EXTREME VALUE INFEERENCE FOR GENERAL HETEROGENEOUS DATA

By

John H. J. Einmahl, Yi He

4 August 2022

ISSN 0924-7815
ISSN 2213-9532
EXTREME VALUE INFERENCE FOR GENERAL HETEROGENEOUS DATA

BY JOHN H.J. EINMAHL\textsuperscript{1,}\textsuperscript{a}, YI HE\textsuperscript{2,}\textsuperscript{b}

\textsuperscript{1}Department of Econometrics & OR and CentER, Tilburg University, Tilburg, The Netherlands, \textsuperscript{a}j.h.j.einmahl@uvt.nl
\textsuperscript{2}Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands, \textsuperscript{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. 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.

\textit{JEL Codes:} C13, C14.

1. Introduction. Consider independent and identically distributed random variables $X_1, \ldots, X_p$, $p \in \mathbb{N}$, from some common distribution function $F$. For this case, the statistical theory of extreme values has been developed comprehensively in the literature, e.g., the monographs Beirlant et al. (2004) and de Haan and Ferreira (2006). In this setting, well-known estimators of the extreme value index $\gamma$ have been introduced in, among others, Hill (1975), Smith (1987), and Dekkers et al. (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 non-identically distributed. Statistics of extremes for identically distributed but (weakly, serially) dependent data has been studied extensively in the literature, see, e.g., Hsing (1991), Drees (2000), Drees and Rootzén (2010), and the monograph Kulik and Soulier (2020).

In this paper, we focus on independent, but non-identically distributed data. Although earlier work on statistics of extremes has been published for this case in more restricted settings, in a general, non-parametric setting the most relevant (and recent) references are Einmahl et al. (2016) and de Haan and Zhou (2021), but see also both papers for more references in 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 $\gamma$, whereas in the second paper a gradually changing $\gamma$ is allowed. Just like in these two papers we consider the case where $\gamma$ 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 et al. (2016). Like in de Haan and Zhou (2021), we allow different extreme value indices for different observations, but we do not require a smooth change of the distribution in $i$ ($i = 1, \ldots, p$): neighboring observations ($X_i$ and $X_{i+1}$) may have (very) different distributions.

Settings where this can be relevant are when computing the Hill estimator for natural hazards across different locations, for a cross-section (on the same day) of daily loss returns of

\textit{MSC2020 subject classifications:} Primary 62G32, 62G20, 62G30; secondary 60G70, 60F17.
\textit{Keywords and phrases:} Extreme value statistics, functional central limit theorem, heterogeneous scales model, Hill estimator, non-identical distributions, weighted tail empirical process.
many stocks, for population sizes of cities in a large country, or for numbers of citations of scientific articles. As a consequence of our setup we are not interested in the average (or local) $\gamma$ but in the $\gamma$ 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 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 $\gamma^2$ for the asymptotic variance. In case $R(1,1) > 0$, the asymptotic variance is smaller and can be expressed in $\gamma$ and the distribution of the latent variables.

The remainder of this paper is organized as follows. In Section 2 we present our general, main results. Section 3 contains the specialization to the heterogeneous scales model. In Section 4 we present a simulation study and in Section 5 we apply the theory to earthquake energies and cross-sectional stock market losses. The proofs of the results in Sections 2 and 3 are deferred to Section 6.

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

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

and their average distribution function by

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

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 non-increasing 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.

This 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 \( T_{\text{emp}} := 1 - F_{\text{emp}} \approx 1 - F_p = T_p \) for large values and large \( p \), then the data exhibit power law behavior in the sense that, for large \( t \) and \( p \),
\[
T_{\text{emp}}(tx) \approx x^{-1/\gamma}T_{\text{emp}}(t).
\]

As shown in Einmahl and He (2022), 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 \( i = 1, \ldots, p \). The heterogeneity thus can go far beyond heteroscedastic extremes as in Einmahl et al. (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 \( p \max\{xf_p(x), T_p(x)\} \leq MT(x) \).

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 \( k = k(p) \in \{1, \ldots, p - 1\} \) be a sequence satisfying:

**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
\[
V_p(x) = \frac{p}{\sqrt{k}} \left( T_{\text{emp}} \left( \frac{u_p}{x^\gamma} \right) - T_p \left( \frac{u_p}{x^\gamma} \right) \right), \quad x \geq 0,
\]
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 \( V_p \):
\[
\text{cov}(V_p(x), V_p(y)) = \frac{1}{k} \sum_{i=1}^{p} \text{cov} \left( \mathbb{1}[X_i^{(p)} > x^{-\gamma}u_p], \mathbb{1}[X_i^{(p)} > y^{-\gamma}u_p] \right)
\[
= \frac{p}{k} T_p \left( u_p (\min\{x, y\})^{-\gamma} \right) - \frac{p}{k} H_p(u_p x^{-\gamma}, u_p y^{-\gamma}),
\]
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

\[
\frac{1}{p \alpha} \sum_{i=1}^{p} \mathbb{P} \left( T_p(X_i^{(p)}) < \alpha x \right) \mathbb{P} \left( T_p(X_i^{(p)}) < \alpha y \right) \to R(x, y),
\]

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

\[
R(x, y) = \lim_{p \to \infty} \frac{1}{p \alpha} \sum_{i=1}^{p} \mathbb{P} \left( T_p(X_i^{(p)}) < \alpha x, T_p(\tilde{X}_i^{(p)}) < \alpha y \right)
\]

where \( \tilde{X}_i^{(p)} \) are independent copies of \( X_i^{(p)} \) and \( \alpha \) 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

\[
R(x, y) = \lim_{p \to \infty} \frac{1}{\alpha} \mathbb{P} \left( U < \alpha x \right) \cdot \lim_{p \to \infty} \mathbb{P} \left( U < \alpha y \right) = x \cdot 0 = 0, U \sim \text{Un}(0, 1),
\]

and therefore the limit \( R \equiv 0 \) is called trivial.

The most interesting results in this paper are for non-trivial 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 non-trivial,
- \( 0 \leq R(x, y) \leq \min\{x, y\} \),
- \( R(ax, ay) = aR(x, y) \) 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 \( 0 \leq \eta < \frac{1}{2} \),

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

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

\[
cov(V(x), V(y)) = \min\{x, y\} - R(x, y).
\]

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 $X_{p-k:p} \leq \ldots \leq X_{p:p}$. Then we estimate the extreme value index $\gamma > 0$ by the Hill estimator:

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

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

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

Using the Skorohod representation theorem we obtain:

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

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

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

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

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

\textbf{COROLLARY 2.2.} \textit{Under the conditions and on the probability space of Corollary 2.1,}

$$\sqrt{k} \left( \frac{X_{p-k:p}}{u_p} \right)^{-1/\gamma} \left( -V(1) + \int_0^1 V(x) \frac{dx}{x} \right) \xrightarrow{a.s.} \gamma V(1) \sim N(0, 1 - R(1, 1)).$$

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.

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

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

The most striking of this result is the smaller (compared with the homogeneous case) limiting variance when $R$ is non-trivial, 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. 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 $0 \leq \eta < \frac{1}{2}$, is implied by $\sqrt{k}A(p/k) \to 0$, where $A$ is the usual auxiliary function in the second-order condition, see, e.g., Theorem 3.2.5 in de Haan and Ferreira (2006).

