EXTREME VALUE INFEERENCE FOR GENERAL HETEROGENEOUS DATA

By

Yi He, John H.J. Einmahl

6 May 2024

ISSN 0924-7815
ISSN 2213-9532
Extreme Value Inference for General Heterogeneous Data

Yi He\(^1\) and John H.J. Einmahl\(^2\)

\(^1\)Amsterdam School of Economics, University of Amsterdam, e-mail: yi.he@uva.nl
\(^2\)Department of Econometrics and Operations Research, Tilburg University, e-mail: j.h.j.einmahl@uvt.nl

**Abstract:** We extend extreme value statistics to the general setting of independent data with possibly very different distributions, whereby the extreme value index of the average distribution can be negative, zero, or positive. We present novel asymptotic theory for the moment estimator, based on a uniform central limit theorem for the underlying weighted tail empirical process. We find that, due to the heterogeneity of the data, the asymptotic variance of the moment estimator can be much smaller than that in the i.i.d. case. We also unravel the improved performance of high quantile and endpoint estimators in this setup. In case of a heavy tail, we ameliorate the Hill estimator by taking an optimal combination of the Hill and the moment estimator. Simulations show the good finite-sample behavior of our limit results. Finally we present applications to the maximum lifespan of monozygotic twins and to the tail heaviness of energies of earthquakes around the globe.

**JEL Codes:** C13, 14.

**MSC2020 subject classifications:** Primary 62G32, 62G20, 62G30; secondary 60G70, 60F17.

**Keywords and phrases:** Endpoint estimation, extreme value statistics, heterogeneous data extremes, moment estimator, monozygotic twins, weighted tail empirical process.

1. Introduction

Consider a random sample of size \( p \in \mathbb{N} \) from some distribution function \( F \), that is, we observe independent and identically distributed random variables \( X_1, \ldots, X_p \) from \( F \). The statistical theory of extremes has been developed comprehensively for this setting; in particular, well-known estimators of the extreme value index \( \gamma \) have been introduced in, among others, Hill (1975) and Dekkers, Einmahl and de Haan (1989); see also the monographs Beirlant et al. (2004) and de Haan and Ferreira (2006). The results for random samples are important, but that setting is too restrictive for various applications. Fortunately, these and other estimators can be used successfully in much more general settings.

Statistics of extremes for dependent but identically distributed data has been studied extensively in the literature, see, e.g., Hsing (1991), Drees (2000), and Kulik and Soulier (2020).

In the present paper, we assume that the data are possibly non-identically distributed (heterogeneous, non-stationary) but independent. Clearly, such a
setting considerably enhances the applicability of the statistical theory. For early work for this situation see, e.g., Davison and Smith (1990), where a linear trend in the parameters of generalized Pareto distributions is considered. In a general, non-parametric setting recent relevant references for heterogeneous extremes are Einmahl et al. (2016), de Haan and Zhou (2021), and Einmahl and He (2023). We refer to the latter paper for a brief discussion of the other two and for some more references.

In Einmahl and He (2023) extreme value statistics is extended to independent data with possibly very different distributions, where the appropriately defined extreme value index of the average distribution is positive. It is shown there that the Hill estimator is asymptotically normal with an asymptotic variance that can be substantially smaller than that in the i.i.d. case. In contrast to that paper, here we allow the extreme value index $\gamma$ to be arbitrary, including negative values and zero. This is a major extension, relevant for many applications, where it is not known beforehand that $\gamma$ is positive like, e.g., to longevity, sport records, flooding, and weather and climate related variables.

The first goal of this paper is to prove a novel limit theorem for the moment estimator (see Dekkers, Einmahl and de Haan (1989)), based on a general uniform central limit theorem for the relevant weighted tail empirical process, that is of independent interest and useful more generally. Here the asymptotic variance can be much smaller than that in the i.i.d. case due to the heterogeneity, which leads to much more accurate statistical inference. In particular in case $\gamma$ is negative, it is interesting and relevant to estimate the right endpoint of the average distribution which, of course, makes no sense for heavy tailed distributions. We provide a new, detailed asymptotic result for the endpoint estimator in the heterogeneous case and show improved performance. This result is a special case of our general result for high quantile estimators. When $\gamma > 0$, we strikingly improve the Hill estimator by taking an optimal combination of the Hill and the moment estimator. All the asymptotic variances depend on a spurious, symmetric tail copula $R$, which actually measures heterogeneity. Two – of the many – interesting applications of the theory developed here to the maximum lifespan of Danish monozygotic twins and to the tail heaviness of global earthquake energies, can be found in Section 5; for various examples, see Sections 3 and 4.

In the remainder of this paper we first present in Section 2 our general results for the weighted tail empirical process, the moment estimator, the improved estimator in case of heavy tails, and the estimators of high quantiles and endpoints. In Section 3 we present as a main, general example an embedded heterogeneous scales model. Section 4 is devoted to simulations and Section 5 to the just mentioned applications. Most proofs of the results in Section 2 are collected in Section 6; the remaining proofs are deferred to the Supplementary Material (Section 7).
2. Asymptotic theory

2.1. Tail empirical process theory

Consider independent random variables $X_1^{(p)}, \ldots, X_p^{(p)} \in \mathbb{R}$, that are not necessarily identically distributed. We are interested in the tail inference on their empirical distribution function

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

which is centered at their average distribution function given by

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

For simplicity, we assume that $F_p$ is continuous for all $p$ throughout.

We generalize the extreme value theory beyond power laws in Einmahl and He (2023) to an arbitrary domain of attraction for both light and heavy tailed heterogeneous data, that may or may not have a finite support.

**Assumption 2.1.** The average survival function $T_p = 1 - F_p$ approaches some limit survival function $T = 1 - F$ in the tail such that $T_p(t)/T(t) \to 1$ for all intermediate sequences $t = t(p) \uparrow x^* := \sup \{x : T(x) > 0\}$ but $pT(t) \to \infty$, where $x^* > 0$ is called the endpoint of $T$. The limit distribution $F$ is in the domain of attraction of an extreme value distribution with index $\gamma \in \mathbb{R}$, that is, there exists a positive function $f$ such that

$$\lim_{t \uparrow x^*} \frac{T(t + xf(t))}{T(t)} = (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0, \quad (1)$$

where the limit should be interpreted as $\exp(-x)$ for $\gamma = 0$ by continuous extension.

When the data are i.i.d. with $T_{pi} = 1 - F_{pi} \equiv T$, the condition gives the foundation of extreme value statistics with various equivalent formulations available in Theorem 1.1.6 in de Haan and Ferreira (2006). For heterogeneous data, the limit relation (1) extends to the average distribution at the intermediate level: for any intermediate threshold $t \uparrow x^*$ and $pT(t) \to \infty$,

$$\frac{T_p(t + xf(t))}{T_p(t)} \to (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0. \quad (2)$$

It is well known in extreme value theory that we can choose

$$f(t) = \begin{cases} 
\gamma t & \gamma > 0, \\
-\gamma (x^* - t) & \gamma < 0, \\
\int_t^{x^*} T(x)dx/T(t) & \gamma = 0,
\end{cases} \quad (3)$$
meaning that, if there is any positive function $f$ satisfying condition (1), then the condition holds for this choice (3). To simplify presentation, we shall take this function $f$ from now on.

While Assumption 2.1 only requires a limit relation at the intermediate level, the following condition controls the extreme observations at higher levels.

**Assumption 2.2.** There exists a positive constant $M > 0$ such that for all sufficiently large $t$ and $p$, $T_p(t) \leq MT(t)$.

This condition is trivial for i.i.d. data with $M = 1$. When the data are heterogeneous, it allows for some ‘outliers’ to escape from the limit (2) when applying the peaks-over-threshold procedures. These extreme observations, however, should not go too far away from the domain to dominate the tail inference. In particular, it implies that $T_p(x^*) = 0$ and thus the right endpoint of $F_p$ is bounded by that of the limit population $F$. We note that this condition is weaker than the stability condition in Einmahl and He (2023), relaxing their smoothness condition based on the average density function. In fact, we do not require the existence of an average density function here.

Our tail inference relies on an intermediate sequence $k = k(p) \in \{1, \ldots, p-1\}$ satisfying:

**Assumption 2.3.** $k \to \infty$ and $k/p \to 0$ as $p \to 0$.

The first part gives a large effective sample size, while the second part entails our target at the tail rather than the central distribution. To define a tail empirical process on a uniform level, we first invert the limit function on the right-hand-side of (2) to obtain the equivalent representation given by, for all intermediate sequences $t \to x^*$ from Assumption 2.1,

$$
\frac{T_p(t + \phi_\gamma(x)f(t))}{T_p(t)} \to x, \quad \phi_\gamma(x) = \frac{x^{-\gamma} - 1}{\gamma}, \quad x > 0,
$$

where $\phi_\gamma(x)$ should be interpreted as $-\log x$ for $\gamma = 0$ via continuous extension. Note that the limit is now pivotal, that is, it does not depend on $\gamma$. Substituting $t = u_p$ the $(1 - k/p)$ generalized quantile of $T_p$, we define the tail empirical process, for $x > 0$

$$
V_p(x) = \frac{p}{\sqrt{k}} \left( T_{\text{emp}}(u_p + \phi_\gamma(x)f(u_p)) - T_p(u_p + \phi_\gamma(x)f(u_p)) \right)
$$

where $T_{\text{emp}} = 1 - F_{\text{emp}}$ is the empirical survival function. A direct calculation gives the covariance structure

$$
cov \left( V_p(x), V_p(y) \right) = \frac{p}{k} \left( T_p(u_p + \phi_\gamma(x)f(u_p)) + T_p(u_p + \phi_\gamma(y)f(u_p)) \right)
$$

where

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

Naturally, we assume the limit of this covariance function exists.
**Assumption 2.4.** For all intermediate sequences \( t = t(p) \to x^* \) with \( pT(t) \to \infty \),
\[
\frac{H_p(t + \phi_\gamma(x)f(t), t + \phi_\gamma(y)f(t))}{T(t)} \to R(x, y).
\]
Like in Einmahl and He (2023), we can show that \( R \equiv 0 \) (called trivial) for i.i.d. data. For non-identically distributed data, the function \( R \) may or may not be trivial, but it is invariance with respect to an increasing transformation of the data \( X_i^{(p)} \) via an alternative rank-based definition.

**Lemma 2.1.** Under Assumption 2.1 with \( \gamma \in \mathbb{R} \), Assumption 2.4 holds if and only if
\[
\frac{1}{p\alpha} \sum_{i=1}^{p} \mathbb{P} \left( t_i^{(p)} < \alpha x \right) \mathbb{P} \left( t_i^{(p)} < \alpha y \right) \to R(x, y), \quad T_i^{(p)} = T_p(X_i^{(p)})
\]
for all intermediate (probability) sequences \( \alpha = \alpha(p) \downarrow 0 \) with \( p\alpha \to \infty \).

This lemma generalizes Lemma 2.1 of Einmahl and He (2023) towards a general \( \gamma \in \mathbb{R} \). The \( R \) function is monotonic and shares some general properties of a symmetric tail copula function including:
- \( R(x, y) > 0 \) for all \( x, y > 0 \) if \( R \) is non-trivial,
- \( 0 \leq R(x, y) = R(y, x) \leq \min \{x, y\} \) for all \( x, y > 0 \),
- \( R(ax, ay) = aR(x, y) \) for all \( a, x, y > 0 \) (homogeneity),
- \( R(x, y) \geq \min \{x, y\} R(1, 1) \) for all \( x, y > 0 \),
- \( R(x, y) \leq \frac{x+y}{2} R(1, 1) \) for all \( x, y > 0 \) (concavity).

Let \( V \) be a centered Gaussian process on \([0, 2]\) with covariance function given by
\[
\text{cov}(V(x), V(y)) = \min \{x, y\} - R(x, y), \quad 0 \leq x, y \leq 2.
\]
Note that
\[
\text{var} \ V(x) = x - R(x, x) = x(1 - R(1, 1)).
\]

**Theorem 2.1.** Under Assumptions 2.1–2.4, the following hold.
(i) If \( \gamma \neq 0 \), for any \( 0 \leq \eta < 1/2 \)
\[
\frac{V_p}{T_n} \sim \frac{V}{I^n} \text{ in } \ell^\infty([0, 2]),
\]
where \( I \) denotes the identity function and \( 0/0 := 0 \); almost all sample paths of \( V/I^n \) are uniformly continuous.
(ii) Part (i) extends to \( \gamma = 0 \) for \( 0 \leq \eta < \beta/2 \) if there exists constants \( M > 0 \) and \( \beta \in (0, 1) \) such that \( T(t + \phi_\gamma(x)f(t)) \leq Mx^\beta T(t) \) for all sufficiently large \( t \) and small \( x \).

The condition in part (ii) is weaker than the usual second-order regular variation condition on \( T \) with a second-order index \( \rho < 0 \) and \( 0 < \beta < -\rho \); see
Chapter 2 of de Haan and Ferreira (2006) for the second-order conditions. Note that this condition holds for all $\beta \in (0, 1)$ in part (i) when $\gamma \neq 0$.

