Estimating the Extreme Value Index

Habilitationsschrift

Bei der Mathematisch-Naturwissenschaftlichen Fakultät der
Universität zu Köln
zur Erlangung der Lehrbefähigung für das Fach
Mathematik

eingereicht von

Holger Drees

Köln, Februar 1998

Contents

List of Symbols

a	Condition 2.1
a^*	(4.10)
$b_{n,k}$	(3.8), (3.31)
β	(1.2)
$\hat{\beta}_n^H$	Example 2.1(ii), (3.1)
$\hat{\beta}_n^M$	Example 2.1(iii)
$\hat{\beta}_n^P$	Example 2.1(i)
$C_{\beta,h}$	(2.20)
$c_{T,\beta}$	Condition 2.8
$D_{\beta,h}$	(2.13)
D_n	(2.17)
δ	(2.3)
E	(3.9)
ε	(3.6), (4.5)
F_β^{-1}	(2.2)
$\bar{F}_\beta^{-1}$	(2.14)
$\mathcal{F}_1$	(4.7)
$\mathcal{F}_2$	(4.21)
G_β	(1.2)
h, $\mathcal{H}$	(2.9)
$\tilde{h}$	(2.8)
$k_n^{(1)}$	(2.30), (3.12)
$\bar{k}_n$	(3.5), (3.28)
k_n^{opt}	(3.13), (3.25)
$\hat{k}_n^{opt}$	Theorems 3.1 and 3.4
$\mathcal{M}(\beta)$	p. 15
$\mu_{T,\beta,\Psi}$	Theorem 2.3
ν_β^*	Theorem 2.4
$\nu_{T,\beta}$	(2.22)

Ξ	(2.36), Lemma 2.5
Φ	Condition 2.1
Ψ	Condition 2.1, (2.4)
$\bar{\Psi}$	(2.15)
$Q_n = Q_{n,k_n}$	(1.4)
$\tilde{Q}_n = \tilde{Q}_{n,k_n}$	(2.7)
q_n	(3.37)
R	Condition 2.1
S_i	(2.5)
$\sigma^2_{\beta,\nu}$	Theorem 2.4
$\sigma^2_{T,\beta}$	(2.23)
T_H	Example 2.1(ii)
T_M	Example 2.1(iii)
T_{ML}	Example 2.3
T_P	Example 2.1(i)
T_{PWM,ν_1,ν_2}	Example 2.1(iv)
T'_β	Conditions 2.4 and 2.8
t_n	(4.8)
t_n^*	(4.12)
V_β	(2.16)
W	Brownian motion
$\omega = F^{-1}(1)$	Lemma 2.5
$\Vert \cdot \Vert_{\beta,h}$	(2.10)
$\sim$	p. 8, Remark 2.5 (i)

1 Introduction

In many statistical applications it is necessary to make inferences about the tail of a distribution, where little or no data is available. For example, if a dike is projected in order to protect a coastal line, then usually it will be higher than any flood recorded yet. Hence the estimation of the minimal height that ensures that the probability of being flooded in a particular year is less than a given small value requires an extrapolation of the underlying distribution beyond the observed data.

As a second example, consider an excess of loss reinsurance treaty, according to which the reinsurance company has to pay the part of each loss exceeding a given deductible (retention), say t_0, to the ceding insurance company. In order to calculate a fair premium for this treaty, one must estimate certain functionals of the tail of the loss distribution beyond the deductible, which has usually been exceeded only by a few losses if the ceding company merely wants to safeguard itself against catastrophic risks, i.e., if t_0 is very high.

Further examples, where mainly the upper (or lower) tail of a distribution is of interest, are given in the monographs by Galambos (1978), Castillo (1988), Embrechts et al. (1997) and Reiss and Thomas (1997). Apparently, in such situations one need assumptions on the distribution of the observed random variables (r.v.'s) X_i, $1 \le i \le n$, in order to make inference about the tail beyond the data. In this treatise we restrict ourselves to the classical model with independent and identically distributed r.v.'s. Moreover, in view of the dike example, it seems natural to assume that the sequence of distributions of the maximum of the observed r.v.'s becomes stable in some sense as the sample size increases. Since the maximum converges to the right endpoint of the parent distribution, one has to standardize it to obtain a non-degenerate limit; due to their simplicity, usually affine transformations are used for this purpose. Hence we arrive at the following condition, which is basic for the classical univariate extreme value theory: The parent distribution function (d.f.) F belongs to the (weak) max-domain of attraction of some (unknown) non-degenerate d.f. G (in short: $F \in D(G)$), that means,

$$\mathcal{L}\Big(a_n^{-1}\big(\max_{1\le i\le n} X_i - b_n\big)\Big) \longrightarrow G \qquad \text{weakly} \tag{1.1}$$

for some normalizing constants $a_n > 0$ and $b_n \in I\!R$, or, equivalently,

$$F^n(a_n x + b_n) \longrightarrow G(x)$$

for all points x of continuity of G. Since the pioneering papers by Fisher and Tippett (1928) and Gnedenko (1943), it is well known that then, up to a scale and location parameter, G equals one of the following extreme value d.f.'s G_β:

$$G_\beta(x) := \begin{cases} \exp(-(1+\beta x)^{-1/\beta}) & \text{if } \quad 1+\beta x > 0,\ \beta \ne 0, \\ \exp(-e^{-x}) & \qquad\quad \beta = 0. \end{cases} \tag{1.2}$$

In case of $F \in D(G_\beta)$, the real parameter β is called *extreme value index* of F.

Pickands (1975) proved that the tail of a d.f. belonging to the domain of attraction of an extreme value d.f. G_β can be approximated in a certain sense by the pertaining generalized Pareto d.f.

$$W_\beta(x) := 1 + \log G_\beta(x) = \begin{cases} 1-(1+\beta x)^{-1/\beta} & \beta \neq 0, \\ 1-\mathrm{e}^{-x} & \beta = 0, \end{cases} \text{if} \quad -1 \leq \log G_\beta(x) \leq 0$$

with suitable scale and location parameter (cf. also Balkema and de Haan (1974)). Analytically, this approximation is reflected by the following necessary and sufficient condition for $F \in D(G_\beta)$, which was established by de Haan (1984):

$$\lim_{t\downarrow 0} \frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} = \frac{x^{-\beta}-1}{\beta} \tag{1.3}$$

for $x > 0$ and some positive, measurable function a, with F^{-1} denoting the quantile function (q.f.) pertaining to F. Note that the right-hand side, which is interpreted as $-\log x$ for $\beta = 0$, equals $W_\beta^{-1}(1-x)$ if $x \in (0,1)$. Thus in the dike example, where one is interested in an extreme quantile $F^{-1}(1-p)$ for small $p \ll 1/n$, loosely speaking, one may approximate this unknown quantity by $F^{-1}(1-t) + a(t)((p/t)^{-\beta} - 1)/\beta$ for some $t > p$. Specifically, choosing $t = k/n$, $1 \leq k \leq n$, one may estimate $F^{-1}(1-k/n)$ by the kth largest order statistic $X_{n-k+1:n}$, and hence one obtains an estimator for $F^{-1}(1-p)$ by replacing the unknown scale factor $a(k/n)$ and the extreme value index by suitable estimators (see Dekkers et al. (1989), de Haan and Rootzén (1993) and Beirlant et al. (1996a), p. 97). Observe that, with np/k being small, the resulting estimator for the extreme quantile is particularly sensitive to the estimation error for β, which enters the formula in the exponent. Therefore, a high accuracy of the estimate of the extreme value index is of particular importance.

There is an extensive literature about the estimation of β (see, e.g., Pickands (1975), Hill (1975), Hall and Welsh (1984, 1985), Smith (1987), Dekkers et al. (1989), Dekkers and de Haan (1989, 1993) and the monographs by Reiss (1989), Beirlant et al. (1996a) and Reiss and Thomas (1997)), and a few papers study more or less restricted families of estimators (Csörgő et al. (1985) and Drees (1995, 1996)). In a series of papers (see, e.g., Falk (1995) and Marohn (1995, 1997)), Falk and Marohn used LAN theory to determine asymptotically efficient estimators for β, but they restricted the sample fraction that the estimator may use (i.e., largest order statistics or exceedances over thresholds) in advance in such a way that the distribution of these r.v.'s under the parent d.f. can be approximated in the variational distance by their distribution under a suitable generalized Pareto law (Reiss (1989), Corollary 5.5.5, and Reiss (1993), Theorem 6.1.2). Since this holds true under the center of localization and the local alternatives as well, actually models were examined that are asymptotically equivalent to parametric models with at most three dimensions.

In fact, a general approach to the construction of estimators of the extreme value index in genuine semiparametric models, where the choice of the sample fraction is part of the estimation problem, is not yet available. It is the main goal of the present work to contribute to a solution of this semiparametric estimation problem.

Because the extreme value index characterizes the behavior of the d.f. F locally at its right endpoint, evidently an estimator $\hat\beta_n$ should utilize only those observations that are large in some sense. Since the seminal work by Pickands (1975), most commonly a certain number, say k_n+1, of largest order statistics $X_{n-k_n:n} \le \ldots \le X_{n-1:n} \le X_{n:n}$ is used. Asymptotically, the number of order statistics shall increase as the sample size increases, while relatively to the size of the whole sample the number shall vanish to ensure that almost surely all order statistics under consideration converge to the right endpoint of the d.f. Hence $(k_n)_{n\in\mathbb{N}}$ shall be an *intermediate sequence*, that is,

$$k_n \longrightarrow \infty \qquad \text{and} \qquad k_n/n \longrightarrow 0.$$

Closely related to this approach is the so-called peak over threshold (POT) method, where all exceedances over a given threshold, i.e., a random number of largest order statistics, are used for estimation. While many of the results derived in Section 2 have counterparts for the POT-method if the average number of exceedances is asymptotic to k_n, the choice of an appropriate threshold seems a bit more difficult than the selection of an "optimal" k_n, since the latter depends only on the so-called second-order behavior of F, whereas the former also depends on the parameter β which is to be estimated (cf., e.g., Reiss (1993), Theorem 6.1.2, and Marohn (1995), (3.2.4) and Theorem 3.1.7).

Finally, one may split the sample in several subsamples and use the maxima of these subsamples for estimation, which, loosely speaking, constitute an i.i.d. sample with d.f. close to G_β if $F \in D(G_\beta)$. While in some situations, especially for time series, such a division of the sample seems quite natural, often this so-called annual maxima method appears artificial (cf. Castillo (1994), Section 5.2). Therefore, we will concentrate on estimators based on a (at least asymptotically) deterministic number of order statistics, though in Chapter 4 arbitrary estimators of β are also considered.

If the estimator $\hat\beta_n$ is based on the k_n+1 largest order statistics, then it can be written as $\hat\beta_n = T_n(Q_n)$ where

$$Q_{n,k_n}(t) := Q_n(t) := F_n^{-1}\Big(1-\frac{k_n}{n}t\Big) = X_{n-[k_nt]:n}, \qquad t\in[0,1], \tag{1.4}$$

is the pertaining empirical tail quantile function and T_n denotes a functional defined on a suitable function space. Here F_n^{-1} denotes the empirical quantile function and $[x]$ is the largest integer less than or equal to x. For notational convenience, usually we will omit the index k_n unless we discuss the choice of the sample fraction used for estimation.

If one wants to determine the limit distribution of $T_n(Q_n)$, then one has to assume that the sequence of functionals T_n converges in some sense. We restrict ourselves to the case where the functional is fixed, i.e.,

$$\hat\beta_n = T(Q_n) \tag{1.5}$$

where T is sufficiently regular. Notice that all estimators investigated in the above mentioned papers are (at least approximately) of this type. Obviously, these estimators can be regarded as extreme value counterparts of the classical statistical functionals, which are defined as functionals on the Skorohod space applied to the empirical d.f. (see Fernholz (1983)). Therefore, estimators of type (1.5) will be addressed as *statistical tail functionals*.

In Chapter 2 we study the asymptotic behavior of statistical tail functionals based on given numbers of order statistics that constitute an intermediate sequence. First, a limit theorem for the empirical tail q.f. is established, from which consistency and asymptotic normality of statistical tail functionals $T(Q_n)$ can be concluded if the pertaining functional T is sufficiently smooth and regular. For this, one has to distinguish whether T is location invariant or not, and, in the latter case, whether β is positive or not. In all cases, we characterize the family of possible limiting distributions and discuss the choice of the functional T. It turns out that for negative extreme value indices location-invariant statistical tail functionals are clearly preferable. Moreover, we introduce a method to construct estimators for the extreme value index with preassigned limit distribution.

In Chapter 3 we discuss the data driven choice of the number of order statistics a fixed statistical tail functional is based on. Inspired by a procedure for the bandwidth selection in nonparametric curve estimation introduced by Lepskii (1991), we propose an adaptive choice of the sample fraction used for estimation, first for the Hill estimator (Section 3.1), and then for location-invariant statistical tail functionals (Section 3.2). In certain semiparametric submodels this choice leads to a minimal asymptotic mean squared error of the pertaining adaptive estimators for β.

While Chapter 2 and 3 mainly deal with the *construction* of estimators of the extreme value index, in the last part we investigate how accurate an *arbitrary sequence* of estimators can estimate β at best. More precisely, we determine optimal rates of convergence uniformly over certain sets of d.f.'s belonging to the domain of attraction of extreme value d.f.'s, thus generalizing a well-known result by Hall and Welsh (1984), who, roughly speaking, considered the case where β is positive and the rate of convergence in (1.1) is algebraic. In contrast, one gets surprisingly sharp bounds if the rate of convergence in (1.1) is poor, that is, slower than any power of n. Moreover, it is proven that suitable statistical tail functionals attain the optimal rates. In particular, again it turns out that location-invariant estimators are preferable if the extreme value index may be negative.

All essential results presented here are contained in the papers Drees (1998a,b) (Chapter 2), Drees and Kaufmann (1998) (Chapter 3) and Drees (1998c) (Chapter 4). The respective references are given in brackets following the number of the result.

Acknowledgements

I am grateful to Prof. H. Milbrodt and Prof. R.-D. Reiss for their encouragement, and to Prof. L. de Haan for his kind hospitality during my visits at the Erasmus University Rotterdam.

2 Statistical tail functionals

In this chapter we examine the asymptotic distributional behavior of statistical tail functionals $T(Q_n)$ defined in the introduction. Recall that, under suitable smoothness conditions, consistency and asymptotic normality of classical statistical functionals follow from the convergence of the empirical d.f. to the underlying d.f. and the convergence of the empirical process to a Brownian bridge, respectively. Following this approach, first we study the asymptotic behavior of the empirical tail q.f. Q_n defined by (1.4). However, in contrast to the empirical d.f., Q_n converges to a non-trivial limit only after a suitable normalization. To see this, recall that $F^{-1}(1-tk_n/n)/F^{-1}(1-k_n/n) \to t^{-\beta}$ if $\beta > 0$, whereas in the general case a shift is needed, too, i.e., $a_n^{-1}(F^{-1}(1-tk_n/n) - F^{-1}(1-k_n/n)) \to (t^{-\beta}-1)/\beta$. Therefore, in order to obtain asymptotic results for $T(Q_n)$ if $F \in D(G_\beta)$ for an arbitrary $\beta \in I\!R$, in Section 2.2 we will assume that the functional T is location and scale invariant:

$$T(az+b) = T(z)$$

for all $a > 0$, $b \in I\!R$ and all z belonging to an appropriate function space. Observe that neither the limit in (1.1), which defines the parameter β of interest, nor the speed of this convergence is changed by a scale transformation or shift, so that certainly the invariance of an estimator of β under such transformations is a natural condition.

From a more practical point of view, it is desirable that an estimator of the dimensionless extreme value index is not influenced by a change of the unit of measure, i.e., a scale transformation of the data. Moreover, in some applications the observations (e.g., sea levels in the diking example, or temperatures) depend on an arbitrarily chosen zero-point, which equally should not affect the estimator.

On the other hand, while all estimators that are considered in the papers mentioned in Chapter 1 are scale invariant, some of them are not invariant under shifts. Observe that the effect of a shift of the q.f. by $\mu \in I\!R$ in case of $\beta > 0$, where one has $F^{-1}(1-t)+\mu = F^{-1}(1-t)(1+o(1))$ as $t \downarrow 0$, is completely different from the effect in the case $\beta < 0$, where the right endpoint $F^{-1}(1)$ is finite. For that reason, one has to investigate the asymptotic behavior of non-location-invariant statistical tail functionals seperately for $\beta > 0$ (Section 2.3) and for $\beta \leq 0$ (Section 2.4). The first three of the following examples are well-known representatives of estimators considered in Sections 2.2–2.4, respectively, while an adaptive version of the last one will enable us to construct estimators for the extreme value index with made-to-order asymptotic behavior (see Theorem 2.6).

Examples 2.1 Let z denote a real function on $[0,1]$.

(i) The scale- and location-invariant functional

$$T_P(z) := \log \frac{z(1/4)-z(1/2)}{z(1/2)-z(1)} \Big/ \log 2 \cdot 1_{\{(z(1/4)-z(1/2))\cdot(z(1/2)-z(1))>0\}}$$

leads to Pickands' (1975) estimator

$$\hat\beta_n^P := T_P(Q_n) = \log \frac{X_{n-[k_n/4]:n} - X_{n-[k_n/2]:n}}{X_{n-[k_n/2]:n} - X_{n-k_n:n}} \Big/ \log 2 \cdot 1_{\{X_{n-[k_n/4]:n} \neq X_{n-[k_n/2]:n} \neq X_{n-k_n:n}\}}.$$

(ii) The non-location-invariant functional

$$T_H(z) := \int_0^1 \log(z(t)/z(1))\, dt$$

if the right-hand side is well defined, and $T_H(z) := 0$ else, yields the Hill (1975) estimator, which in general is consistent only for $\beta > 0$:

$$\hat\beta_n^H := T_H(Q_n) = \frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{X_{n-i+1:n}}{X_{n-k_n:n}} \cdot 1_{\{X_{n-k_n:n} > 0 \text{ or } X_{n:n} < 0\}}.$$

(iii) The moment estimator $\hat\beta_n^M := T_M(Q_n)$, proposed by Dekkers et al. (1989) as a modification of the Hill estimator that is consistent for all $\beta \in I\!R$, is the non-location-invariant statistical tail functional pertaining to

$$T_M(z) = T_H(z) + 1 - \frac{1}{2(1 - T_H^2(z)/T_2(z))}$$

with

$$T_2(z) := \int_0^1 \log^2(z(t)/z(1))\, dt$$

if all terms are defined, and $T_M(z) := 0$ otherwise. Note that $T_2(Q_n) = k_n^{-1}\sum_{i=1}^{k_n} \log^2(X_{n-i+1:n}/X_{n-k_n:n})$ if the right-hand side is well defined.

(iv) Estimators of the type $T_{PWM,\nu_1,\nu_2}(Q_n)$ where

$$T_{PWM,\nu_1,\nu_2}(z) := \frac{\int z\, d\nu_1}{\int z\, d\nu_2} \cdot 1_{\{\int z\, d\nu_2 \neq 0\}}$$

for suitable finite signed Borel measures ν_i on $[0,1]$ can be regarded as generalized probability weighted moment estimators (cf. Hosking and Wallis (1987), p. 341). □

The basis of our investigations is a new approximation of the empirical tail q.f. Q_n that is established in Section 2.1. By the well-known δ–method we conclude the asymptotic normality of a statistical tail functional $T(Q_n)$ under suitable smoothness conditions on T. While, like in the theory of classical statistical functionals, Hadamard differentiability is the key property in the Sections 2.2 and 2.3, in the case $\beta \leq 0$ a more complicated smoothness condition is needed to obtain the asymptotic distribution of a non-location-invariant statistical tail functional. It turns out that in this case to each such functional there exists a location-invariant statistical tail functional with superior asymptotic performance. Therefore, it is clearly advisable to use location-invariant functionals if the sign of β is not known in advance.

All proofs are postponed until Section 2.5

2.1 An approximation of the empirical tail quantile function

As already mentioned in Chapter 1, $(k_n)_{n\in\mathbb{N}}$ has to be an intermediate sequence to ensure consistency of $T(Q_n)$, i.e., $k_n \to \infty$ and $k_n/n \to 0$. To obtain a non-degenerate limiting distribution, however, one has to impose stronger restrictions for k_n depending on the so-called second-order behavior of the underlying d.f. F. Several second-order conditions can be found in literature (see, e.g., Hall (1982), Csörgő et al. (1985), Smith (1987), and Dekkers and de Haan (1989, 1993)), yet the following expansion of the underlying q.f., which was extensively studied by de Haan and Stadtmüller (1996), is the most general one that covers all three cases $\beta >, =, < 0$.

Condition 2.1 *There exist measurable, locally bounded functions a, $\Phi : (0,1) \to (0,\infty)$ and $\Psi : (0,\infty) \to \mathbb{R}$ such that*

$$\frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} = \frac{x^{-\beta}-1}{\beta} + \Phi(t)\Psi(x) + R(t,x) \tag{2.1}$$

for all $t \in (0,1)$ and $x > 0$. (By convention, $(x^{-\beta}-1)/\beta := -\log x$ if $\beta = 0$.) Here

(i) $\Psi \equiv 0$ and $R(t,x) = o(1)$ as $t \downarrow 0$ for all $x > 0$, or

(ii) $x \mapsto \Psi(x)\beta/(x^{-\beta}-1)$ is not constant, $\Phi(t) = o(1)$ and $R(t,x) = o(\Phi(t))$ as $t \downarrow 0$ for all $x > 0$.

Recall that the limiting function

$$F_\beta^{-1}(x) := \frac{x^{-\beta}-1}{\beta} \tag{2.2}$$

is the upper q.f. of a generalized Pareto distribution in von Mises representation. Indeed, up to a scale and location parameter, these are the only q.f.'s that satisfy expansion (2.1) with $\Psi \equiv 0 \equiv R$ (Pereira (1994), Theorem 3.1). This fact reflects their exceptional role in extreme value theory; for further results demonstrating the importance of the generalized Pareto distributions in the context of extreme value approximations we refer to Reiss (1989), Chapter 5.

Condition 2.1(i) is equivalent to our basic assumption $F \in D(G_\beta)$ (see (1.3)). Thus in this case (2.1) is not an extra condition, but just a definition of the remainder term $R(t,x)$. Notice that the second term $\Phi(t)\Psi(x)$ of the expansion can be cancelled by changing the normalizing function a (so that one arrives at case (i)) if Ψ is a multiple of F_β^{-1}.

Otherwise, i.e., in case (ii), Φ has to be δ-varying at 0 for some $\delta \geq 0$:

$$\lim_{t\downarrow 0} \Phi(tx)/\Phi(t) = x^\delta \tag{2.3}$$

for all $x > 0$ (de Haan and Stadtmüller (1996), Theorem 1). Moreover, according to de Haan and Stadtmüller (1996), Remark 3(ii), one has

$$\Psi(x) = c_\Psi \cdot \begin{cases} F_{\beta-\delta}^{-1}(x) & \delta > 0, \\ -x^{-\beta}\log x/\beta \quad \text{if} & \beta \neq \delta = 0, \\ \log^2 x & \beta = \delta = 0, \end{cases} \tag{2.4}$$

for some constant $c_\Psi \neq 0$ if the normalizing function a and the function Φ are chosen suitably. (A possible choice is given in the proof of Lemma 2.1.)

Check that if Condition 2.1(ii) holds for other functions $\tilde{a}$ and $\tilde{\Phi}$ and a function $\tilde{\Psi}$ of type (2.4), then $a(t)/\tilde{a}(t) - 1 = o(\Phi(t))$ and $\Phi(t)\Psi(x) \sim \tilde{\Phi}(t)\tilde{\Psi}(x)$ as $t \downarrow 0$ for all $x > 0$. (Here $f(t) \sim g(t)$ means $\lim_{t\downarrow 0} f(t)/g(t) = 1$.) In particular, one may assume $\Psi = \tilde{\Psi}$ and $\Phi(t) \sim \tilde{\Phi}(t)$.

Though the second-order condition 2.1(ii) is actually stronger than our basic assumption $F \in D(G_\beta)$, it is satisfied for all usual textbook distributions belonging to some max-domain of attraction, for example, by the extreme value d.f.'s ($\delta = 1$), loggamma d.f.'s ($\beta > 0$, $\delta = 0$), normal, Weibull and gamma d.f.'s ($\beta = \delta = 0$) and the Cauchy d.f. ($\beta = 1$, $\delta = 2$).

The relations $R(t,x) = o(1)$ and $R(t,x) = o(\Phi(t))$, respectively, hold locally uniformly for positive x, i.e., uniformly over all $x \in I$ where $I \subset (0,\infty)$ is an arbitrary compact interval (Dekkers and de Haan (1989), Lemma 2.2; de Haan and Stadtmüller (1996), Remark 4(i)). Furthermore, if the functions a, Φ and Ψ are chosen appropriately, then the following weighted uniform convergence result holds in a neighborhood of 0, too.

Lemma 2.1 [Drees (1998a), Lemma 2.1] *Suppose that Condition 2.1(i) or (ii) is satisfied. Then there exist functions a and Φ and a function Ψ satisfying (2.4) such that, in addition, for all $x_0 > 0$ and $\varepsilon > 0$*

$$\sup_{x\in(0,x_0]} x^{\beta+\varepsilon}|R(t,x)| = \begin{cases} o(1), & \textit{case (i)}, \\ o(\Phi(t)), & \textit{case (ii)}. \end{cases}$$

Remark 2.1 If $\beta \geq 0$, then the assertions of Lemma 2.1 hold true for all functions a, Φ and Ψ satisfying (2.1) and (2.4). □

In the sequel we will always assume that the functions a, Φ and Ψ are chosen in accordance with Lemma 2.1.

Most important for our further investigations of the statistical tail functionals $T(Q_n)$ will be an approximation of the empirical tail quantile function. As a first step, we utilize a well-known approximation of the upper order statistics of i.i.d. uniform r.v.'s U_i, $i \in I\!N$. Denote by ξ_i, $i \in I\!N$, independent standard exponential r.v.'s and let

$$S_i := \sum_{j=1}^{i} \xi_j. \tag{2.5}$$

Then by Theorem 5.4.3 of Reiss (1989) one has

$$\|\mathcal{L}((U_{n-i+1:n})_{1\leq i\leq k}) - \mathcal{L}((1 - S_i/n)_{1\leq i\leq k})\| = O(k/n)$$

with $\|\cdot - \cdot\|$ denoting the variational distance. Hence a quantile transformation yields

$$\|\mathcal{L}((X_{n-i+1:n})_{1\leq i\leq k}) - \mathcal{L}((F^{-1}(1 - S_i/n))_{1\leq i\leq k})\| = O(k/n) \tag{2.6}$$

where F^{-1} is some measurable extension of the q.f. to the whole real axis.

