Generalized Pickands Estimators for the Extreme Value Index

Johan Segers[†]

October 10, 2002

Abstract. The Pickands estimator for the extreme value index is generalized in a way that includes all of its previously known variants. A detailed study of the asymptotic behavior of the estimators in the family serves to determine its optimally performing members. These are given by simple, explicit formulas, have the same asymptotic variance as the maximum likelihood estimator in the generalized Pareto model, and are robust to departures from the limiting generalized Pareto model in case the convergence of the excess distribution to its limit is slow. A simulation study involving a wide range of distributions shows the new estimators to compare favorably with the maximum likelihood estimator.

AMS Subject Classification: primary $62\,\mathrm{G}\,32$; secondary $62\,\mathrm{G}\,05$, $62\,\mathrm{G}\,20$

Keywords: Pickands estimator; Extreme value index; Generalized Pareto distribution; Excess distribution; Second-order regular variation; Tail quantile function

1 Introduction

The extreme value index γ of a distribution function (d.f.) F on the real line determines the weight of the right tail of F. Pickands (1975) introduced a simple estimator of γ, which was investigated and improved upon by several authors. Here, all these variants are united into a single family of estimators. A study of their joint asymptotic behavior leads to simple estimation procedures with desirable properties: invariancy under changes of the data in location and scale; efficiency w.r.t. the maximum likelihood estimator in the generalized Pareto model for excesses over a high threshold; ability to cope with the typical bias problems arising from slow convergence in the extreme value asymptotics.

The extreme value d.f. with shape parameter $\gamma \in \mathbb{R}$ is defined by

$$G_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\}, & \text{for } \gamma \neq 0,\ 1 + \gamma x > 0, \\ \exp(-e^{-x}), & \text{for } \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

The d.f. F is said to belong to the domain of attraction of G_γ, notation $F \in D(G_\gamma)$, if the suitably centered and normalized maximum of an independent sample from F converges weakly to G_γ, that is, if there are sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n \to \infty} F^n(a_n x + b_n) = G_\gamma(x), \quad \text{for } x \in \mathbb{R}. \tag{1.1}$$

The generalized Pareto distribution (GPD) with shape parameter $\gamma \in \mathbb{R}$ is defined by

$$H_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma} & \text{for } \gamma \neq 0,\ x \geq 0,\ 1 + \gamma x > 0, \\ e^{-x} & \text{for } \gamma = 0,\ x \geq 0. \end{cases}$$

Let $x_+ = \sup\{x \in \mathbb{R} : F(x) < 1\}$ and for $u \in \mathbb{R}$ such that $F(u) < 1$ denote the excess d.f. $F_u(x) = P(X - u \leq x \mid X > u)$, for $x \geq 0$, where the random variable (r.v.) X has d.f. F. The fundamental result of Pickands (1975) was that $F \in D(G_\gamma)$ if and only if for $u < x_+$ there exists $\sigma(u) > 0$ such that

$$\lim_{u \uparrow x_+} F_u(\sigma(u)x) = H_\gamma(x), \quad \text{for } x \geq 0. \tag{1.2}$$

Equation (1.2) suggests to model the unknown excess d.f. F_u by the parametric family $H_\gamma(\cdot/\sigma)$, where $\gamma \in \mathbb{R}$ and $\sigma > 0$. The parameters γ and σ can be estimated from the excesses in a sample over the threshold u, and the fitted model $H_{\widehat{\gamma}}(\cdot/\widehat{\sigma})$ yields estimates for such tail quantities as the probability of exceeding a level larger than the sample maximum or the quantile corresponding to a return period longer than the time span covered by the data.

Statistically, then, the problem is how to estimate the parameters γ and σ. We restrict our attention to the shape parameter γ, since it is the more crucial one as far as extrapolation from the model is concerned: for $\gamma > 0$ the tails of the GPD are of Pareto-type, for $\gamma = 0$ the GPD coincides with the exponential distribution, and for $\gamma < 0$ the GPD has a finite right endpoint.

Maximum likelihood estimation, considered by Smith (1987), has many advantages, such as efficiency, invariancy under changes of the data in location and scale (imagine converting temperature data from degrees centigrade to Fahrenheit), and extendability to various regression models. Apart from the fact, however, that the estimates must be found by numerical optimization, the method is rather sensitive to discrepancies between F_u and the limiting GPD, which may occur if the rate of convergence in (1.2) is slow, examples being such textbook distributions as the Normal and the Student-t with many degrees of freedom.

Pickands (1975) already proposed a simple, location and scale invariant estimator of γ. Denote by $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ the ascending order statistics from an independent sample of size n from F. Pickands suggested to use

$$\widehat{\gamma}_{n,k}^{\mathrm{Pick}} = \frac{1}{\log 2} \log \left(\frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}} \right), \quad \text{for } k = 1, \ldots, \lfloor n/4 \rfloor,$$

where $\lfloor x \rfloor$ is the largest integer not exceeding the real number x. Dekkers and de Haan (1989) showed consistency and asymptotically normality for all γ. Unfortunately, the estimator is quite volatile as a function of k, and its asymptotic variance is large. Ways of improvement were therefore sought after and discovered by many authors, all of whom realized somehow that the Pickands estimator is a linear combination of log-spacings of order statistics, see Pereira (1994), Falk (1994), Alves (1995), Drees (1995), Yun (2000) and Yun (2002).

The very same idea to form linear combinations of log-spacings of order statistics is the starting point in this paper for a large family of estimators of the extreme value index. Not only does this family include the Pickands estimator and all its variants, certain of its members, hitherto unknown, are as efficient as the maximum likelihood estimator and, moreover, can cope better with the case $\gamma < 0$ and with departures of the excess distribution F_u from the GPD family.

The generalized Pickands estimators make their appearance in Section 2. Section 3 focuses on weak consistency and asymptotic normality. Which members of the family to choose is the topic of Section 4, where the matter is discussed from a theoretical perspective, and of Section 5, where the theory is illustrated by a simulation study. All proofs are deferred to Appendices A and B.

2 Description of the Estimators

What are the generalized Pickands estimators and how do they relate to previously known variants of the Pickands estimator?

Let Λ be the collection of signed Borel measures λ on the interval $(0,1]$ such that

$$\lambda(0,1] = 0, \quad \int \log(1/t)|\lambda|(dt) < \infty, \quad \text{and} \quad \int \log(1/t)\lambda(dt) = 1. \tag{2.1}$$

If not mentioned explicitly, the domain of integration is always understood to be $(0,1]$. For $t \in [0,1]$ abbreviate $\lambda(t) = \lambda((0,t])$. Recall that $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ are the order

statistics of an independent sample from F. Let $\lceil x \rceil$ denote the smallest integer at least as large as $x \in \mathbb{R}$, and, for convenience, set $\log(0) := 0$.

Definition 2.1 *For $k = 1, \ldots, n-1$, $0 < c < 1$, and $\lambda \in \Lambda$, the generalized Pickands estimator $\widehat{\gamma}_{n,k}(c, \lambda)$ is given by*

$$\widehat{\gamma}_{n,k}(c, \lambda) = \int \log(X_{n-\lfloor c\lceil tk\rceil\rfloor,n} - X_{n-\lceil tk\rceil,n})\lambda(dt).$$

The generalized Pickands estimator is a linear combination of log-spacings of order statistics, the spacings being determined by c and k, and the weights by λ. In particular, since $\lceil tk \rceil = j$ if and only if $(j-1)/k < t \leq j/k$, we have

$$\widehat{\gamma}_{n,k}(c, \lambda) = \sum_{j=1}^{k}\{\lambda(j/k) - \lambda((j-1)/k)\}\log(X_{n-\lfloor cj\rfloor,n} - X_{n-j,n}).$$

Example 2.2 Let ε_x denote the point-measure giving mass 1 to x. For $0 < c < 1$, $0 < v < 1$, and $\lambda^{(v)} := (\varepsilon_1 - \varepsilon_v)/\log v$, we find

$$\widehat{\gamma}_{n,k}(c, \lambda^{(v)}) = \frac{1}{\log v}\log\left(\frac{X_{n-\lfloor ck\rfloor,n} - X_{n-k,n}}{X_{n-\lfloor c\lceil vk\rceil\rfloor,n} - X_{n-\lceil vk\rceil,n}}\right). \tag{2.2}$$

We recover Yun's (2001) estimator, special cases of which are the ones of Pickands (1975) $[c = v = 1/2]$, Pereira (1994) and Alves (1995) $[c = v]$, and Yun (2000) $[1/4 < c < 1$ and $v = (4c)^{-1}]$.

Example 2.3 More generally, let again $0 < v < 1$ and let μ be a probability measure on $(0, 1]$ such that $\int \log(1/t)\mu(dt) < \infty$. Define $\lambda^{(v,\mu)} \in \Lambda$ by

$$\int f(t)\lambda^{(v,\mu)}(dt) = \int \frac{f(t) - f(vt)}{\log v}\mu(dt),$$

for bounded measurable functions $f : (0, 1] \to \mathbb{R}$. We have

$$\widehat{\gamma}_{n,k}(c, \lambda^{(v,\mu)}) = \int \frac{1}{\log v}\log\left(\frac{X_{n-\lfloor c\lceil tk\rceil\rfloor,n} - X_{n-\lceil tk\rceil,n}}{X_{n-\lfloor c\lceil vtk\rceil\rfloor,n} - X_{n-\lceil vtk\rceil,n}}\right)\mu(dt),$$

which is a mixture of the $\widehat{\gamma}_{n,k}(c, \lambda^{(v)})$ of (2.2) over different values of k, encompassing the refined Pickands estimators of Drees (1995) and hence also Falk's (1994) convex combinations of Pickands estimators.

We can also form mixtures of the $\widehat{\gamma}_{n,k}(c, \lambda)$ over c, with λ possibly varying with c. The mixture is controlled by a measure ν, about which we assume the following.

Condition 2.4 *Let $0 < c_1 \leq c_2 < 1$. The signed measure ν on $[c_1, c_2] \times (0, 1]$ admits the desintegration $\nu(d(c, t)) = \mu(dc)\lambda_c(dt)$, where μ is a probability measure on $[c_1, c_2]$ and the measures $\lambda_c \in \Lambda$ for $c \in [c_1, c_2]$ are such that $\int_{[c_1,c_2]}\int_{(0,1]}\log(1/t)|\lambda_c|(dt)\mu(dc) < \infty$.*

3

For such ν, the estimator

$$\widehat{\gamma}_{n,k}(\nu) = \int_{[c_1,c_2]\times(0,1]} \log(X_{n-\lfloor c\lceil tk\rceil\rfloor,n} - X_{n-\lceil tk\rceil,n})\nu(d(c,t)) = \int_{[c_1,c_2]} \widehat{\gamma}_{n,k}(c,\lambda_c)\mu(dc) \quad (2.3)$$

is a μ-mixture of the estimators $\widehat{\gamma}_{n,k}(c,\lambda_c)$. If μ degenerates at some $c \in (0,1)$, we recover the special case $\widehat{\gamma}_{n,k}(c,\lambda)$.

3   Large Sample Properties

3.1   Weak Consistency

The mixture $\widehat{\gamma}_{n,k_n}(\nu)$ of (2.3) is, without side conditions, a weakly consistent estimator of the extreme value index γ. Let '$\xrightarrow{p}$' denote convergence in probability. A sequence of integers $\{k_n\}_{n=2}^{\infty}$ is called *intermediate* if $1 \le k_n \le n-1$, $k_n \to \infty$, and $k_n/n \to 0$ as $n \to \infty$.