To apply the tail empirical process theory, as discussed above, one often require a slightly modified tail empirical process by replacing $T_p (u_p + \phi \gamma (x) f(u_p))$ with $\frac{k}{p} x$ according to the tail approximation (4):

$$W_p(x) = \frac{p}{\sqrt{k}} \left( T_{emp} (u_p + \phi \gamma (x) f(u_p)) - \frac{k}{p} x \right) = \sqrt{k} \left( \frac{p}{\sqrt{k}} T_{emp} (u_p + \phi \gamma (x) f(u_p)) - x \right)$$

Using the Skorohod representation theorem we obtain:

**Corollary 2.1.** Consider cases (i) and (ii) of Theorem 2.1, and denote by $\bar{\eta} = 1/2$ when $\gamma \neq 0$ otherwise $\bar{\eta} = \beta/2$ when $\gamma = 0$. If for some $0 \leq \eta < \bar{\eta}$ and $\delta \in (0, 1)$,

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

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

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

Applying the Vervaat (1972) lemma and the delta method gives us the limits of all intermediate quantiles.

**Corollary 2.2 (Intermediate Quantile).** Under the conditions and on the probability space of Corollary 2.1, uniformly for $s$ on any compact subset of $(0, 1+\delta)$

$$\sqrt{k} \left( \frac{X_{p-[k]p} - u_p}{f(u_p)} - s^{-\gamma - 1} \gamma \right) \overset{a.s.}{\rightarrow} s^{-\gamma - 1} V(s).$$

In particular, for $s = 1$,

$$\frac{\sqrt{k} (X_{p-k+p} - u_p)}{f(u_p)} \overset{a.s.}{\rightarrow} V(1).$$

### 2.2. Asymptotic normality of the moment and Hill estimators

For any real-valued extreme value index $\gamma \in \mathbb{R}$, we can decompose it as a sum of its positive and negative parts:

$$\gamma = \gamma_+ + \gamma_- , \quad \gamma_+ = \max\{0, \gamma\} , \quad \gamma_- = \min\{0, \gamma\}.$$

To estimate these two parts, we first introduce the order statistics $X_{1,p} \leq \ldots \leq X_{p,p}$ of the observations $\{X_i^{(p)} : 1 \leq i \leq p\}$, and define the first two moments of the log exceedances by

$$M_p^{(j)} = \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{p-i,p} - \log X_{p-k,p})^j , \quad j = 1, 2.$$
Then we can estimate $\gamma_+$ by the Hill (1975) estimator $\hat{\gamma}_H = M_p^{(1)}$. For $\gamma_-$, as shown in Dekkers, Einmahl and de Haan (1989), we can exploit the limit relation $(M_p^{(1)})^2/M_p^{(2)} \overset{p}{\to} (1 - 2\gamma_-)/(2(1 - \gamma_-))$ to construct a consistent estimator $\hat{\gamma}_-$. Combining both estimators yields the moment estimator of the extreme value index:

$$\hat{\gamma}_M = \hat{\gamma}_H + \hat{\gamma}_- = M_p^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_p^{(1)})^2}{M_p^{(2)}}\right)^{-1}. \quad (10)$$

To analyze the (joint) asymptotic normality of the Hill estimator $\hat{\gamma}_H$ and the moment estimator $\hat{\gamma}_M$, we first establish the joint limit of the first two moments $M_p^{(1)}$ and $M_p^{(2)}$. Define the normalization sequence $s_p = f(u_p)/u_p$, and then $s_p = \gamma$ when $\gamma > 0$ or $s_p \to 0$ when $\gamma \leq 0$ using the auxiliary function $f$ from (3) associated with the limit survival function $T$; see Lemma 1.2.9 of de Haan and Ferreira (2006) for details.

**Theorem 2.2.** Under the conditions of Corollary 2.1 with $\eta > 0$, on the same probability space

$$\sqrt{k} \left( \frac{M_p^{(1)}}{s_p} - \frac{1}{1 - \gamma_-} \right) \overset{p}{\to} M_1, \quad \sqrt{k} \left( \frac{M_p^{(2)}}{s_p^2} - \frac{2}{(1 - \gamma_-)(1 - 2\gamma_-)} \right) \overset{p}{\to} M_2,$$

with

$$M_j = -V(1) \cdot j \int_0^1 \phi_{\gamma_-}^{-1}(x) dx - \int_0^1 V(x) d\phi_{\gamma_-}(x), \quad j = 1, 2,$$

provided that $\sqrt{k}s_p \to 0$ when $\gamma \leq 0$. Since $M_p^{(1)} = \hat{\gamma}_H$ and $s_p \to \gamma_+$, it follows that

$$\sqrt{k} \left( \hat{\gamma}_H - \gamma_+ \right) \overset{p}{\to} \gamma_+ M_1 \sim N(0, \gamma_+^2 (1 - R(1, 1))). \quad (11)$$

Then we can establish the limit of the moment estimator in the next theorem as a linear combination of $M_1$ and $M_2$ using the same delta method as in Corollary 3.2 of Dekkers, Einmahl and de Haan (1989); we do not repeat the proof. However, the joint distribution of our limit variables $M_j$ takes a general form depending on the function $R$ via the covariance structure (7) of our limiting process $V$. Define

$$m_R(x) = (1 + x) \int_0^1 u^x R(1, 1/u) du, \quad x \geq 0. \quad (12)$$

**Theorem 2.3.** Under the conditions of Theorem 2.2 and on the same probability space,

$$\sqrt{k} \left( \hat{\gamma}_M - \gamma \right) = \sqrt{k} \left( \hat{\gamma}_H - \gamma_+ \right) + \sqrt{k} \left( \hat{\gamma}_- - \gamma_- \right) \overset{p}{\to} \Gamma_+ + \Gamma_-$$

where

$$\Gamma_+ = \gamma_+ M_1, \quad \Gamma_- = -2 (1 - \gamma_-)^2 (1 - 2\gamma_-) M_1 + \frac{1}{2} (1 - \gamma_-)^2 (1 - 2\gamma_-)^2 M_2.$$
The limit $\Gamma = \Gamma_+ + \Gamma_- \sim N \left(0, \sigma_\gamma^2 \cdot (1 - \nabla_\gamma)\right)$ with

$$
\sigma_\gamma^2 = \begin{cases} 
\frac{1 + \gamma^2}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)(1-4\gamma)} & \gamma \geq 0 \\
\frac{1 - \gamma^2}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)(1-4\gamma)} & \gamma < 0
\end{cases}
$$

and a heterogeneity effect given by

$$
\nabla_\gamma = \begin{cases} 
w_0^+(\gamma) \cdot R(1,1) + w_1^+(\gamma) \cdot m_R(0) & \gamma \geq 0 \\
w_0^-(\gamma) \cdot R(1,1) + w_1^-(\gamma) \cdot m_R(-\gamma) + w_2^-(\gamma) \cdot m_R(-2\gamma) & \gamma < 0
\end{cases}
$$

where the weights are given by

$$
\begin{bmatrix} w_0^+(\gamma) \\
w_1^+(\gamma) \\
w_2^+(\gamma) \\
w_0^-(\gamma) \\
w_1^-(\gamma) \\
w_2^-(\gamma) \end{bmatrix} = \frac{1}{1 + \gamma^2} \begin{bmatrix} (1 - \gamma)^2 \\
2\gamma 
\end{bmatrix}, \quad w_0^+(\gamma) + w_1^+(\gamma) = 1, \\
$$

and

$$
\begin{bmatrix} w_0^-(\gamma) \\
w_1^-(\gamma) \\
w_2^-(\gamma) \end{bmatrix} = \frac{1}{1 - \gamma + 6\gamma^2} \begin{bmatrix} (1 - 2\gamma)(1 - 3\gamma)(1 - 4\gamma) \\
-4\gamma(1 - 4\gamma) \\
12\gamma(1 - \gamma)(1 - 2\gamma) \end{bmatrix}, \quad w_0^-(\gamma) + w_1^-(\gamma) + w_2^-(\gamma) = 1.
$$

When $R \equiv 0$, the limits reduce to those in the i.i.d. case but still yield new results for a myriad of heterogeneous data models. When $R > 0$ is non-trivial, the heterogeneity effect $\nabla_\gamma$ are weighted averages of $R(1,1)$, $m_R(-\gamma)$, and $m_R(-2\gamma)$; when $\gamma \geq 0$, the last two terms coincide and are equal to $m_R(0)$. While the weights take different forms depending on the sign of $\gamma$, all these parameters, including the homogeneous variance $\sigma_\gamma^2$ and the heterogeneity effect $\nabla_\gamma$, are continuous in $\gamma$.

### 2.3. Improving the Hill estimator in case $\gamma > 0$

Consider the case $\gamma > 0$. The Hill estimator $\hat{\gamma}_H$ is equivariant with respect to power transformation, that is, if we apply a power function to all the positive data with any exponent $1/\theta > 0$ via

$$
X_{i}^{(p)} \mapsto [X_{i}^{(p)}]^{1/\theta}, \quad \forall X_{i}^{(p)} > 0,
$$

the extreme value index $\gamma \mapsto \gamma/\theta$ and Hill estimator $\hat{\gamma}_H \mapsto \hat{\gamma}_H/\theta$ transform in the same way. In contrast, the moment estimator violates this property as the estimator $\hat{\gamma}_-^-$ of the negative part does not change with $\theta$. Applying the moment estimator $\hat{\gamma}_M(\theta)$ after a power transformation (13) to estimate the new extreme value index $\gamma(\theta) = \gamma/\theta$ and then standardize it back to original scale with the extreme value index $\gamma = \theta \gamma(\theta)$ yields the estimator

$$
\hat{\gamma}(\theta) = \theta \hat{\gamma}_M(\theta) = \theta \left(\hat{\gamma}_H/\theta + \hat{\gamma}_-^-ight) = \hat{\gamma}_H + \theta \hat{\gamma}_-^-.
$$
An alternative representation of this new estimator $\hat{\gamma}(\theta)$ is a weighted average of the Hill and moment estimators given by

$$\hat{\gamma}(\theta) = (1 - \theta)\hat{\gamma}_H + \theta\hat{\gamma}_M, \quad \theta \geq 0,$$

(15)

where the weight $\theta$ on the moment estimator can be larger than 1.

The combined estimator $\hat{\gamma}(\theta) \xrightarrow{P} \gamma > 0$ is consistent regardless of the choice $\theta$. We shall show that our new asymptotic theory leads to a non-trivial optimal weight $\theta^*$, denoted by $\theta^*$, for heterogeneous data. In contrast to the i.i.d. case, it is possible to improve the Hill estimator if $R$ is non-trivial. We have the following corollary to Theorem 2.2

**Corollary 2.3.** Under the conditions of Theorem 2.2 with $\gamma > 0$, on the same probability space, for all $\theta \geq 0$,

$$\sqrt{n}(\hat{\gamma}(\theta) - \gamma) \xrightarrow{P} \Gamma_\theta$$

where the limit

$$\Gamma_\theta = \gamma \left( -V(1) + \int_0^1 s^{-1}V(s)ds \right) + \theta \left( V(1) - \int_0^1 (2 + \log s)s^{-1}V(s)ds \right)$$

is a Gaussian variable with mean zero and variance given by

$$\sigma^2(\theta) = \gamma^2(1 - R(1, 1)) - 2\theta \gamma |m_R(0) - R(1, 1)| + \theta^2(1 - R(1, 1)).$$

Whenever the asymptotic distribution of the Hill estimator in (11) is non-degenerate with $R(1, 1) < 1$, the asymptotic variance is a convex quadratic function of $\theta$ with the minimum point given by

$$\theta^* = \gamma \rho_R, \quad \sigma^2(\theta^*) = \sigma^2(0)(1 - \rho_R^2), \quad \rho_R = \frac{m_R(0) - R(1, 1)}{1 - R(1, 1)} \in [0, 1].$$

(16)

When $R$ is trivial, $\theta^* = 0$ and hence no improvement is possible upon the Hill estimator. However, when $R$ is non-trivial, a substantial improvement is possible and we refer to the examples in the next section and in the simulation study.

With any consistent estimator $\hat{R}$ of $R$ satisfying $\rho_{\hat{R}} \xrightarrow{P} \rho_R$, such as that given in Einmahl and He (2023), one may wish to estimate the optimal weight $\theta^*$ consistently by

$$\hat{\theta}^* = \hat{\gamma}_H \rho_{\hat{R}},$$

(17)

yielding the optimally weighted estimator

$$\hat{\gamma}(\hat{\theta}^*) = \hat{\gamma}_H(1 + \rho_{\hat{R}}\hat{\gamma}_M).$$

(18)

The following result, allowing more generally any consistent estimator $\hat{\theta}^*$ of $\theta^*$, follows immediately from Corollary 2.3.
Theorem 2.4. Under the conditions of Theorem 2.2 with γ > 0, for any consistent estimator $\hat{\theta}^*$ $\overset{p}{\to} \theta^*$ with $\theta^*$ defined in (16), on the same probability space

$$\sqrt{k}(\hat{\gamma}(\hat{\theta}^*) - \gamma) \overset{p}{\to} \Gamma_{\theta^*} \sim N(0, \sigma^2(\theta^*)),$$

with $\Gamma_{\theta^*}$ as defined in Corollary 2.3 and the minimal variance

$$\sigma^2(\theta^*) = \gamma^2(1 - R(1, 1))(1 - \rho_R^2).$$

Remark 1. In finite samples, the moment estimator may introduce more estimation bias than the Hill estimator due to the estimator $\hat{\gamma}^$ of $\gamma^ = \min\{\gamma, 0\} = 0$ for $\gamma > 0$. To mitigate this bias, one can substitute $\hat{\gamma}$ with a bias-corrected counterpart, as proposed in the simulation study in Section 4.3.

