The Annals of Statistics
2023, Vol. 51, No. 3, 1331–1356
https://doi.org/10.1214/23-AOS2294
© Institute of Mathematical Statistics, 2023

EXTREME VALUE INFERENCE FOR HETEROGENEOUS POWER LAW DATA

BY JOHN H.J. EINMAHL[1,a] AND YI HE[2,b]

[1]*Department of Econometrics & OR and CentER, Tilburg University,* [a]*j.h.j.einmahl@uvt.nl*
[2]*Amsterdam School of Economics, University of Amsterdam,* [b]*yi.he@uva.nl*

We extend extreme value statistics to independent data with possibly very different distributions. In particular, we present novel asymptotic normality results for the Hill estimator, which now estimates the extreme value index of the average distribution. Due to the heterogeneity, the asymptotic variance can be substantially smaller than that in the i.i.d. case. As a special case, we consider a heterogeneous scales model where the asymptotic variance can be calculated explicitly. The primary tool for the proofs is the functional central limit theorem for a weighted tail empirical process. We also present asymptotic normality results for the extreme quantile estimator. A simulation study shows the good finite-sample behavior of our limit theorems. We also present applications to assess the tail heaviness of earthquake energies and of cross-sectional stock market losses.

1. Introduction. Consider independent and identically distributed random variables $X_1, \dots, X_p$, $p \in \mathbb{N}$, from some common distribution function F. For this case, the statistical theory of extreme values has been developed comprehensively in the literature, for example, the monographs Beirlant et al. (2004) and de Haan and Ferreira (2006). In this setting, well-known estimators of the extreme value index γ have been introduced in, among others, Hill (1975), Smith (1987) and Dekkers, Einmahl and de Haan (1989). The results for the i.i.d. case are important but might be too restrictive for various applications.

Univariate samples can deviate from the i.i.d. assumption by being dependent and/or by being nonidentically distributed. Statistics of extremes for identically distributed but (weakly, serially) dependent data has been studied extensively in the literature; see, for example, Hsing (1991), Drees (2000), Drees and Rootzén (2010) and the monograph Kulik and Soulier (2020).

In this paper, we focus on independent, but nonidentically distributed data. In early work for this case, a trend in the parameters of generalized Pareto distributions was considered, without providing asymptotic theory: in Davison and Smith (1990), a linear trend in both shape and scale parameters was studied, whereas in Coles (2001) a log-linear trend in the scale parameter was explored. In a general, nonparametric setting the most relevant (and recent) references are Einmahl, de Haan and Zhou (2016) and de Haan and Zhou (2021), but see also both papers for various other references for the non-i.i.d. case. The first paper allows for different distributions that are not too different in the sense that all observations have the same extreme value index γ, whereas in the second paper a gradually changing γ is allowed. Just like in these two papers we consider the case where γ is positive, the heavy-tailed case, but the scope of the present paper is quite different since it allows large heterogeneity of the observations, leading to novel limit theorems for the Hill estimator, and thus considerably extending Einmahl, de Haan and Zhou (2016). Like in de Haan and Zhou (2021), we allow

Received August 2022; revised January 2023.
MSC2020 subject classifications. Primary 62G32, 62G20, 62G30; secondary 60G70, 60F17.
Key words and phrases. Extreme value statistics, functional central limit theorem, heterogeneous scales model, Hill estimator, nonidentical distributions, weighted tail empirical process.

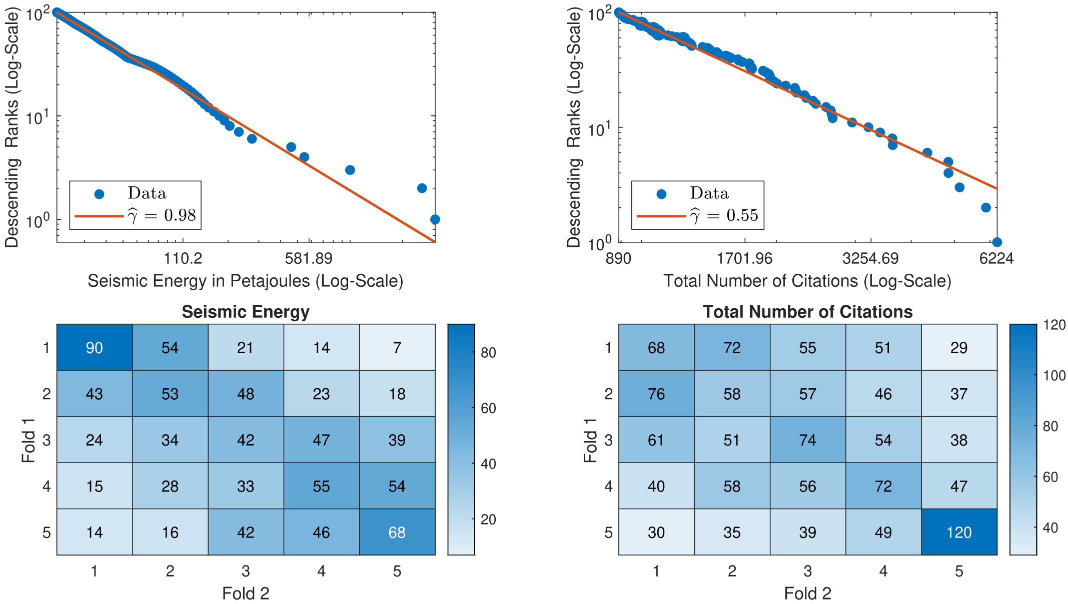

FIG. 1. *Seismic energy of significant global earthquakes from* 1981 *to* 2021 *from the NOAA database and total number of citations from* 1975 *to* 2015 *for statisticians in Ji et al.* (2022)*'s database.*

different extreme value indices for different observations, but we do not require a smooth change of the distribution in i ($i = 1, \dots, p$): neighboring observations (X_i and X_{i+1}) may have (very) different distributions.

Often power laws are discovered and discussed in the applied scientific literature and beyond. In many of these applications, the data are nonidentically distributed. Hence, the existing statistical theory has to be extended to the non-i.i.d. case as indicated above. Settings where this can be relevant are when computing the Hill estimator for natural hazards data across different locations, for a cross-section (on the same day) of daily loss returns of many stocks, for population sizes of cities in a large country or for numbers of citations of scientific articles. In particular, on the top in Figure 1, log-log plots are shown for the top 100 data points in: (1) a seismic energy data set of global significant earthquakes that will be investigated further in Section 6, and (2) a data set of statisticians' total number of citations (excluding self-citations). We observe a linear pattern in these plots that identifies empirical power laws, with Hill estimates of 0.98 and 0.55, respectively. Next, for each data set, we pair the nearest observations on the geographic map (with a minimum time gap of 1 year) in Figure 9 below and on the research map in Figure 1 of Ji et al. (2022), respectively, and then randomly assign one to fold 1 and the other one to fold 2. Within each fold, we label the largest 20% data as class 1, the next 20%–40% as class 2, and so on. The quintile transition matrices at the bottom in Figure 1 count the number of data pairs that occur in the corresponding quintile classes: the sums of relative self-transition frequencies are 1.66 and 1.42, respectively, showing a statistically significant degree of heterogeneity.

As a consequence of our setup, we are not interested in the average (or local) γ but in the γ of the average distribution. Our results reveal that the asymptotic variance of the Hill estimator can be smaller than that in the i.i.d. case, depending on a spurious tail dependence coefficient $R(1, 1)$, which actually measures heterogeneity. A functional central limit theorem for the relevant weighted tail empirical process is crucial for proving the asymptotic normality of the Hill estimator. The limiting process turns out to be a weighted centered Gaussian process that can be substantially "tighter" than the weighted standard Wiener process, which appears in the i.i.d. case. We also use these results to establish the asymptotic normality of the extreme quantile estimator.

In Einmahl and He (2023), consistency of the Hill estimator is shown under weak assumptions, allowing both heterogeneity and dependence of the data. The present paper can be seen as an "asymptotic normality" extension of that paper, thus quantifying the uncertainty in estimation of the Hill estimator and making statistical inference on γ possible.

We highlight as an interesting special case a *heterogeneous scales* model where p latent i.i.d. random variables are multiplied with different, deterministic scales. This relevant and insightful model includes both the case where the above $R(1,1)$ is positive, leading to the novel asymptotic behavior of the Hill estimator, and the case $R(1,1)=0$ leading to the usual γ^2 for the asymptotic variance. In case $R(1,1)>0$, the asymptotic variance is smaller and can be expressed in γ and the distribution of the latent variables.

The remainder of this paper is organized as follows. In Section 2, we present our general results for the Hill estimator. Section 3 contains the specialization to the heterogeneous scales model. The brief Section 4 considers extreme quantile estimation. In Section 5, we present a simulation study and in Section 6 we apply the theory to earthquake energies and cross-sectional stock market losses. Most proofs of the results in Sections 2 and 3 are deferred to Section 7. More simulation results and the remaining proofs are deferred to the Supplementary Material (Einmahl and He (2023)).

2. Asymptotic theory for heterogeneous extremes. Consider independent random variables $X_1^{(p)},\dots,X_p^{(p)}$, for $p\in\mathbb{N}$, that are not necessarily identically distributed. Define their empirical distribution function by

$$F_{\text{emp}}(x)=\frac{1}{p}\sum_{i=1}^{p}\mathbb{1}\big[X_i^{(p)}\le x\big],$$

and their average distribution function by

$$F_p(x)=\mathbb{E}F_{\text{emp}}(x)=\frac{1}{p}\sum_{i=1}^{p}F_{pi}(x),\qquad F_{pi}(x)=\mathbb{P}\big(X_i^{(p)}\le x\big).$$

For simplicity, throughout we assume that there exists a bounded average density function $f_p=F_p'$. (It is possible, although tedious, to weaken this assumption to a suitable smoothness condition on F_p directly). If not indicated differently, for our asymptotic theory it is assumed that the dimension (or sample size) $p\to\infty$.

We consider the situation where the above data arrays obey empirical power laws, which may or may not be generated by heterogeneity of the data.

ASSUMPTION 2.1 (Heavy tail). The average survival function $T_p:=1-F_p$ approaches some nonincreasing function T in the intermediate tail such that $T_p(t)/T(t)\to 1$ for all intermediate threshold sequences $t=t(p)\to\infty$ with $pT(t)\to\infty$. The limit function T is regularly varying with negative index, that is,

$$\frac{T(tx)}{T(t)}\to x^{-1/\gamma},\quad x>0,$$

as $t\to\infty$, where $\gamma>0$ is called the extreme value index.

Observe that the thus defined extreme value index γ is a natural and statistically relevant generalization of that in the i.i.d. case with fixed distribution function F. It is the extreme value index that describes the power-law behavior of the data (often visualized by log-log plots) and that is targeted by the Hill estimator. The assumption implies that, for every intermediate threshold sequence $t = t(p)$,

$$\frac{T_p(tx)}{T_p(t)} \to x^{-1/\gamma}, \quad x > 0.$$

If the empirical survival function $T_{\mathrm{emp}} := 1 - F_{\mathrm{emp}} \approx 1 - F_p = T_p$ for large values and large p, then the data exhibit power-law behavior in the sense that, for large t and p,

$$T_{\mathrm{emp}}(tx) \approx x^{-1/\gamma} T_{\mathrm{emp}}(t).$$

As shown in Einmahl and He (2023), the heavy tail may be due to heterogeneity rather than the tail behavior of the individual random variables. That is, when a high degree of heterogeneity is present, Assumption 2.1 can still hold, even if $X_i^{(p)}$ is light-tailed (e.g., Gaussian) for all $i = 1, \dots, p$; see, for example, (9) and below in Section 5 for specific examples. The heterogeneity thus can go far beyond heteroscedastic extremes as in Einmahl, de Haan and Zhou (2016). Our framework is general and unifies heterogeneous and homogeneous (identically distributed) data.

Recall the average probability density function $f_p = F_p'$. We need a stability condition to control the behavior of extreme observations beyond the intermediate levels.

ASSUMPTION 2.2 (Stability). There exists a positive constant $M < \infty$ such that for all large x and p,

$$\max\{x f_p(x), T_p(x)\} \le M T(x).$$

If in the homogeneous case with $T_p \equiv T$ the von Mises condition holds, Assumption 2.2 is satisfied for $M > \max\{1, 1/\gamma\}$.

Let $k = k(p) \in \{1, \dots, p-1\}$ be a sequence satisfying the following.

ASSUMPTION 2.3 (Intermediate sequence). $k \to \infty$ and $k/p \to 0$ as $p \to \infty$.

Our first key result is a functional central limit theorem for a weighted version of the tail empirical process defined by

$$V_p(x) = \frac{p}{\sqrt{k}}\left(T_{\mathrm{emp}}\left(\frac{u_p}{x^\gamma}\right) - T_p\left(\frac{u_p}{x^\gamma}\right)\right), \quad x \ge 0,$$

where $u_p = Q_p(1 - k/p)$ with Q_p denoting the generalized quantile function corresponding to F_p. It is straightforward to compute the covariance structure of V_p:

$$\begin{aligned}\operatorname{cov}(V_p(x), V_p(y)) &= \frac{1}{k}\sum_{i=1}^{p} \operatorname{cov}(\mathbb{1}[X_i^{(p)} > x^{-\gamma} u_p], \mathbb{1}[X_i^{(p)} > y^{-\gamma} u_p]) \\ &= \frac{p}{k} T_p(u_p(\min\{x, y\})^{-\gamma}) - \frac{p}{k} H_p(u_p x^{-\gamma}, u_p y^{-\gamma}),\end{aligned}$$

where

$$H_p(x, y) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{P}(X_i^{(p)} > x)\mathbb{P}(X_i^{(p)} > y).$$

We assume that the limit of this covariance function exists via the following condition.

