \geq q\}$ is uniformly integrable. $\square$

Proof of Theorem 3.2. Following Theorem 3.1, we only need to show that Condition 5 holds with the function $B_K(\rho, \gamma)$ given in (3.7).

Recall from the discussion in Section 3 that

$$\sqrt{k} (U_n^m - U_{n,Z}^m) = \sqrt{k} \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} (K_m(X_I) - K_m(Z_I))$$

$$\begin{aligned}
&= \sqrt{k} A(m) \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} \xi_m^I \\
&=: \sqrt{k} A(m) U_{n,\xi}^m,
\end{aligned}$$

where $\xi_m^I = K_m(X_I) - K_m(Z_I)$. In what follows, we write ξ_m to indicate a generic ξ_m^I where the index set is not important.

By Hoeffding's decomposition,

$$\text{var}(U_{n,\xi}^m) = \sum_{\ell=1}^m p_{n,m}(\ell) \Psi_{m,\ell}$$

with $p_{n,m}(\ell)$ in (B.2) and $\Psi_{m,\ell} = \text{cov}(\xi_m^I, \xi_m^{I'})$ for index sets $I, I' \subset [n]$ such that $|I'| = |I| = m$ and $|I \cap I'| = \ell$. Below we will show the limit relation

$$\sum_{\ell=1}^m p_{n,m}(\ell) \Psi_{m,\ell} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (\text{D.1})$$

Then it follows from Chebychev's inequality that

$$U_{n,\xi}^m - \mathbb{E}[\xi_m] \xrightarrow{p} 0, \quad \text{as } n \rightarrow \infty.$$

In addition, we will show that

$$\lim_{m \rightarrow \infty} \mathbb{E}[\xi_m] = \mathbb{E} \left[\lim_{m \rightarrow \infty} \xi_m \right] = B_K(\rho, \gamma), \quad (\text{D.2})$$

with $B_K(\rho, \gamma)$ defined as in (3.7). Clearly, the last two limit relations imply that $U_{n,\xi}^m$ converges in probability to $B_K(\rho, \gamma)$, yielding Condition 5. It remains to show (D.1)–(D.2).

We start by proving (D.2). By exploiting the scale and location invariance of K and Lemma A.2 for the $\text{GP}(\gamma)$ distributed random variables, ξ_m^I is equal in distribution to

$$\xi_m := \frac{K(\{h(Y_{q-j:q-1}; Y_{m-q+1:m}^*)\}_{j=1}^q) - K(\{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q)}{A(m)},$$

where $Y_{0:q-1} = 1$, $Y_{q-1:q-1} \geq Y_{q-2:q-1} \geq \dots \geq Y_{1:q-1}$ are Pareto(1) order statistics independent of $Y_{m-q+1:m}^*$ and $h(x; t)$ is the function defined as

$$h(x; t) = \frac{U(tx) - U(t)}{t U'(t)}.$$

Under Condition 3, following Theorem 2.3.12 in [4], we have, for $x > 0$,

$$\lim_{t \rightarrow \infty} h(x; t) = \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{t U'(t)} = h_\gamma(x),$$

and

$$\lim_{t \rightarrow \infty} \frac{h(x; t) - h_\gamma(x)}{A(t)} = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds =: H_{\gamma, \rho}(x). \quad (\text{D.3})$$

By Condition 6 the function K is differentiable and all q first-order partial derivatives $\dot{K}_1, \dots, \dot{K}_q$ are continuous. The mean value theorem for several variables implies that there exists a point $\hat{Z}_m$ on the line segment between the points $\{h(Y_{q-j:q-1}; Y_{m-q+1:m}^*)\}_{j=1}^q$ and $\{Z_{q-j:q-1}\}_{j=1}^q = \{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q$ such that as $m \rightarrow \infty$,

$$\begin{aligned} \xi_m &= \frac{A(Y_{m-q+1:m}^*)}{A(m)} \sum_{j=1}^q \dot{K}_j(\hat{Z}_m) \frac{h(Y_{q-j:q-1}; Y_{m-q+1:m}^*) - h_\gamma(Y_{q-j:q-1})}{A(Y_{m-q+1:m}^*)} \\ &\rightarrow S_q^{-\rho} \sum_{j=1}^q \dot{K}_j(\{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q) H_{\gamma, \rho}(Y_{q-j:q-1}) \end{aligned} \quad (\text{D.4})$$

almost surely, where S_q is a properly constructed Erlang(q) random variable independent of $Y_{q-1:q-1} \geq Y_{q-2:q-1} \geq \dots \geq Y_{1:q-1}$. To see this, note that in a suitable Skorokhod construction, $m^{-1}Y_{m-q+1:m}^* \rightarrow S_q^{-1}$ almost surely. In addition, $A(mx)/A(m) \rightarrow x^\rho$ as $m \rightarrow \infty$. Therefore, it follows by the extended continuous mapping theorem that $A(Y_{m-q+1:m}^*)/A(m) \rightarrow S_q^{-\rho}$ as $m \rightarrow \infty$.