**Example.** 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 $Z_i, i = 1, \ldots, p$, 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” $\gamma$ defined in Assumption 2.1 is equal to $\lim_{p \to \infty} \max_{i=1, \ldots, p} 1/(1 + i/p) = 1$ 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.** 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

$$Y_i^{(p)} = Z X_i^{(p)}, \quad i = 1, \ldots, p,$$

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, e.g., construct the usual asymptotic confidence intervals for $\gamma$. 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 an auxiliary sample of $\tilde{Y}_i^{(p)}, i = 1, \ldots, p$, is available, such that the pairs $(X_i^{(p)}, \tilde{Y}_i^{(p)}), i = 1, \ldots, p$, are independent. Denote the average survival function of the $\tilde{Y}_i^{(p)}, i = 1, \ldots, p$, with $\tilde{T}_p$ and assume it is continuous. We require the following “exceedance similarity” condition.

**ASSUMPTION 2.5 (Exceedance Similarity).** Let $U_i^{(p)} = T_p(X_i^{(p)})$ and $\tilde{V}_i^{(p)} = \tilde{T}_p(\tilde{Y}_i^{(p)})$.

$$\max_{1 \leq i \leq p} \left| \mathbb{P}(\tilde{V}_i^{(p)} < \alpha | U_i^{(p)} < \alpha x) - \mathbb{P}(U_i^{(p)} < \alpha) \right| \to 0$$

for all intermediate sequences $\alpha = \alpha(p) \downarrow 0$ such that $p\alpha \to \infty$ and for $x$ in a neighborhood of 1.

When $\tilde{V}_i^{(p)}$ is independent of $U_i^{(p)}$, the condition is equivalent to

$$\max_{1 \leq i \leq p} \left| \mathbb{P}(\tilde{V}_i^{(p)} < \alpha) - \mathbb{P}(U_i^{(p)} < \alpha) \right| \to 0.$$

Note that this condition is trivially fulfilled for homogeneous data such that $X_i^{(p)} \sim X$ and $\tilde{Y}_i^{(p)} \sim \tilde{Y}$, as $\max_{1 \leq i \leq p} \mathbb{P}(\tilde{V}_i^{(p)} < \alpha) = \max_{1 \leq i \leq p} \mathbb{P}(U_i^{(p)} < \alpha) = \alpha \to 0$. For heterogeneous data, however, $\max_{1 \leq i \leq p} \mathbb{P}(U_i^{(p)} < \alpha)$ does not vanish in general. An ideal situation will be a duplicate sample $\tilde{Y}_i^{(p)} \sim X_i^{(p)}$, where all variables are mutually independent, but this is really not necessary. The condition only requires that the exceedance behavior of the $\tilde{Y}_i^{(p)}$ is
asymptotically independent of that of the $X_i^{(p)}$ and that they share similar rank distributions in the tail.

Let $R_i^X = \sum_{j=1}^p 1[X_i^{(p)} \leq X_j^{(p)}]$ and $R_i^Y = \sum_{j=1}^p 1[\tilde{Y}_i^{(p)} \leq \tilde{Y}_j^{(p)}]$ denote the ranks of $X_i^{(p)}$ and $\tilde{Y}_i^{(p)}$ in their own array. Then we may estimate $R(1,1)$ consistently through the empirical tail dependence coefficient given by

$$\hat{R}(1,1) = \frac{1}{k} \sum_{i=1}^p 1[R_i^X > p - k, R_i^Y > p - k].$$

**Theorem 2.3.** Under Assumptions 2.1–2.5, we have

$$\hat{R}(1,1) \xrightarrow{P} R(1,1).$$

**3. Leading Example: Heterogeneous Scales Model.** In this section, we study a particular type of heterogeneous model related to those in Einmahl et al. (2016) and Einmahl and He (2022). 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 $Z_1, \ldots, Z_p$ be i.i.d. latent continuous random variables. Consider the following independent, but non-identically distributed data

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

where $\mu \in \mathbb{R}$, $Q_\sigma$ is the generalized quantile function of a continuous df $F_\sigma$, with positive left endpoint, and where $\pi$ is an unknown permutation of $1, \ldots, p$. Clearly the $Q_\sigma(1 - \pi(i)/p)$ are scale parameters. (Observe that the permutation $\pi$ does not affect the distribution of the order statistics of the $X_i^{(p)}$ and hence also not that of $V_p$ and $\tilde{\gamma}$, but on the other hand it allows non-smooth 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 (Non-trivial Limit).** Suppose that $U_\sigma$ 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 $\gamma$. If there exists a non-increasing, right-continuous function $h$ on $[0, \infty)$ such that $xg(x) \leq h(x)$ for all $x \geq 0$, such that $0 < \mathbb{E}Z_+^{1/\gamma} \leq \int_0^\infty h(x^\gamma)dx < \infty$, then:

(i) Assumptions 2.1 and 2.2 hold with extreme value index $\gamma$ and

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

(ii) Assumption 2.4 holds with a non-trivial $R$-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 $\lim_{t \to \infty} 1/t^{1/\gamma}S(t) \in (0, \infty)$ exists for some positive extreme value index $\gamma$, and $xg(x) \leq MS(x)$, $x \geq 0$, 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 \( \hat{\gamma} \) 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 \( \gamma^2 \) is the same as in the i.i.d. case, whereas in case of Theorem 3.1 the limiting variance is smaller than \( \gamma^2 \). Observe that in the latter case, although the individual \( Z_i \) can be light-tailed, the large heterogeneity generates a positive \( \gamma \), a heavy tail.

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

The setup of Theorem 3.2 can be compared to that in Einmahl et al. (2016). If the scedasis function \( c \) therein is non-increasing it is equal to \( Q_{\sigma} (1 - \cdot) \) 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 et al. (2016), but note that the condition therein that \( c \), and hence \( Q_{\sigma} \), is bounded is not required here. Indeed, it is natural to allow that the quantile function \( Q_{\sigma}(s) \to \infty \) as \( s \uparrow 1 \).

**Examples.** 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 \( \gamma \) 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 \( \gamma \). Figure 1 depicts the values of \( R(1, 1) \) for various distributions of the \( Z_i \), as a function of the extreme value index \( \gamma \).

**Fig 1.** \( R(1, 1) \) for various distributions of the \( Z_i \)
4. Simulations. We consider three sets of Monte Carlo simulations to illustrate how the asymptotic behavior of the Hill estimator changes for heterogeneous data. First, we fix the extreme value index $\gamma$ but vary the distribution of $Z$ in the heterogeneous scales model. Second, we specify the distribution of $Z$ but change the extreme value index $\gamma$. 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 $k = 50$.

4.1. Heterogeneous scales model with fixed $\gamma$. We generate independent random variables from the heterogeneous scales model

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

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

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

(I) $\text{Un}(1, 1 + 1/\theta)$, for $\theta > 0$, with $R(1, 1) = \frac{6\theta + 2}{6\theta + 3}$;

(II) $\text{Pareto}(1 + \epsilon)$, for $\epsilon > 0$, with $R(1, 1) = \frac{2\epsilon}{2\epsilon + 1}$;

(III) $\text{Weibull}(\tau)$, for $\tau > 0$, with $R(1, 1) = 2^{1/\tau}$.