ASSUMPTION 2.4 (R-function). For all intermediate threshold sequences $t = t(p) \to \infty$ with $pT(t) \to \infty$,

$$\frac{H_p(tx^{-\gamma}, ty^{-\gamma})}{T(t)} \to R(x, y), \quad x, y > 0.$$

The limit function R may or may not be identically zero.

The following lemma gives a rank-based definition of the R-function, from which we can deduce that R is invariant with respect to an increasing transformation of the data $X_i^{(p)}$.

LEMMA 2.1. *Under Assumption 2.1, Assumption 2.4 holds if and only if*

$$\frac{1}{p\alpha}\sum_{i=1}^{p}\mathbb{P}\big(T_p\big(X_i^{(p)}\big) < \alpha x\big)\mathbb{P}\big(T_p\big(X_i^{(p)}\big) < \alpha y\big) \to R(x, y), \tag{1}$$

for all intermediate (probability) sequences $\alpha = \alpha(p) \downarrow 0$ with $p\alpha \to \infty$.

Heterogeneity can lead to spurious correlation. Here, the function R quantifies tail heterogeneity through a measure of spurious tail dependence between two independent copies of heterogeneous data arrays in the sense that

$$R(x, y) = \lim_{p\to\infty}\frac{1}{p\alpha}\sum_{i=1}^{p}\mathbb{P}\big(T_p\big(X_i^{(p)}\big) < \alpha x, T_p\big(\widetilde{X}_i^{(p)}\big) < \alpha y\big), \tag{2}$$

where $\widetilde{X}_i^{(p)}$ are independent copies of $X_i^{(p)}$ and α is any intermediate sequence as in Lemma 2.1. Indeed, when the variables $X_i^{(p)}$ are independent and identically distributed, Assumption 2.1 implies that

$$\begin{aligned} R(x, y) &= \lim_{p\to\infty}\frac{1}{\alpha}\mathbb{P}\big(T_p\big(X_i^{(p)}\big) < \alpha x\big)\cdot\lim_{p\to\infty}\mathbb{P}\big(T_p\big(X_i^{(p)}\big) < \alpha y\big) \\ &= \lim_{p\to\infty}\frac{1}{\alpha}\mathbb{P}(U < \alpha x)\cdot\lim_{p\to\infty}\mathbb{P}(U < \alpha y) \\ &= x\cdot 0 = 0, \quad U \sim \mathrm{Un}(0, 1) \end{aligned}$$

and, therefore, the limit $R \equiv 0$ is called *trivial*.

The most interesting results in this paper are for *nontrivial* functions R, which then play a vital role in the asymptotic theory. However, the trivial R leads to new results for relevant heterogeneous data arrays, too.

By Lemma 2.1, the function R shares the properties of a *symmetric* tail copula function, including:

$R(x, y) > 0$ for all $x, y > 0$ if R is nontrivial,
$0 \le R(x, y) \le \min\{x, y\}$,
$R(ax, ay) = aR(x, y)$ for all $a, x, y > 0$ (homogeneity).

Let $\ell^\infty([a, b])$ denote the set of all uniformly bounded, real functions on an interval $[a, b]$ and let "$\rightsquigarrow$" denote weak convergence. The functional central limit theorem for the weighted version of our tail empirical process V_p is as follows.

THEOREM 2.1. *Under Assumptions 2.1–2.4, for any $0 \le \eta < \frac{1}{2}$,*

$$\frac{V_p}{I^\eta} \rightsquigarrow \frac{V}{I^\eta}, \quad \textit{in } \ell^\infty([0, 2]),$$

where I denotes the identity function, $0/0 := 0$ and V is a centered Gaussian process with continuous sample paths and with covariance function given by

$$\operatorname{cov}\big(V(x), V(y)\big) = \min\{x, y\} - R(x, y). \tag{3}$$

The novel finding is that the R-function measuring heterogeneity plays an essential role in the general limiting process V, which is a standard Wiener process in case of a trivial R.

Next, we show how to apply this theorem to obtain a new and unified limit result for the Hill (1975) estimator of the extreme value index for both heterogeneous and homogeneous data. Denote the $k+1$ upper-order statistics by $X_{p-k:p} \le \cdots \le X_{p:p}$. Then we estimate the extreme value index $\gamma > 0$ by the Hill estimator:

$$\widehat{\gamma} = \frac{1}{k}\sum_{i=0}^{k-1} \log X_{p-i:p} - \log X_{p-k:p}.$$

To apply the tail empirical process theory, we introduce a slightly modified tail empirical process by replacing $T_p(\frac{u_p}{x^\gamma})$ with its approximate value $\frac{k}{p}x$ under regular variation:

$$W_p(x) = \frac{p}{\sqrt{k}}\left(T_{\text{emp}}\left(\frac{u_p}{x^\gamma}\right) - \frac{k}{p}x\right), \qquad x > 0.$$

Using the Skorohod representation theorem, we obtain the following.

COROLLARY 2.1. *Under the conditions of Theorem 2.1 and if for some $0 \le \eta < \frac{1}{2}$ and $\delta > 0$,*

$$\sup_{0<x<1+\delta} \frac{|\sqrt{k}(\frac{p}{k}T_p(\frac{u_p}{x^\gamma}) - x)|}{x^\eta} \to 0, \tag{4}$$

then there exists a probability space carrying probabilistically equivalent versions of W_p and V (still denoted with W_p and V), such that

$$\sup_{0<x<1+\delta} \frac{|W_p(x) - V(x)|}{x^\eta} \xrightarrow{\text{a.s.}} 0.$$

Now applying the Vervaat (1972) lemma yields the asymptotic normality of the intermediate-order statistics.

COROLLARY 2.2. *Under the conditions and on the probability space of Corollary 2.1,*

$$\sqrt{k}\left(\left(\frac{X_{p-k:p}}{u_p}\right)^{-1/\gamma} - 1\right) \xrightarrow{\text{a.s.}} -V(1) \sim \mathrm{N}\big(0, 1 - R(1,1)\big).$$

Our second key result, the asymptotic normality of the Hill estimator, then follows by rewriting it as a functional of the tail empirical process W_p like in Example 5.1.5 in de Haan and Ferreira (2006). Note that the asymptotic variance of the Hill estimator, through the process W_p, depends on the R-function from Assumption 2.4.

THEOREM 2.2. *Under the conditions and on the probability space of Corollary 2.1, if $\eta > 0$,*

$$\sqrt{k}(\widehat{\gamma} - \gamma) \xrightarrow{\text{a.s.}} \gamma\left(-V(1) + \int_0^1 V(x)\frac{dx}{x}\right) \stackrel{d}{=} \gamma V(1) \sim \mathrm{N}\big(0, \gamma^2(1 - R(1,1))\big).$$

The most striking of this result is the smaller (compared with the homogeneous case) limiting variance when R is nontrivial, for both the Hill estimator and the intermediate empirical quantile. The proportion of reduction $R(1, 1)$ is the same for both. In general, a stronger heterogeneity yields a larger $R(1, 1)$, and hence a smaller limiting variance for the Hill estimator and the empirical quantile. Ignoring this would yield oversized asymptotic confidence intervals. In the next section, we calculate the value of $R(1, 1)$ explicitly for some heterogeneous scales models. We show that it can take a wide range of values in [0, 1] depending on the probability distributions of the individual data.

Actually it is well known in empirical process theory that the i.i.d. case leads to the largest variance and that heterogeneity reduces the variance. For this, observe that $pF_{\text{emp}}(x)$ has a Binomial-$(p, F_p(x))$ distribution in the i.i.d. case with variance $pF_p(x)(1 - F_p(x))$, whereas in the heterogeneous case it is a sum of independent Bernoulli random variables with smaller variance $\sum_{i=1}^{p} F_{pi}(x)(1 - F_{pi}(x))$. This is also the case when computing the (co)variance of the tail empirical process V_p. It is H_p there, which leads through its limit R to the reduced variance of the Hill estimator.

Note that condition (4) only depends on the average distribution function $F_p = 1 - T_p$, and hence does not take the heterogeneity into account. In other words, the condition is the same for i.i.d. data from F_p. Now assume for simplicity that in the present setup F_p does not depend on p. Then it can be shown that condition (4), for each $0 \le \eta < \frac{1}{2}$, is implied by $\sqrt{k}A(p/k) \to 0$, where A is the usual auxiliary function in the second-order condition; see, for example, Theorem 3.2.5 in de Haan and Ferreira (2006).

EXAMPLE 1. Let us also consider here a simple example of the $X_i^{(p)}$ to show the scope of the results. Set $X_i^{(p)} = (Z_i \log Z_i)^{1/(1+i/p)}$, where the Z_i, $i = 1, \dots, p$, are i.i.d. standard Pareto distributed. Then $X_i^{(p)}$ has extreme value index $1/(1+i/p) \in [1/2, 1)$. Hence, all $X_i^{(p)}$ have different extreme value indices. The "joint" γ defined in Assumption 2.1 is equal to $\lim_{p\to\infty} \max_{i=1,\dots,p} 1/(1+i/p) = 1$ and $T(x)$ can be chosen to be $1/x$. The function R and in particular $R(1, 1)$ are equal to 0. Hence, although all the $X_i^{(p)}$ have different tail behavior, the Hill estimator has the same asymptotic behavior as in the i.i.d. case, with asymptotic variance $\gamma^2 = 1$.

EXAMPLE 2. Observe that Theorem 2.2 immediately extends to certain dependent data. Suppose that we do not observe the $X_i^{(p)}$ directly, but that we observe

$$Y_i^{(p)} = ZX_i^{(p)}, \quad i = 1, \dots, p,$$

where $Z > 0$ is an unobservable random variable. Then the Hill estimator based on the $Y_i^{(p)}$ is equal to that of the latent $X_i^{(p)}$, and hence Theorem 2.2 applies.

For accurate statistical inference based on Theorem 2.2, we need to know or estimate $R(1, 1)$. When the data are homogeneous, or much more generally, if it is known that $R(1, 1) = 0$, we simply omit the factor $1 - R(1, 1)$ and we can, for example, construct the usual asymptotic confidence intervals for γ. In case $R(1, 1)$ is not known, it is not clear how to estimate it from the $X_i^{(p)}$ only, but a solution exists if a duplicate sample $\widetilde{X}_i^{(p)} \sim X_i^{(p)}$, $i = 1, \dots, p$, is available, where all variables are mutually independent. Let $R_i = \sum_{j=1}^{p} \mathbb{1}[X_j^{(p)} \le X_i^{(p)}]$ and $\widetilde{R}_i = \sum_{j=1}^{p} \mathbb{1}[\widetilde{X}_j^{(p)} \le \widetilde{X}_i^{(p)}]$ denote the ranks of $X_i^{(p)}$ and $\widetilde{X}_i^{(p)}$ in their own array. Then we may estimate $R(x, y)$ consistently through

$$\widehat{R}(x, y) = \frac{1}{k} \sum_{i=1}^{p} \mathbb{1}[R_i > p - kx, \widetilde{R}_i > p - ky]. \tag{5}$$

Note that this rank-based estimator is invariant under, possibly unknown, increasing transformations of the data arrays.

THEOREM 2.3. *Under Assumptions 2.1–2.4, we have for* $T > 0$,

$$\sup_{0 \le x, y \le T} \left| \widehat{R}(x, y) - R(x, y) \right| \xrightarrow{\mathbb{P}} 0.$$

3. Leading example: Heterogeneous scales model. In this section, we study a particular type of heterogeneous model related to those in Einmahl, de Haan and Zhou (2016) and Einmahl and He (2023). We illustrate how the regular variation specified in Assumption 2.1 can emerge naturally as the dimension p grows and why the asymptotic behavior of the Hill estimator may change with the distribution of individual data in general. Let $Z_1, \dots, Z_p$ be i.i.d. latent continuous random variables. Consider the following independent, but nonidentically distributed data

$$X_i^{(p)} = \mu + Q_\sigma\big(1 - \pi(i)/p\big) Z_i, \quad i = 1, \dots, p, \tag{6}$$

where $\mu \in \mathbb{R}$, Q_σ is the generalized quantile function of a continuous distribution function F_σ, with positive left endpoint, and where π is an unknown permutation of $1, \dots, p$. Clearly, the $Q_\sigma(1 - \pi(i)/p)$ are scale parameters. (Observe that the permutation π does not affect the distribution of the order statistics of the $X_i^{(p)}$, and hence also not that of V_p and $\widehat{\gamma}$, but on the other hand, it allows nonsmooth changes of the distribution of $X_i^{(p)}$ in i.) Define the tail quantile function $U_\sigma(t) = Q_\sigma(1 - 1/t)$ and assume throughout that the function $t \mapsto \log U_\sigma(e^t)$ is Lipschitz-continuous on $[0, \infty)$. Denote $Z_+ = \max\{Z, 0\}$, with $Z := Z_1$, and write S and g for the survival function and the probability density function of Z, respectively, hence $\mathbb{E} Z_+^{1/\gamma} = \int_0^\infty S(v^\gamma)\, dv$.

