Approximations to the Tail Empirical Distribution Function with Application to Testing Extreme Value Conditions

Holger Drees*, Laurens de Haan†, Deyuan Li†

Abstract. Weighted approximations to the tail of the distribution function and its empirical counterpart are derived which are suitable for applications in extreme value statistics. The approximation of the tail empirical distribution function is then used to develop an Anderson-Darling type test of the null hypothesis that the distribution function belongs to the domain of attraction of an extreme value distribution.

Running title: Tail empirical distribution function
AMS 2000 subject classification: primary 62G30, 62G32; secondary 62G10
Keywords and phrases: Anderson-Darling test, domain of attraction, limit distribution, maximum likelihood, tail empirical distribution function, weighted approximation

1 Introduction

To assess the risk of extreme events that have not occurred yet, one needs to estimate the distribution function (d.f.) in the far tail. Extreme value theory provides a natural framework for an extrapolation of the distribution function beyond the range of available observations via the so-called Pareto approximation of the tail.

Assume that i.i.d. random variables (r.v.'s) X_i, $1 \le i \le n$, with d.f. F are observed such that

$$\lim_{n\to\infty} P\Big\{a_n^{-1}(\max_{1\le i\le n} X_i - b_n) \le x\Big\} = G_\gamma(x)$$

for all $x \in \mathbb{R}$, with some normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$; in short we write $F \in D(G_\gamma)$.

*Department of Mathematics, SPST, University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany; E-mail: drees@math.uni-hamburg.de

†Department of Econometrics, Faculty of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; E-mail:ldehaan@few.eur.nl, dli@few.eur.nl

Here

$$G_\gamma(x) := \exp\big(-(1+\gamma x)^{-1/\gamma}\big) \tag{1.1}$$

for all $x \in \mathbb{R}$ such that $1+\gamma x > 0$, and $\gamma \in \mathbb{R}$ is the so-called extreme value index. For $\gamma = 0$, the right-hand side of (1.1) is defined as $\exp(-\mathrm{e}^{-x})$.

This extreme value condition can be rephrased in the following way:

$$\lim_{t\to\infty} t\bar{F}(\tilde{a}(t)x + \tilde{b}(t)) = (1+\gamma x)^{-1/\gamma} \tag{1.2}$$

for all x with $1+\gamma x > 0$. Here $\bar{F} := 1-F$, $\tilde{a}$ is some positive normalizing function and $\tilde{b}(t) := U(t)$ with

$$U(t) := \Big(\frac{1}{1-F}\Big)^{\leftarrow}(t) = F^{\leftarrow}\Big(1-\frac{1}{t}\Big)$$

and $F^{\leftarrow}$ denoting the generalized inverse of F. In other words, if X is a r.v. with d.f. F, then

$$\lim_{t\to\infty} P\Big(\frac{X-\tilde{b}(t)}{\tilde{a}(t)} \le x \,\Big|\, X > \tilde{b}(t)\Big) = 1-(1+\gamma x)^{-1/\gamma} =: V_\gamma(x)$$

for $x > 0$, where V_γ is a so-called generalized Pareto distribution. Thus, roughly speaking, we have for large t and $x > \tilde{b}(t)$

$$\bar{F}(x) = P\{X > x\} \approx t^{-1}\Big(1+\gamma\frac{x-\tilde{b}(t)}{\tilde{a}(t)}\Big)^{-1/\gamma}, \tag{1.3}$$

that is, the tail of the d.f. can be approximated by a rescaled tail of a generalized Pareto distribution with suitable scale and location parameter and shape parameter γ. Since the latter can be easily extrapolated beyond the range of the observations, this framework offers an approach for estimating the d.f. F in the far tail.

Condition (1.2) holds for most standard distribution, but not for all distributions. Hence before applying approximation (1.3) one should check whether (1.2) is a reasonable assumption for the data set under consideration. To this end, we do not want to specify the exact parameters of the approximating generalized Pareto distribution beforehand.

A natural way to check the validity of (1.2) is to compare the tail of the empirical d.f. and a generalized Pareto distribution with estimated parameters by some goodness-of-fit test. Here we focus on tests of Anderson-Darling-type; however, using the empirical process approximations that will be established in the paper, similar results can be easily proved for other goodness-of-fit tests.

Davison and Smith (1990) applied such goodness-of-fit tests to the famous River Nidd data, but they used the critical values of the tests for exponentiality. Doing so, they ignored the fact that the exponential distribution is just one of the possible limiting generalized Pareto distributions (cf. p. 4141 of Davison and Smith, 1990) and, in addition, that the parameters of the generalized Pareto distribution must be estimated first. Indeed, we will see that in general the estimation of the shape, scale and location parameters influence the asymptotic distribution of the test statistic. This was already observed in a purely parametric generalized Pareto model by Choulakian and Stephens (2001).

In the classical setting when a simple null hypothesis $F = F_0$ is to be tested, test statistics of Anderson-Darling type can be written in the form

$$\int_0^1 \big(F_n(F_0^{-1}(x)) - x\big)^2 \psi(x)\, dx$$

for a suitable weight function ψ which is unbounded near the boundary of the interval $[0,1]$; here F_n denotes the empirical d.f. defined by

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}, \quad x \in \mathbb{R}.$$

If the null hypothesis is composite (but of parametric form), then F_0 is replaced with a d.f. with estimated parameters.

In the present framework two differences must be taken into account. First, we do not assume that the left hand side and the right hand side of (1.3) are exactly equal, but the unknown d.f. F is only approximated by the "theoretical" generalized Pareto d.f. Second, this approximation is expected to hold only in the right tail, for $x > \tilde{b}(n/k)$ with $k \ll n$, say. In the asymptotic setting, we will assume that $k = k_n$ is an intermediate sequence, that is,

$$\lim_{n\to\infty} k_n = \infty, \quad \lim_{n\to\infty} k_n/n = 0.$$

The first condition is necessary to ensure consistency of the test, while the second condition reflects the restriction to the tail.

To be more specific, here we consider the test statistic

$$T_n := \int_0^1 \left[\frac{n}{k_n}\bar{F}_n\left(\hat{a}(\frac{n}{k_n})\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n})\right) - x\right]^2 x^{\eta-2}\, dx \tag{1.4}$$

with $\bar{F}_n := 1 - F_n$. Here $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ are suitable estimators of the shape, scale and location parameter to be discussed later on, and η is an arbitrary positive constant. Since this test statistic measures a distance between the conditional distribution of the excesses above $\hat{b}(n/k_n)$ and an approximating generalized Pareto distribution (cf. (1.2)), a plot of this statistic as a function of $k = k_n$ may also be a useful tool for determining the point from which on approximation (1.3) is sufficiently accurate.

In the classical setting with simple null hypothesis, the asymptotic distribution of the Anderson-Darling test statistic under the null hypothesis is usually derived from a weighted approximation of the empirical distribution function. In analogy, in Theorem 2.1 we state a weighted approximation to the tail empirical process

$$Y_n(x) := \sqrt{k_n}\Big(\frac{n}{k_n}\bar{F}_n\Big(a\Big(\frac{n}{k_n}\Big)x + b\Big(\frac{n}{k_n}\Big)\Big) - (1+\gamma x)^{-1/\gamma}\Big), \quad x \in \mathbb{R}. \tag{1.5}$$

For the uniform distribution such approximations are well known; see, e.g., Csörgő and Horváth (1993, Theorems 5.1.5 and 5.1.10). For more general d.f.'s $F \in D(G_\gamma)$, one must very carefully choose suitable modifications a and b of the normalizing functions to obtain accurate weighted approximations (cf. Lemma 2.1). Moreover, it turns out that for a certain class of d.f.'s with extreme value index $\gamma = 0$, a qualitatively different result holds. Proposition 2.1 gives an analogous approximation to the corresponding process with estimated parameters in the case $\gamma > -1/2$. The asymptotic normality of T_n then follows easily (Theorem 2.2).

An important step in the proof of approximation of Y_n is to establish a weighted approximation to the (deterministic) tail d.f. $\bar{F}$ or, more precisely, to $t\bar{F}(a(t)x+b(t))-(1+\gamma x)^{-1/\gamma}$, which is proved in Section 3 (see Proposition 3.2). This result, a purely analytical analog to the approximation of the tail empirical quantile function (cf. Lemma 2.1) established by Drees (1998), is very useful in a wider context. For instance, Drees et al. (2003) have derived large deviation results in extreme value theory from this approximation. The Sections 4 and 5 contain the proofs of the main results, while in Section 6 asymptotic critical values are determined and the actual size of the Anderson-Darling type test with nominal size 5% is examined in a simulation study.

2 Main results

Approximation to the Tail Empirical Distribution Function

If i.i.d. uniformly distributed r.v.'s U_i are observed, then (1.2) holds with $\tilde{a}(t) = 1/t$ and $\gamma = -1$. For this particular case, Csörgő and Horváth (1993, Theorems 5.1.5 and 5.1.10) gave a weighted approximation to the left tail analog of the normalized tail empirical process Y_n defined in (1.5). Let

$$U_n(t) := \frac{1}{n}\sum_{i=1}^{n} I_{\{U_i \le t\}}, \quad t \in \mathbb{R},$$

denote the uniform tail empirical d.f. Then there exists a sequence of Brownian motions W_n such that

$$\sup_{t>0} t^{-1/2} \mathrm{e}^{-\epsilon|\log t|} \left| \sqrt{k_n}\Big(\frac{n}{k_n} U_n\Big(\frac{k_n}{n} t\Big) - t\Big) - W_n(t) \right| \xrightarrow{P} 0 \tag{2.1}$$

as $n \to \infty$ for all intermediate sequences k_n, $n \in \mathbb{N}$ (see also Einmahl (1997, Corollary 3.3)).

By the well-known quantile transformation, $(F^{\leftarrow}(1-U_i))_{1\le i\le n}$ has the same distribution as $(X_i)_{1\le i \le n}$. Because $\bar F(x) \le t$ is equivalent to $F^{\leftarrow}(1-t) \le x$, it follows that $\bar F_n$ has the same distribution as

$$x \mapsto 1 - \frac{1}{n}\sum_{i=1}^n 1_{\{F^{\leftarrow}(1-U_i) \le x\}} = \frac{1}{n}\sum_{i=1}^n 1_{\{U_i < \bar F(x)\}} = U_n(\bar F(x) - 0),$$

that is the left hand limit of U_n at $\bar F(x)$. Hence, by the continuity of W_n, we obtain for suitable versions of $\bar F_n$ that

$$\sup_{\{x: z_n(x)>0\}} (z_n(x))^{-1/2} \mathrm{e}^{-\epsilon|\log z_n(x)|} \left| \sqrt{k_n}\Big[\frac{n}{k_n}\bar F_n\Big(\tilde a\big(\frac{n}{k_n}\big)x + \tilde b\big(\frac{n}{k_n}\big)\Big) - z_n(x)\Big] - W_n(z_n(x)) \right| \xrightarrow{P} 0 \tag{2.2}$$

with

$$z_n(x) := \frac{n}{k_n}\bar F\Big(\tilde a\big(\frac{n}{k_n}\big)x + \tilde b\big(\frac{n}{k_n}\big)\Big).$$

In view of (1.2), one may conjecture that (2.2) still holds if $z_n(x)$ is replaced with $(1+\gamma x)^{-1/\gamma}$. However, for this to be justified, one must replace the normalizing functions $\tilde a$ and $\tilde b$ with suitable modifications such that (1.2) holds in a certain uniform sense. Moreover, we must bound the speed at which k_n tends to ∞.

In the sequel, we will focus on distributions which satisfy the following second order refinement of condition (1.2):

$$\lim_{t\to\infty} \frac{t\bar F\big(\tilde a(t)x + \tilde b(t)\big) - (1+\gamma x)^{-1/\gamma}}{\tilde A(t)} = (1+\gamma x)^{-1-1/\gamma} H_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big) \tag{2.3}$$

for all x with $1+\gamma x > 0$, some $\rho \le 0$, a function $\tilde A$ which eventually has constant sign, and

$$H_{\gamma,\rho}(x) := \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right).$$

(Note that, under condition (1.2), $\tilde A(t)$ necessarily tends to 0 as t tends to infinity, because the numerator tends to 0, too.) De Haan and Stadtmüller (1996) proved that (2.3) is equivalent to

$$\lim_{t\to\infty} \frac{\dfrac{U(tx) - \tilde b(t)}{\tilde a(t)} - \dfrac{x^\gamma - 1}{\gamma}}{\tilde A(t)} = H_{\gamma,\rho}(x) \tag{2.4}$$

for all $x > 0$. Moreover, they showed that in (2.3) and (2.4) all possible non-trivial limits must essentially be of the given types, and that $|\tilde{A}|$ is necessarily ρ–varying.

Under this assumption, Drees (1998) and Cheng and Jiang (2001) determined suitable normalizing functions a and b such that convergence (2.4) holds uniformly in the following sense. In what follows, $f(t) \sim g(t)$ means $f(t)/g(t) \to 1$.

Lemma 2.1. *Suppose the second order condition (2.4) holds. Then there exists a function A such that for all $\epsilon > 0$ there is a constant $t_\epsilon > 0$ such that for all t and x with $\min(t, tx) \geq t_\epsilon$*

$$x^{-(\gamma+\rho)}\mathrm{e}^{-\epsilon|\log x|}\left|\frac{\dfrac{U(tx)-b(t)}{a(t)}-\dfrac{x^\gamma-1}{\gamma}}{A(t)}-K_{\gamma,\rho}(x)\right| < \epsilon. \tag{2.5}$$

Here $A(t) \sim \tilde{A}(t)$,

$$a(t) := \begin{cases} ct^\gamma & \text{if} \quad \rho < 0, \\ \gamma U(t) & \text{if} \quad \rho = 0, \;\; \gamma > 0, \\ -\gamma(U(\infty) - U(t)) & \text{if} \quad \rho = 0, \;\; \gamma < 0, \\ U^{**}(t) + U^*(t) & \text{if} \quad \rho = 0, \;\; \gamma = 0 \end{cases}$$

with $c := \lim_{t\to\infty} t^{-\gamma}\tilde{a}(t)$ (which exists if $\rho < 0$),

$$b(t) := \begin{cases} U(t) - a(t)A(t)/(\gamma+\rho) & \text{if } \gamma + \rho \neq 0, \;\; \rho < 0, \\ U(t) & \text{else,} \end{cases}$$

and

$$K_{\gamma,\rho}(x) := \begin{cases} \frac{1}{\gamma+\rho}x^{\gamma+\rho} & \text{if} \quad \rho < 0, \gamma+\rho \neq 0, \\ \log x & \text{if} \quad \rho < 0, \gamma+\rho = 0, \\ \frac{1}{\gamma}x^\gamma \log x & \text{if} \quad \rho = 0 \neq \gamma, \\ \frac{1}{2}\log^2 x & \text{if} \quad \rho = 0 = \gamma, \end{cases}$$

and for any integrable function g the function g^* is defined by

$$g^*(t) := g(t) - \frac{1}{t}\int_0^t g(u)dt.$$

In the sequel, we denote the right endpoint of the support of the generalized Pareto d.f. with extreme value index γ by

$$\frac{1}{(-\gamma)\vee 0} = \begin{cases} -1/\gamma & \text{if} \quad \gamma < 0, \\ \infty & \text{if} \quad \gamma \geq 0, \end{cases}$$

and its left endpoint by

$$-\frac{1}{\gamma \vee 0} = \begin{cases} -\infty & \text{if} \quad \gamma \leq 0, \\ -1/\gamma & \text{if} \quad \gamma > 0. \end{cases}$$

We have the following approximation to the tail empirical process Y_n defined in (1.5):

Theorem 2.1. *Suppose that the second order condition (2.4) holds for some $\gamma \in \mathbb{R}$ and $\rho \leq 0$. Let k_n be an intermediate sequence such that $\sqrt{k_n} A(n/k_n)$, $n \in \mathbb{N}$, is bounded and choose a, b and A as in Lemma 2.1. Then there exist versions of $\bar{F}_n$ and a sequence of Brownian motions W_n such that for all $x_0 > -1/(\gamma \vee 0)$*

(i)

$$\sup_{x_0 \leq x < 1/((-\gamma)\vee 0)} \big((1+\gamma x)^{-1/\gamma}\big)^{-1/2+\epsilon} \Big| Y_n(x) - W_n\big((1+\gamma x)^{-1/\gamma}\big) - \sqrt{k_n} A\big(\frac{n}{k_n}\big)(1+\gamma x)^{-1/\gamma-1} K_{\gamma,\rho}\big((1+\gamma x)^{1/\gamma}\big)\Big| \xrightarrow{P} 0$$

if $\gamma \neq 0$ or $\rho < 0$, and

(ii)

$$\sup_{x_0 \leq x < \infty} \Big(\max\Big(\mathrm{e}^{-x}, \frac{n}{k_n}\bar{F}\Big(a\big(\frac{n}{k_n}\big)x + b\big(\frac{n}{k_n}\big)\Big)\Big)\Big)^{-1/2+\epsilon} \cdot \Big| Y_n(x) - W_n(\mathrm{e}^{-x}) - \sqrt{k_n} A\big(\frac{n}{k_n}\big)\mathrm{e}^{-x}\frac{x^2}{2}\Big| \xrightarrow{P} 0$$

if $\gamma = \rho = 0$.