2.4. Inferring a high quantile and the endpoint

An important application of extreme value theory is the estimation of a very high quantile or the endpoint $x_\tau = Q_p(1 - \tau)$ with a very small $\tau \geq 0$,

where $Q_p$ denotes the generalized quantile function of $F_p$. Via the generalized Pareto model (2) with $t = u_p$ and $t + xf(t) = x_\tau$, we obtain $p\tau/k \approx (1 + \gamma(x_\tau - u_p)/f(u_p))^{-1/\gamma}$ or equivalently

$$x_\tau \approx u_p + f(u_p)\phi, (p\tau/k) = u_p(1 + s_p\phi, (p\tau/k)). \tag{19}$$

To estimate the tail approximation in (19), we can substitute the threshold $u_p$ with $X_{p-k_p}$, the extreme value index $\gamma$ and the normalization sequence $s_p$ by the unified estimator $\hat{\gamma}$ and $\hat{s}_p$ given by,

$$\hat{\gamma} = \begin{cases} \hat{\gamma}(\hat{\theta}^*) & \gamma > 0, \\ \hat{\gamma}_M & \gamma \leq 0, \end{cases} \quad \text{and} \quad \hat{s}_p = \begin{cases} \hat{\gamma}(\hat{\theta}^*) & \gamma > 0, \\ M_p^{(1)}(1 - \hat{\gamma}_-) & \gamma \leq 0. \end{cases} \tag{20}$$

If the sign of $\gamma$ is unknown, one may classify the cases as follows: choose the case $\gamma > 0$ if $\sqrt{k}\hat{\gamma}_H \geq C$, otherwise choose the case $\gamma \leq 0$, for any predetermined constant $C > 0$. We can deduce the asymptotic normality of $\hat{s}_p$ using Theorem 2.2 as follows.

Corollary 2.4. Under the conditions of Theorem 2.2, on the same probability space

$$\sqrt{k}\left(\frac{\hat{s}_p}{s_p} - 1\right) \overset{p}{\to} \Sigma, \quad \Sigma := \begin{cases} M_1 + \rho R\Gamma_+ & \gamma > 0, \\ M_1 - \frac{1}{1 - \gamma}\Gamma_- & \gamma \leq 0. \end{cases} \tag{21}$$

The quantile estimator at the level $1 - \tau$ is therefore given by

$$\hat{x}_\tau = X_{p-k_p}(1 + \hat{s}_p\phi, (p\tau/k)).$$
Note that this estimator reduces to the Weissman (1978) estimator when $\gamma > 0$, namely
\[
\hat{x}_\tau = X_{p-k:p} x_{\gamma} \left(1 + \frac{k}{p} \phi_x \left( \frac{pr}{k} \right) \right) = X_{p-k:p} \left( \frac{k}{p} \right)^{\gamma}.
\] (22)

When $\gamma < 0$, substituting $\tau = 0$ yields an estimator of the endpoint $x_p^* = x_0 = Q_p(1)$:
\[
\hat{x}_0 = X_{p-k:p} \left(1 - \frac{s_p}{\gamma} \right).
\]

We require the following condition to remove the asymptotic bias in the extreme value approximation (19), which is formulated such that it bears some similarity with condition (8).

**Assumption 2.5.** For the sequence of quantile levels $\tau = \tau(p) < k/p$,
\[
\sqrt{k} \left( \frac{p}{k} T_p(u_p + \phi_x(x) f(u_p)) - x \right) / x^{\max \{1+\gamma-,0\}} \rightarrow 0, \quad x = \left(1 + \gamma \frac{x_{\tau} - u_p}{f(u_p)} \right)^{-1/\gamma}.
\]

For $\tau = x = 0$ and $\gamma < 0$, the condition is trivial according to the convention $0/0 := 0$.

Define the following functions for $z > 0$ and $\gamma \in \mathbb{R}$
\[
\omega_\gamma(z) = \left( \int_{-1}^{1} \log sd \phi_x(s) \right)^{-1}, \quad \bar{\omega}_\gamma(z) = \omega_\gamma(z) \phi_x(z)
\]
and let $\omega_\gamma(0) := \lim_{z \downarrow 0} \omega_\gamma(z) = \gamma^2_-$ and $\bar{\omega}_\gamma(0) := \lim_{z \downarrow 0} \bar{\omega}_\gamma(z) = -\gamma_-$ via continuous extension. Write $\varphi_{p,\gamma}(z) = s_p^{-1} \omega_\gamma(z) + \bar{\omega}_\gamma(z)$.

**Theorem 2.5.** Let $pr = o(k)$, but $\log(pr)/\sqrt{k} \rightarrow 0$ if $\gamma \geq 0$. Under the conditions of Theorem 2.2 and Assumption 2.5, on the same probability space,
\[
\varphi_{p,\gamma} \left( \frac{pr}{k} \right) \sqrt{k} \left( \frac{x_{\tau} - 1}{x_{\tau}} \right) \xrightarrow{p} Q, \quad Q = \Gamma + \gamma^2 V(1) - \gamma \Sigma,
\]
where $\Gamma$ is the probabilistic limit of $\sqrt{k}(\gamma - \gamma)$ given by
\[
\Gamma = \begin{cases}
\Gamma_0 & \gamma > 0 \\
\Gamma_0 + \Gamma_- & \gamma \leq 0.
\end{cases}
\]

The limit $Q$ is a centered Gaussian variable not depending on $\tau$ in either case. When $\gamma > 0$, $Q = \Gamma_0$ has variance $\gamma^2(1 - R(1,1))(1 - \rho^2_R)$ as given in Theorem 2.4. When $\gamma \leq 0$, $Q$ has variance $\sigma^2_{Q,\gamma}(1 - \Delta_\gamma)$ with
\[
\sigma^2_{Q,\gamma} = \frac{(1 - \gamma)^2 (1 - 3\gamma + 4\gamma^2)}{(1 - 2\gamma)(1 - 3\gamma)(1 - 4\gamma)}
\]
and a heterogeneity effect given by

\[
\Delta_\gamma = w_0(\gamma) \cdot R(1,1) + w_1(\gamma) \cdot m_R(-\gamma) + w_2(\gamma) \cdot m_R(-2\gamma)
\]

\[
\begin{bmatrix}
w_0(\gamma) \\
w_1(\gamma) \\
w_2(\gamma)
\end{bmatrix} = \frac{1}{1 - 3\gamma + 4\gamma^2} \begin{bmatrix}
(1 - 2\gamma)(1 - 3\gamma)(1 - 4\gamma) \\
-2\gamma(1 - \gamma)(1 - 4\gamma) \\
8\gamma(1 - 2\gamma)^2
\end{bmatrix}, \quad w_0(\gamma) + w_1(\gamma) + w_2(\gamma) = 1.
\]

The theorem can be extended to a more general case, \( p \tau / k \to \nu \in [0,1) \), where the limit

\[
Q = Q(\nu) = \Gamma + \omega_\gamma(\nu) V(1) + \bar{\omega}_\gamma(\nu) \Sigma,
\]

adapts to the level \( \tau \). While the expression of its variance becomes more involved, it remains tractable and in closed-form. In this paper, we focus on the extreme case with \( \nu = 0 \), which corresponds to very high quantiles.

If \( \gamma > 0 \), the asymptotic variance of the quantile estimator is equal to that of \( \tilde{\gamma}(\theta^*) \) and hence minimal.

If \( \gamma < 0 \), the theorem allows us to extrapolate up to an arbitrarily high level. Taking \( \tau = 0 \) gives the following result for the endpoint estimation, via the delta method.

**Corollary 2.5.** Under the conditions of Theorem 2.5 with \( \gamma < 0 \),

\[
\varphi_{p,\gamma}(0) \cdot \sqrt{k \log \frac{x_p}{x_p^*}} \xrightarrow{d} N(0, \sigma_{Q,\gamma}^2 (1 - \Delta_\gamma)),
\]

with \( \varphi_{p,\gamma}(0) = \frac{s_{p\gamma}}{p_{\gamma}} - \gamma \to \infty \). The result remains true if one replaces \( x_p^* \) with \( x^* \) or/and \( \varphi_{p,\gamma}(0) \) with its leading term \( \frac{s_{p\gamma}}{p_{\gamma}} \) because \( s_p \to 0 \).

Finally, the following lemma allows us to estimate \( \varphi_{p,\gamma} \left( \frac{k \tau}{k} \right) \) with the sample counterpart.

**Lemma 2.2.** Under the conditions of Theorem 2.5,

\[
\frac{\hat{\varphi}_{p,\gamma}(p \tau / k)}{\varphi_{p,\gamma}(p \tau / k)} \xrightarrow{p} 1, \quad \hat{\varphi}_{p,\gamma}(z) = \frac{s_{p\gamma}}{p_{\gamma}} - \omega_{p\gamma}(z) + \bar{\omega}_{p\gamma}(z).
\]

**3. Leading example: embedded heterogeneous scales model**

For i.i.d. data with the same survival function \( T_{pi} \equiv T \), an equivalent formulation of the domain of attraction condition depends on the tail quantile function \( U(t) := \inf\{x : T(x) \leq 1/t\} \) via the embedding

\[
X_i^{(p)} = U(Z_i), \quad Z_i \sim \text{Pareto}(1),
\]

as follows; see, e.g., Theorem 1.1.6 of de Haan and Ferreira (2006).

**Assumption 3.1.** There exists a positive function \( a \) such that for \( x > 0 \),

\[
\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},
\]

where for \( \gamma = 0 \) the right-hand-side is interpreted as \( \log x \).
In this section, we generalize the representation (23) to allow for a particular type of heterogeneity whose \( R \)-function can be calculated explicitly. Consider the following embedded heterogeneous scales model:

\[
X_i^{(p)} = U(\sigma_{pi}Z_i), \quad \sigma_{pi} = Q_\sigma(1 - \pi(i)/p), \quad 1 \leq i \leq p,
\]

where \( U \) is some increasing left-continuous transformation, \( Q_\sigma \) is the generalized quantile function of a continuous distribution function \( F_\sigma \) with positive left endpoint, and \( Z_1, \ldots, Z_p \) are i.i.d. latent continuous, positive random variables, not necessarily Pareto-distributed. The function \( \pi \) is an unknown permutation of \( 1, \ldots, p \).

The following theorem shows that heterogeneous scales parameters \( \sigma_{pi} \) can generate extreme value behaviors with a non-trivial \( R \)-function. Denote by \( U_\sigma(t) = Q_\sigma(1 - 1/t) \), the tail quantile function of \( F_\sigma \), and let \( S \) be the survival function of the \( Z_i \).

**Theorem 3.1.** Consider the heterogeneous model (24) satisfying Assumption 3.1, and conditions \( \lim_{t \to \infty} U_\sigma(t)/t \in (0, \infty) \) and \( 0 < \mathbb{E}Z_i < \infty \). If \( \log U_\sigma(e^t) \) is Lipschitz-continuous on \( [0, \infty) \) and \( Z_i \) has a density function \( g \) such that \( \sup_{x \geq 0} x^2 g(x) < \infty \), then:

(i) Assumptions 2.1 and 2.2 hold for

\[
T(x) = \int_0^\infty S\left(\frac{1}{uT_0(x)}\right) dF_\sigma(u)
\]

with the extreme value index \( \gamma \in \mathbb{R} \) of \( U \) given in Assumption 3.1 and endpoint \( x^* = \lim_{t \to \infty} U(t) \in \mathbb{R} \cup \{\infty\} \), with \( T_0(x) = 1/\inf\{t : U(t) > x\} \).