THEOREM 3.1 (Nontrivial limit). *Suppose that* U_σ *obeys a power law such that* $\lim_{t \to \infty} U_\sigma(t)/t^\gamma \in (0, \infty)$ *exists for some positive extreme value index* γ. *If there exists a non-increasing, right-continuous function* h *on* $[0, \infty)$ *such that* $xg(x) \le h(x)$ *for all* $x \ge 0$, *such that* $0 < \mathbb{E} Z_+^{1/\gamma} \le \int_0^\infty h(x^\gamma)\, dx < \infty$, *then*:

(i) *Assumptions 2.1 and 2.2 hold with extreme value index* γ *and*

$$T(x) = \int_0^\infty S\left(\frac{x - \mu}{u}\right) dF_\sigma(u);$$

(ii) *Assumption 2.4 holds with a nontrivial* R*-function given by*

$$R(x, y) = \frac{\int_0^\infty S((v/x)^\gamma) S((v/y)^\gamma)\, dv}{\int_0^\infty S(v^\gamma)\, dv} =: R_\gamma(x, y), \quad x, y > 0.$$

THEOREM 3.2 (Trivial limit). *Suppose that* S *obeys a power law such that* $\lim_{t \to \infty} t^{1/\gamma} S(t) \in (0, \infty)$ *exists for some positive extreme value index* γ, *and* $xg(x) \le MS(x)$, $x \ge 0$, *for some constant* $M < \infty$. *If* $\int_0^\infty x^{1/\gamma}\, dF_\sigma(x) < \infty$, *then the results in Theorem 3.1 remain true except the* R*-function becomes trivial* ($R \equiv 0$).

It should be emphasized that under the conditions of Theorems 3.1 or 3.2, we (obviously) obtain the asymptotic normality of the Hill estimator $\widehat{\gamma}$ through Theorem 2.2, if condition (4) is satisfied for some $\eta, \delta > 0$. In case of Theorem 3.2, although the setup allows substantial heterogeneity, the limiting variance γ^2 is the same as in the i.i.d. case, whereas in case of Theorem 3.1 the limiting variance is smaller than γ^2. Observe that in the latter case, although the individual Z_i can be light-tailed, the large heterogeneity generates a positive γ, a heavy tail.

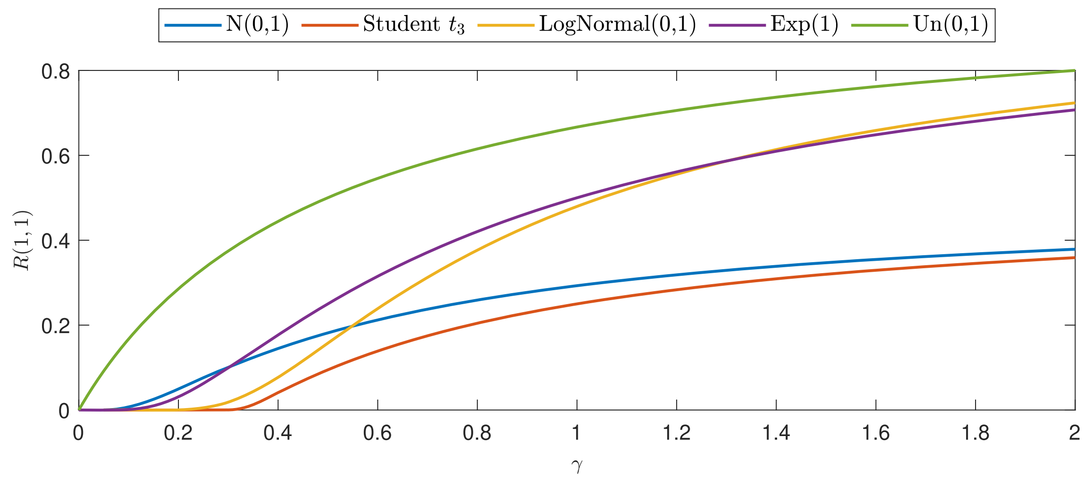

FIG. 2. *$R(1,1)$ for various distributions of the Z_i.*

EXAMPLE 3. If it is known that the Z_i are, for instance, standard normally distributed, then $X_i^{(p)} \sim \mathrm{N}(\mu, Q_\sigma^2(1-\pi(i)/p), i = 1, \ldots, p$. Write $\overline{\Phi}$ for the standard normal survival function, then $R_\gamma(1,1) = \int_0^\infty \overline{\Phi}^2(v^\gamma)\,dv / \int_0^\infty \overline{\Phi}(v^\gamma)\,dv$. Now under the (remaining) assumptions of Theorem 3.1, Assumption 2.3, and (4), we have $R_{\widehat{\gamma}}(1,1) \xrightarrow{\mathbb{P}} R_\gamma(1,1)$, and hence the asymptotic variance of the Hill estimator can be estimated consistently: $\widehat{\gamma}^2(1 - R_{\widehat{\gamma}}(1,1)) \xrightarrow{\mathbb{P}} \gamma^2(1 - R_\gamma(1,1))$.

The setup of Theorem 3.2 can be compared to that in Einmahl, de Haan and Zhou (2016). If the scedasis function c therein is nonincreasing it is equal to $Q_\sigma^{1/\gamma}(1-\cdot)$ here. The condition $\int_0^\infty x^{1/\gamma}\,dF_\sigma(x) < \infty$ then translates to $\int_0^1 c(u)\,du < \infty$ as in Einmahl, de Haan and Zhou (2016), but note that the condition therein that c, and hence Q_σ, is bounded is not required here. Indeed, it is natural to allow that the quantile function $Q_\sigma(s) \to \infty$ as $s \uparrow 1$.

EXAMPLE 4. From Theorem 3.1, we can calculate $R(1,1)$ for various distributions of the Z_i as follows:

- For the Weibull distribution $S(x) = \exp(-x^\tau)$, $\tau > 0$, we find $R(1,1) = 2^{-1/(\tau\gamma)}$. Hence, for the standard exponential distribution we obtain $R(1,1) = 2^{-1/\gamma}$;
- For the Pareto distribution $S(x) = x^{-(1+\epsilon)/\gamma}$, $\epsilon > 0$, $R(1,1)$ does not depend on γ and is equal to $2\epsilon/(1+2\epsilon)$;
- For the uniform-(0, 1) distribution, $R(1,1) = 2\gamma/(2\gamma+1)$.

In general, we can compute $R(1,1)$ (or the entire function R) numerically for a given survival function S and extreme value index γ. Figure 2 above depicts the values of $R(1,1)$ for various distributions of the Z_i, as a function of the extreme value index γ.

4. Extreme quantile estimation. An important application of extreme value theory is the estimation of very high quantiles. In case of heterogeneous data, the notion of quantile may need some clarification. The empirical distribution function F_{emp} is a direct, nonparametric summary of the data and it estimates consistently the average distribution function F_p. Hence, the sample quantiles, the order statistics, estimate the inverse Q_p, which is hence "producing" the relevant quantiles. Similarly, if we are interested in the $(1-\tau)$-th quantile corresponding to our data, then we would like to find a number x_τ such that an "arbitrary" observation exceeds x_τ with probability τ. This probability is equal to $\frac{1}{p}\sum_{i=1}^{p} 1 - F_{pi}(x_\tau) = 1 - F_p(x_\tau)$ and again $x_\tau = Q_p(1-\tau)$.

Now, we would like to estimate the extreme quantile $x_\tau = Q_p(1-\tau)$, where $\tau > 0$ is very small, that is, $\tau = \tau(p) \to 0$, as $p \to \infty$. We estimate such an extreme quantile as usual with

$$\widehat{x}_\tau = X_{p-k:p}\left(\frac{k}{p\tau}\right)^{\widehat{\gamma}}. \tag{7}$$

We require the following condition, which is formulated such that it bears some similarity with condition (4):

$$\sqrt{k}\left(\frac{p}{k}T_p\left(\frac{u_p}{x^\gamma}\right) - x\right)/x \to 0, \quad \text{for } x = (u_p/x_\tau)^{1/\gamma}. \tag{8}$$

THEOREM 4.1. *Assume* $p\tau/k \to \nu \in [0,1)$, $(\log p\tau)/\sqrt{k} \to 0$, *condition* (8) *and assume that the conditions of Corollary* 2.1 *hold for some* $0 < \eta < \frac{1}{2}$, *then*

$$\frac{\sqrt{k}}{\log(k/(p\tau))}\log\frac{\widehat{x}_\tau}{x_\tau} \xrightarrow{d} \mathrm{N}\left(0, \gamma^2\left((1-w)^2(1-R(1,1)) + 2w\int_0^1 (1-R(1/x,1))\,dx\right)\right),$$

where $w = 1/\log(1/\nu)$ *should be read as* 0 *for* $\nu = 0$.

Again, when R is nontrivial, the limiting variance is smaller than in the homogeneous case, leading to shorter confidence intervals for extreme quantiles. When $\nu = 0$, as usually assumed, the limiting variance is equal to that of the Hill estimator: $\gamma^2(1 - R(1,1))$. Taking $\nu = p\tau/k$ instead of 0 in the limiting normal distribution may improve statistical performance in finite samples. The proof of Theorem 4.1 is deferred to the Supplementary Material.

5. Simulations. We consider three sets of Monte Carlo simulations to illustrate how the asymptotic behavior of the Hill estimator and that of the extreme quantile estimator change for heterogeneous data. First, we fix the extreme value index γ but vary the distribution of Z in the heterogeneous scales model. Second, we specify the distribution of Z but change the extreme value index γ. Third, we study three miscellaneous examples. In all cases, we generate 5000 replications of heterogeneous data arrays of a large dimension $p = 1000$ and take $k = 50$.

5.1. *Heterogeneous scales model with fixed* γ. We generate independent random variables from the heterogeneous scales model

$$X_i^{(p)} = Q_\sigma(1 - i/p)Z_i = \left(\frac{p}{i}\right)^\gamma Z_i \tag{9}$$

with the quantile function $Q_\sigma(u) = (1-u)^{-\gamma}$ of the Pareto distribution with extreme value index γ.

We fix $\gamma = 1$ and generate i.i.d. latent variables Z_i from three classes of light(er) tailed distributions:

(I) Un$(1, 1 + 1/\theta)$, for $\theta > 0$, with $R(1,1) = \frac{6\theta+2}{6\theta+3}$;
(II) Pareto$(1 + \epsilon)$, for $\epsilon > 0$, with $R(1,1) = \frac{2\epsilon}{2\epsilon+1}$;
(III) Weibull(κ), for $\kappa > 0$, with $R(1,1) = 2^{-1/\kappa}$.

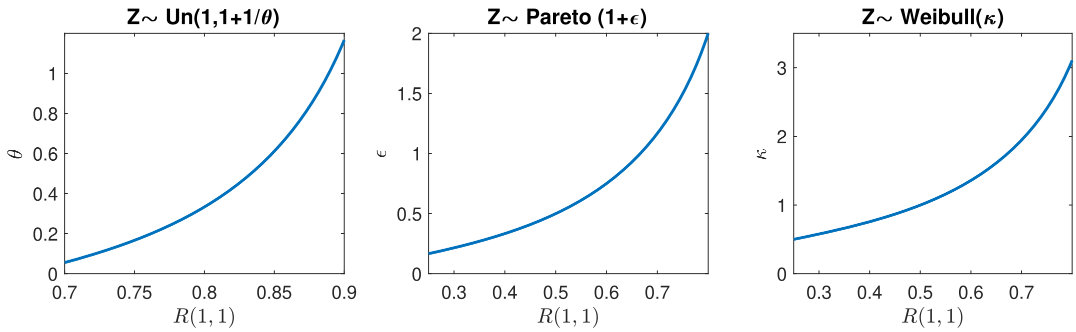

FIG. 3. *Parameter as a function of* $R(1,1)$.

To render comparable results across these classes, we control for $R(1,1)$ and solve the corresponding parameter in every case. Note that the ranges of $R(1,1)$ are different: $(2/3,1)$ for case I and $(0,1)$ for cases II and III. Figure 3 shows the parameters as a function of $R(1,1)$.

Figure 4 compares the variance of $\sqrt{k}(\widehat{\gamma}-\gamma)=\sqrt{k}(\widehat{\gamma}-1)$ over the replications in Monte Carlo simulations with the asymptotic variance in Theorem 3.1. Recall that for i.i.d. data, the asymptotic variance of $\sqrt{k}(\widehat{\gamma}-\gamma)$ is equal to the (much) larger value $\gamma^2=1$. The variance curves match very well for all three distributions, showing that our asymptotic theory yields a good approximation for finite samples. The variance of the Hill estimator decreases with the corresponding parameter θ, ϵ or τ, respectively. In fact, the asymptotic variance $1-R(1,1)\to 0$ is vanishing for all cases as $\theta,\epsilon,\tau\to\infty$. When these parameter values approach ∞, the Z_i, and hence the Hill estimator are becoming deterministic.

Figure 5 compares the boxplots of the Hill estimator for the heterogeneous data (blue) with those for i.i.d. data (red) generated from the average distribution F_p. Indeed, the Hill estimator shows a much smaller spread for heterogeneous data, while the median relative errors are close to 0 in both setups. Boxplots for the extreme quantile estimator $\widehat{x}_\tau$ for $\tau=1/200$ show similar patterns and are available in the Supplementary Material.