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 2 shows the parameters as a function of $R(1, 1)$.

![Graph showing parameters as a function of R(1,1)](image)

Figure 3 compares the variance of $\sqrt{k}(\hat{\gamma} - \gamma) = \sqrt{k}(\tilde{\gamma} - 1)$ 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 $\sqrt{k}(\hat{\gamma} - \gamma)$ 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 $\theta$, $\epsilon$ or $\tau$ 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 $\infty$, the $Z_i$ and hence the Hill estimator are becoming deterministic.

Figure 4 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.
4.2. Heterogeneous scales model with different $\gamma$. We again simulate heterogeneous data $X_i^{(p)}$ according to the heterogeneous scales model (8). We change the value of the extreme value index $\gamma$, and generate i.i.d. $Z_i$ from:

(I) the standard uniform distribution with $R(1, 1) = \frac{2\gamma}{2\gamma + 1}$;

(II) the standard normal distribution with $R(1, 1) = \int_0^\infty \Phi^2(v^\gamma)dv / \int_0^\infty \Phi(v^\gamma)dv$;

Fig 3. Variance of $\sqrt{K(\hat{\gamma} - \gamma)}$ over simulation replications versus its theoretical limit as a function of $R(1, 1)$

Fig 4. Boxplots of Estimation Error $\hat{\gamma} - 1$
(III) the standard exponential distribution with $R(1, 1) = 2^{-1/\gamma}$.

Figure 5 shows the variance of $\sqrt{k(\hat{\gamma} - \gamma)}$ over simulation replications and the asymptotic variance. We also plot the asymptotic variance $\gamma^2$ 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. According to our theory, the heterogeneous data lead to a smaller variance than $\gamma^2$, and the difference $\gamma^2 - \gamma^2(1 - R(1, 1)) = \gamma^2 R(1, 1)$ grows with $\gamma$ (as well as with $R(1, 1)$). The variance of the relative estimation error $\hat{\gamma}/\gamma - 1$, however, is decreasing in $R(1, 1)$. Figure 6 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 $R(1, 1)$.

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

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

(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 $Q_p(1 - i/p) = \log(p/i)$ can be seen as a scedasis $c(i/p)$, that violates the boundedness condition on $c$ in Einmahl et al. (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 the first example below Theorem 2.2 for more details about Case III. For all cases, the extreme value indices of the average distribution function $F_p$ are equal to $\gamma = 1$.

In Figure 7 Probability-Probability (PP) plots are shown of 5000 simulation replications of $\sqrt{k(\hat{\gamma} - 1)}$ 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 PP-plots are very close to the diagonal, showing again that the asymptotic theory for the Hill estimator works well for finite samples.
5. **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. 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 8 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 $-1/\hat{\gamma}$, where the Hill estimate $\hat{\gamma} = 0.9763$ indicates a very heavy tail.

Next, we split the sample into halves and estimate $R(1, 1)$ by $\hat{R}(1, 1)$ in (6) of the two folds. We pair the nearest observations in the spatial distance but with a minimum time gap of 1 year, randomly assigning one to fold 1 and the other one to fold 2, as illustrated on the right in Figure 8. The rather large estimate of $\hat{R}(1, 1) = 0.41$ suggests a substantial spatial heterogeneity of energies and a non-trivial limit of the Hill estimator. Using Theorem 2.2, we then obtain a relatively narrow 95% asymptotic confidence interval $(0.8293, 1.1233)$ for $\gamma$, indicated by dashed lines in the log-log plot in Figure 8 (keeping the $y$-intercept unchanged).

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 9 shows the Hill estimates and confidence intervals of the extreme value index 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 $R(1, 1)$: one uses the daily losses for the same firms 25 working days before the estimation date as auxiliary sample and then formula (6), and the other one uses the heterogeneous scales model with $Z$ a Student-$t_3$ variable (satisfying the cubic power law). Observe from Figure 1 that the latter estimator $R_{\hat{\gamma}}(1, 1)$ takes values in $[0, 0.2)$ for $\hat{\gamma} < 0.8$. 
6. Proofs.

6.1. 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

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

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

$$\frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}(X^{(p)}_i > tx^{-\gamma}) \mathbb{P}(X^{(p)}_i > ty^{-\gamma})$$

$$\leq \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}(T_p(X^{(p)}_i) \leq T_p(tx^{-\gamma})) \mathbb{P}(T_p(X^{(p)}_i) \leq T_p(ty^{-\gamma}))$$

$$\leq \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}(T_p(X^{(p)}_i) \leq (1 + \varepsilon)\alpha x) \mathbb{P}(T_p(X^{(p)}_i) \leq (1 + \varepsilon)\alpha y),$$

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

$$\liminf_{p \to \infty} \frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P}(T_p(X^{(p)}_i) \leq \alpha x) \mathbb{P}(T_p(X^{(p)}_i) \leq \alpha y) \geq \frac{1}{1 + 2\varepsilon} R(x, y).$$

Similarly, we obtain that

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

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 6.1.** 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 $t \geq t_0$,

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

**Proof.** The case $x \geq 1$ (or more generally $x \geq c$ 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 6.2.** Under Assumptions 2.1 and 2.2, there exists $p_0 > 0$ such that $\lim_{z \to \infty} z^r T_p(z) = 0$ for any $0 \leq r < 1/\gamma$ and all $p \geq p_0$.

**Proof.** By Assumption 2.2, it suffices to prove that $\lim_{z \to \infty} z^r T(z) = 0$. Let $0 \leq r < 1/\gamma$, and take a $\delta \in (0, 1/\gamma - r)$. By Lemma 6.1, there exists some $t_0 = t_0(\delta)$, such that for all $z \geq t_0$,

$$z^r T(z) \leq C z^r T(t_0) \left( \frac{z}{t_0} \right)^{-1/\gamma + \delta} = CT(t_0) t_0^{1/\gamma - \delta} z^{r - 1/\gamma + \delta} \to 0, \quad z \to \infty,$$

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

**Lemma 6.3.** Let $S$ be any survival function with probability density function $g = -S'$, such that $tg(t) \leq h(t)$, $t > t_0 > 0$, for some non-increasing function $h$ on $[t_0, \infty)$. Then

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

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 $\phi'(z) = e^z g(e^z) \leq h(e^z)$ for $z > \log t_0$. By the mean-value theorem and the monotonicity of $h$, for $t > t_0$ and $x > 1$

$$S(t) - S(tx) = \phi(\log t + \log x) - \phi(\log t) \leq \sup_{0 \leq \delta \leq \log x} \phi'(\log t + \delta) \log x$$

$$\leq \sup_{0 \leq \delta \leq \log x} h(\exp(\log t + \delta)) \log x = h(t) \log x.$$

Observe that for the second part we have $tg(t) \leq h(t)$ 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 6.4 (Fidis convergence).** For any $0 \leq \eta < \frac{1}{2}$,

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

provided the asymptotic tightness of $V_p/I^n$. 
PROOF. It suffices to prove the finite-dimensional (fidis) convergence: for any given \( m \) and all fixed \( 0 < x_1 < \ldots < x_m \leq 2 \) and \( \eta \in (0, 1/2) \),

\[
(V_p(x_1)/x_1^n, \ldots, V_p(x_m)/x_m^n) \distr (V(x_1)/x_1^n, \ldots, V(x_m)/x_m^n).
\]