Theorem 3.1 *Let $F \in D(G_\gamma)$ for some real γ and let ν be as in Condition 2.4. For every intermediate sequence k_n we have $\widehat{\gamma}_{n,k_n}(\nu) \xrightarrow{p} \gamma$ as $n \to \infty$.*

3.2   Second-Order Conditions

Before we can treat the results on asymptotic normality in the next paragraph, we need a small digression on second-order theory. The tail quantile function U of the d.f. F is given by

$$U(x) = \begin{cases} \inf\{y \in \mathbb{R} : F(y) \ge 1 - 1/x\}, & \text{for } x > 1, \\ 0, & \text{for } 0 < x \le 1. \end{cases}$$

The extreme value condition $F \in D(G_\gamma)$ is equivalent to the existence of a positive, measurable function a defined on a neighborhood of infinity such that for all $y > 0$

$$\lim_{x\to\infty} \frac{U(xy) - U(x)}{a(x)} = \frac{y^\gamma - 1}{\gamma} =: h_\gamma(y), \quad (3.1)$$

to be read as $\log(y)$ in case $\gamma = 0$ (de Haan, 1984). The convergence in (3.1) takes place locally uniformly in $y \in (0,\infty)$ and the auxiliary function a is necessarily regularly varying with index γ, notation $a \in \mathcal{R}_\gamma$, that is,

$$\lim_{x\to\infty} \frac{a(xy)}{a(x)} = y^\gamma, \quad \text{for } y > 0. \quad (3.2)$$

which also holds locally uniformly in $y \in (0,\infty)$, see Bingham et al. (1987), Theorem 1.5.2.

It will turn out in Section 4 that optimal performance within the generalized Pickands family requires that $c = c_n \uparrow 1$ as $n \to \infty$, for which we need to enforce (1.1) to density convergence. Moreover, we need to quantify the speed of convergence by imposing a precise rate. The following second-order regular variation condition on the tail quantile density captures these two features in one stroke.

4

Condition 3.2 *For some $x_0 > 1$ the tail quantile function U is absolutely continuous on $[x_0, \infty)$ with density u. There are $\gamma \in \mathbb{R}$, $\rho \leq 0$, $d \in \mathbb{R}$, and $A \in \mathcal{R}_\rho$ with $\lim_{x \to \infty} A(x) = 0$ such that, denoting $a(x) = x\,u(x)$, we have*

$$\lim_{x \to \infty} \frac{\log a(xy) - \log a(x) - \gamma \log y}{A(x)} = dh_\rho(y), \quad for\ y > 0. \tag{3.3}$$

Remark 3.3 Condition 3.2 implies that $a(x) = x\,u(x)$ is regularly varying with index γ, which, in turn, is necessary and sufficient for (1.1) to be strengthened to locally uniform convergence of densities, see Pickands (1986), Theorem 4.1.

Remark 3.4 Condition 3.2 actually imposes a rate of convergence in (3.1). Define

$$h(y; x) \;\; = \;\; \frac{U(xy) - U(x)}{a(x)} \quad \text{for } y > 0,\ x \geq x_0, \tag{3.4}$$

$$H_{\gamma,\rho}(y) \;\; = \;\; \int_1^y w^{\gamma-1} h_\rho(w) dw = \frac{1}{\rho}\left(\frac{y^{\rho+\gamma} - 1}{\rho + \gamma} - \frac{y^\gamma - 1}{\gamma} \right), \quad \text{for } y > 0. \tag{3.5}$$

with the appropriate limits in case $\gamma = 0$, $\rho = 0$, or $\gamma + \rho = 0$. Condition 3.2 now implies

$$\lim_{x \to \infty} \frac{h(y; x) - h_\gamma(y)}{A(x)} = dH_{\gamma,\rho}(y), \quad \text{for } y > 0. \tag{3.6}$$

The latter fits within the framework of second-order generalized variation of second order (de Haan and Stadtmüller, 1996) and appears already in Pereira (1994) and Drees (1995).

The pace at which the intermediate sequence k_n tends to infinity should be in accordance with the rate of convergence in the extreme value condition as quantified by the rate function A.

Condition 3.5 *Let A and d be as in condition 3.2. The intermediate sequence k_n satisfies*

- *case $d \neq 0$: $\lim_{n \to \infty} k_n^{1/2} A(n/k_n)$ exists;*

- *case $d = 0$: $\sup_{n \geq 2} k_n^{1/2} A(n/k_n) < \infty$.*

In both cases, we denote $r := \lim_{n \to \infty} d k_n^{1/2} A(n/k_n)$.

3.3 Asymptotic Normality

First, we mention a useful approximation to the joint distribution of the upper order statistics of a sample. Let ξ_i, $i \geq 1$, be independent standard exponential r.v. with partial sums

$$S_j = \sum_{i=1}^{j} \xi_i, \quad \text{for } j = 1, 2, \ldots \tag{3.7}$$

5

Recall that $X_{n-k,n} \leq \cdots \leq X_{n,n}$ are the $k+1$ largest observations of an independent sample from a d.f. F with tail quantile function U. For any intermediate sequence k_n, we have

$$\left\| \mathcal{L}\left((X_{n-j,n})_{j=0}^{k_n} \right) - \mathcal{L}\left((U(n/S_{j+1}))_{j=0}^{k_n} \right) \right\| = O(k_n/n), \quad n \to \infty, \qquad (3.8)$$

where $\mathcal{L}(\cdot)$ denotes the law of a random vector and $\|\cdot - \cdot\|$ is the variational distance of two distributions (Theorem 5.4.3 of Reiss, 1989).

Recall the class Λ in (2.1). For $0 < c < 1$ and $\lambda \in \Lambda$, let

$$\widetilde{\gamma}_{n,k_n}(c, \lambda) = \int \log\{U(n/S_{\lfloor cj \rfloor+1}) - U(n/S_{j+1})\}\lambda(dt), \qquad (3.9)$$

with $j = \lceil tk_n \rceil$ and S_j as in (3.7). Abbreviate 'with probability 1' to 'wp1'. Recall the function $H_{\gamma,\rho}$ defined in (3.5). For $0 < c < 1$, let the difference operator Δ_c act on functions $f : (0, 1] \to \mathbb{R}$ through

$$\Delta_c f(t) := f(ct) - f(t), \quad \text{for } 0 < t \leq 1.$$

Theorem 3.6 *Assume Conditions 3.2 and 3.5. Let $\Lambda' \subset \Lambda$ be such that for some $0 < \varepsilon < 1/2$ the integral $\int t^{-1/2-\varepsilon}|\lambda|(dt)$ is uniformly bounded over Λ'. Let $0 < c_0 \leq c_n < 1$ be such that $k_n^\eta(1 - c_n) \to \infty$ for some $0 < \eta < \varepsilon/(2 + 4\varepsilon)$. On a suitable probability space, there exist r.v. $\{S_j, j \geq 1\}$ as in (3.7) and a standard Wiener process W such that wp1*

$$\sup_{c_0 \leq c \leq c_n, \, \lambda \in \Lambda'} \left| k_n^{1/2}\{\widetilde{\gamma}_{n,k_n}(c, \lambda) - \gamma\} - Z_n(c, \gamma, \lambda) - rB(c, \gamma, \rho, \lambda) \right| \to 0,$$

as $n \to \infty$, where

$$Z_n(c, \gamma, \lambda) = \int \frac{t^\gamma \Delta_c\{t^{-\gamma-1}W_n(t)\}}{h_\gamma(1/c)}\lambda(dt),$$

$$B(c, \gamma, \rho, \lambda) = \int \frac{t^\gamma \Delta_c H_{\gamma,\rho}(1/t)}{h_\gamma(1/c)}\lambda(dt).$$

and $W_n(t) = -k_n^{-1/2}W(k_n t)$ is a standard Wiener process as well.

Observe that for a standard Wiener process W, numbers $0 < c < 1$, $\gamma \in \mathbb{R}$, and a measure $\lambda \in \Lambda$ such that $\int t^{-1/2-\varepsilon}|\lambda|(dt) < \infty$ for some $\varepsilon > 0$, the r.v.

$$Z(c, \gamma, \lambda) = \int \frac{t^\gamma \Delta_c\{t^{-\gamma-1}W(t)\}}{h_\gamma(1/c)}\lambda(dt)$$

is mean-zero normally distributed with variance

$$v(c, \gamma, \lambda) = \int\int \sigma_{c,\gamma}(s, t)\lambda(ds)\lambda(dt),$$

6

where

$$\sigma_{c,\gamma}(s,t) = E\left[\frac{s^\gamma \Delta_c\{s^{-\gamma-1}W(s)\}}{h_\gamma(1/c)}\frac{t^\gamma \Delta_c\{t^{-\gamma-1}W(t)\}}{h_\gamma(1/c)}\right] \tag{3.10}$$

$$= \frac{1}{stc^{\gamma+1}h_\gamma^2(1/c)}[(c^{-\gamma}+c^{\gamma+1})(s\wedge t) - s\wedge(ct) - (cs)\wedge t].$$

Theorem 3.6 and Reiss's approximation (3.8) lead to asymptotic normality of the generalized Pickands estimators. Let the '$\Rightarrow$' denote convergence in distribution.

Corollary 3.7 *Assume Conditions 3.2 and 3.5. If $0 < c < 1$ and if $\lambda \in \Lambda$ is such that $\int t^{-1/2-\varepsilon}|\lambda|(dt) < \infty$ for some ε, then*

$$k_n^{1/2}\{\widehat{\gamma}_{n,k_n}(c,\lambda) - \gamma\} \Rightarrow N(rB(c,\gamma,\rho,\lambda), v(c,\gamma,\lambda)).$$

Corollary 3.7 can be extended to asymptotic normality of the vector $(k_n^{1/2}\{\widehat{\gamma}_{n,k_n}(c_i,\lambda_i) - \gamma\})_{i=1}^m$, with asymptotic covariance matrix computable from Theorem 3.6. In the same way, we can obtain asymptotic normality for the mixture $\widehat{\gamma}_{n,k}(\nu)$, with ν as in Condition 2.4 being such that for some $\varepsilon > 0$ we have $\int_{[c_1,c_2]}\int_{(0,1]} t^{-1/2-\varepsilon}|\lambda_c|(dt)\mu(dc) < \infty$. We skip the details, however, since well performing estimation procedures can be based upon the $\widehat{\gamma}_{n,k_n}(c,\lambda)$ alone.

4 Optimal members and adaptive procedures

Which member to choose within the huge family of generalized Pickands estimators? We will determine the measure $\lambda_{c,\gamma}$ that minimizes the asymptotic variance $v(c,\gamma,\lambda)$ of $\widehat{\gamma}_{n,k_n}(c,\lambda)$, and we will also determine the measure $\lambda_{c,\gamma,\rho}$ that minimizes the asymptotic variance under the constraint that the asymptotic bias vanishes, that is, $B(c,\gamma,\rho,\lambda) = 0$. In both cases, it is optimal to let $c = c_n \uparrow 1$. The measures $\lambda_{c,\gamma}$ and $\lambda_{c,\gamma,\rho}$ depend on the unknown parameters γ and ρ, which in practice have to be replaced by consistent estimators.

Let Λ_0 be the linear space of all signed Borel measures on the interval $(0,1]$ such that $\lambda(0,1] = 0$ and $\int t^{-1/2-\varepsilon}|\lambda|(dt) < \infty$ for some $\varepsilon > 0$. For $0 < c < 1$, $\delta > -1/2$, and $\rho \le 0$, define the signed measure $\lambda_{\delta,\rho}^c$ on $(0,1]$ through its 'distribution function'

$$\lambda_{\delta,\rho}^c(0,t] = (1-c^{1+\delta})t^{1-\rho}\sum_{i=0}^{j-1}c^{(\delta+\rho)i}, \quad \text{for } t \in [c^j, c^{j-1}) \text{ and } j = 1, 2, \dots$$

$$= \begin{cases} (1-c^{1+\delta})t^{1-\rho}\dfrac{1-c^{(\delta+\rho)j}}{1-c^{\delta+\rho}} & \text{if } \delta + \rho \ne 0, \\ (1-c^{1+\delta})t^{1-\rho} & \text{if } \delta + \rho = 0. \end{cases}$$

The positive part of the measure $\lambda_{\rho,\delta}^c$ is absolutely continuous w.r.t. Lebesgue measure, while its negative part puts point masses at each c^j for $j = 0, 1, 2, \dots$

7

Theorem 4.1 *Let $0 < c < 1$, $\gamma \in \mathbb{R} \setminus \{-1/2\}$, and $\rho < 0$. Set $\delta = |\gamma + 1/2| - 1/2$.*

(i) *The measure $\lambda_{c,\gamma} = \lambda_{\delta,0}^c$ minimizes $v(c, \gamma, \lambda)$ over $\lambda \in \Lambda \cap \Lambda_0$, and the minimal value is*

$$V(c, \gamma) = \begin{cases} \dfrac{\gamma^2(1 - c^{1+\gamma})^2}{c(1 - c^\gamma)^2}, & \text{for } \gamma > -1/2 \text{ and } \gamma \neq 0, \\[2ex] \dfrac{(1 - c)^2}{c(\log c)^2}, & \text{for } \gamma = 0, \\[2ex] \gamma^2, & \text{for } \gamma < -1/2. \end{cases}$$

(ii) *The measure*

$$\lambda_{c,\gamma,\rho} = \frac{(1 - \rho)^2}{\rho^2}\lambda_{c,\gamma} - \frac{(1 - \rho)(1 - 2\rho)}{\rho^2}\lambda_{\delta,\rho}^c$$

minimizes $v(c, \gamma, \lambda)$ over $\lambda \in \Lambda \cap \Lambda_0$ under the constraint $B(c, \gamma, \rho, \lambda) = 0$, and the minimal value is $V(c, \gamma, \rho) = (1 - 1/\rho)^2 V(c, \gamma)$.