(ii) Assumption 2.4 holds with a non-trivial \( R \)-function given by

\[
R(x, y) = \frac{\int_0^\infty S(v/x)S(v/y)dv}{\int_0^\infty S(v)dv}, \quad x, y > 0.
\]

Under this theorem, we can compute the function \( \lambda(u) := R(1, 1/u) \) for calculating the parameter \( R(1, 1) = \lambda(1) \) and the moment parameters \( m_R \) in (12) of various distributions of \( Z_i \) as follows (we consider only distributions of unit scale without loss of generality):

- Beta\((a, 1)\), \( a > 0 \), with \( S(x) = 1 - x^a \) on \( (0, 1) \), and \( \lambda(u) = 1 - u^a/(2a + 1) \) for \( u \in (0, 1] \).
- Uniform\((1, 1 + 1/\xi)\), \( \xi > 0 \), with \( S(x) = 1 + \xi - \xi x \) on \((1, 1 + 1/\xi)\), and

\[
\lambda(u) = 1 - \frac{(\max\{u(1 + \xi) - \xi, 0\})^3}{(6\xi + 3)u^2}, \quad u \in (0, 1].
\]
- Lognormal\((0, \sigma)\), \( \sigma > 0 \), with \( S(x) = 1 - \Phi(\log x/\sigma) \) on \((0, \infty)\), and

\[
\lambda(u) = \Phi \left( \left( -\log u/\sigma - \sigma \right)/\sqrt{2} \right) + \frac{1}{u} \Phi \left( \left( \log u/\sigma - \sigma \right)/\sqrt{2} \right).
\]
- Weibull($\kappa$), $\kappa > 0$, with $S(x) = \exp(-x^\kappa)$ on $(0, \infty)$, and $\lambda(u) = (1+u^\kappa)^{-1/\kappa}$.
- Pareto($1+\epsilon$), $\epsilon > 0$, with $S(x) = x^{-(1+\epsilon)}$ on $[1, \infty)$, and $\lambda(u) = 1-\epsilon u/(2\epsilon+1)$ for $u \in (0, 1]$.

The next theorem gives the cases with trivial $R \equiv 0$, comprising the i.i.d. model $(23)$ as well as many heterogeneous models.

**Theorem 3.2.** Consider the heterogeneous model $(24)$ satisfying Assumption 3.1, and conditions $\lim_{t \to \infty} tS(t) \in (0, \infty)$ and $\int_0^\infty xdf_{\sigma}(x) < \infty$. If $\log U_{\sigma}(\epsilon t)$ is Lipschitz-continuous on $[0, \infty)$ and $Z_i$ has a density function $g$ such that $xg(x) \leq MS(x)$, $x \geq 0$, for some constant $M < \infty$, then the results in Theorem 3.1 remain true, but the $R$-function becomes trivial ($R \equiv 0$).

Clearly, under the conditions of Theorems 3.1 and 3.2 we can proceed with obtaining the results in Section 2 for a myriad of heterogeneous samples.

### 4. Simulations

We consider three sets of Monte Carlo simulations with 50000 replications from the embedded heterogeneous scales model in Section 3 for different domains of attraction. We start with bounded observations in the case $\gamma < 0$ and illustrate how the heterogeneity influences the asymptotic variance of our extreme value index and endpoint estimators. In the second set, we consider the case where the unknown $\gamma = 0$. Finally, we consider the heavy-tailed case $\gamma > 0$, and illustrate the improvements by our optimal estimator in Section 2.3 to the Hill estimator. To render comparable results across different distributions, we control for $R(1, 1)$ and solve the corresponding distribution parameters. Throughout we set $p = 3500$, and choose the same $k = 200$ for simplicity.

#### 4.1. Inference for $\gamma < 0$

Consider heterogeneous variables with a population endpoint $x^*_p \equiv 1$ given by

$$X_i^{(p)} = 1 - \frac{1}{3} \left( \frac{p}{i} Z_i \right)^{\gamma}, \quad \gamma \in \{-0.1, -0.2, -0.3\}, \quad Z_i \sim \text{Pareto}(1+\epsilon), \quad (25)$$

where the scaling factor $1/3$ calibrates the parameter $s_p$ to that in our human lifespan application below. Indeed, $\gamma$ is the extreme value index corresponding to $F_p$. Note that the $R$-function is invariant with respect to $\gamma$ but depends on $\epsilon > 0$, with $R(1, 1) = 2\epsilon/(2\epsilon + 1)$.

Figure 1 depicts the extreme value index estimation error $\hat{\gamma}_M - \gamma$ for two data generating processes: (1) the heterogeneous data described in $(25)$; (2) i.i.d. data from their average distribution function $F_p$. Figure 2 compares the endpoint estimation error of our extreme-value-theory based estimator $\tilde{x}_0$. To render comparable results over different scale parameters $s_p$ depending on $\epsilon$, we report the rescaled error $\varphi_{p, \gamma}(0) \log \tilde{x}_0$ for the endpoint estimation. In both figures, the i.i.d. data generate similar behaviors across different distributions.
Fig 1. Boxplots of extreme value index estimation errors $\hat{\gamma}_M - \gamma$.

Fig 2. Boxplots of endpoint estimation errors $\varphi_{p, \gamma}(0) \log \hat{x}_0$.

of $Z_1$, as the scales heterogeneity dominates the tail behavior. The boxes are consistently shorter for heterogeneous data, according to our asymptotic theory, and the reduction becomes substantial as the heterogeneity measure $R(1,1)$ increases.
4.2. Estimation of $\gamma$ when it is equal to 0

Next, consider heterogeneous variables with $\gamma = 0$ and $x^* = \lim_{p \to \infty} x^*_p = \infty$, given by

$$X_i(p) = \log \left( \frac{p}{i} Z_i \right),$$

for independent $Z_i$ from some common distributions:

(I) $\text{Unif}(1, 1 + 1/\xi)$ for $\xi > 0$ with $R(1, 1) = (6\xi + 2)/(6\xi + 3)$.

(II) $\text{Lognormal}(0, \sigma)$ for $\sigma > 0$ with $R(1, 1) = 2\Phi(-\sigma/\sqrt{2})$.

(III) $\text{Pareto}(1 + \epsilon)$ for $\epsilon > 0$ with $R(1, 1) = 2\epsilon/(2\epsilon + 1)$.

Under Theorem 2.3, when $\gamma = 0$, the moment estimator $\hat{\gamma}_M$ has the asymptotic variance $1 - R(1, 1)$, decreasing with $R(1, 1)$ as depicted by the boxplots in Figure 3. This behavior is different from that for the i.i.d. data, for which the moment estimator maintains similar scales over all cases.

![Boxplots](image)

**Fig 3.** Boxplots of the moment estimator $\hat{\gamma}_M$ for $\gamma = 0$.

4.3. Improving tail estimation for $\gamma > 0$

We generate heterogeneous variables

$$X_i^{(p)} = \left[ \left( \frac{p}{i} - \frac{1}{2} \right) Z_i \right]^\gamma$$

with a fixed $\gamma = 1$ and independent $Z_i$ from some common distribution from the following classes:
(I) Beta(\(a, 1\)), for \(\nu > 0\), with \(R(1, 1) = \frac{2a}{a+1}\) and \(\theta^\ast = \frac{a}{a+1}\);

(II) Weibull(\(\kappa\)), for \(\kappa > 0\), with \(R(1, 1) = 2^{-1/\kappa}\) and

\[
\theta^\ast = \left(\int_0^1 (1 + x^\kappa)^{-1/\kappa} \, dx - 2^{-1/\kappa}\right) / \left(1 - 2^{-1/\kappa}\right);
\]

(III) Pareto(1 + \(\epsilon\)) for \(\epsilon > 0\) with \(R(1, 1) = 2^{-\epsilon}\) and \(\theta^\ast = \frac{\epsilon}{1+\epsilon}\).

The optimal weight \(\theta^\ast\) for our weighted estimator in Section 2.3 increases with \(R(1, 1)\) in all cases.

Through unreported simulations, we discovered that the weighted estimator often exhibits more bias than the Hill estimator in finite samples, see Remark 1. This discrepancy arises because the \(M_p^{(j)}\) in (9) tend to bias towards to the Riemann sum \(\gamma j \frac{1}{k} \sum_{i=0}^{k-1} (\log((k+1)/(i+1)))^j\) rather than its limit \(\gamma j \int_0^1 (-\log x)^j dx = \gamma j \cdot j\) for \(j \in \{1, 2\}\), under power-law approximations. To mitigate this bias, we recommend normalizing the moments \(M_p^{(j)}\) by their finite-sample power-law approximation, given by:

\[
\hat{M}_p^{(j)} = \frac{j \cdot \sum_{i=0}^{k-1} (\log X_{p-i:p} - \log X_{p-k:p})^j}{\sum_{i=0}^{k-1} (-\log(i+1) + \log(k+1))^j}, \quad j = 1, 2,
\]

yielding the modified moment estimator

\[
\hat{\gamma}_M^\ast = \hat{\gamma}_H + \hat{\gamma}_M^\ast = \hat{\gamma}_H + 1 - \frac{1}{2} \left(1 - \frac{(\hat{M}_p^{(1)})^2}{\hat{M}_p^{(2)}}\right)^{-1}.
\]

In the sequel, we shall always use these modified estimators for the weighted estimator \(\hat{\gamma}(\theta)\) via either (14) or (15), given by

\[
\hat{\gamma}(\theta) = \hat{\gamma}_H + \theta \hat{\gamma}_M^\ast = (1 - \theta) \hat{\gamma}_H + \theta \hat{\gamma}_M^\ast.
\]

Observe that the normalized moments \(\hat{M}_p^{(j)}\) are asymptotically indistinguishable from the \(M_p^{(j)}\) because the Riemann sums approach their limit at a faster rate than \(k^{-1/2}\), meaning that they are interchangeable in our asymptotic theory in Section 2.2.

Figure 4 shows that our improved estimator \(\hat{\gamma}(\theta^\ast)\) consistently achieves smaller mean absolute error and mean squared error against the Hill estimator \(\hat{\gamma}_H\). We use the error statistics for the Hill estimator using i.i.d. data drawn from the average distribution function \(F_p\) as benchmarks, and report the relative errors as ratios of these benchmarks. According to Theorem 2.2, the Hill estimator has a smaller asymptotic variance for heterogeneous data (dashed line) than for i.i.d. data, showing average errors decreasing with \(R(1, 1)\). Our improved estimator (solid line) reduces the errors further with non-trivial improvements according to our Theorem 2.4.

5. Real-life examples

This section presents two real-life examples.
5.1. Human life span

First, we analyze the life spans of $p = 3544$ individuals forming 1772 Danish monozygotic twins born between 1870 and 1910 that died in Denmark, as documented in the Danish Twin Registry. Here we assume that the two life spans of twins are identically distributed, which makes it possible to estimate $R$, see Theorem 2.3 in Einmahl and He (2023). The probability-probability plot on the left in Figure 5 shows a good alignment of the twin life spans. The life span heterogeneity across twins is evident from the quintile transition matrix on the right, with a statistically significant sum of relative self-transition frequencies of 1.42, meaning that twins tend to stay in the same quintile groups. We then estimate $R(1,1)$ and obtain, for $k = 200$, $R(1,1) = 0.31$.

We estimate the extreme value index of the (average) life span distribution with $\hat{\gamma}_M = -0.1779$, for $k = 200$, with an asymptotic 95% upper confidence bound of $-0.0644 < 0$, according to Theorem 2.3. This implies a negative extreme value index and hence a finite endpoint of the life span distribution. We obtain as the estimate of this endpoint $\hat{x}_0 = 113.7$ years, suggesting the possibility of supercentenarians (age at least 110 years) beyond our observations.

Combining Corollary 2.5 and Lemma 2.2, we obtain 124.4 as an asymptotic 95% upper confidence bound of the life span endpoint, which is 0.7 years lower than 125.1, obtained from the i.i.d. model. This result for Danish monozygotic twins is similar to those for the entire Dutch population, see Einmahl, Einmahl and
Finally, we examine the improved accuracy of extreme earthquake estimation achieved by applying our new estimator to a global dataset comprising the 3191 most significant earthquakes spanning the last 100 years (1924-2023), provided by the National Oceanic and Atmospheric Administration (NOAA). We correct the magnitudes, denoted by $M$, for rounding errors and convert them to earthquake energies $E$ using $E = 10^{1.5(M-6.8)}$. Employing the sample splitting strategy in Einmahl and He (2023), we obtain $\rho_B = 0.4910$ and the estimated optimal weight $b^* = 0.4674$ on the moment estimator via (17). Then we improve the full-sample estimate of extreme value index for earthquake energies to be $\hat{\gamma}(b^*) = 0.9832$ via (18); the Hill estimator yields $\hat{\gamma}_H = 0.9519$, slightly lower.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\gamma}$</th>
<th>S.E. of $\hat{\gamma}$</th>
<th>95% CI for $\gamma$</th>
<th>$M_\tau$</th>
<th>S.E. of $M_\tau$</th>
<th>95% CI for $M_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>0.9832</td>
<td>0.0503</td>
<td>(0.8846, 1.0818)</td>
<td>9.2467</td>
<td>0.0624</td>
<td>(9.1243, 9.3690)</td>
</tr>
<tr>
<td>Hill</td>
<td>0.9519</td>
<td>0.0559</td>
<td>(0.8423, 1.0615)</td>
<td>9.1987</td>
<td>0.0698</td>
<td>(9.0637, 9.3337)</td>
</tr>
<tr>
<td>i.i.d. Hill</td>
<td>0.9519</td>
<td>0.0673</td>
<td>(0.8206, 1.0838)</td>
<td>9.1987</td>
<td>0.0829</td>
<td>(9.0362, 9.3612)</td>
</tr>
</tbody>
</table>

Plugging the new estimate $\hat{\gamma}(b^*) = 0.9832$ into the Weissman estimator (22) yields an improved estimate of the 100-year earthquake energy $\hat{x}_\tau = 4677.30$ for $\tau = 1/3191$, as much as 18% higher than the estimate $\hat{x}_\tau = 3962.93$ based on the Hill estimator. These energy estimates are equivalent to magnitudes of 9.2467 and 9.1987, respectively. Table 1 shows that the confidence intervals taking advantage of spatial heterogeneity are substantially shorter than those based on the i.i.d. setup. Our improved estimators have smaller standard errors, leading to shorter confidence intervals, according to Theorem 2.5.