In view of (D.4), we show that the sequence $\{\xi_m\}_{m \geq m_0}$ is uniformly integrable for sufficiently large integer m_0 . In particular, we will show that

$$\sup_{m \geq m_0} \mathbb{E}[|\xi_m|^{1+\epsilon/2}] < \infty$$

for some positive integer m_0 , with $\epsilon > 0$ as given in Condition 7. Write $\eta = 1 + \epsilon/2$. We deal with ξ_m on the event

$$D_m = \{Y_{m-q+1:m}^* > t_0\}$$

and its complement D_m^c for $t_0 > 1$ to be defined later.

Firstly, we handle $\mathbb{E}(|\xi_m|^\eta; D_m^c)$. By independence of $Y_{m-q+1:m}^*$ and $Y_1, \dots, Y_{q-1}$, we have, by the Cauchy-Schwarz inequality, for any $t_0 > 1$,

$$\begin{aligned} \mathbb{E}(|\xi_m|^\eta; D_m^c) &\leq \frac{2^\eta \mathbb{E}[|K_m(X_I)|^\eta + |K_m(Z_I)|^\eta; D_m^c]}{|A(m)|^\eta} \\ &\leq \frac{2^\eta}{|A(m)|^\eta} \left\{ [\mathbb{E}(|K_m(X_I)|^{2\eta})]^{1/2} + [\mathbb{E}(|K_m(Z_I)|^{2\eta})]^{1/2} \right\} [\mathbb{P}(D_m^c)]^{1/2} \rightarrow 0, \end{aligned}$$

as $m \rightarrow \infty$. To obtain the final step, we use the following three facts: $\mathbb{E}(|K_m(Z_I)|^{2\eta})$ is finite and does not depend on m , $\sup_{m \geq q} \mathbb{E}(|K_m(X_I)|^{2\eta}) = O(m^a)$ for some $a > 0$ by Condition 7 and $\mathbb{P}(D_m^c)$ tends to zero much faster than any power of $|A(m)|$ (exponentially versus polynomially).

Next, we handle $\mathbb{E}(|\xi_m|^\eta; D_m)$. Consider again the expansion (D.4) of ξ_m by the mean value theorem with

$$\hat{Z}_m = (1 - \hat{a})\{h(Y_{q-j:q-1}; Y_{m-q+1:m}^*)\}_{j=1}^q + \hat{a}\{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q$$

for some random $0 < \hat{a} < 1$. Since $q = 3$, we abbreviate $(\hat{Z}_{3-j:2})_{j=1}^3 := \hat{Z}_m$, where $\hat{Z}_{2:2} > \hat{Z}_{1:2} > \hat{Z}_{0:2} = 0$.

Fix $\delta > 0$ to be determined later. Since $|A|$ is regularly varying, the inequality in Lemma D.1 implies that there exists some $t_0(\delta) > 0$ such that

$$|\xi_m|^\eta \mathbb{I}_{D_m} \leq (1 + \delta) \left(\frac{Y_{m-2:m}^*}{m} \right)^{\eta(\rho \pm \delta)} \left| \sum_{j=1}^3 \dot{K}_j(\hat{Z}_m) \frac{\tilde{h}(Y_{3-j:2})}{A(Y_{m-2:m}^*)} \right|^\eta$$

where $(x)^{\pm\delta}$ is short-hand for $\max\{x^\delta, x^{-\delta}\}$ and where $\tilde{h}(Y_{i:2}) = h(Y_{i:2}; Y_{m-2:m}^*) - h_\gamma(Y_{i:2})$. With straightforward calculation, we obtain

$$\dot{K}_1(x) = \frac{g' \left(\ln \left(\frac{x_1 - x_2}{x_2 - x_3} \right) \right)}{x_1 - x_2}, \dot{K}_3(x) = \frac{g' \left(\ln \left(\frac{x_1 - x_2}{x_2 - x_3} \right) \right)}{x_2 - x_3}, \dot{K}_2(x) = -\dot{K}_1(x) - \dot{K}_3(x).$$

By exploiting the derivative structure, we have

$$\begin{aligned} & \sum_{j=1}^3 \dot{K}_j(\hat{Z}_m) \tilde{h}(Y_{3-j:2}) \\ &= \dot{K}_1(\hat{Z}_m) (\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})) + \dot{K}_3(\hat{Z}_m) (\tilde{h}(Y_{0:2}) - \tilde{h}(Y_{1:2})) \\ &= g' \left(\ln \left(\frac{\hat{Z}_{2:2} - \hat{Z}_{1:2}}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} \right) \right) \left(\frac{\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})}{\hat{Z}_{2:2} - \hat{Z}_{1:2}} + \frac{\tilde{h}(Y_{0:2}) - \tilde{h}(Y_{1:2})}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} \right). \end{aligned}$$

It follows that

$$\begin{aligned} |\xi_m|^\eta \mathbb{I}_{D_m} &\leq (1 + \delta) \left(\frac{Y_{m-2:m}^*}{m} \right)^{\eta(\rho \pm \delta)} \left| g' \left(\ln \left(\frac{\hat{Z}_{2:2} - \hat{Z}_{1:2}}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} \right) \right) \right|^\eta \\ &\quad \cdot \left| \sum_{j=1}^2 \frac{\tilde{h}(Y_{3-j:2}) - \tilde{h}(Y_{2-j:2})}{(\hat{Z}_{3-j:2} - \hat{Z}_{2-j:2}) A(Y_{m-2:m}^*)} \right|^\eta. \end{aligned}$$