For a fixed value of $\gamma \neq -1/2$, the function $V(c, \gamma)$ is decreasing in c, and the minimal asymptotic variance is

$$V(\gamma) = V(1, \gamma) = \lim_{c\uparrow 1} V(c, \gamma) = \begin{cases} (1 + \gamma)^2, & \text{for } \gamma \geq -1/2, \\ \gamma^2, & \text{for } \gamma < -1/2. \end{cases} \tag{4.1}$$

Similarly, $\lim_{c\uparrow 1} V(c, \gamma, \rho) = (1 - 1/\rho)^2 V(1, \gamma)$ is the optimal asymptotic variance for asymptotically unbiased estimation. The case $\gamma = -1/2$ is an exception, because formally taking the limit as $\gamma \to -1/2$ results in a measure $\lambda_{c,-1/2}$ for which $\int t^{-1/2}|\lambda_{c,-1/2}|(dt) = \infty$, so that $Z(c, -1/2, \lambda_{c,-1/2})$ is not defined.

The problem that the optimal $\lambda_{c,\gamma}$ and $\lambda_{c,\gamma,\rho}$ depend on the unknown γ and ρ is solved by substituting them with weakly consistent estimators. Since, however, the second-order parameter ρ is difficult to estimate, we restrict the remainder of this Section to the unconstrained adaptive estimator $\widehat{\gamma}_{n,k}(c, \lambda_{c,\overline{\gamma}})$, where $\overline{\gamma}$ is some initial estimate for γ. The simulation study of Section 5 will illustrate the finite-sample properties of the constrained adaptive estimator $\widehat{\gamma}_{n,k}(c, \lambda_{c,\overline{\gamma},\overline{\rho}})$, where $\overline{\rho} < 0$ is fixed and may or may not be equal to the true ρ.

Theorem 4.2 *Assume Conditions 3.2 and 3.5 with $\gamma \neq -1/2$ and $r = 0$. Let $0 < \varepsilon < |\gamma + 1/2| \wedge 1/2$, $0 < \eta < \varepsilon/(2 + 4\varepsilon)$, and $0 < c_0 \leq c_n < 1$ such that $k_n^\eta(1 - c_n) \to \infty$ as $n \to \infty$. On the probability space of Theorem 3.6, if $\overline{\gamma}_n \xrightarrow{p} \gamma$, then*

$$\left| k_n^{1/2}\{\widetilde{\gamma}_{n,k_n}(c_n, \lambda_{c_n,\overline{\gamma}_n}) - \gamma\} - Z_n(c_n, \gamma, \lambda_{c_n,\gamma}) \right| \xrightarrow{p} 0,$$

and $Z_n(c_n, \gamma, \lambda_{c_n,\gamma}) \sim N(0, V(c_n, \gamma))$.

Corollary 4.3 *Under the assumptions of Theorem 4.2 and with $c_n \to c \in (0, 1]$ we have for each weakly consistent estimator $\overline{\gamma}_n = \overline{\gamma}_n\left((X_{n-j,n})_{j=0}^{k_n}\right)$ of γ*

$$k_n^{1/2}\{\widehat{\gamma}_{n,k_n}(c_n, \lambda_{c_n,\overline{\gamma}_n}) - \gamma\} \Rightarrow N(0, V(c, \gamma)).$$

The transition from the special version $\widetilde{\gamma}_{n,k_n}$ (Theorem 4.2) to the estimator $\widehat{\gamma}_{n,k_n}$ is accomplished by Reiss's approximation (3.8).

For $\gamma < -1/2$, we have $V(c, \gamma) = V(\gamma) = \gamma^2$, that is, the choice of c is asymptotically irrelevant. For $\gamma > -1/2$, we need to let $c_n \uparrow 1$ to obtain the minimal variance $V(\gamma) = (1 + \gamma)^2$. However, for c relatively far away from 1, the asymptotic variance $V(c, \gamma)$ is for practical purposes undistinguishable from the optimal one: for instance, for $\gamma \leq 2$ we have $V(0.75, \gamma) \leq 1.04 V(\gamma)$.

Remark 4.4 Drees (1998) developed a general theory for estimators of γ that are location and scale invariant and that can be written as Hadamard differentiable, continuous functionals of the empirical tail quantile function. Although the functional corresponding to the two-stage generalized Pickands estimators of Corollary 4.3 is not Hadamard differentiable, the similarity with Theorem 4.1 and Example 4.2 of Drees (1998) is striking: the asymptotic variances $V(\gamma)$ and $(1 - 1/\rho)^2 V(\gamma)$ are the minimal ones found by Drees (1998), so that it is reasonable to conjecture that his results are valid for in fact a larger class of estimators – see by the way the Remark after his Theorem 3.2.

Remark 4.5 How does the asymptotic variance of the estimator $\widehat{\gamma}_{n,k_n}(c, \lambda_{c,\gamma})$ compare with the ones of other estimators? The maximum likelihood estimator for γ in the GPD-model (Smith, 1987) has asymptotic variance $V(\gamma) = (1 + \gamma)^2$ if $\gamma > -1/2$. Drees's (1995) adaptive procedure based on mixtures of ordinary Pickands estimators has asymptotic variance $2^{-1}(\log 2)^{-2} V(1/2, \gamma) \doteq 1.04 \, V(1/2, \gamma)$. Finally, the moment estimator of Dekkers, Einmahl and de Haan (1989) is not invariant w.r.t. changes in location of the data. For $\gamma \geq 0$ its asymptotic variance is $1 + \gamma^2$, which is smaller than $V(\gamma)$; for $\gamma < 0$ it is a complicated expression which is appreciably larger than $V(\gamma)$.

Remark 4.6 An alternative to minimizing the asymptotic variance, with or without the constraint of zero asymptotic bias, would be minimizing the asymptotic mean square error. Such an approach is difficult to implement in practice, however, since the corresponding tuning parameters will not only depend on ρ but also on the second-order rate function A, which is difficult to estimate.

5   Simulations

A simulation study served to compare the following three estimators: (a) the unconstrained adaptive generalized Pickands estimator $\widehat{\gamma}_{n,k}(c, \lambda_{c,\overline{\gamma}})$ at $c = 0.75$ and with initial estimate $\overline{\gamma} = \widehat{\gamma}_{n,k}(c, \lambda_{c,0})$; (b) the constrained adaptive generalized Pickands estimator $\widehat{\gamma}_{n,k}(c, \lambda_{c,\overline{\gamma},\overline{\rho}})$ at $c = 0.75$ and $\overline{\rho} = -1$, and with initial estimate $\overline{\gamma} = \widehat{\gamma}_{n,k}(c, \lambda_{c,0,\overline{\rho}})$; (c) the maximum likelihood estimator in the GPD model fitted to the excesses over the $k + 1$ largest order statistic.

The distributions involved in the experiments (Table 1) were chosen to cover a wide range of possibilities for (γ, ρ). Observe that for the GPD we have $\rho = -\infty$, that, is Condition 3.2 holds for every $\rho \leq 0$. From each of the distributions, 500 samples of size

9

Table 1: Distributions used in simulation study.

	Distribution	γ	ρ
$\gamma > 0$	• GPD with $\gamma = 0.5$ and $\sigma = 1$	γ	$-\infty$
	• Fréchet $F(x) = \exp(-x^{-\alpha})$, $x > 0$, with $\alpha = 4$	$1/\alpha$	-1
	• Burr $F(x) = 1 - (x^{-\rho/\gamma} + 1)^{1/\rho}$, $x \geq 0$, with $(\gamma, \rho) \in \{(0.5, -1), (0.25, -0.5)\}$	γ	ρ
	• Student-t with $\nu \in \{10, 20\}$ degrees of freedom	$1/\nu$	$-2/\nu$
$\gamma = 0$	• Exponential $F(x) = 1 - e^{-x}$, $x \geq 0$	0	$-\infty$
	• Gumbel $F(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$	0	-1
	• Logistic $F(x) = 1 - 2/(1 + e^x)$, $x \geq 0$	0	-1
	• Weibull $F(x) = 1 - \exp(-x^\tau)$, $x \geq 0$, with $\tau = 0.5$	0	0
	• Normal	0	0
	• Lognormal	0	0
$\gamma < 0$	• GPD with $\gamma = -0.5$ and $\sigma = 1$	γ	$-\infty$
	• Extreme Value Weibull $F(x) = \exp(-|x|^\alpha)$, $x \leq 0$, with $\alpha = 4$	$-1/\alpha$	-1
	• Beta, density $(B(\alpha, \beta))^{-1} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$, with $(\alpha, \beta) \in \{(2, 2), (5, 5)\}$	$-1/\beta$	$-1/\beta$
	• Reversed Burr $F(x) = 1 - (|x|^{-\rho/\gamma} + 1)^{1/\rho}$, $x < 0$, with $\gamma \in \{-0.5, -0.25\}$ and $\rho = -0.5$	γ	ρ

$n = 500$ were generated and the estimates were computed at $k = 20, 30, \ldots, 250$. The median absolute errors of the estimates in case of distributions with $\gamma > 0$, $\gamma = 0$, and $\gamma < 0$ are shown in Figures 1, 2, and 3 respectively (at the back of the paper).

For ρ far away from zero (fast convergence of the excess distribution to the GPD) the unconstrained Pickands estimator performed better than the constrained one, while for ρ close to zero (slow convergence) the opposite was true. The unconstrained Pickands estimator performed in almost all cases at least as well as the maximum likelihood estimator.

Nearly always did the constrained Pickands estimator attain its best performance at the largest k considered, corresponding here to as much as 50% of the sample size. Moreover, its behavior as a function of k was remarkably stable, particularly for large k, which is good news for the practitioner, whose concerns which threshold to choose might thus become less urgent.

A   Proofs

We provide the technical arguments to the results of the paper. For the proof of Theorem 3.6, consult Appendix B.

Theorem 3.1 is an immediate consequence of Reiss's approximation (3.8) and the following property.