Fig 5. Danish monozygotic twins born in 1870-1910.
6. Proofs from Section 2

6.1. Proof of Lemma 2.1

The proof is completely analogous to that of Lemma 2.1 in Einmahl and He (2023), by replacing $tx^{-\gamma}$ and $ty^{-\gamma}$ therein with the more general forms $t + \phi(x)f(t)$ and $t + \phi(y)f(t)$ respectively. We omit details. □

6.2. Proof of Theorem 2.1

Define
\[ \zeta_i^{(p)} = \frac{1}{1 - F_p(X_i^{(p)})} = \frac{1}{T_p(X_i^{(p)})}, \quad 1 \leq i \leq p, \]
with standard Pareto average survival function:
\[ \frac{1}{p} \sum_{i=1}^{p} \mathbb{P} \left( \zeta_i^{(p)} > t \right) = \frac{1}{p} \sum_{i=1}^{p} \mathbb{P} \left( X_i^{(p)} > T_p^{-1}(t^{-1}) \right) = T_p(T_p^{-1}(t^{-1})) = t^{-1}, \quad t > 1, \]
where $T_p^{-1}(u) = \inf \{ x : T_p(x) \leq u \} = \inf \{ x : F_p(x) \geq 1 - u \}$ denotes the generalized inverse of $T_p$. Define the tail empirical process of $\{ \zeta_i^{(p)} \}$ by
\[ G_p(x) = \frac{p}{\sqrt{k}} \left( \frac{1}{p} \sum_{i=1}^{p} \mathbb{1} \left[ \zeta_i^{(p)} > \frac{p}{kx} \right] - \frac{p}{kx} \right), \quad x > 0. \]

Applying Theorem 2.1 in Einmahl and He (2023) to the transformed data $\zeta_i^{(p)}$ yields the following functional central limit theorem. Note that Assumptions 2.1 and 2.2 therein (special cases of Assumptions 2.1 and 2.2 here) are trivial for the standard Pareto distribution and the $R$-function remains unchanged according to Lemma 2.1.

**Lemma 6.1.** Under Assumptions 2.3 and 2.4, for any $0 \leq \eta < 1/2$,
\[ \frac{G_p}{T^{\eta}} \Rightarrow V \quad \text{in } L^\infty([0, 3]), \]
where $I$ denotes the identity function and $0/0 := 0$.

Applying the Skorohod representation theorem to this result, there exist a probability space carrying probabilistically equivalent versions of $G_p$ and $V$ (still denoted with $G_p$ and $V$) such that
\[ \sup_{0 < x \leq 3} \frac{|G_p(x) - V(x)|}{x^{\eta}} \overset{a.s.}{\longrightarrow} 0. \quad (26) \]
Lemma 6.2. Let \( t = t(p) \) be any intermediate sequence from Assumption 2.1. Then the limit relation (2) holds uniformly for \( x \) on any compact subinterval of \((0, \infty)\).

Proof. The pointwise convergence follows from Assumption 2.1: for any fixed \( x > 0 \)

\[
I_p(x) := \frac{TP(t + \phi_\gamma(x)f(t))}{TP(t)} = \frac{TP(t + \phi_\gamma(x)f(t))}{TP(t)} \frac{T(t) TP(t)}{TP(t)} \frac{T(t + \phi_\gamma(x)f(t))}{TP(t)} \to 1.1x,
\]

as \( t + \phi_\gamma(x)f(t) \) is also an intermediate sequence. But the function \( I_p \) is increasing for every \( p \) and the limit (the identity function) is a uniformly continuous function, and thus the convergence is uniform.

The following lemma allows us to control the observations beyond the intermediate level, which is essential for establishing the convergence of the weighted empirical process near the origin.

Lemma 6.3. Let \( t = t(p) \) be any intermediate sequence from Assumption 2.1 and suppose that also Assumption 2.2 holds.

(i) If \( \gamma \neq 0 \), for all \( \iota \in (0, 1) \), there exists a constant \( C \) such that for large \( p \)

\[
\frac{TP(t + \phi_\gamma(x)f(t))}{TP(t)} \leq Cx^\iota, \quad 0 \leq x \leq 2.
\]

(ii) When \( \gamma = 0 \), the result remains true for all \( \iota \in (0, \beta] \) with the constant \( \beta \in (0, 1) \) given in part (ii) of Theorem 2.1 (under the assumption there: \( T(t + \phi_\gamma(x)f(t)) \leq Mx^\beta T(t) \)).

Proof. We have \( TP(t)/T(t) \to 1 \). Hence under Assumption 2.2 we only need to prove the lemma for \( T \) instead of \( TP \). Part (i) follows from Lemma 7.1 in Einmahl and He (2023); for the case \( \gamma < 0 \), use \( \bar{T} \) defined by \( \bar{T}(x) = T(x^* - 1/x) \). Case (ii) follows immediately from the assumption involving \( \beta \in (0, 1) \) in part (ii) of Theorem 2.1.

Take \( V_p(x) = G_p(I_p(x)) \) with \( G_p \) from (26) and \( I_p(x) = \frac{TP(u_p + \phi_\gamma(x)f(u_p))}{TP(u_p)} \) as in the proof of Lemma 6.2, with \( t = u_p \). By the triangle inequality,

\[
\frac{|V_p(x) - V(x)|}{x^\eta} = \frac{|G_p(I_p(x)) - V(x)|}{x^\eta} \leq \frac{|G_p(I_p(x)) - V(I_p(x))|}{x^\eta} + \frac{|V(I_p(x)) - V(x)|}{x^\eta} =: \Delta_1(x) + \Delta_2(x).
\]
Take any $\eta' \in (\eta, 1/2)$. Then

$$\Delta_1(x) = \frac{|G_p(I_p(x)) - V(I_p(x))|}{I_p^\eta(x)} \eta' .$$

Uniformly for $x \in (0, 2]$, the perturbed location $I_p(x) \leq I_p(2) \to 2 < 3$ by the monotonicity of $I_p$ and the domain of attraction condition (2). Hence, for all large $p$,

$$\sup_{0 < x \leq 2} \frac{|G_p(I_p(x)) - V(I_p(x))|}{I_p^\eta(x)} \leq \sup_{0 < x < 3} \frac{|G_p(x) - V(x)|}{x^{\eta'}} \xrightarrow{a.s.} 0$$

by using (26). On the other hand, applying Lemma 6.3 with $t = u_p$, by choosing a $\eta'$ close enough to $1/2$ and $t$ sufficiently large (close to 1 for $\gamma \neq 0$ but to $\beta$ for $\gamma = 0$) such that $t - \eta/\eta' > 0$, we have $I_p(x)/x^{\eta'/\eta'} \leq Cx^{t-\eta/\eta'} \leq C2^{t-\eta/\eta'} \leq \infty$, uniformly for all $x \in (0, 2]$. Thus $\sup_{0 < x \leq 2} \Delta_1(x) \xrightarrow{a.s.} 0$.

Next we consider $\Delta_2$. For every $\delta \in (0, 1)$,

$$\sup_{\delta \leq x \leq 2} \frac{|V(I_p(x)) - V(x)|}{x^{\eta'}} \leq \sup_{\delta \leq x \leq 2} \frac{|V(I_p(x)) - V(x)|}{x^{\eta'}} \xrightarrow{a.s.} 0,$$

because of the fact that $V$ has uniformly continuous sample paths and Lemma 6.2.

Let $\varepsilon \in (0, 1)$. We have

$$\sup_{0 < x \leq \delta} \frac{|V(I_p(x)) - V(x)|}{x^{\eta'}} \leq \sup_{0 < x \leq \delta} \frac{|V(I_p(x))|}{x^{\eta'}} + \sup_{0 < x \leq \delta} \frac{|V(x)|}{x^{\eta'}}.$$

Since $V/I^\eta$ is almost surely continuous and takes the value 0 at 0, we have for small enough $\delta$ that $\mathbb{P}(\sup_{0 < x \leq \delta} |V(x)|/x^{\eta} \geq \varepsilon) \leq \varepsilon$. Let $\eta < \eta' < 1/2$ if $\gamma \neq 0$, and $\eta' < 1/2$ if $\gamma = 0$. Then for small enough $\delta_1$ and large $p$,

$$\mathbb{P}\left(\sup_{0 < x \leq \delta} \frac{|V(I_p(x))|}{x^{\eta'}} \geq \varepsilon\right) \leq \mathbb{P}\left(\sup_{0 < x \leq \delta} \frac{|V(I_p(x))|}{(I_p(x))^{\eta'}} \sup_{0 < x \leq \delta} \frac{(I_p(x))^{\eta'}}{x^{\eta'}} \geq \varepsilon\right) \leq \varepsilon.$$

Here again the a.s. continuity of $V/I^\eta$ is used and for the second supremum in the last probability we use Lemma 6.3.

\section*{6.3. Proof of Theorem 2.2}

We invoke the probability space of (26); the probability space used in Corollary 2.1 can and will be taken the same. Unless specified otherwise, the inequalities (only) hold with probability 1 (for large $p$).

We start from rewriting

$$M_p^{(j)} = -\frac{p}{k} \int_{X_p-k_p}^{x^*} (\log x - \log X_{p-k:p})^j dT_{\text{emp}}(x)$$

$$= \int_{X_p-k_p}^{x^*} \frac{p}{k} T_{\text{emp}}(x) d(\log x - \log X_{p-k:p})^j = \tilde{M}_p^{(j)} + \varepsilon_p^{(j)}$$
where
\[
\tilde{M}^{(j)}_p = - \int_0^1 \left[ \frac{p}{k} T_{\text{emp}}(X_{p-k:p}(1 + \phi_\gamma(x)s_p)) \right] d (\log (1 + \phi_\gamma(x)s_p))^j
\] (27)
and
\[
\varepsilon^{(j)}_p = \begin{cases} 0 & \gamma \geq 0 \\ \int_{x^*}^{x^*} \left[ \frac{p}{k} T_{\text{emp}}(x) \right] d (\log x - \log X_{p-k:p})^j & \gamma < 0 \end{cases}
\]
with
\[
x^* = \min \{x^*, X_{p-k:p}(1 + \phi_\gamma(0)s_p)\} = \min \left\{1, \frac{X_{p-k:p}}{u_p}\right\} \cdot x^*.
\] (28)

We first show that the perturbation \(\varepsilon^{(j)}_p\) is negligible.

**Lemma 6.4.** \(\sqrt{k} \varepsilon^{(j)}_p \overset{a.s.}{\longrightarrow} 0\).

**Proof.** We only need to consider \(\gamma < 0\). By the monotonicity of \(T_{\text{emp}}\) and \((\log x - \log X_{p-k:p})^j\) on \(x \in [\tilde{x}^*, x^*]\),
\[
\sqrt{k} \varepsilon^{(j)}_p \leq \frac{p}{k} T_{\text{emp}}(\tilde{x}^*) \cdot \sqrt{k} \left| \left( \frac{x^*}{X_{p-k:p}} \right)^j - \left( \frac{x^*}{u_p} \right)^j \right| = \frac{p}{k} T_{\text{emp}}(\tilde{x}^*) \cdot \Delta^{(j)}_{p,\varepsilon}.
\] (29)

Combining (28) with Corollaries 2.1 and 2.2 gives that, almost surely
\[
\frac{p}{k} T_{\text{emp}}(\tilde{x}^*) = \left( \gamma \cdot \frac{\tilde{x}^* - x^*}{f(u_p)} \right)^{-1/\gamma} + o(1) = \left( \gamma \cdot \min \left\{0, \frac{X_{p-k:p} - u_p x^*}{f(u_p)} \right\} \right)^{-1/\gamma} + o(1) \rightarrow 0.
\]
It remains show that the second term \(\Delta^{(j)}_{p,\varepsilon} = O(1)\), almost surely, on the right-hand-side of (29). First, recall that \(s_p \rightarrow 0\) when \(\gamma < 0\) and therefore
\[
\frac{X_{p-k:p}}{u_p} - 1 = \frac{X_{p-k:p} - u_p x^*}{f(u_p)} \cdot s_p \overset{a.s.}{\longrightarrow} 0 \cdot 0 = 0.
\]
Then
\[
\Delta^{(1)}_{p,\varepsilon} = \frac{\sqrt{k}}{s_p} \left| \log \frac{X_{p-k:p}}{u_p} \right| = O(1) \cdot \sqrt{k} \left| \frac{X_{p-k:p} - u_p x^*}{f(u_p)} \right| = O(1).
\]
Similarly,
\[
\Delta^{(2)}_{p,\varepsilon} \leq \frac{\sqrt{k}}{s_p} \left( \log \frac{X_{p-k:p}}{u_p} \right)^2 + 2 \frac{\sqrt{k}}{s_p} \left| \log \frac{X_{p-k:p}}{u_p} \right| \cdot \frac{1}{s_p} \left| \frac{x^*}{u_p} \right| = O(k^{-1/2}) + O(1) \cdot \frac{1}{s_p} \left| \frac{x^*}{u_p} - 1 \right| = O(k^{-1/2}) + O(1)O(1).
\]
Next, we normalize the leading term (27) and rewrite it by

$$\frac{\widehat{M}^{(j)}}{s_p} = -\int_0^1 \left[ \frac{P}{k} T_{\text{emp}}(u_p + \phi_\gamma(\hat{I}(x))) f(u_p) \right] d\psi_{p,\gamma}^j(x),$$

where

$$\psi_{p,\gamma}(x) = \frac{\log(1 + \phi_\gamma(x) s_p)}{s_p} \leq \phi_{\gamma^{-}}(x), \quad x \in (0,1),$$

and the perturbed identity function $\hat{I}(x) \geq 0$ is the solution of the equation

$$\begin{cases}
X_{p-k,p}(1 + \phi_\gamma(x) \cdot s_p) = u_p(1 + \phi_\gamma(\hat{I}(x)) \cdot s_p) & \gamma \geq 0, \\
\min \{ x^*, X_{p-k,p}(1 + \phi_\gamma(x) \cdot s_p) \} = u_p(1 + \phi_\gamma(\hat{I}(x)) \cdot s_p) & \gamma < 0.
\end{cases}$$

Using the upper bound of $\psi_{p,\gamma}$ in (31), it is easy to obtain the following lemma.