We use the Cramér-Wold device in conjunction with the Lindeberg central limit theorem; we can ignore the weights \( x_j^n \) 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 \( V_p/I^n \). We shall prove this for \( 0 < \eta < 1/2 \), 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 \( Y_{p,i}, i = 1, \ldots, p \), and define

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

which is bounded by

\[
\|Y_{p,i}\|_{\mathcal{F}} := \sup_{0 \leq x \leq 2} \frac{1}{\sqrt{k}} \mathbb{I} \left[ X_i^{(p)} > u_p x^{-\gamma} \right] / x^n \leq \frac{1}{\sqrt{k}} \left( \frac{X_i^{(p)}}{u_p} \right)^{\eta/\gamma}.
\]

**Lemma 6.5.** For any \( \eta \in (0, 1/2) \) and \( \lambda > 0 \),

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

**Proof.** It follows from (9) that

\[
\sum_{i=1}^{p} \mathbb{E} \left( \|Y_{p,i}\|_{\mathcal{F}} \mathbb{I} \left[ \|Y_{p,i}\|_{\mathcal{F}} > \lambda \right] \right) \leq \sum_{i=1}^{p} \int_{u_p(\lambda \sqrt{k})^{\gamma/\eta}}^{\infty} \frac{1}{\sqrt{k}} \left( \frac{x}{u_p} \right)^{\eta/\gamma} dF_p(x)
\]

\[
= p \int_{u_p(\lambda \sqrt{k})^{\gamma/\eta}}^{\infty} \frac{1}{\sqrt{k}} \left( \frac{x}{u_p} \right)^{\eta/\gamma} dF_p(x) = \lambda p \int_{1}^{\infty} z^{\eta/\gamma} dF_p(u_p(\lambda \sqrt{k})^{\gamma/\eta} z)
\]

\[
= \frac{p}{k} T_p(u_p) \cdot \lambda k \left[ - \int_{1}^{\infty} z^{\eta/\gamma} d \frac{T_p(u_p(\lambda \sqrt{k})^{\gamma/\eta} z)}{T_p(u_p)} \right].
\]

Integration by parts using Lemma 6.2 and then applying Assumptions 2.1 and 2.2, there exists \( M > 0 \) such that for all large \( p \),

\[
- \int_{1}^{\infty} z^{\eta/\gamma} d \frac{T_p(u_p(\lambda \sqrt{k})^{\gamma/\eta} z)}{T_p(u_p)} = \frac{T_p(u_p(\lambda \sqrt{k})^{\gamma/\eta})}{T_p(u_p)} + \int_{1}^{\infty} \frac{T_p(u_p(\lambda \sqrt{k})^{\gamma/\eta} z)}{T_p(u_p)} dz^{\eta/\gamma}
\]

\[
\leq 2M \left( \frac{T(u_p(\lambda \sqrt{k})^{\gamma/\eta})}{T(u_p)} + \int_{1}^{\infty} \frac{T(u_p(\lambda \sqrt{k})^{\gamma/\eta} z)}{T(u_p)} dz^{\eta/\gamma} \right).
\]
Now applying Lemma 6.1, for any sufficiently small $\delta > 0$, and for all $p \geq p_0$ (because $u_p \to \infty$) with $p_0$ depending on $\delta$,

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

$$
\int_1^\infty \frac{T(u_p(\lambda \sqrt{k})^{\gamma/\eta}z)}{T(u_p)} dz^{\eta/\gamma} \leq C \int_1^\infty \left\{ (\lambda \sqrt{k})^{\gamma/\eta}z \right\}^{-1/\gamma + \delta} dz^{\eta/\gamma}.
$$

Combining these bounds we obtain

$$
\frac{T(u_p(\lambda \sqrt{k})^{\gamma/\eta})}{T(u_p)} + \int_1^\infty \frac{T(u_p(\lambda \sqrt{k})^{\gamma/\eta}z)}{T(u_p)} dz^{\eta/\gamma}
$$

$$
\leq C \left\{ (\lambda \sqrt{k})^{\gamma/\eta} \right\}^{-1/\gamma + \delta} \left\{ 1 + \int_1^\infty z^{-1/\gamma + \delta} dz^{\eta/\gamma} \right\} = C \left\{ (\lambda \sqrt{k})^{\gamma/\eta} \right\}^{-1/\gamma + \delta} \frac{1 - \delta \gamma}{1 - \eta - \delta \gamma}.
$$

Hence, for sufficiently small $\delta > 0$

$$
\sum_{i=1}^p E \left( \| Y_{p,i} \|_F \mathbb{I} \left[ \| Y_{p,i} \|_F > \lambda \right] \right) \leq 2MC \cdot \lambda k \cdot \left\{ (\lambda \sqrt{k})^{\gamma/\eta} \right\}^{-1/\gamma + \delta} \frac{1 - \delta \gamma}{1 - \eta - \delta \gamma}
$$

$$
= 2MC \lambda^{1-1/\eta+\delta \gamma/\eta} k^{1-1/(2\eta)+\delta \gamma/(2\eta)} \frac{1 - \delta \gamma}{1 - \eta - \delta \gamma} \to 0,
$$

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

$$
\sum_{i=1}^p E \sup_{x,y \in F_a} (Y_{p,i}(x) - Y_{p,i}(y))^2 \leq \varepsilon^2.
$$

**Proof.** We have

$$
\sum_{i=1}^p E \sup_{x,y \in F_a} (Y_{p,i}(x) - Y_{p,i}(y))^2 \leq 4 \sum_{i=1}^p E \sup_{x \in F_a} Y_{p,i}(x)^2
$$

$$
= 4 \sum_{i=1}^p E \sup_{x \in F_a} \mathbb{I} \left[ X_i^p > u_p x^{-\gamma} \right] / x^{2\eta} = 4 \sum_{i=1}^p E \left( \frac{X_i^{(p)}}{u_p} \right)^{2\eta/\gamma} \mathbb{I} \left[ X_i^{(p)} > u_p (2a)^{-\gamma} \right]
$$

$$
= 4p \sum_{i=1}^p \int_{u_p(2a)^{-\gamma}}^\infty \left( \frac{x}{u_p} \right)^{2\eta/\gamma} dF_p(x) = 4 \int_{(2a)^{-\gamma}}^\infty z^{2\eta/\gamma} dF_p(z) / T_p(u_p).
$$

Like in the proof of Lemma 6.5, using integration by parts and Lemma 6.1 we obtain that for some $M$ not depending on $\varepsilon$ and small $\delta > 0$ and $p \geq p_0$ with $p_0$ only depending on $\delta$ but not $\varepsilon$,

$$
\sum_{i=1}^p E \sup_{x,y \in F_a} (Y_{p,i}(x) - Y_{p,i}(y))^2
$$

$$
\leq 8MC \frac{1 - \delta \gamma}{1 - 2\eta - \delta \gamma} \cdot (2a)^{1-2\eta-\delta \gamma} = 8MC \frac{1 - \delta \gamma}{1 - 2\eta - \delta \gamma} \cdot 2^{1-2\eta-\delta \gamma} e^{6-\frac{3}{1/2-\eta}} \delta \gamma \leq \varepsilon^2,
$$