Proposition A.1 *Let $F \in D(G_\gamma)$ and let ν be as in condition 2.4. For every intermediate sequence k_n and every sequence $\{s_j\}_{j \geq 1}$ of positive numbers such that $\lim_{j \to \infty} s_j/j = 1$, we have*

$$\lim_{n \to \infty} \int_{[c_1,c_2]} \int_{(0,1]} \log\{U(n/s_{\lfloor c \lceil tk_n \rceil \rfloor + 1}) - U(n/s_{\lceil tk_n \rceil + 1})\} \lambda_c(dt) \mu(dc) = \gamma. \qquad (1.1)$$

Proof. By assumption $\lambda_c(0,1] = 0$ and $\int \log(1/t) \lambda_c(dt) = 1$ for every $c \in [c_1, c_2]$. Abbreviate $j = \lceil tk_n \rceil$ and let γ_n be the double integral on the left-hand side of (1.1). For a as in (3.1), we have

$$\begin{aligned}
\gamma_n - \gamma &= \int_{[c_1,c_2]} \int_{(0,1]} \log\left(\frac{t^\gamma a(n/s_{j+1})}{a(n/k_n)}\right) \lambda_c(dt)\mu(dc) \\
&\quad + \int_{[c_1,c_2]} \int_{(0,1]} \log\left(\frac{U(n/s_{\lfloor cj \rfloor + 1}) - U(n/s_{j+1})}{h_\gamma(1/c) a(n/s_{j+1})}\right) \lambda_c(dt)\mu(dc). \qquad (1.2)
\end{aligned}$$

Since $\lim_{n \to \infty} s_{\lceil tk_n \rceil + 1}/k_n = t$ for $0 < t \leq 1$ and $\lim_{j \to \infty} s_{\lfloor cj \rfloor + 1}/s_{j+1} = c$ for $c \in [c_1, c_2]$, the integrands of both integrals converge to zero due to the local uniformity in (3.1) and (3.2). It remains to bound them by integrable functions uniformly over n. Let

$$0 < x_\ell = \inf_{j \geq 1} s_j/j \leq \sup_{j \geq 1} s_j/j = x_u < \infty.$$

For every $t \in (0,1]$ and every n, we have

$$\frac{1}{2x_u} \leq \frac{k_n}{s_{\lceil tk_n \rceil + 1}} = \frac{k_n}{\lceil tk_n \rceil + 1} \frac{\lceil tk_n \rceil + 1}{s_{\lceil tk_n \rceil + 1}} \leq \frac{1}{tx_\ell}.$$

Moreover, for $\varepsilon > 0$, we can find $n(\varepsilon)$ such that for all $n \geq n(\varepsilon)$ and all $j = 1, \ldots, k_n$ the Potter bounds

$$e^{-\varepsilon} \min\left\{\left(\frac{k_n}{s_{j+1}}\right)^{\gamma+\varepsilon}, \left(\frac{k_n}{s_{j+1}}\right)^{\gamma-\varepsilon}\right\} \leq \frac{a(n/s_{j+1})}{a(n/k_n)} \leq e^\varepsilon \max\left\{\left(\frac{k_n}{s_{j+1}}\right)^{\gamma+\varepsilon}, \left(\frac{k_n}{s_{j+1}}\right)^{\gamma-\varepsilon}\right\}$$

hold, see Bingham et al. (1987), Theorem 1.5.6. Together, we see that there are positive constants A_1 and A_2 such that for all $n \geq n(\varepsilon)$ and all $t \in (0,1]$,

$$\left|\log\left(\frac{t^\gamma a(n/s_{j+1})}{a(n/k_n)}\right)\right| \leq A_1 + A_2 \log(1/t).$$

This is sufficient for the first term in (1.2). For the second term, the conditions on $\{s_j\}_{j \geq 1}$ and the monotonicity of U imply by virtue of the extreme value condition (3.1) that

$$\limsup_{n \to \infty} \sup_{0 < t \leq 1, \, c_0 \leq c \leq c_1} \left|\log\left(\frac{U(n/s_{\lfloor cj \rfloor + 1}) - U(n/s_{j+1})}{h_\gamma(1/c) a(n/s_{j+1})}\right)\right| < \infty.$$

$\square$

Next, let us turn attention to the proof of Theorem 4.1. The optimal measures $\lambda_{c,\gamma}$ and $\lambda_{c,\gamma,\rho}$ are elegantly characterized by a functional equation, solving which allows actual computation of the measures. The functional equation is derived in the Characterization Lemma A.2.

First, some notation needs to be introduced or recalled. For $0 < c < 1$, $\gamma \in \mathbb{R}$, $\lambda, \lambda' \in \Lambda_0$, $s, t \in (0, 1]$, and a standard Wiener process W, define

$$
\begin{aligned}
Y_{c,\gamma}(t) &= \frac{t^\gamma \Delta_c\{t^{-\gamma-1}W(t)\}}{h_\gamma(1/c)}, \\
\sigma_{c,\gamma}(s,t) &= E\left[Y_{c,\gamma}(s)Y_{c,\gamma}(t)\right], \\
Z(c,\gamma,\lambda) &= \int Y_{c,\gamma}(t)\lambda(dt), \\
\langle \lambda, \lambda' \rangle &= \langle \lambda, \lambda' \rangle_{c,\gamma} = E\left[Z(c,\gamma,\lambda)Z(c,\gamma,\lambda')\right] = \int\int \sigma_{c,\gamma}(s,t)\lambda(ds)\lambda'(dt), \\
\|\lambda\| &= \|\lambda\|_{c,\gamma} = (\langle \lambda, \lambda \rangle)^{1/2} = [\operatorname{Var}Z(c,\gamma,\lambda)]^{1/2}
\end{aligned}
$$

Observe that if $\lambda \in \Lambda \cap \Lambda_0$, then by Corollary 3.7,

$$
k_n^{1/2}\{\widehat{\gamma}_{n,k_n}(c,\lambda) - \gamma\} \Rightarrow N\left(rB(c,\gamma,\rho,\lambda), \|\lambda\|^2\right).
$$

So our task becomes: find $\lambda \in \Lambda \cap \Lambda_0$ that minimizes $\|\lambda\|$, and in case unbiased estimation is desired, under the constraint $B(c,\gamma,\rho,\lambda) = 0$. For $\lambda \in \Lambda_0$, abbreviate $\lambda(t) = \lambda(0,t]$ with $t > 0$, where λ is trivially extended over $(1,\infty)$.

Lemma A.2 (Characterization) *Let $0 < c < 1$, $\gamma \in \mathbb{R}$, and $\rho < 0$.*

(i) *If $\lambda \in \Lambda_0$ is such that $\int_0^1 \lambda(t)t^{-1}dt = 1$ and if for some $\alpha > 0$ we have*

$$
(c^{-\gamma} + c^{\gamma+1})\lambda(t) - \lambda(ct) - c\lambda(t/c) = \alpha t, \quad for\ 0 < t < 1, \tag{1.3}
$$

then $\lambda \in \Lambda$ and $\|\lambda\|^2 = c^{-\gamma-1}h_\gamma^{-2}(1/c)\alpha$ is minimal over $\Lambda \cap \Lambda_0$.

(ii) *If $\lambda \in \Lambda_0$ is such that $\int_0^1 \lambda(t)t^{-1}dt = 1$ and $\int_0^1 \lambda(t)t^{-\rho-1}dt = 0$, and if for some $\alpha_1 > 0$ and $\alpha_2 \in \mathbb{R}$ we have*

$$
(c^{-\gamma} + c^{\gamma+1})\lambda(t) - \lambda(ct) - c\lambda(t/c) = \alpha_1 t + \alpha_2 t^{1-\rho}, \quad for\ 0 < t < 1, \tag{1.4}
$$

then $\lambda \in \Lambda$, $B(c,\gamma,\rho,\lambda) = 0$, and $\|\lambda\|^2 = c^{-\gamma-1}h_\gamma^{-2}(1/c)\alpha_1$ is minimal over $\Lambda \cap \Lambda_0$ under the constraint $B(c,\gamma,\rho,\cdot) = 0$.

Proof. Let $\lambda \in \Lambda_0$. By Fubini's theorem, we find $\int \log(1/t)\lambda(dt) = \int_0^1 \lambda(t)t^{-1}dt$, so that $\lambda \in \Lambda$ if and only if the latter integral equals 1. And since

$$
\frac{d}{dt}\left[t^\gamma \Delta_c H_{\gamma,\rho}(1/t)\right] = -t^{-1-\rho}h_{\gamma+\rho}(1/c), \tag{1.5}
$$

12

[use the functional equation $\rho H_{\gamma,\rho} + h_\gamma = h_{\gamma+\rho}$], we have

$$B(c, \gamma, \rho, \lambda) = \int \frac{t^\gamma \Delta_c H_{\gamma,\rho}(1/t)}{h_\gamma 1/c} \lambda(dt) = \frac{h_{\gamma+\rho}(1/c)}{h_\gamma(1/c)} \int_0^1 \lambda(t) \frac{dt}{t^{1+\rho}}.$$

Therefore, $B(c, \gamma, \rho, \lambda) = 0$ if and only if the right-hand integral vanishes.

(i) Suppose that $\lambda \in \Lambda$ is such that for some $\beta_0 \in \mathbb{R}$ and $\beta_1 > 0$,

$$\int \sigma_{c,\gamma}(s, t) \lambda(ds) = \beta_0 + \beta_1 \log(1/t), \quad \text{for } 0 < t \leq 1. \tag{1.6}$$

Then for all $\lambda' \in \Lambda \cap \Lambda_0$, we have $\langle \lambda, \lambda' \rangle = \int \left(\int \sigma_{c,\gamma}(s, t) \lambda(ds) \right) \lambda'(dt) = \beta_1$. In particular, $\langle \lambda, \lambda' - \lambda \rangle = 0$, so that $\|\lambda'\|^2 = \|\lambda\|^2 + \|\lambda' - \lambda\|^2 \geq \|\lambda\|^2 = \beta_1$, that is, λ minimizes $\|\cdot\|^2$ over $\Lambda \cap \Lambda_0$. Hence, it is enough to show that (1.3) implies (1.6). Put

$$w(t) = \int \frac{1}{s \vee t} \lambda(ds) = \int_{t \wedge 1}^1 \lambda(s) \frac{ds}{s^2}, \quad \text{for } t > 0.$$

By formula (3.10) for $\sigma_{c,\gamma}(s, t)$, we see that

$$\int \sigma_{c,\gamma}(s, t) \lambda(ds) = \frac{1}{h_\gamma^2(1/c)} \left[(c^{-2\gamma-1} + 1) w(t) - c^{-\gamma} w(ct) - c^{-\gamma-1} w(t/c) \right].$$

Use $-t^2(d/dt)w(t) = \lambda(t)$ to see that under (1.3), the derivative of $\int \sigma_{c,\gamma}(s, t) \lambda(ds)$ equals $-\beta_1/t$ with $\alpha = c^{\gamma+1} h_\gamma^2(1/c) \beta_1$.

(ii) Suppose next that $\lambda \in \Lambda$ satisfies $B(c, \gamma, \rho, \lambda) = 0$ and is such that for some $\beta_0, \beta_1 \in \mathbb{R}$ and $\beta_2 > 0$,

$$\int \sigma_{c,\gamma}(s, t) \lambda(ds) = \beta_0 + \beta_1 t^\gamma \Delta_c \{H_{\gamma,\rho}(1/t)\} + \beta_2 \log(1/t), \quad \text{for } 0 < t \leq 1. \tag{1.7}$$

Then for all $\lambda' \in \Lambda \cap \Lambda_0$ for which $B(c, \gamma, \rho, \lambda') = 0$, we have $\langle \lambda, \lambda' \rangle = \beta_2$. Therefore, $\|\lambda\|^2 = \beta_2$ is minimal over $\Lambda \cap \Lambda_0$ under the constraint $B(c, \gamma, \rho, \cdot) = 0$. To obtain (1.7) from (1.4), proceed in the same way as in step (i). This time, $\alpha_1 = c^{\gamma+1} h_\gamma^2(1/c) \beta_2$. $\qquad\square$

Proof of Theorem 4.1. Use the following formulas to check that the proposed measures $\lambda_{c,\gamma}$ and $\lambda_{c,\gamma,\rho}$ satisfy the requirements of the characterization lemma A.2. For $0 < c < 1$, $\delta > -1/2$, $\rho \leq 0$, and $\varepsilon < (1/2 + \delta) \wedge (1/2 - \rho)$, we have

$$\int_0^1 \lambda_{\delta,\rho}^c(t) \frac{dt}{t} = \frac{1}{1-\rho},$$

$$\int_0^1 \lambda_{\delta,\rho}^c(t) \frac{dt}{t^{1+\rho}} = \frac{1 - c^{1+\delta}}{(1 - 2\rho)(1 - c^{1+\delta-\rho})},$$

$$\int_0^1 \lambda_{\delta,0}^c(t) \frac{dt}{t^{1+\rho}} = \frac{1}{(1 - \rho)(1 - c^{1+\delta-\rho})},$$

$$\int_{(0,1]} t^{-1/2-\varepsilon} |\lambda_{\delta,\rho}^c|(dt) = \frac{1 - c^{1+\delta}}{1 - c^{1/2+\delta-\varepsilon}} \left(\frac{1-\rho}{1/2 - \varepsilon - \rho} + 1 \right); \tag{1.8}$$