**Lemma 6.5.** The following hold for all $\varepsilon > 0$ and $j \in \{1, 2\}$:

(i) $\left| \int_{x_{\varepsilon,0}}^1 x^\varepsilon d\psi_{p,\gamma}^j(x) \right| \leq C < \infty$, for some constant $C$ that only depends on $j$, $\varepsilon$ and $\gamma^{-}$.

(ii) $\int_{x_{\varepsilon,0}}^x x^\varepsilon d\psi_{p,\gamma}^j(x) \to 0$ for any sequence $x_p \to 0$.

The results remain true for $\varepsilon = 0$ when $\gamma < 0$.

**Proof.** First consider $\varepsilon > 0$. Observe that $\psi_{p,\gamma}(1) = 0$ and

$$\lim_{x \downarrow 0} x^\varepsilon \psi_{p,\gamma}^j(x) \leq \lim_{x \downarrow 0} x^\varepsilon \phi_{\gamma^+}^j(x) \leq \lim_{x \downarrow 0} x^\varepsilon |\log x|^j = 0.$$ 

Part (i) follows by integration by parts and the upper bound (31), via $\left| \int_{x_{\varepsilon,0}}^1 x^\varepsilon d\psi_{p,\gamma}^j(x) \right| \leq \int_{x_{\varepsilon,0}}^1 \phi_{\gamma^-}^j(x) dx \varepsilon < \infty$. Similarly, for part (ii), integration by parts also gives that for all large $p$

$$\left| \int_{0}^{x_p} x^\varepsilon d\psi_{p,\gamma}^j(x) \right| \leq x_p \psi_{p,\gamma}^j(x_p) + \int_{x_{\varepsilon,0}}^{x_p} \psi_{p,\gamma}^j(x) dx \varepsilon \leq x_p \phi_{\gamma}^j(x_p) + \int_{0}^{x_p} \phi_{\gamma^-}^j(x) dx \varepsilon \to 0,$$

where we recall that $\int_{0}^{1} \phi_{\gamma^-}^j(x) dx \varepsilon < \infty$ in the last step.

Now consider $\varepsilon = 0$ with $\gamma < 0$. We have

$$\int_{0}^{1} d\psi_{p,\gamma}^j(x) = \psi_{p,\gamma}^j(0) - \psi_{p,\gamma}^j(1) = \psi_{p,\gamma}^j(0) < \infty,$$

and similarly, as $x_p \to 0$,

$$\int_{0}^{x_p} d\psi_{p,\gamma}^j(x) = \psi_{p,\gamma}^j(0) - \psi_{p,\gamma}^j(x_p) \to \phi_{\gamma}^j(0) - \phi_{\gamma}^j(0) = 0.$$

□
The next lemma gives more precise limits.

**Lemma 6.6.** The following holds almost surely for any positive sequence $a_p \to 0$.

(i) $\psi_{p,\gamma}(x)/\phi_{\gamma,\gamma}(x) \to 1$ uniformly for $x \in (\exp(-a_p/s_p) \mathbb{I}[\gamma = 0], 1)$

(ii) $\bar{I}(x)/x \to 1$ uniformly for $x \in (x_p, 1)$ where: $x_p = 0$ for $\gamma > 0$, $x_p = \exp(-\sqrt{k}a_p/s_p)$ for $\gamma = 0$, and $x_p = (\sqrt{k} \cdot a_p)^{\frac{1}{2}}$ for $\gamma < 0$.

(iii) Consider $\gamma = 0$. Let $\varepsilon \in (0, 1)$ be arbitrary and $x_p \to 0$ be any vanishing sequence from part (ii), then for all $x \in (0,x_p]$ and large $p$, $\bar{I}(x) \leq x^{1-\varepsilon}$.

**Proof.** Part (i) is trivial for $\gamma > 0$ as the inequality in (31) is an equality. The case for $\gamma \leq 0$ follows from fact that $\log(1 + t)/t \uparrow 1$ with all $t = \phi_{\gamma}(x)s_p \downarrow 0$ uniformly. For part (ii), we obtain from (32) for $\bar{c}(x) = \bar{I}(x)/x$

$$\phi_{\gamma}(\bar{c}(x)) = \min \left\{ \frac{X_{p-k:p} - u_p}{f(u_p)} \cdot (1 + \phi_{\gamma}(x) \cdot s_p) \cdot x^{\gamma}, \phi_{\gamma}(0) \right\}, \quad x > 0, \quad (33)$$

where $\phi_{\gamma}(0) = \infty$ can be omitted for $\gamma \geq 0$ but $\phi_{\gamma}(0) = -1/\gamma$ for $\gamma < 0$. Applying Corollary 2.2 yields that, uniformly for $x \in (x_p, 1)$

$$\phi_{\gamma}(\bar{c}(x)) - \phi_{\gamma}(1) = \phi_{\gamma}(\bar{c}(x)) = O \left( k^{-1/2} x^\gamma + k^{-1/2} \phi_{\gamma}(x) \cdot s_p \cdot x^\gamma \right) = o(1),$$

whose asymptotic order is dominated by the first term inside the minimum operator in (33). Then apply the delta method to obtain part (ii).

Next, for part (iii) with $\gamma = 0$, we can rewrite (32) to get that

$$\phi_{\gamma}(\bar{I}(x)) = \frac{X_{p-k:p} - u_p}{f(u_p)} + \frac{X_{p-k:p}}{u_p} \phi_{\gamma}(x)$$

Note that the first term $O(k^{-1/2})$ according to Corollary 2.2 and

$$\frac{X_{p-k:p}}{u_p} - 1 = \frac{X_{p-k:p} - u_p}{f(u_p)} \cdot s_p = O(k^{-1/2}) \cdot o(1) \to 0.$$

Therefore, by taking a sufficiently small $\delta \in (0,\varepsilon)$, for all large $p$ and $x \in (0,x_p]$,

$$\phi_{\gamma}(\bar{I}(x)) \geq -\delta^2 + (1 - \delta/2)\phi_{\gamma}(x) \geq (1 - \delta)\phi_{\gamma}(x) \geq \phi_{\gamma}(x^{1-\varepsilon}).$$

The stated result follows by the monotonicity of $\phi_{\gamma}$. \hfill $\square$

By replacing the tail empirical survival function by its limit in (30), we can construct a stochastic approximation $\bar{M}_{p}^{(j)}$ of $\bar{M}_{p}^{(j)}$ given by

$$\frac{\bar{M}_{p}^{(j)}}{s_p} = - \int_{0}^{1} \bar{I}(x) d\psi_{p,\gamma}^{j}(x), \quad (34)$$

satisfying

$$\sqrt{k} \left( \frac{\bar{M}_{p}^{(j)}}{s_p} - \frac{\bar{M}_{p}^{(j)}}{s_p} \right) = - \int_{0}^{1} W_{p} \left( \bar{I}(x) \right) d\psi_{p,\gamma}^{j}(x).$$
The following lemma allows us to substitute the stochastic process $W_p$ with its limit $V$, integrator $\psi^j_{p,\gamma}$ with its limit $\phi^j_{\gamma,-}$, as well as to remove the perturbation in the identity function $\bar{I}(x)$.

**Lemma 6.7.** We have

$$\sqrt{k} \left( \frac{\bar{M}_p^{(j)}}{s_p} - \frac{\bar{M}_p^{(j)}}{s_p} \right) = - \int_0^1 V(x) \, d\phi^j_{\gamma,-}(x) - \bar{\Delta}_p^{(j)} \quad \bar{\Delta}_p^{(j)} \overset{P}{\to} 0.$$

**Proof.** First, with $\tau = \inf\{x > 0 : \bar{I}(x) > 0\}$,

$$\bar{\Delta}_p^{(j)} = \int_\tau^1 \left[ W_p(\bar{I}(x)) - V(\bar{I}(x)) \right] d\psi^j_{p,\gamma}(x) + \int_0^1 \left[ V(\bar{I}(x)) - V(x) \right] d\psi^j_{p,\gamma}(x)$$

$$+ \left\{ \int_0^1 V(x) d\psi^j_{p,\gamma}(x) - \int_0^1 V(x) d\phi^j_{\gamma,-}(x) \right\} =: \sum_{i=1}^3 \bar{\Delta}_p^{(j)}.$$

We shall show that $\bar{\Delta}_p^{(j)} \overset{a.s.}{\to} 0$ and $\bar{\Delta}_p^{(j)} \overset{P}{\to} 0$.

To apply Corollary 2.1 with $\eta, \delta > 0$ therein, rewrite

$$\bar{\Delta}_p^{(j)} = \int_\tau^1 \left[ W_p(\bar{I}(x)) - V(\bar{I}(x)) \right] (\bar{I}(x))^{\eta} d\psi^j_{p,\gamma}(x).$$

Since $\bar{I}(x) \in (0, 1 + \delta)$ for $x \in (\tau, 1)$ and large $p$ almost surely by Lemma 6.6, part (ii), the first term in the integrand is vanishing uniformly and we only need to show

$$\int_\tau^1 (\bar{I}(x))^{\eta} d\psi^j_{p,\gamma}(x) = \int_0^1 (\bar{I}(x))^{\eta} d\psi^j_{p,\gamma}(x) = O(1). \quad (35)$$

Now take a sequence $a_p \to 0$ vanishing slow enough in part (ii) of Lemma 6.6 such that the induced sequence $x_p \to 0$. Together with the monotonicity of the integrator $\psi^j_{p,\gamma}(x)$ and Lemma 6.5,

$$\left| \int_0^1 (\bar{I}(x))^{\eta} d\psi^j_{p,\gamma}(x) \right| \leq \left| \int_{x_p}^1 2x^{\eta} d\psi^j_{p,\gamma}(x) \right| + R_1 = O(1) + R_1,$$

where

$$R_1 = \begin{cases} O(1) \cdot \left| \int_{x_p}^1 x^{(1-\gamma)} d\psi^j_{p,\gamma}(x) \right| = o(1) & \gamma \geq 0 \\ O(x_p^{\eta}) \cdot \left| \int_0^{x_p} d\psi^j_{p,\gamma}(x) \right| = o(1) & \gamma < 0. \end{cases}$$

Next we consider $\bar{\Delta}_p^{(j)}$. For every $\delta \in (0, 1)$, by the triangle inequality

$$\left| \bar{\Delta}_p^{(j)} \right| = \left| \int_\delta^1 V(\bar{I}(x)) - V(x) \right| \, d\psi^j_{p,\gamma}(x) + \left| \int_0^\delta V(x) \, d\psi^j_{p,\gamma}(x) \right|$$

$$+ \left| \int_0^\delta V(\bar{I}(x)) \, d\psi^j_{p,\gamma}(x) \right| =: R_{2,1} + R_{2,2} + R_{2,3}.$$
By the uniform continuity of the sample paths of $V$ and part (i) of Lemma 6.5, for any $\eta > 0$,

$$R_{2,1} \leq \sup_{\delta < x \leq 1} \frac{|V(\bar{I}(x)) - V(x)|}{\delta^\eta} \cdot \left| \int_0^1 x^\eta d\psi^j_{p,\gamma}(x) \right| = o(1) \cdot O(1) \to 0.$$ 

Furthermore,

$$R_{2,2} \leq \sup_{0 < x \leq \delta} \frac{|V(x)|}{x^\eta} \cdot \left| \int_0^\delta x^\eta d\psi^j_{p,\gamma}(x) \right|.$$ 

Recall that $V/I^n$ is a.s. continuous and takes the value 0 at 0, and therefore the first term $\sup_{0 < x \leq \delta} |V(x)|/x^\eta$ can be made arbitrarily small with high probability. Since the second term $\int_0^\delta x^\eta d\psi^j_{p,\gamma}(x)$ is bounded, we also have that $R_{2,2}$ is arbitrarily small with high probability. Similarly, for a possibly different $\eta \geq 0$,

$$R_{2,3} \leq \sup_{\bar{I}(0) \leq x \leq \bar{I}(\delta)} \frac{|V(z)|}{z^\eta} \cdot \left| \int_0^\delta \left(\bar{I}(x)\right)^\eta d\psi^j_{p,\gamma}(x) \right|,$$

where the first term can be again be arbitrarily small with high probability and the second term is bounded almost surely by (35).