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 6.7.** 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 \( \varepsilon \) such that for every \( p \geq p_0 \) and every \( \ell \)

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

**Proof.** For all \( i \in \{1, \ldots, p\} \),

\[
\mathbb{E} \sup_{x, y \in \mathcal{F}_\ell} (Y_{p,i}(x) - Y_{p,i}(y))^2 \leq \mathbb{E} \left( \sup_{x \in \mathcal{F}_\ell} Y_{p,i}(x) - \inf_{x \in \mathcal{F}_\ell} Y_{p,i}(x) \right)^2
\]

\[
\leq \mathbb{E} \left( \frac{1}{\sqrt{k}} \mathbb{I} \left[ X_i^{(p)} > \frac{u_p}{(2\theta^\ell)^\gamma} \right] / (2\theta^\ell + 1)^\eta - \frac{1}{\sqrt{k}} \mathbb{I} \left[ X_i^{(p)} > \frac{u_p}{(2\theta^\ell + 1)^\gamma} \right] (2\theta^\ell)^\eta \right)^2
\]

\[
= \frac{1}{k^{4\eta}} \mathbb{E} \left( \mathbb{I} \left[ X_i^{(p)} > \frac{u_p}{(2\theta^\ell)^\gamma} \right] \left( \frac{1}{2\theta^\ell + 1} - \frac{1}{2\theta^\ell} \right) + \mathbb{I} \left[ \frac{u_p}{(2\theta^\ell)^\gamma} < X_i^{(p)} \leq \frac{u_p}{(2\theta^\ell + 1)^\gamma} \right] \right) ^2
\]

\[
\leq \frac{2}{k^{4\eta}} \left\{ T_p \left( \frac{u_p}{(2\theta^\ell)^\gamma} \right) - \frac{1}{2\theta^\ell} \right\} + \left\{ T_p \left( \frac{u_p}{(2\theta^\ell)^\gamma} \right) - T_p \left( \frac{u_p}{(2\theta^\ell + 1)^\gamma} \right) \right\} \frac{1}{2\theta^\ell}
\]

Hence

\[
\sum_{i=1}^{p} \mathbb{E} \sup_{x, y \in \mathcal{F}_\ell} (Y_{p,i}(x) - Y_{p,i}(y))^2
\]

\[
\leq \frac{2}{4^{\eta}} \frac{p}{k} T_p \left( \frac{u_p}{(2\theta^\gamma)^\gamma} \right) \frac{1}{2\theta^\ell} \left( \frac{1}{2\theta^\ell} - 1 \right)^2 + \frac{2}{4^{\eta}} \frac{p}{k} \left\{ T_p \left( \frac{u_p}{(2\theta^\gamma)^\gamma} \right) - \frac{1}{2\theta^\ell} \right\} \frac{1}{2\theta^\ell}
\]

\[
= J_1 + J_2.
\]

Using Assumption 2.2 and Lemma 6.1, there exists constants \( M, C > 0 \) not depending on \( \varepsilon \) such that for small \( \delta > 0 \) and large \( p \)

\[
J_1 \leq 4^{1-\eta} M T_p \left( \frac{u_p}{(2\theta^\gamma)^\gamma} \right) \frac{1}{2^\ell} \left( \frac{1}{2^\ell} - 1 \right)^2 \leq 8MC\theta^{\ell(1-2\eta-\delta)} \left( \frac{1}{\theta^\ell} - 1 \right)^2 \leq 8MC \left( \frac{1}{\theta^{1/2}} - 1 \right)^2.
\]

On the other hand, applying Lemma 6.3 and using Assumption 2.2, for large \( t \) and \( x > 1 \)

\[
T_p(t) - T_p(tx) \leq MT(t) \log x, \text{ for some constant } M.
\]

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

\[
J_2 \leq \frac{2M\gamma p}{4^{\eta}} T \left( \frac{u_p}{(2\theta^\gamma)^\gamma} \right) \log (1/\theta) \frac{1}{2^\ell} \leq \frac{4M\gamma}{4^{\eta}} T \left( \frac{u_p}{(2\theta^\gamma)^\gamma} \right) \frac{1}{2^\ell} \log (1/\theta)
\]

where for second inequality Assumption 2.1 is used. Lemma 6.1 yields, for some constants \( C > 0 \) not depending \( \varepsilon \), for \( \delta > 0 \) small

\[
J_2 \leq \frac{4M\gamma C}{4^{\eta}} 2^{1-\delta} \theta^{\ell(1-2\eta-\gamma \delta)} \log (1/\theta) \leq 8M\gamma C \log (1/\theta).
\]
To conclude
\[
\sum_{i=1}^{p} \mathbb{E} \sup_{x,y \in F(i)} (Y_{p,i}(x) - Y_{p,i}(y))^2 \\
\leq 8MC \max\{1, \gamma\} \left\{ \left( \frac{1}{\theta^{1/2}} - 1 \right)^2 + \log(1/\theta) \right\} \leq 8MC \max\{1, \gamma\} (\varepsilon^2 + 2\varepsilon^3) \leq \varepsilon^2,
\]
for small \( \varepsilon > 0 \). \( \square \)

**PROOF OF THEOREM 2.1.** We have
\[
F = [0, 2] = [0, 2a] \cup \left( \bigcup_{\ell=0}^{[\log a/\log \theta]} [2\theta^\ell, 2\theta^{\ell+1}] \right).
\]
The number of elements of this covering is bounded by \( \varepsilon^{-4} \). The theorem follows now from Theorem 3 in Einmahl and Segers (2021), the conditions of which are verified in Lemmas 6.4–6.7 and by using this bound \( \varepsilon^{-4} \).

**PROOF OF THEOREM 2.2.** We have (see Example 5.1.5 in de Haan and Ferreira, 2006)
\[
\sqrt{k}(\hat{\gamma} - \gamma) = \sqrt{k} \int_{X_{p-k:p}/u_p}^1 \frac{p_k T_{emp}(s u_p)}{s} \frac{ds}{s} + \sqrt{k} \int_{X_{p-k:p}/u_p}^\infty \left( \frac{p_k T_{emp}(s u_p)}{s} - s^{-1/\gamma} \right) \frac{ds}{s}.
\]
On the probability space of Corollary 2.1 (with \( \eta > 0 \)), the second term on the right in (10) is equal to
\[
\gamma \sqrt{k} \int_0^1 \left( \frac{p_k T_{emp}(u_p x^\gamma)}{x^\gamma} - x \right) \frac{dx}{x} = \gamma \int_0^1 V(x) \frac{dx}{x} + \gamma \int_0^1 \frac{W_p(x) - V(x)}{x^{1-\eta}} \frac{dx}{x^{1-\eta}}.
\]
By Corollary 2.2, \( X_{p-k:p}/u_p \xrightarrow{p} 1 \), and the first term on the right in (10) is equal to
\[
\int_{X_{p-k:p}/u_p}^1 W_p(s^{-1/\gamma}) \frac{ds}{s} + \sqrt{k} \int_{X_{p-k:p}/u_p}^1 s^{-1/\gamma} \frac{ds}{s} \xrightarrow{a.s.} o(1) + \gamma \sqrt{k} \left( \left( \frac{X_{p-k:p}}{u_p} \right)^{-1/\gamma} - 1 \right) \xrightarrow{a.s.} -\gamma V(1),
\]
where the first term vanishes due to Corollary 2.1 and the second term converges because of Corollary 2.2. Hence we obtain that
\[
\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{a.s.} \gamma \left( -V(1) + \int_0^1 V(x) \frac{dx}{x} \right),
\]
which is a centered normal random variable.
To complete the proof we will show that
\[
\text{var} \left( -V(1) + \int_0^1 V(x) \frac{dx}{x} \right) = \text{var}(V(1)).
\]
We have
\[
\text{cov} \left( -V(1) + \int_0^1 V(x) \frac{dx}{x}, -V(1) + \int_0^1 V(y) \frac{dy}{y} \right) \\
= \text{var}(V(1)) - 2 \int_0^1 \text{cov}(V(x), V(1)) \frac{dx}{x} + \int_0^1 \int_0^1 \text{cov}(V(x), V(y)) \frac{dx}{x} \frac{dy}{y} \\
= \text{var}(V(1)) - 2 \int_0^1 \text{cov}(V(x), V(1)) \frac{dx}{x} + 2 \int_0^1 \int_0^1 \text{cov}(V(x), V(y)) \frac{dx}{x} \frac{dy}{y}.
\]