5.2. *Heterogeneous scales model with different* γ. We again simulate heterogeneous data $X_i^{(p)}$ according to the heterogeneous scales model (9). We change the value of the extreme value index γ, and generate i.i.d. Z_i from:

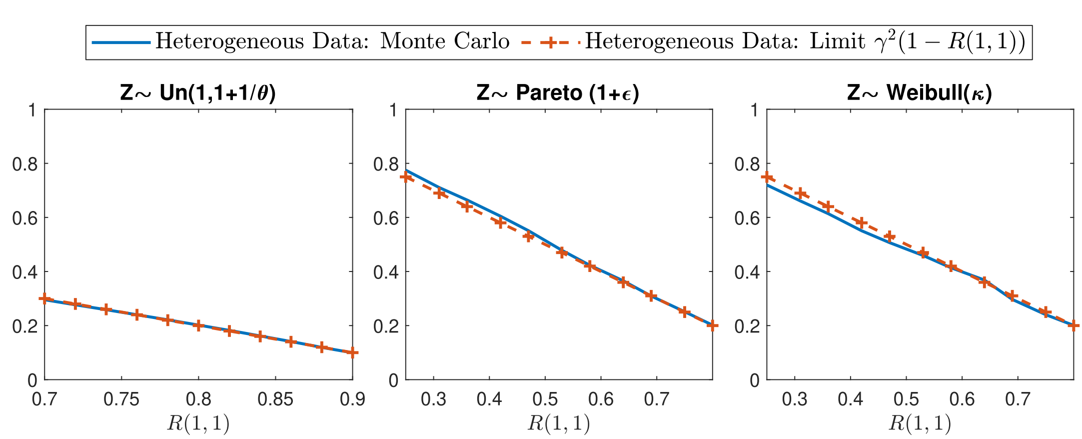

FIG. 4. *Variance of* $\sqrt{k}(\widehat{\gamma}-\gamma)$ *over simulation replications and its theoretical limit as a function of* $R(1,1)$.

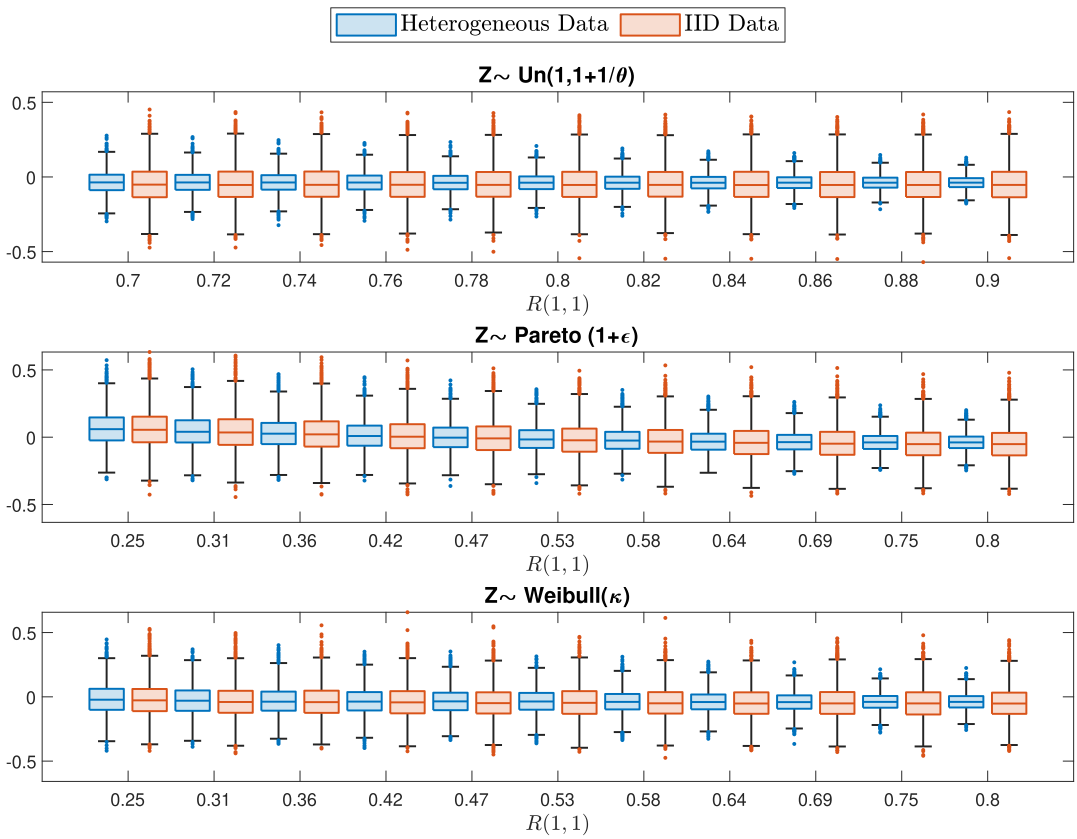

FIG. 5. *Boxplots of estimation error* $\widehat{\gamma} - 1$.

(I) the standard uniform distribution with $R(1,1) = \frac{2\gamma}{2\gamma+1}$;
(II) the standard normal distribution with $R(1,1) = \int_0^\infty \bar{\Phi}^2(v^\gamma)\,dv / \int_0^\infty \bar{\Phi}(v^\gamma)\,dv$;
(III) the standard exponential distribution with $R(1,1) = 2^{-1/\gamma}$.

Figure 6 shows the variance of $\sqrt{k}(\widehat{\gamma} - \gamma)$ over simulation replications and the asymptotic variance. We also plot the asymptotic variance γ^2 for i.i.d. data as a benchmark. The finite-sample variances of the Hill estimator are again close to their limits in Theorem 3.1.

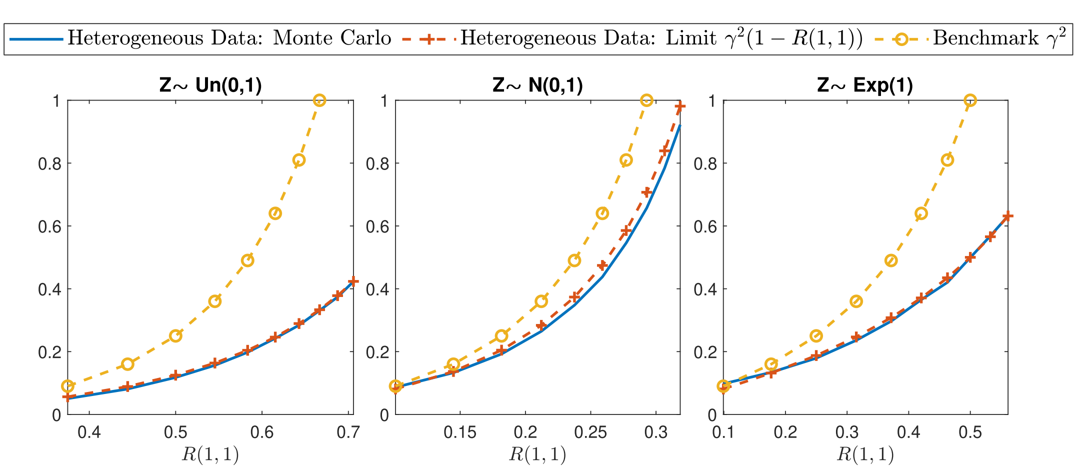

FIG. 6. *Variance of* $\sqrt{k}(\widehat{\gamma} - \gamma)$ *over simulation replications and its theoretical limit.*

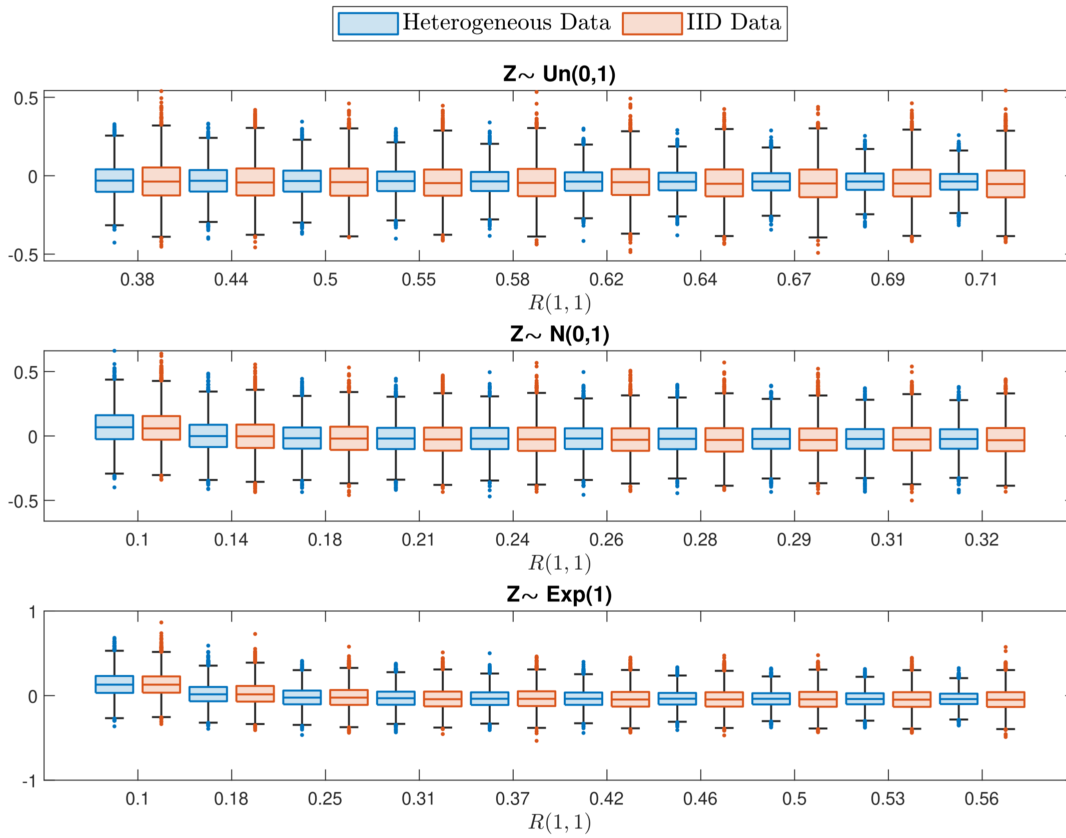

FIG. 7. *Boxplots of relative estimation error* $\widehat{\gamma}/\gamma - 1$.

According to our theory, the heterogeneous data lead to a smaller variance than γ^2, and the difference $\gamma^2 - \gamma^2(1 - R(1,1)) = \gamma^2 R(1,1)$ grows with γ (as well as with $R(1,1)$). The variance of the relative estimation error $\widehat{\gamma}/\gamma - 1$, however, is decreasing in $R(1,1)$. Figure 7 shows the corresponding boxplots for all three cases. The spread for heterogeneous data is consistently smaller than that for i.i.d. data and indeed decreases with $R(1,1)$. Again, boxplots for the extreme quantile estimator $\widehat{x}_\tau$, for $\tau = 1/200$, show similar patterns and are available in the Supplementary Material.

5.3. *Miscellaneous examples.* Let Z_i be i.i.d. latent variables from the standard Pareto distribution. We consider three examples:

(I) $X_i^{(p)} = (1 + \log(p/i))Z_i$;

(II) $X_i^{(p)} = (p/i)Z_i^{\gamma(i/p)}$ with extreme value index function $\gamma(u) = 1/(2(1+u)) \in (1/4, 1/2]$;

(III) $X_i^{(p)} = (Z_i \log Z_i)^{\gamma(i/p)}$ with extreme value index function $\gamma(u) = 1/(1+u) \in (1/2, 1]$.

Case I is another heterogeneous scales model, but one with a trivial limit as in Theorem 3.2. Note that $Q_\sigma(1 - i/p) = 1 + \log(p/i)$ can be seen as a scedasis $c(i/p)$, that violates the boundedness condition on c in Einmahl, de Haan and Zhou (2016). Cases II and III assign different extreme value indices $\gamma(i/p)$ to individual observations, but in Case III there is no scale factor dominating the individual $\gamma(i/p)$; see Example 1 for more details about Case III. For all cases, the extreme value indices of the average distribution function F_p are equal to $\gamma = 1$.