13

if additionally $\delta = |\gamma + 1/2| - 1/2$, then for $0 < t < 1$,

$$(c^{\gamma} + c^{\gamma+1})\lambda^c_{\delta,\rho}(t) - \lambda^c_{\delta,\rho}(ct) - c\lambda^c_{\delta,\rho}(t/c) = (1 - c^{1+\delta})(c^{-\delta} - c^{1-\rho})t^{1-\rho}.$$

$\square$

Proof of Theorem 4.2. Let $\varepsilon_1 > 0$ be small enough, to be determined later, and let $\varepsilon_2 >$ be arbitrary. We have

$$P\left(\left|k_n^{1/2}\{\widetilde{\gamma}_{n,k_n}(c_n, \lambda_{c_n,\overline{\gamma}_n}) - \gamma\} - Z_n(c_n, \gamma, \lambda_{c_n,\gamma})\right| > \varepsilon_2\right) \le P_{n,1} + P_{n,2} + P_{n,3},$$

where $P_{n,1} = P(|\overline{\gamma}_n - \gamma| > \varepsilon_1)$, and $P_{n,2}$ and $P_{n,3}$ are respectively equal to

$$P\left(\sup_{|\gamma'-\gamma|\le\varepsilon_1} \left|k_n^{1/2}\{\widetilde{\gamma}_{n,k_n}(c_n, \lambda_{c_n,\gamma'}) - \gamma\} - Z_n(c_n, \gamma, \lambda_{c_n,\gamma'})\right| > \varepsilon_2/2\right),$$

$$P\left(\sup_{|\gamma'-\gamma|\le\varepsilon_1} |Z_n(c_n, \gamma, \lambda_{c_n,\gamma'}) - Z_n(c_n, \gamma, \lambda_{c_n,\gamma})| > \varepsilon_2/2\right).$$

The term $P_{n,1}$ converges to zero by weak consistency of $\overline{\gamma}_n$.

Regarding the term $P_{n,2}$, we want to apply Theorem 3.6 to the family

$$\Lambda' = \{\lambda_{c,\gamma'} :\ c \in [c_0, 1),\ |\gamma' - \gamma| \le \varepsilon_1\}.$$

This is allowed if $\varepsilon_1 > 0$ is small enough so that $\min_{|\gamma'-\gamma|\le\varepsilon_1} |\gamma' + 1/2| > \varepsilon$, for then (1.8) implies

$$\sup_{c_0\le c<1, |\gamma'-\gamma|\le\varepsilon_1} \int t^{-1/2-\varepsilon}|\lambda_{c,\gamma'}|(dt) < \infty.$$

The term $P_{n,3}$ remains. Define the signed Borel measure $\nu_{c,\gamma,\gamma'}$ on $(0,1]$ by

$$\int f(t)\nu_{c,\gamma,\gamma'}(dt) = \int \frac{t^{\gamma}\Delta_c\{t^{-\gamma}f(t)\}}{h_\gamma(1/c)}\lambda_{c,\gamma'}(dt),$$

for bounded measurable functions $f : (0,1] \to \mathbb{R}$. Observe that

$$Z_n(c, \gamma, \lambda_{c,\gamma'}) = \int \frac{t^{\gamma}\Delta_c\{t^{-\gamma-1}W_n(t)\}}{h_\gamma(1/c)}\lambda_{c,\gamma'}(dt) = \int \frac{W_n(t)}{t}\nu_{c,\gamma,\gamma'}(dt).$$

Hence we have

$$|Z_n(c_n, \gamma, \lambda_{c_n,\gamma'}) - Z_n(c_n, \gamma, \lambda_{c_n,\gamma})| \le \sup_{0<t\le 1} t^{-1/2+\varepsilon}|W_n(t)| \cdot \int t^{-1/2-\varepsilon}|\nu_{c,\gamma,\gamma'} - \nu_{c,\gamma,\gamma}|(dt).$$

By the equicontinuity property of Lemma A.3 below, we have $P_{n,3} \to 0$ if $\varepsilon_1 \downarrow 0$. $\square$

Lemma A.3 *For $0 < c < 1$, $\gamma \in \mathbb{R}$, $\gamma' \in \mathbb{R} \setminus \{-1/2\}$, and $\varepsilon < |\gamma' + 1/2| \wedge 1/2$, we have*

$$\lim_{\gamma''\to\gamma'} \sup_{c_0\le c<1} \int t^{-1/2-\varepsilon}|\nu_{c,\gamma,\gamma''} - \nu_{c,\gamma,\gamma'}|(dt) = 0.$$

The proof of Lemma A.3, although elementary, is long, involved and not very informative, so we omit it.

14

B   Proof of Theorem 3.6

Let $0 < c < 1$ and $\lambda \in \Lambda$, and abbreviate $j = \lceil tk_n \rceil$. We have

$$k_n^{1/2}\{\widehat{\gamma}_{n,k_n}(c,\lambda) - \gamma\} = \int k_n^{1/2} \log\left(\frac{t^\gamma(X_{n-\lfloor cj \rfloor,n} - X_{n-j,n})}{a(n/k_n)h_\gamma(1/c)}\right)\lambda(dt).$$

Theorem B.1 below presents a strong approximation of the integrand in the previous equation by a Gaussian process, but with the order statistics $X_{n-j,n}$ replaced by the $U(n/S_{j+1})$ of Reiss's approximation (3.8). Theorem 3.6 then follows as an immediate corollary to Theorem B.1.

Theorem B.1 *Assume conditions 3.2 and 3.5. Let $0 < \varepsilon < 1/2$ and let $0 < c_0 \leq c_n < 1$ be such that $k_n^\eta(1-c_n) \to \infty$ as $n \to \infty$ for some $0 < \eta < \varepsilon/(2+4\varepsilon)$. On a suitable probability space, there exist r.v. S_j $(j \geq 1)$ as in (3.7) and a standard Wiener process W such that wp1*

$$\sup_{0<t\leq 1,\, c_0\leq c\leq c_n} |f_n(t;c,\gamma) - g_n(t;c,\gamma,\rho,r)| \to 0, \quad \text{as } n \to \infty,$$

where, abbreviating $j = \lceil tk_n \rceil$,

$$
\begin{aligned}
f_n(t;c,\gamma) &= k_n^{1/2}\log\left(\frac{t^\gamma\{U(n/S_{\lfloor cj \rfloor+1}) - U(n/S_{j+1})\}}{a(n/k_n)h_\gamma(1/c)}\right), \\
g_n(t;c,\gamma,\rho,r) &= \frac{t^\gamma\Delta_c\{t^{-\gamma-1}W_n(t) + rH_{\gamma,\rho}(1/t)\}}{h_\gamma(1/c)},
\end{aligned}
$$

and $W_n(t) = -k_n^{-1/2}W(tk_n)$ is also a standard Wiener process.

The so-called 'Hungarian construction' of Komlós et al. (1975, 1976) sets the stage for the proof of Theorem B.1. On a suitable probability space, there exist independent standard exponential r.v.'s $\xi_1, \xi_2, \ldots$ and a standard Wiener process W such that the partial sums $S_j = \sum_{i=1}^j \xi_i$ satisfy wp1

$$\max_{1\leq j\leq k} |S_j - j - W(j)| = O(\log k), \quad k \to \infty. \tag{2.1}$$

The increments of W are controlled by

$$\sup_{s,t\in[0,T],|s-t|\leq 1} |W(s) - W(t)| = O\left((\log T)^{1/2}\right) \quad \text{wp1}, \quad T \to \infty, \tag{2.2}$$

(Csörgő and Horváth, 1993, Theorem A.1.1), and by the reflection principle, the magnitude of W is

$$\sup_{t\in[0,k]} |W(t)| = O\left((k\log k)^{1/2}\right) \quad \text{wp1}, \quad k \to \infty. \tag{2.3}$$

Thanks to the strong law of large numbers (SLLN), we have wp1

$$0 < x_\ell = \inf_{j \geq 1} S_j/j \leq \sup_{j \geq 1} S_j/j \leq x_u < \infty.$$

To study the function f_n of interest, decompose it as

$$f_n(t; c, \gamma) = k_n^{1/2}(\mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n + \mathrm{IV}_n),$$

where, with the convenient abbreviation $j = \lceil tk_n \rceil$,

$$
\begin{aligned}
\exp \mathrm{I}_n &= \frac{a(n/S_{j+1})}{a(n/k_n)} \left(\frac{S_{j+1}}{k_n} \right)^\gamma, \\
\exp \mathrm{II}_n &= \frac{h(S_{j+1}/S_{\lfloor cj \rfloor+1}; n/S_{j+1})}{h_\gamma(S_{j+1}/S_{\lfloor cj \rfloor+1})}, \\
\exp \mathrm{III}_n &= \left(\frac{tk_n}{S_{j+1}} \right)^\gamma, \\
\exp \mathrm{IV}_n &= \frac{h_\gamma(S_{j+1}/S_{\lfloor cj \rfloor+1})}{h_\gamma(1/c)};
\end{aligned}
$$

consult (3.4) for the definition of $h(y; x)$. The terms I_n and II_n will lead to a deterministic bias term, while the terms III_n and IV_n will lead to a mean-zero Gaussian term. Also, I_n and III_n depend on t, while II_n and IV_n depend on t and c. We will devote a proposition to each of the terms separately.

Proposition B.2 *For every $\varepsilon > 0$, we have wp1,*

$$\sup_{0 < t \leq 1} t^\varepsilon |k_n^{1/2} \mathrm{I}_n(t) - rh_\rho(1/t)| \to 0.$$

Proof. Let $m := \sup_{n \geq 2} k_n^{1/2} A(n/k_n) < \infty$. We have

$$t^\varepsilon |k_n^{1/2} \mathrm{I}_n(t) - rh_\rho(1/t)| \leq P_n(t) + Q_n(t) + R_n(t),$$

where

$$
\begin{aligned}
P_n(t) &= m \left(\frac{tk_n}{S_{j+1}} \right)^\varepsilon \left(\frac{S_{j+1}}{k_n} \right)^\varepsilon \left| \frac{\mathrm{I}_n(t)}{A(n/k_n)} - dh_\rho(k_n/S_{j+1}) \right|, \\
Q_n(t) &= m|d| t^\varepsilon \left| h_\rho(k_n/S_{j+1}) - h_\rho(1/t) \right|, \\
R_n(t) &= \left| dk_n^{1/2} A(n/k_n) - r \right| t^\varepsilon h_\rho(1/t).
\end{aligned}
$$

By the SLLN, we have $tk_n/S_{j+1} \leq (j+1)/S_{j+1} \leq 1/x_\ell$ for all $0 < t \leq 1$; moreover $k_n/S_{j+1} \geq j/S_{j+1} \geq 1/(2x_u) =: y_0$. Hence for all $0 < t \leq 1$,

$$
\begin{aligned}
P_n(t) &\leq mx_\ell^{-\varepsilon} \sup_{y \geq y_0} y^{-\varepsilon} \left| \frac{1}{A(n/k_n)} \log \left(\frac{a(yn/k_n)}{y^\gamma a(n/k_n)} \right) - dh_\rho(y) \right| \\
&\to 0, \quad n \to \infty,
\end{aligned}
$$

16

by Lemma B.3 below.

Next, let $1/k_n < t_n \leq 1$ be such that $t_n \downarrow 0$ and $t_n k_n \to \infty$ as $n \to \infty$, and observe that $h_\rho(k_n/S_{j+1}) - h_\rho(1/t) = t^{-\rho}h_\rho(tk_n/S_{j+1})$. By the Strong Law of Large Numbers, we have uniformly in $t_n \leq t \leq 1$,

$$t^{\varepsilon - \rho}|h_\rho(tk_n/S_{j+1})| \to 0.$$

For $1/k_n \leq t \leq t_n$, the ratio tk_n/S_{j+1} is bounded away from 0 and ∞, while $t^{\varepsilon - \rho} \leq t_n^{\varepsilon - \rho} \to 0$. For $0 < t \leq 1/k_n$, we have $j = \lceil tk_n \rceil = 1$ and

$$t^{\varepsilon - \rho}|h_\rho(tk_n/S_2)| = \frac{(tk_n)^\varepsilon}{k_n^{\varepsilon - \rho}}\,|h_\rho(1/S_2) - h_\rho(1/tk_n)| \to 0$$

uniformly. This deals with the term $Q_n(t)$.