Now
\[
\int_0^1 \text{cov}(V(x), V(1)) \frac{dx}{x} = \int_0^1 (1 - R(1, x^{-1})) \, dx,
\]
and also, by the change of variables \(x/y = u\),
\[
\int_0^1 \int_0^y \text{cov}(V(x), V(y)) \frac{dx}{x} \frac{dy}{y} = \int_0^1 \int_0^y (1 - R(1, y/x)) \, dx \frac{dy}{y} = \int_0^1 \int_0^1 (1 - R(1, u^{-1})) \, dy \, du = \int_0^1 (1 - R(1, u^{-1})) \, du.
\]

\[\square\]

**Proof of Theorem 2.3.** Let \(U_{1,p} \leq U_{2,p} \ldots \leq U_{p,p}\) be the order statistics of \(\{U_{i,p}\}\), and define \(\bar{V}_{i,p}\) similarly. Note that, with probability 1, we can rewrite that
\[
\bar{R}(x,y) = \frac{1}{k} \sum_{i=1}^p \mathbb{1} \left[ U_{i,p} < U_{k,p}, \bar{V}_{i,p} < \bar{V}_{k,p} \right] = V_{p,k}(x,y) \left( \frac{p}{k} U_{k,p}, \frac{p}{k} \bar{V}_{k,p} \right),
\]
where
\[
V_{p,k}(x,y) = \frac{1}{k} \sum_{i=1}^p \mathbb{1} \left[ U_{i,p} < kx/p, \bar{V}_{i,p} < ky/p \right].
\]

Using Assumption 2.5 and (1), for \(x, y\) in a neighborhood of 1,
\[
\mathbb{E} V_{p,k}(x,y) = \frac{1}{k} \sum_{i=1}^p \mathbb{P} \left( U_{i,p} < kx/p \right) \mathbb{P} \left( U_{i,p} < ky/p \right) + o(1) \cdot \frac{1}{k} \sum_{i=1}^p \mathbb{P} \left( U_{i,p} < kx/p \right)
\rightarrow R(x,y) + 0 = R(x,y).
\]
Furthermore, for every \(x, y\) in a neighborhood of 1,
\[
\text{var}(V_{p,k}(x,y)) \leq \frac{1}{k^2} \sum_{i=1}^p \mathbb{P} \left( U_{i,p} < kx/p, \bar{V}_{i,p} < ky/p \right) = \frac{1}{k} \mathbb{E} V_{p,k}(x,y) \rightarrow 0.
\]
This implies that \(V_{p,k}(x,y) \xrightarrow{p} R(x,y)\) pointwise, for \(x, y\) in a neighborhood of 1. The convergence is then uniform by continuity of \(R\) and the monotonicity of \(V_{p,k}\) and \(R\).

It remains to show that
\[
P \frac{p}{k} U_{k,p} \overset{p}{\rightarrow} 1, \quad P \frac{p}{k} \bar{V}_{k,p} \overset{p}{\rightarrow} 1.
\]
Observe that Theorem 2.1 implies that $X_{p-k:p}/u_p \xrightarrow{p} 1$. This implies that, for any $c > 1$, with probability tending to 1
\[
\frac{p}{k} T_p(X_{p-k:p}) \geq \frac{p}{k} T_p(cu_p) = \frac{p}{k} T_p(u_p) \cdot \frac{T_p(cu_p)}{T_p(u_p)} \to 1 \cdot c^{-1/\gamma},
\]
and similarly
\[
\frac{p}{k} T_p(X_{p-k:p}) \leq \frac{p}{k} T_p(cu_p) = \frac{p}{k} T_p(u_p) \cdot \frac{T_p(cu_p)}{T_p(u_p)} \to 1 \cdot c^{1/\gamma}.
\]
As $c$ can be arbitrarily close to 1, $\frac{p}{k} U_{k,p} = \frac{p}{k} T_p(X_{p-k:p}) \xrightarrow{p} 1$. Finally we prove the second part of (11). Define $\tilde{U}_p \left( \frac{p}{xk} \right)$ as the $(1 - xk/p)$-th quantile of the average distribution function $\tilde{F}_p = 1 - T_p$. Since, for every $x$ in a neighborhood of 1,
\[
\text{var} \left( \frac{1}{k} \sum_{i=1}^{p} \mathbb{1} \left[ Y_i^{(p)} > \tilde{U}_p \left( \frac{p}{xk} \right) \right] \right) - \frac{p}{k} \tilde{T}_p \left( \tilde{U}_p \left( \frac{p}{xk} \right) \right) = \frac{1}{k} \sum_{i=1}^{p} \mathbb{1} \left[ Y_i^{(p)} > \tilde{U}_p \left( \frac{p}{xk} \right) \right] - x \xrightarrow{p} 0,
\]
we have
\[
\frac{1}{k} \sum_{i=1}^{p} \mathbb{1} \left[ Y_i^{(p)} > \tilde{U}_p \left( \frac{p}{xk} \right) \right] - \frac{p}{k} \tilde{T}_p \left( \tilde{U}_p \left( \frac{p}{xk} \right) \right) = \frac{1}{k} \sum_{i=1}^{p} \mathbb{1} \left[ Y_i^{(p)} > \tilde{U}_p \left( \frac{p}{xk} \right) \right] - x \xrightarrow{p} 0.
\]
For any $c > 1$, with probability tending to 1, $\tilde{U}_p \left( \frac{p}{ck} \right) \leq \tilde{Y}_{p-k:p} \leq \tilde{U}_p \left( \frac{cp}{k} \right)$ by monotonicity. It follows that $c^{-1} \frac{p}{k} \tilde{T}_p \left( \tilde{U}_p \left( \frac{cp}{k} \right) \right) \leq \frac{p}{k} \tilde{T}_p \left( \tilde{Y}_{p-k:p} \right) \leq \frac{p}{k} \tilde{T}_p \left( \tilde{U}_p \left( \frac{cp}{k} \right) \right) = c$. Since $c$ can be arbitrarily close to 1, this implies that $\frac{p}{k} \tilde{V}_{k,p} = \frac{p}{k} \tilde{T}_p \left( \tilde{Y}_{p-k:p} \right) \xrightarrow{p} 1$. 