By Lemma D.3 below, any positive power of

$$\left| \sum_{j=1}^2 \frac{\tilde{h}(Y_{3-j:2}) - \tilde{h}(Y_{2-j:2})}{(\hat{Z}_{3-j:2} - \hat{Z}_{2-j:2}) A(Y_{m-2:m}^*)} \right|$$

can be bounded by an integrable random variable not depending on m , provided $Y_{m-2:m}^* > t_0$ for some sufficiently large t_0 . Consequently, to show that $\sup_{m > m_0} \mathbb{E} |\xi_m|^\eta \mathbb{I}_{D_m} < \infty$ by the Hölder's inequality, it suffices to prove that

$$\sup_{m > m_0} \mathbb{E} \left[\left| g' \left(\ln \left(\frac{\hat{Z}_{2:2} - \hat{Z}_{1:2}}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} \right) \right) \right|^{r\eta} \left(\frac{Y_{m-2:m}^*}{m} \right)^{r\eta(\rho \pm \delta)} \right] < \infty$$

for some $r > 1$.

Lemma D.2 shows that $(Y_{m-2:m}^*/m)^{r\eta(\rho\pm\delta)}$ is uniformly integrable for any $r > 1$, with choosing δ such that $\rho \pm \delta \leq 3/(r\eta)$. Therefore, it remains to prove that there exists some $r > 1$ such that

$$\sup_{m>m_0} \mathbb{E} \left| g' \left(\ln \left(\frac{\hat{Z}_{2:2} - \hat{Z}_{1:2}}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} \right) \right) \right|^{r\eta} < \infty. \quad (\text{D.5})$$

Note that the relation (D.5) holds trivially if part (b) of Condition 6 holds. We show that it also holds if part (a) of Condition 6 holds. Write $\frac{b}{c} := \frac{Z_{2:2} - Z_{1:2}}{Z_{1:2} - Z_{0:2}}$ and $\frac{d}{e} := \frac{h(Y_{2:2}; Y_{m-2:m}^*) - h(Y_{1:2}; Y_{m-2:m}^*)}{h(Y_{1:2}; Y_{m-2:m}^*) - h(Y_{0:2}; Y_{m-2:m}^*)}$. Then $\frac{\hat{Z}_{2:2} - \hat{Z}_{1:2}}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} = \frac{\hat{a}b + (1-\hat{a})d}{\hat{a}c + (1-\hat{a})e}$ is bounded by $\frac{b}{c}$ and $\frac{d}{e}$. By the assumed monotonicity of g' , we have

$$\begin{aligned} \left| g' \left(\ln \left(\frac{\hat{Z}_{2:2} - \hat{Z}_{1:2}}{\hat{Z}_{1:2} - \hat{Z}_{0:2}} \right) \right) \right|^{r\eta} &\leq \left| g' \left(\ln \left(\frac{Z_{2:2} - Z_{1:2}}{Z_{1:2} - Z_{0:2}} \right) \right) \right|^{r\eta} \\ &\quad + \left| g' \left(\ln \left(\frac{h(Y_{2:2}; Y_{m-2:m}^*) - h(Y_{1:2}; Y_{m-2:m}^*)}{h(Y_{1:2}; Y_{m-2:m}^*) - h(Y_{0:2}; Y_{m-2:m}^*)} \right) \right) \right|^{r\eta}. \end{aligned}$$

Together with Condition 6, it completes the proof of the relation (D.5).

By combining the uniform integrability of $|\xi_m|^\eta \mathbb{I}_{D_m}$ and $|\xi_m|^\eta \mathbb{I}_{D_m^c}$, we conclude that ξ_m is uniformly integrable. It follows that

$$\begin{aligned} \lim_{m \rightarrow \infty} \mathbb{E}[\xi_m] &= \mathbb{E} \left[\lim_{m \rightarrow \infty} \xi_m \right] \\ &= \mathbb{E} [S_3^{-\rho}] \mathbb{E} \left[\sum_{j=1}^3 \dot{K}_{(j)} (\{h_\gamma(Y_{3-j:2})\}_{j=1}^3) H_{\gamma,\rho}(Y_{3-j:2}) \right] = B_K(\rho, \gamma), \end{aligned}$$

which completes the proof of (D.2).

Finally, we turn to prove (D.1). Note that all steps above in conjunction with Lemma D.3 can be adapted to prove uniform integrability of ξ_m^2 , from which it follows that

$$\lim_{m \rightarrow \infty} \mathbb{E}[\xi_m^2] = \mathbb{E} [S_3^{-2\rho}] \mathbb{E} \left[\left(\sum_{j=1}^3 \dot{K}_{(j)} (\{h_\gamma(Y_{3-j:2})\}_{j=1}^3) H_{\gamma,\rho}(Y_{3-j:2}) \right)^2 \right] =: \phi.$$

By classical theory on U-statistics [19, Chapter 5.2], we have

$$\sum_{\ell=1}^m p_{n,m}(\ell) \Psi_{m,\ell} \leq \frac{m}{n} \Psi_{m,m},$$

where $\Psi_{m,m}$ is the variance of ξ_m . The relation (D.1) follows immediately from $\mathbb{E}[\xi_m] \rightarrow B_K(\rho, \gamma)$ and $\mathbb{E}[\xi_m^2] \rightarrow \phi$ as $m \rightarrow \infty$. $\square$