In the next step, we give a strong approximation of

$$\tilde{Q}_{n,k_n}(t) := \tilde{Q}_n(t) := F^{-1}(1 - S_{[k_n t]+1}/n), \qquad t \in [0,1], \tag{2.7}$$

which by (2.6) has the same asymptotic distributional behavior as Q_n, because $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence. Since $F^{-1}(1-t)$ diverges as $t \downarrow 0$ if the right endpoint of F is infinite (e.g., if $\beta > 0$), it is plausible that one has to introduce weight functions to tie down the empirical tail q.f. at $t = 0$. Let

$$\tilde{h}(t) := (t/\log\log(3/t))^{1/2}, \quad t \in [0,1], \tag{2.8}$$

and

$$\mathcal{H} := \{h : [0,1] \to [0,\infty) \,|\, h \text{ continuous}, \, h_{|(0,1]} > 0, \, \lim_{t\downarrow 0} h(t)/\tilde{h}(t) = 0\}. \tag{2.9}$$

Finally, for each $\beta \in \mathbb{R}$ and $h \in \mathcal{H}$ we define a seminorm on the space of real functions on the unit interval by

$$\|z\|_{\beta,h} := \sup_{t\in[0,1]} t^\beta h(t)|z(t)|. \tag{2.10}$$

(By convention $0 \cdot \infty = 0$.)

Theorem 2.1 [Drees (1998a), Theorem 2.1] *Under Condition 2.1(i) or (ii) (with a, Φ and Ψ according to Lemma 2.1) there exist versions of $\tilde{Q}_n$ and a standard Brownian motion W such that almost surely for all $h \in \mathcal{H}$ and $\varepsilon > 0$*

$$\begin{aligned}
&\left\| \left(\frac{\tilde{Q}_n(t) - \tilde{D}_n}{a(k_n/n)} - \left(F_\beta^{-1}(t) - t^{-(\beta+1)} k_n^{-1} W(k_n t) + \Phi(k_n/n)\Psi(t) \right) \right)_{t\in[0,1]} \right\|_{\beta,h} \\
&= \quad o(k_n^{-1/2}) + \begin{cases} O\left(\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2} |R(k_n/n, x)| \right), & \textit{case (i)}, \\ o(\Phi(k_n/n)), & \textit{case (ii)}, \end{cases}
\end{aligned} \tag{2.11}$$

where

$$\tilde{D}_n = \begin{cases} F^{-1}(1 - k_n/n) & \qquad \qquad \beta \geq -1/2, \\ F^{-1}(1 - S_1/n) + a(k_n/n)\Big(\dfrac{1}{\beta} + \dfrac{c_\Psi}{\beta - \delta}\Phi(k_n/n) \cdot 1_{\{\delta>0\}}\Big) & \textit{if} \quad \beta < -1/2. \end{cases} \tag{2.12}$$

Observe that $-k_n^{-1/2} W(k_n \cdot)$ are standard Brownian motions, too. Thus, in view of (2.6) and Theorem 2.1, the normalized empirical tail q.f. has a non-degenerate limiting distribution if k_n tends to infinity sufficiently slowly. More precisely, corresponding to the two Conditions 2.1(i) and (ii), we will consider the following conditions:

Condition 2.2 *Assume that $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence such that*

(i) $\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2} |R(k_n/n, x)| = o(k_n^{-1/2})$ *for some $\varepsilon > 0$, or*

(ii) $k_n^{1/2}\Phi(k_n/n) \to \lambda \in [0,\infty]$.

If Condition 2.1(ii) holds (and hence 2.1(i) with remainder term $\tilde{R}(t,x) := \Phi(t)\Psi(x) + R(t,x)$), then Condition 2.2(ii) with $\lambda > 0$ allows to use more upper order statistics for estimation than Condition 2.2(i). This is due to the fact that in this case expansion (2.1) contains some extra information about the "direction" Ψ in which the tail of F^{-1} deviates from the ideal generalized Pareto q.f. F_β^{-1}.

In view of Theorem 2.1,

$$D_{\beta,h} := \Big\{ z : [0,1] \to \bar{\mathbb{R}} \mid \lim_{t\downarrow 0} t^\beta h(t) z(t) = 0,\ (t^\beta h(t) z(t))_{t\in[0,1]} \in D[0,1] \Big\}, \tag{2.13}$$

equipped with the weighted supremum seminorm $\|\cdot\|_{\beta,h}$, is the appropriate function space in which weak convergence of Q_n is established.

Since F_β^{-1} and Ψ need not belong to $D_{\beta,h}$ if $\beta < -1/2$, one has to consider certain modifications of these functions. Let

$$\bar{F}_\beta^{-1}(t) := F_\beta^{-1}(t) + 1/\beta \cdot 1_{\{\beta\neq 0\}} = \begin{cases} t^{-\beta}/\beta & \text{if } \beta \neq 0, \\ -\log t & \text{if } \beta = 0, \end{cases} \tag{2.14}$$

and for Ψ satisfying (2.4)

$$\bar{\Psi}(x) := \Psi(x) + \begin{cases} c_\Psi/(\beta-\delta) & \text{if } \beta \neq \delta > 0, \\ 0 & \text{if } \beta = \delta \text{ or } \delta = 0 \end{cases} = c_\Psi \cdot \begin{cases} \bar{F}_{\beta-\delta}^{-1}(x) & \text{if } \delta > 0, \\ -\bar{F}_\beta^{-1}(x)\log x & \text{if } \delta = 0. \end{cases} \tag{2.15}$$

Finally, define a Gaussian process V_β by

$$V_\beta(t) := t^{-(\beta+1)} W(t), \quad t \in [0,1], \tag{2.16}$$

where again W denotes a standard Brownian motion.

Corollary 2.1 [Drees (1998a), Corollary 2.1] *With*

$$D_n := \begin{cases} F^{-1}(1-k_n/n) - a(k_n/n)\Big(\dfrac{1_{\{\beta\neq 0\}}}{\beta} + \dfrac{c_\Psi}{\beta-\delta}\Phi(k_n/n)\cdot 1_{\{\beta\neq\delta>0\}}\Big) & \text{if } \beta \geq -1/2, \\ X_{n:n} & \text{if } \beta < -1/2, \end{cases} \tag{2.17}$$

the following assertions hold for all $h \in \mathcal{H}$:

(i) Under the Conditions 2.1(i) and 2.2(i) one has

$$k_n^{1/2}\Big(\frac{Q_n - D_n}{a(k_n/n)} - \bar{F}_\beta^{-1}\Big) \longrightarrow V_\beta$$

weakly in $D_{\beta,h}$.

(ii) Suppose Condition 2.1(ii) holds with a, Φ and Ψ according to Lemma 2.1. Then

$$\Phi(k_n/n) = O(k_n^{-1/2}) \quad \Longrightarrow \quad k_n^{1/2}\Big(\frac{Q_n - D_n}{a(k_n/n)} - \bar{F}_\beta^{-1} - \Phi(k_n/n)\bar{\Psi}\Big) \longrightarrow V_\beta$$

weakly in $D_{\beta,h}$. In particular, under Condition 2.2(ii) one has

$$\lambda < \infty \qquad \Longrightarrow \qquad k_n^{1/2}\Big(\frac{Q_n - D_n}{a(k_n/n)} - \bar{F}_\beta^{-1}\Big) \longrightarrow V_\beta + \lambda\bar{\Psi}$$

and, moreover,

$$\lambda = \infty \qquad \Longrightarrow \qquad \Phi(k_n/n)^{-1}\Big(\frac{Q_n - D_n}{a(k_n/n)} - \bar{F}_\beta^{-1}\Big) \longrightarrow \bar{\Psi}$$

weakly in $D_{\beta,h}$.

Remarks 2.2 (i) Note that (2.11) holds almost surely *simultaneously for all $h \in \mathcal{H}$*. Therefore, one may derive a result analogous to Corollary 2.1 in terms of weak convergence in the space $D_\beta := \bigcap_{h\in\mathcal{H}} D_{\beta,h}$ equipped with the topology that is induced by the family $\|\cdot\|_{\beta,h}$, $h \in \mathcal{H}$. Though that approach allows of a slight generalization of the results of the next sections, we do not go into details, since the improvements are negligible for practical purposes (cf. Remark 2.3(ii)). Moreover, additional problems concerning the measurability of the random functions under consideration would arise, as the locally convex vector space D_β is not metrizable.

(ii) Using the Theorems 5.2.1 and 5.2.5 of Csörgő and Horváth (1993), which give bounds and approximations, respectively, of the weighted uniform tail quantile process, one can prove a stochastic approximation of Q_n directly under slightly different conditions on the weight function h. However, the present approach is more suitable for obtaining uniform results of the type needed in Section 3.2 (cf. proof of Lemma 3.6), since it only involves one sequence of exponential partial sums instead of a triangle scheme of Gaussian processes. Likewise, one may use $\mathcal{L}((X_{n-i+1:n})_{1\le i\le n}) = \mathcal{L}((F^{-1}(1 - S_i/S_{n+1}))_{1\le i\le n})$ to obtain an almost sure or stochastic approximation (depending on the rate of convergence of k_n) for suitable versions of Q_n, yet this would lead to more complicated proofs. □

Strong and stochastic approximations of the empirical tail q.f. were already given by Csörgő and Horváth (1993, p. 386 ff.) and Einmahl (1992, Section 3), among others, yet their results are not suitable for our purposes, because there the normalizing factor ($1/a(k_n/n)$ in Corollary 2.1) is a function of the argument t which depends on the unknown d.f. F. Furthermore, these results require stronger conditions, for example, the existence of a regularly varying Lebesgue density of F. Other related results are the famous approximations of the uniform empirical q.f. in terms of weighted supremum norms established by O'Reilly and their extensions to the case of non-uniform r.v.'s (see Shorack and Wellner (1986), Theorems 11.5.2 and 18.1.2).

2.2 Asymptotics for location-invariant statistical tail functionals

In this section we investigate the asymptotic distributional behavior of scale- and location-invariant statistical tail functionals $T(Q_n)$. For the time being, the extreme value index

β is fixed. In accordance with Corollary 2.1, we consider functionals T that are defined on span$(D_{\beta,h},1)$ for some $\beta \in I\!R$ and $h \in \mathcal{H}$, i.e., the space of all affine transformations of functions belonging to $D_{\beta,h}$.

Because of the invariance properties of T, roughly speaking, in case (ii) with $\lambda < \infty$ Corollary 2.1 yields

$$T(Q_n) \approx^d T(\bar{F}_\beta^{-1} + k_n^{-1/2}(V_\beta + \lambda \bar{\Psi})) \tag{2.18}$$

where "$\approx^d$" means "approximately equal in distribution". Hence one may expect consistency of $T(Q_n)$ only if $T(\bar{F}_\beta^{-1}) = \beta$. In the sequel we will always assume that the following conditions are fulfilled:

Condition 2.3 *For some $\beta \in I\!R$ and $h \in \mathcal{H}$ the functional $T : span(D_{\beta,h},1) \to I\!R$ satisfies*

(i) $T_{|D_{\beta,h}}$ is $\mathcal{B}(D_{\beta,h}), \mathcal{B}(I\!R)$-measurable (where $\mathcal{B}$ denotes the appertaining Borel-σ-field),

(ii) $T(az+b) = T(z) \qquad$ for all $z \in D_{\beta,h},\ a > 0,\ b \in I\!R,$

(iii) $T(\bar{F}_\beta^{-1}) = \beta$.

Then the statistical tail functional $T(Q_n)$ is weakly consistent and asymptotically normal, respectively, depending on the degree of smoothness of T.

Theorem 2.2 [Drees (1998a), Theorem 3.1] *If $F \in D(G_\beta)$, $(k_n)_{n\in I\!N}$ is an intermediate sequence, T satisfies Condition 2.3 and, in addition, $T_{|D_{\beta,h}}$ is continuous at $\bar{F}_\beta^{-1}$, then*

$$T(Q_n) \longrightarrow \beta \qquad \textit{in probability.}$$

To obtain asymptotic normality, one clearly has to require some kind of differentiability of the functional $T_{|D_{\beta,h}}$. As in the asymptotic theory of classical statistical functionals (see Fernholz (1983) and Gill (1989)), it turns out that the concept of Hadamard (or compact) differentiability fits best to the present situation. Recall that a map T between topological vector spaces B_i, $i = 1,2$, is called *Hadamard differentiable tangentially to some subset $S \subset B_1$ at $x \in B_1$* if and only if there exists a continuous linear map $T'_x : B_1 \to B_2$, the Hadamard derivative, such that

$$\frac{T(x + t_n y_n) - T(x)}{t_n} \longrightarrow T'_x(y) \tag{2.19}$$

for all sequences $t_n \downarrow 0$ and $y_n \in B_1$ converging to $y \in S$ (Gill (1989), p. 102). Note that this condition is weaker than Fréchet differentiability, but stronger than Gateaux differentiability. In the present context it is sufficient to consider sequences $(y_n)_{n\in I\!N}$ converging to some function in

$$C_{\beta,h} := \{z \in D_{\beta,h} \mid z_{|(0,1]} \in C(0,1]\}, \tag{2.20}$$

because the paths of the limiting processes in Corollary 2.1 belongs to $C_{\beta,h}$ with probability 1. This leads to the following assumption:

Condition 2.4 *$T_{|D_{\beta,h}}$ is Hadamard differentiable tangentially to $C_{\beta,h}$ at $\bar{F}_\beta^{-1}$ with derivative T'_β.*

Of course, though T'_β is defined on $D_{\beta,h}$, it is well determined only on $C_{\beta,h}$.

For the description of the weak limits of the standardized empirical tail functional, it is convenient to identify the Hadamard derivative T'_β with an appropriate measure. To be more concrete, due to the Riesz representation theorem, there exists a unique signed measure, say $\nu_{T,\beta}$, on $[0,1]$ satisfying

$$\int_{[0,1]} \frac{1}{h(t)t^\beta} \, |\nu_{T,\beta}|(dt) < \infty \tag{2.21}$$

such that

$$T'_\beta(z) = \int_{[0,1]} z(t) \, \nu_{T,\beta}(dt) \tag{2.22}$$

for all $z \in C_{\beta,h}$. (Here $|\nu_{T,\beta}|$ denotes the total variation of $\nu_{T,\beta}$.)

Theorem 2.3 [Drees (1998a), Theorem 3.2] *Assume that $T : span(D_{\beta,h}, 1) \to I\!R$ satisfies the Conditions 2.3 and 2.4. Let*

$$\sigma^2_{T,\beta} := \int_{[0,1]^2} (st)^{-(\beta+1)} \min(s,t) \, \nu^2_{T,\beta}(ds, dt) \tag{2.23}$$

and

$$\mu_{T,\beta,\Psi} := \int_{[0,1]} \bar{\Psi}(t) \, \nu_{T,\beta}(dt).$$

(i) Under the Conditions 2.1(i) and 2.2(i) one obtains

$$\mathcal{L}\Big(k_n^{1/2}(T(Q_n) - \beta)\Big) \longrightarrow \mathcal{N}(0, \sigma^2_{T,\beta}) \qquad \textit{weakly.}$$

(ii) If the Conditions 2.1(ii) and 2.2(ii) are fulfilled, then

$$\lambda < \infty \qquad \Longrightarrow \qquad \mathcal{L}\Big(k_n^{1/2}(T(Q_n) - \beta)\Big) \longrightarrow \mathcal{N}(\lambda\mu_{T,\beta,\Psi}, \sigma^2_{T,\beta}) \quad \textit{weakly,}$$

and

$$\lambda = \infty \qquad \Longrightarrow \qquad \Phi(k_n/n)^{-1}(T(Q_n) - \beta) \longrightarrow \mu_{T,\beta,\Psi} \quad \textit{in probability.}$$

Remarks 2.3 (i) For some applications it is useful to define T only on a scale-invariant subset $\bar{D}_{\beta,h} \subset D_{\beta,h}$ such that $\bar{F}_\beta^{-1} \in \bar{D}_{\beta,h}$ and $P\{Q_n \in span(\bar{D}_{\beta,h}, 1)\} \to 1$; for example, one may restrict oneself to non-increasing functions belonging to $D_{\beta,h}$ (cf. Example 2.3). Then for the assertions of Theorem 2.3 it is sufficient to assume (2.19) for functions $\bar{F}_\beta^{-1} + t_n y_n \in \bar{D}_{\beta,h}$ only.

(ii) As already indicated in Remark 2.2(i), one may weaken Condition 2.4 slightly by requiring Hadamard differentiability merely on $D_\beta = \bigcap_{h\in\mathcal{H}} D_{\beta,h}$ equipped with the topology induced by $(\|\cdot\|_{\beta,h})_{h\in\mathcal{H}}$. However, since a linear map $D_\beta \to I\!R$ is continuous if and only if it is even continuous with respect to $\|\cdot\|_{\beta,h}$ for some $h \in \mathcal{H}$ (Dieudonné (1970), 12.14.11), one would obtain the same possible limiting distributions as in Theorem 2.3. In particular, one may not expect that under these weaker conditions there exist more efficient statistical tail functionals.

(iii) In general it is impossible to derive asymptotic confidence intervals directly from Theorem 2.3, since the parameters of the limit distribution are unknown. While it is easy to estimate the variance of the limit distribution consistently if it depends on the extreme value index β continuously, the estimation of the asymptotic bias $\Phi(k_n/n)\mu_{T,\beta,\Psi}$ is more involved. In Chapter 3 asymptotic confidence intervals are constructed if the number of order statistices used for estimation is chosen adaptively. □

In the following example Theorem 2.3 is applied to the Pickands and the generalized probability weighted moment estimators considered in the Examples 2.1(i) and (iv).

Examples 2.2 (i) The functional T_P pertaining to the Pickands estimator can be defined on $\bigcup_{\beta\in I\!R, h\in\mathcal{H}} D_{\beta,h}$. It fulfills the Conditions 2.3 and 2.4 with Hadamard derivative

$$T'_{P,\beta}(z) = \frac{\beta}{(2^\beta-1)\log 2}\Big(z(1) - (1+2^{-\beta})z(1/2) + 2^{-\beta}z(1/4)\Big)$$

and appertaining signed measure

$$\nu_{P,\beta} = \frac{\beta}{(2^\beta-1)\log 2}\Big(\varepsilon_1 - (1+2^{-\beta})\varepsilon_{1/2} + 2^{-\beta}\varepsilon_{1/4}\Big),$$

where ε_x denotes the Dirac-measure with mass 1 at x. Thus, under the second-order Condition 2.1(ii) and Condition 2.2(ii) with $\lambda < \infty$, Theorem 2.3 yields

$$\mathcal{L}\Big(k_n^{1/2}(T_P(Q_n) - \beta)\Big) \longrightarrow \mathcal{N}(\lambda\mu_{P,\beta,\Psi}, \sigma^2_{P,\beta})$$

with

$$\mu_{P,\beta,\Psi} = \frac{\beta}{(2^\beta-1)\log 2}(-(1+2^{-\beta})\Psi(1/2) + 2^{-\beta}\Psi(1/4))$$

and

$$\sigma^2_{P,\beta} = \Big(\frac{\beta}{(2^\beta-1)\log 2}\Big)^2(1+2^{2\beta+1}).$$

This result was already established by Dekkers and de Haan (1989) for $\lambda = 0$; moreover, it is contained as a special case in Drees (1995), Theorem 2.1.

(ii) For the functional T_{PWM,ν_1,ν_2} defined in Example 2.1(iv), Condition 2.3 reads as

(a) $|\nu_i|\,[0,1] < \infty, \quad \int (h(t)t^\beta)^{-1}\,|\nu_i|(dt) < \infty, \quad i = 1,2,$

(b) $\nu_i[0,1] = 0, \quad i = 1,2,$

(c) $\int \bar{F}_\beta^{-1}\,d\nu_2 \neq 0, \quad \int \bar{F}_\beta^{-1}\,d\nu_1 = \beta\cdot\int \bar{F}_\beta^{-1}\,d\nu_2.$

Then T_{PWM,ν_1,ν_2} is Hadamard differentiable at $\bar{F}_\beta^{-1}$ and the measure pertaining to the derivative is

$$\nu_{PWM,\beta} = (\nu_1 - \beta\nu_2)/\int \bar{F}_\beta^{-1}\,d\nu_2.$$

(For all $\beta \in I\!R$ these conditions are satisfied, for example, if $\nu_1[0,t] = tf(t)$ and $\nu_2[0,t] = \int_0^t f(x)\,dx$ where f is some right continuous function such that $f(1) = 0 = \int_0^1 f(x)\,dx \neq \int_0^1 \bar{F}_\beta^{-1}(x)f(x)\,dx$ and $\int (h(t)t^\beta)^{-1}\,d(tf(t))$ exists in $I\!R$.) □

Next, we examine the limiting distribution of statistical tail functionals $T(Q_n)$ if the Conditions 2.3 and 2.4 are satisfied for all parameter values β belonging to a given open set. Then Condition 2.3(ii) and (iii) imply certain restrictions for the possible Hadamard derivatives of T and thus for the limiting distributions.

Lemma 2.2 [Drees (1998a), Lemma 4.1] *Suppose that $T : span(\bigcup_{\beta\in I} D_{\beta,h}, 1) \to I\!R$ satisfies the Conditions 2.3 and 2.4 for all β belonging to some open set $I \subset I\!R$. Then the signed measures $\nu_{T,\beta}$ pertaining to the Hadamard derivatives T'_β, $\beta \in I$, satisfy*

(i) $\int \bar F_\beta^{-1}\, d\nu_{T,\beta} = 0$,

(ii) $\nu_{T,\beta}[0,1] = 0$ *if the constant function 1 belongs to* $D_{\beta,h}$,

(iii) $\int \bar F_\beta^{-1}(t) \log t\, \nu_{T,\beta}(dt) = \begin{cases} -1 & \text{if } \beta \neq 0, \\ -2 & \text{if } \beta = 0. \end{cases}$

Remarks 2.4 (i) The constant function 1 belongs to $D_{\beta,h}$ for all $\beta \geq -1/2$ and all $h \in \mathcal{H}$. Moreover, if $\lim_{t\downarrow 0} t^{-\varepsilon} h(t)/\tilde h(t) = \infty$ for all $\varepsilon > 0$ (e.g., $h(t) = \tilde h(t)/\log(2/t)$), then the function 1 does not belong to $D_{\beta,h}$ for $\beta < -1/2$.

(ii) Obviously, one might generalize Lemma 2.2 to the case where $h \in \mathcal{H}$ depends on β, too, i.e., $T : span(\bigcup_{\beta\in I} D_{\beta,h_\beta}, 1) \to I\!R$ where $h_\beta \in \mathcal{H}$ for all $\beta \in I$. □

Now one can characterize those statistical functionals T satisfying the Conditions 2.3 and 2.4 that yield a minimal asymptotic variance by calculating the signed measures pertaining to their Hadamard derivatives. Observe that in the situation of Theorem 2.3(i) a minimal asymptotic variance leads to an asymptotically minimal expected normalized loss $E(\ell(k_n^{1/2}|T(Q_n) - \beta|))$ if $\ell : [0,\infty) \to [0,\infty)$ is a bounded, almost everywhere continuous and increasing function; in particular, the coverage probabilities $P\{|T(Q_n) - \beta| \leq c k_n^{-1/2}\}$, $c \in (0,\infty)$, are asymptotically maximal.

For $\beta \in I\!R$ denote by $\mathcal{M}(\beta)$ the set of all signed measures satisfying (2.21) and the conditions established in Lemma 2.2(i)–(iii) for some $h \in \mathcal{H}$.

Theorem 2.4 [Drees (1998a), Theorem 4.1] *For $\nu \in \mathcal{M}(\beta)$ let*

$$\sigma^2_{\beta,\nu} := \int_{[0,1]^2} (st)^{-(\beta+1)} \min(s,t)\, \nu^2(ds,dt).$$

Then

$$\inf_{\nu\in\mathcal{M}(\beta)} \sigma^2_{\beta,\nu} = \begin{cases} (\beta+1)^2 & \text{if } \beta \geq -1/2, \\ \beta^2 & \text{if } \beta < -1/2. \end{cases} \tag{2.24}$$

Moreover, for $\beta \neq -1/2$ the infimum is attained at

$$\nu^*_\beta = f_\beta \cdot \lambda_{|(0,1)} + \rho_\beta \cdot \varepsilon_1 \tag{2.25}$$

with

$$f_\beta(t) := \begin{cases} (\beta+1)^2(t^\beta - (2\beta+1)t^{2\beta})/\beta & \beta > -1/2, \beta \neq 0, \\ -(\log t + 2) & \text{if} \quad \beta = 0, \\ \beta t^\beta & \beta < -1/2, \end{cases}$$

$0 < t < 1$, *and*

$$\rho_\beta := \begin{cases} \beta+1 & \beta > -1/2, \\ -\beta & \text{if} \quad \beta < -1/2. \end{cases}$$

For $\beta = -1/2$ the infimum is not attained. (Here $\lambda_{|(0,1)}$ denotes the uniform distribution on $(0,1)$ and ε_1 the Dirac measure at 1.)

Note that for $\beta > -1/2$ the right-hand side of (2.24) equals the asymptotic variance of the maximum likelihood estimator in a generalized Pareto model (Smith (1987), Theorems 3.2, 5.1 and 9.3). For $\beta < -1/2$ the ML-estimator has asymptotic variance β^2 as given in (2.24) if first the unknown right endpoint of the distribution is replaced by a suitable estimator, e.g., by the maximum $X_{n:n}$. This choice leads to the Hill type estimator $k_n^{-1}\sum_{i=1}^{k_n}\log((X_{n-i+1:n} - X_{n:n})/(X_{n-k_n:n} - X_{n:n}))$; the Hadamard differentiability of the corresponding functional $T(z) := \int_0^1 \log((z(t)-z(0+))/(z(1)-z(0+)))\,dt$ can be proven in exactly the same way as the differentiability of the Hill functional in Example 2.6 below.

The following example demonstrates that also for $\beta > 0$ the ML-estimator is a statistical tail functional satisfying Conditions 2.3 and 2.4, thus proving that its asymptotic normality follows from Theorem 2.3.