Finally, since $\sup_{0 < t \leq 1} t^\varepsilon h_\rho(1/t) < \infty$, also $R_n(t) \to 0$ uniformly in t. □

Lemma B.3 *For a and A as in (3.3), we have for all $\varepsilon > 0$ and all $y_0 > 0$,*

$$\limsup_{\substack{x \to \infty \\ y \geq y_0}} y^{-\varepsilon}\left|\frac{1}{A(x)}\log\left(\frac{a(xy)}{y^\gamma a(x)}\right) - dh_\rho(y)\right| = 0.$$

Proof. For every $y_1 > y_0$, we have as $x \to \infty$,

$$\frac{\log a(xy) - \log a(x) - \gamma \log y}{A(x)} \to dh_\rho(y), \quad \text{uniformly in } y \in [y_0, y_1].$$

By Theorem 3.1.3 of Bingham et al. (1987), there exist positive constants K and x_1 large enough such that, with $\ell(x) = x^{-\gamma}a(x)$,

$$\left|\frac{\log \ell(xy) - \log \ell(x)}{A(x)}\right| \leq Ky^{\varepsilon/2}, \quad \text{for } x \geq x_1,\ y \geq 1.$$

Hence for $y \geq y_1 \geq 1$ and $x \geq x_1$, we have

$$y^{-\varepsilon}\left|\frac{1}{A(x)}\log\left(\frac{a(xy)}{y^\gamma a(x)}\right) - dh_\rho(y)\right| \leq Ky^{-\varepsilon/2} + dy^{-\varepsilon}h_\rho(y).$$

Since $h_\rho(y) \leq h_0(y) = \log y$, the result is obtained by letting $y_1 \to \infty$. □

Proposition B.4 *For every $\varepsilon > 0$ and $0 < c_0 < 1$, we have wp1,*

$$\sup_{0 < t \leq 1,\, c_0 \leq c < 1} t^\varepsilon \left|k_n^{1/2}\mathrm{II}_n(t) - rt^{-\rho}\frac{H_{\gamma,\rho}(1/c)}{h_\gamma(1/c)}\right| \to 0.$$

Proof. Suppose that $d \neq 0$; the proof for $d = 0$ is simpler. We have, abbreviating $j = \lceil tk_n \rceil$ and $S(j,c) = S_{j+1}/S_{\lfloor cj \rfloor + 1}$,

$$
\begin{aligned}
k_n^{1/2}\mathrm{II}_n(t,c) &= dk_n^{1/2}A(n/k_n)\cdot\frac{A(n/S_{j+1})}{A(n/k_n)}\cdot\frac{1}{dA(n/S_{j+1})}\log\left(\frac{h_\gamma(S(j,c);n/S_{j+1})}{h_\gamma(S(j,c))}\right)\\
&= P_n \cdot Q_n(t) \cdot R_n(t,c)
\end{aligned}
$$

17

say. By assumption, $\lim_{n\to\infty} P_n = r$. Next we claim that

$$Q_n(t) = t^{-\rho} + t^{-\varepsilon/2}\delta_n(t),$$

$$R_n(t,c) = \frac{H_{\gamma,\rho}(S(j,c))}{h_\gamma(S(j,c))} + \varepsilon_n(t,c),$$

$$\frac{H_{\gamma,\rho}(S(j,c))}{h_\gamma(S(j,c))} = \frac{H_{\gamma,\rho}(1/c)}{h_\gamma(1/c)} + t^{-\varepsilon/2}\eta_n(t,c),$$

where $\delta_n(t)$, $\varepsilon_n(t,c)$ and $\eta_n(t,c)$ converge to 0 as $n \to \infty$ uniformly in $0 < t \le 1$ and $c_0 \le c < 1$. Since $\rho \le 0$ and $1 < y \mapsto H_{\gamma,\rho}(y)/h_\gamma(y)$ is increasing and converges to 0 as $y \downarrow 1$, this claim is sufficient to prove the proposition.

First, let $0 < t_1 \le 1$. Since by the Strong Law of Large Numbers $k_n/S_{j+1} \to 1/t$ uniformly in $t_1 \le t \le 1$, the Uniform Convergence Theorem applied to $A \in \mathcal{R}_\rho$ gives

$$A(n/S_{j+1})/A(n/k_n) \to t^{-\rho} \quad \text{as } n \to \infty, \text{ uniformly in } t_1 \le t \le 1.$$

Also, $k_n/S_{j+1} > 1/(tx_u)$ for all n and all $0 < t \le 1$, and hence by Potter's Theorem, there exist a constant $C > 0$ and an integer n_0 such that

$$A(n/S_{j+1})/A(n/k_n) \le Ct^{-\rho-\varepsilon/4}, \quad \text{for all } n \ge n_0 \text{ and } 0 < t \le 1.$$

Hence we have

$$\limsup_{n\to\infty} \sup_{0<t\le 1} t^{\varepsilon/2}|A(n/S_{j+1})/A(n/k_n) - t^{-\rho}| \le \sup_{0<t\le t_1} t^{\varepsilon/2}(Ct^{\rho-\varepsilon/4} + t^{-\rho}).$$

But $\rho \le 0$, so letting $t_1 \downarrow 1$ gives the claim for $\delta_n(t)$.

The claim for $\varepsilon_n(t,c)$ follows directly from Lemma B.5 below and the fact that

$$1 \le \inf_{j\ge 1,\, c_0\le c\le 1} S(j,c) \le \sup_{j\ge 1,\, c_0\le c\le 1} S(j,c) < \infty$$

by the Strong Law of Large Numbers. And since $S(j,c) \to 1/c$ as $j \to \infty$ uniformly in $c_0 \le c \le 1$, the claim for $\eta_n(t,c)$ follows easily. $\qquad\square$

Lemma B.5 *Under condition 3.2 and with $h(y;x)$ as in (3.4), we have for every $y_0 > 1$,*

$$\frac{1}{A(x)}\log\left(\frac{h(y;x)}{h_\gamma(y)}\right) \to d\frac{H_{\gamma,\rho}(y)}{h_\gamma(y)} \quad \text{as } x \to \infty, \text{ uniformly in } 1 < y \le y_0.$$

Proof. From (3.4), we have

$$h(y;x) = \int_1^y \frac{a(xw)}{a(x)}\frac{dw}{w} = \int_1^y w^{\gamma-1}\{1 + A(x)[dh_\rho(w) + r(x,w)]\}dw,$$

where $r(x,w) \to 0$ as $x \to \infty$ uniformly in $1 \le w \le y_0$. But then

$$\frac{h(y;x)}{h_\gamma(y)} = 1 + A(x)\left[d\frac{H_{\gamma,\rho}(y)}{h_\gamma(y)} + R(x,y)\right]$$

where $R(x,y) \to 0$ as $x \to \infty$ uniformly in $1 < y \le y_0$. Since the function $1 < y \mapsto H_{\gamma,\rho}(y)/h_\gamma(y)$ is increasing and converges to 0 as $y \downarrow 1$, we can take a Taylor expansion of the logarithm around 1, proving the lemma. $\qquad\square$

Proposition B.6 *For every $\varepsilon > 0$, we have wp1,*

$$\sup_{0 < t \le 1} t^{1/2+\varepsilon} \left| k_n^{1/2} \mathrm{III}_n(t) - \gamma \frac{W_n(t)}{t} \right| \to 0.$$

Proof. Without loss of generality, we suppose that $\varepsilon < 1/2$. Rewrite the argument of the supremum as

$$k_n^{-\varepsilon} (tk_n)^{1/2+\varepsilon} |\gamma| \left| \log\left(\frac{S_{j+1}}{tk_n} \right) - \frac{W(tk_n)}{tk_n} \right|,$$

where $j = \lceil tk_n \rceil$. Since $k_n^{-\varepsilon} \to 0$, it is now clearly sufficient to prove that wp1,

$$\sup_{x > 0} x^{1/2+\varepsilon} \left| \log\left(\frac{S_{\lceil x \rceil + 1}}{x} \right) - \frac{W(x)}{x} \right| < \infty; \tag{2.4}$$

let us do so. Since $\sup_{0 < x \le 1} x^{-1/2+\varepsilon} |W(x)| < \infty$, the supremum over $0 < x \le 1$ in (2.4) is finite. Further, denoting $Z_j = S_j - j - W(j)$, we have

$$
\begin{aligned}
\frac{S_{\lceil x \rceil + 1}}{x} &= 1 + \frac{\lceil x \rceil + 1 - x}{x} + \frac{W(x)}{x} + \frac{W(\lceil x \rceil + 1) - W(x)}{x} + \frac{Z_{\lceil x \rceil + 1}}{x} \\
&= 1 + \frac{W(x)}{x} + O\left(x^{-1} (\log x)^{1/2} \right), \quad x \to \infty
\end{aligned}
$$

by (2.1) and (2.2). Further, $W(x)/x = O\left((x^{-1} \log x)^{1/2} \right)$ as $x \to \infty$ by (2.3), and so a Taylor expansion of the logarithm around 1 gives

$$\log\left(\frac{S_{\lceil x \rceil + 1}}{x} \right) = \frac{W(x)}{x} + O\left(x^{-1} \log x \right), \quad x \to \infty.$$

This completes the proof of (2.4) and thus of the proposition. $\qquad\square$

Proposition B.7 *Let $0 < \varepsilon < 1/2$ and $0 < c_0 \le c_n < 1$ be such that $k_n^\eta (1 - c_n) \to \infty$ as $n \to \infty$ for some $0 < \eta < \varepsilon/(2 + 4\varepsilon)$. Then wp1,*

$$\sup_{0 < t \le 1,\, c_0 \le c \le c_n} t^{1/2+\varepsilon} \left| k_n^{1/2} \mathrm{IV}_n(t, c) - \frac{1}{|h_\gamma(c)|} \left[\frac{W_n(ct)}{ct} - \frac{W_n(t)}{t} \right] \right| \to 0.$$

Proof. Let $4\eta < \nu < 2\varepsilon/(1 + 2\varepsilon)$ and $j_n = \lceil k_n^\nu \rceil$. Since

$$\frac{W_n(ct)}{ct} - \frac{W_n(t)}{t} = k_n^{1/2} \left[\frac{W(tk_n)}{tk_n} - \frac{W(ctk_n)}{ctk_n} \right],$$

our quantity of interest is bounded by the maximum of

$$\sup_{0 < t \le j_n/k_n,\, c_0 \le c \le c_n} t^{1/2+\varepsilon} k_n^{1/2} \left| \mathrm{IV}_n(t, c) \right|,$$

19

$$\sup_{0<t\le j_n/k_n,\, c_0\le c\le c_n} \frac{(tk_n)^{1/2+\varepsilon}}{k_n^\varepsilon |h_\gamma(c)|} \left| \frac{W(j+1)}{j+1} - \frac{W(\lfloor cj\rfloor+1)}{\lfloor cj\rfloor+1} \right|,$$

$$\sup_{j_n/k_n\le t\le 1,\, c_0\le c\le c_n} k_n^{-\varepsilon}(tk_n)^{1/2+\varepsilon} \left| \mathrm{IV}_n(t,c) - \frac{1}{|h_\gamma(c)|}\left[\frac{W(j+1)}{j+1} - \frac{W(\lfloor cj\rfloor+1)}{\lfloor cj\rfloor+1} \right] \right|,$$

$$\sup_{0<t\le 1,\, c_0\le c\le c_n} \frac{(tk_n)^{1/2+\varepsilon}}{k_n^\varepsilon |h_\gamma(c)|} \left| \left[\frac{W(j+1)}{j+1} - \frac{W(\lfloor cj\rfloor+1)}{\lfloor cj\rfloor+1} \right] - \left[\frac{W(tk_n)}{tk_n} - \frac{W(ctk_n)}{ctk_n} \right] \right|,$$

which we call $R_{n,1}$, $R_{n,2}$, $R_{n,3}$ and $R_{n,4}$ respectively. Every term requires a different approach.