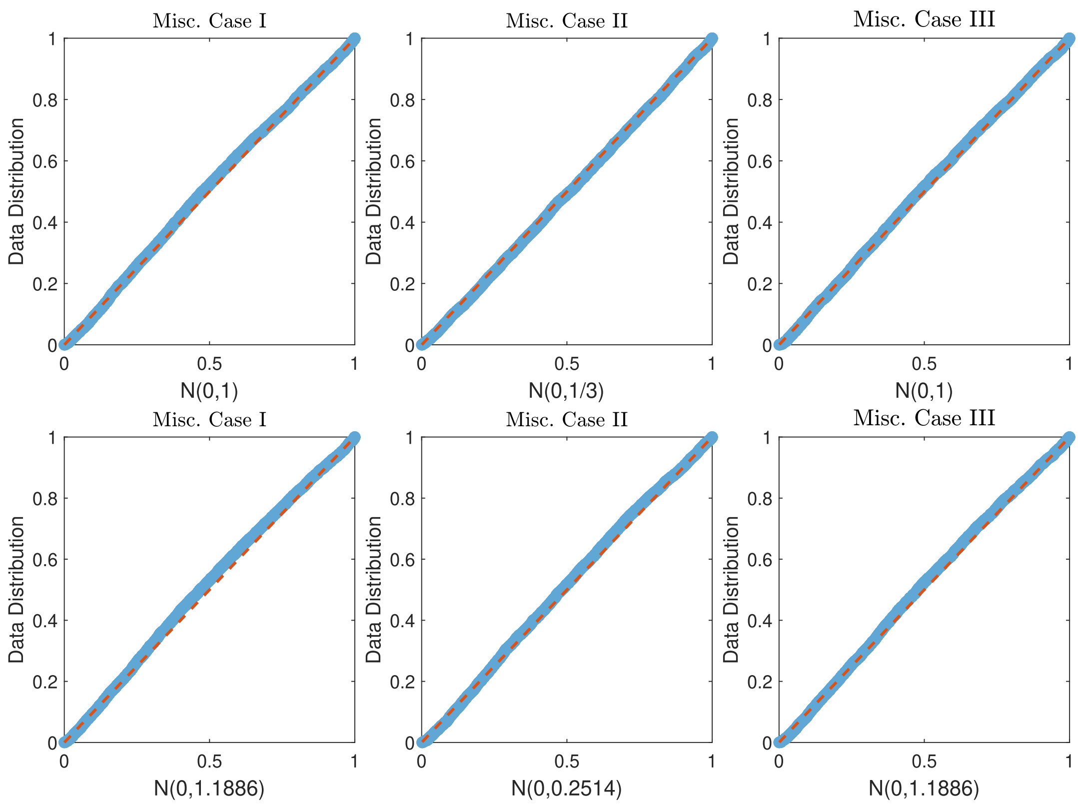

FIG. 8. *Probability-Probability (P-P) Plots for* $\sqrt{k}(\widehat{\gamma}-1)$ *in the top row and for* $\frac{\sqrt{k}}{\log(k/(p\tau))}\log\frac{\widehat{x}_\tau}{x_\tau}$ *with* $\tau = 1/200$ *in the bottom row.*

In the first row of Figure 8, Probability-Probability (PP) plots are shown of 5000 simulation replications of $\sqrt{k}(\widehat{\gamma}-1)$ against the limiting normal distribution. For Cases I and III, the R-function is trivial, and hence the limiting distribution is standard normal, whereas for Case II we find $R(1,1) = 2/3$, and hence the limiting variance is equal to $1/3$. The second row shows the plots for the quantile estimation error $\frac{\sqrt{k}}{\log(k/(p\tau))}\log\frac{\widehat{x}_\tau}{x_\tau}$, for $\tau = 1/200$. For Cases I and III, $R \equiv 0$ is trivial and taking $\nu = p\tau/k = 0.1$ gives an asymptotic variance of 1.1886. For Case II, we find $R(1/x, 1) = 1 - \frac{x}{3}$, $0 < x \leq 1$, which gives an asymptotic variance of 0.2514. The PP-plots are very close to the diagonal, showing again that our asymptotic theory works well for finite samples.

6. Real-life examples. This section presents two real-life examples. Our first example is a global data set of the 1858 most significant earthquakes from 1981 to 2021 provided by the National Oceanic and Atmospheric Administration depicted in Figure 1 already. To avoid the ties in the magnitudes (caused by rounding at 1 decimal), for each group of repeated values we add equally-spaced corrections on the interval $(-0.05, 0.05)$ to the data. Then for each earthquake with (corrected) magnitude of M, we compute its seismic energy, in petajoules, using the Gutenberg–Richter energy–magnitude relationship given by $E = 10^{1.5(M-6.8)}$. On the left in Figure 9 is the log-log plot showing the data ranks for the $k = 100$ largest seismic energies in descending order as a function of the data values, on logarithmic axes. The observations concentrate around a straight line according to the Gutenberg–Richter law in seismology with a slope of $-1/\widehat{\gamma}$, where the Hill estimate $\widehat{\gamma} = 0.9763$ indicates a very heavy tail. Our sample spans 40 years with an average number of 46.45 earthquakes per year. The estimate (7) of

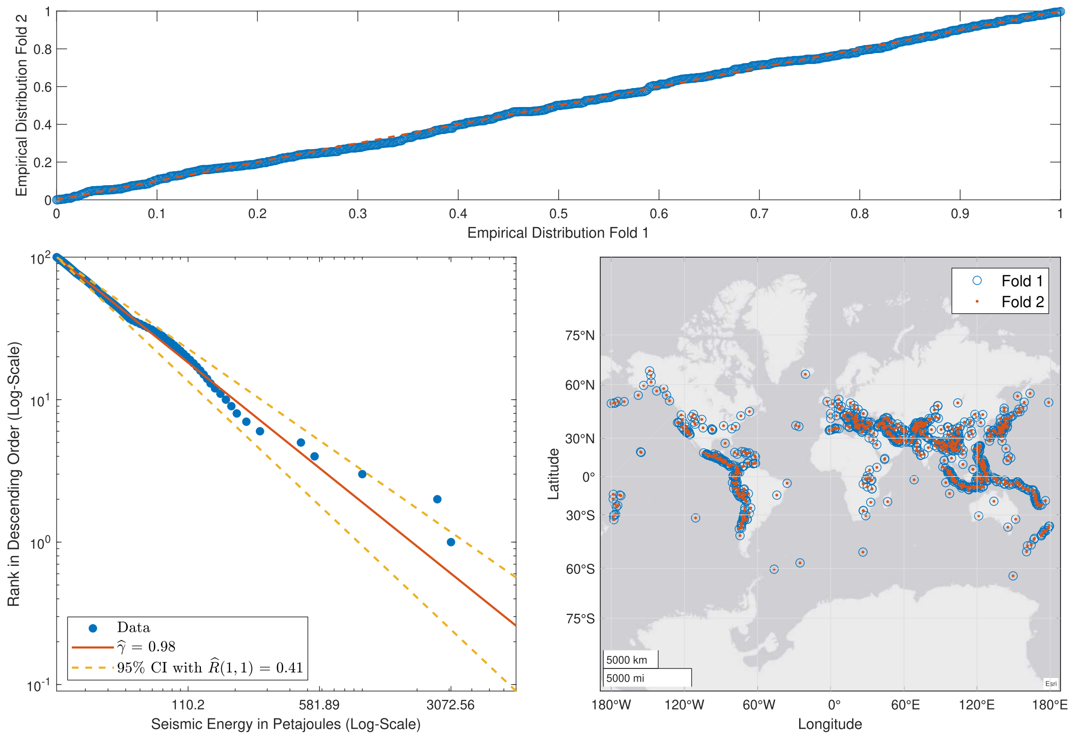

FIG. 9. *Global significant earthquakes from* 1981 *to* 2021.

the extreme quantile x_τ for the energy of a 100-year earthquake with $\tau = 1/4645$ is equal to $\widehat{x_\tau} = 4528.83$, corresponding to a magnitude of 9.24.

We split the sample into halves according to Figure 1 and estimate the R-function by $\widehat{R}$ in (5) of the two folds. On the bottom right of Figure 9 is the geographic map and we indicate the locations of the earthquakes in different folds by circles and dots, respectively. The P-P plot above the map shows a good alignment of the data distributions between the two folds, which agrees with our assumption of duplicate samples in Theorem 2.3. The rather large estimate of $\widehat{R}(1, 1) = 0.41$ suggests a substantial spatial heterogeneity of energies and a nontrivial limit of the Hill estimator. Using Theorem 2.2, we then obtain a relatively narrow 95% asymptotic confidence interval (0.8293, 1.1233) for γ, indicated by dashed lines in the log-log plot in Figure 9 (keeping the y-intercept unchanged). We also obtain the 95% asymptotic confidence interval for x_τ corresponding to the magnitude interval (8.9940, 9.4807), which is 24% narrower than the interval (8.9189, 9.5557) based on the i.i.d. model.

In the second example, we estimate the extreme value index for the cross-sectional distributions of daily losses on NYSE/AMEX/NASDAQ stocks with share codes 10 and 11 (i.e., ordinary common shares) in the last quarter of 2019. We use only the $p = 1000$ firms with largest lagged market values on each day. We choose $k \leq 50$ as large as possible but not more than 10% of the number of positive observations.

Figure 10 shows the estimates and confidence intervals of the extreme value index and extreme quantile for $\tau = 0.01$ (known as 99% value-at-risk in finance) for 64 days of cross-sectional stock loss data in our data set. Interestingly, the estimated extreme value indices are usually (much) higher than the benchmark value 1/3 from the cubic power law (see, e.g., Gabaix et al. (2003)) for individual stocks, suggesting that heterogeneity rather than individuals drive the cross-sectional tails most of the time. The daily confidence intervals match well for two different estimators of the R-function: one uses the daily losses for the same firms 25 working days before the estimation date as auxiliary sample and then formula

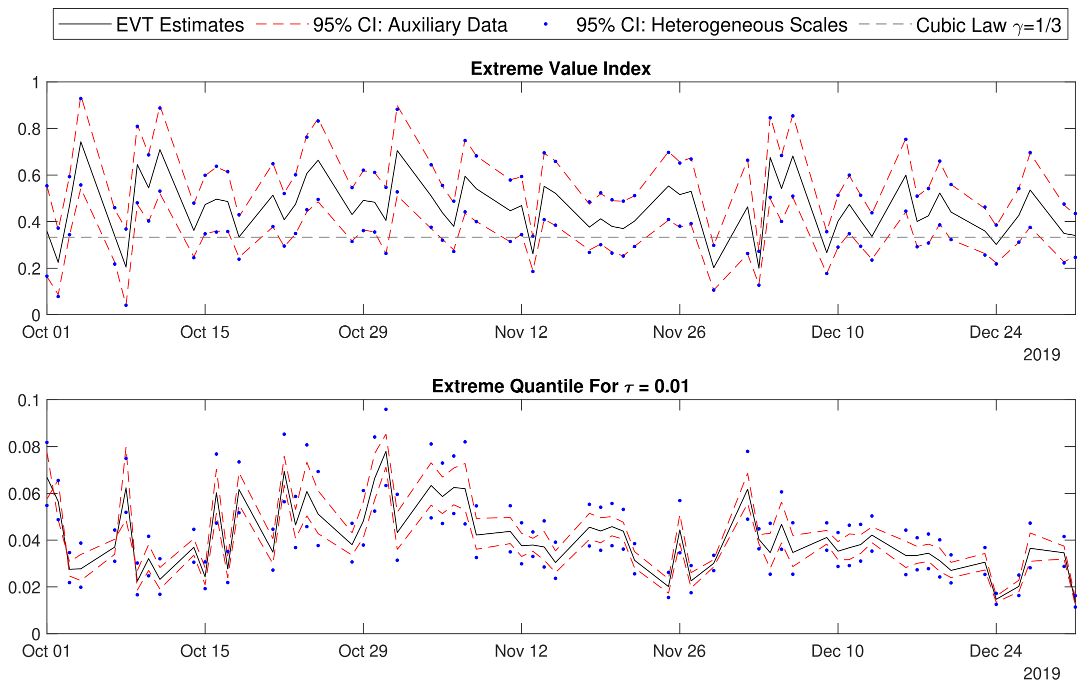

FIG. 10. *Daily estimates for cross-sectional stock loss data.*

(5), and the other one uses the heterogeneous scales model with Z a Student-t_3 variable (satisfying the cubic power law).

7. Proofs.

7.1. *Proofs from Section* 2. PROOF OF LEMMA 2.1. We only prove that Assumption 2.4 implies (1); the proofs of the converse is analogous and omitted. Let $\alpha = \alpha(p) \downarrow 0$ be any intermediate sequence in (1) such that $p\alpha \to \infty$. Take the intermediate threshold sequence $t = Q(1-\alpha) \to \infty$ where Q denotes the quantile function of T, then $T(t)/\alpha \to 1$ by regular variation. Substituting t into Assumption 2.4 gives that

$$R(x, y) = \lim_{p\to\infty} \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}\big(X_i^{(p)} > tx^{-\gamma}\big)\mathbb{P}\big(X_i^{(p)} > ty^{-\gamma}\big).$$

Let $\varepsilon > 0$ be small. For large p, by monotonicity of T_p,

$$\begin{aligned}
&\frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}\big(X_i^{(p)} > tx^{-\gamma}\big)\mathbb{P}\big(X_i^{(p)} > ty^{-\gamma}\big) \\
&\quad \le \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}\big(T_p\big(X_i^{(p)}\big) \le T_p\big(tx^{-\gamma}\big)\big)\mathbb{P}\big(T_p\big(X_i^{(p)}\big) \le T_p\big(ty^{-\gamma}\big)\big) \\
&\quad \le \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}\big(T_p\big(X_i^{(p)}\big) \le (1+\varepsilon)\alpha x\big)\mathbb{P}\big(T_p\big(X_i^{(p)}\big) \le (1+\varepsilon)\alpha y\big),
\end{aligned}$$

where the last line follows from Assumption 2.1. Changing the variables gives that

$$\liminf_{p\to\infty} \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}\big(T_p\big(X_i^{(p)}\big) \le \alpha x\big)\mathbb{P}\big(T_p\big(X_i^{(p)}\big) \le \alpha y\big) \ge \frac{1}{1+2\varepsilon} R(x, y).$$

Similarly, we obtain that

$$\limsup_{p\to\infty}\frac{1}{p\alpha}\sum_{i=1}^{p}\mathbb{P}\big(T_p\big(X_i^{(p)}\big)\le\alpha x\big)\mathbb{P}\big(T_p\big(X_i^{(p)}\big)\le\alpha y\big)\le\frac{1}{1-2\varepsilon}R(x,y).$$

Since $\varepsilon>0$ can be arbitrarily small, (1) follows. □

Next, we prove the functional central limit theorem for the weighted version of V_p. We first present some useful lemmas.