Example 2.3 Recall that for sufficiently large n the normalized order statistics $(X_{n-i+1:n} - X_{n-k_n:n})_{1\le i\le k_n}$ are approximately distributed as the order statistics of k_n generalized Pareto r.v.'s with d.f. $x \mapsto 1-(1+\beta/\sigma x)^{-1/\beta}$ where σ is a suitable scale parameter (cf. Pickands (1975), Theorem 7). Hence it is reasonable to estimate β and σ by maximizing the likelihood of the "observations" $(X_{n-i+1:n} - X_{n-k_n:n})_{1\le i\le k_n}$ in the generalized Pareto model. The ML-estimator of β can be computed as the first coordinate of a solution (β,σ) of the equations

$$\eta(Q_n,\beta,\sigma) = (\eta_1,\eta_2)(Q_n,\beta,\sigma) = (0,0)$$

where

$$\begin{aligned} \eta_1(z,\beta,\sigma) &:= \int_0^1 \frac{1}{1+\beta/\sigma(z(t)-z(1))}\,dt - \frac{1}{\beta+1} \qquad \text{and} \\ \eta_2(z,\beta,\sigma) &:= \int_0^1 \log(1+\beta/\sigma(z(t)-z(1)))\,dt - \beta. \end{aligned}$$

Now assume $\beta_0 > 0$ and, for the sake of simplicity, $t^{-3/4}h(t) \to \infty$ as $t \downarrow 0$. Consider $\eta(\bar F_{\beta_0}^{-1},\cdot,\cdot)$ as an element of the space $C^2(\Theta)$ of continuous functions that map $\Theta := [\beta_0/2, 2\beta_0] \times [1/2, 2]$ into $I\!R^2$ equipped with the supremum norm. It is easily seen that $\eta(\bar F_{\beta_0}^{-1},\cdot,\cdot)$ is differentiable at its zero $(\beta_0, 1)$ with invertible Jacobian matrix

$$J_{\eta,\beta_0} = \frac{\beta_0}{\beta_0+1}\begin{pmatrix} 1/((\beta_0+1)(2\beta_0+1)) & 1/(2\beta_0+1) \\ -1 & -1 \end{pmatrix}.$$

Hence, according to the implicit function theorems 1.4.7 and 1.4.2 in the book by Rieder (1994), there exist a neighborhood U of $\eta(\bar F_{\beta_0}^{-1},\cdot,\cdot)$ and a map $\tilde T: U\to\Theta$ such that $\xi(\tilde T(\xi))=0$ for all $\xi\in U$. Moreover, $\tilde T$ is unique on the subset of continuously differentiable functions in U and it is Hadamard differentiable at $\eta(\bar F_{\beta_0}^{-1},\cdot,\cdot)$ with derivative

$$\tilde T'_{\beta_0}(\zeta) = -J_{\eta,\beta_0}^{-1}\zeta(\beta_0,1).$$

It remains to prove that the map

$$\eta\ :\ \bar D_{\beta_0,h} := \{z\in D_{\beta_0,h}\,|\, z \text{ non-increasing}\}\longrightarrow C^2(\Theta),\quad z\longmapsto \eta(z,\cdot,\cdot)$$

is Hadamard differentiable at $\bar F_{\beta_0}^{-1}$ tangentially to $C_{\beta_0,h}$ with derivative

$$\begin{aligned}\eta'_{\beta_0}\ &:\ C_{\beta_0,h}\longrightarrow C^2(\Theta)\\ \eta'_{\beta_0}(z)\,(\beta,\sigma) &:= \Big(-\int_0^1 \frac{\beta(z(t)-z(1))}{\sigma(1+\beta/(\beta_0\sigma)(t^{-\beta_0}-1))^2}\,dt, \int_0^1 \frac{\beta(z(t)-z(1))}{\sigma(1+\beta/(\beta_0\sigma)(t^{-\beta_0}-1))}\,dt\Big)\end{aligned}$$

(cf. Remark 2.3(i)). For then, by the chain rule, the ML-functional $T_{ML}=pr_1\circ\tilde T\circ\eta:\bar D_{\beta_0,h}\to I\!R$, which maps z to the first coordinate of the unique solution of $\eta(z,\beta,\sigma)=0$ in the neighborhood of $\bar F_{\beta_0}^{-1}$, is Hadamard differentiable at $\bar F_{\beta_0}^{-1}$ tangentially to $C_{\beta_0,h}$ with derivative

$$\begin{aligned}T'_{ML,\beta_0}(z) &= pr_1(\tilde T'_{\beta_0}(\eta'_{\beta_0}(z))) \;=\; -pr_1\Big(J_{\eta,\beta_0}^{-1}\eta'_{\beta_0}(z)(\beta_0,1)\Big)\\ &= \frac{(\beta_0+1)^2}{\beta_0}(2\beta_0+1,1)\begin{pmatrix} -\int_0^1\beta_0 t^{2\beta_0}(z(t)-z(1))\,dt\\ \int_0^1 \beta_0 t^{\beta_0}(z(t)-z(1))\,dt\end{pmatrix}\\ &= \frac{(\beta_0+1)^2}{\beta_0}\int_0^1 (t^{\beta_0}-(2\beta_0+1)t^{2\beta_0})(z(t)-z(1))\,dt,\end{aligned}$$

from which it is immediate that $\nu^*_{\beta_0}$ defined by (2.25) is the pertaining measure.

We restrict ourselves to proving the Hadamard differentiability of the second coordinate of η, because the proof for the first component is even simpler. To this end, consider a sequence $z_n := \bar F_{\beta_0}^{-1}+\lambda_n y_n\in\bar D_{\beta_0,h}$, where $\lambda_n\to 0$ and $y_n\to y\in C_{\beta_0,h}$, and let $s_n:=\lambda_n^{5/4}$ and $\rho:=\beta/\sigma>0$. Then by the monotonicity of z_n

$$\begin{aligned}&\int_0^{s_n}\log(1+\rho(z_n(t)-z_n(1)))-\log(1+\rho(\bar F_{\beta_0}^{-1}(t)-\bar F_{\beta_0}^{-1}(1)))\,dt\\ &\quad\geq\ -\int_0^{s_n}\log(1+\rho/\beta_0(t^{-\beta_0}-1))\,dt = o(\lambda_n) \qquad\text{and}\qquad (2.26)\\ &\int_0^{s_n}\log(1+\rho(z_n(t)-z_n(1)))-\log(1+\rho(\bar F_{\beta_0}^{-1}(t)-\bar F_{\beta_0}^{-1}(1)))\,dt\\ &\quad=\ \int_0^{s_n}\log\Big(1+\lambda_n\rho\frac{y_n(t)-y_n(1)}{1+\rho F_{\beta_0}^{-1}(t)}\Big)\,dt\leq\int_0^{s_n}\lambda_n\rho\frac{|y_n(t)-y_n(1)|}{1+\rho F_{\beta_0}^{-1}(t)}\,dt = o(\lambda_n)\quad (2.27)\end{aligned}$$

uniformly for $(\beta,\sigma)\in\Theta$.

Furthermore, because of $|\log(1+x)-x| \le x^2/(2(1+x))$ for $x>-1$ and

$$\sup_{s_n\le t\le 1} \frac{\lambda_n |y_n(t)-y_n(1)|}{1+\rho \bar F_{\beta_0}^{-1}(t)} = O\Big(\sup_{s_n\le t\le 1}\frac{\lambda_n}{h(t)}\Big) = o(\lambda_n s_n^{-3/4}) = o(1),$$

one has for sufficiently large n

$$\begin{aligned} &\Big|\int_{s_n}^1 \log(1+\rho(z_n(t)-z_n(1))) - \log(1+\rho \bar F_{\beta_0}^{-1}(t)) - \lambda_n\rho\frac{y_n(t)-y_n(1)}{1+\rho \bar F_{\beta_0}^{-1}(t)}\,dt\Big| \\ &\le \int_{s_n}^1 \Big(\lambda_n\rho\frac{y_n(t)-y_n(1)}{1+\rho \bar F_{\beta_0}^{-1}(t)}\Big)^2 dt = O\Big(\lambda_n^2\int_{s_n}^1 \frac{1}{h^2(t)}\,dt\Big) = O(\lambda_n^2 s_n^{-1/2}) = o(\lambda_n) \quad (2.28)\end{aligned}$$

uniformly for $(\beta,\sigma)\in\Theta$. Combining (2.26)–(2.28), one arrives at

$$\Big|\eta_2(z_n,\beta,\sigma)-\eta_2(\bar F_{\beta_0}^{-1},\beta,\sigma)-\lambda_n\int_0^1 \frac{\beta(y_n(t)-y_n(1))}{\sigma(1+\rho \bar F_{\beta_0}^{-1}(t))}\,dt\Big| = o(\lambda_n)$$

uniformly for $(\beta,\sigma)\in\Theta$, from which the asserted Hadamard differentiability is obvious. □

In the case $-1/2<\beta<0$ one has to deal with additional technical difficulties. For then the zero $(\beta_0,1)$ of the map $\eta(\bar F_{\beta_0}^{-1},\cdot,\cdot)$ lies at the boundary of the domain so that the usual implicit function theorems do not apply. We do not try to overcome this problem by extending $\eta(\bar F_{\beta_0}^{-1},\cdot,\cdot)$ differentiably to an open neighborhood of $(\beta_0,1)$, because below we give a general method to construct estimators of β with prescribed asymptotic behavior, for example, with minimal asymptotic variance.

So far we have only considered the case where the limit distribution is centered. However, Theorem 2.3 shows that under Condition 2.1(ii) in general one obtains a faster rate of convergence if one chooses k_n such that the asymptotic bias is not negligible any longer. In that situation, usually the asymptotic mean squared error (AMSE) is regarded as an appropriate measure of the quality of an estimator for β (see, e.g., Hall and Welsh (1985), Csörgő et al. (1985), Smith (1987), Dekkers and de Haan (1993), Pereira (1994) and Alves (1995)), though there is no simple decision theoretical justification for this criterion unless convergence of the second moment of the standardized estimation error holds. In this context, it is assumed that the sequence $(k_n)_{n\in\mathbb{N}}$ is chosen such that the AMSE is minimized. Alternatively, one may consider the "minimal length" of symmetric intervals around the estimator that cover the true parameter β with probability converging to 1. (A uniform version of this optimality criterion is considered in Chapter 4 in more detail.) It turns out that, for both criteria, the asymptotic performance of statistical tail functionals strongly depends on whether δ equals 0 or not.

Let us start with the case $\delta=0$. Then under the conditions of Theorem 2.3(ii) the asymptotic bias is proportional to $\mu_{T,\beta,\Psi}=\int_{[0,1]}\bar\Psi(t)\,\nu_{T,\beta}(dt)$, where again $\nu_{T,\beta}$ denotes the measure pertaining to the Hadamard derivative of T at $\bar F_\beta^{-1}$. On the other hand, the definition of $\bar\Psi$ and Lemma 2.2(iii) imply

$$\mu_{T,\beta,\Psi} = c_\Psi\cdot\begin{cases} 1 & \text{if } \beta\ne 0, \\ 2 & \phantom{\text{if }}\beta=0\end{cases} \tag{2.29}$$

and thus the following result is proven.

Theorem 2.5 [Drees (1998a), Theorem 4.2] *If the Conditions 2.1(ii) and 2.2(ii) hold for $\delta = 0$, then all statistical tail functionals $T(Q_n)$ with T satisfying the Conditions 2.3 and 2.4 have the same asymptotic bias if they are based on the same number of largest order statistics.*

Even more surprising, the error of these estimators is asymptotically the same if the number $k_n + 1$ of largest order statistics is chosen in an optimal way.

Corollary 2.2 [Drees (1998a), Corollary 4.1] *Denote by $(k_n^{(1)})_{n\in\mathbb{N}}$ a sequence such that*

$$\lim_{n\to\infty} \Phi(k_n^{(1)}/n)(k_n^{(1)})^{1/2} = 1. \tag{2.30}$$

If the Conditions 2.1(ii) and 2.2(ii) hold with $\delta = 0$, then the asymptotic mean squared error of $T(Q_n)$ is minimal for any sequence $(k_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \frac{k_n^{(1)}}{k_n} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{\Phi(k_n^{(1)}/n)}{\Phi(k_n/n)} = 1. \tag{2.31}$$

In this case

$$(k_n^{(1)})^{1/2}(T(Q_{n,k_n}) - \beta) \longrightarrow c_\Psi \cdot \begin{cases} 1 & \beta \neq 0, \\ 2 \quad \text{if} & \beta = 0. \end{cases} \tag{2.32}$$

Moreover, $\liminf_{n\to\infty} P\{|T(Q_{n,k_n}) - \beta| \leq c(k_n^{(1)})^{-1/2}\} < 1$ *for all* $c < |c_\Psi|(1 + 1_{\{\beta=0\}})$*, and* $\liminf_{n\to\infty} P\{|T(Q_{n,k_n}) - \beta| \leq c(k_n^{(1)})^{-1/2}\} = 1$ *for all* $c > |c_\Psi|(1 + 1_{\{\beta=0\}})$ *if and only if* $(k_n)_{n\in\mathbb{N}}$ *satisfies (2.31).*

Remarks 2.5 (i) Condition (2.30) is equivalent to $f(n/k_n^{(1)}) \sim n^{1/2}$ (i.e., $n^{-1/2}f(n/k_n^{(1)}) \to 1$) where $f(t) := t^{1/2}/\Phi(1/t)$. By Bingham et al. (1987), Theorem 1.5.12, there exists an 'asymptotic inverse' g of f such that this in turn is equivalent to $k_n^{(1)} \sim n/g(n^{1/2})$. In particular, the sequence $(k_n^{(1)})_{n\in\mathbb{N}}$ is asymptotically uniquely determined and, moreover, a $(2\delta/(2\delta+1))$−varying function of n.

(ii) According to the uniform convergence theorem for regularly varying functions, for all $i \in \mathbb{N}$ there exists a natural n_i such that for all $n \geq n_i$

$$\sup_{x\in[1,i]} \left| \frac{\Phi(xk_n^{(1)}/n)}{\Phi(k_n^{(1)}/n)} - 1 \right| \leq 1/i.$$

W.l.o.g. one may assume that $n_i \uparrow \infty$ as $i \to \infty$. Then obviously the sequence $(k_n)_{n\in\mathbb{N}}$ defined by $k_n := [ik_n^{(1)}]$ if $n_i \leq n < n_{i+1}$ satisfies (2.31). Note that this sequence is not asymptotically unique, since (2.31) holds for every sequence $(\tilde{k}_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \tilde{k}_n/k_n = \lim_{n\to\infty} k_n^{(1)}/\tilde{k}_n = 0$. □

So if $\delta = 0$ then the choice (2.31) minimizes the AMSE and, simultaneously, it minimizes asymptotically the minimal length of a symmetric interval with center $T(Q_{n,k_n})$ that contains the true value β with probability converging to 1. For this choice all statistical tail functionals under consideration have the same asymptotic behavior. Moreover, the rate of convergence of the AMSE and the minimal length of the symmetric interval is slower than every power of $1/n$.

Example 2.4 The normal d.f. satisfies Condition 2.1(ii) with $\Phi(t) = |\log t|^{-1}$ and $\Psi(x) = -\log^2 x/4$. Hence $k_n^{(1)} \sim \log^2 n$ and a choice for k_n is asymptotically optimal if and only if $\log^2 n = o(k_n)$ and $\log k_n = o(\log n)$, e.g., $k_n = \log^a n$ for some $a > 2$. Then $\Phi(k_n/n)^{-1}(T(Q_n) - \beta) \sim \log n(T(Q_n) - \beta) \to -1/2$. □

The situation changes completely if $\delta > 0$. Then, in view of Theorem 2.3(ii), the optimal choice for k_n leads to an asymptotic root mean squared error of order $(k_n^{(1)})^{-1/2}$ with $k_n^{(1)}$ satisfying (2.30) if the bias term $\mu_{T,\beta,\Psi}$ does not equal 0.

Otherwise, i.e., in case of $\mu_{T,\beta,\Psi} = 0$, from Theorem 2.3(ii) and a diagonal procedure similar to the one used in Remark 2.5(ii) it follows that for a suitable choice for k_n one obtains a faster rate of convergence. For according to Theorem 2.3(ii) for all $m \in I\!N$ there exists $n_m \in I\!N$ such that for all $n \geq n_m$

$$\sup_{x\in I\!R} \Big| P\Big\{[mk_n^{(1)}]^{1/2}(T(Q_{n,[mk_n^{(1)}]}) - \beta) \leq x\Big\} - \mathcal{N}(0, \sigma_{T,\beta}^2)(-\infty, x]\Big| \leq 1/m.$$

Since one may assume that $n_m \uparrow \infty$, with $k_n := [mk_n^{(1)}]$ if $n_m \leq n < n_{m+1}$ one has

$$\mathcal{L}\Big(k_n^{1/2}(T(Q_{n,k_n}) - \beta)\Big) \longrightarrow \mathcal{N}(0, \sigma_{T,\beta}^2) \qquad \text{weakly;} \tag{2.33}$$

in particular, the AMSE equals $k_n^{-1}\sigma_{T,\beta}^2$, which is of smaller order than the minimal AMSE in case of $\mu_{T,\beta,\Psi} \neq 0$.

Likewise, if $a_n = o((k_n^{(1)})^{-1/2})$, but $k_n^{-1/2} = o(a_n)$, then $\lim_{n\to\infty} P\{|T(Q_{n,\tilde{k}_n}) - \beta| \leq a_n\} = 0$ for all intermediate sequences $(\tilde{k}_n)_{n\in I\!N}$ if $\mu_{T,\beta,\Psi} \neq 0$, but $\lim_{n\to\infty} P\{|T(Q_{n,k_n}) - \beta| \leq a_n\} = 1$ if $\mu_{T,\beta,\Psi} = 0$. (If one assumes a certain expansion for T in the neighborhood of $\bar{F}_\beta^{-1}$ instead of merely the Hadamard differentiability, then one can give more explicit conditions on the sequence $(k_n)_{n\in I\!N}$ in terms of $\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n, x)|$ and $\Phi(k_n/n)$ that ensure (2.33).)

Thus, for fixed parameters β and δ, a statistical tail functional can have a minimal AMSE for optimal k_n only if the bias term $\mu_{T,\beta,\Psi}$ vanishes. This fact demonstrates that no statistical tail functional minimizes the AMSE for suitably chosen k_n uniformly for all d.f.'s satisfying Condition 2.1(ii), because $\mu_{T,\beta,\Psi} = 0$ for all $\delta > 0$ clearly implies that $\nu_{T,\beta}$ is the zero measure, which is a contradiction to Lemma 2.2(iii). In fact, we have even shown that no statistical tail functional converges with the optimal rate uniformly for all such d.f.'s. (Note that there is a vague resemblance with the well-known superefficiency, because a better rate can be achieved only for discrete parameter values δ, not for whole intervals, but we do not discuss this analogy in detail.)

While it is impossible to achieve $\mu_{T,\beta,\Psi} = 0$ for all $\delta > 0$, instead one may choose a *finite* set $D \subset (0, \infty)$ and a family of measures $\nu_{\beta,D}$ such that the bias vanishes if Condition 2.1(ii) holds with $\delta \in D$. However, the following example, which deals with the case $D = \{\delta_0\}$, shows that this approach has a serious disadvantage if the sample size is moderate.

Example 2.5 Fix some $\delta_0 > 0$. Then the asymptotic bias $\mu_{T,\beta,\Psi} = \int \bar{\Psi}\, d\nu_{T,\beta}$ given in Theorem 2.3 vanishes if the measure $\nu_{T,\beta}$ pertaining to the Hadamard derivative of T at

$\bar F_\beta^{-1}$ belongs to $\mathcal{M}(\beta,\delta_0) := \{\nu \in \mathcal{M}(\beta) \mid \int \bar F_{\beta-\delta_0}^{-1}\, d\nu = 0\}$ and Condition 2.1(ii) holds with $\delta = \delta_0$.

It suggests itself to minimize the asymptotic variance over $\mathcal{M}(\beta,\delta_0)$ using the methods of the proof of Theorem 2.4. It turns out that again the infimum is attained if $\beta \neq -1/2$. For example, if $\beta < -1/2$ then

$$\inf_{\nu\in\mathcal{M}(\beta,\delta_0)} \sigma^2_{\beta,\nu} = \sigma^2_{\beta,\nu^*(\beta,\delta_0)}$$

where

$$\begin{aligned}
\nu^*(\beta,\delta_0) &= f(\beta,\delta_0)\cdot\lambda_{|(0,1)} + \rho(\beta,\delta_0)\cdot\varepsilon_1,\\
f(\beta,\delta_0)(t) &= \frac{(\delta_0+1)^2}{\delta_0^2}\beta\Big(-(2\delta_0+1)t^{\beta+\delta_0}+t^\beta\Big), \quad 0<t<1,\\
\rho(\beta,\delta_0) &= \frac{\delta_0+1}{\delta_0}\beta.
\end{aligned}$$

Because many d.f.'s (e.g., the extreme value d.f.'s and the logistic d.f.) satisfy the second-order Condition 2.1(ii) with $\delta = 1$, one might feel tempted to use a functional T such that $\nu_{T,\beta} = \nu^*(\beta,1)$. Notice, however, that then the variance of the limiting normal distribution is four times as high as the minimal variance $\sigma^2_{\beta,\nu^*_\beta}$. So although, according to the discussion given above, $T(Q_n)$ is asymptotically superior to a statistical tail functional $\tilde T(\tilde Q_n)$ with $\nu_{\tilde T,\beta} = \nu^*_\beta$ if $\delta = 1$, because it may utilize more order statistics, this advantage has a favorable effect only for an extremely large sample size n, since the optimal number k_n for the functional T should be 16 times as high as the optimal $\tilde k_n$ for $\tilde T$. Moreover, the situation is even worse for smaller δ_0, because in general one has

$$\inf_{\nu\in\mathcal{M}(\beta,\delta_0)} \sigma^2_{\beta,\nu} = \frac{(\delta_0+1)^2}{\delta_0^2}\inf_{\nu\in\mathcal{M}(\beta)} \sigma^2_{\beta,\nu}. \qquad \square$$

So apparently there is no statistical tail functional that can be regarded as optimal in all situations. While in some econometric applications with very large sample size the approach described in the preceding example may be adequat if one has prior knowledge about the true second-order parameter δ, in other cases it might be more appropriate, e.g., to minimize the maximal AMSE under the assumption that δ belongs to some interval, and each criterion leads to a different *characterization* of an "optimal" functional T via its Hadamard derivatives. Hence clearly there is need of a flexible procedure to *construct* functionals T with Hadamard derivatives pertaining to a prescribed smooth family of signed measures ν_β. The following theorem shows that adaptive generalized probability weighted moment estimators (cf. Examples 2.1(iv) and 2.2(ii)) with made-to-order asymptotic behavior can be defined.

Theorem 2.6 [Drees (1998a), Theorem 4.3] *For some open set $I \subset I\!R$ and some $1/2-$varying function $h \in \mathcal{H}$ let $(\nu_\beta)_{\beta\in I}$ be a family of signed measures satisfying (2.21),* *the conditions described in Lemma 2.2(i)–(iii) (with $\nu_{T,\beta}$ replaced by ν_β) and the following continuity condition:*

$$\lim_{\beta_n\to\beta}\int_{[0,1]}\frac{1}{h(t)t^\beta}\,|\nu_{\beta_n}-\nu_\beta|(dt)=0 \tag{2.34}$$

for all $\beta\in I\setminus\{-1/2\}$. Furthermore, denote by $\tilde T$ a functional that is continuous at $\bar F_\beta^{-1}$ and satisfies Condition 2.3 (with T replaced by $\tilde T$) for all $\beta\in I$. Define for all $z\in span(1,\cup_{\beta\in I}D_{\beta,h})$

$$\begin{aligned} H_\beta(t) &:= t^\beta\int_{[0,t]}s^{-\beta}\,\nu_\beta(ds),\\ T_\beta(z) &:= \frac{\int_{[0,1]}z(t)\,dH_\beta(t)}{\int_{[0,1]}z(t)H_\beta(t)/t\,dt}\end{aligned}$$

(with the convention $x/0:=0$ for all $x\in I\!R$) and

$$T(z):=\begin{cases} T_{\tilde T(z)}(z) & \text{if } \tilde T(z)\ge -1/2,\\ T_{\tilde T(z)}(z-z(0+)) & \text{if } \tilde T(z)<-1/2\end{cases} \tag{2.35}$$

if the right-hand side is well defined, and $T(z):=0$ else. Then T fulfills the Conditions 2.3 and 2.4 and its Hadamard derivatives T'_β correspond to the prescribed measures ν_β for all $\beta\in I\setminus\{-1/2\}$.

Remark 2.6 Note that it is not a real restriction to assume that h is 1/2–varying at 0, because for all $h\in\mathcal H$ there exists a 1/2–varying function $\bar h\in\mathcal H$ with $h\le\bar h$, so that (2.21), (2.34) and the conditions of Lemma 2.2 also hold for $\bar h$ instead of h. To be more concrete, let $t_n:=\inf\{t\in[0,1]\,|\,h(t)/\tilde h(t)\ge 1/n\}$, $n\ge1$, which is positive by the definition of $\mathcal H$, where, by convention, $\inf\emptyset:=1$. Next define $\bar t_n:=\min(t_n/2,\bar t_{n-1}/(n+1))$ recursively, and denote by g the function on $[0,1]$ that is linear on $[\bar t_n,\bar t_{n-1}]$ with $g(\bar t_n)=1/(n-1)$ for $n\ge2$ and $g(t)=\max(1,\sup_{0<s\le1}h(s)/\tilde h(s))$ for $t\in[\bar t_1,1]$.

Then it is easily seen that $\bar h:=g\cdot\tilde h$ belongs to $\mathcal H$. Moreover, for $t\in[\bar t_{n+1},\bar t_n]$, $n\ge1$, one has $\bar h(t)\ge g(\bar t_{n+1})\tilde h(t)=\tilde h(t)/n>h(t)$, and $\bar h(t)\ge\tilde h(t)\sup_{0<s\le1}h(s)/\tilde h(s)\ge h(t)$ for $t\in[\bar t_1,1]$. Finally, g is slowly varying at 0 and hence $\bar h$ is 1/2–varying, as asserted, because for $x\ge1$, $n\ge\max(x,3)$ and $t\in[\bar t_{n+1},\bar t_n]$ one has $tx\le\bar t_{n-1}$, so that $1\le g(tx)/g(t)\le n/(n-2)$. □

Check that for $z=Q_n$ and $I=I\!R$ the right-hand side of (2.35) is always defined. Examples of families possessing the continuity property (2.34) are the variance minimizing measures $(\nu^*_\beta)_{\beta\ne-1/2}$ and $(\nu^*(\beta,\delta_0))_{\beta\ne-1/2}$. In particular, Theorem 2.6 yields an adaptive statistical tail functional with minimal variance if, for example, one chooses the Pickands estimator $T_P(Q_n)$ as the initial estimator $\tilde T(Q_n)$ of the extreme value index. In this case the particular choice of $\nu^*_{-1/2}$ does not matter asymptotically since $P\{T_P(Q_n)=-1/2\}\to 0$.

2.3 Non-location-invariant statistical tail functionals: the case $\beta > 0$

In this section we examine the asymptotic behavior of statistical tail functionals $T(Q_n)$ with T satisfying the following conditions:

Condition 2.5 *For some $\beta > 0$ and $h \in \mathcal{H}$ the functional $T : D_{\beta,h} \to I\!R$ fulfills*

(i) T is $\mathcal{B}(D_{\beta,h}), \mathcal{B}(I\!R)$-measurable,

(ii) $T(az) = T(z) \qquad \forall z \in D_{\beta,h},\ a > 0,$

(iii) $T(\bar{F}_\beta^{-1}) = \beta$.