6.2. Proofs from Section 3. Since the permutation $\pi$ 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 $H(x, y) = \int_{0}^{\infty} S \left( \frac{u}{y} \right) S \left( \frac{y}{u} \right) dF_{\sigma}(u)$.

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

**Lemma 6.8.** Let $X_1$ and $X_2$ be non-negative, 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,
\[
\lim_{t \to \infty} t^{1/\gamma} T(t) = \tilde{c}, \quad \text{for some } \tilde{c} \in (0, \infty).
\]

**Lemma 6.9.** If $T$ satisfies (12), then $U_\sigma(t) \leq C t^{\gamma}$, $t \geq 1$, for some constant $C$, and hence $1 - H_\sigma(t) = T_\sigma(t) \leq C^{1/\gamma} t^{1-1/\gamma}$, $t > 0$. We also have $S(t) \leq C^{1/\gamma} t^{-1/\gamma}$ for $t > 0$. 


PROOF. There exists \( \varepsilon > 0 \) such that \( S(\varepsilon) > 0 \). Furthermore, for some constant \( M \)

\[
M e^{-1/\gamma} (U_\sigma(t))^{-1/\gamma} \geq T(\varepsilon U_\sigma(t)) \geq T_\sigma(U_\sigma(t)) S(\varepsilon) = t^{-1} S(\varepsilon),
\]

where the first inequality is due to (12). Rewriting the inequality completes the proof of the first part. The result for \( S \) can be seen to hold by interchanging \( T_\sigma \) and \( S \) in the definition of \( T \). \( \square \)

**Lemma 6.10.** 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 \( \gamma \) and \( T \) as above if

\[
\int xg(x) \leq MT(x), \quad x \geq 0,
\]

for some constant \( M < \infty \).

PROOF. Let \( t = t(p) \rightarrow \infty \) be any intermediate threshold sequence such that \( pT(t) \rightarrow \infty \). It is possible to take a small (probability) sequence \( \beta = \beta(p) = o(T(t)) \) but \( p\beta \rightarrow \infty \). Consider the truncated function \( T^+(t) = \frac{1}{p} \sum_{i=[p\beta]+1}^{p} S\left(\frac{t}{Q_\sigma(1-i/p)}\right) \) and \( T^+(t) = \int_{[p\beta]/p}^{1} S\left(\frac{t}{Q_\sigma(1-u)}\right) du \). Since survival probabilities are bounded by 1,

\[
|T_p(t) - T^+_p(t)| + |T - T^+(t)| \leq \beta + \beta = o(T(t)).
\]

For Assumption 2.1, it remains to show that

\[
|T^+_p(t) - T^+(t)| = o(T(t)).
\]

For \( i = [p\beta] + 1, \ldots, p \), Lemma 6.3 gives that

\[
\int_{(i-1)/p}^{i/p} S\left(\frac{t}{Q_\sigma(1-u)}\right) - S\left(\frac{t}{Q_\sigma(1-i/p)}\right) du \leq M \int_{(i-1)/p}^{i/p} T\left(\frac{t}{Q_\sigma(1-u)}\right) \log \frac{Q_\sigma(1-u)}{Q_\sigma(1-i/p)} du.
\]

By the Lipschitz-continuity of \( \log U_\sigma(e^t) \), for some constants \( K, M \) and for \( [p\beta] \leq (i-1) \leq pu \leq i \),

\[
\log \frac{Q_\sigma(1-u)}{Q_\sigma(1-i/p)} \leq K \log \frac{i}{i-1} \leq M \frac{i}{pu}.
\]

Together with Lemma 6.9 and (12), for some constant \( C \)

\[
\int_{(i-1)/p}^{i/p} S\left(\frac{t}{Q_\sigma(1-u)}\right) - S\left(\frac{t}{Q_\sigma(1-i/p)}\right) du \leq \frac{M}{p} \int_{(i-1)/p}^{i/p} T\left(\frac{t}{Q_\sigma(1-u)}\right) du \leq \frac{C}{p} \int_{(i-1)/p}^{i/p} \left(\frac{tu^\gamma}{1}\right) du = \frac{Ct^{-1/\gamma}}{p} \int_{(i-1)/p}^{i/p} \frac{1}{u^2} du.
\]

Summing up the bounds over \( i \),

\[
|T^+_p(t) - T^+(t)| \leq \frac{Ct^{-1/\gamma}}{p} \int_{(i-1)/p}^{i/p} \frac{1}{u^2} du \leq \frac{2Ct^{-1/\gamma}}{p\beta} = o(T(t)).
\]

\( \square \)

**Lemma 6.11.** 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 \( x \geq 0 \)

\[
xg(x) \leq h(x), \quad \text{and} \quad \int_0^\infty h(x/u) dF_\sigma(u) \leq MT(x).
\]
PROOF. By monotonicity of \( h \) and \( Q_\sigma \), for all \( x > 0 \)
\[
xf_p(x) = \frac{1}{p} \sum_{i=1}^{p} g\left(\frac{x}{Q_\sigma(1-i/p)}\right) \frac{x}{Q_\sigma(1-i/p)} \leq \frac{1}{p} \sum_{i=1}^{p} h\left(\frac{x}{Q_\sigma(1-i/p)}\right)
\]
\[
\leq \int_{0}^{1} h\left(\frac{x}{Q_\sigma(1-u)}\right) du = \int_{0}^{\infty} h\left(\frac{x}{u}\right) dF_\sigma(u) \leq MT(x).
\]
Similarly, using monotonicity of \( S \), for \( x > 0 \)
\[
T_p(x) := \frac{1}{p} \sum_{i=1}^{p} S\left(\frac{x}{Q_\sigma(1-i/p)}\right) \leq \int_{0}^{1} S\left(\frac{x}{Q_\sigma(1-u)}\right) du = T(x).
\]

\[\square\]

**PROOF OF THEOREM 3.1.** By Lemma 6.8, we know that (12) holds and we claim that, for some constants \( C, M \) and \( x \geq 0 \)
\[
\int_{0}^{\infty} h(x/u)dF_\sigma(u) \leq Ch(0)T_\sigma(x) \leq MT(x),
\]
(13)
\[
xg(x) \leq MT(x).
\]
Part (i) then follows from Lemmas 6.10 and 6.11. To prove (13), we introduce a survival function \( \hat{h}(x) = \frac{h(x)}{h(0)} \) on the positive half-line. Then we have
\[
\int_{0}^{\infty} h(x/u)dF_\sigma(u) = h(0) \cdot \int_{0}^{\infty} h(x/u)dF_\sigma(u) =: h(0) \cdot \hat{T}(x).
\]
Observe that \( \hat{T} \) is the survival function of the product of two independent random variables with distribution functions \( F_\sigma \) and \( 1-\hat{h} \), respectively. Applying Lemma 6.8 yields that \( \lim_{x \to \infty} x^{-1/\gamma} \hat{T}(x) < \infty \). Therefore, for some large constants \( C_1 \) and \( C \),
\[
\hat{T}(x) \leq \min\{C_1x^{-1/\gamma}, 1\} \leq C_1 \min\{x^{-1/\gamma}, 1\} \leq CT_\sigma(x),
\]
and hence the first inequality in (13) follows. Similar arguments yield the second inequality. For (14) note that \( xg(x) \leq h(x) \leq C \min\{x^{-1/\gamma}, 1\} \leq MT(x). \)