LEMMA 7.1. *Let T be regularly varying with index $-1/\gamma$ for some $\gamma>0$. For arbitrary $\delta>0$, there exists $t_0=t_0(\delta)$ and $C=C(\delta)$ such that for $t\ge t_0$,*

$$\frac{T(tx)}{T(t)}\le Cx^{-1/\gamma}\cdot\max\{x^{\delta},x^{-\delta}\},\qquad x>0.$$

PROOF. The case $x\ge1$ (or more generally $x\ge c$ for any given $c>0$) is due to Potter (1942); see also part 5 of Proposition B.1.9 in de Haan and Ferreira (2006)). The case $0<x<1$ follows from part 7 of Proposition B.1.9 in de Haan and Ferreira (2006). □

The next two lemmas give some important consequences of the stability condition.

LEMMA 7.2. *Under Assumptions 2.1 and 2.2, there exists $p_0>0$ such that $\lim_{z\to\infty}z^r\times T_p(z)=0$ for any $0\le r<1/\gamma$ and all $p\ge p_0$.*

PROOF. By Assumption 2.2, it suffices to prove that $\lim_{z\to\infty}z^rT(z)=0$. Let $0\le r<1/\gamma$, and take a $\delta\in(0,1/\gamma-r)$. By Lemma 7.1, there exists some $t_0=t_0(\delta)$, such that for all $z\ge t_0$,

$$z^rT(z)\le Cz^rT(t_0)\left(\frac{z}{t_0}\right)^{-1/\gamma+\delta}=CT(t_0)t_0^{1/\gamma-\delta}z^{r-1/\gamma+\delta}\to0,\qquad z\to\infty,$$

as the exponent $r-1/\gamma+\delta<0$. □

LEMMA 7.3. *Let S be any survival function with probability density function $g=-S'$, such that $tg(t)\le h(t)$, $t>t_0>0$, for some nonincreasing function h on $[t_0,\infty)$. Then*

$$S(t)-S(tx)\le h(t)\log x,\qquad t>t_0,x>1.$$

In particular, when $\log S(e^t)$ is Lipschitz-continuous with a Lipschitz constant K, the result holds for $h(t)=KS(t)$.

PROOF. Consider the function $\phi(z)=-S(e^z)$ with derivative $\phi'(z)=e^zg(e^z)\le h(e^z)$ for $z>\log t_0$. By the mean-value theorem and the monotonicity of h, for $t>t_0$ and $x>1$,

$$\begin{aligned}S(t)-S(tx)&=\phi(\log t+\log x)-\phi(\log t)\\&\le\sup_{0\le\delta\le\log x}\phi'(\log t+\delta)\log x\\&\le\sup_{0\le\delta\le\log x}h\big(\exp(\log t+\delta)\big)\log x=h(t)\log x.\end{aligned}$$

Observe that for the second part we have $tg(t)\le h(t)$ with $h(t)=KS(t)$. □

For convenience of presentation, we assume the conditions of Theorem 2.1 hold throughout the remainder of this subsection. All asymptotic results hold as $p\to\infty$ unless specified otherwise. The following lemma establishes the finite-dimensional (fidis) convergence for V_p.

LEMMA 7.4 (Fidis convergence). *For any* $0 \le \eta < \frac{1}{2}$,

$$\frac{V_p}{I^\eta} \rightsquigarrow \frac{V}{I^\eta}, \quad \text{in } \ell^\infty([0,2]),$$

provided the asymptotic tightness of V_p/I^η.

PROOF. It suffices to prove the finite-dimensional (fidis) convergence: for any given m and all fixed $0 < x_1 < \cdots < x_m \le 2$ and $\eta \in [0, 1/2)$,

$$\big(V_p(x_1)/x_1^\eta, \dots, V_p(x_m)/x_m^\eta\big) \xrightarrow{d} \big(V(x_1)/x_1^\eta, \dots, V(x_m)/x_m^\eta\big).$$

We use the Cramér–Wold device in conjunction with the Lindeberg central limit theorem; we can ignore the weights x_j^η here. The Lindeberg condition is satisfied since the indicators constituting T_{emp} are bounded by definition. We omit the standard details but note that $\text{cov}(V_p(x), V_p(y)) \to \text{cov}(V(x), V(y))$ by Assumptions 2.1, 2.3 and 2.4. □

It remains to verify the asymptotic tightness of V_p/I^η. We shall prove this for $0 < \eta < \frac{1}{2}$, as then the case $\eta = 0$ follows. For the proof of the asymptotic tightness, we use Theorem 3 in Einmahl and Segers (2021), which is a corollary to Theorem 2.11.9 in van der Vaart and Wellner (1996). For the clearness of notation, we relabel the $Z_{n,i}$ there by $Y_{p,i}$, $i = 1, \dots, p$, and define

$$Y_{p,i} = \frac{1}{\sqrt{k}} \mathbb{1}\big[X_i^{(p)} > u_p x^{-\gamma}\big]/x^\eta, \quad x \in \mathcal{F} = [0, 2],$$

which is bounded by

$$\|Y_{p,i}\|_{\mathcal{F}} := \sup_{0 \le x \le 2} \frac{1}{\sqrt{k}} \mathbb{1}\big[X_i^{(p)} > u_p x^{-\gamma}\big]/x^\eta \le \frac{1}{\sqrt{k}} \bigg(\frac{X_i^{(p)}}{u_p}\bigg)^{\eta/\gamma}. \tag{10}$$

LEMMA 7.5. *For any* $\eta \in (0, 1/2)$ *and* $\lambda > 0$,

$$\sum_{i=1}^p \mathbb{E}\big[\|Y_{p,i}\|_{\mathcal{F}} \mathbb{1}\big[\|Y_{p,i}\|_{\mathcal{F}} > \lambda\big]\big] \to 0.$$

PROOF. It follows from (10) that

$$\begin{aligned}
&\sum_{i=1}^p \mathbb{E}\big(\|Y_{p,i}\|_{\mathcal{F}} \mathbb{1}\big[\|Y_{p,i}\|_{\mathcal{F}} > \lambda\big]\big) \\
&\quad \le \sum_{i=1}^p \int_{u_p(\lambda\sqrt{k})^{\gamma/\eta}}^\infty \frac{1}{\sqrt{k}} \bigg(\frac{x}{u_p}\bigg)^{\eta/\gamma} dF_{pi}(x) \\
&\quad = p \int_{u_p(\lambda\sqrt{k})^{\gamma/\eta}}^\infty \frac{1}{\sqrt{k}} \bigg(\frac{x}{u_p}\bigg)^{\eta/\gamma} dF_p(x) \\
&\quad = \lambda p \int_1^\infty z^{\eta/\gamma}\, dF_p\big(u_p(\lambda\sqrt{k})^{\gamma/\eta} z\big) \\
&\quad = \frac{p}{k} T_p(u_p) \cdot \lambda k \bigg[-\int_1^\infty z^{\eta/\gamma}\, d\frac{T_p(u_p(\lambda\sqrt{k})^{\gamma/\eta} z)}{T_p(u_p)}\bigg]
\end{aligned}$$

Integration by parts using Lemma 7.2 and then applying Assumptions 2.1 and 2.2, there exists $M > 0$ such that for all large p,

$$
\begin{aligned}
&-\int_1^\infty z^{\eta/\gamma}\, d\frac{T_p(u_p(\lambda\sqrt{k})^{\gamma/\eta}z)}{T_p(u_p)} \\
&= \frac{T_p(u_p(\lambda\sqrt{k})^{\gamma/\eta})}{T_p(u_p)} + \int_1^\infty \frac{T_p(u_p(\lambda\sqrt{k})^{\gamma/\eta}z)}{T_p(u_p)}\, dz^{\eta/\gamma} \\
&\le 2M\biggl(\frac{T(u_p(\lambda\sqrt{k})^{\gamma/\eta})}{T(u_p)} + \int_1^\infty \frac{T(u_p(\lambda\sqrt{k})^{\gamma/\eta}z)}{T(u_p)}\, dz^{\eta/\gamma}\biggr).
\end{aligned}
$$

Now applying Lemma 7.1, for any sufficiently small $\delta > 0$, and for all $p \ge p_0$ (because $u_p \to \infty$) with p_0 depending on δ,

$$
\frac{T(u_p(\lambda\sqrt{k})^{\gamma/\eta})}{T(u_p)} \le C\bigl\{(\lambda\sqrt{k})^{\gamma/\eta}\bigr\}^{-1/\gamma+\delta}, \quad \text{and}
$$

$$
\int_1^\infty \frac{T(u_p(\lambda\sqrt{k})^{\gamma/\eta}z)}{T(u_p)}\, dz^{\eta/\gamma} \le C\int_1^\infty \bigl\{(\lambda\sqrt{k})^{\gamma/\eta}z\bigr\}^{-1/\gamma+\delta}\, dz^{\eta/\gamma}.
$$

Combining these bounds, we obtain

$$
\begin{aligned}
&\frac{T(u_p(\lambda\sqrt{k})^{\gamma/\eta})}{T(u_p)} + \int_1^\infty \frac{T(u_p(\lambda\sqrt{k})^{\gamma/\eta}z)}{T(u_p)}\, dz^{\eta/\gamma} \\
&\le C\bigl\{(\lambda\sqrt{k})^{\gamma/\eta}\bigr\}^{-1/\gamma+\delta}\biggl\{1 + \int_1^\infty z^{-1/\gamma+\delta}\, dz^{\eta/\gamma}\biggr\} \\
&= C\bigl\{(\lambda\sqrt{k})^{\gamma/\eta}\bigr\}^{-1/\gamma+\delta}\frac{1-\delta\gamma}{1-\eta-\delta\gamma}.
\end{aligned}
$$

Hence, for sufficiently small $\delta > 0$,

$$
\begin{aligned}
&\sum_{i=1}^p \mathbb{E}\bigl(\|Y_{p,i}\|_{\mathcal{F}}\mathbb{1}\bigl[\|Y_{p,i}\|_{\mathcal{F}} > \lambda\bigr]\bigr) \\
&\le 2MC\cdot\lambda k\cdot\bigl\{(\lambda\sqrt{k})^{\gamma/\eta}\bigr\}^{-1/\gamma+\delta}\frac{1-\delta\gamma}{1-\eta-\delta\gamma} \\
&= 2MC\lambda^{1-1/\eta+\delta\gamma/\eta}k^{1-1/(2\eta)+\delta\gamma/(2\eta)}\frac{1-\delta\gamma}{1-\eta-\delta\gamma} \to 0,
\end{aligned}
$$

where the last step uses that $k \to \infty$ and its exponent $1 - \frac{1}{2\eta} + \frac{\delta\gamma}{2\eta} < 0$ for small $\delta > 0$. □

LEMMA 7.6. *Let $\varepsilon > 0$ be small and define $a = \varepsilon^{3/(1/2-\eta)}$ and $\mathcal{F}_a = [0, 2a]$. Then there exists a constant p_0 not depending on ε such that for every $p \ge p_0$,*

$$
\sum_{i=1}^p \mathbb{E}\sup_{x,y\in\mathcal{F}_a}\bigl(Y_{p,i}(x) - Y_{p,i}(y)\bigr)^2 \le \varepsilon^2.
$$

PROOF. We have

$$
\sum_{i=1}^p \mathbb{E}\sup_{x,y\in\mathcal{F}_a}\bigl(Y_{p,i}(x) - Y_{p,i}(y)\bigr)^2 \le 4\sum_{i=1}^p \mathbb{E}\sup_{x\in\mathcal{F}_a} Y_{p,i}^2(x)
$$

$$
\begin{aligned}
&= \frac{4}{k}\sum_{i=1}^{p}\mathbb{E}\sup_{x\in\mathcal{F}_a}\mathbb{1}\big[X_i^p > u_p x^{-\gamma}\big]/x^{2\eta} \\
&= \frac{4}{k}\sum_{i=1}^{p}\mathbb{E}\bigg(\frac{X_i^{(p)}}{u_p}\bigg)^{2\eta/\gamma}\mathbb{1}\big[X_i^{(p)} > u_p(2a)^{-\gamma}\big] \\
&= \frac{4p}{k}\int_{u_p(2a)^{-\gamma}}^{\infty}\bigg(\frac{x}{u_p}\bigg)^{2\eta/\gamma}\,dF_p(x) \\
&= -4\int_{(2a)^{-\gamma}}^{\infty} z^{2\eta/\gamma}\,d\frac{T_p(u_p z)}{T_p(u_p)}.
\end{aligned}
$$

Like in the proof of Lemma 7.5, using integration by parts and Lemma 7.1 we obtain that for some M not depending on ε and small $\delta > 0$ and $p \ge p_0$ with p_0 only depending on δ but not ε,

$$
\begin{aligned}
&\sum_{i=1}^{p}\mathbb{E}\sup_{x,y\in\mathcal{F}_a}\big(Y_{p,i}(x) - Y_{p,i}(y)\big)^2 \\
&\quad\le 8MC\frac{1-\delta\gamma}{1-2\eta-\delta\gamma}\cdot(2a)^{1-2\eta-\delta\gamma} \\
&\quad= 8MC\frac{1-\delta\gamma}{1-2\eta-\delta\gamma}\cdot 2^{1-2\eta-\delta\gamma}\varepsilon^{6-\frac{3}{1/2-\eta}\delta\gamma} \le \varepsilon^2,
\end{aligned}
$$

where the last step holds for all small $\varepsilon > 0$, because $6 - \frac{3}{1/2-\eta}\delta\gamma > 2$ for small $\delta > 0$. □