Hence it is no longer assumed that T is location invariant, but β is restricted to the positive half axis. Note that in this case the constant functions belong to $D_{\beta,h}$, so that it is sufficient to define T on this function space.

In the preceding section we have seen that the asymptotic bias of a location-invariant statistical tail functional is caused by the term $\Phi(k_n/n)\bar{\Psi}$ in the approximation of the empirical tail quantile function. If one considers non-location-invariant functionals, then a second bias term, namely, $D_n/a(k_n/n)$, has to be taken into account. Hence it is crucial to know which of these two terms is of larger order.

Observe that for $\beta > 0$ one has

$$\frac{D_n}{a(k_n/n)} = \Xi(k_n/n) \quad \text{where} \quad \Xi(t) := \frac{F^{-1}(1-t)}{a(t)} - \frac{1}{\beta} - \frac{c_\Psi}{\beta - \delta}\Phi(t) \cdot 1_{\{\beta \neq \delta > 0\}}. \tag{2.36}$$

Under Condition 2.1(ii), roughly speaking, $\Phi(t)$ is of smaller order than $\Xi(t)$ as $t \downarrow 0$ if $\beta \leq \delta$, and it is of larger order for $\beta > \delta$. However, there is one exception to this rule. According to de Haan and Stadtmüller (1996), Theorem 2, if Condition 2.1(ii) holds for some $\delta-$varying function Φ with $\delta > 0$, then there exist constants $d_1 > 0$ and $d_2 \in I\!R$ and a function $g : (0,1) \to I\!R$ that is $(\delta - \beta)$–varying at 0 if $\delta \neq \beta$, and Π–varying at 0 with auxiliary function L, say, (i.e., $(g(tx) - g(t))/L(t) \to -\log x$ as $t \downarrow 0$) if $\beta = \delta$ such that

$$F^{-1}(1-t) = d_1 t^{-\beta} + d_2 + g(t). \tag{2.37}$$

(Note that, for the sake of simplicity, we do not require that a regularly or Π–varying function is positive, as it is often assumed, but it has eventually constant sign.) Now we have the following relation between Φ and Ξ.

Lemma 2.3 *Under Condition 2.1 one has*

$$\lim_{t \downarrow 0} \Xi(t) = 0. \tag{2.38}$$

If Condition 2.1(ii) holds with a, Φ and Ψ according to Lemma 2.1, then

(i) $\Phi(t) = o(\Xi(t))$ and Ξ is β–varying at 0 $\qquad$ if $\beta = \delta$ $\quad$ or $\quad$ $\beta < \delta$ and $d_2 \neq 0$, and

(ii) $\Xi(t) = o(\Phi(t)) \qquad$ if $\beta > \delta$ $\quad$ or $\quad$ $\beta < \delta$ and $d_2 = 0$.

In view of the convergence of Ξ and Theorem 2.1, the following analog to Theorem 2.2 is obvious.

Theorem 2.7 *If $F \in D(G_\beta)$, $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence and T satisfies Condition 2.5 and is continuous at $\bar{F}_\beta^{-1}$, then*

$$T(Q_n) \longrightarrow \beta \qquad \textit{in probability.}$$

Corresponding to the two cases considered in Lemma 2.3, we need two different conditions on k_n to obtain a non-trivial weak limit for the statistical tail functionals under consideration.

Condition 2.6 *Assume that $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence satisfying*

(i) $\max(k_n^{1/2}, \Xi(k_n/n)^{-1}) \sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n,x)| \longrightarrow 0$ *for some $\varepsilon > 0$ and* $k_n^{1/2}\Xi(k_n/n) \longrightarrow \rho \in [0,\infty]$, *or*

(ii) $k_n^{1/2}\Phi(k_n/n) \longrightarrow \lambda \in [0,\infty]$.

Now we are ready to conclude the following counterpart of Theorem 2.3 from Theorem 2.1 and Lemma 2.3 by the same methods as used in Section 2.2.

Theorem 2.8 [Drees (1998b), Theorem 3.1] *Assume that T satisfies the Conditions 2.4 and 2.5, let $h \in \mathcal{H}$ and define $\nu_{T,\beta}$, $\sigma^2_{T,\beta}$ and $\mu_{T,\beta,\Psi}$ as in Section 2.2.*

(i) Suppose that Condition 2.6(i) is satisfied and
Condition 2.1(i) or
Condition 2.1(ii) with $\beta = \delta$ or with $\beta < \delta$ and $d_2 \neq 0$ (cf. (2.37)).
Then

$$\rho < \infty \qquad \Longrightarrow \qquad \mathcal{L}\Big(k_n^{1/2}(T(Q_n)-\beta)\Big) \longrightarrow \mathcal{N}(\rho\nu_{T,\beta}[0,1], \sigma^2_{T,\beta}) \quad \textit{weakly,}$$

and

$$\rho = \infty \qquad \Longrightarrow \qquad \Xi(k_n/n)^{-1}(T(Q_n)-\beta) \longrightarrow \nu_{T,\beta}[0,1] \quad \textit{in probability.}$$

(ii) If Condition 2.1(ii) holds with $\beta > \delta$ or $\beta < \delta$ and $d_2 = 0$, then under Condition 2.6(ii)

$$\lambda < \infty \qquad \Longrightarrow \qquad \mathcal{L}\Big(k_n^{1/2}(T(Q_n)-\beta)\Big) \longrightarrow \mathcal{N}(\lambda\mu_{T,\beta,\Psi}, \sigma^2_{T,\beta}) \quad \textit{weakly,}$$

and

$$\lambda = \infty \qquad \Longrightarrow \qquad \Phi(k_n/n)^{-1}(T(Q_n)-\beta) \longrightarrow \mu_{T,\beta,\Psi} \quad \textit{in probability.}$$

Under the conditions of Lemma 2.3(i) the rate of convergence of statistical tail functionals T with $\nu_{T,\beta}[0,1] \neq 0$ is slower than the rate for location-invariant estimators for suitably chosen k_n. On the other hand, the rates are the same under the assumptions of Lemma 2.3(ii), but the variance $\sigma^2_{T,\beta}$ of the limit distribution (or its mean squared error ${\mu_{T,\beta,\Psi}}^2 + \sigma^2_{T,\beta}$) is smaller if a suitable functional T is chosen that is not location invariant. For, in the present situation, together with the location invariance the condition on $\nu_{T,\beta}$ derived in Lemma 2.2(ii) is dropped, too.

Lemma 2.4 [Drees (1998b), Lemma 3.1] *If T satisfies the Conditions 2.4 and 2.5 for all β of some open set $I \subset (0,\infty)$, then*

(i) $\int \bar{F}_\beta^{-1}\, d\nu_{T,\beta} = 0 \quad \forall \beta \in I$,

(ii) $\int \bar{F}_\beta^{-1} \cdot \log\, d\nu_{T,\beta} = -1 \quad \forall \beta \in I$.

Since the set of admissible measures is smaller if the condition corresponding to the location invariance is added, one may expect that the minimal asymptotic variance is smaller in the present situation. In fact, whereas in Theorem 2.4 it was shown that the minimal asymptotic variance under location invariance equals $(\beta+1)^2$, using the same methods one can easily see that

$$\inf_{\nu \in \mathcal{M}^+(\beta)} \sigma^2_{\beta,\nu} = \beta^2$$

where $\mathcal{M}^+(\beta)$ denotes the set of all signed measures satisfying (2.21) and the conditions given in Lemma 2.4(i) and (ii). Moreover, the infimum is attained at

$$\nu_\beta^+ := f_\beta^+ \cdot \lambda_{|(0,1)} - \beta \cdot \varepsilon_1 \qquad \text{with} \qquad f_\beta^+(t) := \beta t^\beta$$

where again $\lambda_{|(0,1)}$ denotes the uniform distribtuion on (0,1) and ε_1 the Dirac measure at 1.

Observe that β^2 is the asymptotic variance of the Hill estimator. Indeed in Reiss (1989), Section 9.4, and Marohn (1995), Section 4, it is proven that under more restrictive conditions the Hill estimator is asymptotically optimal if the bias is negligible. In the following example it is shown that the asymptotic normality of Hill's estimator, which was established previously in several papers (see, e.g., Hall (1982), Davis and Resnick (1984), and Haeusler and Teugels (1985)), can be derived from Theorem 2.8.

Example 2.6 Recall from Example 2.1(ii) the definition of the Hill functional

$$T_H(z) := \int_0^1 \log(z(t)/z(1))\, dt$$

if the right-hand side is well defined and finite, and $T_H(z) := 0$ otherwise.

Obviously, T_H is scale invariant and $T_H(\bar{F}_\beta^{-1}) = \beta$. Now fix some $h \in \mathcal{H}$ that is increasing and regularly varying at 0 with exponent 1/2 (e.g., $h(t) = (t/(1-\log t))^{1/2}$) and define

$$\bar{D}_{\beta,h} := \{z \in D_{\beta,h} \mid z \text{ positive and non-increasing}\}.$$

Since $P\{Q_n \in \bar{D}_{\beta,h}\} = P\{X_{n-k_n:n} > 0\} \to 1$, it suffices to verify that $T_{H|\bar{D}_{\beta,h}}$ is Hadamard differentiable at $\bar{F}_\beta^{-1}$ tangentially to $C_{\beta,h}$ (cf. Remark 2.3(i)).

For this, let $\lambda_n \to 0$ and $y_n \in D_{\beta,h}$ denote a sequence converging to some $y \in C_{\beta,h}$ such that $\bar{F}_\beta^{-1} + \lambda_n y_n \in \bar{D}_{\beta,h}$ for all $n \in I\!N$. Then

$$T_H(\bar{F}_\beta^{-1} + \lambda_n y_n) - T_H(\bar{F}_\beta^{-1}) = \int_0^1 \log(1 + \lambda_n x_n(t))\, dt$$

where

$$x_n(t) := \beta \frac{t^\beta y_n(t) - y_n(1)}{1 + \beta\lambda_n y_n(1)}.$$

The monotonicity of $\bar{F}_\beta^{-1} + \lambda_n y_n$ implies $\log(1+\lambda_n x_n(t)) \geq -\beta \log(s/t) + \log(1+\lambda_n x_n(s))$ for all $t \leq s$ and hence

$$\int_0^s \log(1 + \lambda_n x_n(t))\, dt \geq s\Big(- \beta + \log(1 + \lambda_n x_n(s))\Big). \tag{2.39}$$

Moreover, for $s_n = \lambda_n^{7/6}$ one has

$$\sup_{t\in[s_n,1]} \log(1 + \lambda_n x_n(t)) \leq \lambda_n \sup_{t\in[s_n,1]} |x_n(t)| = O(\lambda_n/h(s_n)) = o(1). \tag{2.40}$$

Using (2.39), (2.40) and $\log(1+x) \leq x$ one obtains

$$\Big|\int_0^{s_n} \log(1 + \lambda_n x_n(t)) - \lambda_n x_n(t)\, dt\Big| = O(s_n) = o(\lambda_n).$$

Furthermore, a combination of (2.40) and $\log(1+x) = x + O(x^2)$ as $x \to 0$ shows that

$$\begin{aligned}
\Big|\int_{s_n}^1 \log(1 + \lambda_n x_n(t)) - \lambda_n x_n(t)\, dt\Big| &= O\Big(\int_{s_n}^1 (\lambda_n x_n(t))^2\, dt\Big) \\
&= O\Big(\lambda_n^2 \sup_{t\in[s_n,1]} |x_n(t)| \int_0^1 1/h(t)\, dt\Big) \\
&= o(\lambda_n).
\end{aligned}$$

To sum up, we have proven that

$$\Big|\int_0^1 \log(1 + \lambda_n x_n(t)) - \lambda_n x_n(t)\, dt\Big| = o(\lambda_n),$$

which in turn is equivalent to the Hadamard differentiability of $T_{H|\bar{D}_{\beta,h}}$ at $\bar{F}_\beta^{-1}$ with $T'_{H,\beta}(y) = \beta\int_0^1 t^\beta y(t) - y(1)\, dt$, because obviously $|\int_0^1 x_n(t)\, dt - T'_{H,\beta}(y)| = o(1)$. Notice that the signed measure pertaining to $T'_{H,\beta}$ is the one that minimizes the asymptotic variance. □

In a similar way one may check that the moment estimator proposed by Dekkers et al. (1989) and kernel type estimators (Csörgő et al. (1985)) satisfy the Conditions 2.4 and 2.5 and are therefore asymptotically normal. While for the moment estimator this was already proven in the above mentioned paper under comparable second-order conditions, in case of kernel type statistical tail functionals $T(Q_{n,k_n})$, for *all* intermediate sequences $(k_n)_{n\in I\!N}$ there exist constants b_{n,k_n} such that $k_n^{1/2}(T(Q_{n,k_n})-\beta-b_{n,k_n})$ converges to a centered normal r.v. if the q.f. F^{-1} is normalized regularly varying (see (3.6)), i.e., no additional second-order condition is needed for asymptotic normality in this case (Csörgő et al. (1985)). This particular feature of kernel type estimators can be explained by their linearity as a function of the logarithm of the order statistics under consideration, which in combination with condition (3.6) yields the desired asymptotic normality. On the other hand, under (3.6) convergence of $k_n^{1/2}(T_H(Q_{n,k_n})-\beta)$ towards a non-centered normal r.v. for a *regularly varying sequence* $(k_n)_{n\in I\!N}$ implies a second-order condition that is closely related to Condition 2.1(ii) (see Geluk et al. (1996), Theorem 4.3, and Lemma 3.2 below). Furthermore, if F^{-1} is not *normalized* regularly varying, then in general asymptotic normality of the Hill estimator does not hold without additional conditions on k_n; for a detailed discussion we refer to Csörgő and Viharos (1995).

If one considers the AMSE (with optimally chosen k_n) instead of just the asymptotic variance, then from the discussion following Example 2.4 it is immediate that in the situation of Lemma 2.3(i) the AMSE can be minimal only if $\nu_{T,\beta}[0,1]=0$. Thus the signed measure $\nu_{T,\beta}$ has to satisfy the conditions established in Lemma 2.2, and the arguments given in Section 2.2 show that again there is no functional T fulfilling the Conditions 2.5 and 2.4 that minimizes the AMSE uniformly for all d.f.'s $F\in D(G_\beta)$ satisfying Condition 2.1(ii). Moreover, under the Conditions 2.5 and 2.4 all statistical tail functionals have the same degenerate limiting distribution given in Corollary 2.2 if Condition 2.1(ii) holds for some slowly varying function Φ and k_n is chosen in a optimal way, i.e., according to (2.31).

Finally, it is worth mentioning that the construction given in Theorem 2.6 works in the present situation, too. To be more precise, let $I\subset(0,\infty)$ be an open set and let $(\nu_\beta)_{\beta\in I}$ denote a family of signed measures that satisfy (2.21), the conditions established in Lemma 2.4 and the continuity condition (2.34). Define H_β and T_β as in Theorem 2.6 and let $\tilde{T}: D_{\beta,h}\to I\!R$ be a functional that satisfies Condition 2.5 and is continuous at $\bar{F}_\beta^{-1}$ (e.g., $\tilde{T}=T_H$). Then

$$T(z):=\begin{cases} T_{\tilde{T}(z)}(z) & \text{if } \tilde{T}(z)\in I,\\ 0 & \phantom{\text{if }} T(z)\notin I \end{cases}$$

defines a functional T satisfying the Conditions 2.4 and 2.5 such that $\nu_{T,\beta}=\nu_\beta$ for all $\beta\in I$. So again these adaptive generalized probability weighted moment estimators enable us to construct statistical tail functionals with prescribed limiting distribution.

2.4 Non-location-invariant statistical tail functionals: the case $\beta\le 0$

In contrast to the preceding sections, for $\beta\le 0$ in general not the upper Pareto q.f. $\bar{F}_\beta^{-1}$ is the leading term of the approximation for Q_n any longer, but the constant term $D_n/a(k_n/n)$ (see Corollary 2.1 for the definition).

Lemma 2.5 *Let $\omega := F^{-1}(1)$ denote the right endpoint of the underlying d.f. and define*

$$\Xi(t) := \begin{cases} F^{-1}(1-t)/a(t) + \frac{c_\Psi}{\delta}\Phi(t)\cdot 1_{\{\delta>0\}} & \beta = 0, \\ & \text{if} \\ \omega/a(t) & \beta < 0. \end{cases}$$

(i) For suitably chosen functions a, Φ and Ψ in accordance with Lemma 2.1 we have

$$\frac{D_n}{a(k_n/n)} = \begin{cases} \Xi(k_n/n) & \beta \geq -1/2, \\ & \text{if} \\ \Xi(k_n/n) + o_P(k_n^{-1/2}) & \beta < -1/2. \end{cases} \tag{2.41}$$

(ii) If $\beta < 0$ and $\omega \neq 0$ or $\beta = 0$, then $|\Xi(t)| \to \infty$ as $t \downarrow 0$.
If, in addition, Condition 2.1(ii) holds for some $\delta-$varying function Φ, then
$\Phi(t) = o(1/\Xi(t))$ *for* $|\beta| < \delta$, *and*
$1/\Xi(t) = o(\Phi(t))$ *for* $|\beta| > \delta$.

Remark 2.7 Under Condition 2.1(ii) with $|\beta| = \delta$, $\Phi(t)$ can be of smaller, larger or the same order as $1/\Xi(t)$. □

With the exception of the case $\beta < 0$ and $\omega = 0$, in which the limiting distribution of a non-location-invariant statistical tail functional can be established as in Section 2.2, the term $D_n/a(k_n/n)$ tends to ∞ or $-\infty$, depending on the sign of ω. For the sake of definiteness, in the sequel we will assume that $\omega > 0$, but the case $\omega < 0$ can be treated likewise.

Since for $\beta < -1/2$ in general $D_{\beta,h}$ does not include the constant functions, we again consider functionals T that are defined on $span(D_{\beta,h}, 1)$. This function space will be equipped with the σ-field $\mathcal{A}_{\beta,h}$ that is generated by all sets of the form $\{z + c | z \in B, c \in [c_1, c_2]\}$ with $B \in \mathcal{B}(D_{\beta,h})$ and $-\infty < c_1 < c_2 < \infty$.

Condition 2.7 *Assume that $T : span(D_{\beta,h}, 1) \to I\!R$ satisfies for $\beta \leq 0$ and some $h \in \mathcal{H}$*

(i) T is $\mathcal{A}_{\beta,h}, \mathcal{B}(I\!R)$-measurable,

(ii) $T(az) = T(z) \quad \forall z \in span(D_{\beta,h}, 1),\ a > 0.$

Because $\Xi(t) \to \infty$, instead of continuity or Hadamard differentiability of T at $\bar{F}_\beta^{-1}$, here we need smoothness conditions on T that determine its local behavior in a neighborhood of a constant function in order to obtain consistency or asymptotic normality of the pertaining statistical tail functional $T(Q_n)$.

Theorem 2.9 *If $F \in D(G_\beta)$ for some $\beta \leq 0$, $(k_n)_{n\in I\!N}$ is an intermediate sequence, the functional T satisfies Condition 2.7 and for some $h \in \mathcal{H}$*

$$T(1 + \rho_n z_n) \longrightarrow \beta \qquad \text{if} \quad \rho_n \to 0 \ \text{and} \ z_n \to \bar{F}_\beta^{-1} \text{ in } D_{\beta,h},$$

then

$$T(Q_n) \to \beta \qquad \text{in probability.}$$

In order to obtain a non-trivial limiting distribution for the standardized statistical tail functional, we need an approximation of $T(1+\rho_n z_n)-\beta$ as z_n approaches $\bar{F}_\beta^{-1}$ from a direction tangentially to $C_{\beta,h}$.

Condition 2.8 *For all $\rho_n, \lambda_n \to 0$ and $y_n \to y \in C_{\beta,h}$*

$$T(1+\rho_n(\bar{F}_\beta^{-1}+\lambda_n y_n)) = \beta + \rho_n c_{T,\beta} + \lambda_n T'_\beta(y) + o(\rho_n+\lambda_n)$$

where $c_{T,\beta} \in I\!R$ and $T'_\beta : C_{\beta,h} \to I\!R$ denotes some continuous, linear functional.

Note that, despite the notation, T'_β is not a proper derivative. At first glance Condition 2.8 seems rather peculiar, but Corollary 2.1 suggests that nevertheless it is appropriate in the present situation. Furthermore, below we will see that, besides the moment estimator due to Dekkers et al. (1989), there is a quite general class of non-location-invariant statistical tail functionals of that type. It is worth mentioning that there are straightforward generalizations of Condition 2.8 that lead to similar results on the asymptotic behavior of $T(Q_n)$. For example, in the right-hand side of the expansion one may replace ρ_n with $g_\beta(\rho_n)$ where $g_\beta(t)$ is some function converging to 0 as t tends to 0.

In the present situation we need three conditions on the sequence $(k_n)_{n\in I\!N}$, since Φ and $1/\Xi$ may have the same order if $|\beta| = \delta$. To keep these conditions as simple as possible, we restrict ourselves to the case of a non-degenerate limiting distribution of the standardized estimator and leave the rest to the reader, though this excludes the optimal choice for k_n if $\beta = 0$ or $\delta = 0$.

Condition 2.9 *Let $(k_n)_{n\in I\!N}$ be some intermediate sequence satisfying*

(i) $k_n^{1/2} \sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n,x)| \longrightarrow 0,$

(ii) $k_n^{1/2}/\Xi(k_n/n) \longrightarrow \rho \in [0,\infty),$ *or*

(iii) $k_n^{1/2}\Phi(k_n/n) \longrightarrow \lambda \in [0,\infty).$

Using the notation of Section 2.2, we have the following counterpart of the Theorems 2.3 and 2.8.

Theorem 2.10 [Drees (1998b), Theorem 4.1] *Suppose T satisfies the Conditions 2.7 and 2.8.*

(i) If Condition 2.1(i) holds or (ii) with $|\beta| < \delta$, then under Condition 2.9(i) and (ii)

$$\mathcal{L}\Big(k_n^{1/2}(T(Q_n)-\beta)\Big) \longrightarrow \mathcal{N}(\rho c_{T,\beta}, \sigma^2_{T,\beta}) \quad \textit{weakly.}$$

(ii) Under the Conditions 2.1(ii) with $|\beta| > \delta$ and 2.9(iii) one has

$$\mathcal{L}\Big(k_n^{1/2}(T(Q_n)-\beta)\Big) \longrightarrow \mathcal{N}(\lambda\mu_{T,\beta,\Psi}, \sigma^2_{T,\beta}) \quad \textit{weakly.}$$

(iii) Condition 2.1(ii) with $|\beta| = \delta$ in combination with Condition 2.9(ii) and (iii) implies

$$\mathcal{L}\Big(k_n^{1/2}(T(Q_n)-\beta)\Big) \longrightarrow \mathcal{N}(\rho c_{T,\beta} + \lambda\mu_{T,\beta,\Psi}, \sigma^2_{T,\beta}) \quad \textit{weakly.}$$

Example 2.7 Consider the functional T_M pertaining to the moment estimator as defined in Example 2.1(iii), which is obviously scale invariant. Here we give a sketch of a proof of Condition 2.8 under the additional assumption that the supremum of y_n is bounded, which holds always true if $\beta < -1/2$ and $h \in \mathcal{H}$ is chosen appropriately. The general case can be treated in a similar way as the Hill estimator in Example 2.6.

A Taylor expansion of the logarithm yields

$$\begin{aligned} T_H(1+\rho_n(\bar F_\beta^{-1}+\lambda_n y_n)) &= \rho_n\Big(\int_0^1 \bar F_\beta^{-1}(t)-\bar F_\beta^{-1}(1)\,dt+\lambda_n\int_0^1 y(t)-y(1)\,dt \\ &\quad -\frac{\rho_n}{2}\int_0^1(\bar F_\beta^{-1})^2(t)-(\bar F_\beta^{-1})^2(1)\,dt+o(\rho_n+\lambda_n)\Big) \end{aligned}$$

and

$$\begin{aligned} T_2(1+\rho_n(\bar F_\beta^{-1}+\lambda_n y_n)) &= \rho_n^2\Big(\int_0^1(\bar F_\beta^{-1}(t)-\bar F_\beta^{-1}(1))^2\,dt+2\lambda_n\int_0^1(\bar F_\beta^{-1}(t)-\bar F_\beta^{-1}(1))\cdot \\ &\cdot(y(t)-y(1))\,dt-\rho_n\int_0^1(\bar F_\beta^{-1}(t)-\bar F_\beta^{-1}(1))((\bar F_\beta^{-1})^2(t)-(\bar F_\beta^{-1})^2(1))\,dt+o(\rho_n+\lambda_n)\Big). \end{aligned}$$

From this it follows by straightforward but lengthy computations that

$$\begin{aligned} T_M(1+\rho_n(\bar F_\beta^{-1}+\lambda_n y_n)) &= \beta-\rho_n\frac{\beta(\beta+1)}{(1-\beta)(1-3\beta)}+\lambda_n(1-\beta)(1-2\beta) \\ &\cdot\Big((1-\beta)\int_0^1((1-2\beta)\bar F_\beta^{-1}(t)-2)y(t)\,dt+y(1)\Big)+o(\rho_n+\lambda_n). \end{aligned}$$

Hence Condition 2.8 is satisfied with

$$c_{T_M,\beta} = -\frac{\beta(\beta+1)}{(1-\beta)(1-3\beta)}$$

and

$$\nu_{T_M,\beta} = (1-\beta)(1-2\beta)\Big((1-\beta)((1-2\beta)\bar F_\beta^{-1}-2)\lambda_{|(0,1)}+\varepsilon_1\Big).$$

By Theorem 2.10 the asymptotic normality of the moment estimator follows under assumptions comparable to those made by Dekkers et al. (1989).

Note that $\nu_{T_M,0}$ is the signed measure that minimizes the asymptotic variance if $\beta=0$ and location invariance is assumed (see Theorem 2.4). Generally, $\nu_{T_M,\beta}$ satisfies the conditions given in Lemma 2.2 for location-invariant statistical tail functionals. Later on we will see that under a very weak extra condition this always holds true for functionals satisfying Conditions 2.7 and 2.8. $\square$

The following example indicates a general method to construct functionals satisfying the Conditions 2.7 and 2.8.

Example 2.8 Let T_i, $i=1,2$, be scale-invariant measurable functionals satisfying the expansion

$$T_i(1+\rho_n z) = \rho_n T_i'(z)+\rho_n^2\tilde T_i(z)+o(\rho_n^2)$$

uniformly for $\|z\|_{\beta,h} \leq 1$ where T_i' is a continuous and linear and $\tilde{T}_i$ some functional that is continuous at $\bar{F}_\beta^{-1}$. (A typical example of such a functional is $T(z) = \int f(z(t)/z(1))\,\mu(dt)$, where f is some sufficiently smooth function with $f(1) = 0$, and μ is some signed measure on $\mathcal{B}([0,1])$ that does not concentrate too much mass in a neighborhood of 0.)