Lemma D.3. Recall $\tilde{h}(Y_{i:2}) = h(Y_{i:2}; Y_{m-2:m}^*) - h_\gamma(Y_{i:2})$ and $\hat{Z}_m = (1 - \hat{a})\{h(Y_{3-j:2}; Y_{m-2:m}^*)\}_{j=1}^3 + \hat{a}\{h_\gamma(Y_{3-j:2})\}_{j=1}^3$ for some $0 < \hat{a} < 1$, where $\hat{a}$ is a random variable. Recall that $\{\hat{Z}_{3-j:2}\}_{j=1}^3 := \hat{Z}_m$ where $\hat{Z}_{2:2} \geq \hat{Z}_{1:2} \geq \hat{Z}_{0:2}$ and $D_m = \{Y_{m-q+1:m}^* > t_0\}$. For sufficiently large $t_0 > 1$, we have

$$\sup_{m \geq 3} \sum_{j=1}^2 \left| \frac{\tilde{h}(Y_{3-j:2}) - \tilde{h}(Y_{2-j:2})}{(\hat{Z}_{3-j:2} - \hat{Z}_{2-j:2}) A(Y_{m-2:m}^*)} \right| \mathbb{I}_{D_m} \leq B$$

for a random variable B satisfying $\mathbb{E}[|B|^r] < \infty$ for any $r > 0$.

Proof. For brevity, we only deal with the first term in the summation:

$$\left| \frac{\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})}{(\hat{Z}_{2:2} - \hat{Z}_{1:2}) A(Y_{m-2:m}^*)} \right|.$$

The second term can be handled using the same inequalities even though $Y_{0:2} = 1$.

Recall the function $A(t)$ from Condition 3 as

$$A(t) = \frac{tU''(t)}{U'(t)} - \gamma + 1.$$

Following the proof of Theorem 2.3.12 in [4], we can write

$$\frac{\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})}{A(Y_{m-2:m}^*)} = \int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s \frac{A(Y_{m-2:m}^* u)}{A(Y_{m-2:m}^*)} \frac{U'(Y_{m-2:m}^* u)}{U'(Y_{m-2:m}^*)} u^{-\gamma} du ds.$$

By regular variation of the absolute value of A and U' with indices ρ and $\gamma - 1$ respectively and using Potter's bounds in Lemma D.1, we get that for any $\delta_2 > 0$ there exists a $t_0(\delta_2)$ such that if $Y_{m-2:m}^* > t_0(\delta_2)$,

$$\begin{aligned} & \int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s \frac{A(Y_{m-2:m}^* u)}{A(Y_{m-2:m}^*)} \frac{U'(Y_{m-2:m}^* u)}{U'(Y_{m-2:m}^*)} u^{-\gamma} du ds \\ & \leq \int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s (1 + \delta_2) u^{\rho-1+\delta_2} du ds \\ & = \frac{1 + \delta_2}{\rho + \delta_2} (h_{\gamma+\rho+\delta_2}(Y_{2:2}) - h_{\gamma+\rho+\delta_2}(Y_{1:2})) - \frac{1 + \delta_2}{\rho + \delta_2} (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})), \end{aligned}$$

provided $\rho + \delta_2 \neq 0$, and similarly,

$$\begin{aligned} & \int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s \frac{A(Y_{m-2:m}^* u)}{A(Y_{m-2:m}^*)} \frac{U'(Y_{m-2:m}^* u)}{U'(Y_{m-2:m}^*)} u^{-\gamma} du ds \\ & \geq \frac{1 - \delta_2}{\rho - \delta_2} (h_{\gamma+\rho-\delta_2}(Y_{2:2}) - h_{\gamma+\rho-\delta_2}(Y_{1:2})) - \frac{1 - \delta_2}{\rho - \delta_2} (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})). \end{aligned}$$

Accordingly, it suffices to prove that, for any $r > 0$ and $\gamma \in \mathbb{R}$, the term

$$\left| \frac{h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})}{\hat{Z}_{2:2} - \hat{Z}_{1:2}} \right|^r$$

can be bounded by an integrable function not depending on m .

Note that

$$\begin{aligned} \hat{Z}_{2:2} - \hat{Z}_{1:2} &= (1 - \hat{a}) (h(Y_{2:2}; Y_{m-2:m}^*) - h(Y_{1:2}; Y_{m-2:m}^*)) \\ &\quad + \hat{a} (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})). \end{aligned}$$