LEMMA 7.7. *Let $\varepsilon > 0$ be small. Define $\theta = 1 - \varepsilon^3$ and $\mathcal{F}_{(\ell)} = [2\theta^{\ell+1}, 2\theta^{\ell}]$, $\ell = 0.1, 2, \ldots$. Then there exists a constant p_0 not depending on ε such that for every $p \ge p_0$ and every ℓ,*

$$
\sum_{i=1}^{p}\mathbb{E}\sup_{x,y\in\mathcal{F}_{(\ell)}}\big(Y_{p,i}(x) - Y_{p,i}(y)\big)^2 \le \varepsilon^2.
$$

PROOF. For all $i \in \{1, \ldots, p\}$,

$$
\begin{aligned}
&\mathbb{E}\sup_{x,y\in\mathcal{F}_{(\ell)}}\big(Y_{p,i}(x) - Y_{p,i}(y)\big)^2 \\
&\quad\le \mathbb{E}\Big(\sup_{x\in\mathcal{F}_{(\ell)}} Y_{p,i}(x) - \inf_{x\in\mathcal{F}_{(\ell)}} Y_{p,i}(x)\Big)^2 \\
&\quad\le \mathbb{E}\bigg(\frac{1}{\sqrt{k}}\mathbb{1}\bigg[X_i^{(p)} > \frac{u_p}{(2\theta^{\ell})^{\gamma}}\bigg]\Big/(2\theta^{\ell+1})^{\eta} - \frac{1}{\sqrt{k}}\mathbb{1}\bigg[X_i^{(p)} > \frac{u_p}{(2\theta^{\ell+1})^{\gamma}}\bigg]\Big/(2\theta^{\ell})^{\eta}\bigg)^2 \\
&\quad= \frac{1}{k4^{\eta}}\mathbb{E}\bigg(\mathbb{1}\bigg[X_i^{(p)} > \frac{u_p}{(2\theta^{\ell})^{\gamma}}\bigg]\bigg(\frac{1}{\theta^{(\ell+1)\eta}} - \frac{1}{\theta^{\ell\eta}}\bigg) \\
&\qquad + \mathbb{1}\bigg[\frac{u_p}{(2\theta^{\ell})^{\gamma}} < X_i^{(p)} \le \frac{u_p}{(2\theta^{\ell+1})^{\gamma}}\bigg]\Big/\theta^{\ell\eta}\bigg)^2 \\
&\quad\le \frac{2}{k4^{\eta}}\bigg\{T_{pi}\bigg(\frac{u_p}{(2\theta^{\ell})^{\gamma}}\bigg)\frac{1}{\theta^{2\ell\eta}}\bigg(\frac{1}{\theta^{\eta}} - 1\bigg)^2 + \bigg(T_{pi}\bigg(\frac{u_p}{(2\theta^{\ell})^{\gamma}}\bigg) - T_{pi}\bigg(\frac{u_p}{(2\theta^{\ell+1})^{\gamma}}\bigg)\bigg)\frac{1}{\theta^{2\ell\eta}}\bigg\}.
\end{aligned}
$$

Hence,

$$\sum_{i=1}^{p} \mathbb{E} \sup_{x,y\in\mathcal{F}_{(\ell)}} \left(Y_{p,i}(x)-Y_{p,i}(y)\right)^2$$
$$\le \frac{2}{4^\eta}\frac{p}{k}T_p\left(\frac{u_p}{(2\theta^\ell)^\gamma}\right)\frac{1}{\theta^{2\ell\eta}}\left(\frac{1}{\theta^\eta}-1\right)^2+\frac{2}{4^\eta}\frac{p}{k}\left(T_p\left(\frac{u_p}{(2\theta^\ell)^\gamma}\right)-T_p\left(\frac{u_p}{(2\theta^{\ell+1})^\gamma}\right)\right)\frac{1}{\theta^{2\ell\eta}}$$
$$=: J_1+J_2.$$

Using Assumption 2.2 and Lemma 7.1, there exists constants $M, C>0$ not depending on ε such that for small $\delta>0$ and large p,

$$J_1 \le 4^{1-\eta}M\frac{T(\frac{u_p}{(2\theta^\ell)^\gamma})}{T(u_p)}\frac{1}{\theta^{2\ell\eta}}\left(\frac{1}{\theta^\eta}-1\right)^2$$
$$\le 8MC\theta^{\ell(1-2\eta-\gamma\delta)}\left(\frac{1}{\theta^\eta}-1\right)^2$$
$$\le 8MC\left(\frac{1}{\theta^{1/2}}-1\right)^2.$$

On the other hand, applying Lemma 7.3 and using Assumption 2.2, for large t and $x>1$,

$$T_p(t)-T_p(tx)\le MT(t)\log x, \quad \text{for some constant} M.$$

Hence, setting $t=\frac{u_p}{(2\theta^\ell)^\gamma}$ and $x=\theta^{-\gamma}$,

$$J_2 \le \frac{2M\gamma}{4^\eta}\frac{p}{k}T\left(\frac{u_p}{(2\theta^\ell)^\gamma}\right)\log(1/\theta)\frac{1}{\theta^{2\ell\eta}} \le \frac{4M\gamma}{4^\eta}\frac{T(\frac{u_p}{(2\theta^\ell)^\gamma})}{T(u_p)}\frac{1}{\theta^{2\ell\eta}}\log(1/\theta),$$

where for second inequality Assumption 2.1 is used. Lemma 7.1 yields, for some constants $C>0$ not depending ε, for $\delta>0$ small

$$J_2 \le \frac{4M\gamma C}{4^\eta}2^{1-\delta\gamma}\theta^{\ell(1-2\eta-\gamma\delta)}\log(1/\theta)\le 8M\gamma C\log(1/\theta).$$

To conclude,

$$\sum_{i=1}^{p} \mathbb{E} \sup_{x,y\in\mathcal{F}_{(\ell)}} \left(Y_{p,i}(x)-Y_{p,i}(y)\right)^2$$
$$\le 8MC\max\{1,\gamma\}\left\{\left(\frac{1}{\theta^{1/2}}-1\right)^2+\log(1/\theta)\right\}$$
$$\le 8MC\max\{1,\gamma\}(\varepsilon^6+2\varepsilon^3)\le\varepsilon^2,$$

for small $\varepsilon>0$. □

PROOF OF THEOREM 2.1. We have

$$\mathcal{F}=[0,2]=[0,2a]\cup\left(\bigcup_{\ell=0}^{\lfloor\log a/\log\theta\rfloor}\left[2\theta^{\ell+1},2\theta^\ell\right]\right).$$

The number of elements of this covering is bounded by ε^{-4}. The theorem follows now from Theorem 3 in Einmahl and Segers (2021), the conditions of which are verified in Lemmas 7.4–7.7 and by using this bound ε^{-4}. □

PROOF OF THEOREM 2.2. We have (see Example 5.1.5 in de Haan and Ferreira (2006))

$$\sqrt{k}(\widehat{\gamma}-\gamma)=\sqrt{k}\int_{X_{p-k:p}/u_p}^{1}\frac{p}{k}T_{\text{emp}}(su_p)\frac{ds}{s}+\sqrt{k}\int_1^\infty\left(\frac{p}{k}T_{\text{emp}}(su_p)-s^{-1/\gamma}\right)\frac{ds}{s}. \tag{11}$$

On the probability space of Corollary 2.1 (with $\eta>0$), the second term on the right-hand side in (11) is equal to

$$\begin{aligned}&\gamma\sqrt{k}\int_0^1\left(\frac{p}{k}T_{\text{emp}}\left(\frac{u_p}{x^\gamma}\right)-x\right)\frac{dx}{x}\\&=\gamma\int_0^1 V(x)\frac{dx}{x}+\gamma\int_0^1\frac{W_p(x)-V(x)}{x^\eta}\frac{dx}{x^{1-\eta}}\\&\xrightarrow{\text{a.s.}}\gamma\int_0^1 V(x)\frac{dx}{x}.\end{aligned}$$

By Corollary 2.2, $X_{p-k:p}/u_p\xrightarrow{\mathbb{P}}1$, and the first term on the right-hand side in (11) is equal to

$$\begin{aligned}&\int_{X_{p-k:p}/u_p}^{1}W_p(s^{-1/\gamma})\frac{ds}{s}+\sqrt{k}\int_{X_{p-k:p}/u_p}^{1}s^{-1/\gamma}\frac{ds}{s}\\&\overset{\text{a.s.}}{=}o(1)+\gamma\sqrt{k}\left(\left(\frac{X_{p-k:p}}{u_p}\right)^{-1/\gamma}-1\right)\\&\xrightarrow{\text{a.s.}}-\gamma V(1),\end{aligned}$$

where the first term vanishes due to Corollary 2.1 and the second term converges because of Corollary 2.2. Hence, we obtain that

$$\sqrt{k}(\widehat{\gamma}-\gamma)\xrightarrow{\text{a.s.}}\gamma\left(-V(1)+\int_0^1 V(x)\frac{dx}{x}\right),$$

which is a centered normal random variable.

To complete the proof, we will show that

$$\operatorname{var}\left(-V(1)+\int_0^1 V(x)\frac{dx}{x}\right)=\operatorname{var}(V(1)).$$

We have

$$\begin{aligned}&\operatorname{cov}\left(-V(1)+\int_0^1 V(x)\frac{dx}{x},-V(1)+\int_0^1 V(y)\frac{dy}{y}\right)\\&=\operatorname{var}(V(1))-2\int_0^1\operatorname{cov}(V(x),V(1))\frac{dx}{x}+\int_0^1\int_0^1\operatorname{cov}(V(x),V(y))\frac{dx}{x}\frac{dy}{y}\\&=\operatorname{var}(V(1))-2\int_0^1\operatorname{cov}(V(x),V(1))\frac{dx}{x}+2\int_0^1\int_0^y\operatorname{cov}(V(x),V(y))\frac{dx}{x}\frac{dy}{y}.\end{aligned}$$

Now

$$\int_0^1\operatorname{cov}(V(x),V(1))\frac{dx}{x}=\int_0^1(1-R(1,x^{-1}))\,dx,$$

and also, by the change of variables $x/y=u$,

$$\int_0^1\int_0^y\operatorname{cov}(V(x),V(y))\frac{dx}{x}\frac{dy}{y}=\int_0^1\int_0^y(1-R(1,y/x))\,dx\frac{dy}{y}$$

$$= \int_0^1 \int_0^1 (1 - R(1, u^{-1}))\, dy\, du$$

$$= \int_0^1 (1 - R(1, u^{-1}))\, du. \qquad \square$$

The proof of Theorem 2.3 is deferred to the Supplementary Material.

7.2. *Proofs from Section* 3. Since the permutation π does not change T_p and H_p, we can and will omit it in the proofs. We will only prove the results for the case $\mu = 0$. Extending the proofs to $\mu \neq 0$ is straightforward (but tedious), as the influence of a location shift is asymptotically negligible for a heavy tail. Accordingly, define the limit functions $T(x) = \int_0^\infty S(x/u)\, dF_\sigma(u)$ and $H(x, y) = \int_0^\infty S(\frac{x}{u})S(\frac{y}{u})\, dF_\sigma(u)$.

We need an elementary lemma for the tail of the product of two independent random variables; see, for example, Lemma 4.2 in Jessen and Mikosch (2006).

LEMMA 7.8. *Let X_1 and X_2 be nonnegative, independent random variables. If $\lim_{x\to\infty} x^{1/\gamma}\mathbb{P}(X_1 > x) = c \in (0, \infty)$ for some $\gamma > 0$ and $\mathbb{E}X_2^{1/\gamma} < \infty$, then*

$$\lim_{x\to\infty} x^{1/\gamma}\mathbb{P}(X_1 X_2 > x) = c \cdot \mathbb{E}X_2^{1/\gamma}.$$

From this lemma, we can deduce that, under the conditions of Theorem 3.1 or 3.2,

$$\lim_{t\to\infty} t^{1/\gamma} T(t) = \widetilde{c}, \quad \text{for some } \widetilde{c} \in (0, \infty). \tag{12}$$

LEMMA 7.9. *If T satisfies (12), then $U_\sigma(t) \le Ct^\gamma$, $t \ge 1$, for some constant C, and hence $1 - F_\sigma(t) =: T_\sigma(t) \le C^{1/\gamma} t^{-1/\gamma}$, $t > 0$. We also have $S(t) \le C^{1/\gamma} t^{-1/\gamma}$ for $t > 0$.*

LEMMA 7.10. *Suppose that $\log U_\sigma(e^t)$ is Lipschitz-continuous on $[0, \infty)$ and T satisfies (12). Assumption 2.1 holds with extreme value index γ and T as above if*

$$xg(x) \le MT(x), \quad x \ge 0,$$

for some constant $M < \infty$.

LEMMA 7.11. *Assumption 2.2 holds with the limit function T if there exist a non-increasing function h on $[0, \infty)$ and a constant M such that for all $x \ge 0$,*

$$xg(x) \le h(x), \quad \textit{and} \quad \int_0^\infty h(x/u)\, dF_\sigma(u) \le MT(x).$$

The proofs of Lemmas 7.9–7.11 are deferred to the Supplementary Material.