If one has $T_2'(\bar{F}_\beta^{-1}) \neq 0$ and $T_1'(\bar{F}_\beta^{-1})/T_2'(\bar{F}_\beta^{-1}) = \beta$, then it is easily seen that

$$T(z) := T_1(z)/T_2(z)$$

satisfies Condition 2.8 with

$$c_{T,\beta} = \frac{\tilde{T}_1(\bar{F}_\beta^{-1}) - \beta\tilde{T}_2(\bar{F}_\beta^{-1})}{T_2'(\bar{F}_\beta^{-1})} \qquad \text{and} \qquad T_\beta'(y) = \frac{T_1'(y) - \beta T_2'(y)}{T_2'(\bar{F}_\beta^{-1})}. \qquad \square$$

According to Theorem 2.10, the rate of convergence of estimators $T(Q_n)$ with T satisfying the Conditions 2.7 and 2.8 such that $c_{T,\beta} \neq 0$ is slower than the rate for location-invariant statistical tail functionals if Condition 2.1(ii) holds with $|\beta| < \delta$. Yet, in contrast to the situation examined in Section 2.3, the following lemma shows that asymptotically there is no advantage in using functionals that are not location invariant in the case $|\beta| > \delta$ either.

Lemma 2.6 [Drees (1998b), Lemma 4.1] *If T satisfies the Conditions 2.7 and 2.8, then*

(i) $\int \bar{F}_\beta^{-1}\, d\nu_{T,\beta} = 0$,

(ii) $\nu_{T,\beta}[0,1] = 0$.

(iii) If, in addition,

$$T(1 + \rho_n \bar{F}_{\beta_n}^{-1}) = \beta_n + \rho_n c_{T,\beta_n} + o(\rho_n) \tag{2.42}$$

for all $\rho_n \downarrow 0$ and $\beta_n - \beta = O(\rho_n)$, then one has

$$\int \bar{F}_\beta^{-1} \cdot \log\, d\nu_{T,\beta} = \begin{cases} -1 & \beta \neq 0, \\ -2 & \beta = 0. \end{cases} \quad \textit{if}$$

The additional condition (2.42) is very mild, since, roughly speaking, it means that in case $y_n \equiv 0$ the Condition 2.8 holds locally uniformly for $\beta \leq 0$. Note that for $\beta \geq -1/2$ the conditions given in Lemma 2.6 are exactly the conditions established in Lemma 2.2 for location-invariant functionals, while in the case $\beta < -1/2$ they are even stronger. Hence for each functional satisfying Condition 2.8 there exists a location-invariant statistical tail functional with the same limit distribution if $|\beta| > \delta$ and a faster rate of convergence towards the true parameter if $|\beta| < \delta$, provided that the signed measures pertaining to the functionals T_β' satisfy the continuity condition (2.34). In view of this result, we recommend to use statistical tail functionals that are location invariant if one expects that β may be non-positive.

2.5 Proofs

Proof of Lemma 2.1. Since the simpler case (i) can be treated in a similar way, we restrict ourselves to the case (ii), in which we can use the representation of the q.f. established by de Haan and Stadtmüller (1996), Theorem 2.

First assume $\beta = \delta = 0$. Then, according to Omey and Willekens (1988), Theorem 2.3, $g(t) := F^{-1}(1-t) - t\int_t^{1/2} F^{-1}(1-s)/s^2\,ds$ is Π–varying at 0 with slowly varying auxiliary function $L(t) := g(t) - t\int_t^1 g(s)/s^2\,ds$. More precisely, $g(t) = L(t) + \int_t^1 L(s)/s\,ds$ and the uniform convergence theorem for regularly varying functions (see de Haan (1975), Corollary 1.2.1 4.) imply

$$\begin{aligned}
&\sup_{0<x\le\max(x_0,1)} x^\varepsilon\Big|\frac{g(tx)-g(t)}{L(t)} + \log x\Big| \\
&= \sup_{0<x\le\max(x_0,1)}\Big|\frac{L(tx)(tx)^\varepsilon}{L(t)t^\varepsilon} - x^\varepsilon + x^\varepsilon\int_x^1\Big(\frac{L(ts)(ts)^\varepsilon}{L(t)t^\varepsilon} - s^\varepsilon\Big)s^{-(1+\varepsilon)}\,ds\Big| \\
&= o(1) \qquad (2.43)
\end{aligned}$$

for all $\varepsilon > 0$. Using $F^{-1}(1-t) = g(t) + \int_t^{1/2} g(s)/s\,ds$ we conclude likewise

$$\begin{aligned}
&\sup_{0<x\le x_0} x^\varepsilon\Big|\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{g(t)+L(t)} + \log x - \frac{L(t)}{2(g(t)+L(t))}\log^2 x\Big| \\
&= \frac{|L(t)|}{|g(t)+L(t)|}\sup_{0<x\le x_0} x^\varepsilon\Big|\frac{g(tx)-g(t)}{L(t)} + \log x + \int_x^1\Big(\frac{g(ts)-g(t)}{L(t)} + \log s\Big)/s\,ds\Big| \\
&= o\Big(\frac{|L(t)|}{|g(t)+L(t)|}\Big),
\end{aligned}$$

i.e., the assertion with $a(t) = g(t) + L(t)$ and $\Phi(t) = |L(t)|/|g(t)+L(t)|$ for sufficiently small $t > 0$, and $\Psi(x) = \pm\frac{1}{2}\log^2 x$.

In case of $\beta > 0$ and $\delta = 0$, the q.f. admits the representation $F^{-1}(1-t) = t^{-\beta}g(t)$ where g is Π–varying at 0 with auxiliary function L, say. Thus

$$\begin{aligned}
&\sup_{0<x\le x_0} x^{\beta+\varepsilon}\Big|\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{\beta t^{-\beta}g(t)} - \frac{x^{-\beta}-1}{\beta} + \frac{x^{-\beta}L(t)}{\beta g(t)}\log x\Big| \\
&= \frac{|L(t)|}{\beta|g(t)|}\sup_{0<x\le x_0} x^\varepsilon\Big|\frac{g(tx)-g(t)}{L(t)} + \log x\Big| = o\Big(\frac{|L(t)|}{|g(t)|}\Big)
\end{aligned}$$

by (2.43), i.e., the assertion with $a(t) = \beta t^{-\beta}g(t) = \beta F^{-1}(1-t)$ and $\Phi(t) = |L(t)|/|g(t)|$ in a neighborhood of 0, and $\Psi(x) = \pm\frac{1}{\beta}x^{-\beta}\log x$. Similarly, Lemma 2.1 may be proven in case of $\beta < 0$ and $\delta = 0$, where $F^{-1}(1-t) = \omega - t^{-\beta}g(t)$ for some Π–varying function g, $a(t) = |\beta|t^{-\beta}g(t) = |\beta|(\omega - F^{-1}(1-t))$ and Φ and Ψ as before.

Finally, consider the case $\delta > 0$. Then $F^{-1}(1-t) = c(t^{-\beta}-1)/\beta + g(t)$ where $(g(tx)-g(t))/L(t) \to F^{-1}_{\beta-\delta}(x)$ as $t \downarrow 0$, that is, g is Π–varying at 0 if $\beta = \delta$, and $g + d$ is $(\delta-\beta)$–varying at 0 for some $d \in I\!R$ if $\beta \neq \delta$, so that one can choose $L(t) = (\beta-\delta)(g(t)+d)$.

Hence, (2.43) and the uniform convergence theorem, respectively, lead to

$$\begin{aligned}
&\sup_{0<x\le x_0} x^{\beta+\varepsilon}\Big|\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{ct^{-\beta}}-\frac{x^{-\beta}-1}{\beta}-\frac{L(t)}{ct^{-\beta}}F^{-1}_{\beta-\delta}(x)\Big| \\
&= \frac{|L(t)|}{ct^{-\beta}}\sup_{0<x\le x_0} x^{\beta+\varepsilon}\Big|\frac{g(tx)-g(t)}{L(t)}-F^{-1}_{\beta-\delta}(x)\Big| \;=\; o\Big(\frac{|L(t)|}{ct^{-\beta}}\Big).
\end{aligned}$$

Thus the assertion is proven in this case, too, with $a(t)=ct^{-\beta}$ and $\Phi(t)=|L(t)|/(ct^{-\beta})$ for small $t>0$, and $\Psi=\pm F^{-1}_{\beta-\delta}$. □

The *proof of Theorem 2.1* will be given in several steps. For the time being, we replace the supremum over all $t\in[0,1]$ by the supremum over all $t\in(\eta/k_n,1]$, where $0<\eta<1$ if $\beta<-1/2$, and $\eta=0$ if $\beta\ge-1/2$. Check that for $\gamma\ge\beta+1/2$

$$\sup_{t\in(\eta/k_n,1]}\Big(\frac{k_nt}{[k_nt]+1}\Big)^{\gamma} \;=\; O(1), \tag{2.44}$$

$$\sup_{t\in(\eta/k_n,1]}\Big|\Big(\frac{k_nt}{[k_nt]+1}\Big)^{\gamma}-1\Big|([k_nt]+1) \;=\; O(1). \tag{2.45}$$

An application of Lemma 2.1 yields

$$\begin{aligned}
&\sup_{t\in(\eta/k_n,1]} t^{\beta+1/2}\Big|\frac{\tilde Q_n(t)-F^{-1}(1-k_n/n)}{a(k_n/n)}-\frac{(S_{[k_nt]+1}/k_n)^{-\beta}-1}{\beta}-\Phi(k_n/n)\Psi(S_{[k_nt]+1}/k_n)\Big| \\
&\le \sup_{x\in[S_1/k_n,S_{k_n+1}/k_n]} x^{\beta+1/2}\Big|\frac{F^{-1}(1-xk_n/n)-F^{-1}(1-k_n/n)}{a(k_n/n)}-\frac{x^{-\beta}-1}{\beta}-\Phi(k_n/n)\Psi(x)\Big| \\
&\quad\cdot\sup_{t\in(\eta/k_n,1]}\Big(\frac{k_nt}{S_{[k_nt]+1}}\Big)^{\beta+1/2} \\
&= \begin{cases} O\Big(\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n,x)|\Big), & \text{case (i)},\\ o(\Phi(k_n/n)), & \text{case (ii)}\end{cases} \qquad \text{a.s.}
\end{aligned} \tag{2.46}$$

Besides (2.44) for the last equation it has been utilized that by the strong law of large numbers a.s. there exist $0<x_l<x_u<\infty$ such that $S_i/i\in(x_l,x_u)$ for all $i\in I\!N$.

In a second step, we establish an approximation of $((S_{[k_nt]+1}/k_n)^{-\beta}-1)/\beta$ using the well-known "Hungarian construction" of Komlós et al. (1975, 1976). According to that construction, there exist a standard Brownian motion W and a sequence of standard exponential r.v.'s ξ_i, $i\in I\!N$, such that

$$\max_{1\le i\le k}|S_i-i-W(i)|=O(\log k) \qquad \text{a.s.} \tag{2.47}$$

with S_i defined by (2.5). Furthermore,

$$\sup_{s,t\in[0,T],|t-s|\le1}|W(t)-W(s)|=O(\log^{1/2}T) \qquad \text{a.s.} \tag{2.48}$$

as $T\to\infty$ (Csörgő and Horváth (1993), Theorem A.1.1), and by the reflection principle

$$\sup_{t\in[0,k]}|W(t)|=O\Big((k\log k)^{1/2}\Big) \qquad \text{a.s.} \tag{2.49}$$

Proposition 2.1 *Almost surely for all $h \in \mathcal{H}$*

$$\sup_{t\in(\eta/k,1]} h(t)t^{\beta}\Big|\frac{(S_{[kt]+1}/k)^{-\beta}-1}{\beta}-\frac{t^{-\beta}-1}{\beta}+t^{-\beta}\frac{W(kt)}{kt}\Big| = o(k^{-1/2}).$$

Proof. We restrict ourselves to the case $\beta \neq 0$, since the proof runs along the same lines in the other case. Let $I(i,k) := [(i-1)/k, i/k) \cap (\eta/k, 1]$. Then

$$\begin{aligned}
&\sup_{t\in(\eta/k_n,1]} h(t)t^{\beta}\Big|\frac{(S_{[kt]+1}/k)^{-\beta}-1}{\beta}-\frac{t^{-\beta}-1}{\beta}+t^{-\beta}\frac{W(kt)}{kt}\Big| \\
\le\ & \max_{1\le i\le k}\sup_{t\in I(i,k)} h(t)t^{\beta}\Big|\frac{(S_i/k)^{-\beta}-1}{\beta}-\frac{(i/k)^{-\beta}-1}{\beta}+\Big(\frac{i}{k}\Big)^{-\beta}\frac{W(i)}{i}\Big| \\
&+\max_{1\le i\le k}\sup_{t\in I(i,k)} h(t)t^{\beta}\Big|\Big(\frac{i}{k}\Big)^{-\beta}-t^{-\beta}\Big|/\beta+\max_{1\le i\le k}\sup_{t\in I(i,k)} h(t)t^{\beta}\Big|\Big(\frac{i}{k}\Big)^{-\beta}\frac{W(i)}{i}-t^{-\beta}\frac{W(kt)}{kt}\Big| \\
=:\ & T_1+T_2+T_3.
\end{aligned}$$

A Taylor expansion, (2.47) and (2.49) yield $\max_{[k^{4/5}]\le i\le k}|(S_i/i)^{-\beta}-(1-\beta W(i)/i)| = O(k^{-3/5}\log k)$ a.s. and hence

$$\max_{[k^{4/5}]\le i\le k}\sup_{t\in I(i,k)} h(t)t^{\beta}\Big|\frac{(S_i/k)^{-\beta}-1}{\beta}-\frac{(i/k)^{-\beta}-1}{\beta}+\Big(\frac{i}{k}\Big)^{-\beta}\frac{W(i)}{i}\Big| = o(k^{-1/2}) \quad \text{a.s.} \quad (2.50)$$

In particular, $\sup_{i\in I\!N} i^{1/2}|(S_i/i)^{-\beta}-(1-\beta W(i)/i)| < \infty$ a.s., and thus by (2.44) and (2.9)

$$\begin{aligned}
&\max_{1\le i\le [k^{4/5}]}\sup_{t\in I(i,k)} h(t)t^{\beta}\Big|\frac{(S_i/k)^{-\beta}-1}{\beta}-\frac{(i/k)^{-\beta}-1}{\beta}+\Big(\frac{i}{k}\Big)^{-\beta}\frac{W(i)}{i}\Big| \\
\le\ & \max_{1\le i\le [k^{4/5}]}\sup_{t\in I(i,k)} h(t)t^{-1/2}\Big(\frac{i}{kt}\Big)^{-(\beta+1/2)}k^{-1/2}i^{1/2}\Big|\Big(\frac{S_i}{i}\Big)^{-\beta}-\Big(1-\beta\frac{W(i)}{i}\Big)\Big|/|\beta| \\
=\ & o(k^{-1/2}) \qquad \text{a.s.} \qquad\qquad (2.51)
\end{aligned}$$

Combining (2.50) and (2.51), one obtains $T_1 \to 0$ a.s.

The mean value theorem and some straightforward calculations yield

$$T_2 = O\Big(\big(\sup_{t\in I(1,k)} h(t)t^{-1/2}(kt)^{\min(\beta,0)+1/2}+\max_{2\le i\le k}\sup_{t\in I(i,k)} h(t)t^{-1/2}i^{-1/2}\big)k^{-1/2}\Big) \qquad \text{a.s.}$$

Consequently, for any sequence $i_k \to \infty$, $i_k/k \to 0$

$$T_2 = O\Big(\big(\sup_{t\in(0,i_k/k]} h(t)t^{-1/2}+\sup_{t\in[i_k/k,1]} h(t)t^{-1/2}i_k^{-1/2}\big)k^{-1/2}\Big) = o(k^{-1/2}) \qquad \text{a.s.}$$

Finally we consider T_3. From (2.44), (2.48) and (2.9) it follows

$$\begin{aligned}
&\max_{1\le i\le k}\sup_{t\in I(i,k)} h(t)t^{\beta}(i/k)^{-\beta}i^{-1}|W(i)-W(kt)| \\
=\ & O\Big(\Big(\max_{1\le i\le \log^2 k}+\max_{\log^2 k<i\le k}\Big)\sup_{t\in I(i,k)} h(t)t^{-1/2}i^{-1/2}k^{-1/2}|W(i)-W(kt)|\Big) \\
=\ & o\Big((\log\log k)^{-1/2}k^{-1/2}(\log\log k)^{1/2}\Big)+O\Big((\log^2 k)^{-1/2}k^{-1/2}\log^{1/2} k\Big) \\
=\ & o(k^{-1/2}) \qquad \text{a.s.} \qquad\qquad (2.52)
\end{aligned}$$

On the other hand, using (2.45) for $\gamma = \beta + 1$, the law of iterated logarithm, the strong law of large numbers, and again (2.9), one obtains

$$
\begin{aligned}
&\max_{1\le i\le k} \sup_{t\in I(i,k)} h(t)t^{\beta}|(i/k)^{-\beta}i^{-1} - t^{-(\beta+1)}k^{-1}||W(kt)| \\
&= O\Big(\sup_{t\in(0,1/k]} \frac{h(t)}{\tilde h(t)}\Big(\frac{\log\log(3/(kt))}{\log\log(3/t)}\Big)^{1/2}\frac{|W(kt)|}{(kt\log\log(3/(kt)))^{1/2}}k^{-1/2} \\
&\qquad + \max_{2\le i\le k}\sup_{t\in I(i,k)} h(t)t^{-1/2}\frac{|W(kt)|}{kt}(kt)^{-1/2}k^{-1/2}\Big) \\
&= o(k^{-1/2}) \qquad \text{a.s.,}
\end{aligned}
\tag{2.53}
$$

where the last step follows from an argument similar to the proof of $T_2 \to 0$ a.s. In view of (2.52) and (2.53), $T_3 \to 0$ a.s., and hence the proof of Proposition 2.1 is complete. □

Proposition 2.2 *Under Condition 2.1(ii) with Ψ of type (2.4) one has almost surely for all $h \in \mathcal{H}$*

$$\sup_{t\in(\eta/k,1]} h(t)t^{\beta}\Big|\Psi(S_{[kt]+1}/k) - \Psi(t)\Big| = o(1).$$

Proof. First check that

$$
\begin{aligned}
&\sup_{t\in(\eta/k,1/k]} h(t)t^{\beta}\Big|\Psi(S_1/k) - \Psi(t)\Big| \\
&= \begin{cases} O(\sup_{t\in(\eta/k,1/k]} h(t)t^{-1/2}k^{-1/2}\log k) & \\ O(\sup_{t\in(0,1/k]} h(t)\log^2 t) & \end{cases} \text{if} \quad \begin{matrix}\beta < 0 \\ \beta \ge 0\end{matrix} \\
&= o(1) \qquad \text{a.s.}
\end{aligned}
$$

Furthermore, Ψ is differentiable with a continuous derivative Ψ' satisfying $|\Psi'(x)| \le const\cdot x^{-(\beta+1)}(1+|\log x|)$ for all $0 < x \le 2$. Therefore, the mean value theorem, (2.47) and (2.9) yield almost surely

$$\sup_{t\in[1/k,1]} h(t)t^{\beta}|\Psi(S_{[kt]+1}/k) - \Psi(t)| = O\Big(\sup_{t\in[1/k,1]} h(t)(1+|\log t|)|S_{[kt]+1}/(kt) - 1|\Big) = o(1)$$

□

In view of (2.46), Proposition 2.1 and 2.2, the assertion of Theorem 2.1 is proven in the case $\beta \ge -1/2$. The following proposition will complete the proof for $\beta < -1/2$, too.

Proposition 2.3 *Assume that $\beta < -1/2$ and that the conditions of Theorem 2.1 hold. Then $\Psi(0) := \lim_{t\downarrow 0}\Psi(t) = -c_\Psi/(\beta-\delta)\cdot 1_{\{\delta>0\}}$, and one has almost surely for all $h \in \mathcal{H}$*

$$
\begin{aligned}
&\sup_{t\in(0,1/k_n)} h(t)t^{\beta}\left|\frac{\tilde Q_n(t) - \tilde D_n}{a(k_n/n)} - \Big(F_\beta^{-1}(t) - t^{-(\beta+1)}k_n^{-1}W(k_nt) + \Phi(k_n/n)\Psi(t)\Big)\right| \\
&= o\Big(k_n^{-1/2} + \Phi(k_n/n)1_{\{\Psi\not\equiv 0\}}\Big)
\end{aligned}
\tag{2.54}
$$

and

$$
\begin{aligned}
&\sup_{t\in[1/k_n,1]} h(t)t^\beta \left|\frac{\tilde D_n - F^{-1}(1-k_n/n)}{a(k_n/n)}\right| \\
&= o(k_n^{-1/2}) + \begin{cases} O\Big(\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n,x)|\Big), & \textit{case (i)},\\ o(\Phi(k_n/n)), & \textit{case (ii)}.\end{cases}
\end{aligned}
\tag{2.55}
$$

Proof. The formula for $\Psi(0)$ is obvious from (2.4). Because $\tilde Q_n(t) = \tilde Q_n(1/(2k_n)) = F^{-1}(1-S_1/n)$ for all $t\in(0,1/k_n)$, the left-hand side of (2.54) is equal to

$$
\begin{aligned}
&\sup_{t\in(0,1/k_n)} h(t)t^\beta \left|-t^{-\beta}/\beta + \Phi(k_n/n)(\Psi(0)-\Psi(t)) + t^{-(\beta+1)}k_n^{-1}W(k_nt)\right| \\
&= O\Big(\sup_{t\in(0,1/k_n)} \Big(h(t) + \Phi(k_n/n)h(t)|\log t| 1_{\{\Psi\not\equiv 0\}} \\
&\qquad + k_n^{-1/2}\frac{h(t)}{\tilde h(t)}\Big(\frac{\log\log(3/(k_nt))}{\log\log(3/t)}\Big)^{1/2} \frac{|W(k_nt)|}{(k_nt\log\log(3/(k_nt)))^{1/2}}\Big) \\
&= o\Big(k_n^{-1/2} + \Phi(k_n/n)1_{\{\Psi\not\equiv 0\}}\Big) \qquad \text{a.s.}
\end{aligned}
$$

by the law of iterated logarithm and (2.9).

To prove (2.55), first note that $\bar h : t\mapsto t^{-\beta}\sup_{s\in[t,1]}(h(s)s^\beta)\in\mathcal H$ if $\beta < -1/2$ and $h\in\mathcal H$. Thus by (2.46), Proposition 2.1 and 2.2

$$
\begin{aligned}
&\sup_{t\in[1/k_n,1]} h(t)t^\beta \left|\frac{\tilde D_n - F^{-1}(1-k_n/n)}{a(k_n/n)}\right| \\
&\le \bar h(1/(2k_n))(2k_n)^{-\beta}\left|\frac{\tilde Q_n(1/(2k_n)) - F^{-1}(1-k_n/n)}{a(k_n/n)} + 1/\beta - \Phi(k_n/n)\Psi(0)\right| \\
&= \bar h(1/(2k_n))\Big|1/\beta - 2W(1/2) + \Phi(k_n/n)(2k_n)^{-\beta}(\Psi(1/(2k_n)) - \Psi(0))\Big| \\
&\quad + o(k_n^{-1/2}) + \begin{cases} O\Big(\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n,x)|\Big), & \text{case (i)},\\ o(\Phi(k_n/n)), & \text{case (ii)}\end{cases} \\
&= o(k_n^{-1/2}) + \begin{cases} O\Big(\sup_{x\in(0,1+\varepsilon]} x^{\beta+1/2}|R(k_n/n,x)|\Big), & \text{case (i)},\\ o(\Phi(k_n/n)), & \text{case (ii)},\end{cases}
\end{aligned}
$$

where in the last step

$$
\bar h(1/(2k_n))(2k_n)^{-\beta}(\Psi(1/(2k_n) - \Psi(0)) = O(\bar h(1/(2k_n))\log k_n) = o(1)
$$

has been utilized. □

Proof of Corollary 2.1. Define $\bar D_n$ by (2.12) with $F^{-1}(1-S_1/n)$ replaced by $X_{n:n}$. Then $D_n = \bar D_n + a(k_n/n)(F_\beta^{-1} - \bar F_\beta^{-1} + \Phi(k_n/n)(\Psi - \bar\Psi))$, and hence the assertions follow immediately from (2.6) and Theorem 2.1. □

Proof of Theorem 2.2. According to Theorem 2.1, $(\tilde{Q}_n - \tilde{D}_n)/a(k_n/n)$ converges to $\bar{F}_\beta^{-1}$ a.s. in $D_{\beta,h}$ By Condition 2.3(ii), (iii) and the continuity of $T_{|D_{\beta,h}}$ at $\bar{F}_\beta^{-1}$ it follows that $T(\tilde{Q}_n) \to \beta$ a.s. and thus by (2.6) the assertion. □

Proof of Theorem 2.3. The assertions follow from Corollary 2.1 by the well-known δ-method (see, e.g., Gill (1989), Theorem 3). For example, since the function space $C_{\beta,h}$ is separable, according to the Skorohod-Dudley-Wichura representation theorem (see Gill (1989), Theorem 2), there exist versions $\bar{Q}_n$ of Q_n, versions $\bar{D}_n$ of D_n and $\bar{V}_\beta$ of V_β such that in case (ii) with $\lambda < \infty$

$$y_n := k_n^{1/2}\Big(\frac{\bar{Q}_n - \bar{D}_n}{a(k_n/n)} - \bar{F}_\beta^{-1}\Big) \longrightarrow \bar{V}_\beta + \lambda\bar{\Psi} \qquad \text{a.s.}$$

Therefore, (2.19) with $x = \bar{F}_\beta^{-1}$ and $t_n = k_n^{-1/2}$ implies

$$k_n^{1/2}\Big(T\Big(\frac{\bar{Q}_n - \bar{D}_n}{a(k_n/n)}\Big) - T(\bar{F}_\beta^{-1})\Big) \longrightarrow T'_\beta(\bar{V}_\beta + \lambda\bar{\Psi}) \qquad \text{a.s.}$$

and thus by Condition 2.3

$$k_n^{1/2}(T(Q_n) - \beta) \longrightarrow \int V_\beta + \lambda\bar{\Psi}\, d\nu_{T,\beta} \qquad \text{weakly.}$$

Now the assertion is an immediate consequence of Shorack and Wellner (1986), Proposition 2.2.1. □

Proof of Lemma 2.2. By the scale invariance one has

$$0 = T(\bar{F}_\beta^{-1} + n^{-1}\bar{F}_\beta^{-1}) - T(\bar{F}_\beta^{-1}) = n^{-1}T'_\beta(\bar{F}_\beta^{-1}) + o(n^{-1})$$

and thus

$$0 = T'_\beta(\bar{F}_\beta^{-1}) = \int \bar{F}_\beta^{-1}\, d\nu_{T,\beta},$$

i.e., (i). In a similar way the location invariance of T implies (ii).