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 6.8 with (12),
\[
zT(U_\sigma(z)) = \frac{T(U_\sigma(z))}{T_\sigma(U_\sigma(z))} \to \mathbb{E}Z_+^{1/\gamma} = \int_{0}^{\infty} S(v^{\gamma})dv \in (0, \infty).
\]
Following the definition of \( R \) in Assumption 2.4, we need to show that
\[
zH_p(U_\sigma(z)x^{-\gamma}, U_\sigma(z)y^{-\gamma}) \to \int_{0}^{\infty} S(v^{\gamma}x^{-\gamma})S(v^{\gamma}y^{-\gamma})dv, \quad x,y > 0.
\]
In fact, we only need to show that, as \( t \to \infty \),
\[
zH(U_\sigma(z)x^{-\gamma}, U_\sigma(z)y^{-\gamma}) \to \int_{0}^{\infty} S(v^{\gamma}x^{-\gamma})S(v^{\gamma}y^{-\gamma})dv
\]
because then, very similar to the proof of Lemma 6.10, we can show that
\[
zH_p(U_\sigma(z)x^{-\gamma}, U_\sigma(z)y^{-\gamma}) = zH(U_\sigma(z)x^{-\gamma}, U_\sigma(z)y^{-\gamma}) + o(1).
\]
Take another intermediate threshold sequence $\lambda = \lambda(z)$ such that $\lambda \to \infty$ but $z/\lambda \to \infty$ as $z \to \infty$. We have

$$zH(U_\sigma(z)x^{-\gamma}, U_\sigma(z)y^{-\gamma}) = \int_0^z S\left(\frac{U_\sigma(z)x^{-\gamma}}{U_\sigma(z/v)}\right) S\left(\frac{U_\sigma(z)y^{-\gamma}}{U_\sigma(z/v)}\right) dv$$

$$= \int_0^\lambda S\left(\frac{U_\sigma(z)x^{-\gamma}}{U_\sigma(z/v)}\right) S\left(\frac{U_\sigma(z)y^{-\gamma}}{U_\sigma(z/v)}\right) dv + \int_0^z S\left(\frac{U_\sigma(z)x^{-\gamma}}{U_\sigma(z/v)}\right) S\left(\frac{U_\sigma(z)y^{-\gamma}}{U_\sigma(z/v)}\right) dv$$

$$=: J_1(x, y) + J_2(x, y).$$

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 $z/\lambda$,

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

By monotonicity of $S$,

$$J_1(x, y) \leq \int_0^\infty S\left((1 - \varepsilon)v^{\gamma}x^{-\gamma}\right) S\left((1 - \varepsilon)v^{\gamma}y^{-\gamma}\right) dv = (1 - \varepsilon)^{-1/\gamma} \int_0^\infty S\left(v^{\gamma}x^{-\gamma}\right) S\left(v^{\gamma}y^{-\gamma}\right) dv,$$

and, on the other hand,

$$J_1(x, y) \geq \int_0^\lambda S\left((1 + \varepsilon)v^{\gamma}x^{-\gamma}\right) S\left((1 + \varepsilon)v^{\gamma}y^{-\gamma}\right) dv$$

$$\to \int_0^\infty S\left((1 + \varepsilon)v^{\gamma}x^{-\gamma}\right) S\left((1 + \varepsilon)v^{\gamma}y^{-\gamma}\right) dv = (1 + \varepsilon)^{-1/\gamma} \int_0^\infty S\left(v^{\gamma}x^{-\gamma}\right) S\left(v^{\gamma}y^{-\gamma}\right) dv.$$

It follows that $J_1(x, y) \to \int_0^\infty S\left(v^{\gamma}x^{-\gamma}\right) S\left(v^{\gamma}y^{-\gamma}\right) dv$ as $\varepsilon$ 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 $\infty$ for $z \geq 1$. Hence, for some large constant $C$,

$$J_2(x, y) \leq \int_\lambda^\infty S\left(Cv^{\gamma}x^{-\gamma}\right) S\left(Cv^{\gamma}y^{-\gamma}\right) dv$$

$$= C^{-1/\gamma} \int_{C^{1/\gamma}/\lambda}^\infty S\left(v^{\gamma}x^{-\gamma}\right) S\left(v^{\gamma}y^{-\gamma}\right) dv \to 0,$$

since $\int_\lambda^\infty S\left(v^{\gamma}x^{-\gamma}\right) S\left(v^{\gamma}y^{-\gamma}\right) dv \leq \int_0^\infty S\left(v^{\gamma}x^{-\gamma}\right) dv = x \int_0^\infty S\left(v^{\gamma}\right) dv < \infty$.

Now, for any intermediate threshold sequence $t = t(p)$, using the power-law approximation of $U_\sigma$, we can find two intermediate threshold sequences $U_\sigma(z_\pm)$ with $z_\pm = z_\pm(p) \to \infty$ such that $U_\sigma(z_-) \leq t \leq U_\sigma(z_+)$ 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 $H_p(t x^{-\gamma}, t y^{-\gamma})/T(t) \to R(x, y)$.

**Proof of Theorem 3.2.** Part (i) follows from Lemmas 6.10 and 6.11, since (12) holds due to Lemma 6.8 again and the condition in Lemma 6.11 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 $Q_\sigma$,

$$H_p(t x^{-\gamma}, t y^{-\gamma}) = \frac{1}{p} \sum_{i=1}^p S\left(\frac{tx^{-\gamma}}{Q_\sigma(1 - i/p)}\right) S\left(\frac{ty^{-\gamma}}{Q_\sigma(1 - i/p)}\right)$$

$$\leq \int_0^1 S\left(\frac{tx^{-\gamma}}{Q_\sigma(1 - u)}\right) S\left(\frac{ty^{-\gamma}}{Q_\sigma(1 - u)}\right) du = H(t x^{-\gamma}, t y^{-\gamma}).$$
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 6.9, for some constant \( C \)

\[
\frac{H(tx^{-\gamma}, ty^{-\gamma})}{T(t)} = \int_0^\lambda S \left( \frac{tx^{-\gamma}}{u} \right) S \left( \frac{ty^{-\gamma}}{u} \right) dF_\sigma(u) + \int_\lambda^\infty S \left( \frac{tx^{-\gamma}}{u} \right) S \left( \frac{ty^{-\gamma}}{u} \right) dF_\sigma(u)
\]

\[
\leq S \left( \frac{tx^{-\gamma}}{\lambda} \right) \int_0^\lambda S \left( \frac{ty^{-\gamma}}{u} \right) dF_\sigma(u) + \int_\lambda^\infty (Ct^{-1/\gamma} xu^{1/\gamma} \cdot 1) dF_\sigma(u)
\]

\[
\leq S \left( \frac{tx^{-\gamma}}{\lambda} \right) \frac{T(ty^{-\gamma})}{T(t)} + x \cdot \frac{Ct^{-1/\gamma}}{T(t)} \cdot \int_\lambda^\infty u^{1/\gamma} dF_\sigma(u) \to 0 \cdot y + x \cdot \frac{C}{c} \cdot 0 = 0,
\]

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

**REFERENCES**