Let $S(j,c) = S_{j+1}/S_{\lfloor cj\rfloor+1}$. First of all, we have

$$R_{n,1} \le \frac{j_n^{1/2+\varepsilon}}{k_n^\varepsilon} \sup_{c_0\le c\le c_n} \left\{ \max_{1\le j\le j_n} |\log h_\gamma(S(j,c))| + |\log h_\gamma(1/c)| \right\}.$$

Now on the one hand, $\sup\{S(j,c) : j\ge 1,\ c_0\le c<1\} < \infty$, while on the other hand $S(j,c) \ge 1 + \xi_{j+1}/S_j$ for all $0 < c < 1$. Hence as $n\to\infty$, the asymptotic equivalence $h_\gamma(x) \sim x - 1$ for $x\to 1$ implies

$$\max_{1\le j\le j_n} |\log h_\gamma(S(j,c))| = O\left(\max_{j=1,\dots,j_n} |\log(\xi_{j+1}/j)| \right).$$

However, by the Borel-Cantelli lemma, $P(\xi_j \le j^{-2}$ or $\xi_j \ge j$ infinitely often$) = 0$, and hence wp1,

$$\max_{j=1,\dots,j_n} |\log(\xi_{j+1}/j)| = O(\log j_n) = O(\log k_n), \quad n\to\infty.$$

The same order of magnitude is found in

$$\sup_{c_0\le c\le c_n} |\log h_\gamma(1/c)| = O(|\log(1-c_n)|) = O(\log k_n), \quad n\to\infty.$$

Together, we have wp1,

$$R_{n,1} = O\left(\frac{j_n^{1/2+\varepsilon}\log k_n}{k_n^\varepsilon} \right) = O\left(\frac{\log k_n}{k_n^{\varepsilon-\nu(1/2+\varepsilon)}} \right) \to 0, \quad n\to\infty,$$

by the choice of ν.

Secondly, the bound on the supremum of W given by (2.3) shows that wp1,

$$R_{n,2} = O\left(\frac{j_n^\varepsilon (\log j_n)^{1/2}}{k_n^\varepsilon (1-c_n)} \right), \quad n\to\infty.$$

However, since $\varepsilon(1-\nu) > \varepsilon/(1+2\varepsilon)$, we have $R_{n,2} \to 0$ by the choice of c_n.

Thirdly, the term $R_{n,3}$ is bounded by

$$\sup_{j\ge j_n,\, c_0\le c\le c_n} j^{1/2} \left| \log\left(\frac{h_\gamma(S(j,c))}{h_\gamma(1/c)} \right) - \frac{1}{|h_\gamma(c)|}\left[\frac{W(j+1)}{j+1} - \frac{W(\lfloor cj\rfloor+1)}{\lfloor cj\rfloor+1} \right] \right|.$$

20

Since
$$\frac{(1-c_n)^2 j_n^{1/2}}{\log j_n} \sim \frac{(1-c_n)^2 k_n^{\nu/2}}{\nu \log k_n} \to \infty, \quad n \to \infty,$$
our job for $R_{n,3}$ will be accomplished if we can show that whenever $0 < c_0 \le c_j' < 1$ is such that $(1-c_j')^2 j^{1/2}/\log j \to \infty$ as $j \to \infty$, then also wp1,
$$\sup_{c_0 \le c \le c_j'} j^{1/2} \left| \log\left(\frac{h_\gamma(S(j,c))}{h_\gamma(1/c)}\right) - \frac{1}{|h_\gamma(c)|}\left[\frac{W(j+1)}{j+1} - \frac{W(\lfloor cj \rfloor + 1)}{\lfloor cj \rfloor + 1}\right] \right| \to 0.$$

Now, the function h_γ satisfies the functional equations
$$h_\gamma(xy) = x^\gamma h_\gamma(y) + h_\gamma(x) \quad \text{and} \quad x^\gamma h_\gamma(1/x) = -h_\gamma(x), \quad \text{for } x, y > 0.$$

From these and from the fact that $h_\gamma(x) < 0$ for $0 < x < 1$, it follows that
$$\frac{h_\gamma(S(j,c))}{h_\gamma(1/c)} = 1 + \frac{h_\gamma(cS(j,c))}{|h_\gamma(c)|}.$$

Denote $S_j = j + W(j) + Z_j$. The properties of the Hungarian construction and the Wiener process W as expressed by (2.1), (2.2), and (2.3) yield the first-order asymptotic expansion
$$\begin{aligned}
cS(j,c) &= \frac{c(j+1)}{\lfloor cj \rfloor + 1} \cdot \frac{1 + (j+1)^{-1}W(j+1) + (j+1)^{-1}Z(j+1)}{1 + (\lfloor cj \rfloor + 1)^{-1}W(\lfloor cj \rfloor + 1) + (\lfloor cj \rfloor + 1)^{-1}Z(\lfloor cj \rfloor + 1)} \\
&= 1 + \frac{W(j+1)}{j+1} - \frac{W(\lfloor cj \rfloor + 1)}{\lfloor cj \rfloor + 1} + O(j^{-1}\log j), \quad j \to \infty,
\end{aligned}$$
uniformly in $c_0 \le c \le 1$ and wp1. The bound (2.3) and the Taylor expansion $h_\gamma(1+x) = x + O(x^2)$ as $x \to 0$ give
$$h_\gamma(cS(j,c)) = \frac{W(j+1)}{j+1} - \frac{W(\lfloor cj \rfloor + 1)}{\lfloor cj \rfloor + 1} + O(j^{-1}\log j), \quad j \to \infty,$$
uniformly in $c_0 \le c \le 1$. Repeat the argument for $\log(1+x)$ as $x \to 0$ to get
$$\log\left(\frac{h_\gamma(S(j,c))}{h_\gamma(1/c)}\right) = \frac{1}{|h_\gamma(c)|}\left[\frac{W(j+1)}{j+1} - \frac{W(\lfloor cj \rfloor + 1)}{\lfloor cj \rfloor + 1}\right] + O\left(\frac{\log j}{j(1-c_j')^2}\right), \quad j \to \infty,$$
uniformly in $c_0 \le c \le c_j'$. The remainder term is $o(j^{-1/2})$ by assumption.

Fourthly and finally, the term $R_{n,4}$ is not larger than
$$\frac{1}{k_n^\varepsilon |h_\gamma(c_n)|} \sup_{x>0,\, c \in [c_0,1]} x^{1/2+\varepsilon} \left| \left[\frac{W(\lceil x \rceil + 1)}{\lfloor x \rfloor + 1} - \frac{W(\lfloor c\lceil x \rceil \rfloor + 1)}{\lfloor c\lceil x \rceil \rfloor + 1}\right] - \left[\frac{W(x)}{x} - \frac{W(cx)}{cx}\right] \right|.$$

Again by assumption, $k_n^\varepsilon |h_\gamma(c_n)| \to \infty$, so it remains to show that the supremum is finite wp1. But the latter is readily reduced to the assertion
$$\sup_{x>0,\, c_0 \le c \le 1} x^{1/2+\varepsilon} \left| \frac{W(\lfloor c\lceil x \rceil \rfloor + 1)}{\lfloor c\lceil x \rceil \rfloor + 1} - \frac{W(cx)}{cx} \right| < \infty \quad \text{wp1}, \tag{2.5}$$

21

which we now prove. The supremum over $0 < x \le 1$ is finite since

$$\sup_{0 < x \le 1} x^{-1/2 + \varepsilon} |W(x)| < \infty \quad \text{wp1}.$$

For $x \ge 1$, the triangle inequality gives

$$\left| \frac{W(\lfloor c \lceil x \rceil \rfloor + 1)}{\lfloor c \lceil x \rceil \rfloor + 1} - \frac{W(cx)}{cx} \right|$$

$$\le |W(\lfloor c \lceil x \rceil \rfloor + 1)| \left| \frac{1}{\lfloor c \lceil x \rceil \rfloor + 1} - \frac{1}{cx} \right| + \frac{1}{cx} |W(\lfloor c \lceil x \rceil + 1 \rfloor) - W(cx)|.$$

Now the properties (2.2) and (2.3) of the sample paths of W entail as $x \to \infty$ and as always wp1

$$|W(\lfloor c \lceil x \rceil + 1 \rfloor)| = O\left((x \log x)^{1/2} \right),$$

$$\left| \frac{1}{\lfloor c \lceil x \rceil \rfloor + 1} - \frac{1}{cx} \right| = O(x^{-2}),$$

$$|W(\lfloor c \lceil x \rceil \rfloor + 1) - W(cx)| = O\left((\log x)^{1/2} \right),$$

uniformly in $c_0 \le c \le 1$. Assertion (2.5) follows. □

Proof of Theorem B.1. After the laborious estimates of the previous pages, the pieces of the puzzle fall together. Collect the approximations of Propositions B.2, B.4, B.6, and B.7 to get

$$\sup_{0 < t \le 1, \, c_0 \le c \le c_n} t^{1/2 + \varepsilon} |f_n(t; c, \gamma) - \overline{g}_n(t; c, \gamma, \rho, r)| \to 0 \quad \text{wp1},$$

where

$$\overline{g}_n(t; c, \gamma, \rho, r) = r \left(h_\rho(1/t) + t^{-\rho} \frac{H_{\gamma, \rho}(1/c)}{h_\gamma(1/c)} \right) + \gamma \frac{W_n(t)}{t} + \frac{1}{|h_\gamma(c)|} \left(\frac{W_n(ct)}{ct} - \frac{W_n(t)}{t} \right).$$

The functional relation

$$H_{\gamma, \rho}(xy) = x^{\gamma + \rho} H_{\gamma, \rho}(y) + H_{\gamma, \rho}(x) + x^\gamma h_\gamma(y) h_\rho(x), \quad \text{for } x, y > 0$$

confirms $\overline{g}_n = g_n$. □

References

Alves, M.I.F., 1995. Estimation of the tail parameter in the domain of attraction of an extremal distribution. J. Statist. Plann. Inference 45, 143–173.

Bingham, N.C., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. Cambridge University Press, Cambridge.

Csörgő, M., Horváth, L., 1993. Weighted approximations in probability and statistics. Wiley, Chichester.

Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833–1855.

Dekkers, A.L.M., de Haan, L., 1989. On the estimation of the extreme value index and large quantile estimation. Ann. Statist. 17, 1795–1832.

Drees, H., 1995. Refined Pickands estimators for the extreme value index. Ann. Statist. 23, 2059–2080.

Drees, H., 1998. On Smooth Statistical Tail Functionals. Scand. J. Statist. 25, 187–210.

Falk, M., 1994. Efficiency of convex combinations of Pickands' estimator of the extreme value index. J. Nonparametric Statist. 4, 133–147.

de Haan, L., 1984. Slow variation and characterization of domains of attraction. In: J. Tiago de Oliveira (Ed.) Statistical Extremes and Applications, Reidel, Dordrecht, 31–38.

de Haan, L., Stadtmüller, U., 1996. Generalized Regular Variation of Second Order. J. Austral. Math. Soc. (Series A) 61, 381–395.

Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent r.v.'s, and the sample d.f. (I). Probab. Theory Related Fields 32, 111–131.

Komlós, J., Major, P., Tusnády, G., 1976. An approximation of partial sums of independent r.v.'s, and the sample d.f. (II). Probab. Theory Related Fields 34, 33–58.

Pereira, T.T., 1994. Second order behaviour of domains of attraction and the bias of generalized Pickands' estimator. In: J. Galambos, J. Lechner and E. Simiu (Eds.) Extreme Value Theory and Applications III, NIST, 165–177.

Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.

Pickands, J., 1986. The continuous and differentiable domains of attraction of the extreme value distributions. Ann. Probab. 14, 996–1004.

Reiss, R.-D., 1989. Approximate distributions of order statistics. Springer, New York.

Smith, R.L., 1987. Estimating tails of probability distributions. Ann. Statist. 15, 1174–1207.

Yun, S., 2000. A class of Pickands-type estimators for the extreme value index. J. Statist. Plann. Inference 83, 113–124.