It remains to check that $\Delta^{(j)}_{p,3} \xrightarrow{p} 0$. By the Markov inequality, it suffices to show that

$$\mathbb{E}[\Delta^{(j)}_{p,3}]^2 \leq \int_0^1 \left(\psi^j_{p,\gamma}(x) - \phi^j_{\gamma}(x)\right)^2 dx \to 0,$$

where the inequality holds because the covariance in (7) does not exceed that of a standard Brownian motion. The convergence to 0 is trivial for $\gamma > 0$ because $\psi_{p,\gamma} \equiv \phi_0$. For $\gamma \leq 0$, the result remains true by using the elementary inequality $|(\log(1 + z))^j/z^j - 1| \leq z$ with $z = s_p\phi_\gamma(x) > 0$, for $j \in \{1, 2\}$, or equivalently

$$|\psi^j_{p,\gamma}(x) - \phi^j_{\gamma}(x)| \leq s_p\phi^{j+1}_\gamma(x), \quad x \in (0, 1], \quad j \in \{1, 2\}, \quad (36)$$

so that, since $s_p \to 0$,

$$\int_0^1 (\psi^j_{p,\gamma}(x) - \phi^j_{\gamma}(x))^2 dx \leq s_p^2 \int_0^1 \phi^{2j+2}_\gamma(x) dx \to 0.$$ 

\[\square\]

Next, we consider a further approximation $m_p^{(j)}$ of $\bar{M}_p^{(j)}$ in (34) given by

$$\frac{m_p^{(j)}}{s_p^1} = -\int_0^1 x d\psi^j_{p,\gamma}(x), \quad (37)$$

satisfying

$$\sqrt{k} \left(\frac{\bar{M}_p^{(j)}}{s_p^1} - \frac{m_p^{(j)}}{s_p^1} \right) = -\int_0^1 I(x)d\psi^j_{p,\gamma}(x), \quad \mathbb{E}(x) = \sqrt{k}(\bar{I}(x) - x). \quad (38)$$

Replacing $I(x)$ with its limit yields the following the lemma.
Lemma 6.8. For all \( j \in \{1, 2\} \),
\[
\sqrt{k} \left( \frac{M_p(j)}{s_p^j} - \frac{m_p(j)}{s_p^j} \right) = -V(1) \cdot j \cdot \int_0^1 \phi_{j-1}^\gamma(x)dx - \Delta_p(j), \quad \Delta_p(j) \xrightarrow{a.s.} 0.
\]

Proof. Define \( B_p := \frac{\sqrt{k}(X_{p-k:p} - u_p)}{f(u_p)} \). We have
\[
\Delta_p(j) = \int_0^1 \{ I(x) + B_p(1 + \phi(x)s_p)x^\gamma \} d\psi_{p,\gamma}(x) + (B_p - V(1)) \cdot j \cdot \int_0^1 \psi_{p,\gamma}^{j-1}(x)dx
\]
\[
+ V(1) \cdot j \cdot \left\{ \int_0^1 \psi_{p,\gamma}^{j-1}(x)dx - \int_0^1 \phi_{j-1}^\gamma(x)dx \right\} =: \sum_{i=1}^3 \Delta_p(j).
\]
We immediately obtain \( \Delta_p,j \xrightarrow{a.s.} 0 \), by combining part (i) of Lemma 6.6 and inequality (31): with probability 1, \( V(1) = O(1) \) and \( \int_0^1 \psi_{p,\gamma}^{j-1}(x)dx \rightarrow \int_0^1 \phi_{j-1}^\gamma(x)dx \) by the dominated convergence theorem. Furthermore, we have \( \Delta_p,2 \xrightarrow{a.s.} 0 \) because \( B_p \xrightarrow{a.s.} V(1) \) from Corollary 2.2 and \( \int_0^1 \psi_{p,\gamma}^{j-1}(x)dx \leq \int_0^1 \phi_{j-1}^\gamma(x)dx < \infty \).

It remains to show \( \Delta_p,3 \xrightarrow{a.s.} 0 \). Recall \( \tau = \inf\{x > 0 : \hat{I}(x) > 0\} \). When \( \gamma \geq 0 \), we have \( \tau = 0 \). When \( \gamma < 0 \), via (32), one can show that \( \tau = 0 \) if \( X_{p-k:p} \leq u_p \) (or equivalently \( B_p \leq 0 \)) because \( \hat{I}(x) \geq x > 0 \) for all \( x > 0 \); otherwise, if \( X_{p-k:p} > u_p \) (or equivalently \( B_p > 0 \)), \( \tau > 0 \) solves the equation
\[
X_{p-k:p}(1 + \phi(\tau)s_p) = x^\gamma.
\]
In summary, for all \( \gamma \in \mathbb{R} \),
\[
\sqrt{k}\tau^{-\gamma} = \frac{u_p}{X_{p-k:p}} \cdot (-\gamma + s_p) \cdot \max\{B_p, 0\} = O(1).
\]

Then we decompose:
\[
\Delta_p,1 = \int_0^1 \{ I(x) + B_p(1 + \phi(x)s_p)x^\gamma \} d\psi_{p,\gamma}(x)
\]
\[
- \sqrt{k} \int_0^\tau x d\psi_{p,\gamma} - B_p \cdot j \cdot \int_0^\tau \psi_{p,\gamma}^{j-1}(x)dx =: \sum_{i=1}^3 \Delta_p,1,i.
\]
Note that \( \tilde{c}(x) := \hat{I}(x)/x > 0 \) for \( x > \tau \), and then we can rewrite (33) as follows:
\[
\sqrt{k}(\phi(\gamma(\tilde{c}(x)) - \phi(1)) = B_p(1 + \phi(x)s_p)x^\gamma.
\]
Using the mean-value theorem, it follows that for all \( x \in (\tau, 1) \)
\[
\left| \sqrt{k}(\tilde{c}(x) - 1) + B_p x^\gamma(1 + \phi(x)s_p) \right| \leq |B_p| \left| [\tilde{c}(x)]^{\gamma+1} - 1 \right| x^{\gamma}(1 + \phi(x)s_p).
\]
Using this inequality and monotonicity of \( \psi_{p,\gamma} \),
\[
\left| \Delta_p,1 \right| \leq j \cdot |B_p| \int_0^1 |[\tilde{c}(x)]^{\gamma+1} - 1| \cdot \psi_{p,\gamma}^{j-1}(x)dx \leq j \cdot |B_p| \int_0^1 |[\tilde{c}(x)]^{\gamma+1} - 1| \phi_{j-1}^\gamma(x)dx.
\]
Since $B_p = O(1)$, we only need to show that the second factor is $o(1)$. This is trivial for $\gamma > 0$ by combining part (ii) of Lemma 6.6 and $\int_0^1 \phi_{\gamma}^{-1}(x)dx < \infty$. When $\gamma \leq 0$, taking a sequence $a_p \to 0$ vanishing slow enough such that $x_p \to 0$ in part (ii) of Lemma 6.6, we also have that $\int_{x_p}^1 \left| [\bar{c}(x)]^{\gamma+1} - 1 \right| \phi_{\gamma}^{-1}(x)dx = o(1)$ and thus we only need to check the reminder $\int_0^{x_p} \left| [\bar{c}(x)]^{\gamma+1} - 1 \right| \phi_{\gamma}^{-1}(x)dx = o(1)$, or equivalently, $R := \int_0^{x_p} [\bar{c}(x)]^{\gamma+1} \phi_{\gamma}^{-1}(x)dx = o(1)$. For $\gamma = 0$ this follows from part (iii) of Lemma 6.6 and $\int_0^1 x^{-\varepsilon} \phi_{\gamma}^{-1}(x)dx < \infty$ for any $\varepsilon \in (0, 1)$. When $\gamma < 0$, by the monotonicity of $\bar{I}$ and $\phi_\gamma$,

$$R \leq \left[ \bar{I}(2x_p) \right]^{\gamma+1} \phi_{\gamma}^{-1}(0) \int_0^{x_p} x^{-\gamma-1}dx = O(x_p^\gamma \cdot O(1)) \cdot O(x_p^{-\gamma}) = O(x_p) = o(1).$$

The proof is complete for $\gamma \geq 0$ already with $\Delta_{p,1,2}^{(j)} = \Delta_{p,1,3}^{(j)} = 0$. For $\gamma < 0$, it remains to check $\Delta_{p,1,2}^{(j)} \xrightarrow{a.s.} 0$ and $\Delta_{p,1,3}^{(j)} \xrightarrow{a.s.} 0$. The latter convergence is obvious as $B_p = O(1)$, $\int_0^1 \phi_{p,\gamma}^{-1}(x)dx < \infty$, and $\tau \to 0$ almost surely. On the other hand, recalling (31) and monotonicity of $\phi_\gamma$, almost surely,

$$\left| \Delta_{p,1,2}^{(j)} \right| = j \cdot \sqrt{k} \int_0^\tau \frac{x^{-\gamma}}{1 + \phi_\gamma(x)s_p} \psi_{p,\gamma}^{-1}(x)dx \leq j \cdot \sqrt{k} \int_0^\tau x^{-\gamma} \phi_{\gamma}^{-1}(x)dx = j \cdot \phi_{\gamma}^{-1}(0) \cdot \sqrt{k} \int_0^\tau x^{-\gamma}dx = O(\sqrt{k} \tau^{-\gamma+1}) \to 0,$$

where the last step is due to (39).

Now, combining Lemmas 6.7 and 6.8 yields

$$\sqrt{k} \left( M_p^{(j)} / s_p^{(j)} - m_p^{(j)} / s_p^{(j)} \right) \xrightarrow{a.s.} -V(1) \cdot j \cdot \int_0^1 \phi_{\gamma}^{-1}(x)dx - \int_0^1 V(x) d\phi_{\gamma}^{-1}(x).$$

Observe that on the right-hand-side these are the desired limits in Theorem 2.2. To complete the proof it remains to replace the deterministic approximation $m_p^{(j)} / s_p^{(j)}$ in (37) with its limit $\theta^{(j)} = - \int_0^1 x d\phi_{\gamma}^{-1}(x)$. This is trivial for $\gamma > 0$ in which case $m_p^{(j)} / s_p^{(j)} = \theta^{(j)}$. For $\gamma \leq 0$, via integration by parts and inequality (36), for $j = 1, 2$,

$$\sqrt{k} \left| m_p^{(j)} / s_p^{(j)} - \theta^{(j)} \right| \leq \sqrt{k} \int_0^1 \left| \psi_{p,\gamma}(x) - \phi_{\gamma}^{-1}(x) \right| dx \leq \sqrt{k} s_p \int_0^1 \phi_{\gamma}^{j+1}dx \to 0$$

as $\sqrt{k} s_p \to 0$. \qed

Acknowledgments

We are very grateful to the Danish Twin Registry, in particular to Prof. Jacob Hjelmborg and Lisbeth Aagaard Larsen, for kindly providing the life spans of the Danish monozygotic twins.
7. Supplementary Material

This supplementary material provides the proofs of Theorem 2.5, Lemma 2.2, and Theorems 3.1 and 3.2 in the main document.