For any $\delta_1 > 0$, there exists $t_0(\gamma, \delta_1) > 1$ such that, if $Y_{m-2:m}^* > t_0(\gamma, \delta_1)$, following the Potter's bounds in Lemma D.1, we have that

$$\begin{aligned} h(Y_{2:2}; Y_{m-2:m}^*) - h(Y_{1:2}; Y_{m-2:m}^*) &= \int_{Y_{1:2}}^{Y_{2:2}} \frac{U'(tY_{m-2:m}^*)}{U'(Y_{m-2:m}^*)} dt \\ &\geq (1 - \delta_1) \int_{Y_{1:2}}^{Y_{2:2}} t^{\gamma-1-\delta_1} dt = (1 - \delta_1) (h_{\gamma-\delta_1}(Y_{2:2}) - h_{\gamma-\delta_1}(Y_{1:2})), \end{aligned}$$

provided $\gamma - \delta_1 \neq 0$. Define $\tilde{Y} = Y_{2:2}/Y_{1:2}$. Recalling that $0 < \hat{a} < 1$, we get that

$$\begin{aligned} &\hat{Z}_{2:2} - \hat{Z}_{1:2} \\ &\geq (1 - \hat{a})(1 - \delta_1) (h_{\gamma-\delta_1}(Y_{2:2}) - h_{\gamma-\delta_1}(Y_{1:2})) + \hat{a} (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})) \\ &\geq (1 - \hat{a})(1 - \delta_1) (h_{\gamma-\delta_1}(Y_{2:2}) - h_{\gamma-\delta_1}(Y_{1:2})) + \hat{a}(1 - \delta_1) (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})) \\ &\geq (1 - \delta_1) Y_{2:2}^{-\delta_1} (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})), \end{aligned}$$

where the last inequality follows from the fact that for $0 < x < y < \infty$, for $\gamma \in \mathbb{R}$ and for $0 \leq \delta < \infty$, we have

$$\begin{aligned} y^\delta (h_{\gamma-\delta}(y) - h_{\gamma-\delta}(x)) &= y^\delta \int_x^y t^{\gamma-\delta-1} dt = \int_x^y t^{\gamma-1} (y/t)^\delta dt \\ &\geq \int_x^y t^{\gamma-1} dt = h_\gamma(y) - h_\gamma(x). \end{aligned}$$

We conclude with the bound

$$\left| \frac{h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})}{\hat{Z}_{2:2} - \hat{Z}_{1:2}} \right| \leq \frac{h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})}{(1 - \delta_1) Y_{2:2}^{-\delta_1} (h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}))} \leq \frac{Y_{2:2}^{\delta_1}}{(1 - \delta_1)},$$

from which the lemma follows by choosing $\delta_1 < (2r)^{-1}$. $\square$

Appendix E: Proofs for Section 4

To show Proposition 4.1, we use known formulas for the expectation of exponentiated Generalized Pareto random variables and their order statistics in terms of the digamma function ψ .

Lemma E.1. Let $\gamma \in \mathbb{R}$ and let Z_1, Z_2 be two independent $\text{GP}(\gamma)$ random variables, with order statistics $Z_{1:2} \leq Z_{2:2}$. We have

$$\begin{aligned} \mathbb{E}[\ln Z_1] &= \begin{cases} \ln(1/\gamma) + \psi(1) - \psi(1/\gamma) & \text{if } \gamma > 0 \\ \ln(-1/\gamma) + \psi(1) - \psi(1 - 1/\gamma) & \text{if } \gamma < 0 \\ \psi(1) & \text{if } \gamma = 0, \end{cases} \\ \mathbb{E}[\ln Z_{2:2}] &= \begin{cases} \ln(1/\gamma) + \psi(1) + \psi(2/\gamma) - 2\psi(1/\gamma) & \text{if } \gamma > 0 \\ \ln(-1/\gamma) + \psi(1) + \psi(1 - 2/\gamma) - 2\psi(1 - 1/\gamma) & \text{if } \gamma < 0 \\ \psi(1) + \ln(2) & \text{if } \gamma = 0, \end{cases} \\ \mathbb{E}[\ln Z_{1:2}] &= \begin{cases} \ln(1/\gamma) + \psi(1) - \psi(2/\gamma) & \text{if } \gamma > 0 \\ \ln(-1/\gamma) + \psi(1) - \psi(1 - 2/\gamma) & \text{if } \gamma < 0 \\ \psi(1) - \ln(2) & \text{if } \gamma = 0. \end{cases} \end{aligned}$$

As a consequence, for all $\gamma \in \mathbb{R}$, we have

$$\mathbb{E}[K_P(Z_{2:2}, Z_{1:2}, 0)] = \gamma. \quad (\text{E.1})$$

Proof. The formulas for $\mathbb{E}[\ln Z_1]$, $\mathbb{E}[\ln Z_{1:2}]$ and $\mathbb{E}[\ln Z_{2:2}]$ follow directly from the moments of exponentiated GP random variables and their order statistics, see [13, Section 2.3].

Let Y_1, Y_2 be two independent standard Pareto random variables and let $Y_{1:2} \leq Y_{2:2}$ denote their order statistics as. Recall that we can represent the $\text{GP}(\gamma)$ random variables Z_1, Z_2 and their corresponding order statistics $Z_{1:2} \leq Z_{2:2}$ as $Z_i = h_\gamma(Y_i)$, $Z_{1:2} = h_\gamma(Y_{1:2})$ and $Z_{2:2} = h_\gamma(Y_{2:2})$. But then

$$Z_{2:2} - Z_{1:2} = h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) = Y_{1:2}^\gamma h_\gamma(Y_{2:2}/Y_{1:2}).$$