To verify (iii), observe that for $\beta \neq 0$

$$n(\bar{F}_{\beta+n^{-1}}^{-1} - \bar{F}_\beta^{-1}) \longrightarrow -\bar{F}_\beta^{-1} \cdot (1/\beta + \log)$$

in $D_{\beta,h}$. By the Conditions 2.3(iii) and 2.4 it follows

$$n^{-1} = T(\bar{F}_{\beta+n^{-1}}^{-1}) - T(\bar{F}_\beta^{-1}) = n^{-1}T'_\beta(-\bar{F}_\beta^{-1} \cdot (1/\beta + \log)) + o(n^{-1})$$

and hence by (i) and the linearity of T'_β

$$-1 = T'_\beta(\bar{F}_\beta^{-1})/\beta + T'_\beta(\bar{F}_\beta^{-1} \cdot \log) = \int \bar{F}_\beta^{-1}(t) \log t\ \nu_{T,\beta}(dt).$$

The proof of (iii) in the case $\beta = 0$ runs along the same lines. □

Proof of Theorem 2.4. We give the proof for the case $\beta \geq -1/2$ and $\beta \neq 0$. The other cases can be treated similarly. For $\nu \in \mathcal{M}(\beta)$ let

$$N_{\nu,\beta}(t) := -\int_{(t,1]} s^{-(\beta+1)}\,\nu(ds). \tag{2.56}$$

Then by (2.21) and integration by parts we obtain

$$\sigma^2_{\beta,\nu} = \int_0^1 N^2_{\nu,\beta}(t)\,dt. \tag{2.57}$$

Furthermore, the Conditions 2.2(i)-(iii) are equivalent to

$$\int_0^1 N_{\nu,\beta}(t)\,dt = 0 \tag{2.58}$$

$$\int_0^1 N_{\nu,\beta}(t)t^\beta\,dt = 0 \tag{2.59}$$

$$\int_0^1 N_{\nu,\beta}(t)\log t\,dt = \beta \tag{2.60}$$

In case of $\beta > -1/2$ one can easily determine the function $N^*_\beta \in L_2[0,1]$ with minimal L_2-norm (2.57) subject to the constraints (2.58)–(2.60) by standard methods of Hilbert space theory. By (2.56) it follows that ν^*_β defined by (2.25) minimizes the asymptotic variance $\sigma^2_{\beta,\nu}$. Direct calculations yield (2.24) in the case $\beta > -1/2$ and $\beta \neq 0$.

If $\beta = -1/2$ (and thus $(t \mapsto t^\beta) \notin L_2[0,1]$), then again by standard methods it can be verified that the infimum of (2.57) subject to the constraints (2.58) and (2.60) equals 1/4 and hence

$$\inf_{\nu\in\mathcal{M}(-1/2)} \sigma^2_{-1/2,\nu} \geq 1/4. \tag{2.61}$$

On the other hand, for all $\varepsilon > 0$ the infimum of the L_2-norm (2.57) on $\mathcal{F}_\varepsilon := \{g \in L_2[0,1]\,|\,g_{|[0,\varepsilon]} \equiv 0,\ \int_0^1 g(t)dt = \int_0^1 g(t)t^{-1/2}dt = 0,\ \int_0^1 g(t)\log t\,dt = -1/2\}$ is attained. Straightforward calculations show that $\lim_{\varepsilon\downarrow 0}\inf_{g\in\mathcal{F}_\varepsilon}\int_0^1 g^2(t)\,dt = 1/4$ and thus $\inf_{\nu\in\mathcal{M}(-1/2)}\sigma^2_{-1/2,\nu} \leq 1/4$. Combining this with (2.61), one arrives at (2.24).

Finally, suppose that $\inf_{\nu\in\mathcal{M}(-1/2)}\sigma^2_{-1/2,\nu}$ is attained at some $\nu(-1/2)$. Then the function $N_{\nu(-1/2),-1/2}$ has minimal L_2-norm subject to the constraints (2.58)–(2.60). If $g \in L_2[0,1]$ satisfies

$$\int_0^1 g(t)\,dt = \int_0^1 g(t)\log t\,dt = 0 \tag{2.62}$$

and $\int_0^1 |g(t)|t^{-1/2}\,dt < \infty$, then for all $\varepsilon > 0$ there exists a function $g_\varepsilon \in L_2[0,1]$ satisfying (2.62) such that $\int_0^1 g_\varepsilon(t)t^{-1/2}\,dt = 0$ and $\int_0^1 (g(t) - g_\varepsilon(t))^2\,dt < \varepsilon^2$. Hence

$$\begin{aligned}
0 &\leq \int_0^1 \Big(N_{\nu(-1/2),-1/2}(t) \pm \varepsilon g_\varepsilon(t)\Big)^2 dt - \int_0^1 N^2_{\nu(-1/2),-1/2}(t)\,dt \\
&= \pm 2\varepsilon\int_0^1 N_{\nu(-1/2),-1/2}(t)g(t)dt + \varepsilon^2\Big(\int g^2_\varepsilon(t)\,dt \pm 2\int N_{\nu(-1/2),-1/2}(t)(g_\varepsilon(t) - g(t))dt/\varepsilon\Big) \\
&= \pm 2\varepsilon\int_0^1 N_{\nu(-1/2),-1/2}(t)g(t)\,dt + O(\varepsilon^2).
\end{aligned}$$

Consequently, $N_{\nu(-1/2),-1/2}(t) = a + b \log t$ for some constants $a, b \in I\!R$ and $\lambda_{|[0,1]}$-a.e. t, which is a contradiction to (2.58) and (2.59). $\Box$

Proof of Corollary 2.2. Under the assumptions of Theorem 2.3(ii) the asymptotic mean squared error of $T(Q_n)$ is of order $O(\max(k_n^{-1}, \Phi(k_n/n)^2))$. Since for $k_n = k_n^{(1)}$ the asymptotic variance and squared bias are of the same order, under condition (2.31) the asymptotic variance is of smaller order and hence the AMSE bahaves asymptotically as $\mu_{T,\beta,\Psi}\Phi(k_n^{(1)}/n)$. If $k_n = o(k_n^{(1)})$ then the asymptotic variance and thus the AMSE is of larger order, and if k_n is of the same order as $k_n^{(1)}$ then an additional variance term enters the AMSE. Finally, if $k_n^{(1)} = o(k_n)$ but the second condition of (2.31) is not satisfied, then the asymptotic bias is larger. Hence the condition (2.31) characterizes the optimal choice for k_n. Now (2.32) follows immediately from Theorem 2.3(ii) and (2.29).

If $\liminf_{n\to\infty} k_n/k_n^{(1)} = \tau < \infty$, then $\Phi(k_n/n)k_n^{1/2} \to \tau^{1/2}$ along a subsequence and hence

$$\liminf_{n\to\infty} P\{|T(Q_{n,k_n}) - \beta| \le c(k_n^{(1)})^{-1/2}\} \le \mathcal{N}(\tau^{1/2}\mu_{T,\beta,\Psi}, \sigma^2_{T,\beta})\,[-c\tau^{1/2}, c\tau^{1/2}] < 1$$

for all $c < \infty$. Secondly, if $k_n/k_n^{(1)} \to \infty$ and $\limsup_{n\to\infty} \Phi(k_n/n)/\Phi(k_n^{(1)}/n) = \tau \ge 1$, then

$$\begin{aligned}
&\liminf_{n\to\infty} P\{|T(Q_{n,k_n}) - \beta| \le c(k_n^{(1)})^{-1/2}\} \\
&= \liminf_{n\to\infty} P\{\Phi(k_n/n)^{-1}|T(Q_{n,k_n}) - \beta| \le c\Phi(k_n^{(1)}/n)/\Phi(k_n/n)\} \\
&= \left\{ \begin{array}{ll} 1 & \\ & \text{if} \\ 0 & \end{array} \begin{array}{l} c/\tau > |\mu_{T,\beta,\Psi}|, \\ \\ c/\tau < |\mu_{T,\beta,\Psi}|, \end{array} \right.
\end{aligned}$$

so that the choice $\tau = 1$ is optimal and the assertion is proven. $\Box$

Proof of Theorem 2.6. The location invariance in the case $\tilde{T}(z) < -1/2$ and the scale invariance of T are obvious. If $\tilde{\beta} := \tilde{T}(z) \ge -1/2$, then $H_{\tilde{\beta}}(1) = 0 = \int_0^1 H_{\tilde{\beta}}(t)/t\,dt$ by Lemma 2.2(i),(ii) and integration by parts. Hence T is location invariant in this case, too.

Next check that, in view of

$$dH_\beta(t) = \nu_\beta(dt) + \beta H_\beta(t)/t\,dt \tag{2.63}$$

and

$$\int_0^1 \bar{F}_\beta^{-1}(t) H_\beta(t)/t\,dt = 1, \tag{2.64}$$

it follows from Lemma 2.2(i) that $T(\bar{F}_\beta^{-1}) = T_\beta(\bar{F}_\beta^{-1}) = \beta$.

Now note that the continuity property (2.34) for (ν_β) implies the analogous property for the measures $(H_\beta(t)/t\,dt)$ and thus for (dH_β), too. To see this, observe that by (2.34) for all $\beta \in I \setminus \{-1/2\}$ there exists an $\varepsilon > 0$ such that $\int_{[0,1]} t^{-(\beta+\varepsilon)}/h(t)\,|\nu_\beta|(dt) < \infty$. Therefore, it is easily seen that the first two terms of the right-hand side of the inequality

$$\begin{aligned}
\int_0^1 \frac{1}{h(t)t^\beta} \cdot \frac{|H_{\beta_n} - H_\beta|}{t}\,dt \;\; &\le \;\; \int_0^1 \frac{1}{h(t)t} |t^{\beta_n-\beta} - 1| \cdot \Big| \int_{[0,t]} s^{-\beta}\,\nu_\beta(ds) \Big|\,dt \\
&+ \int_0^1 \frac{t^{\beta_n-\beta-1}}{h(t)} \int_{[0,t]} |s^{-\beta_n} - s^{-\beta}|\,|\nu_\beta|(ds)\,dt + \int_0^1 \frac{t^{\beta_n-\beta-1}}{h(t)} \int_{[0,t]} s^{-\beta_n}\,|\nu_\beta - \nu_{\beta_n}|(ds)\,dt
\end{aligned}$$

vanish as $\beta_n \to \beta$. The convergence of the last term to 0 can be proven using integration by parts, Karamata's theorem and (2.34).

To prove the Hadamard differentiability, consider sequences $\lambda_n \to 0$ and $y_n \to y \in C_{\beta,h}$. One has $\int_0^1 \bar{F}_\beta^{-1}(t) H_{\tilde{\beta}_n}(t)/t\, dt > 0$ for sufficiently large n, because $\tilde{\beta}_n := \tilde{T}(\bar{F}_\beta^{-1} + \lambda_n y_n) \to \beta$. Thus again using integration by parts one obtains

$$T_{\tilde{\beta}_n}(\bar{F}_\beta^{-1}) = -\frac{\int_0^1 (\bar{F}_\beta^{-1})'(t) H_{\tilde{\beta}_n}(t)\, dt}{\int_0^1 \bar{F}_\beta^{-1}(t) H_{\tilde{\beta}_n}(t)/t\, dt} = \beta,$$

where in the case $\beta = 0$ integration by parts and Lemma 2.2 (i) and (ii) have been utilized. Again by the continuity property of (dH_β) and $(H_\beta(t)/t\, dt)$, Lemma 2.2(i) and (2.64) it follows that for $\beta > -1/2$ and sufficiently large n

$$\begin{aligned}
&T(\bar{F}_\beta^{-1} + \lambda_n y_n) - \beta \\
&= \frac{\int_0^1 (\bar{F}_\beta^{-1} + \lambda_n y_n)(t)\, dH_{\tilde{\beta}_n}(t)}{\int_0^1 (\bar{F}_\beta^{-1} + \lambda_n y_n)(t) H_{\tilde{\beta}_n}(t)/t\, dt} - \frac{\int_0^1 \bar{F}_\beta^{-1}(t)\, dH_{\tilde{\beta}_n}(t)}{\int_0^1 \bar{F}_\beta^{-1}(t) H_{\tilde{\beta}_n}(t)/t\, dt} \\
&= \lambda_n \frac{\int_0^1 y(t)\, dH_\beta(t) \cdot \int_0^1 \bar{F}_\beta^{-1}(t) H_\beta(t)/t\, dt - \int_0^1 y(t) H_\beta(t)/t\, dt \cdot \int_0^1 \bar{F}_\beta^{-1}(t)\, dH_\beta(t) + o(1)}{\left(\int_0^1 \bar{F}_\beta^{-1}(t) H_\beta(t)/t\, dt\right)^2 + o(1)} \\
&= \lambda_n \int_0^1 y(t)\, \nu_\beta(dt) + o(\lambda_n),
\end{aligned}$$

i.e., the asserted Hadamard differentiability.

For $\beta < -1/2$ the $1/2-$variation of h at 0 yields $z(0+) = 0$ for all $z \in D_{\beta,h}$. Consequently, the proof works in this case, too. □

Proof of Lemma 2.3. The first assertion is obvious from $a(t) \sim \beta F^{-1}(1-t)$ as $t \downarrow 0$ (de Haan and Stadtmüller (1996), Proposition 1(ii)).

Next check that the assertions (i) and (ii) do not depend on the particular choice of a, Φ and Ψ. For if $\tilde{a}$, $\tilde{\Phi}$ and $\tilde{\Psi}$ are other functions in accordance with Lemma 2.1, then in view of the discussion following (2.4) it is easily seen that the w.l.o.g. $\tilde{\Phi}(t) \sim \Phi(t)$ and $\tilde{\Xi}(t) = \Xi(t) + o(\Phi(t))$.

If Condition 2.1(ii) for $\delta > 0$ and hence (2.37) hold, then straightforward calculations show that Lemma 2.1 is satisfied for $a(t) = \beta d_1 t^{-\beta}$ and $c_\Psi \Phi(t) = L(t)/a(t)$ where for $\beta \neq \delta$ we define $L(t) := (\beta - \delta) g(t)$ (cf. proof of Lemma 2.1). Hence one has

$$\Xi(t) = (d_2 + g(t) \cdot 1_{\{\beta = \delta > 0\}})/a(t),$$

and the assertions follow immediate from the fact that $L(t) = o(d_2 + g(t))$ if $\beta = \delta$ (Geluk and de Haan (1987), Remark 1 on page 25).

In the case $\delta = 0$ the q.f. is of the type $F^{-1}(1-t) = t^{-\beta} g(t)$ where g is Π–varying at 0 with some auxiliary function L (de Haan and Stadtmüller (1996), Theorem 2). Thus one can take $a(t) = \beta t^{-\beta} g(t)$ and $c_\Psi \Phi(t) = L(t)/g(t)$, which implies $\Xi \equiv 0$ (cf. proof of Lemma 2.1). □

Proof of Lemma 2.5. We have to distinguish several cases. First assume $\beta = 0$. Then (2.41) holds by definition and $|\Xi(t)| \to \infty$ follows from Geluk and de Haan (1987), Remark 1 on page 25. If, in addition, Condition 2.1(ii) is satisfied with $\delta > 0$, then according to de Haan and Stadtmüller (1996), Theorem 2, $\Xi(t) \sim -\log t$ and hence $\Phi(t) = o(1/\Xi(t))$.

Next we turn to the case $-1/2 \leq \beta < 0$. If merely Condition 2.1(i) is assumed, then one may choose $a(t) = |\beta|(\omega - F^{-1}(1-t))$ and (2.41) is obvious. Under Condition 2.1(ii) with $\delta = 0$ there exists a function $g \in \Pi_L^0$ (i.e., $(g(tx) - g(t))/L(t) \to -\log x$) such that $\omega - F^{-1}(1-t) = t^{-\beta} g(t)$ (de Haan and Stadtmüller (1996), Theorem 2). Then $a(t) = |\beta| t^{-\beta} g(t)$ is an admissible choice in the sense of Lemma 2.1, which leads to $D_n/a(k_n/n) = \Xi(k_n/n)$. Likewise, for $\delta > 0$ one has $F^{-1}(1-t) = \omega + dt^{-\beta}/\beta + g(t)$ for some $d > 0$ and $(\delta - \beta)-$varying function g, so that $a(t) = dt^{-\beta}$ and $c_\Psi \Phi(t) = g(t) t^\beta (\beta - \delta)/d$ yield (2.41) (cf. proof of Lemma 2.1).

For the case $\beta < -1/2$ w.l.o.g. we may assume $X_{n:n} = F^{-1}(1 - U_{1:n})$ where U_i are i.i.d. uniformly on $(0,1)$ distributed r.v.'s. Thus the $(-\beta)-$variation of the function $a(t) \sim |\beta|(\omega - F^{-1}(1-t))$ (de Haan and Stadtmüller (1996), Proposition 1(ii) and Remark 2(i)) in combination with the Potter bounds (Bingham et al. (1987), Theorem 1.5.6) implies that for all $0 < \varepsilon < -1/2 - \beta$

$$\frac{\omega}{a(k_n/n)} - \frac{D_n}{a(k_n/n)} \sim \frac{\omega - F^{-1}(1 - U_{1:n})}{|\beta|(\omega - F^{-1}(1 - k_n/n))} = O\Big(\Big(\frac{\max(1, nU_{1:n})}{k_n}\Big)^{-\beta-\varepsilon}\Big) = o_P(k_n^{-1/2}),$$

where for the last equation $nU_{1:n} = O_P(1)$ has been used.

Finally, observe that for an arbitrary $\beta < 0$ the assertion (ii) follows immediately from the $(-\beta)-$variation of a if $\omega \neq 0$. □

Proof of Theorem 2.9. By Theorem 2.1 one has $y_n := (\tilde{Q}_n - \tilde{D}_n)/a(k_n/n) - \bar{F}_\beta^{-1} \to 0$ a.s. in $D_{\beta,h}$. Hence

$$T(\tilde{Q}_n) = T(\tilde{Q}_n/\tilde{D}_n) = T(1 + a(k_n/n)\tilde{D}_n^{-1}(\bar{F}_\beta^{-1} + y_n)) \to \beta \qquad \text{a.s.},$$

for it is easily shown along the lines of the proof of Lemma 2.5 that $\tilde{D}_n/a(k_n/n) \to \infty$ a.s. Now the assertion follows immediately from (2.6). □

Proof of Theorem 2.10. According to the Skorohod-Dudley-Wichura representation theorem, under the conditions of Theorem 2.10 there are versions of Q_n, D_n and V_β such that almost surely

$$\frac{Q_n}{D_n} = 1 + \frac{a(k_n/n)}{D_n}\Big(\bar{F}_\beta^{-1} + k_n^{-1/2}(\lambda\bar{\Psi} + V_\beta + o(1))\Big)$$

in $D_{\beta,h}$, with $\lambda = 0$ if $|\beta| < \delta$ or merely Condition 2.1(i) is assumed. Consequently, by the Conditions 2.7 and 2.8

$$T(Q_n) = \beta + \frac{a(k_n/n)}{D_n} c_{T,\beta} + k_n^{-1/2} T'_\beta(\lambda\bar{\Psi} + V_\beta) + o\Big(\frac{a(k_n/n)}{D_n} + k_n^{-1/2}\Big) \quad \text{a.s.},$$

which in combination with Lemma 2.5 implies the assertion. □

Proof of Lemma 2.6. By the Conditions 2.7 and 2.8 one has

$$\begin{aligned}\beta + \rho_n c_{T,\beta} + o(\rho_n) &= T(1 + \rho_n \bar{F}_\beta^{-1}) \\ &= T(1 + \rho_n(1+\rho_n^2)(\bar{F}_\beta^{-1} + \rho_n/(1+\rho_n^2))) \\ &= \beta + \rho_n c_{T,\beta} + \rho_n T'_\beta(1) + o(\rho_n)\end{aligned}$$

and hence $T'_\beta(1) = 0$, i.e., assertion (ii). Likewise, Condition 2.8 applied to $\lambda_n y_n = \rho_n \bar{F}_\beta^{-1}$ yields (i).

Now assume that (2.42) holds and $\beta < 0$. Then $(\bar{F}_{\beta_n}^{-1} - \bar{F}_\beta^{-1})/(\beta_n - \beta) \to -\bar{F}_\beta^{-1} \cdot (\log + 1/\beta)$ in $D_{\beta,h}$, and hence (2.42) and an application of Condition 2.8 to $\lambda_n y_n = \bar{F}_{\beta_n}^{-1} - \bar{F}_\beta^{-1}$ give

$$(\beta_n - \beta)(1 + T'_\beta(\bar{F}_\beta^{-1} \cdot (\log + 1/\beta))) + \rho_n(c_{T,\beta_n} - c_{T,\beta}) = o(\rho_n)$$

for all $\beta_n - \beta = O(\rho_n)$. So the choice $\beta_n - \beta = o(\rho_n)$ shows that the map $\beta \mapsto c_{T,\beta}$ is continuous and thus

$$(\beta_n - \beta)(1 + T'_\beta(\bar{F}_\beta^{-1} \cdot (\log + 1/\beta))) = o(\rho_n)$$

for all $\beta_n - \beta = O(\rho_n)$, which in view of (i) implies (iii).

Likewise, assertion (iii) can be proven in the case $\beta = 0$ by utilizing $2\beta_n^{-1}(\bar{F}_{\beta_n}^{-1} - 1/\beta_n - \bar{F}_0^{-1}) \to -\bar{F}_0^{-1} \cdot \log$ as $\beta_n \uparrow 0$, which yields for $\beta_n = O(\rho_n)$

$$\begin{aligned}&\beta_n + \frac{\rho_n}{1 - \rho_n/\beta_n} c_{T,\beta_n} + o(\rho_n) \;=\; T\Big(1 + \frac{\rho_n}{1-\rho_n/\beta_n} \bar{F}_{\beta_n}^{-1}\Big) \;=\; T\Big(1 + \rho_n(\bar{F}_{\beta_n}^{-1} - 1/\beta_n)\Big) \\ &= \; T\Big(1 + \rho_n(\bar{F}_0^{-1} - \frac{\beta_n}{2}\bar{F}_0^{-1} \cdot \log + o(\beta_n))\Big) \;=\; \rho_n c_{T,0} - \frac{\beta_n}{2} T'_0(\bar{F}_0^{-1} \cdot \log) + o(\rho_n).\end{aligned}$$

Hence letting $\beta_n = o(\rho_n)$ gives $c_{T,0} = 0$, so that with $\rho_n = -\lambda\beta_n$ one has for all $\lambda > 0$

$$\beta_n\Big(1 - \frac{\lambda}{1+\lambda} c_{T,\beta_n} + \frac{1}{2} T'_0(\bar{F}_0^{-1} \cdot \log)\Big) = o(\beta_n),$$

that is, $c_{T,\beta_n} \to 0$ and $T'_0(\bar{F}_0^{-1} \cdot \log) = -2$. □

3 Selecting the optimal sample fraction – a sequential approach

In Chapter 2 we discussed how to choose the functional T such that the pertaining statistical tail functional $T(Q_{n,k_n})$ is optimal in some sense if the number $k_n + 1$ of order statistics used for estimation is chosen suitably. However, often the selection of k_n for a given functional, which is examined in the present chapter, is more crucial than the particular choice of T. In view of the Theorems 2.3, 2.8 and 2.10, it is obvious that a small number k_n will lead to a large asymptotic variance, while a large value for k_n will usually yield a large asymptotic bias. So apparently for an optimal choice of k_n one has to balance these contrary effects.

In order to be more specific, first one has to define "optimality" in this context. Here we take up the often used criterion that the asymptotic mean squared error (AMSE) of the estimator under consideration, which is informally motivated as a measure of concentration of the limiting distribution, shall be minimized.

The literature on methods of a data-driven choice of k_n is most comprehensive for the Hill estimator

$$\hat{\beta}_{n,k} := \hat{\beta}^H_{n,k} := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} \tag{3.1}$$

if the right-hand side is well defined, and $\hat{\beta}_{n,k} := 0$ else. (For notational covenience, the superscript H is omitted in the rest of this section.) Yet even in this case there are only few results that actually prove asymptotic optimality of such a procedure.

The probably most widely used approach is purely heuristic: One plots the graph, the so-called Hill plot, that maps the number k to the Hill estimator $\hat{\beta}_{n,k}$ and chooses k in a region where this graph is "stable" in some sense (Falk et al. (1994), Section 6.1, and Dekkers and de Haan (1989); see also Resnick and Stărică (1997) for a modified version). Since this method is completely subjective, it is impossible to determine its performance mathematically. Furthermore, while the procedure may yield reasonable estimates if the tail of the underlying distribution is close to a Pareto tail, in other cases it fails completely, because there is no obvious region of stability. For example, two "typical" Hill plots for samples of size 500 drawn from Student's t–distribution with 4 degrees of freedom, which belongs to the domain of attraction of $G_{1/4}$, are shown in Figure 1 below for $1 \le k \le 100$. Following a region, where the Hill plots vary rapidly, up to minor fluctuations the graphs are almost increasing for $k \ge 20$ with no stable part in between these two regions. In fact, simulations show that the mean squared error of $\hat{\beta}^H_{n,k}$ is minimal for about $k = 11$, a choice that can hardly be motivated by Hill plots.