Remark 2.1. *If, in particular, $\sqrt{k_n} A(n/k_n)$ tends to 0, then the bias term $\sqrt{k_n} A(n/k_n)$ $(1 + \gamma x)^{-1/\gamma-1} K_{\gamma,\rho}((1+\gamma x)^{1/\gamma})$ is asymptotically negligible. In order for this statement to be true, it is sufficient to assume that the left-hand side of (2.3) remains bounded (rather than the present limit requirement) provided that k_n tends to infinity sufficiently slowly.*

The assertion in Theorem 2.1(ii) is wrong if the maximum of e^{-x} and $n/k_n \bar{F}(a(n/k_n)x + b(n/k_n))$ is replaced with just one of these two terms. In particular, unlike in the case $\gamma \neq 0$ or $\rho > 0$, here one cannot use a power of the pertaining generalized Pareto distribution $(1+\gamma x)^{-1/\gamma}$. Hence the asymptotic behavior of the tail empirical d.f. in the case $\gamma = \rho = 0$ is qualitatively different from the behavior in the case (i). This is due to the fact that in the case $\gamma \neq 0$ or $\rho < 0$ the tail behavior of F is essentially determined by the parameters γ and ρ, while in the case $\gamma = \rho = 0$ tail behaviors as diverse as $\bar{F}(x) \sim \exp(-\log^2 x)$, $\bar{F}(x) \sim \exp(-\sqrt{x})$ and $\bar{F}(x) \sim \exp(-x^2)$, say, are possible (cf. Example 3.1).

Nevertheless, also in the case $\gamma = \rho = 0$ results similar to the one in case (i) hold if $\max\big(\mathrm{e}^{-x},\ (n/k_n)\bar F(a(n/k_n)x + b(n/k_n))\big)$ is replaced with some weight function converging to 0 much slower than e^{-x} as x tends to ∞:

Corollary 2.1. *Under the conditions of Theorem 2.1 with $\gamma = \rho = 0$ one has for all $\tau > 0$*

$$\sup_{x_0 \le x < \infty} \max(1, x^\tau) \left| Y_n(x) - W_n(\mathrm{e}^{-x}) - \sqrt{k_n} A\big(\frac{n}{k_n}\big) \mathrm{e}^{-x} \frac{x^2}{2} \right| \xrightarrow{P} 0.$$

The proofs of Theorem 2.1 and Corollary 2.1 are given in section 4.

According to these results, the standardized tail empirical d.f.

$$Y_n((x^{-\gamma} - 1)/\gamma) = \sqrt{k_n}\Big(\frac{n}{k_n}\bar F_n\Big(a\big(\frac{n}{k_n}\big)\frac{x^{-\gamma}-1}{\gamma} + b\big(\frac{n}{k_n}\big)\Big) - x\Big), \quad x \in (0,1],$$

converges to a Brownian motion plus a bias term if k_n tends to ∞ not too fast. This may be used to construct a test for $F \in D(G_\gamma)$. However, to this end, first the unknown parameters $\gamma, a(n/k_n)$ and $b(n/k_n)$ must be replaced with suitable estimators. The following result is an analog to Theorem 2.1(i) and Corollary 2.1 for the process with estimated parameters in the case $\gamma > -1/2$.

Proposition 2.1. *Suppose that the conditions of Theorem 2.1 are satisfied for some $\gamma > -1/2$. Let $\hat\gamma_n$, $\hat a(n/k_n)$ and $\hat b(n/k_n)$ be estimators such that*

$$\sqrt{k_n}\Big(\hat\gamma_n - \gamma,\ \frac{\hat a(n/k_n)}{a(n/k_n)} - 1,\ \frac{\hat b(n/k_n) - b(n/k_n)}{a(n/k_n)}\Big) - \big(\Gamma(W_n),\ \alpha(W_n),\ \beta(W_n)\big) \xrightarrow{P} 0 \tag{2.6}$$

for some measurable real-valued functionals Γ, α and β of the Brownian motions W_n used in Theorem 2.1. Let

$$L_n^{(\gamma)}(x) := \begin{cases} \frac{1}{\gamma}x\big(\frac{1}{\gamma}\Gamma(W_n) - \alpha(W_n)\big) + \frac{1}{\gamma}\Gamma(W_n)x\log x & \\ \qquad -\frac{1}{\gamma}x^{1+\gamma}\big(\gamma\beta(W_n) + \frac{1}{\gamma}\Gamma(W_n) - \alpha(W_n)\big) & \textit{if} \quad \gamma \ne 0, \\ x\big(-\beta(W_n) - \frac{1}{2}\Gamma(W_n)\log^2 x + \alpha(W_n)\log x\big) & \textit{if} \quad \gamma = 0, \end{cases}$$

and

$$h(x) = \begin{cases} x^{-1/2+\epsilon} & \textit{if} \quad \gamma \ne 0 \textit{ or } \rho < 0, \\ (1 + |\log x|)^\tau & \textit{if} \quad \gamma = \rho = 0. \end{cases}$$

Then, for the versions of $\bar F_n$ used in Theorem 2.1 and every $\epsilon > 0$ and $\tau > 0$, one has

$$\sup_{0<x\le 1} h(x)\Big|\sqrt{k_n}\,\Big[\frac{n}{k_n}\bar F_n\Big(\hat a\big(\frac{n}{k_n}\big)\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n} + \hat b\big(\frac{n}{k_n}\big)\Big) - x\Big] - W_n(x) - L_n^{(\gamma)}(x) - \sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)\Big| \xrightarrow{P} 0. \tag{2.7}$$

Remark 2.2. *(i) If $\gamma < -1/2$, a rate of convergence of $k_n^{-1/2}$ for the estimators in (2.6) is not sufficient to ensure the approximation (2.7). To see this, note that in this case $\hat{b}(n/k_n) - b(n/k_n)$ is of larger order than $k_n^{-1/2}(n/k_n)^{\gamma-\epsilon}$ and hence also of larger order than the difference between the i_nth largest order statistic and the right endpoint $F^{\leftarrow}(1)$ for some sequence $i_n \to \infty$ not too fast, leading, for small $x > 0$, to a non-negligible difference between $\bar{F}_n(a(n/k_n)(x^{-\gamma} - 1)/\gamma + b(n/k_n))$ and the corresponding expression with estimated parameters.*

(ii) Typically the functionals Γ, α and β depend on the underlying d.f. F only through γ if the estimators $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ use only the largest $k_n + 1$ order statistics and $\sqrt{k_n}A(n/k_n) \to 0$. This justifies the notation $L_n^{(\gamma)}$ for the limiting function occurring in (2.7) in that case. However, if $\sqrt{k_n}A(n/k_n) \to c \neq 0$ then $L_n^{(\gamma)}$ will also depend on c; for simplicity, we ignore this dependence in the notation.

Example 2.1. In Proposition 2.1 one may use the so-called maximum likelihood estimator in a generalized Pareto model discussed by Smith (1987). Denote the jth order statistic by $X_{j,n}$. Since the excesses $X_{n-i+1,n} - X_{n-k_n,n}$, $1 \le i \le k_n$ over the random threshold $X_{n-k_n,n}$ are approximately distributed according to a generalized Pareto distribution with shape parameter γ and scale parameter $\sigma_n := a(n/k_n)$ if $F \in D(G_\gamma)$ and k_n is not too big, γ and σ_n are estimated by the pertaining maximum likelihood estimators $\hat{\gamma}_n$ and $\hat{\sigma}_n$ in an exact generalized Pareto model for the excesses. They can be calculated as the solutions to the equations

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k} \log\Big(1 + \frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})\Big) &= \gamma \\
\frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 + \frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})} &= \frac{1}{\gamma+1}.
\end{aligned}$$

In Theorem 2.1 of Drees et al. (2003) it is proved that $\hat{\gamma}_n$, $\hat{a}(n/k_n) := \hat{\sigma}_n$ and $\hat{b}(n/k_n) := X_{n-k_n,n}$ satisfy (2.6) with

$$\begin{aligned}
\Gamma(W_n) &= -\frac{(\gamma+1)^2}{\gamma}\big((2\gamma+1)S_n - R_n\big) + (\gamma+1)W_n(1), \\
\alpha(W_n) &= -\frac{\gamma+1}{\gamma}\big(R_n - (\gamma+1)(2\gamma+1)S_n\big) - (\gamma+2)W_n(1), \\
\beta(W_n) &= W_n(1),
\end{aligned}$$

where

$$R_n := \int_0^1 t^{-1} W_n(t)\, dt,$$

$$S_n := \int_0^1 t^{\gamma-1} W_n(t)\, dt,$$

provided $\sqrt{k_n} A(n/k_n) \to 0$; if $\sqrt{k_n} A(n/k_n) \to c > 0$ then additional bias terms enter the formulas. As usual, for $\gamma = 0$, these expressions are to be interpreted as their limits as γ tends to 0, that is,

$$\begin{aligned} \Gamma(W_n) &= -\int_0^1 (2 + \log t) t^{-1} W_n(t)\, dt + W_n(1), \\ \alpha(W_n) &= \int_0^1 (3 + \log t) t^{-1} W_n(t)\, dt - 2W_n(1), \\ \beta(W_n) &= W_n(1). \end{aligned}$$

(Applying Vervaat's (1972) lemma to the approximation to the tail empirical distribution function given in Theorem 2.1, restricted to a compact interval bounded away from 0, and then using a Taylor expansion of $t \mapsto (t^{-\gamma} - 1)/\gamma$ shows that the Brownian motions used by Drees et al. (2003) are indeed the Brownian motions used in Proposition 2.1 multiplied with -1.)

Hence one may apply Proposition 2.1 to obtain the asymptotics of the tail empirical distribution function with estimated parameters. $\square$

A Test for the Extreme Value Condition

It is easy to devise tests for $F \in D(G_\gamma)$ with $\gamma > -1/2$ using approximation (2.7). For example, using the following limit theorem, the critical values of the Anderson-Darling type test can be calculated which rejects the null hypothesis if $k_n T_n$ (defined in (1.4)) is too large.

Theorem 2.2. *Under the conditions of Proposition 2.1 with $\sqrt{k_n} A(n/k_n) \to 0$ one has*

$$k_n T_n - \int_0^1 \Big(W_n(x) + L_n^{(\gamma)}(x) \Big)^2 x^{\eta-2}\, dx \xrightarrow{P} 0 \tag{2.8}$$

for all $\eta > 0$ if $\gamma \neq 0$ or $\rho < 0$, and all $\eta \geq 1$ if $\gamma = \rho = 0$.

Remark 2.3. *Note that $\int_0^1 \Big(W_n(x) + L_n^{(\gamma)}(x) \Big)^2 x^{\eta-2}\, dx > 0$ a.s., because $L_n^{(\gamma)}$ is a differentiable function on $(0, 1]$ while W_n has almost surely continuous, non-differentiable sample paths.*

Since the continuous distribution of $\int_0^1 (W_n(x) + L_n^{(\gamma)}(x))^2 x^{\eta-2} dx$ does not depend on n, for fixed $\gamma > -1/2$ its quantiles $Q_{p,\gamma}$ defined by $P\{\int_0^1 (W_n(x) + L_n^{(\gamma)}(x))^2 x^{\eta-2} dx \leq Q_{p,\gamma}\} = p$ can be obtained by simulations (see Section 6). Then the test rejecting $F \in D(G_\gamma)$ if $k_n T_n > Q_{1-\bar\alpha,\gamma}$ has asymptotic size $\bar\alpha \in (0, 1)$. (Likewise, one can consider a 'two-sided' test that rejects the hypothesis if $k_n T_n < Q_{\bar\alpha/2,\gamma}$ or $k_n T_n > Q_{1-\bar\alpha/2,\gamma}$. However, this test seems intuitively less appealing because usually the tail empirical d.f. will be poorly approximated by a generalized Pareto d.f. under the alternative hypothesis.)

If one wants to test $F \in D(G_\gamma)$ for an arbitrary unknown $\gamma > -1/2$, one may use the test rejecting the null hypothesis if $k_n T_n > Q_{1-\bar{\alpha},\tilde{\gamma}_n}$ for some estimator $\tilde{\gamma}_n$ which is consistent for γ if $F \in D(G_\gamma)$. If the functionals Γ, α and β determining the limit distributions of $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ are continuous functions of γ (like the ones obtained in Example 2.1), then also $L_n^{(\gamma)}(x)$ and hence the quantiles $Q_{p,\gamma}$ are continuous functions of γ. Thus the test has asymptotic size $\bar{\alpha}$.

Remark 2.4. *A natural alternative to the maximum likelihood estimators discussed in Example 2.1 is to estimate $a(n/k_n), b(n/k_n)$ and γ by those values a, b and γ for which the test statistic*

$$\int_0^1 \left[\frac{n}{k_n}\bar{F}_n\left(a\frac{x^{-\gamma}-1}{\gamma}+b\right)-x\right]^2 x^{\eta-2}\,dx$$

is minimized. Note, however, that the minimization of the integral in the threedimensional space $(0,\infty)\times\mathbb{R}\times(-1/2,\infty)$ can be a daunting task, because it is a discontinuous function of its arguments.

However, recall that, in fact, for (2.8) to hold we have not merely assumed that $F \in D(G_\gamma)$ but also that the second order condition (2.4) holds and, for the particular k_n used in the definition of the test statistic T_n, in addition we have assumed that $A(t) \to 0$ sufficiently fast such that $\sqrt{k_n}A(n/k_n) \to 0$. Hence, we actually test the subset of the null hypothesis $F \in D(G_\gamma)$ described by these additional assumptions. This, however, is exactly what is needed in statistical applications. For instance, note that typically the very same assumptions are made when confidence intervals for extreme quantiles or for exceedance probability over high thresholds are calculated. Therefore, for this purpose, one must not only check whether $F \in D(G_\gamma)$ but whether the Pareto approximation is sufficiently accurate for the number of order statistics used for estimation! Moreover, if one lets k vary, then the test statistic can also be used to find the largest k for which the Pareto approximation of the tail distribution beyond $X_{n-k:n}$ is justified.

Remark 2.5. *If one first tests for $F \in D(G_\gamma)$ and, in the case of acceptance, then calculates confidence intervals of the interesting extreme value parameters, then the confidence bounds should be adjusted for this pre-testing. For example, to construct an adjusted confidence interval for γ, first determine (by simulation) a constant $r^{(\gamma)}$ such that*

$$P\Big\{\int_0^1 \big(W_n(x)+L_n^{(\gamma)}(x)\big)^2 x^{\eta-2}\,dx \le Q_{1-\bar{\alpha},\gamma},\ |\Gamma(W_n)| \le r^{(\gamma)}\Big\} = \beta$$

for some pre-specified $\beta < 1-\bar{\alpha}$. As above, let $\tilde{\gamma}_n$ be any consistent estimator of γ. Then, under the conditions of Theorem 2.2, the probability that the hypothesis is accepted and $\gamma \in I_n :=$

$[\hat{\gamma}_n - k_n^{-1/2} r^{(\tilde{\gamma}_n)},\ \hat{\gamma}_n + k_n^{-1/2} r^{(\tilde{\gamma}_n)}]$ *converges to* β*. In this sense,* I_n *is a confidence interval with asymptotic confidence level* β*.*

A test for a similar hypothesis, but based on the tail empirical quantile function instead of the tail empirical distribution function, has been discussed by Dietrich et al. (2002). That test does not require $\gamma > -1/2$ but, on the other hand, $U(\infty) > 0$ and a slightly different second order condition were assumed.