Yun, S., 2002. On a generalized Pickands estimator of the extreme value index. J. Statist. Plann. Inference 102, 389–409.
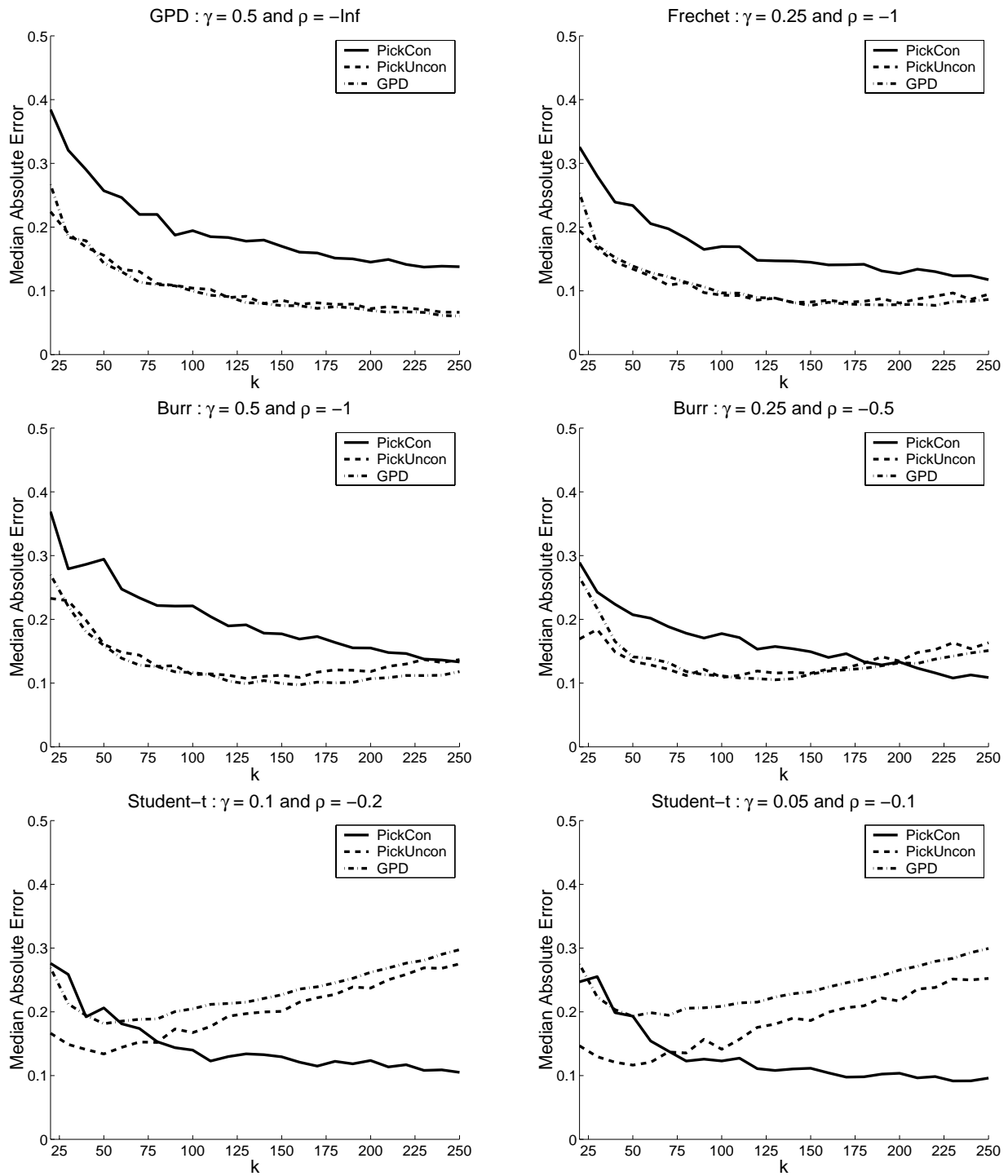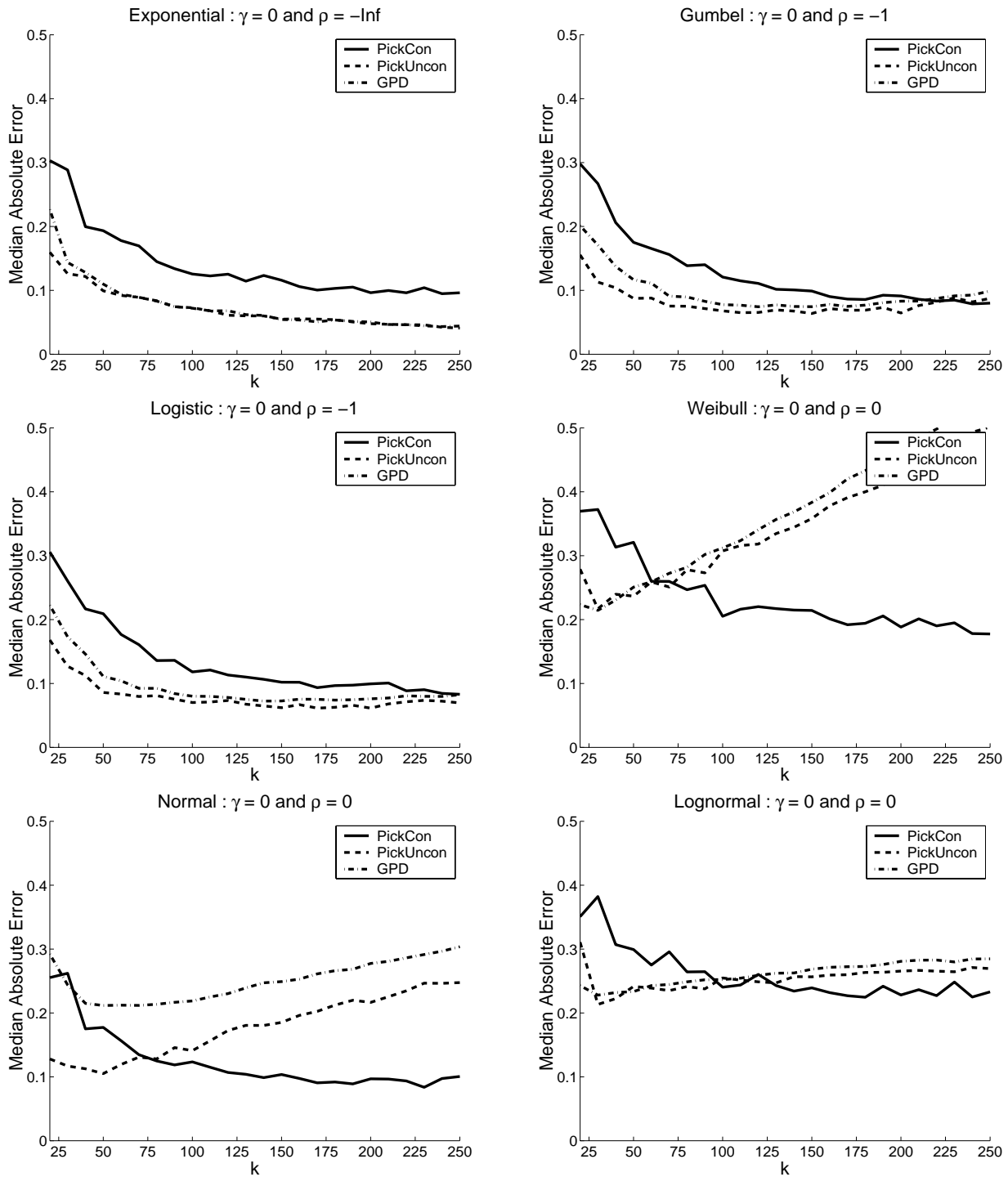
Figure 1: Median absolute errors of unconstrained adaptive generalized Pickands estimator $(--)$, constrained adaptive generalized Pickands estimator $(—)$ and maximum likelihood estimator in GPD model $(-\cdot-)$ for samples from GPD, Fréchet, and Student-t distributions.

Figure 2: Median absolute errors of unconstrained adaptive generalized Pickands estimator ($--$), constrained adaptive generalized Pickands estimator (—) and maximum likelihood estimator in GPD model ($-\cdot-$) for samples from Exponential, Gumbel, Logistic, Weibull, Normal, and Lognormal distribution.
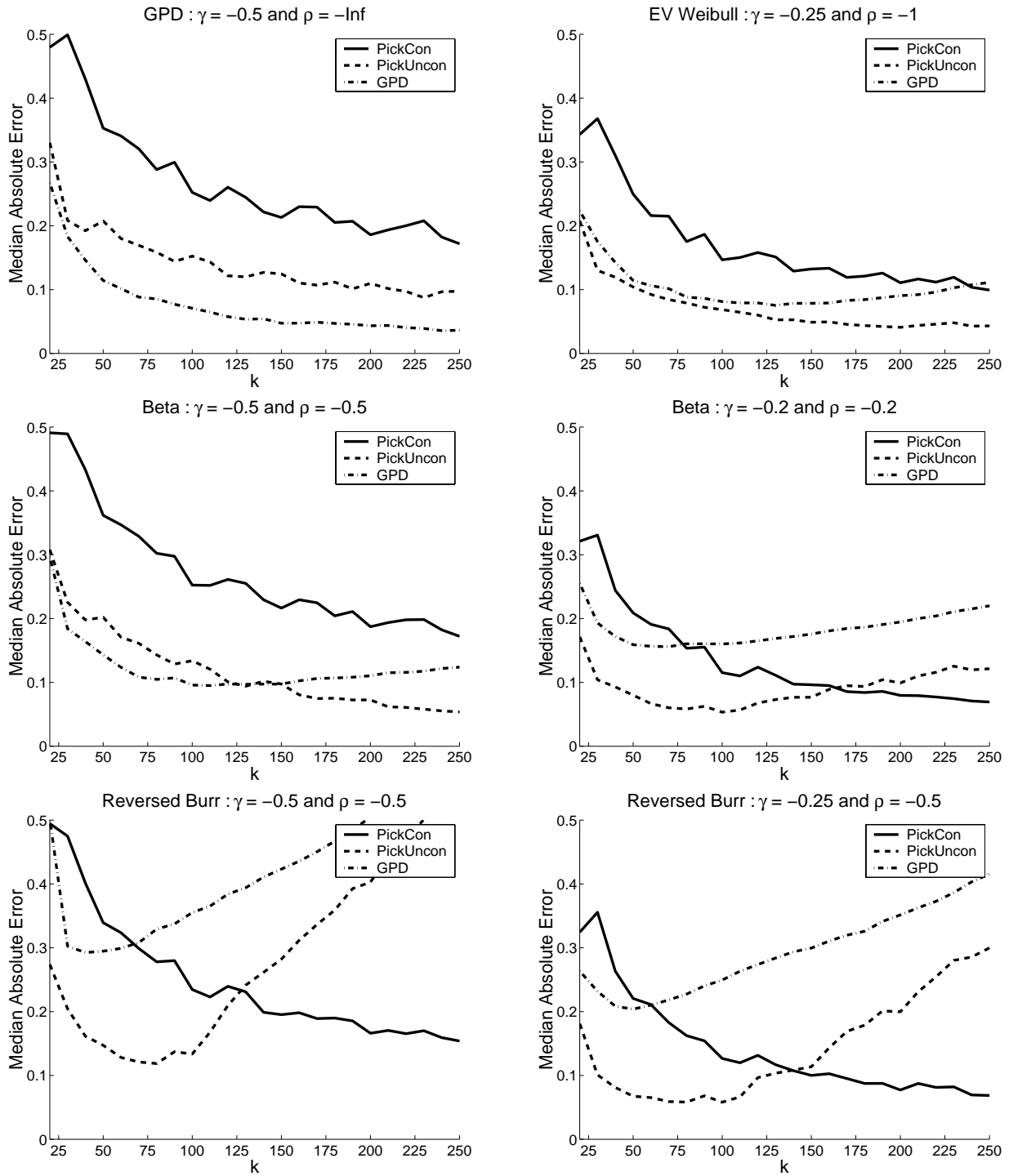
25

Figure 3: Median absolute errors of unconstrained adaptive generalized Pickands estimator $(--)$, constrained adaptive generalized Pickands estimator $(—)$ and maximum likelihood estimator in GPD model $(-\cdot-)$ for samples from GPD, EV Weibull, Beta, and Reversed Burr distributions.