Further ways of selecting the sample fraction used by the Hill estimator were suggested by Hill (1975) and Beirlant et al. (1996a,b), among others, yet despite the mathematical motivation for these methods, no optimality results were given. In fact, Hall and Welsh (1985) demonstrated that in general Hill's procedure yields too large a number k_n. Moreover, they showed that the AMSE of the Hill estimator $\hat{\beta}_{n,k_n}$ is minimized for

$$k_n \sim \Big(\frac{C^{2\rho}(\rho+1)^2}{2D^2\rho^3}\Big)^{1/(2\rho+1)} n^{2\rho/(2\rho+1)} \tag{3.2}$$

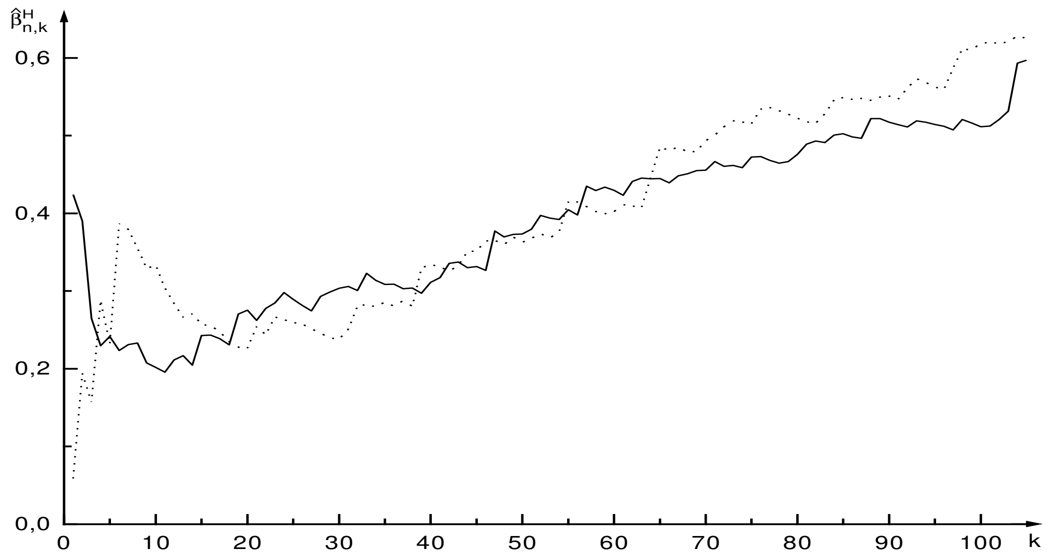

Figure 1: Hill plots for two t_4–samples ($n = 500$)

if the underlying d.f. belongs to the so-called Hall model, i.e.,

$$1 - F(x) = Cx^{-1/\beta}(1 + Dx^{-\rho/\beta} + o(x^{-\rho/\beta})) \tag{3.3}$$

with unknown parameters $\rho > 0$, $C > 0$ and $D \neq 0$. Then they took advantage of the particular structure of the bias of $\hat\beta_{n,i}$ as a function of i to construct an estimator of k_n^{opt} if some prior information about ρ is known. More precisely, it was proven that, with

$$\hat\rho := \Big|\log\Big|\frac{\hat\beta^{-1}_{n,t_1} - \hat\beta^{-1}_{n,s}}{\hat\beta^{-1}_{n,t_2} - \hat\beta^{-1}_{n,s}}\Big| \Big/ \log\frac{t_1}{t_2}\Big| \qquad \text{and} \qquad \hat\lambda_0 := \Big|(2\hat\rho)^{-1/2}\Big(\frac{n}{t_1}\Big)^{\hat\rho}\frac{\hat\beta^{-1}_{n,t_1} - \hat\beta^{-1}_{n,s}}{\hat\beta_{n,s}}\Big|^{2/(2\hat\rho+1)},$$

$$\hat k_n := [\hat\lambda_0 n^{2\hat\rho/(2\hat\rho+1)}]$$

is a consistent estimator of k_n^{opt} in the sense that $\hat k_n/k_n^{opt} \to 1$ in probability if $t_i = [n^{\tau_i}]$, $i = 1, 2$, and $s = [n^\sigma]$ with $0 < 2\rho(1-\tau_1) < \sigma < 2\rho/(2\rho+1) < \tau_1 < \tau_2 < 1$; here again $[x]$ denotes the largest integer smaller than or equal to x. Hence, in this case, $\hat\beta_{n,\hat k_n}$ is asymptotically as efficient as $\hat\beta_{n,k_n^{opt}}$ (Hall and Welsh (1985), Theorem 4.1). Note that the number of order statistics used by $\hat\rho$ must not be too small. In fact, this holds true for all estimators of ρ that are based on a fixed number of order statistics independent of the sample. To see this, consider a d.f. F with density f satisfying

$$f(x) = \frac{C}{\beta}x^{-(1/\beta+1)}\Big(1 + D(\rho+1)x^{-\rho/\beta} + o(x^{-\rho/\beta})\Big),$$

which is the counterpart of (3.3) for the density. Then, according to Reiss (1989), Corollary 5.5.5, asymptotically one cannot distinguish between the distribution of the $k_n = o(n^{2\rho/(2\rho+1)})$ largest order statistics with parent d.f.'s F and $W_{C,\beta}(x) = 1 - Cx^{-1/\beta}$, respectively. Hence it is impossible to estimate ρ consistently by using the k_n largest order statistics only.

Another natural approach is to select the optimal sample fraction by a resampling procedure, yet the usual bootstrap method fails, as it underestimates the bias of the Hill estimator seriously. To overcome this problem, Hall (1990) proposed to use resamples whose size is of a smaller order than the original sample size, but this procedure requires that the parameter ρ is exactly known and the sample size n is very large. Recently, Danielsson et al. (1996) used a combination of subsample bootstrap estimates for the difference of two estimators of β^2 based on bootstrap sample sizes of different order to obtain a consistent estimator of the optimal number of order statistics that does not need any restriction on ρ. Similar methods were developed for the moment and the Pickands estimator by de Haan at al. (1997), where for each case one has to find two estimators of the extreme value index whose asymptotic behavior is related to the behavior of the estimator under consideration.

As an alternative we propose a sequential approach to construct a consistent estimator of k_n^{opt} that works without any prior knowledge about the underlying d.f. of type (3.3). Theorem 3.1 of Hall and Welsh (1985) and the Theorem of Mason and Turova (1994) in combination with a law of iterated logarithms (w.r.t. convergence in probability) suggest that the maximum random fluctuation of $i^{1/2}(\hat\beta_{n,i} - \beta)$, $1 \le i \le k_n$, is of the order $(\log\log n)^{1/2}$, i.e.,

$$\max_{1\le i\le k_n} i^{1/2}|\hat\beta_{n,i} - \beta - b_{n,i}| = O_P((\log\log n)^{1/2}) \tag{3.4}$$

for all intermediate sequences $(k_n)_{n\in I\!N}$, where $b_{n,i} \in I\!R$, $1 \le i \le k_n$, denote some bias terms of the Hill estimators. Now define a "stopping time" for the sequence of Hill estimators by

$$\bar k_n(r_n) := \min\Big\{k \in \{1,\ldots,n\} \mid \max_{1\le i\le k} i^{1/2}|\hat\beta_{n,i} - \hat\beta_{n,k}| > r_n\Big\}, \tag{3.5}$$

where the thresholds r_n constitute a sequence that is of larger order than $(\log\log n)^{1/2}$, but of smaller order than $n^{1/2}$. This definition is inspired by a somewhat related statistic that is used to choose an optimal local bandwidth in nonparametric curve estimation (see Lepskii (1991) and Lepskii et al. (1997)). In view of (3.4), $(\bar k_n(r_n))_{n\in I\!N}$ is asymptotically equivalent to the deterministic sequence $(\tilde k_n(r_n))_{n\in I\!N}$ defined by (3.5) with $\hat\beta_{n,i}$ and $\hat\beta_{n,k}$ replaced by the corresponding bias terms $b_{n,i}$ and $b_{n,k}$, respectively. As, loosely speaking, $b_{n,i} \sim \text{const.}(i/n)^\rho$, this fact can be used to construct consistent estimators of ρ and k_n^{opt} based on $\bar k_n(r_n)$ and $\bar k_n(r_n^\xi)$ with $\xi \in (0,1)$.

Before the details of this programme are worked out, first we show in Section 3.1 that the second moment of the standardized error of a slightly modified version of Hill's estimator behaves asymptotically as the AMSE, thus giving a rigorous decision theoretical justification for our optimality criterion, which previously was only informally motivated. Then, in order to avoid the distinction of cases made in Theorem 2.8, instead of Condition 2.1(ii) we consider a different condition on the underlying q.f., which was used earlier by Csörgő et al. (1985); in particular, the relationship to Condition 2.1(ii) is clarified.

After this preparation, we introduce our "sequential" estimator, which is weakly consistent for the optimal number k_n^{opt} in the Hall model and, if the sequence of thresholds r_n

does not grow too fast, under more general second-order conditions as well. Finally, we indicate how this approach can be used to construct adaptive Hill estimators with minimal AMSE if there is no asymptotically unique optimal sequence $(k_n^{opt})_{n\in I\!N}$ (cf. Corollary 2.2 and Remark 2.5(ii)). For a simulation study in which the performance of the adaptive Hill estimator for moderate sample sizes is investigated we refer to Drees and Kaufmann (1998).

In Section 3.2 we illustrate by means of the example of scale- and location-invariant statistical tail functionals that, with essentially no changes, our approach also works for quite general estimators of the extreme value index that are merely based on upper order statistics. Again all proofs are collected in the last section of the chapter.

3.1 An optimal adaptive Hill estimator

In Section 2.3 we have seen that the Hill estimator is asymptotically normal if $F \in D(G_\beta)$ and Condition 2.2(i) is satisfied. If one wants to determine an intermediate sequence $(k_n)_{n\in I\!N}$ such that the pertaining sequence of Hill estimators $(\hat\beta_{n,k_n})_{n\in I\!N}$ has minimal AMSE, then one needs a second-order condition like, e.g., Condition 2.1(ii). In order to avoid the distinction of the two cases that was necessary in Theorem 2.8, following Csörgő et al. (1985) instead we will assume that the q.f. F^{-1} is normalized regularly varying, that is,

$$F^{-1}(1-t) = ct^{-\beta}\exp\Big(\int_t^1 \frac{\varepsilon(s)}{s}\,ds\Big), \qquad 0 < t \le t_0, \tag{3.6}$$

for some $t_0 \in (0,1)$, a constant $c > 0$ and a bounded measurable function ε satisfying $\varepsilon(t) \to 0$ as $t \downarrow 0$. Note that (3.6) is equivalent to the well-known von Mises condition:

$$\lim_{x\to\infty} \frac{1-F(x)}{xf(x)} = \beta$$

where f is a suitable Lebesgue density of F (restricted to its tail if necessary). For, using (3.6), it is easily seen that F^{-1} is eventually strictly increasing and absolutely continuous with

$$-\frac{1}{F^{-1}(1-t)f(F^{-1}(1-t))} = (\log F^{-1}(1-t))' = -\frac{\beta+\varepsilon(t)}{t}$$

for almost all sufficiently small t. Now the von Mises condition follows immediately by letting $t = 1 - F(x)$. The converse implication was proven by Csörgő et al. (1985, Lemma 6). Consequently, (3.6) is sufficient for $F \in D(G_\beta)$, and it is also necessary if f is eventually monotone (Bingham et al. (1987), Theorem 1.7.2).

Under condition (3.6) the suitably standardized Hill estimator is asymptotically normal for *all* intermediate sequences $(k_n)_{n\in I\!N}$:

$$\mathcal{L}\Big(k_n^{1/2}(\hat\beta_{n,k_n} - \beta - b_{n,k_n})\Big) \longrightarrow \mathcal{N}(0,\beta^2) \qquad \text{weakly}, \tag{3.7}$$

where

$$b_{n,k} := \int_0^1 \varepsilon\Big(\frac{k}{n}t\Big)\,dt = E\Big(\frac{k}{n}\Big) \tag{3.8}$$

with

$$E(t) := t^{-1} \int_0^t \varepsilon(s)\, ds \tag{3.9}$$

(Csörgő et al. (1985), Theorem 1). The following lemma shows that also convergence of the second moment holds if the Hill estimator is slightly modified. To be more concrete, define for an arbitrary $x_0 > 0$

$$\tilde{\beta}_{n,k_n} := \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{\max(X_{n-i+1:n}, x_0)}{\max(X_{n-k_n:n}, x_0)}.$$

Since $P\{X_{n-k_n:n} \leq x_0\} \to 0$ for all intermediate sequences $(k_n)_{n\in\mathbb{N}}$, obviously (3.7) remains true if $\hat{\beta}_{n,k_n}$ is replaced by $\tilde{\beta}_{n,k_n}$.

Lemma 3.1 *If (3.6) is satisfied and $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence, then there are versions of $\tilde{\beta}_{n,k_n}$ and a centered normal r.v. Y with variance β^2 such that*

$$\lim_{n\to\infty} P\Big(\Big(k_n^{1/2}(\tilde{\beta}_{n,k_n} - \beta - b_{n,k_n}) - Y\Big)^2\Big) = 0.$$

In particular,

$$P\Big((\tilde{\beta}_{n,k_n} - \beta)^2\Big) \sim b_{n,k_n}^2 + \beta^2 k_n^{-1} =: AMSE_{n,k_n}.$$

Here $P(Z)$ denotes the expectation of the r.v. Z and again $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$.

Thus an intermediate sequence $(k_n)_{n\in\mathbb{N}}$ is asymptotically optimal in our sense if and only if it minimizes the mean squared error of the modified Hill estimator asymptotically. Without further conditions on the q.f., in general, such an optimal sequence can be calculated only numerically, since the asymptotc bias b_{n,k_n} can behave very irregularly. Therefore, in the sequel we assume that the function E, which determines the asymptotic bias, is regularly varying at 0. Before we characterize the asymptotically optimal sequences $(k_n)_{n\in\mathbb{N}}$ in that case, first the relation between this condition and the second-order Condition 2.1(ii) is clarified. A somewhat related result is given in de Haan et al. (1997), Theorem A.

Lemma 3.2 *Suppose that the q.f. F^{-1} admits a representation of type (3.6).*

(i) If Condition 2.1(ii) holds for some $\delta-$varying function Φ, then E is $\rho-$varying at 0 for some $\rho \geq 0$. Moreover, one has $\rho = \beta$ under the conditions of Lemma 2.3(i) and $\rho = \delta$ else, i.e., under the conditions of Lemma 2.3(ii).

(ii) Conversely, if E is $\rho-$varying at 0 for some $0 \leq \rho \neq \beta$, then Condition 2.1(ii) is satisfied with $\Phi = |E|$.

Remarks 3.1 (i) According to Lemma 3.2(i) all q.f.'s considered in Lemma 2.3(i) are subsumed under the case where E is $\beta-$varying at 0. Furthermore, a comparison of the proofs of these two lemmas shows that in that case $E(t) \sim -\beta^2/(\beta+1)\Xi(t)$ with $\Xi(t)$ defined by (2.36). This fact explains why, in contrast to Theorem 2.8, no distinction of cases is necessary in the above cited result on the asymptotic normality of Hill's estimator.

(ii) By a refinement of the arguments given in the proof of Lemma 3.2 in combination with Theorem 2 of de Haan and Stadtmüller (1996), it can be shown that, under the von Mises condition, the seond-order Condition 2.1(ii) and the regular variation of E at 0 are satisfied by the same q.f.'s, except if for some $d > 0$ the function $t \mapsto F^{-1}(1-t) - dt^{-\beta}$ is slowly varying at 0 but does not belong to the class Π^0, in which case E is $\beta-$varying at 0, but Condition 2.1(ii) is not fulfilled. □

Example 3.1 The d.f. F satisfies the Hall condition (3.3) if and only if

$$F^{-1}(1-t) = C^\beta t^{-\beta}\Big(1 + \beta DC^{-\rho}t^\rho + o(t^\rho)\Big) \tag{3.10}$$

(Csörgő et al. (1985), p. 1055). Under condition (3.6) it follows that $c\exp(\int_0^1 \varepsilon(s)/s\,ds) = C^\beta$ and $-\int_0^t \varepsilon(s)/s\,ds \sim \beta DC^{-\rho}t^\rho$, so that one obtains from equation (3.35) in the proof of Lemma 3.2 that

$$E(t) \sim -\beta DC^{-\rho}\frac{\rho}{\rho+1}t^\rho. \tag{3.11}$$

Conversely, using (3.36) (see proof of Lemma 3.2), it follows from (3.11) that

$$\int_t^1 \frac{\varepsilon(s)}{s}\,ds = E(1) + \int_0^1 \frac{E(s)}{s}\,ds - E(t) - \int_0^t \frac{E(s)}{s}\,ds = E(1) + \int_0^1 \frac{E(s)}{s}\,ds + \beta DC^{-\rho}t^\rho + o(t^\rho).$$

Hence $\int_0^t \varepsilon(s)/s\,ds \sim -\beta DC^{-\rho}t^\rho$, from which (3.10) with $c\exp(\int_0^1 \varepsilon(s)/s\,ds) = C^\beta$ is obvious. □

Next we characterize the asymptotically optimal number of order statistics used by the Hill estimator. In case of $\rho > 0$ the optimal sequence $(k_n)_{n\in\mathbb{N}}$ is asymptotically unique, whereas for $\rho = 0$ one obtains an analogue to the characterization given in Corollary 2.2.

Lemma 3.3 [Drees and Kaufmann (1998), Propositions 1 and 2] *Suppose that (3.6) holds with E being $\rho-$varying at 0 for some $\rho \geq 0$. Let $(k_n^{(1)})_{n\in\mathbb{N}}$ be an intermediate sequence such that*

$$\lim_{n\to\infty} |E(k_n^{(1)}/n)|(k_n^{(1)})^{1/2} = 1. \tag{3.12}$$

(i) For $\rho > 0$, the $AMSE_{n,k_n}$ is asymptotically minimal if and only if

$$k_n \sim k_n^{opt} := \left[\Big(\frac{\beta^2}{2\rho}\Big)^{1/(2\rho+1)}k_n^{(1)}\right] \tag{3.13}$$

or, equivalently,

$$\lim_{n\to\infty} |E(k_n/n)|k_n^{1/2} = \beta(2\rho)^{-1/2}. \tag{3.14}$$

For this choice one has

$$\mathcal{L}\Big(k_n^{1/2}(\hat\beta_{n,k_n} - \beta)\Big) \longrightarrow \mathcal{N}(\mathrm{sgn}(E)\beta(2\rho)^{-1/2}, \beta^2) \qquad \textit{weakly.} \tag{3.15}$$

(Here sgn(E) := 1 if E is positive on $(0,\iota)$ for some $\iota > 0$, and sgn(E) := −1 else.)

(ii) *In case of* $\rho = 0$, *the sequence* $(k_n)_{n\in\mathbb{N}}$ *minimizes* $AMSE_{n,k_n}$ *asymptotically if and only if*

$$\lim_{n\to\infty} \frac{k_n^{(1)}}{k_n} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{E(k_n^{(1)}/n)}{E(k_n/n)} = 1. \tag{3.16}$$

In this case the limiting distribution is degenerated:

$$\frac{\hat\beta_{n,k_n} - \beta}{E(k_n/n)} \longrightarrow 1 \qquad \text{in probability.}$$

Remarks 3.2 (i) The existence of a sequence $(k_n^{(1)})_{n\in\mathbb{N}}$ according to (3.12) and of sequences $(k_n)_{n\in\mathbb{N}}$ satisfying (3.16) follows from the arguments of the Remarks 2.5.

(ii) Based on Lemma 2.8 of Dekkers and de Haan (1993), a different characterization of the optimal number of order statistics can be established, which is less suitable for our purposes. □

Example 3.2 In the Hall model (3.3) one can plug (3.11) into (3.12) to obtain $k_n^{(1)} \sim (C^\rho(\rho+1)/|\beta\rho D|)^{2/(2\rho+1)} n^{2\rho/(2\rho+1)}$, so that (3.13) is equivalent to Hall's and Welsh's characterization (3.2) of the optimal number of order statistics (provided (3.6) holds). □

As explained in the introduction of this chapter, in case of $\rho > 0$ we aim at constructing a consistent estimator of k_n^{opt} that is based on the "stopping times" $\bar k_n(r_n)$ defined by (3.5). In a first step, we verify (3.4) in the general setup (3.6).

Lemma 3.4 [Drees and Kaufmann (1998), Lemma 1] *If (3.6) holds, then for all intermediate sequences* $(k_n)_{n\in\mathbb{N}}$ *one has*

$$\max_{1\le i\le k_n} i^{1/2}|\hat\beta_{n,i} - \beta - b_{n,i}| = O_P((\log\log n)^{1/2}).$$

Since the "random part" of $i^{1/2}|\hat\beta_{n,i} - \hat\beta_{n,k}|$ is stochastically bounded by a term of order $(\log\log n)^{1/2}$, obviously the bias $i^{1/2}|b_{n,i} - b_{n,k}|$ must be the dominant part if $i^{1/2}|\hat\beta_{n,i} - \hat\beta_{n,k}|$ is of larger order than $(\log\log n)^{1/2}$. In particular, this is true for $k_n = \bar k_n(r_n)$ with $(\log\log n)^{1/2} = o(r_n)$, so that $\bar k_n(r_n)$ is asymptotically equivalent to a deterministic sequence. More precisely, the following result holds true.

Corollary 3.1 [Drees and Kaufmann (1998), Corollary 1] *Suppose (3.6) is satisfied and* E *is* ρ*–varying at 0 with* $\rho > 0$. *If* $r_n = o(n^{1/2})$ *and* $(\log\log n)^{1/2} = o(r_n)$, *then*

$$\frac{\bar k_n(r_n)}{\kappa_n} \longrightarrow 1 \qquad \text{in probability}$$

for any deterministic sequence $(\kappa_n)_{n\in\mathbb{N}}$ *satisfying*

$$\lim_{n\to\infty} \frac{|E(\kappa_n/n)|\kappa_n^{1/2}}{r_n} = \frac{(2\rho+1)^{1+1/(2\rho)}}{2\rho}. \tag{3.17}$$

Note that again $r_n = o(n^{1/2})$ and the regular variation of E imply the existence of an asymptotically unique intermediate sequence $(\kappa_n)_{n\in\mathbb{N}}$ satisfying (3.17) (cf. Remarks 2.5).

Since the asymptotic behavior of the optimal sequence $(k_n^{opt})_{n\in\mathbb{N}}$ as well as the asymptotics of the "stopping time" $\bar{k}_n(r_n)$ is determined by the function E in a similar way (cf. (3.14) and (3.17)), it is plausible that one can construct a sequential estimator for k_n^{opt} based on $\bar{k}_n(r_n)$. Moreover, any additional conditions on the sequence of thresholds r_n will depend only on the behavior of the function E in a neighborhood of 0. Hence most of the following arguments and calculations will be of purely analytical nature.

For the time being, in order to introduce the estimator $\hat{k}_n^{opt}$, we concentrate on the Hall model (3.3), where the function E is particularly regular, namely, $E(t) \sim \text{const.}t^\rho$. Therefore, by Corollary 3.1, $\bar{k}_n(r_n) \sim \text{const.}(r_n n^\rho)^{2/(2\rho+1)}$, so that $(\bar{k}_n(r_n^\xi)/(\bar{k}_n(r_n))^\xi)^{1/(1-\xi)}$ with $\xi \in (0,1)$ has the optimal order $n^{2\rho/(2\rho+1)}$ and can serve as the basis of a consistent estimator of k_n^{opt}.

Theorem 3.1 [Drees and Kaufmann (1998), Theorem 1] *Suppose that (3.6) and (3.3) hold, and let r_n be a threshold such that $(\log\log n)^{1/2} = o(r_n)$ and $r_n = o(n^{1/2})$.*

(i) The estimators

$$\hat{\rho}_{n,\lambda}^{(1)} := \log \frac{\max\limits_{1\le i\le[\lambda\bar{k}_n(r_n)]} i^{1/2}|\hat{\beta}_{n,i} - \hat{\beta}_{n,[\lambda\bar{k}_n(r_n)]}|}{\max\limits_{1\le i\le\bar{k}_n(r_n)} i^{1/2}|\hat{\beta}_{n,i} - \hat{\beta}_{n,\bar{k}_n(r_n)}|} \Big/ \log\lambda - \frac{1}{2}$$

and

$$\hat{\rho}_{n,\lambda}^{(2)} := \log \frac{\hat{\beta}_{n,[\lambda^2\bar{k}_n(r_n)]} - \hat{\beta}_{n,[\lambda\bar{k}_n(r_n)]}}{\hat{\beta}_{n,[\lambda\bar{k}_n(r_n)]} - \hat{\beta}_{n,\bar{k}_n(r_n)}} \Big/ \log\lambda,$$

with $\lambda \in (0,1)$, are weakly consistent for ρ.

(ii) If $\hat{\rho}_n$ is any consistent estimator of ρ and $\tilde{\beta}_n$ is a consistent initial estimator of β, then for $\xi \in (0,1)$ and $(\log\log n)^{1/(2\xi)} = o(r_n)$

$$\hat{k}_n^{opt} := \left[(2\hat{\rho}_n + 1)^{-1/\hat{\rho}_n}(2\tilde{\beta}_n^2\hat{\rho}_n)^{1/(2\hat{\rho}_n+1)}\left(\frac{\bar{k}_n(r_n^\xi)}{(\bar{k}_n(r_n))^\xi}\right)^{1/(1-\xi)}\right]$$

is a consistent estimator of k_n^{opt} given by (3.13), in the sense that $\hat{k}_n^{opt}/k_n^{opt} \to 1$ in probability. In particular, $\hat{\beta}_{n,\hat{k}_n^{opt}}$ has the same limiting distribution as $\hat{\beta}_{n,k_n^{opt}}$.

Observe that no initial estimators for the parameters C and D are needed, while one can use the Hill estimator based on an arbitrary intermediate sequence as a consistent initial estimator for β. It is worth mentioning that the assertions of Theorem 3.1 hold true in the full Hall model, i.e., without assumption (3.6). For then an upper bound on the random part of the error can be established, which is analogous to (3.30), and hence the arguments of Section 3.2 yield the desired result.

In view of (3.15), Theorem 3.1 yields asymptotic confidence intervals for β, since the sign of E locally at 0 can be easily estimated from the sign of $\hat{\beta}_{n,i} - \hat{\beta}_{n,\bar{k}_n(r_n)}$ for suitable $i < \bar{k}_n(r_n)$:

Corollary 3.2 *Let*

$$Sg_n := \begin{cases} 1 & \max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}(\hat{\beta}_{n,\bar{k}_n(r_n)} - \hat{\beta}_{n,i}) \ge \max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}(\hat{\beta}_{n,i} - \hat{\beta}_{n,\bar{k}_n(r_n)}), \\ -1 \quad \text{if} & \max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}(\hat{\beta}_{n,\bar{k}_n(r_n)} - \hat{\beta}_{n,i}) < \max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}(\hat{\beta}_{n,i} - \hat{\beta}_{n,\bar{k}_n(r_n)}). \end{cases} \tag{3.18}$$

Then, under the conditions of Theorem 3.1,

$$\lim_{n\to\infty} P\{Sg_n = \operatorname{sgn}(E)\} = 1. \tag{3.19}$$

Consequently, for all $\alpha \in (0,1)$

$$\lim_{n\to\infty} P\Big\{\beta \in \Big[\hat{\beta}_{n,\hat{k}_n^{opt}}(1 - (\hat{k}_n^{opt})^{-1/2}(Sg_n(2\hat{\rho}_n)^{-1/2} + u_{1-\alpha/2})), \\ \hat{\beta}_{n,\hat{k}_n^{opt}}(1 + (\hat{k}_n^{opt})^{-1/2}(-Sg_n(2\hat{\rho}_n)^{-1/2} + u_{1-\alpha/2}))\Big]\Big\} = 1-\alpha,$$

where $u_{1-\alpha/2}$ denotes the $(1-\alpha/2)$–quantile of the standard normal distribution.