The test based on the statistic $k_n T_n$ becomes particularly simple if Γ, α and β are the zero functional, that is, the standardized estimation errors $\hat{\gamma}_n - \gamma$, $\hat{a}(n/k_n)/a(n/k_n)$ and $(\hat{b}(n/k_n) - b(n/k_n))/a(n/k_n)$ converge at a faster rate than $k_n^{-1/2}$. This can be achieved by using suitable estimators based on m_n largest order statistics with $k_n = o(m_n)$ and $\sqrt{m_n}A(n/m_n) \to 0$. (For example, γ may be estimated by the estimator given in Example 2.1 with m_n instead of k_n, and $b(n/k_n)$ by a quantile estimator of the type described in de Haan and Rootzén (1993).) In that case the limit distribution $\int_0^1 W_n^2(x) x^{\eta-2} dx$ of the test statistic $k_n T_n$ does not depend on γ, so that no consistent estimator $\tilde{\gamma}_n$ for γ is needed. However, this approach has two disadvantages. Firstly, in practice it is often not an easy task to choose k_n such that the bias is negligible (i.e. $\sqrt{k_n}A(n/k_n) \to 0$). It is even more delicate to choose two numbers k_n and m_n such that k_n is much smaller than m_n but not too small and, at the same time, the bias of the estimators of the parameters is still not dominating when these are based on m_n order statistics. Secondly, while this approach may lead to a test whose actual size is closer to the nominal value $\bar{\alpha}$, the power of the test will probably higher if one choose a larger value for k_n, e.g. $k_n = m_n$, because the larger k_n the larger will typically be the test statistic $k_n T_n$ if the tail empirical d.f. is not well approximated by a generalized Pareto d.f. For these reasons, in the simulation study we will focus on the case where the tail empirical d.f. and the estimators $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ are based on the same number of largest order statistics.

3 Tail Approximation to the Distribution Function

A substantial part of the proof of Theorem 2.1 consists of proving an approximation to the tail of the (deterministic) distribution function. In the sequel, we will use the notation

$$\begin{aligned} g_t(x) &:= t\bar{F}(a(t)x + b(t)) \\ g(x) &:= (1 + \gamma x)^{-1/\gamma}. \end{aligned}$$

Then g_t converges to g pointwise as t tends to infinity by the basic assumption $F \in D(G\gamma)$ (cf. (1.2)). The following propositions give weighted approximations to the difference $g_t - g$, that are analogous to the approximation (2.5) for the quantile function. Indeed, because $g_t^{\leftarrow}(1/x) = (U(tx) - b(t))/a(t)$, inequality (2.5) gives a weighted approximation to $g_t^{\leftarrow} - g^{\leftarrow}$, which will be used to derive a similar approximation for $g_t - g$.

It is intuitively clear that under the second order condition, that describes the behavior of the right tail of F, an approximation to $g_t(x) - g(x)$ can hold uniformly only for certain values of x for which $a(t)x + b(t)$ belongs to the tail of the support of F. More precisely, for all $c, \delta > 0$, we define sets

$$D_{t,\rho} := D_{t,\rho,\delta,c} := \begin{cases} \{x : g_t(x) \le ct^{1-\delta}\} & \text{if } \rho < 0, \\ \{x : g_t(x) \le |A(t)|^{-c}\} & \text{if } \rho = 0. \end{cases}$$

Check that, in particular, eventually $[x_0, \infty) \subset D_{t,\rho}$ for all $x_0 > -1/(\gamma \vee 0)$.

Proposition 3.1. *Suppose that the second order relation (2.4) holds for some $\gamma \in \mathbb{R}$ and $\rho \le 0$. For $\epsilon > 0$, define*

$$w_t(x) := \begin{cases} g_t^{\rho-1}(x) \exp(-\epsilon |\log g_t(x)|) & \text{if } \gamma \neq 0 \text{ or } \rho \neq 0, \\ \min\left((g_t(x))^{-1} \exp(-\epsilon |\log g_t(x)|), \mathrm{e}^{x-\epsilon|x|}\right) & \text{if } \gamma = \rho = 0. \end{cases}$$

Then, for all $\epsilon, \delta, c > 0$,

$$\sup_{x \in D_{t,\rho}} w_t(x) \Big| \frac{g_t(x) - g(x)}{A(t)} - g_t^{1+\gamma}(x) K_{\gamma,\rho}(1/g_t(x)) \Big| \to 0$$

as $t \to \infty$.

Note that $\mathrm{e}^{x-\epsilon|x|} = (g(x))^{-1} \exp(-\epsilon |\log g(x)|)$ if $\gamma = \rho = 0$. Hence, in this case, we weight with the minimum of the weight function used in the other cases and the analog where g_t is replaced with the limiting function g.

Next we establish an analogous result where $g_t(x)$ is replaced with $g(x)$. To this end, let for $\delta, c > 0$

$$\tilde{D}_{t,\rho} := \tilde{D}_{t,\rho,\delta,c} := \begin{cases} \{x : g(x) \le ct^{1-\delta}\} & \text{if } \rho < 0, \\ \{x : g(x) \le |A(t)|^{-c}\} & \text{if } \rho = 0, \end{cases}$$

and, for $\gamma \neq 0$ or $\rho < 0$,

$$w(x) := g^{\rho-1}(x) \exp(-\epsilon |\log g(x)|).$$

Proposition 3.2. *If the second order relation (2.4) holds for some $\gamma \in \mathbb{R}$ and $\rho \le 0$, then*

$$\sup_{x \in D_{t,\rho}} w_t(x) \Big| \frac{g_t(x) - g(x)}{A(t)} - g^{1+\gamma}(x) K_{\gamma,\rho}(1/g(x)) \Big| \to 0 \tag{3.1}$$

as $t \to \infty$. Moreover, if $\gamma \neq 0$ or $\rho < 0$, then

$$\sup_{x \in \tilde{D}_{t,\rho}} w(x) \Big| \frac{g_t(x) - g(x)}{A(t)} - g^{1+\gamma}(x) K_{\gamma,\rho}(1/g(x)) \Big| \to 0, \tag{3.2}$$

and for $\gamma = \rho = 0$

$$\sup_{x \in \tilde{D}_{t,0}} w_t(x) \Big| \frac{g_t(x) - g(x)}{A(t)} - \mathrm{e}^{-x} \frac{x^2}{2} \Big| \to 0 \tag{3.3}$$

for all $\delta, c > 0$ as $t \to \infty$.

At first glance, it is somewhat surprising that the results look differently in the case $\gamma = \rho = 0$ in that one needs a more complicated weight function, namely the minimum of a function of the standardized tail d.f. $g_t(x)$ and the corresponding function of the limiting exponential d.f. The following example shows that indeed the straightforward analog to the assertion in the case $\gamma \neq 0$ or $\rho < 0$ does not hold, because, in the case $\gamma = \rho = 0$, $g_t(x)$ and $g(x)$ may behave quite differently for large x, despite of the pointwise convergence of $g_t(x)$ towards $g(x)$.

Example 3.1. Here we give an example of a d.f. satisfying (2.4) such that

$$\sup_{x \in \tilde{D}_{t,0}} \mathrm{e}^{x - \epsilon |x|} \Big| \frac{g_t(x) - g(x)}{A(t)} - \mathrm{e}^{-x} \frac{x^2}{2} \Big| \tag{3.4}$$

does not tend to 0 for any $c, \epsilon > 0$.

Let $F(x) := 1 - \mathrm{e}^{-\sqrt{x}}$, $x > 0$, and $a(t) := 2 \log t$, $b(t) := \log^2 t$, $A(t) := 1/\log t$. Then $U(x) = \log^2 x$ satisfies the second order condition (2.4):

$$\frac{1}{A(t)} \Big(\frac{U(tx) - U(t)}{a(t)} - \log x \Big) \to \frac{\log^2 x}{2}$$

as $t \to \infty$. Moreover

$$g_t(x) = t \exp\Big(- \sqrt{2x \log t + \log^2 t} \Big) = \exp\Big(- \log t \big(\sqrt{1 + 2x/\log t} - 1 \big) \Big).$$

Hence, for $x = x(t) = \lambda(t) \log t / 2$ with $\lambda(t) \to \infty$ as $t \to \infty$, one obtains

$$g_t(x) = \exp\Big(- \log t \sqrt{\lambda(t)} (1 + o(1)) \Big),$$

$$\mathrm{e}^{-x} \frac{x^2}{2} = \frac{1}{8} \exp\Big(2(\log\log t + \log \lambda(t)) - \frac{1}{2} \lambda(t) \log t \Big) = o(g_t(x)), \qquad \text{and}$$

$$g(x) = o(g_t(x)),$$

so that

$$\frac{g_t(x)-g(x)}{A(t)} - \mathrm{e}^{-x}\frac{x^2}{2} = \frac{g_t(x)}{A(t)}(1+o(1)).$$

However, this contradicts the convergence of (3.4) to 0 as $t\to\infty$:

$$\begin{aligned}
&(\mathrm{e}^{-x})^{-1+\epsilon}\Big|\frac{g_t(x)-g(x)}{A(t)} - \mathrm{e}^{-x}\frac{x^2}{2}\Big| \\
&= (\mathrm{e}^{-x})^{-1+\epsilon}\frac{g_t(x)}{A(t)}(1+o(1)) \\
&= \exp\Big(\frac{1-\epsilon}{2}\lambda(t)\log t - \sqrt{\lambda(t)}\log t(1+o(1)) + \log\log t\Big)(1+o(1)) \\
&\to \infty.
\end{aligned}$$

Likewise one can show that $F(x)=1-\mathrm{e}^{-x^2}$, $x>0$, satisfies the second order condition (2.4) but that

$$\sup_{x\in D_{t,0}} (g_t(x))^{-1}\exp(-\epsilon|\log g_t(x)|)\Big|\frac{g_t(x)-g(x)}{A(t)} - \mathrm{e}^{-x}\frac{x^2}{2}\Big| \to \infty.$$

□

For the proofs of the propositions, we need some auxiliary results. In what follows, we assume that the conditions of Proposition 3.1 are met. Recall that

$$\Delta_t(x) := g_t^{\leftarrow}(x) - g^{\leftarrow}(x) = \frac{U(t/x)-b(t)}{a(t)} - \frac{x^{-\gamma}-1}{\gamma}.$$

Lemma 3.1. *For each $\epsilon>0$, there exists $\tilde t_\epsilon>0$ such that*

$$\sup_{x\le t/\tilde t_\epsilon} x^{\gamma+\rho}\mathrm{e}^{-\epsilon|\log x|}|\Delta_t(x)| = O(A(t))$$

as $t\to\infty$.

Proof: We focus on the case $\gamma=\rho=0$; the assertion can be proved by the same arguments in the other cases. From Lemma 2.1 we know that, for each $\delta>0$, there exists t_δ such that for $t, t/x\ge t_\delta$

$$\mathrm{e}^{-\epsilon|\log x|}|\Delta_t(x)| \le \mathrm{e}^{-\epsilon|\log x|}|A(t)|\Big(\frac{\log^2 x}{2} + \delta\mathrm{e}^{\delta|\log x|}\Big).$$

Choose $\delta<\epsilon$ and $\tilde t_\epsilon = t_\delta$ to obtain the assertion, since $\sup_{x>0}\mathrm{e}^{-\epsilon|\log x|}\log^2 x<\infty$. □

In the case $\rho=0$, we must deal with those very large values of x separately for which $g_t(x)/g(x)$ is not necessarily close to 1. For this purpose, let

$$B_{t,\rho} := B_{t,\rho,\delta,c} := \begin{cases} (0, ct^{1-\delta}] & \text{if } \rho<0,\\ [|A(t)|^c, |A(t)|^{-c}] = \{x : |\log x|\le c|\log|A(t)||\} & \text{if } \rho=0,\end{cases}$$

with $\delta, c > 0$. Then $D_{t,\rho} = \{x : g_t(x) \in B_{t,\rho}\}$ for $\rho < 0$, while $\{x : g_t(x) \in B_{t,0}\}$ is a strict subset of $D_{t,0}$.

Corollary 3.1. *For all $c, \delta > 0$,*

$$\sup_{x \in B_{t,\rho}} x^\gamma |\Delta_t(x)| \to 0$$

as $t \to \infty$.

Proof:

For $\rho < 0$, choose $\epsilon \leq |\rho|$ in Lemma 3.1 to obtain

$$\sup_{x \leq 1} x^\gamma |\Delta_t(x)| \leq \sup_{x \leq 1} x^{\gamma+\rho} \mathrm{e}^{-\epsilon|\log x|} |\Delta_t(x)| = O(A(t)) = o(1).$$

For all $c, \delta, \tilde{t}_\epsilon > 0$, eventually $ct^{1-\delta}$ is smaller than $t/\tilde{t}_\epsilon$. Hence, by Lemma 3.1,

$$\sup_{1 \leq x \leq ct^{1-\delta}} x^\gamma |\Delta_t(x)| \leq O(A(t)) \cdot \sup_{1 \leq x \leq ct^{1-\delta}} x^{\epsilon-\rho} = O\big(A(t) \cdot t^{(1-\delta)(\epsilon-\rho)}\big) \to 0$$

if $(1-\delta)(\epsilon-\rho) < -\rho$ (which is satisfied for sufficient small $\epsilon > 0$), since $A(t)$ is ρ-varying and hence $A(t) = o(t^{\eta+\rho})$ for all $\eta > 0$.

In the case $\rho = 0$, one has eventually $|A(t)|^c > \tilde{t}_\epsilon/t$, because $|A|$ is slowly varying. Hence, $x > \tilde{t}_\epsilon/t$ for all $x \in B_{t,0}$ and for all $\epsilon \in (0, 1/c)$

$$\sup_{x \in B_{t,0}} x^\gamma |\Delta_t(x)| \leq O(A(t)) \cdot \sup_{x \in B_{t,0}} \mathrm{e}^{\epsilon|\log x|} = O\big(A(t)\mathrm{e}^{\epsilon c|\log|A(t)||}\big) \to 0.$$

□

If F is not eventually strictly increasing, then $g_t^{\leftarrow}(g_t(x))$ may be strictly smaller than x. In this case, we need an upper bound on the difference.

Lemma 3.2. *For all $\epsilon > 0$*

$$\sup_{x \in D_{t,\rho}} g_t^{\gamma+\rho}(x) \exp(-\epsilon|\log g_t(x)|) |g_t^{\leftarrow}(g_t(x)) - x| = o(A(t))$$

$$\sup_{x : g_t(x) \in B_{t,\rho}} g_t^\gamma(x) |g_t^{\leftarrow}(g_t(x)) - x| \to 0.$$

Proof:

De Haan and Stadtmüller (1996) proved that $U(U^{\leftarrow}(y)) - y = o\big(a(U^{\leftarrow}(y))A(U^{\leftarrow}(y))\big)$ as $y \to \infty$. (Note that l. 2 on p. 391 of de Haan and Stadtmüller (1996) contains two errors; the correct formula is $\lim_{t \to f(\infty)}[f(\phi(t)) - t]/a_1(\phi(t)) = 0$.) Because of $U^{\leftarrow}(a(t)x + b(t)) = t/g_t(x)$, it follows that for $v := g_t(x)$ one has

$$g_t^{\leftarrow}(g_t(x)) - x = \frac{U\big(U^{\leftarrow}(a(t)x + b(t))\big) - (a(t)x + b(t))}{a(t)} = o\Big(\frac{a(t/v)}{a(t)} A(t/v)\Big)$$

uniformly for $v \in B_{t,\rho}$. By the Potter bounds (see Bingham et al. (1987), Theorem 1.5.6)

$$\frac{A(t/v)}{A(t)} \leq 2v^{-\rho}\mathrm{e}^{\epsilon|\log v|/2}, \qquad \frac{a(t/v)}{a(t)} \leq 2v^{-\gamma}\mathrm{e}^{\epsilon|\log v|/2}$$

for sufficiently large t, because $t/v \to \infty$ uniformly for all $x \in D_{t,\rho}$. Now the first assertion is obvious and the second assertion follows by the arguments given in the proof of Corollary 3.1. □

Proof of Proposition 3.1. :