Since $Y_{2:2}/Y_{1:2}$ is equal in distribution to Y_1 , we find

$$\begin{aligned} u_1(\gamma) &:= \mathbb{E} \left[\ln \left(\frac{Z_{2:2} - Z_{1:2}}{Z_{1:2}} \right) \right] \\ &= \mathbb{E}[\ln Y_{1:2}^\gamma] + \mathbb{E}[\ln Z_1] - \mathbb{E}[\ln Z_{1:2}] \\ &= \begin{cases} \gamma/2 - \psi(1/\gamma) + \psi(2/\gamma) & \text{if } \gamma > 0 \\ \gamma/2 - \psi(1 - 1/\gamma) + \psi(1 - 2/\gamma) & \text{if } \gamma < 0 \\ \ln(2) & \text{if } \gamma = 0. \end{cases} \end{aligned}$$

More straightforwardly, we have

$$\begin{aligned} u_2(\gamma) &:= \mathbb{E} \left[\ln \left(\frac{Z_{2:2}}{Z_{1:2}} \right) \right] = \mathbb{E}[\ln Z_{2:2}] - \mathbb{E}[\ln Z_{1:2}] \\ &= \begin{cases} 2(\psi(2/\gamma) - \psi(1/\gamma)) & \text{if } \gamma > 0 \\ 2(\psi(1 - 2/\gamma) - \psi(1 - 1/\gamma)) & \text{if } \gamma < 0 \\ 2\ln(2) & \text{if } \gamma = 0. \end{cases} \end{aligned}$$

It follows that $\mathbb{E}[K_P(Z_{2:2}, Z_{2:1}, 0)] = 2u_1(\gamma) - u_2(\gamma) = \gamma$, as required. $\square$

Proof of Proposition 4.1. We have

$$\begin{aligned}\mathbb{E}[K_{P,m}(Z_1, \dots, Z_m)] &= \mathbb{E}[K_P(Z_{m:m}, Z_{m-1:m}, Z_{m-2:m})] \\ &= \mathbb{E}[K_P(Z_{2:2}, Z_{1:2}, 0)] = \gamma,\end{aligned}$$

where we used Lemma A.2 and the location-scale invariance of K_P in the second step and Lemma E.1 in the last step. $\square$

Proof of Theorem 4.2. We apply Theorem 3.2. To this end, we need to show that Conditions 1, 2, 3, 4, 6 and 7 hold. Conditions 2, 3, and 4 hold by assumption.

One can check that the kernel K_P is differentiable, has continuous first-order partial derivatives and is of the form (3.5) for an auxiliary function g that satisfies $|g'(x)| = |2 - e^x/(e^x + 1)| \leq 2$. Therefore, Condition 6 holds too. It remains to show that $K_{P,m}$ satisfies Conditions 1 and 7.

Condition 1 can be verified as follows: for any $p > 0$, with the same notation as in the proof of Lemma E.1,

$$\begin{aligned}\mathbb{E}[|K_{P,m}(Z_1, \dots, Z_m)|^p] &= \mathbb{E}\left[\left|2 \ln\left(\frac{Z_{2:2} - Z_{1:2}}{Z_{1:2}}\right) - \ln\left(\frac{Z_{2:2}}{Z_{1:2}}\right)\right|^p\right] \\ &= \mathbb{E}[|2 \ln[Y_{1:2}^\gamma] + 2 \ln[h_\gamma(Y_{2:2}/Y_{1:2})] - \ln[h_\gamma(Y_{1:2})] - \ln[h_\gamma(Y_{2:2})]|^p] < \infty,\end{aligned}$$

because exponential random variables and exponentiated GP random variables (and its order statistics) have finite moments of all orders [13, Corollary 3].

To show Condition 7, we rely on Condition 8. Note that $K_{P,m}$ is a linear combination of the logarithms of the spacings $X_{m:m} - X_{m-1:m}$, $X_{m:m} - X_{m-2:m}$ and $X_{m-1:m} - X_{m-2:m}$. We have

$$\begin{aligned}-\ln(X_{m:m} - X_{m-2:m}) \mathbb{1}_{\{X_{m:m} - X_{m-2:m} < 1\}} \\ \leq -\ln(X_{m:m} - X_{m-1:m}) \mathbb{1}_{\{X_{m:m} - X_{m-1:m} < 1\}}\end{aligned}$$

and for $0 \leq j < i \leq 2$,

$$\begin{aligned}\ln(X_{m-j:m} - X_{m-i:m}) \mathbb{1}_{\{X_{m-j:m} - X_{m-i:m} > 1\}} \\ \leq \ln(X_{m:m} - X_{m-2:m}) \mathbb{1}_{\{X_{m:m} - X_{m-2:m} > 1\}}.\end{aligned}$$

Consequently, to verify Condition 7 we only have to show that

$$\mathbb{E}[(-\ln(X_{m-j:m} - X_{m-j-1:m}))^p \mathbb{1}_{\{X_{m-j:m} - X_{m-j-1:m} < 1\}}] = O(m^a) \quad (\text{E.2})$$

for $j \in \{0, 1\}$, and

$$\mathbb{E}[(\ln(X_{m:m} - X_{m-2:m}))^p \mathbb{1}_{\{X_{m:m} - X_{m-2:m} > 1\}}] = O(m^a) \quad (\text{E.3})$$

for some $p > 2$ and $a > 0$.