7.1. Proof of Theorem 2.5

Let \( z_p = p\tau/k \) and \( \bar{x}_\tau = u_p + f(u_p)\phi_\gamma(z_p) \) denote the tail approximation from (19). We first show that \( x_\tau \) and \( \bar{x}_\tau \) are indistinguishable in our statements, because

\[ \frac{\varphi_{p,\gamma}(z_p)}{x_\tau} \sqrt{k} (x_\tau - \bar{x}_\tau) \to 0. \]  

(41)

Consider \( \gamma \geq 0 \). We obtain from Assumption 2.5 and the delta method,

\[ \sqrt{k} \left( \phi_\gamma \left( \left( z_p \left( 1 + \gamma \frac{x_\tau - u_p}{f(u_p)} \right)^{1/\gamma} \right)^{-1} \right) - \phi_\gamma(1^{-1}) \right) \to 0, \]

that is,

\[ \sqrt{k} \frac{z_p}{f(u_p)} (x_\tau - \bar{x}_\tau) \to 0, \]

and therefore

\[ \sqrt{k} \frac{z_p}{f(u_p)} (x_\tau - \bar{x}_\tau) \to 0, \]

(42)

Note that it implies that \( x_\tau/\bar{x}_\tau \to 1 \). Moreover,

\[ \varphi_{p,\gamma}(z_p) \frac{\sqrt{k} (x_\tau - \bar{x}_\tau)}{x_\tau} = \varphi_{p,\gamma}(z_p) \cdot \sqrt{k} \frac{x_\tau - \bar{x}_\tau}{x_\tau} \]

\[ = o \left( \frac{z_p^{-\gamma}s_p}{1 + s_p\phi_\gamma(z_p)} \right) = o \left( \gamma + \frac{(1 - \gamma)s_p}{1 + s_p\phi_\gamma(z_p)} \right) = o(1). \]

(43)

Consider \(-1 < \gamma < 0 \). We have

\[ \sqrt{k} |z_p^{-\gamma} - x^{-\gamma}| \leq \sqrt{k} |x^{-\gamma-1} z_p - x| \to 0, \]

that is, just rewriting the first term,

\[ \sqrt{k} \left| \frac{x_\tau - \bar{x}_\tau}{x^\gamma - u_p} \right| \to 0. \]

(44)
Then, recalling that \( s_p = f(u_p)/u_p = -\gamma(x^*-u_p)/u_p \) and \( u_p, x_r \to x^* \in (0, \infty) \),
\[
\varphi_{p, \gamma}(z_p) \frac{\sqrt{k(x_r - \bar{x}_r)}}{x_r} = O\left(s_p^{-1}\sqrt{k(x_r - \bar{x}_r)}\right) = O\left(\sqrt{k} \frac{x_r - \bar{x}_r}{x^* - u_p}\right) = o(1).
\]

For the easiest case \( \gamma \leq -1 \), using a Taylor expansion
\[
\sqrt{k}|z_p^-\gamma - x^-\gamma| \leq -\gamma \max\{x^-\gamma-1, z_p^-\gamma-1\} \cdot \sqrt{k}|z_p - x| \to 0,
\]
because \( x, z_p \to 0 \) and, by Assumption 2.5, \( \sqrt{k}|z_p - x| \to 0 \). This leads to (44) again and the rest is the same as for the case \(-1 < \gamma < 0\). The proof of (41) is now complete.

Hence, it suffices to show the theorem by replacing \( x_r \) with \( \bar{x}_r \). Decompose that
\[
\varphi_{p, \gamma}(z_p) \frac{\sqrt{k}(\bar{x}_r - \bar{x}_r)}{x_r} = \omega_{\gamma}(z_p) \frac{\sqrt{k}(\bar{x}_r - \bar{x}_r)}{x_r} = \omega_{\gamma}(z_p) \cdot \sqrt{k} \frac{X_{p-k:p} - u_p}{f(u_p)} + \omega_{\gamma}(z_p) \cdot \sqrt{k} \frac{X_{p-k:p} - u_p}{f(u_p)}
\]
\[
+ \omega_{\gamma}(z_p) \cdot \sqrt{k} \left( \frac{\hat{s}_p}{s_p} - 1 \right) + \frac{X_{p-k:p}}{u_p} \cdot \omega_{\gamma}(z_p) \sqrt{k} (\phi_{\gamma}(z_p) - \phi(z_p))
\]
\[
= : I + II + III + IV.
\]

As \( z_p \downarrow 0, w_{\gamma}(z_p) \to w_{\gamma}(0) = \gamma^2, \bar{w}_{\gamma}(z_p) \to \bar{w}_{\gamma}(0) = -\gamma_- \). Together with Corollaries 2.2 and 2.4, we immediately have
\[
\begin{align*}
I & \xrightarrow{a.s.} \gamma^2 \mathbb{V}(1), & II & \xrightarrow{p} -\gamma_- \cdot 1 \cdot \gamma_+ \cdot \mathbb{V}(1) = 0, & III & \xrightarrow{p} -\gamma_- \cdot \Sigma
\end{align*}
\]
where \( \gamma_- \cdot \gamma_+ = 0 \), and
\[
IV = (1 + o_p(1)) \cdot (1 + o_p(1)) \cdot \omega_{\gamma}(z_p) \sqrt{k} (\phi_{\gamma}(z_p) - \phi(z_p)) \xrightarrow{p} 1 \cdot 1 \cdot \Gamma.
\]
if we can show that
\[
\omega_{\gamma}(z_p) \sqrt{k} (\phi_{\gamma}(z_p) - \phi(z_p)) \xrightarrow{p} \Gamma.
\] (45)

In what follows, we shall show (45) in two different cases with \( \gamma \geq 0 \) and \( \gamma < 0 \) respectively. We begin with the easier case \( \gamma < 0 \), in which
\[
\omega_{\gamma}(z_p) \to \omega_{\gamma}(0) = \gamma^2 > 0,
\]
meaning that we only need to show
\[
\gamma^2 \sqrt{k} (\phi_{\gamma}(z_p) - \phi(z_p)) = \left[1 - z_p^{-\gamma}\right] \frac{\gamma}{\gamma} \sqrt{k}(\gamma - \gamma) + \gamma \cdot \sqrt{k}[z_p^{-\gamma} - z_p^{-\gamma}] \xrightarrow{p} \Gamma. (46)
\]
By the consistency of the moment estimator, \( -\hat{\gamma} \xrightarrow{P} -\gamma > 0 \), implying that \( \mathbb{P}(-\hat{\gamma} > -\gamma/2 > 0) \to 1 \). But conditioning on the event \( -\hat{\gamma} > -\gamma/2 > 0 \), \( z_p^{-\hat{\gamma}} \leq z_p^{-\gamma/2} \) and by using Taylor expansion, for all large \( p \) and small \( z_p \),

\[
|z_p^{-\hat{\gamma}} - z_p^{-\gamma}| \leq z_p^{-\gamma/4} |\hat{\gamma} - \gamma|.
\]

Hence, \( z_p^{-\hat{\gamma}} = O_p\left(z_p^{-\gamma/4}\right) \xrightarrow{p} 0 \) and

\[
\sqrt{k}[z_p^{-\hat{\gamma}} - z_p^{-\gamma}] = O_p(z_p^{-\gamma/4}) \cdot O_p(\sqrt{k}(\hat{\gamma} - \gamma)) = o_p(1) \cdot O_p(1) \xrightarrow{p} 0,
\]

and (46) follows.

Now consider the case \( \gamma \geq 0 \) and \( \log(p\tau) = o(\sqrt{k}) \). Rewrite the left-hand-side of (45) by

\[
\omega_\gamma(z_p)\sqrt{k}(\phi_\gamma(z_p) - \phi_\gamma(z_p)) = \omega_\gamma(z_p) \cdot \int_{z_p}^{1} \frac{\exp((\gamma - \hat{\gamma})\log s) - 1}{(\gamma - \hat{\gamma})\log s} \log s \phi_\gamma(s) \cdot \sqrt{k}(\hat{\gamma} - \gamma).
\]

Note that uniformly for \( z_p \leq s \leq 1 \),

\[
|(\gamma - \hat{\gamma})\log s| \leq \sqrt{k}(\hat{\gamma} - \gamma) \cdot \frac{\log z_p}{\sqrt{k}} = O_p(1) \cdot \frac{\log(p\tau)}{\sqrt{k}} \xrightarrow{k} 0.
\]

Then using the limit \( \frac{e^y - 1}{y} \to 1 \) for all \( y = (\gamma - \hat{\gamma})\log s \) with \( s \in [z_p, 1] \), we can show that

\[
\omega_\gamma(z_p)\sqrt{k}(\phi_\gamma(z_p) - \phi_\gamma(z_p)) = \omega_\gamma(z_p) \cdot \omega_\gamma^{-1}(z_p) \cdot (1 + o_p(1)) \cdot \sqrt{k}(\hat{\gamma} - \gamma)
\]

and (45) follows.

7.2. Proof of Lemma 2.2

Let \( \tau_p = \frac{p\tau}{k} \). Note that

\[
\frac{\hat{\phi}_{p,\hat{\gamma}}(p\tau/k)}{\varphi_{p,\gamma}(p\tau/k)} = \frac{\omega_\gamma(z_p)}{\hat{\omega}_\gamma(z_p)} \left( \frac{\hat{s}_p}{s_p} \right)^{-1} \cdot \frac{1 + \hat{s}_p\phi_\gamma(z_p)}{1 + s_p\phi_\gamma(z_p)}.
\]

Recall that \( \hat{s}_p/s_p \xrightarrow{P} 1 \) from Corollary 2.4. Furthermore \( s_p \to 0, \hat{\gamma} \xrightarrow{P} \gamma \leq 0 \) and \( z_p \to 0, 1 + \hat{s}_p\phi_\gamma(z_p) = 1 + \hat{s}_p/s_p \cdot s_p\phi_\gamma(z_p) \to 1 + 1 \cdot 0 \cdot (-1/\gamma) = 1 \) and \( 1 + s_p\phi_\gamma(z_p) \to 1 + 0 \cdot (-1/\gamma) = 1 \). Hence, it suffices to show that

\[
\frac{\hat{\omega}_\gamma(z_p)}{\omega_\gamma(z_p)} \xrightarrow{P} 1.
\]

We only prove the case for \( \gamma < 0 \); the case for \( \gamma \geq 0 \) follows from Corollary 4.3.2 in de Haan and Ferreira (2006) by noting their \( q_\gamma(z) = 1/\omega(1/z) \). Since \( \omega_\gamma(z_p) \to \omega_\gamma(0) = \gamma < 0 \) as \( z_p \to 0 \), it suffices to show

\[
(\omega_\gamma(z_p))^{-1} - (\omega_\gamma(z_p))^{-1} \xrightarrow{p} 0.
\]
or equivalently
\[ - \int_{z_p}^{1} s^{-\gamma^{-1}} \log s ds + \int_{z_p}^{1} s^{-\gamma} \log s ds \leq \int_{0}^{1} |s^{-\gamma^{-1}} - s^{-\gamma}| (-\log s) ds \xrightarrow{p} 0. \]

But using Taylor expansion, for all \(0 < s \leq 1\)
\[ |s^{-\gamma^{-1}} - s^{-\gamma}| \leq \max\{s^{-\gamma^{-1}}, s^{-\gamma}\} \log s \cdot (\hat{\gamma} - \gamma). \]

Hence, as \(\hat{\gamma} \xrightarrow{p} \gamma < 0\),
\[ \int_{0}^{1} |s^{-\gamma^{-1}} - s^{-\gamma}| (-\log s) ds \leq |\gamma - \hat{\gamma}| \cdot \max\left\{ \int_{0}^{1} s^{-\gamma^{-1}} (-\log s)^2 ds, \int_{0}^{1} s^{-\gamma} (-\log s)^2 ds \right\} \]
\[ = |\gamma - \hat{\gamma}| \cdot \max\{-2\hat{\gamma}^{-3}, -2\gamma^{-3}\} \xrightarrow{p} 0 \cdot (-2\gamma^{-3}) = 0, \]
where the first equality in the last line holds with probability tending to 1 with \(\hat{\gamma} < 0\).

### 7.3. Proof of Theorem 3.1

First consider the array of embedded variables \(Y_i(p) = \sigma_p Z_i\), whose average distribution \(T_p^*\) satisfies Assumption 2.1 with an extreme value index \(\gamma^* = 1\) and the limit distribution \(T^*(x) = \int_{0}^{\infty} S(x/u) dF(u)\) by Lemma 7.9 of Einmahl and He (2023). Observe that the average distributions \(T_p\) and \(T_p^*\) and their limits are related via the following equations
\[ T_p(x) = T_p^*(1/T_0(x)), \quad T(x) = T^*(1/T_0(x)) \]

and thus Assumption 2.1 for \(X_i(p)\) follows. Combining the power law \(T^*(t) \sim ct^{-1}\), as \(t \to \infty\), and Assumption 3.1 yields (1): there exists some positive function \(f(t)\) such that
\[ \lim_{t \to \infty} \frac{T(t + xf(t))}{T(t)} = \lim_{t \to \infty} \frac{T_0(t + xf(t))}{T_0(t)} = (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0, \]
where the last step is due to Theorem B.2.21 in de Haan and Ferreira (2006).

Second, the stability condition, Assumption 2.2 with \(M = 1\), follows from the monotonicity of \(Q_\sigma\), because
\[ T_p(x) = \frac{1}{p} \sum_{i=1}^{p} S \left( \frac{1}{Q_\sigma(1 - i/p)T_0(x)} \right) \leq \int_{0}^{1} S \left( \frac{1}{Q_\sigma(1 - u)T_0(x)} \right) du = T(x). \]

It remains to show the existence of \(R\) and to verify its expression. By Lemma 2.1, we know the arrays \(X_i(p)\) and \(Y_i(p)\) share the same \(R\)-function, while the latter has the desired \(R\)-function by Theorem 3.1 of Einmahl and He (2023). Note that the proof therein for the \(R\)-function does not require the stability condition.
7.4. Proof of Theorem 3.2

The proof is analogous to that of Theorem 3.1, except using Theorem 3.2 instead of Theorem 3.1 from Einmahl and He (2023) when deriving the R-function. We omit the details.

References


Vervaat, W. (1972). Functional central limit theorems for processes with