Let $v = g_t(x)$. Then, by Taylor's formula,

$$v - g(g_t^{\leftarrow}(v)) = g(g^{\leftarrow}(v)) - g(g_t^{\leftarrow}(v)) = -\big(g'(g^{\leftarrow}(v))\Delta_t(v) + \frac{1}{2}g''(w)\Delta_t^2(v)\big)$$

for some w between $g^{\leftarrow}(v)$ and $g_t^{\leftarrow}(v)$. Note that $g' = -g^{1+\gamma}$ and hence $-g'(g^{\leftarrow}(v)) = v^{1+\gamma}$. Since $g'' = (1+\gamma)g^{1+2\gamma}$ is monotonic in its domain, $g''(w)$ is between $g''(g^{\leftarrow}(v)) = (1+\gamma)v^{1+2\gamma}$ and

$$\begin{aligned} g''(g_t^{\leftarrow}(v)) &= (1+\gamma)\Big(1+\gamma\Big(\frac{v^{-\gamma}-1}{\gamma} + \Delta_t(v)\Big)\Big)^{-(1+2\gamma)/\gamma} \\ &= (1+\gamma)v^{1+2\gamma}(1+\gamma v^{\gamma}\Delta_t(v))^{-(1/\gamma+2)} \\ &= v^{1+2\gamma}O(1) \end{aligned}$$

as $t \to \infty$ uniformly for $v \in B_{t,\rho}$ by Corollary 3.1. Hence, by Lemma 3.1 and Corollary 3.1

$$v - g(g_t^{\leftarrow}(v)) - v^{1+\gamma}\Delta_t(v) = O(v^{1+2\gamma}\Delta_t^2(v)) = v^{1-\rho}\mathrm{e}^{\epsilon|\log v|}o(A(t))$$

uniformly for $v \in B_{t,\rho}$. In view of (2.5), this in turn implies

$$\Big|\frac{v - g(g_t^{\leftarrow}(v))}{A(t)} - v^{1+\gamma}K_{\gamma,\rho}(1/v)\Big| \leq v^{1+\gamma}\Big|\frac{\Delta_t(v)}{A(t)} - K_{\gamma,\rho}(1/v)\Big| + v^{1-\rho}\mathrm{e}^{\epsilon|\log v|}o(1) = v^{1-\rho}\mathrm{e}^{\epsilon|\log v|}o(1). \tag{3.5}$$

Likewise, Taylor's formula and Lemma 3.2 yield

$$|g(g_t^{\leftarrow}(v)) - g(x)| = |g'(w)|v^{-(\gamma+\rho)}\mathrm{e}^{\epsilon|\log v|}o(A(t))$$

for some $w \in (g_t^{\leftarrow}(v), x)$. As above, the monotonicity of g' implies that $|g'(w)|$ is between

$$|g'(g_t^{\leftarrow}(v))| = \big|g(g^{\leftarrow}(v) + \delta_t(v))\big|^{1+\gamma} = v^{1+\gamma}\big|1+\gamma v^{\gamma}\Delta_t(v)\big|^{-(1+1/\gamma)} \qquad = v^{1+\gamma}O(1)$$

and

$$\begin{aligned} |g'(x)| &= \big|g(g^{\leftarrow}(v) + \Delta_t(v) + g_t^{\leftarrow}(v) - x)\big|^{1+\gamma} \\ &= v^{1+\gamma}\big|1+\gamma v^{\gamma}(\Delta_t(v) + g_t^{\leftarrow}(v) - x)\big|^{-(1+1/\gamma)} \\ &= v^{1+\gamma}O(1) \end{aligned}$$

by Corollary 3.1 and Lemma 3.2, and so

$$|g(g_t^{\leftarrow}(v)) - g(x)| = v^{1-\rho} \mathrm{e}^{\epsilon|\log v|} o(A(t))$$

uniformly for $v \in B_{t,\rho}$. Combining this with (3.5), we obtain

$$w_t(x) \Big| \frac{g_t(x) - g(x)}{A(t)} - g_t^{1+\gamma}(x) K_{\gamma,\rho}(1/g_t(x)) \Big| \to 0 \tag{3.6}$$

as $t \to \infty$, uniformly for $v \in B_{t,\rho}$.

Now, for $\rho < 0$ the assertion follows from $D_{t,\rho} = \{x : g_t(x) \in B_{t,\rho}\}$.

In the case $\rho = 0$, it remains to prove that convergence (3.6) also holds uniformly for all x such that $g_t(x) \le |A(t)|^c$, where we may assume $c > 2/\epsilon$.

To this end, first note that for $v = g_t(x)$ and sufficiently large t

$$w_t(x) |g_t^{1+\gamma}(x) K_{\gamma,0}(1/g_t(x))| \le v^{-1+\epsilon} v^{1+\gamma} v^{-\gamma-\epsilon/2} \le |A(t)|^{c\epsilon/2}$$

which tends to 0 uniformly for $v \le |A(t)|^c$. Likewise,

$$w_t(x) \Big| \frac{g_t(x)}{A(t)} \Big| \le v^{-1+\epsilon} \Big| \frac{v}{A(t)} \Big| \le |A(t)|^{c\epsilon - 1} \to 0$$

uniformly for $v \le |A(t)|^c$. Thus, it suffices to verify that $w_t(x) g(x) = o(A(t))$ uniformly for $g_t(x) \le |A(t)|^c$. For this purpose, we exploit the special choice of the normalizing functions a and b given in Lemma 2.1 and treat the cases $\gamma >, <, = 0$ separately.

First we consider the case $\gamma > 0$. Then $a(t)x + b(t) = (1 + \gamma x) U(t) \to \infty$ uniformly for all x satisfying $g_t(x) = t\bar{F}(a(t)x + b(t)) \le |A(t)|^c \to 0$. Because $\bar{F}$ is regularly varying with exponent $-1/\gamma$, $t \ge 1/\bar{F}((1-\epsilon)U(t)) \ge 1/(2\bar{F}(U(t)))$ for sufficiently small $\epsilon > 0$. By the Potter bounds it follows that

$$g_t(x) \ge \frac{\bar{F}\big((1+\gamma x)U(t)\big)}{2\bar{F}(U(t))} \ge \frac{1}{4}(1+\gamma x)^{-1/(\gamma(1-\epsilon/2))} = \frac{1}{4} g^{1/(1-\epsilon/2)}(x) \tag{3.7}$$

and hence

$$w_t(x) g(x) \le g_t^{\epsilon-1}(x) (4 g_t(x))^{1-\epsilon/2} \le 4 |A(t)|^{c\epsilon/2} = o(A(t))$$

uniformly for $g_t(x) \le |A(t)|^c$, i.e. the assertion.

Likewise, in the case $\gamma < 0$, one can conclude the assertion from the inequality

$$g_t(x) \ge \frac{\bar{F}\big(U(\infty) - (1+\gamma x)(U(\infty) - U(t))\big)}{2\bar{F}\big(U(\infty) - (U(\infty) - U(t))\big)} \ge \frac{1}{4} g^{1/(1-\epsilon/2)}(x). \tag{3.8}$$

Finally, consider the case $\gamma = 0$. Then

$$w_t(x)g(x) \le \mathrm{e}^{-\epsilon x} \le \exp\big(-\epsilon g_t^{\leftarrow}(|A(t)|^c)\big) = g^{\epsilon}\big(g_t^{\leftarrow}(|A(t)|^c)\big)$$

for all $g_t(x) \le |A(t)|^c$. Apply (3.5) with $v = |A(t)|^c$ and $\epsilon = 1/2$ to obtain

$$\frac{g(g_t^{\leftarrow}(v))}{A(t)} = \Big|\frac{v}{A(t)} - vK_{0,0}(1/v)\Big| + o(v^{1/2}) = O\big(|A(t)|^{c-1} + |A(t)|^{c/2}\big).$$

Thus

$$\frac{w_t(x)g(x)}{A(t)} \le |A(t)|^{\epsilon-1}\Big|\frac{g(g_t^{\leftarrow}(|A(t)|^c))}{A(t)}\Big|^{\epsilon} = O\big(|A(t)|^{\epsilon(c-1)+\epsilon-1} + |A(t)|^{\epsilon-1+c\epsilon/2}\big) \to 0,$$

and the proof is complete. □

Proof of Proposition 3.2:

As in the proof of Proposition 3.1, let $v := g_t(x)$. We consider three cases.

Case (i): $\rho < 0$.

Inequality (3.5) and Corollary 3.1 imply

$$\sup_{x\in D_{t,\rho}}\Big|\frac{g(x)}{g_t(x)} - 1\Big| \le \sup_{v\in B_{t,\rho}}\frac{A(t)}{v}\big|v^{1+\gamma}K_{\gamma,\rho}(1/v) + v^{1-\rho}\mathrm{e}^{\epsilon|\log v|}o(1)\big| \to 0. \tag{3.9}$$

Hence, for $\gamma + \rho \ne 0$, by the definition of $K_{\gamma,\rho}$

$$\begin{aligned}&\sup_{x\in D_{t,\rho}} w_t(x)\Big|g^{1+\gamma}(x)K_{\gamma,\rho}(1/g(x)) - g_t^{1+\gamma}(x)K_{\gamma,\rho}(1/g_t(x))\Big|\\ &= \sup_{x\in D_{t,\rho}} \mathrm{e}^{-\epsilon|\log g_t(x)|}\frac{1}{|\gamma+\rho|}\Big|\Big(\frac{g(x)}{g_t(x)}\Big)^{1-\rho} - 1\Big| \\ &\to 0.\end{aligned} \tag{3.10}$$

If $\gamma + \rho = 0$, then the left-hand side of (3.10) equals

$$\sup_{x\in D_{t,\rho}} \mathrm{e}^{-\epsilon|\log g_t(x)|}\Big|\Big(\frac{g(x)}{g_t(x)}\Big)^{1+\gamma}\log\Big(\frac{g(x)}{g_t(x)}\Big) + \Big(\Big(\frac{g(x)}{g_t(x)}\Big)^{1+\gamma} - 1\Big)\log g_t(x)\Big| \to 0.$$

Now (3.1) is immediate from Proposition 3.1. By (3.9), $w(x)/w_t(x)$ tends to 1 uniformly for $x \in D_{t,\rho}$. Moreover, $g(x) \le ct^{1-\delta}$ implies $g_t(x) \le 2ct^{1-\delta}$ and so $\tilde{D}_{t,\rho,\delta,c} \subset D_{t,\rho,\delta,2c}$ for sufficient large t. Thus (3.2) follows immediately from (3.1).

Case (ii): $\rho = 0$, $\gamma \ne 0$.

Define

$$D_{t,0}^1 := \big\{x : |A(t)|^c \le g_t(x) \le |A(t)|^{-c}\big\} = \{x : g_t(x) \in B_{t,0}\},$$

$$D_{t,0}^2 := \big\{x : g_t(x) \le |A(t)|^c\big\},$$

so that $D_{t,0} = D^1_{t,0} \cup D^2_{t,0}$. As in case (i), (3.5) and Corollary 3.1 imply

$$\sup_{x \in D^1_{t,0}} \Big| \frac{g(x)}{g_t(x)} - 1 \Big| \le \sup_{v \in B_{t,0}} \frac{A(t)}{v} \big| v^{1+\gamma} K_{\gamma,0}(1/v) + v \mathrm{e}^{\epsilon |\log v|} o(1) \big| \to 0. \tag{3.11}$$

Hence

$$\begin{aligned} &\sup_{x \in D^1_{t,0}} w_t(x) \Big| g^{1+\gamma}(x) K_{\gamma,0}(1/g(x)) - g_t^{1+\gamma}(x) K_{\gamma,0}(1/g_t(x)) \Big| \\ &= \frac{1}{|\gamma|} \sup_{x \in D^1_{t,0}} \mathrm{e}^{-\epsilon |\log g_t(x)|} \Big| \frac{g(x)}{g_t(x)} \log \frac{g(x)}{g_t(x)} + \frac{g(x)}{g_t(x)} \log g_t(x) \Big| \qquad (3.12) \\ &\to 0. \end{aligned}$$

Note that

$$\sup_{x \in D^2_{t,0}} w_t(x) \big| g_t^{1+\gamma}(x) K_{\gamma,0}(1/g_t(x)) \big| = \sup_{x \in D^2_{t,0}} \frac{1}{|\gamma|} \mathrm{e}^{-\epsilon |\log g_t(x)|} |\log g_t(x)| \to 0. \tag{3.13}$$

Moreover, by (3.7) and (3.8),

$$\begin{aligned} w_t(x) |g^{1+\gamma}(x) K_{\gamma,0}(1/g(x))| &= \frac{1}{|\gamma|} g_t^{\epsilon-1}(x) g(x) |\log g(x)| \\ &= O\big(g^{(\epsilon-1)/(1-\epsilon/2)+1}(x) |\log g(x)| \big) \\ &= O\big(g^{\epsilon/(2-\epsilon)}(x) |\log g(x)| \big) \\ &\to 0 \qquad (3.14) \end{aligned}$$

uniformly for $x \in D^2_{t,0}$, which proves (3.1) in this case.

Recall from (3.11) that $g_t/g \to 1$ uniformly for $x \in D^1_{t,0}$ for all $c > 0$. In particular, one has for sufficiently large t and $x = g^{\leftarrow}(|A(t)|^{-c})$ that $g_t(x) \le 2|A(t)|^{-c} \le |A(t)|^{-2c}$. Since g and g_t are decreasing functions, it follows that

$$\tilde{D}_{t,0} = \tilde{D}_{t,0,1,c} = \{x : g(x) \le |A(t)|^{-c}\} \subset \{x : g_t(x) \le |A(t)|^{-2c}\} = D_{t,0,1,2c}.$$

Hence (3.1) implies

$$\sup_{x \in \tilde{D}_{t,0}} (g_t(x))^{-1} \exp(-\eta |\log g_t(x)|) \Big| \frac{g_t(x) - g(x)}{A(t)} - g^{1+\gamma}(x) K_{\gamma,0}(1/g(x)) \Big| \to 0$$

for all $\eta > 0$.

It remains to prove that, for sufficiently small $\eta > 0$, $w/((g_t(x))^{-1} \exp(-\eta |\log g_t(x)|))$ is uniformly bounded on $\tilde{D}_{t,0}$. In view of (3.11), the boundedness holds uniformly on $D^1_{t,0,1,2c}$. On the other hand, similarly as in (3.7) and (3.8), the Potter bounds yield $g_t(x) \le 2g^{1-\epsilon/2}(x)$ for sufficiently large t and all $x \in D^2_{t,0}$. Therefore,

$$\frac{w(x)}{(g_t(x))^{-1} \exp(-\eta |\log g_t(x)|)} \le 2 g^{\epsilon - 1 - (\eta-1)(1-\epsilon/2)}(x) \to 0 \tag{3.15}$$

uniformly for $x \in D^2_{t,0}$ if $\eta < \epsilon/2$, and (3.2) is proved in case (ii).

Case (iii): $\gamma = \rho = 0$.

The convergences (3.11) and (3.12) can be established as in the case (ii). Moreover, because $g(x) \le g(g_t^{\leftarrow}(|A(t)|^c)) \to 0$ by (3.11), one has uniformly for $x \in D^2_{t,0}$,

$$\begin{aligned} w_t(x)|g_t(x)K_{0,0}(1/g_t(x))| &\le g_t^{\epsilon}(x)\log^2 g_t(x) \to 0 \\ w_t(x)|g(x)K_{0,0}(1/g(x))| &\le g^{\epsilon}(x)\log^2 g(x) \to 0, \end{aligned}$$

which proves (3.1). Assertion (3.3) follows by the very same arguments as in the case (ii). $\square$

The following approximations and bounds to g_t are direct consequences of the above propositions.

Corollary 3.2. *Suppose $x_0 > -1/(\gamma \vee 0)$.*

(i) If $\rho < 0$, then

$$\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} \left| \frac{g(x)}{g_t(x)} - 1 \right| \;\to\; 0.$$

(ii) If $\rho = 0$ and $\gamma \ne 0$, then for all $\eta > 0$

$$\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} \frac{|g_t(x) - g(x)|}{g^{1-\eta}(x)} \;\to\; 0$$

as $t \to \infty$ and thus

$$\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} \frac{g_t(x)}{g^{1-\eta}(x)} \qquad \textit{is bounded.}$$

(iii) If $\gamma = \rho = 0$, then for all $\eta, c > 0$

$$\sup_{x_0 \le x < -c\log|A(t)|} \frac{|g_t(x) - g(x)|}{g^{1-\eta}(x)} \;\to\; 0$$

as $t \to \infty$ and so

$$\sup_{x_0 \le x < -c\log|A(t)|} \frac{g_t(x)}{g^{1-\eta}(x)} \qquad \textit{is bounded.}$$

Proof:

(i) By (3.9), one has for all $\delta \in (0,1)$ and $c > 0$

$$\left[x_0,\ \frac{1}{(-\gamma)\vee 0}\right) \subset \left\{x:\ g(x) \le \frac{c}{2}t^{1-\delta}\right\} \subset \{x : g_t(x) \le ct^{1-\delta}\} = D_{t,\rho}$$

for sufficiently large t. Hence, again by (3.9),

$$\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} \Big| \frac{g(x)}{g_t(x)} - 1 \Big| \to 0.$$

(ii) By a similar argument as in (i), one concludes $[x_0, 1/((-\gamma) \vee 0)) \subset D_{t,0}$. Hence Proposition 3.2 with $\epsilon = \eta$ implies

$$\begin{aligned}
&\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} \frac{1}{A(t)} \Big| \frac{g_t(x) - g(x)}{g^{1-\eta}(x)} - A(t) g^{\gamma+\eta}(x) K_{\gamma,0}(1/g(x)) \Big| \\
&= \sup_{x_0 \le x < 1/((-\gamma)\vee 0)} g^{\eta-1}(x) \Big| \frac{g_t(x) - g(x)}{A(t)} - g^{1+\gamma}(x) K_{\gamma,0}(1/g(x)) \Big| \\
&\to\ 0.
\end{aligned}$$

Because $A(t) \to 0$ and $g(x)$ is bounded for $x \ge x_0$, the assertions are immediate from the definition of $K_{\gamma,0}$.