First, for the expectation in (E.2), note that the distribution function $\tilde{F}_j$ of the spacing $X_{m-j:m} - X_{m-j-1:m}$ for $j \in \{0, 1\}$ is given by

$$\tilde{F}_j(r) = 1 - \int_{-\infty}^{+\infty} (m-j-1) \binom{m}{m-j-1} f(x) F^{m-j-2}(x) (1 - F(x+r))^{j+1} dx.$$

Moreover, note that $|(1 - F(x))^2 - (1 - F(a))^2| \leq 4|F(x) - F(a)|$ for all $x, a \in \mathcal{E}$ and $\tilde{F}_j(0) = 0$. We get, for any $0 < x_0 < 1$ and $r_0 = (-\ln x_0)^p$, that

$$\begin{aligned} & \mathbb{E} [(-\ln(X_{m-j:m} - X_{m-j-1:m}))^p \mathbb{1}_{\{X_{m-j:m} - X_{m-j-1:m} < x_0\}}] \\ &= \int_{r_0}^{\infty} \mathbb{P}(X_{m-j:m} - X_{m-j-1:m} < e^{-r^{1/p}}) \, dr \\ &= \int_{r_0}^{\infty} (\tilde{F}_j(e^{-r^{1/p}}) - \tilde{F}_j(0)) \, dr \\ &\leq (m-j-1) \binom{m}{m-j-1} \\ &\quad \cdot \int_{r_0}^{\infty} \int_{-\infty}^{+\infty} f(x) F^{m-j-2}(x) \left| (1 - F(x + e^{-r^{1/p}}))^{j+1} - (1 - F(x))^{j+1} \right| \, dx \, dr \\ &\leq c \binom{m}{m-j-1} \int_{r_0}^{\infty} r^{-(1+\epsilon)} \, dr, \end{aligned}$$

for some constant $c > 0$. Here the last step follows from (4.4) in Condition 8. With straightforward calculation, we get that as $m \rightarrow \infty$,

$$\mathbb{E} [(-\ln(X_{m-j:m} - X_{m-j-1:m}))^p \mathbb{1}_{\{X_{m-j:m} - X_{m-j-1:m} < 1\}}] = O(m^2).$$

Secondly, for the expectation in (E.3), consider the following five events:

$$\begin{aligned} A &:= \{X_{m:m} - X_{m-2:m} > 1\}, \\ A_1 &:= \{X_{m:m} > 1, X_{m-2:m} < -1\}, \\ A_2 &:= \{X_{m:m} \leq 1, X_{m-2:m} < -1\}, \\ A_3 &:= \{X_{m:m} > 1, X_{m-2:m} \geq -1\}, \\ A_4 &:= \{X_{m:m} \leq 1, X_{m-2:m} \geq -1\}. \end{aligned}$$

The following inequalities can be verified in a straightforward way:

$$\begin{aligned} (\ln(X_{m:m} - X_{m-2:m}))^p \mathbb{1}_{A \cap A_1} &\leq 2^p [(\ln(X_{m:m}))^p + (\ln(-X_{m-2:m}))^p + 1] \mathbb{1}_{A \cap A_1}, \\ (\ln(X_{m:m} - X_{m-2:m}))^p \mathbb{1}_{A \cap A_2} &\leq 2^p [(\ln(-X_{m-2:m}))^p + 1] \mathbb{1}_{A \cap A_2}, \\ (\ln(X_{m:m} - X_{m-2:m}))^p \mathbb{1}_{A \cap A_3} &\leq 2^p [(\ln(X_{m:m}))^p + 1] \mathbb{1}_{A \cap A_3}, \\ (\ln(X_{m:m} - X_{m-2:m}))^p \mathbb{1}_{A \cap A_4} &\leq \mathbb{1}_{A \cap A_4}. \end{aligned}$$

Therefore, we only need to show that the following two expectations

$$\mathbb{E} [(\ln(X_{m:m}))^p \mathbb{1}_{\{X_{m:m} > 1\}}] \quad \text{and} \quad \mathbb{E} [(\ln(-X_{m-2:m}))^p \mathbb{1}_{\{X_{m-2:m} < -1\}}],$$

are of the order $O(m^a)$ for some constant $a > 0$.

We start from the second expectation concerning $X_{m-2:m}$. There exists $x_0 > 1$ such that for all $m > 3$, the variable $(-X_{m-2:m}) \mathbb{1}_{\{X_{m-2:m} < -x_0\}}$ is stochastically dominated by $(-X) \mathbb{1}_{\{X < -x_0\}}$, where X has distribution F . Therefore, in view of Condition 8, we have

$$\begin{aligned} \mathbb{E} [(\ln(-X_{m-2:m}))^p \mathbf{1}_{\{X_{m-2:m} < -1\}}] \\ \leq \mathbb{E} [(\ln(-X))^p \mathbf{1}_{\{X < -x_0\}}] + (\ln(x_0))^p = O(1). \end{aligned}$$

Finally, for the first expectation concerning $X_{m:m}$, choose $0 < \theta < 1/\gamma$. Note that there exists $x_0 = x_0(p) > 1$ such that $(\ln(x))^p$ is a concave function in $x > x_0$. In view of Jensen's inequality, we have

$$\mathbb{E} [(\ln(X_{m:m}))^p \mathbf{1}_{\{X_{m:m} > x_0\}}] \leq \left(\frac{1}{\theta} \ln (\mathbb{E} [X_{m:m}^\theta \mathbf{1}_{\{X_{m:m} > x_0\}}]) \right)^p.$$