PROOF OF THEOREM 3.1. By Lemma 7.8, we know that (12) holds and we claim that, for some constants C, M and $x \ge 0$,

$$\int_0^\infty h(x/u)\, dF_\sigma(u) \le Ch(0)T_\sigma(x) \le MT(x), \tag{13}$$

$$xg(x) \le MT(x). \tag{14}$$

Part (i) then follows from Lemmas 7.10 and 7.11. To prove (13), we introduce a survival function $\widetilde{h}(x) = \frac{h(x)}{h(0)}$ on the positive half-line. Then we have

$$\int_0^\infty h(x/u)\, dF_\sigma(u) = h(0) \cdot \int_0^\infty \widetilde{h}(x/u)\, dF_\sigma(u) =: h(0) \cdot \widetilde{T}(x).$$

Observe that $\widetilde{T}$ is the survival function of the product of two independent random variables with distribution functions F_σ and $1-\widetilde{h}$, respectively. Applying Lemma 7.8 yields that $\lim_{x\to\infty} x^{-1/\gamma}\widetilde{T}(x)<\infty$. Therefore, for some large constants C_1 and C,

$$\widetilde{T}(x)\le \min\{C_1x^{-1/\gamma},1\}\le C_1\min\{x^{-1/\gamma},1\}\le CT_\sigma(x),$$

and hence the first inequality in (13) follows. Similar arguments yield the second inequality. For (14), note that $xg(x)\le h(x)\le C\min\{x^{-1/\gamma},1\}\le MT(x)$.

It is only left to show the existence of R and to verify its expression. First, consider any intermediate threshold sequence of the form $t=U_\sigma(z)$ with $z=z(p)\to\infty$ such that $T(U_\sigma(z))\to 0$ and $pT(U_\sigma(z))\to\infty$. By part (i) of this theorem and Lemma 7.8 with (12),

$$zT\big(U_\sigma(z)\big)=\frac{T(U_\sigma(z))}{T_\sigma(U_\sigma(z))}\quad\to\quad \mathbb{E}Z_+^{1/\gamma}=\int_0^\infty S\big(v^\gamma\big)\,dv\in(0,\infty).$$

Following the definition of R in Assumption 2.4, we need to show that

$$zH_p\big(U_\sigma(z)x^{-\gamma},U_\sigma(z)y^{-\gamma}\big)\to\int_0^\infty S\big(v^\gamma x^{-\gamma}\big)S\big(v^\gamma y^{-\gamma}\big)\,dv,\qquad x,y>0.$$

In fact, we only need to show that, as $t\to\infty$,

$$zH\big(U_\sigma(z)x^{-\gamma},U_\sigma(z)y^{-\gamma}\big)\to\int_0^\infty S\big(v^\gamma x^{-\gamma}\big)S\big(v^\gamma y^{-\gamma}\big)\,dv$$

because then, very similar to the proof of Lemma 7.10, we can show that

$$zH_p\big(U_\sigma(z)x^{-\gamma},U_\sigma(z)y^{-\gamma}\big)=zH\big(U_\sigma(z)x^{-\gamma},U_\sigma(z)y^{-\gamma}\big)+o(1).$$

Take another intermediate threshold sequence $\lambda=\lambda(z)$ such that $\lambda\to\infty$ but $z/\lambda\to\infty$ as $z\to\infty$. We have

$$\begin{aligned}
&zH\big(U_\sigma(z)x^{-\gamma},U_\sigma(z)y^{-\gamma}\big)\\
&\quad=\int_0^z S\left(\frac{U_\sigma(z)x^{-\gamma}}{U_\sigma(z/v)}\right)S\left(\frac{U_\sigma(z)y^{-\gamma}}{U_\sigma(z/v)}\right)dv\\
&\quad=\int_0^\lambda S\left(\frac{U_\sigma(z)x^{-\gamma}}{U_\sigma(z/v)}\right)S\left(\frac{U_\sigma(z)y^{-\gamma}}{U_\sigma(z/v)}\right)dv\\
&\qquad+\int_\lambda^z S\left(\frac{U_\sigma(z)x^{-\gamma}}{U_\sigma(z/v)}\right)S\left(\frac{U_\sigma(z)y^{-\gamma}}{U_\sigma(z/v)}\right)dv\\
&\quad=: J_1(x,y)+J_2(x,y).
\end{aligned}$$

We shall show that $J_1(x,y)$ converges to the required limit and $J_2(x,y)\to 0$. Let $\varepsilon>0$ be small. For all large z and z/λ,

$$\sup_{0<v\le\lambda}\left|\frac{U_\sigma(z)}{U_\sigma(z/v)v^\gamma}-1\right|=\sup_{0<v\le\lambda}\left|\frac{U_\sigma(z)z^{-\gamma}}{U_\sigma(z/v)(z/v)^{-\gamma}}-1\right|<\varepsilon.$$

By monotonicity of S,

$$\begin{aligned}
J_1(x,y)&\le\int_0^\infty S\big((1-\varepsilon)v^\gamma x^{-\gamma}\big)S\big((1-\varepsilon)v^\gamma y^{-\gamma}\big)\,dv\\
&=(1-\varepsilon)^{-1/\gamma}\int_0^\infty S\big(v^\gamma x^{-\gamma}\big)S\big(v^\gamma y^{-\gamma}\big)\,dv,
\end{aligned}$$

and, on the other hand,

$$
\begin{aligned}
J_1(x,y) &\ge \int_0^{\lambda} S\big((1+\varepsilon)v^{\gamma}x^{-\gamma}\big)S\big((1+\varepsilon)v^{\gamma}y^{-\gamma}\big)\,dv \\
&\to \int_0^{\infty} S\big((1+\varepsilon)v^{\gamma}x^{-\gamma}\big)S\big((1+\varepsilon)v^{\gamma}y^{-\gamma}\big)\,dv \\
&= (1+\varepsilon)^{-1/\gamma}\int_0^{\infty} S\big(v^{\gamma}x^{-\gamma}\big)S\big(v^{\gamma}y^{-\gamma}\big)\,dv.
\end{aligned}
$$

It follows that $J_1(x,y)\to\int_0^{\infty}S(v^{\gamma}x^{-\gamma})S(v^{\gamma}y^{-\gamma})\,dv$ as ε can be arbitrarily small.

Next, we show that $J_2(x,y)\to 0$. Recall that $U_\sigma(z)/z^{\gamma}$ is bounded away from 0 and ∞ for $z\ge 1$. Hence, for some large constant C,

$$
\begin{aligned}
J_2(x,y) &\le \int_{\lambda}^{\infty} S\big(Cv^{\gamma}x^{-\gamma}\big)S\big(Cv^{\gamma}y^{-\gamma}\big)\,dv \\
&= C^{-1/\gamma}\int_{C^{1/\gamma}\lambda}^{\infty} S\big(v^{\gamma}x^{-\gamma}\big)S\big(v^{\gamma}y^{-\gamma}\big)\,dv \to 0,
\end{aligned}
$$

since $\int_0^{\infty}S(v^{\gamma}x^{-\gamma})S(v^{\gamma}y^{-\gamma})\,dv\le\int_0^{\infty}S(v^{\gamma}x^{-\gamma})\,dv = x\int_0^{\infty}S(v^{\gamma})\,dv<\infty$.

Now, for any intermediate threshold sequence $t=t(p)$, using the power-law approximation of U_σ, we can find two intermediate threshold sequences $U_\sigma(z_\pm)$ with $z_\pm=z_\pm(p)\to\infty$ such that $U_\sigma(z_-)\le t\le U_\sigma(z_+)$ and $U_\sigma(z_+)/U_\sigma(z_-)\to 1$. Then by monotonicity of H_p and T and the squeeze theorem it readily follows that $H_p(tx^{-\gamma},ty^{-\gamma})/T(t)\to R(x,y)$. □

PROOF OF THEOREM 3.2. Part (i) follows from Lemmas 7.10 and 7.11, since (12) holds due to Lemma 7.8 again and the condition in Lemma 7.11 is trivial with $h(x)=MS(x)$ therein. The proof is very similar to that for Theorem 3.1 and we omit the details.

It remains to verify that R is trivial. Take any $x,y>0$. Let $t=t(p)\to\infty$ be an arbitrary intermediate threshold sequence such that $pT(t)\to\infty$. By monotonicity of S and Q_σ,

$$
\begin{aligned}
H_p\big(tx^{-\gamma},ty^{-\gamma}\big) &= \frac{1}{p}\sum_{i=1}^{p} S\Big(\frac{tx^{-\gamma}}{Q_\sigma(1-i/p)}\Big)S\Big(\frac{ty^{-\gamma}}{Q_\sigma(1-i/p)}\Big) \\
&\le \int_0^1 S\Big(\frac{tx^{-\gamma}}{Q_\sigma(1-u)}\Big)S\Big(\frac{ty^{-\gamma}}{Q_\sigma(1-u)}\Big)\,du \\
&= H\big(tx^{-\gamma},ty^{-\gamma}\big).
\end{aligned}
$$

It suffices to show that $H(tx^{-\gamma},ty^{-\gamma})/T(t)\to 0$ as $t\to\infty$.

Take another threshold sequence $\lambda=\lambda_p\to\infty$ but $t/\lambda\to\infty$. Using monotonicity of S and Lemma 7.9, for some constant C,

$$
\begin{aligned}
\frac{H(tx^{-\gamma},ty^{-\gamma})}{T(t)} &= \frac{\int_0^{\lambda}S(\frac{tx^{-\gamma}}{u})S(\frac{ty^{-\gamma}}{u})\,dF_\sigma(u)}{T(t)} + \frac{\int_{\lambda}^{\infty}S(\frac{tx^{-\gamma}}{u})S(\frac{ty^{-\gamma}}{u})\,dF_\sigma(u)}{T(t)} \\
&\le \frac{S(\frac{tx^{-\gamma}}{\lambda})\int_0^{\lambda}S(\frac{ty^{-\gamma}}{u})\,dF_\sigma(u)}{T(t)} + \frac{\int_{\lambda}^{\infty}(Ct^{-1/\gamma}xu^{1/\gamma}\cdot 1)\,dF_\sigma(u)}{T(t)} \\
&\le S\Big(\frac{tx^{-\gamma}}{\lambda}\Big)\frac{T(ty^{-\gamma})}{T(t)} + x\cdot\frac{Ct^{-1/\gamma}}{T(t)}\cdot\int_{\lambda}^{\infty}u^{1/\gamma}\,dF_\sigma(u) \\
&\to 0\cdot y + x\cdot\frac{C}{\tilde{c}}\cdot 0 = 0,
\end{aligned}
$$

where we also recall (12) for the last line. □

Acknowledgment. We are grateful to the Editor, the Associate Editor and two referees for various thoughtful comments that greatly helped improving the article.

SUPPLEMENTARY MATERIAL

Supplementary material for "Extreme value inference for heterogeneous power law data" (DOI: 10.1214/23-AOS2294SUPP; .pdf). The supplementary material consist of two sections. Section 1 provides more simulation results for the extreme quantile estimator. Section 2 provides proofs of Theorems 2.3 and 4.1 and Lemmas 7.9–7.11.

REFERENCES

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

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. Springer, London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds (with discussion). *J. Roy. Statist. Soc. Ser. B* **52** 393–442. MR1086795

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

DE HAAN, L. and ZHOU, C. (2021). Trends in extreme value indices. *J. Amer. Statist. Assoc.* **116** 1265–1279. MR4309271 https://doi.org/10.1080/01621459.2019.1705307

DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855. MR1026315 https://doi.org/10.1214/aos/1176347397

DREES, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. MR1810875 https://doi.org/10.1214/aoap/1019487617

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

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

EINMAHL, J. H. J. and HE, Y. (2023). Extreme value estimation for heterogeneous data. *J. Bus. Econom. Statist.* **41** 255–269. MR4522166 https://doi.org/10.1080/07350015.2021.2008408

EINMAHL, J. H and HE, Y. (2023). Supplement to "Extreme value inference for heterogeneous power law data." https://doi.org/10.1214/23-AOS2294SUPP

EINMAHL, J. H. J. and SEGERS, J. (2021). Empirical tail copulas for functional data. *Ann. Statist.* **49** 2672–2696. MR4338379 https://doi.org/10.1214/21-aos2050

GABAIX, X., GOPIKRISHNAN, P., PLEROU, V. and STANLEY, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature* **423** 267–270. https://doi.org/10.1038/nature01624

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

HSING, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569. MR1126337 https://doi.org/10.1214/aos/1176348261

JESSEN, A. H. and MIKOSCH, T. (2006). Regularly varying functions. *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **80** 171–192. MR2281913 https://doi.org/10.2298/PIM0694171J

JI, P., JIN, J., KE, Z. T. and LI, W. (2022). Co-citation and co-authorship networks of statisticians. *J. Bus. Econom. Statist.* **40** 469–485. MR4410871 https://doi.org/10.1080/07350015.2021.1978469

KULIK, R. and SOULIER, P. (2020). *Heavy-Tailed Time Series*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR4174389 https://doi.org/10.1007/978-1-0716-0737-4

POTTER, H. SA. (1942). The mean values of certain Dirichlet series, II. *Proc. Lond. Math. Soc.* **2** 1–19.

SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174–1207. MR0902252 https://doi.org/10.1214/aos/1176350499

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

VERVAAT, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. Verw. Gebiete* **23** 245–253. MR0321164 https://doi.org/10.1007/BF00532510