(iii) The proof is similar to the one of (ii). Note that $w_t(x)/\mathrm{e}^{(1-\epsilon)x} \to 1$ uniformly for $x_0 \le x < -c \log |A(t)|$ by (3.11). □

4 Tail Approximation to the Empirical Distribution Function

For the proof of Theorem 2.1, we need an additional lemma.

Lemma 4.1. *Let W denote a Brownian motion.*

(i) If $\gamma \neq 0$ or $\rho < 0$, then

$$\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} g^{-1/2+\epsilon}(x) |W(g_t(x)) - W(g(x))| \to 0 \quad a.s.$$

as $t \to \infty$.

(ii) If $\rho = \gamma = 0$, then

$$\sup_{x_0 \le x} \big(\max(g(x), g_t(x)) \big)^{-1/2+\epsilon} |W(g_t(x)) - W(g(x))| \to 0 \quad \text{a.s.}$$

as $t \to \infty$.

Proof:

(i) By Corollary 3.2 $g_t(x) - g(x) \to 0$ uniformly for $x_0 \le x < 1/((-\gamma) \vee 0)$. Using Levy's modulus of continuity of the Brownian motion (see Csörgő and Horváth (1993), Theorem A.1.2) one gets for sufficiently large t

$$\begin{aligned}
g^{-1/2+\epsilon}(x) |W(g_t(x)) - W(g(x))| &\le 2 g^{-1/2+\epsilon}(x) \big| (g_t(x) - g(x)) \log |g_t(x) - g(x)| \big|^{1/2} \\
&\le 2 \Big| \frac{g_t(x) - g(x)}{g^{(1-2\epsilon)/(1-\epsilon)}(x)} \Big|^{(1-\epsilon)/2} \\
&\to 0 \qquad \text{a.s.}
\end{aligned}$$

again by Corollary 3.2.

(ii) As in (i), one can conclude from Corollary 3.2 that

$$\sup_{x_0 \le x < -c \log |A(t)|} g^{-1/2+\epsilon}(x) |W(g_t(x)) - W(g(x))| \to 0 \quad \text{a.s.}$$

Moreover, the law of the iterated logarithm yields

$$g^{-1/2+\epsilon}(x)|W(g(x))| = O\big((\log\log(1/g(x)))^{1/2} g^{\epsilon}(x)\big) \to 0 \quad \text{a.s.}$$

uniformly on $\{x : x > -c \log |A(t)|\} = \{x : g(x) < |A(t)|^c\}$. Since $g_t(x) \to 0$ uniformly on this set, likewise one obtains

$$g_t^{-1/2+\epsilon}(x)|W(g_t(x))| \to 0 \quad \text{a.s.,}$$

and hence assertion (ii). □

Proof of Theorem 2.1:

We focus on the case $\gamma \neq 0$ or $\rho < 0$, because the other case can be treated similarly. We have to prove that the following expression tends to 0 uniformly for $x_0 \le x < 1/((-\gamma) \vee 0)$:

$$\begin{aligned}
I \quad := \quad & g^{-1/2+\epsilon}(x) \Big| \sqrt{k_n} \left[\frac{n}{k_n} \bar{F}_n \Big(a(\frac{n}{k_n})x + b(\frac{n}{k_n}) \Big) - g(x) \right] \\
& \qquad -W_n(g(x)) - \sqrt{k_n} A\big(\frac{n}{k_n}\big) g^{1+\gamma}(x) K_{\gamma,\rho}(1/g(x)) \Big| \\
\le \quad & \frac{g^{-1/2+\epsilon}(x)}{g_{n/k_n}^{-1/2+\epsilon/2}(x)} g_{n/k_n}^{-1/2+\epsilon/2}(x) \Big| \sqrt{k_n} \Big[\frac{n}{k_n} \bar{F}_n \big(a(\frac{n}{k_n})x + b(\frac{n}{k_n}) \big) - g_{n/k_n}(x) \Big] - W_n(g_{n/k_n}(x)) \Big| \\
& + \frac{g^{-1/2+\epsilon}(x)}{w(x)} w(x) \sqrt{k_n} A\big(\frac{n}{k_n}\big) \Big| \frac{g_{n/k_n}(x) - g(x)}{A(n/k_n)} - g^{1+\gamma}(x) K_{\gamma,\rho}(1/g(x)) \Big| \\
& + g^{-1/2+\epsilon}(x) |W_n(g_{n/k_n}(x)) - W_n(g(x))| \\
=: \quad & I_1 + I_2 + I_3.
\end{aligned}$$

By (2.2) (with a and b instead of $\tilde{a}$ and $\tilde{b}$ such that $z_n(x)$ is replaced with $g_{n/k_n}(x)$) and Corollary 3.2, one has $\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} I_1 \xrightarrow{P} 0$. From Proposition 3.2 and the fact that $g^{-1/2+\epsilon}(x)/w(x) \le g^{1/2-\rho}(x)$ is bounded uniformly for $x_0 \le x < 1/((-\gamma)\vee 0)$, it follows that $\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} I_2 \to 0$. Finally Lemma 4.1 shows that $\sup_{x_0 \le x < 1/((-\gamma)\vee 0)} I_3 \xrightarrow{P} 0$. □

Proof of Corollary 2.1:

Because of Theorem 2.1(ii) and $\max(1, x^\tau) = o(\mathrm{e}^{(1/2-\epsilon)x})$ as $x \to \infty$ for all $\tau > 0$ and $\epsilon \in (0, 1/2)$, it suffices to prove that $\sup_{x_0 \le x < \infty} g_{n/k_n}^{1/2-\epsilon}(x) \max(1, x^\tau) = O(1)$.

According to Lemma 2.2 of Resnick (1987), there exists a function $\bar{a}$ such that $a(t)/\bar{a}(t) \to 1$ as $t \to \infty$ and $F^t\big(\bar{a}(t)x + b(t)\big) \geq 1 - (1+\delta)^3(1+\delta x)^{-1/\delta}$ for all $\delta > 0$, sufficiently large t and $x \geq x_0$. Thus, by the mean value theorem, there exists $\theta_{t,x} \in (0,1)$ such that

$$
\begin{aligned}
g_t(x) = t\bar{F}\big(\bar{a}(t)x + b(t)\big) &\leq t\Big(1 - \Big(1 - (1+\delta)^3(1+\delta x)^{-1/\delta}\Big)^{1/t}\Big) \\
&= (1+\delta)^3(1+\delta x)^{-1/\delta}\Big(1 - \theta_{t,x}(1+\delta)^3(1+\delta x)^{-1/\delta}\Big)^{1/t-1} \\
&\leq 2(1+\delta x)^{-1/\delta}
\end{aligned}
$$

if $x \geq 0$ and $\delta > 0$ is sufficiently small. Since by the locally uniform convergence in (1.2) $\sup_{x_0 \leq x < 0} g_{n/k_n}^{1/2-\epsilon}(x)\max(1, x^\tau) = O(1)$, it follows that

$$
\sup_{x_0 \leq x < \infty} g_{n/k_n}^{1/2-\epsilon}(x)\max(1, x^\tau) = O(1) + 2\sup_{0 \leq x < \infty}\Big(1 + \delta x\Big)^{-(1/2-\epsilon)/\delta}\max(1, x^\tau) = O(1)
$$

if δ is chosen smaller than $(1/2 - \epsilon)/\tau$. $\square$

5 Tail Empirical Process With Estimated Parameters: Proofs

In this section we prove the approximation to the tail empirical process with estimated parameters stated in Proposition 2.1 and the limit theorem 2.2 for the test statistic T_n. In the sequel, we will use the abbreviations $a := a(n/k)$, $\hat{a} := \hat{a}(n/k)$, $b := b(n/k)$ and $\hat{b} := \hat{b}(n/k)$ wherever this is convenient. Let

$$
g_{a,b,\gamma}(x) := \Big(1 + \gamma\frac{x-b}{a}\Big)^{-1/\gamma}, \qquad g_{a,b,\gamma}^{\leftarrow}(x) := a\frac{x^{-\gamma}-1}{\gamma} + b \tag{5.1}
$$

and

$$
y_n(x) := g_{a,b,\gamma}(g_{\hat{a},\hat{b},\hat{\gamma}}^{\leftarrow}(x)).
$$

According to Theorem 2.1,

$$
\sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n(g_{a,b,\gamma}^{\leftarrow}(v)) - v\Big] \approx W_n(v) + \sqrt{k_n}A\big(\frac{n}{k_n}\big)v^{1+\gamma}K_{\gamma,\rho}(1/v)
$$

with respect to a suitable weighted supremum norm. In particular, for $v = y_n(x)$ the left-hand side equals $k_n^{1/2}\big(n/k_n\bar{F}_n(\hat{a}(x^{-\hat{\gamma}_n}-1)/\hat{\gamma}_n+\hat{b}) - y_n(x)\big)$, that is the first term in (2.7) minus $k_n^{1/2}(y_n(x)-x)$. It is easily seen that $y_n(x)$ converges x pointwise. Refined calculations show that $k_n^{1/2}(y_n(x) - x) \to L_n^{(\gamma)}(x)$. Hence $W_n(y_n(x)) + k_n^{1/2}A(n/k_n)(y_n(x))^{1+\gamma}K_{\gamma,\rho}(1/y_n(x))$ can be approximated by $W_n(x) + k_n^{1/2}A(n/k_n)x^{1+\gamma}K_{\gamma,\rho}(1/x)$. So the main problem in the proof of Proposition 2.1 is to show that these approximations hold uniformly in a suitable sense.

We start with the uniform approximation of $y_n(x)$ which will serve as the basis for all the other approximations. Define

$$\begin{aligned} A_{n,k_n} &:= \frac{\hat{a}(n/k_n)}{a(n/k_n)}, \\ B_{n,k_n} &:= \frac{\hat{b}(n/k_n) - b(n/k_n)}{a(n/k_n)} \end{aligned}$$

so that

$$y_n(x) = \Big(1 + \gamma\big(B_{n,k_n} + A_{n,k_n}\frac{x^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n}\big)\Big)^{-1/\gamma}.$$

Recall from (2.6) that

$$\begin{aligned} A_{n,k_n} &= 1 + k_n^{-1/2}\alpha(W_n) + o_P(k_n^{-1/2}), \\ B_{n,k_n} &= k_n^{-1/2}\beta(W_n) + o_P(k_n^{-1/2}), \\ \hat{\gamma}_n &= \gamma + k_n^{-1/2}\Gamma(W_n) + o_P(k_n^{-1/2}). \end{aligned} \tag{5.2}$$

Lemma 5.1. *Suppose (5.2) holds. Let $\lambda_n > 0$ be such that $\lambda_n \to 0$, and $k_n^{-1/2}\lambda_n^{\gamma} \to 0$ if $\gamma < 0$, or $k_n^{-1/2}\log^2\lambda_n \to 0$ if $\gamma = 0$.*
(i) If $\gamma > 0$ then, for all $\epsilon > 0$, $x^{-1/2+\epsilon}\big(\sqrt{k_n}(y_n(x) - x) - L_n^{(\gamma)}(x)\big) \xrightarrow{P} 0$, and $x^{\epsilon-1}(y_n(x) - x) \xrightarrow{P} 0$ as $n \to \infty$ uniformly for $x \in (0,1]$.
(ii) If $-1/2 < \gamma \le 0$ then, for all $\epsilon > 0$, $x^{-1/2+\epsilon}\big(\sqrt{k_n}(y_n(x)-x)-L_n^{(\gamma)}(x)\big) \xrightarrow{P} 0$ and $(y_n(x) - x)/x \xrightarrow{P} 0$ as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

Proof: For $\gamma \neq 0$, define $\delta_n := 1 + \gamma B_{n,k_n} - A_{n,k_n}\gamma/\hat{\gamma}_n$, and $\Delta_n := \Delta_{n,x} := \delta_n\hat{\gamma}_n/(\gamma A_{n,k_n}x^{-\hat{\gamma}_n})$, so that $\delta_n = O_P(k_n^{-1/2})$.

(i) By the mean value theorem there exist $\theta_{n,x} \in (0,1)$ such that

$$\begin{aligned} y_n(x) &= \Big(1 + \gamma\big(B_{n,k_n} + A_{n,k_n}\frac{x^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n}\big)\Big)^{-1/\gamma} \\ &= \Big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n} + \delta_n\Big)^{-1/\gamma} \\ &= \Big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}\Big)^{-1/\gamma}(1 + \Delta_n)^{-1/\gamma} \\ &= \big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}\big)^{-1/\gamma} - \frac{1}{\gamma}\Big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}(1 + \theta_{n,x}\Delta_n)\Big)^{-1/\gamma-1}\delta_n \\ &= \big(\frac{\gamma}{\hat{\gamma}_n}A_{n,k_n}x^{-\hat{\gamma}_n}\big)^{-1/\gamma} - \frac{1}{\gamma}x^{\hat{\gamma}_n(1/\gamma+1)}\delta_n(1 + o_P(1)) \end{aligned} \tag{5.3}$$

where the $o_P(1)$-term tends to 0 uniformly for $x \in (0,1]$. Hence again by the mean value theorem and (5.2), for some $\theta_{n,x} \in (0,1)$,

$$
\begin{aligned}
&y_n(x) - x \\
&= \Big(\big(\frac{\gamma}{\hat\gamma_n} A_{n,k_n} \big)^{-1/\gamma} - 1 \Big) x^{\hat\gamma_n/\gamma} + (x^{\hat\gamma_n/\gamma} - x) - \frac{1}{\gamma} x^{\hat\gamma_n(1/\gamma+1)} \delta_n (1 + o_p(1)) \\
&= -\frac{1}{\gamma}(1+o_p(1)) \big(\frac{\gamma}{\hat\gamma_n} A_{n,k_n} - 1 \big) x^{\hat\gamma_n/\gamma} + x^{1+\theta_{n,x}(\hat\gamma_n/\gamma-1)} \log x \big(\frac{\hat\gamma_n}{\gamma} - 1 \big) - \frac{1}{\gamma} x^{\hat\gamma_n(1/\gamma+1)} \delta_n (1+o_p(1)) \\
&= \frac{1}{\gamma}(1+o_p(1)) x^{\hat\gamma_n/\gamma} \Big(\frac{\hat\gamma_n - \gamma}{\hat\gamma_n} A_{n,k_n} - (A_{n,k_n} - 1) \Big) + x^{1+\theta_{n,x}(\hat\gamma_n/\gamma-1)} \log x \frac{\hat\gamma_n - \gamma}{\gamma} \\
&\qquad - \frac{1}{\gamma} x^{\hat\gamma_n(1/\gamma+1)} \Big(\gamma B_{n,k_n} + \frac{1}{\hat\gamma_n}(\hat\gamma_n - \gamma) - \frac{\gamma}{\hat\gamma_n}(A_{n,k_n} - 1) \Big)(1 + o_p(1)). \qquad (5.4)
\end{aligned}
$$

Now the first assertion is a straightforward consequence of (5.2). For example,

$$
\begin{aligned}
&x^{-1/2+\epsilon} \sqrt{k_n} \frac{1}{\gamma}(1+o_P(1)) x^{\hat\gamma_n/\gamma} \Big(\frac{\hat\gamma_n - \gamma}{\hat\gamma_n} A_{n,k_n} - (A_{n,k_n} - 1) \Big) \\
&= \frac{1}{\gamma} x^{-1/2+\epsilon} \exp\big((\hat\gamma_n/\gamma - 1) \log x \big) x \Big(\sqrt{k_n} \frac{\hat\gamma_n - \gamma}{\hat\gamma_n} A_{n,k_n} - \sqrt{k_n}(A_{n,k_n} - 1) \Big)(1 + o_P(1)) \\
&= \frac{1}{\gamma} x^{-1/2+\epsilon} x \Big(\frac{\Gamma(W_n)}{\gamma} - \alpha(W_n) \Big)(1 + o_P(1))
\end{aligned}
$$

uniformly for $x \in (0,1]$.