For $\theta < 1/\gamma$, Theorem 5.3.2 in [4] yields that as $m \rightarrow \infty$,

$$\mathbb{E} [X_{m:m}^\theta \mathbf{1}_{\{X_{m:m} > 1\}}] = (U(m))^\theta = o(m),$$

which is sufficient to bound the first expectation by $O(m^a)$ with any $a > 0$. Combining the two expectations completes the proof of (E.3) and thus that of Condition 7. $\square$

Proof of Lemma 4.3. Put $S_{ij} = \ln(X_{n-i+1:n} - X_{n-j+1:n})$ for integer $1 \leq i < j \leq n$. For $I \subset [n]$ with $|I| = m$, both $K_{m,1}(X_I)$ and $K_{m,2}(X_I)$ are linear combinations of such log-spacings S_{ij} . For a given pair (i, j) of indices, the number of such subsets I in which S_{ij} appears can be counted:

- There are $\binom{n-j}{m-2}$ sets I such that $X_{n-i+1:n}$ and $X_{n-j+1:n}$ are respectively the largest and 2nd largest order statistics of X_I .
- There are $(i-1)\binom{n-j}{m-3}$ sets I such that $X_{n-i+1:n}$ and $X_{n-j+1:n}$ are respectively the 2nd and 3rd largest order statistics of X_I .
- There are $(j-i-1)\binom{n-j}{m-3}$ sets I such that $X_{n-i+1:n}$ and $X_{n-j+1:n}$ are respectively the largest and 3rd largest order statistics of X_I .

Since $n-j+1$ cannot be lower than $m-2$ (because there need to be at least $m-3$ observations lower than $X_{n-j+1:n}$), we find

$$U_{n,P}^m = 2U_{n,1}^m - U_{n,2}^m$$

with

$$\begin{aligned} U_{n,1}^m &= \binom{n}{m}^{-1} \sum_{j=2}^{n-m+3} \sum_{i=1}^{j-1} \left[\binom{n-j}{m-2} - (i-1) \binom{n-j}{m-3} \right] S_{ij}, \\ U_{n,2}^m &= \binom{n}{m}^{-1} \sum_{j=2}^{n-m+3} \sum_{i=1}^{j-1} \left[(j-i-1) \binom{n-j}{m-3} - (i-1) \binom{n-j}{m-3} \right] S_{ij}. \end{aligned}$$

It follows that

$$U_{n,P}^m = \binom{n}{m}^{-1} \sum_{j=2}^{n-m+3} \left[2 \binom{n-j}{m-2} - (j-2) \binom{n-j}{m-3} \right] \sum_{i=1}^{j-1} S_{ij},$$

which can be further simplified to the stated formula. $\square$

Acknowledgments

We are grateful to the constructive comments by the referees and the associate editor, which stimulated us to improve the structure of the paper.

References

- [1] BÜCHER, A. and SEGERS, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Statist.* **12** 1098–1125. [MR3780041](#)
- [2] CHVÁTAL, V. (1979). The tail of the hypergeometric distribution. *Discrete Mathematics* **25** 285–287. [MR0534946](#)
- [3] DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **52** 393–425. [MR1086795](#)
- [4] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York. [MR2234156](#)
- [5] DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723. [MR3961227](#)
- [6] DREES, H., FERREIRA, A. and DE HAAN, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Prob.* 1179–1201. [MR2071420](#)
- [7] FERREIRA, A. and DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298. [MR3285607](#)
- [8] GRIMSHAW, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* **35** 185–191. [MR1225094](#)
- [9] HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43. [MR0015746](#)
- [10] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325. [MR0026294](#)
- [11] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](#)
- [12] LEE, A. J. (1990). *U-Statistics. Theory and Practice*. Marcel Dekker, Inc., New York. [MR1075417](#)
- [13] LEE, S. and KIM, J. H. (2019). Exponentiated generalized Pareto distribution: Properties and applications towards extreme value theory. *Communications in Statistics – Theory and Methods* **48** 2014–2038. [MR3962098](#)
- [14] OORSCHOT, J. and ZHOU, C. (2020). All Block Maxima method for estimating the extreme value index. *arXiv preprint arXiv:2010.15950*.
- [15] PICKANDS III, J. (1975). Statistical Inference Using Extreme Order Statistics. *Ann. Statist.* **3** 119–131. [MR0423667](#)
- [16] ROBERT, C. Y., SEGERS, J. and FERRO, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020. [MR2540849](#)

- [17] SEGERS, J. (2001). Extremes of a random sample: limit theorems and statistical applications, PhD thesis, Katholieke Universiteit Leuven.
- [18] SEGERS, J. and TEUGELS, J. (2000). Testing the Gumbel hypothesis by Galton's ratio. *Extremes* **3** 291–303. [MR1856202](#)
- [19] SERFLING, R. J. (2009). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York. [MR0595165](#)
- [20] VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge. [MR1652247](#)
- [21] ZHOU, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivar. Anal.* **100** 794–815. [MR2478199](#)