Moreover, in view of (5.4),

$$
\begin{aligned}
x^{\epsilon-1}(y_n(x) - x) &= -\frac{1}{\gamma}(1+o_p(1)) \big(\frac{\gamma}{\hat\gamma_n} A_{n,k_n} - 1 \big) x^{\hat\gamma_n/\gamma - 1 + \epsilon} + x^{\epsilon + \theta_{n,x}(\hat\gamma_n/\gamma - 1)} \log x \big(\frac{\hat\gamma_n}{\gamma} - 1 \big) \\
&\qquad - \frac{1}{\gamma} x^{\hat\gamma_n - 1 + \epsilon + \hat\gamma_n/\gamma} \delta_n (1 + o_p(1)) \\
&\xrightarrow{P} 0
\end{aligned}
$$

as $n \to \infty$ uniformly for $x \in (0,1]$.

(ii) First we consider the case $\gamma = 0$. Then

$$
\begin{aligned}
&y_n(x) - x \\
&= \exp\Big(- \big(B_{n,k_n} + A_{n,k_n} \frac{x^{-\hat\gamma_n} - 1}{\hat\gamma_n} \big) \Big) - x \qquad (5.5) \\
&= x \Big(\exp\Big(- \big(B_{n,k_n} + A_{n,k_n} \big(\frac{x^{-\hat\gamma_n} - 1}{\hat\gamma_n} + \log x \big) - (A_{n,k_n} - 1) \log x \big) \Big) - 1 \Big).
\end{aligned}
$$

An application of the mean value theorem to $\gamma \mapsto x^{-\gamma}$ together with (5.2) yields

$$
\frac{x^{-\hat\gamma_n} - 1}{\hat\gamma_n} = -\log x + \frac{1}{2} \hat\gamma_n \log^2 x \exp\big(-\theta_{n,x} \hat\gamma_n \log x \big) \qquad (5.6)
$$

for some $\theta_{n,x} \in (0,1)$. It follows that

$$
\frac{x^{-\hat\gamma_n} - 1}{\hat\gamma_n} + \log x \xrightarrow{P} 0
$$

as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$, since then $k_n^{-1/2}\log^2 x \le k_n^{-1/2}\log^2 \lambda_n \to 0$ and likewise $\hat\gamma_n \log x \xrightarrow{P} 0$. Hence

$$B_{n,k_n} + A_{n,k_n}\big(\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n} + \log x\big) - (A_{n,k_n}-1)\log x \xrightarrow{P} 0,$$

and by (5.5), (5.2) and (5.6)

$$\begin{aligned} & x^{-1/2+\epsilon}\sqrt{k_n}(y_n(x)-x) \\ = \;& -x^{-1/2+\epsilon}x\Big(\sqrt{k_n}B_{n,k_n} + \sqrt{k_n}A_{n,k_n}\big(\frac{x^{-\hat\gamma_n}-1}{\hat\gamma_n} + \log x\big) - \sqrt{k_n}(A_{n,k_n}-1)\log x\Big)(1+o_P(1)) \\ = \;& -x^{-1/2+\epsilon}x\Big(\beta(W_n) + \frac{1}{2}\Gamma(W_n)\log^2 x - \alpha(W_n)\log x\Big)(1+o_P(1)) \end{aligned}$$

uniformly for $x \in [\lambda_n, 1]$, that is, the first assertion.

Likewise one concludes from (5.5), (5.2) and (5.6) that $(y_n(x)-x)/x$ tends to 0 uniformly for $x \in [\lambda_n, 1]$.

Next assume $-1/2 < \gamma < 0$.
Because $\delta_n = O_P(k_n^{-1/2})$ and, by the definition of λ_n and (5.2),

$$k_n^{-1/2}x^{\hat\gamma_n} \le k_n^{-1/2}\lambda_n^{\hat\gamma_n} = o_P\Big(\exp\big(\log\lambda_n(\hat\gamma_n - \gamma)\big)\Big) = o_P(1),$$

$\Delta_n \to 0$ in probability uniformly for $x \in [\lambda_n, 1]$. Therefore, the first assertion can be established as in the case $\gamma > 0$.

Furthermore, according to (5.4),

$$\begin{aligned} \frac{y_n(x)-x}{x} &= -\frac{1}{\gamma}(1+o_p(1))\big(\frac{\gamma}{\hat\gamma_n}A_{n,k_n}-1\big)x^{\hat\gamma_n/\gamma-1} + x^{\theta_{n,x}(\hat\gamma_n/\gamma-1)}\log x\big(\frac{\hat\gamma_n}{\gamma}-1\big) \\ &\quad -\frac{1}{\gamma}x^{\hat\gamma_n+\hat\gamma_n/\gamma-1}\delta_n(1+o_p(1)) \\ &= O_P(k^{-1/2})x^{-\epsilon} \\ &\xrightarrow{P} 0 \end{aligned}$$

as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$ by the choice of λ_n, if $0 < \epsilon < -\gamma$. □

Next we examine the expressions $W_n(y_n(x))$ and $(y_n(x))^{1+\gamma}K_{\gamma,\rho}(1/y_n(x))$.

Lemma 5.2. *Under the conditions of Lemma 5.1 one has for all $\epsilon > 0$:*
(i) If $\gamma > 0$, then $x^{-1/2+\epsilon}\big(W_n(y_n(x)) - W_n(x)\big) \xrightarrow{P} 0$ as $n \to \infty$ uniformly for $x \in (0, 1]$.
(ii) If $-1/2 < \gamma \le 0$, then $x^{-1/2+\epsilon}\big(W_n(y_n(x)) - W_n(x)\big) \xrightarrow{P} 0$ as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

Proof:

Similarly as in the proof of Lemma 4.1, the assertion follows from Lemma 5.1 using Levy's modulus of continuity of the Brownian motion:

$$x^{-1/2+\epsilon}|W_n(y_n(x)) - W_n(x)| = O\Big(x^{-1/2+\epsilon}\big|(y_n(x)-x)\log|y_n(x)-x|\big|^{1/2}\Big) \xrightarrow{P} 0$$

uniformly over the ranges of x-values specified in the assertion. $\square$

Lemma 5.3. *Under the conditions of Lemma 5.1 one has for all $\epsilon > 0$:*
(i) For $\gamma > 0$

$$x^{-1/2+\epsilon}\Big((y_n(x))^{\gamma+1}K_{\gamma,\rho}(1/y_n(x)) - x^{\gamma+1}K_{\gamma,\rho}(1/x)\Big) \xrightarrow{P} 0$$

as $n \to \infty$ uniformly for $x \in (0,1]$.
(ii) For $-1/2 < \gamma \le 0$

$$x^{-1/2+\epsilon}\Big((y_n(x))^{\gamma+1}K_{\gamma,\rho}(1/y_n(x)) - x^{\gamma+1}K_{\gamma,\rho}(1/x))\Big) \xrightarrow{P} 0$$

as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

Proof:

(i) We only consider the case $\gamma > 0 = \rho$; the assertion can be proved similarly in the case $\gamma > 0 > \rho$. Equation (5.3) implies

$$\log\frac{y_n(x)}{x} = O\Big(\log\Big(\frac{\gamma}{\hat\gamma_n}A_{n,k_n}\Big)\Big) + O\Big(\Big(\frac{\hat\gamma_n}{\gamma}-1\Big)\log x\Big) + O(\Delta_{n,x}) = O_P\big(k_n^{-1/2}(1+|\log x|)\big)$$

uniformly for $x \in (0,1]$. Hence, by the definition of $K_{\gamma,0}$ and Lemma 5.1(i),

$$\begin{aligned}
&x^{-1/2+\epsilon}\big((y_n(x))^{\gamma+1}K_{\gamma,0}(1/y_n(x)) - x^{\gamma+1}K_{\gamma,0}(1/x)\big)\\
&= x^{-1/2+\epsilon}\Big(-\frac{y_n(x)\log(y_n(x))}{\gamma} + \frac{x\log x}{\gamma}\Big)\\
&= -\frac{1}{\gamma}\Big(x^{-1/2+\epsilon}y_n(x)\log\frac{y_n(x)}{x} + x^{-1/2+\epsilon}(y_n(x)-x)\log x\Big)\\
&= -\frac{1}{\gamma}\Big(x^{\epsilon-1}y_n(x)x^{1/2}O_P\big(k_n^{-1/2}(1+|\log x|)\big) + x^{\epsilon-1}(y_n(x)-x)x^{1/2}\log x\Big)\\
&\xrightarrow{P} 0
\end{aligned}$$

as $n \to \infty$ uniformly for $x \in (0,1]$.

(ii) In the case $\gamma = 0 > \rho$, according to the definition of $K_{0,\rho}$, Lemma 5.1(ii) and the mean value theorem, there exists $\theta_{n,x} \in (0,1)$ such that

$$\begin{aligned} x^{-1/2+\epsilon}\Big(y_n(x)K_{0,\rho}(1/y_n(x)) - xK_{0,\rho}(1/x)\Big) &= x^{-1/2+\epsilon}\Big(\frac{(y_n(x))^{1-\rho}}{\rho} - \frac{x^{1-\rho}}{\rho}\Big) \\ &= \frac{1-\rho}{\rho}x^{1/2+\epsilon}\frac{y_n(x)-x}{x}\big(x+\theta_{n,x}(y_n(x)-x)\big)^{-\rho} \\ &\xrightarrow{P} 0 \end{aligned}$$

as $n \to \infty$ uniformly for $x \in [\lambda_n, 1]$.

Likewise, the assertion can be proved in the other cases . $\square$

Remark 5.1. *The assertions of Lemma 5.1(ii) with weight function $x^{\epsilon-1-\gamma}$ instead of $x^{-1/2+\epsilon}$, and of the Lemmas 5.2(ii) and 5.3(ii) also hold true for $-1 < \gamma \le 0$.*

Lemma 5.4. *Suppose $p_n \to 0$, $np_n \to 0$, and $k_n^{-1/2}\log^2(np_n) \to 0$ as $n \to \infty$. Define $g^{\leftarrow}_{a,b,\gamma}$ as in* (5.1). *Then, under the conditions of Proposition 2.1 for $-1/2 < \gamma \le 0$, $P\{g^{\leftarrow}_{\hat a,\hat b,\hat\gamma_n}(np_n/k_n) \le X_{n,n}\} \to 0$ as $n \to \infty$.*

Proof: According to Theorem 1 of de Haan and Stadtmüller (1996), one has

$$\frac{\frac{a(tx)}{x^\gamma a(t)} - 1}{A(t)} \to \frac{x^\rho - 1}{\rho}$$

as $t \to \infty$. By similar arguments as used by Drees (1998) and Cheng and Jiang (2001) it follows that, for all $0 < \epsilon < 1/2$, there exists $t_\epsilon > 0$ such that for all $t \ge t_\epsilon$ and $x \ge 1$

$$\left|\frac{\frac{a(tx)}{x^\gamma a(t)} - 1}{A(t)} - \frac{x^\rho - 1}{\rho}\right| \le \epsilon x^{\rho+\epsilon}.$$

Hence

$$\frac{a(n)}{k_n^\gamma a(n/k_n)} = 1 + A\Big(\frac{n}{k_n}\Big)\frac{k_n^\rho - 1}{\rho} + o\Big(A\Big(\frac{n}{k_n}\Big)k_n^{\rho+\epsilon}\Big) \to 1, \tag{5.7}$$

because $\rho \le 0$ and $\sqrt{k_n}A(n/k_n) = O(1)$.

Now we distinguish two cases.

Case (i): $-1/2 < \gamma < 0$.

Then

$$\begin{aligned} \frac{g^{\leftarrow}_{\hat a,\hat b,\hat\gamma_n}(np_n/k_n) - X_{n,n}}{a(n/k_n)} &= -\frac{1}{\gamma}\Big(\frac{\hat a}{a} - 1\Big) + \frac{1}{\hat\gamma_n}\frac{\hat a}{a}\Big(\frac{k_n}{np_n}\Big)^{\hat\gamma_n} + \frac{\hat a}{a}\Big(\frac{1}{\gamma} - \frac{1}{\hat\gamma_n}\Big) \\ &\quad + \frac{\hat b - b}{a} - \Big(\frac{b(n) - b(n/k_n)}{a(n/k_n)} + \frac{1}{\gamma}\Big) - \frac{X_{n,n} - b(n)}{a(n)} \cdot \frac{a(n)}{a(n/k_n)} \\ &=: T_1 + T_2 + T_3 + T_4 - T_5 - T_6. \end{aligned}$$

Assumption (5.2) implies $T_1 + T_3 + T_4 = O_P(k_n^{-1/2}) = o_p(k_n^\gamma)$ and

$$T_2 = O_P\Big(\Big(\frac{k_n}{np_n}\Big)^\gamma \exp\Big((\hat\gamma_n - \gamma)\log\frac{k_n}{np_n}\Big)\Big) = O_P\Big(\Big(\frac{k_n}{np_n}\Big)^\gamma\Big) = o_P(k_n^\gamma),$$

because $np_n \to 0$ and $k_n^{-1/2}\log(np_n) \to 0$.

Since, in view of (5.7) and the definition of $b(n)$,

$$\frac{U(n) - b(n)}{a(n/k_n)} = \frac{a(n)}{a(n/k_n)} \cdot \frac{A(n)}{\gamma+\rho} 1\{\rho < 0\} = o(k_n^\gamma),$$

approximation (2.5) yields

$$T_5 = \frac{k_n^\gamma - 1}{\gamma} + o\Big(k_n^{\gamma+\rho+\epsilon} A\Big(\frac{n}{k_n}\Big)\Big) + o(k_n^\gamma) + \frac{1}{\gamma} = \frac{k_n^\gamma}{\gamma} + o(k_n^\gamma).$$

Finally, $k_n^{-\gamma} T_6$ converges to G_γ in distribution because of $F \in D(G_\gamma)$ and (5.7).

Summing up, one obtains

$$\frac{g^{\leftarrow}_{\hat a,\hat b,\hat\gamma_n}(np_n/k_n) - X_{n,n}}{k_n^\gamma a(n/k_n)} \xrightarrow{d} -\Big(M + \frac{1}{\gamma}\Big)$$

for a G_γ-distributed r.v. M. Now the assertion follows from the fact that $-(M + 1/\gamma) > 0$ a.s.

Case (ii): $\gamma = 0$.

By similar arguments as in the first case one obtains

$$\begin{aligned}
\frac{g^{\leftarrow}_{\hat a,\hat b,\hat\gamma_n}(np_n/k_n) - X_{n,n}}{a(n/k_n)} &= \Bigg(\frac{\big(\frac{k_n}{np_n}\big)^{\hat\gamma_n} - 1}{\hat\gamma_n} - \log\frac{k_n}{np_n}\Bigg)\frac{\hat a}{a} + \Big(\frac{\hat a}{a} - 1\Big)\log k_n + \frac{\hat a}{a}\log\frac{1}{np_n} \\
&\quad + \frac{\hat b - b}{a} - \Big(\frac{b(n) - b(n/k_n)}{a(n/k_n)} - \log k_n\Big) - \frac{X_{n,n} - b(n)}{a(n)} \cdot \frac{a(n)}{a(n/k_n)} \\
&= o_P(1) + o_P(1) + \log\frac{1}{np_n}(1 + o_P(1)) + O_P(k_n^{-1/2}) + o(1) + O_P(1) \\
&= \log\frac{1}{np_n}(1 + o_P(1)) \\
&\xrightarrow{P} \infty
\end{aligned}$$

from which the assertion is obvious. □

Proof of Proposition 2.1:

We must prove that the following expression tends to 0 uniformly for $x \in (0,1]$:

$$\begin{aligned}
I &:= x^{-1/2+\epsilon}\Big(\sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n\big(g^{\leftarrow}_{\hat{a},\hat{b},\hat{\gamma}_n}(x)\big)-x\Big]-W_n(x)-L_n^{(\gamma)}(x)-\sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}(1/x)\Big)\\
&= \frac{x^{-1/2+\epsilon}}{(y_n(x))^{-1/2+\epsilon/2}}(y_n(x))^{-1/2+\epsilon/2}\Big(\sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n\big(g^{\leftarrow}_{a,b,\gamma}(y_n(x))\big)-y_n(x)\Big]\\
&\qquad -W_n(y_n(x))-\sqrt{k_n}A\big(\frac{n}{k_n}\big)(y_n(x))^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{y_n(x)}\big)\Big)\\
&\quad +x^{-1/2+\epsilon}\Big(\sqrt{k_n}(y_n(x)-x)-L_n^{(\gamma)}(x)\Big)\\
&\quad +x^{-1/2+\epsilon}\Big(W_n(y_n(x))-W_n(x)\Big)\\
&\quad +x^{-1/2+\epsilon}\Big(\sqrt{k_n}A\big(\frac{n}{k_n}\big)(y_n(x))^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{y_n(x)}\big)-\sqrt{k_n}A\big(\frac{n}{k_n}\big)x^{\gamma+1}K_{\gamma,\rho}\big(\frac{1}{x}\big)\Big)\\
&:= I_1+I_2+I_3+I_4
\end{aligned}$$

Now we distinguish three cases.

Case (i): $\gamma > 0$.
By Lemma 5.1(i), $\sup_{x\in(0,1]} x^{-1/2+\epsilon}/(y_n(x))^{-1/2+\epsilon/2}$ is stochastically bounded. Combining this with Theorem 2.1, we obtain $\sup_{x\in(0,1]}|I_1| \to 0$ in probability as $n \to \infty$. An application of Lemma 5.1(i), Lemma 5.2(i), and Lemma 5.3(i) gives

$$\sup_{x\in(0,1]}|I_2| \xrightarrow{d} 0, \qquad \sup_{x\in(0,1]}|I_3| \xrightarrow{P} 0, \qquad \sup_{x\in(0,1]}|I_4| \xrightarrow{P} 0,$$

respectively. Hence $\sup_{x\in(0,1]}|I| \to 0$ in probability as $n \to \infty$.

Case (ii) $-1/2 < \gamma < 0$, or $\gamma = 0$ and $\rho < 0$.
Let $\lambda_n := 1/(k_n \log k_n)$. Obviously $\lambda_n \to 0$, $k_n^{-1/2}\lambda_n^{\gamma} \to 0$ and $k_n^{-1/2}\log^2\lambda_n \to 0$ as $n \to \infty$, and hence the Lemmas 5.1, 5.2 and 5.3 apply. Like in case (i), we obtain $\sup_{x\in(\lambda_n,1]}|I| \to 0$ in probability as $n \to \infty$.

It remains to prove that $\sup_{x\in(0,\lambda_n]}|I| \to 0$ in probability. To this end, let $p_n := 1/(n\log k_n)$, and so $np_n \to 0$ and $k_n^{-1/2}\log^2(np_n) \to 0$ as $n \to \infty$. Thus, for $x \in (0,\lambda_n]$,

$$g^{\leftarrow}_{\hat{a},\hat{b},\hat{\gamma}_n}(x) \geq g^{\leftarrow}_{\hat{a},\hat{b},\hat{\gamma}_n}(\lambda_n) = g^{\leftarrow}_{\hat{a},\hat{b},\hat{\gamma}_n}(np_n/k_n).$$

It follows from Lemma 5.4 that

$$P\Big\{\sup_{x\in(0,\lambda_n]} x^{-1/2+\epsilon}\frac{n}{\sqrt{k_n}}\bar{F}_n\big(g^{\leftarrow}_{\hat{a},\hat{b},\hat{\gamma}_n}(x)\big) \neq 0\Big\} = P\big\{g^{\leftarrow}_{\hat{a},\hat{b},\hat{\gamma}_n}(np_n/k_n) < X_{n,n}\big\} \to 0 \tag{5.8}$$

as $n \to \infty$.

Furthermore, it is easy to check that

$$
\begin{aligned}
x^{-1/2+\epsilon}\sqrt{k_n}x \to 0, &\qquad x^{-1/2+\epsilon}W_n(x) \xrightarrow{P} 0,\\
x^{-1/2+\epsilon}L_n^{(\gamma)}(x) \xrightarrow{P} 0, &\quad x^{-1/2+\epsilon}\sqrt{k_n}Ax^{\gamma+1}K_{\gamma,\rho}(\frac{1}{x}) \to 0
\end{aligned}
\tag{5.9}
$$

uniformly for $x \in (0, \lambda_n]$ as $n \to \infty$. For example, the second convergence is an immediate consequence of the law of the iterated logarithm, and in the case $-1/2 < \gamma < 0$

$$
\begin{aligned}
\sup_{x\in(0,\lambda_n]} x^{-1/2+\epsilon}|L_n^{(\gamma)}(x)| \quad &\leq \quad \sup_{x\in(0,\lambda_n]} \frac{1}{|\gamma|}x^{1/2+\epsilon}\Big|\frac{1}{\gamma}\Gamma(W_n) - \alpha(W_n)\Big| + \sup_{x\in(0,\lambda_n]} \frac{1}{|\gamma|}|\Gamma(W_n)|x^{1/2+\epsilon}\log x\\
&\quad + \sup_{x\in(0,\lambda_n]} \frac{1}{|\gamma|}x^{1/2+\gamma+\epsilon}\Big|\gamma\beta(W_n) + \frac{1}{\gamma}\Gamma(W_n) - \alpha(W_n)\Big|\\
&\xrightarrow{P} \quad 0.
\end{aligned}
$$

In view of (5.8) and (5.9), the assertion $\sup_{x\in(0,\lambda_n]}|I| \to 0$ in probability is immediate.

Case (iii): $\gamma = \rho = 0$.

According to Lemma 5.1, $y_n(x)/x \to 1$ in probability uniformly for $x \in [\lambda_n, 1]$ with $\lambda_n := 1/(k_n \log k_n)$, and hence

$$
\frac{(1+|\log x|)^\tau}{(1+|\log y_n(x)|)^\tau} = \Big(\frac{1+|\log x|}{1+|\log x| + o_P(1)}\Big)^\tau = O_P(1)
$$

uniformly for $x \in [\lambda_n, 1]$. Therefore, one can argue as in case (ii) (using Corollary 2.1 instead of Theorem 2.1) to establish the assertion. □

Proof of Theorem 2.2:

By Proposition 2.1 one has

$$
\begin{aligned}
&\left(\sqrt{k_n}\Big[\frac{n}{k_n}\bar{F}_n\Big(\hat{a}(\frac{n}{k_n})\frac{x^{-\hat{\gamma}_n}-1}{\hat{\gamma}_n} + \hat{b}(\frac{n}{k_n})\Big) - x\Big]\right)^2\\
&= \left(W_n(x) + L_n^{(\gamma)}(x) + \sqrt{k_n}A(\frac{n}{k_n})x^{\gamma+1}K_{\gamma,\rho}(1/x) + \frac{o_p(1)}{h(x)}\right)^2
\end{aligned}
\tag{5.10}
$$

Using the law of iterated logarithm, it is readily checked that

$$
\begin{aligned}
\int_0^1 \big(W_n(x) + L_n^{(\gamma)}\big)^2 x^{\eta-2}\,dx &= O_P(1)\\
\int_0^1 \Big(x^{\gamma+1}K_{\gamma,\rho}(1/x)\Big)^2 x^{\eta-2}\,dx &< \infty\\
\int_0^1 \frac{x^{\eta-2}}{h^2(x)}\,dx &< \infty
\end{aligned}
$$

for $\eta > 0$, and $\eta \geq 1$ if $\gamma = \rho = 0$, provided ϵ or τ are chosen appropriately. Hence the assertion is an immediate consequence of (5.10) and $\sqrt{k_n}A(n/k_n) \to 0$. □

$\gamma =$ p	2	1.5	1	0.5	0.25	0	-0.25	-0.375	-0.45	-0.49	-0.499
0.995	0.545	0.513	0.507	0.525	0.553	0.621	0.672	0.739	0,739	0,889	0,909
0.99	0.477	0.462	0.459	0.474	0.494	0.554	0.604	0.667	0.726	0.774	0.795
0.975	0.408	0.389	0.383	0.390	0.409	0.459	0.510	0.558	0.590	0.641	0.657
0.95	0.349	0.337	0.330	0.337	0.355	0.390	0.431	0.468	0.500	0.539	0.552
0.9	0.289	0.281	0.278	0.285	0.295	0.318	0.355	0.381	0.405	0.435	0.444
0.5	0.151	0.148	0.147	0.149	0.154	0.162	0.178	0.189	0.199	0.207	0.211
0.1	0.083	0.082	0.081	0.082	0.085	0.089	0.095	0.099	0.103	0.105	0.106
0.05	0.071	0.070	0.070	0.071	0.073	0.078	0.080	0.083	0.087	0.089	0.090
0.025	0.062	0.062	0.062	0.063	0.064	0.068	0.071	0.073	0.076	0.077	0.078
0.01	0.053	0.054	0.054	0.055	0.056	0.059	0.060	0.062	0.066	0.066	0.067
0.005	0.048	0.049	0.049	0.050	0.051	0.052	0.054	0.055	0.059	0.058	0.059

Table 1: Quantiles $Q_{p,\gamma}$ of the limit distribution of $k_n T_n$.

6 Simulations

First we want to calculate the limiting distribution of the test statistic $k_n T_n$ defined by (1.4), where we use the maximum likelihood estimator $\hat{\gamma}_n$, $\hat{a}(n/k_n)$ and $\hat{b}(n/k_n)$ described in Example 2.1. Here we have chosen $\eta = 1$, thus giving maximal weight to deviations in the extreme tail region that is possible in the framework of Theorem 2.2 for all values of $\gamma > -1/2$.

To simulate $\int_0^1 (W_n(x) + L_n^{(\gamma)}(x))^2 x^{-1}\, dx$, the Brownian motion W_n on the unit interval is simulated on a grid with 50 000 points. Then the integral is approximated by a Riemann sum for the extreme value indices $\gamma = 2, 1.5, 1, 0.5, 0.25, 0, -0.25, -0.375, -0.45, -0.49$ and -0.499. Note that for $\gamma \le -1/2$ the term $L_n^{(\gamma)}$ is not defined since the integral $S_n = \int_0^1 t^{\gamma-1} W_n(t)\, dt$ defined in Example 2.1 does not exist. The empirical quantiles of the integral statistic obtained in 20 000 runs are reported in Table 1. It is not surprising that the extreme upper quantiles increase rapidly as $\gamma < 0$ decreases, since $|S_n| \to \infty$ in probability as $\gamma \downarrow -1/2$, and thus the limit distribution of $k_n T_n$ converges weakly to ∞, too.

Next we investigate the finite sample behavior of the test described in Section 2, that rejects the hypothesis that $F \in D(G_\gamma)$ for some $\gamma > -1/2$ if $k_n T_n$ exceeds $\hat{Q}_{1-\bar{\alpha},\tilde{\gamma}_n}$. Here we use the maximum likelihood estimator for γ also as the pilot estimator, that is, $\tilde{\gamma}_n = \hat{\gamma}_n$. Since we have approximately determined the quantiles $Q_{p,\gamma}$ only for 11 different values of γ, we use linear interpolation to approximate the quantiles for intermediate values of γ, that is, for $\tilde{\gamma}_n \in [\gamma_1, \gamma_2]$ we define

$$\hat{Q}_{p,\tilde{\gamma}_n} = Q_{p,\gamma_1} + \frac{\tilde{\gamma}_n - \gamma_1}{\gamma_2 - \gamma_1}(Q_{p,\gamma_2} - Q_{p,\gamma_1})$$

where Q_{p,γ_i} denote the quantiles given in Table 1. Moreover, we define $\hat{Q}_{p,\tilde{\gamma}_n} := Q_{p,2}$ if $\tilde{\gamma}_n > 2$. (If one wants to perform the test for a single data set then it seems more natural to simulate the quantile $Q_{p,\tilde{\gamma}_n}$ directly, but for our simulation study this approach is much too computer intensive. However, a comparison of directly simulated quantiles with the approximations obtained by the above linear interpolation for $\gamma = 1.75, 1.25, 0.75, -0.1, -0.2, -0.3, -0.35$ and -0.45 showed a sufficient accuracy of the linear approximation.)

As usually in extreme value theory, the choice of the number k_n of order statistics used for the inference is a crucial point. Here we consider $k_n = 20, 40, \ldots, 100$ for sample size $n = 200$, and $k_n = 50, 100, \ldots, 400$ for sample size $n = 1000$.

We have drawn 10 000 samples from each of the following distribution functions belonging to the domain of attraction of G_γ for some $\gamma > -1/2$:

- Cauchy distribution $\quad (\gamma = 1,\ \rho = -2)$:

  $$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x, \quad x \in \mathbb{R}.$$

- Log-gamma distribution $LG(\gamma, m)$ $\quad (\rho = 0)$ $\qquad$ with density

  $$f(x) = \frac{1}{\gamma^m \Gamma(m)} \log^{m-1}(x) x^{-(1/\gamma+1)} 1_{[1,\infty)}(x),$$

  for $\gamma = 0.5$, and $m = 2$ and $m = 10$.

- Burr(β, τ, λ) distribution $\quad (\gamma = 1/(\tau\lambda),\ \rho = -1/\lambda)$:

  $$F(x) = 1 - \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda, \quad x > 0,$$

  with $(\beta, \tau, \lambda) = (1, 2, 2)$.

- Extreme Value distribution $EV(\gamma)$ $\quad (\gamma \in \mathbb{R}, \rho = -1)$:

  $$F(x) = \exp\big(-(1+\gamma x)^{-1/\gamma}\big), \quad 1 + \gamma x > 0,$$

  with $\gamma = 0.25$ and $\gamma = 0$.

- Weibull(λ, τ) distribution $\quad (\gamma = 0, \rho = 0)$:

$$F(x) = 1 - \exp(-\lambda x^\tau), \quad x > 0,$$

with $(\lambda, \tau) = (1, 0.5)$.

- Standard normal distribution $\quad (\gamma = 0, \rho = 0)$.

- Reversed Burr(β, τ, λ) distribution $\quad (\gamma = -1/(\tau\lambda),\ \rho = -1/\lambda)$:

$$F(x) = 1 - \left(\frac{\beta}{\beta + (x_+ - x)^{-\tau}}\right)^\lambda, \quad x < x_+,$$

with $(\beta, \tau, \lambda) = (1, 4, 1)$ and $x_+ = 1$.

We used the software package XTREMES with its implemented routine for calculating the maximum likelihood estimates. In some simulations either the algorithm could not find any solution to the likelihood equations, or the maximum likelihood estimate of γ is less than -0.499, so that the test cannot be applied. The relative frequency of simulations in which this happened are given in the Tables 2–5; for all other values of k_n not mentioned in these tables, the test could be performed in all simulations.

For the reversed Burr distribution, one gets estimates of γ less than -0.499 in at least 1% of the simulations for all values of k_n and in about one third of all simulations if $n = 200$, and for the normal distribution in more than 7% of the simulations for $n = 200$, while for all other distributions this happened only if a small proportion of the data is used for the inference. It is obvious that the problem of pilot estimates of γ being smaller than $-1/2$ becomes more and more acute as the true extreme value index approaches $-1/2$; this is particularly true for small sample sizes.

In the Tables 6 and 7 the empirical size of the test with nominal size $\bar{\alpha} = 0.05$ is reported, that is, the relative frequency of simulations in which the hypothesis is rejected. These frequencies are based only on those simulations in which the test could actually be applied. The numbers are given in **bold face** for those k_n for which the empirical mean squared error of $\hat{\gamma}_n$ is minimal. Note that for these 'optimal' sample fractions the bias and the variance of $\hat{\gamma}_n$ are balanced. For smaller k_n, (usually) the bias is dominated by the variance, which indicates that the deviation of the true distribution of the excesses from the ideal generalized Pareto distribution is small. As we have expected (cf. the discussion after Remark 2.4), the empirical size of the test is close to or smaller than its nominal size for this range of k, while for most distributions the actual size increases

k_n	Cauchy $\gamma = 1$, $\rho = -2$	LG(0.5,2) $\gamma = 0.5$, $\rho = 0$	LG(0.5,10) $\gamma = 0.5$, $\rho = 0$	Burr(1,2,2) $\gamma = 0.25$, $\rho = -0.5$	EV(0.25) $\gamma = 0.25$, $\rho = -1$	EV(0) $\gamma = 0$, $\rho = -1$	Weib.(1,0.5) $\gamma = 0$, $\rho = 0$	normal $\gamma = 0$, $\rho = 0$	RBurr(1,4,1) $\gamma = -0.25$ $\rho = -1$
20	0.2	0.7	0.2	2.6	2.1	5.8	1.7	10.8	17.3
40						0.1		0.8	2.0
60						0.1		0.2	0.7
80								0.1	0.7
100								0.2	1.4

Table 2: Percentage of simulations in which no maximum likelihood estimate was found for sample size $n = 200$; empty entries correspond to 0

k_n	Cauchy	LG(0.5,2)	LG(0.5,10)	Burr(1,2,2)	EV(0.25)	EV(0)	Weib.(1,0.5)	normal	RBurr(1,4,1)
20	0.4	1.4	0.6	5.8	4.9	12.6	4.2	21.4	30.7
40				0.6	0.5	3.6	0.1	14.0	30.5
60				0.1	0.1	0.9		8.8	28.6
80						0.3		7.1	33.7
100						0.1		7.6	46.0

Table 3: Percentage of simulations in which $\hat{\gamma}_n < -0.499$ for sample size $n = 200$; empty entries correspond to 0

k_n	Cauchy	LG(0.5,2)	LG(0.5,10)	Burr(1,2,2)	EV(0.25)	EV(0)	Weib.(1,0.5)	normal	RBurr(1,4,1)
50								0.2	0.6
100									0.1

Table 4: Percentage of simulations in which no maximum likelihood estimate was found for sample size $n = 1000$; empty entries correspond to 0

k_n	Cauchy	LG(0.5,2)	LG(0.5,10)	Burr(1,2,2)	EV(0.25)	EV(0)	Weib.(1,0.5)	normal
50				0.2	0.1	1.2	0.1	4.9
100								0.5

k_n	50	100	150	200	250	300	350	400
RBurr(1,4,1)	14.5	4.3	2.0	1.1	0.9	0.9	1.1	2.0

Table 5: Percentage of simulations in which $\hat{\gamma}_n < -0.499$ for sample size $n = 1000$; empty entries correspond to 0

k_n	Cauchy	LG(0.5,2)	LG(0.5,10)	Burr(1,2,2)	EV(0.25)	EV(0)	Weib.(1,0.5)	normal	RBurr(1,4,1)
20	4.7	4.1	4.1	3.7	3.5	3.0	3.5	2.5	1.9
40	5.3	4.8	**4.9**	4.2	4.0	3.6	**5.3**	3.0	2.7
60	**6.4**	5.1	5.2	4.5	4.5	3.7	6.7	**3.8**	**3.1**
80	10.9	5.4	5.2	**5.3**	4.8	4.1	9.4	5.2	5.6
100	24.7	5.3	5.5	7.4	**5.5**	**5.4**	14.4	87	12.9

Table 6: Empirical size in % of the test with nominal size $\bar{\alpha} = 5\%$ for sample size $n = 200$.

k_n	Cauchy	LG(0.5,2)	LG(0.5,10)	Burr(1,2,2)	EV(0.25)	EV(0)	Weib.(1,0.5)	normal	RBurr(1,4,1)
50	4.9	4.3	**4.7**	4.4	4.3	3.7	**4.6**	3.5	3.0
100	5.3	4.8	5.1	4.8	4.4	4.3	5.3	4.4	3.5
150	5.5	4.6	5.2	5.0	4.7	4.5	6.8	**4.9**	**3.9**
200	5.9	4.6	5.5	5.6	4.4	4.5	9.1	6.7	5.3
250	**7.6**	4.8	6.2	**6.5**	5.0	5.2	12.4	9.8	8.5
300	11.5	4.9	6.7	7.5	5.3	**5.9**	17.4	14.0	14.7
350	19.4	4.7	7.5	9.7	**6.2**	7.5	24.3	21.3	24.9
400	34.4	5.1	8.6	12.7	7.1	10.0	34.4	30.7	40.6

Table 7: Empirical size in % of the test with nominal size $\bar{\alpha} = 5\%$ for sample size $n = 1000$.

rapidly if a larger number of order statistics is used, thus indicating the growing deviation from the generalized Pareto model.

Note that the test behaves differently for the log-gamma distributions. The $LG(0.5, 2)$ is somewhat special: here first the bias of $\hat{\gamma}_n$ increases as k_n increases but for $k_n \geq 350$ (and $n = 1000$) it decreases again until it almost vanishes for $k_n = 875$, where the mean squared error is minimized. Of course, such an 'irregular' behavior cannot be taken into account by the asymptotic extreme value analysis which always assumes that k_n/n tends to 0 and thus cannot deal with k-values close to n.

For $LG(0.5, 10)$, the empirical size of the test exceeds the nominal size if k is larger than the 'optimal' value, but it grows very slowly. Such a behavior can be observed for some d.f.'s satisfying the second order condition (2.4) with $\rho = 0$ (or ρ close to 0). Note that in this case the function $A(t)$, that (for $t = n/k_n$) describes the rate of convergence of the first order term of the deviation from the generalized Pareto model, is slowly varying and hence it decreases very slowly as t increases. Therefore, an increase of k_n leads to just a small increase of the model deviation

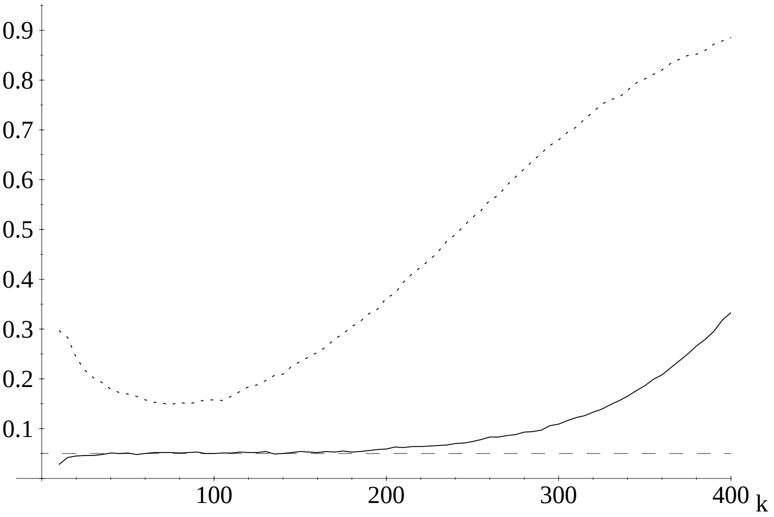

Figure 1: Empirical size of the test with nominal size $\bar{\alpha} = 0.05$ as a function of k_n for Cauchy samples of size $n = 1000$.

which is difficult to detect by the test. Such d.f.'s are infamous for causing problems to data-driven choices of k (see e.g. Drees and Kaufmann (1998)).

In view of these results, we may conclude that, for most d.f.'s, the test indeed indicates the range of k-values for which extreme value estimators are only moderately biased. Unfortunately, for very small k the test seems too conservative, in particular for the normal and the reversed Burr distribution and, to a lesser extent, also for the Gumbel distribution EV(0).

This conclusion is also supported by Figure 1 that displays the empirical size of the test versus k for the Cauchy distribution and $k = 10, 15, 20, \ldots, 400$. In addition, the ratio of the absolute bias to the root mean squared error of $\hat{\gamma}_n$ is plotted by the dotted line. When the bias contributes less than (about) 30% to the root mean squared error, then the test rejects the hypothesis with probability 0.05 or less. As the influence of the bias on the total error increases, the probability of a rejection rises above the 5%-line and increases more and more rapidly as k increases.

At first glance, it might be surprising that, unlike estimators of γ, the test behaves almost equally well for small and large values of $|\rho|$. However, recall that for the actual size to be close to the nominal value it is not important how accurate the estimators are but only how precise the Gaussian approximation for the tail empirical distribution function with estimated parameters is. While the rate of convergence of estimators of the extreme value index deteriorates as ρ tends to 0, this is not necessarily true for the accuracy of the normal approximation.

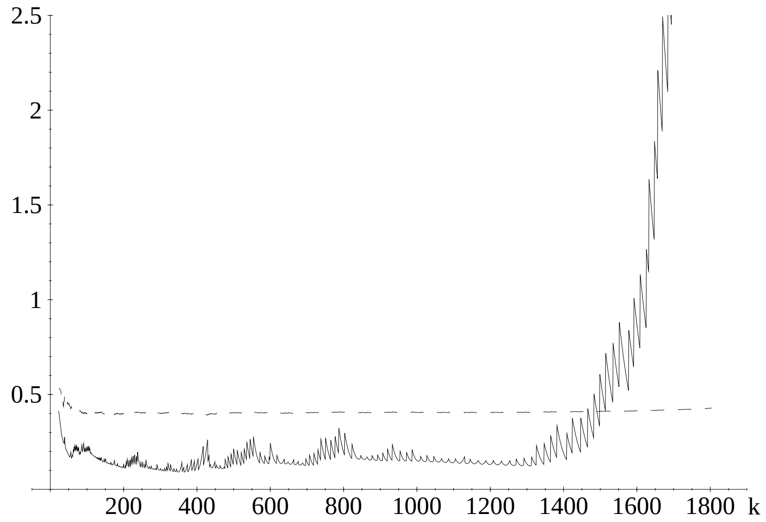

Figure 2: Test statistic kT_n (solid line) and estimated critical values of the test with nominal size 0.05 (dashed line) versus k.

7 Application: Sea Level Data

De Haan (1990) analyzed high tide water levels recorded at five different stations along the Dutch coastline. Here we examine the levels observed at Delfzijl between 1881 and 1988. After some preprocessing (cf. de Haan (1990)), one arrives at 1873 observations that may be regarded as approximately independent. In Figure 2, the test statistic kT_n is plotted versus k together with the (estimated) critical value of the test with nominal size 0.05. It turns out that the null hypothesis is clearly accepted for $29 \leq k \leq 1465$. (For most $k \leq 28$ the maximum likelihood estimator does not exist or $\hat{\gamma}_n < -0.499$; the sawtooth shape of the curve is due to a rounding of the water level given in centimeter to the next integer.)

A comparison with Figure 3 that displays a plot of $\hat{\gamma}_n$ versus k reveals that the test statistic starts to increase at the same point where the curve of $\hat{\gamma}_n$ shows a clear downward trend, indicating a growing negative bias of the estimator. This effect is in line with the findings of our simulation study and the discussion after Remark 2.4.

Indeed, a qq-plot of the k upper order statistics $X_{n-i+1,n}$, $1 \leq i \leq k$, versus the quantiles $X_{n-k+1:n} + \hat{a}(n/k)((i/k)^{-\hat{\gamma}_n} - 1)/\hat{\gamma}_n$ of the fitted generalized Pareto distribution shows a very good fit for $k = 1200$ (Figure 4, left graph), while for $k = 1600$ the plot clearly deviates from the diagonal (and, in fact, from any straight line), i.e. the largest 1600 observations cannot be fitted well by a generalized Pareto distribution (Figure 4, right graph).

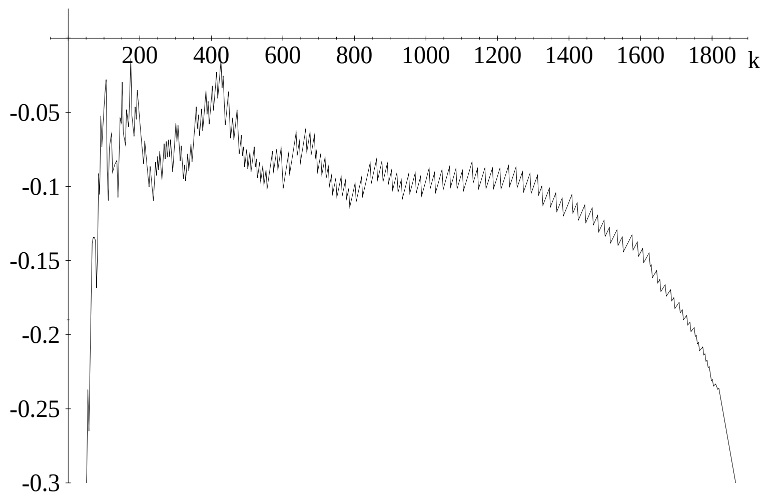

Figure 3: Maximum likelihood estimator $\hat{\gamma}_n$ versus k.

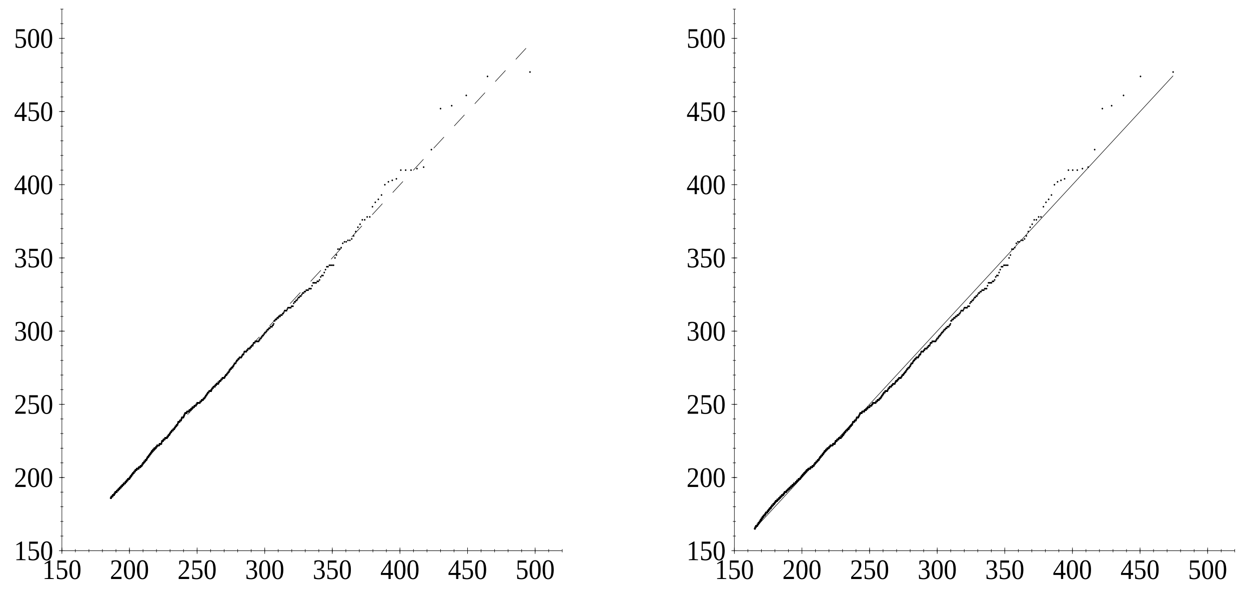

Figure 4: QQ-plot of the k upper order statistics versus the fitted generalized Pareto quantiles for $k = 1200$ (left) and $k = 1600$ (right).

Acknowledgment: Part of the work of the first two authors was done while visiting the Stochastics Center at Chalmers University Gothenburg. Grateful acknowledgement is made for hospitality particularly to Holger Rootzén. During that time Holger Drees was supported by the European Union TMR grant ERB-FMRX-CT960095. His work was also partly supported by the Netherlands Organization for Scientific Research through the Netherlands Mathematical Research Foundation and by the Heisenberg program of the DFG. Detailed comments by an anonymous referee led to a significant improvement of the presentation; in particular, the proofs of the Lemmas 4.1 and 5.2 became much simpler.

References

[1] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[2] Choulakian, V. and Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* **43**, 478–484.

[3] Cheng, S. and Jiang, C. (2001). The Edgeworth expansion for distributions of extreme values. *Sci. China* **44**, 427–437.

[4] Csörgő, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics.* Wiley, Chichester.

[5] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc. B* **52**, 393–442.

[6] Dietrich, D., de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions. *Extremes* **5**, 71–85.

[7] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics* **23**, 277–281.

[8] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* **20**, 393–403.

[9] Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.* **25**, 187–210.

[10] Drees, H., Ferreira, A. and de Haan, L. (2003). On maximum likelihood estimation of the extreme value index. To appear in *Ann. Appl. Probab.*

[11] Drees, H., de Haan, L. and Li, D. (2003). On large deviations for extremes. *Statist. Probab. Letters*, **64**, 51–62.

[12] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Proc. Appl.* **75**, 149–172.

[13] Einmahl, J.H.J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stoch. Proc. Appl.* **70**, 31–58.

[14] de Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statist. Neerlandica* **44**, 45–68.

[15] de Haan, L. and Resnick, S.I. (1998). On the asymptotic normality of the Hill estimator. *Stochastic Models* **14**, 849–867.

[16] de Haan, L. and Rootzén, H. (1993). On the estimation of high quantiles. *J. Statist. Planning Inference* **35**, 1–13.

[17] de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A* **61**, 381–395.

[18] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[19] Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–1207.

[20] Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **23**, 245–253.