Likewise, one-sided asymptotic confidence intervals can be constructed.

A central role in the proof of Theorem 3.1 plays the regularity property $E(\lambda_n k_n/n) \sim \lambda_n^\rho E(k_n/n)$ of the function E in the Hall model. For general ρ–varying functions E, this holds true only if λ_n does not converge to ∞ too fast. In order to specify the additional conditions, recall that by Karamata's representation there exist a constant $d \ne 0$ and a function $\eta : (0,1) \to I\!R$ with $\lim_{t\downarrow 0} \eta(t) = 0$ such that

$$E(t) \sim dt^\rho \exp\Big(\int_t^1 \frac{\eta(s)}{s}\, ds\Big). \tag{3.20}$$

Theorem 3.2 [Drees and Kaufmann (1998), Theorem 2] *If (3.6) holds with function E according to (3.20), $(\log\log n)^{1/2} = o(r_n)$, $r_n = o(n^{1/2})$ and, in addition,*

$$\lim_{n\to\infty} \sup_{0<t\le (r_n^2/n)^{1/(2\rho+1)-\iota}} |\eta(t)| \log r_n = 0 \tag{3.21}$$

for some $\iota > 0$, then $\hat{\rho}_{n,\lambda}^{(1)}$ and $\hat{\rho}_{n,\lambda}^{(2)}$ are weakly consistent estimators for ρ. Moreover, $\hat{k}_n^{opt}$ estimates the optimal number consistently, i.e., $\hat{k}_n^{opt}/k_n^{opt} \to 1$ in probability.

As shown in Corollary 3.2, one may derive asymptotic confidence intervals for β from Theorem 3.2, too.

Remarks 3.3 (i) Recall that one can take

$$\eta(t) := \rho - \frac{E(t)}{\int_0^t E(s)/s\, ds} = \rho + \frac{E(t)}{E(t) - \int_0^t \varepsilon(s)/s\, ds}$$

for sufficiently small t (de Haan (1975), Remark 1.2.3). If ε is ρ–varying at 0, too, then another possible choice is

$$\eta(t) := \rho + 1 - \frac{\varepsilon(t)}{E(t)}.$$

In both cases η measures the accuracy of the pertaining Karamata approximation (cf. de Haan (1975), Theorem 1.2.1).

(ii) If η converges to 0 extremely slowly, then it might be impossible to fulfill (3.21) and $(\log\log n)^{1/(2\xi)} = o(r_n)$ at the same time. In that case, however, the rate of convergence in (3.15) is extremely poor, too, so that the minimization of $AMSE_{n,k_n}$ is not a reasonable criterion from a practical point of view. □

Example 3.3 Suppose that

$$E(t) \sim dt^\rho \exp\Big(\int_t^{t_0} \frac{1}{s\log s}\, ds\Big) \sim \tilde{d}t^\rho |\log t|.$$

Then condition (3.21) reads as

$$\lim_{n\to\infty} \frac{\log r_n}{\log(r_n^2/n)} = 0 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \frac{\log r_n}{\log n} = 0,$$

which is satisfied by all slowly varying functions r_n of n. □

Finally, we turn to the case $\rho = 0$, that is, E is assumed to be slowly varying at 0. In this situation, the problem is of a different nature, since, as mentioned above, the criterion that the AMSE shall be asymptotically minimized does not determine the rate of convergence of the sequence $(k_n)_{n\in I\!N}$.

In the sequel we restrict ourselves to showing that an estimator $\hat{k}_{n,\vartheta}^{opt}$ closely related to $\hat{k}_n^{opt}$ can be defined that satisfies (3.16) in probability, although a different type of estimator may be more appropriate in the present situation. This approach is motivated by the fact that usually it is not known in advance whether ρ equals 0.

For the examination of the asymptotic behavior of $\bar{k}_n(r_n)$ we need stronger conditions, namely, $E \in \Pi_g^0$, i.e., there exists an auxiliary function g, which is slowly varying at 0, such that

$$\lim_{t\downarrow 0} \frac{E(tx) - E(t)}{g(t)} = -\log x$$

for all $x > 0$. Then we have the following counterpart of Corollary 3.1.

Corollary 3.3 [Drees and Kaufmann (1998), Lemma 2] *If (3.6) is satisfied for some $E \in \Pi_g^0$, $r_n = o(n^{1/2})$ and $(\log\log n)^{1/2} = o(r_n)$, then*

$$\frac{\bar{k}_n(r_n)}{\kappa_n} \longrightarrow 1 \qquad \textit{in probability}$$

for any deterministic sequence $(\kappa_n)_{n\in I\!N}$ satisfying

$$\frac{|g(\kappa_n/n)|\kappa_n^{1/2}}{r_n} \longrightarrow \frac{\mathrm{e}}{2}. \tag{3.22}$$

Since E is slowly varying, due to Karamata's representation there exists a function $\eta(t) \to 0$ as $t \downarrow 0$ such that

$$E(t) \;\sim\; d\exp\Big(\int_t^1 \frac{\eta(s)}{s}\, ds\Big). \tag{3.23}$$

The functions E, g and η are linked by

$$g(t) \;\sim\; E(t)\eta(t), \tag{3.24}$$

because, according to the proof of Theorem 1.2.1 and by Theorem 1.4.1 of de Haan (1975), $\eta(t^{-1}) \sim (E(t^{-1}) - t^{-1}\int_1^t E(s^{-1})ds)/E(t^{-1}) \sim -g(t^{-1})/E(t^{-1})$ as $t \to \infty$. In particular, η is slowly varying at 0, too. For the sake of simplicity, we assume in the main part of the following counterpart of Theorem 3.2 that $|\eta|$ is eventually decreasing and that r_n grows more slowly than any power of n. However, it is worth mentioning that the same assertions can be proven under weaker conditions.

Theorem 3.3 [Drees and Kaufmann (1998), Theorem 3] *If (3.6) and (3.23) are satisfied and $(r_n)_{n\in I\!N}$ is a sequence of thresholds such that $(\log\log n)^{1/2} = o(r_n)$ and $r_n = o(n^{1/2})$, then $\rho^{(1)}_{n,\lambda}$ and $\rho^{(2)}_{n,\lambda}$ (defined in Theorem 3.1) converge to $\rho = 0$ in probability.*

Suppose that, in addition, $(\log\log n)^{1/(2\xi)} = o(r_n)$, $\log r_n = o(\log n)$ and $|\eta|$ is an eventually decreasing function such that

(i) $\limsup_{n\to\infty} |\eta(n^{\iota-1})| \log r_n < \infty$ *and*

(ii) $\lim_{n\to\infty} \eta(n^{\iota-1}) \log|\eta(n^{-1})| = 0$

for some $\iota > 0$. Then for all $\vartheta > 0$

$$\hat{k}^{opt}_{n,\vartheta} := \left[\vartheta\left(\frac{\bar{k}_n(r_n^{\xi})}{(\bar{k}_n(r_n))^{\xi}}\right)^{1/(1-\xi)}\right]$$

fulfills (3.16) in probability. Consequently,

$$\frac{\hat{\beta}_{n,\hat{k}^{opt}_{n,\vartheta}} - \beta}{E(k^{(1)}_n/n)} \;\longrightarrow\; 1$$

in probability, i.e., $\hat{k}^{opt}_{n,\vartheta}$ is an asymptotically optimal number of order statistics used by the Hill estimator.

Note that (ii) is a very weak condition, which ensures that η is not too "wiggly". Condition (i) is the counterpart of (3.21) in Theorem 3.2.

Examples 3.4 (i) Let $E(t) = d|\log t|^{-\alpha}$ for some $\alpha > 0$. Then one can choose $\eta(t) := \alpha/\log t$ for sufficiently small t, and the conditions of Theorem 3.3 are satisfied for all slowly varying sequences $(r_n)_{n\in I\!N}$, because $\log r_n/\log n \to 0$ and $\log\log n/\log n \to 0$.

(ii) Suppose $E(t) = d\exp(-|\log t|^{\alpha})$ with $\alpha \in (0,1)$, so that $\eta(t) = -\alpha(-\log t)^{\alpha-1}$ satisfies the conditions (i) and (ii) of Theorem 3.3 if $\log r_n = O(\log^{1-\alpha} n)$, e.g., for $r_n = \log^{\tau} n$ with $\tau > 0$. □

3.2 Optimal sample fraction for statistical tail functionals

As mentioned above, an important strong point of the sequential approach is the fact that it can be utilized for a large class of estimators of the extreme value index, including quite general statistical tail functionals. For the sake of clarity, we restrict ourselves to scale- and location-invariant functionals, yet similar results can be also proven for the estimators considered in the Sections 2.3 and 2.4. Essentially in this section we work out the ideas outlined in the discussion following Example 3 of Drees and Kaufmann (1998) (see also the last paragraph of Section 1 of that paper).

Recall from (1.4) that the empirical tail q.f. based on the $i+1$ largest order statistics is denoted by $Q_{n,i}$. We consider estimators

$$\hat{\beta}_{n,i} := T(Q_{n,i})$$

where $T : span(D_{\beta,h}, 1) \to I\!R$ is a functional satisfying Condition 2.3.

First, in analogy to Lemma 3.3, we give a characterization of those intermediate sequences $(k_n)_{n\in I\!N}$ minimizing the AMSE $(\mu_{T,\beta,\Psi}\Phi(k_n/n))^2 + \sigma^2_{T,\beta}k_n^{-1}$ of $\hat{\beta}_{n,k_n}$ if the second-order condition 2.1(ii) is satisfied for some function Φ that is regularly varying at 0 with positive second-order parameter δ. For this purpose, we have to assume that the bias term $\mu_{T,\beta,\Psi}$ does not vanish (cf. the discussion following Example 2.4). Recall that the case $\delta = 0$ was already considered in Corollary 2.2.

Lemma 3.5 *Suppose that the functional T satisfies the Conditions 2.3 and 2.4. Moreover, assume that Condition 2.1(ii) holds for a function Φ that is δ–varying at 0 for some $\delta > 0$, and that $\mu_{T,\beta,\Psi} = T'_\beta(\bar{\Psi}) \neq 0$. Let the sequence $(k_n^{(1)})_{n\in I\!N}$ satisfy (2.30), that is,*

$$\lim_{n\to\infty} \Phi(k_n^{(1)}/n)(k_n^{(1)})^{1/2} = 1.$$

Then the AMSE of $\hat{\beta}_{n,k_n} = T(Q_{n,k_n})$ is asymptotically minimal if and only if

$$k_n \sim k_n^{opt} := \Big[\Big(\frac{\sigma^2_{T,\beta}}{2\delta\mu^2_{T,\beta,\Psi}}\Big)^{1/(2\delta+1)}k_n^{(1)}\Big] \tag{3.25}$$

or, equivalently,

$$\lim_{n\to\infty} \Phi(k_n/n)k_n^{1/2} = (2\delta)^{-1/2}\Big|\frac{\sigma_{T,\beta}}{\mu_{T,\beta,\Psi}}\Big|. \tag{3.26}$$

For this choice one has

$$\mathcal{L}\Big(k_n^{1/2}(\hat{\beta}_{n,k_n} - \beta)\Big) \longrightarrow \mathcal{N}(\mathrm{sgn}(\mu_{T,\beta,\Psi})(2\delta)^{-1/2}|\sigma_{T,\beta}|, \sigma^2_{T,\beta}) \qquad \textit{weakly.} \tag{3.27}$$

Next we examine the "stopping time"

$$\bar{k}_n(r_n) := \min\Big\{k \in \{1,\dots,n\} \mid \max_{1\le i\le k} i^{1/2}|\hat{\beta}_{n,i} - \hat{\beta}_{n,k}| > r_n\Big\}. \tag{3.28}$$

As we have seen in the case of the Hill estimator, the asymptotic behavior of $\bar{k}_n(r_n)$ is determined by two important properties of the estimation error of $\hat{\beta}_{n,i}$: First, the random part of the standardized error is at most of the stochastic order $(\log\log n)^{1/2}$ uniformly for all $1 \le i \le k_n$ if $(k_n)_{n\in I\!N}$ is an intermediate sequence; and second, the deterministic part of the error, that is, the asymptotic bias $b_{n,i}$, is a regularly varying function of i/n.

While under Condition 2.1(ii) the statistical tail functionals considered here possess the latter property, in general one must add a further term, which is of smaller order than the asymptotic bias, to the stochastic bound of the standardized random error part. In order to establish such a bound *uniformly for* $1 \le i \le k_n$, we need stronger smoothness assumptions than merely Hadamard differentiability, which is sufficient only for weak convergence of a single sequence of estimators.

Condition 3.1 *The functional $T_{|D_{\beta,h}}$ is Fréchet differentiable at $\bar F_\beta^{-1}$ with derivative $T'_\beta : D_{\beta,h} \to I\!R$ and, in addition, it is bounded on bounded subsets of $D_{\beta,h}$.*

Remark 3.4 Analogously to Remark 2.3(i), it is sufficient to require the Fréchet differentiability of $T_{|\bar D_{\beta,h}}$ at $\bar F_\beta^{-1}$ if the scale-invariant subset $\bar D_{\beta,h} \subset D_{\beta,h}$ satisfies $\bar F_\beta^{-1} \in \bar D_{\beta,h}$ and $P\{Q_{n,i} \in span(\bar D_{\beta,h}, 1)\ \forall 1 \le i \le k_n\} \to 1$, that is,

$$\frac{T(\bar F_\beta^{-1} + \lambda z) - T(\bar F_\beta^{-1})}{\lambda} \longrightarrow T'_\beta(z)$$

as $\lambda \downarrow 0$ uniformly for all $z \in D_{\beta,h}$ such that $\|z\|_{\beta,h} \le 1$ and $\bar F_\beta^{-1} + \lambda z \in \bar D_{\beta,h}$. □

It is well known that Fréchet differentiability is a much stronger restriction for classical statistical functionals than Hadamard differentiability; as Gill (1989, page 100 f.) puts it: "Fréchet differentiability is so strong that hardly any interesting statistical functionals are differentiable at all". In contrast to this, all functionals considered in Section 2.2, namely, the Pickands functional T_P, the probability weighted moment functional T_{PWM,ν_1,ν_2} and its "adaptive" version defined in (2.35), and the ML-functional examined in Example 2.3, are Fréchet differentiable. (To prove the last assertion use Theorem 1.4.7 of Rieder (1994) instead of Theorem 1.4.2.)

Moreover, under this differentiability assumption boundedness can be always achieved by a modification of the functional $T_{|D_{\beta,h}}$ without changing the asymptotic distribution of the pertaining statistical tail functional $T(Q_{n,k_n})$. To see this, note that by the Fréchet differentiability $T_{|D_{\beta,h}}$ is bounded on a ball with center $\bar F_\beta^{-1}$ and sufficiently small radius r. Hence, by the scale and location invariance, it is also bounded on the set $S := \{a(\bar F_\beta^{-1} + z) + b \in D_{\beta,h} \,|\, a > 0, b \in I\!R, \|z\|_{\beta,h} < r\}$ if the constant functions belong to $D_{\beta,h}$, and $S := \{a(\bar F_\beta^{-1} + z) \in D_{\beta,h} \,|\, a > 0, \|z\|_{\beta,h} < r\}$ else. Consequently, the functional $\tilde T : span(D_{\beta,h}, 1) \to I\!R$ defined by $\tilde T(z) := T(z)1_S(z)$ for $z \in D_{\beta,h}$, and $\tilde T(z+b) := \tilde T(z)$ for $z \in D_{\beta,h}$ and $b \in I\!R$ if $1 \notin D_{\beta,h}$, satisfies the Conditions 2.3 and 3.1, while the pertaining statistical tail functionals $\tilde T(Q_{n,k_n})$ and $T(Q_{n,k_n})$ have the same limiting distributions, because $\tilde T$ is equal to T on a neighborhood of $\bar F_\beta^{-1}$.

So apparently Condition 3.1 is only slightly more restrictive than Condition 2.4.

Now we are ready to state the following counterpart of the central Lemma 3.4, which is given for the r.v.'s $\tilde\beta_{n,i} = T(\tilde Q_{n,i})$ instead of $\hat\beta_{n,i}$ with $\tilde Q_{n,i}$ defined by (2.7). Recall that, in view of (2.6), the r.v.'s $\tilde\beta_{n,i}$ and the statistical tail functionals $\hat\beta_{n,i}$ are linked via

$$\|\mathcal{L}((\tilde\beta_{n,i})_{1\le i\le k_n}) - \mathcal{L}((\hat\beta_{n,i})_{1\le i\le k_n})\| \longrightarrow 0. \tag{3.29}$$

Lemma 3.6 [cf. Drees and Kaufmann (1998), (35)] *Suppose that the functional T satisfies the Conditions 2.3 and 3.1, and that the second-order Condition 2.1(ii) is fulfilled. Then for every intermediate sequence $(k_n)_{n\in I\!N}$ we have*

$$\max_{1\le i\le k_n} i^{1/2}\Big|\tilde\beta_{n,i}-\beta-b_{n,i}-rem_{n,i}\Big| = O\Big((\log\log k_n)^{1/2}\Big) \qquad a.s. \tag{3.30}$$

where

$$b_{n,i} := T'_\beta(\bar\Psi)\Phi(i/n) = \mu_{T,\beta,\Psi}\Phi(i/n) \tag{3.31}$$

and the r.v.'s $rem_{n,i}$ satisfy

$$\max_{1\le i\le k_n}\frac{rem_{n,i}}{\Phi(i/n)} = o(1) \qquad a.s.$$

Though Lemma 3.6 is weaker than the corresponding Lemma 3.4 for the Hill estimator, it is nevertheless sufficient to prove a counterpart of Corollary 3.1, since, besides the regularity of the asymptotic bias, there we have only used the dominance of the bias over the random error part if $\max_{1\le i\le k}|\hat\beta_{n,i}-\hat\beta_{n,k}|$ is of larger order than $(\log\log n)^{1/2}$. Of course, in the present situation again one has to assume that the bias term $b_{n,i}$ does not vanish.

Corollary 3.4 *Suppose T satisfies the Conditions 2.3 and 3.1 and the second-order condition 2.1(ii) holds for some δ–varying function Φ with $\delta>0$. If $\mu_{T,\beta,\Psi}\neq 0$ and the sequence of thresholds r_n is such that $r_n=o(n^{1/2})$ and $(\log\log n)^{1/2}=o(r_n)$, then*

$$\frac{\bar k_n(r_n)}{\kappa_n} \longrightarrow 1 \qquad \textit{in probability}$$

for any deterministic sequence $(\kappa_n)_{n\in I\!N}$ satisfying

$$\lim_{n\to\infty}\frac{\Phi(\kappa_n/n)\kappa_n^{1/2}}{r_n} = \frac{(2\delta+1)^{1+1/(2\delta)}}{2\delta|\mu_{T,\beta,\Psi}|}.$$

The existence and asymptotic uniqueness of the sequence $(\kappa_n)_{n\in I\!N}$ is ensured by the same arguments as in the Remarks 2.5.

Using the Lemmas 3.5 and 3.6 and Corollary 3.4, one can readily transfer the Theorems 3.1 and 3.2 and Corollary 3.2 to the present situation. Recall that the central condition of Theorem 3.1, namely, $E(t)\sim \text{const.}t^\delta$ for some $\delta>0$, is equivalent to the Hall condition and also to representation (3.10) of the q.f., provided (3.6) holds. Before we state the main results of this section, we establish a similar representation of the q.f. that is equivalent to the corresponding condition $\Phi(t)\sim\text{const.}t^\delta$ in the present setup.

Lemma 3.7 *Condition 2.1(ii) holds with $\Phi(t)\sim const.t^\delta$, $\delta>0$, if and only if there are constants $b,d\in I\!R$ and $c>0$ such that*

$$F^{-1}(1-t) = b + c\bar F_\beta^{-1}(t) + d\bar F_{\beta-\delta}^{-1}(t)(1+o(1)) \tag{3.32}$$

as $t\downarrow 0$.

Theorem 3.4 *Suppose the q.f. is of type (3.32) and T satisfies the Conditions 2.3 and 3.1 such that $\mu_{T,\beta,\Psi} = T'_\beta(\bar\Psi) \neq 0$. Moreover, let $(r_n)_{n\in I\!N}$ be a sequence such that $(\log\log n)^{1/2} = o(r_n)$ and $r_n = o(n^{1/2})$.*

(i) The estimators

$$\hat\delta^{(1)}_{n,\lambda} := \log \frac{\max\limits_{1\le i\le[\lambda\bar k_n(r_n)]} i^{1/2}|\hat\beta_{n,i} - \hat\beta_{n,[\lambda\bar k_n(r_n)]}|}{\max\limits_{1\le i\le \bar k_n(r_n)} i^{1/2}|\hat\beta_{n,i} - \hat\beta_{n,\bar k_n(r_n)}|} \Big/ \log\lambda - \frac{1}{2}$$

and

$$\hat\delta^{(2)}_{n,\lambda} := \log \frac{\hat\beta_{n,[\lambda^2\bar k_n(r_n)]} - \hat\beta_{n,[\lambda\bar k_n(r_n)]}}{\hat\beta_{n,[\lambda\bar k_n(r_n)]} - \hat\beta_{n,\bar k_n(r_n)}} \Big/ \log\lambda,$$

with $\lambda\in(0,1)$ and $\bar k_n(r_n)$ defined by (3.28), are weakly consistent for δ.

(ii) Assume that the function $\beta\mapsto\sigma^2_{T,\beta}$ is continuous. If $\hat\delta_n$ denotes an arbitrary consistent estimator of δ and $\tilde\beta_n$ is a consistent initial estimator of β, then for $\xi\in(0,1)$ and $(\log\log n)^{1/(2\xi)} = o(r_n)$

$$\hat k^{opt}_n := \Big[(2\hat\delta_n+1)^{-1/\hat\delta_n}(2\hat\delta_n\sigma^2_{T,\tilde\beta_n})^{1/(2\hat\delta_n+1)}\Big(\frac{\bar k_n(r_n^\xi)}{(\bar k_n(r_n))^\xi}\Big)^{1/(1-\xi)}\Big]$$

is a consistent estimator of k^{opt}_n given by (3.25), in the sense that $\hat k^{opt}_n/k^{opt}_n \to 1$ in probability. In particular, $\hat\beta_{n,\hat k^{opt}_n}$ has the same limiting distribution as $\hat\beta_{n,k^{opt}_n}$.

Observe that, similarly as in Theorem 3.1, no initial estimator of the bias term $\mu_{T,\beta,\Psi}$ is needed. The continuity condition (2.34) implies the continuity of $\beta\mapsto\sigma^2_{T,\beta}$, which enables the construction of consistent estimators for the limiting variance $\sigma^2_{T,\beta}$.

Corollary 3.5 *Under the conditions of Theorem 3.4, one has for Sg_n defined by (3.18)*

$$\lim_{n\to\infty} P\{Sg_n = \mathrm{sgn}(\mu_{T,\beta,\Psi})\} = 1.$$

Thus for all $\alpha\in(0,1)$

$$\lim_{n\to\infty} P\Big\{\beta\in\Big[\hat\beta_{n,\hat k^{opt}_n} - (\hat k^{opt}_n)^{-1/2}|\sigma_{T,\hat\beta_{n,\hat k^{opt}_n}}|(Sg_n(2\hat\delta_n)^{-1/2} + u_{1-\alpha/2})),$$
$$\hat\beta_{n,\hat k^{opt}_n} + (\hat k^{opt}_n)^{-1/2}|\sigma_{T,\hat\beta_{n,\hat k^{opt}_n}}|(-Sg_n(2\hat\delta_n)^{-1/2} + u_{1-\alpha/2}))\Big]\Big\} = 1-\alpha,$$

with $u_{1-\alpha/2}$ denoting the $(1-\alpha/2)$–quantile of the standard normal distribution.

Theorem 3.5 *Suppose that the conditions on T, $(r_n)_{n\in I\!N}$ and $\sigma^2_{T,\beta}$ imposed in Theorem 3.4 are fulfilled. In addition, assume that Condition 2.1(ii) holds with*

$$\Phi(t)\sim dt^\delta\exp\Big(\int_t^1 \frac{\eta(s)}{s}\,ds\Big),$$

where $d,\delta>0$ and $lim_{t\downarrow 0}\eta(t) = 0$. If (3.21) is satisfied for some $\iota>0$, then the assertions of Theorem 3.4 hold true.

To sum up, for a positive second-order parameter δ our sequential estimator for the optimal number of order statistics works for quite general estimators of the extreme value index β under conditions that are comparable with those assumed in Section 3.1.

In contrast to this, to prove an analog to Theorem 3.3 for statistical tail functionals we would need essentially stronger conditions, including more restrictive assumptions about the remainder term $R(t,x)$ in Condition 2.1(ii). For, in case of $\delta = 0$, the maximal weighted deviation of the asymptotic bias of two estimators based on a different number of order statistics, i.e., $\max_{1\le i\le k} i^{1/2}|b_{n,i} - b_{n,k}|$, is of smaller order than the weighted asymptotic bias $k^{1/2}b_{n,k}$ itself (cf. proof of Corollary 3.3), so that the terms $rem_{n,i}$ need not be negligible any longer, and hence one cannot conclude from Lemma 3.6 that the bias dominates the random error part.

3.3 Proofs

Proof of Lemma 3.1. It suffices to prove that

$$\limsup_{n\to\infty} P\Big(\big(k_n^{1/2}(\tilde\beta_{n,k_n} - \beta - b_{n,k_n})\big)^4\Big) < \infty$$

(see Billingsley (1986), Theorems 25.6 and 16.13 and the remark following the latter). W.l.o.g. we assume that $X_i = F^{-1}(1-U_i)$ for i.i.d. uniformly on $(0,1)$ distributed r.v.'s U_i. Since $E(t) \to 0$ as $t \downarrow 0$, for all sufficiently small $a > 0$ with $F^{-1}(1-a) > x_0$ and sufficiently large n one has

$$\begin{aligned}
&P\Big(\big(k_n^{1/2}(\tilde\beta_{n,k_n} - \beta - E(k_n/n))\big)^4 1_{\{X_{n-k_n:n} < F^{-1}(1-a)\}}\Big)\\
&\quad\le\; k_n^2 \max\Big(P\Big(\tilde\beta^4_{n,k_n} 1_{\{U_{k_n+1:n} > a\}}\Big), 2\beta^4 P\{U_{k_n+1:n} > a\}\Big)\\
&\quad\le\; k_n^2 \max\Big(\Big(P\Big(\log^8 \max(F^{-1}(1-U_{1:n})/x_0, 1)\Big) P\{U_{k_n+1:n} > a\}\Big)^{1/2},\\
&\qquad\qquad 2\beta^4 P\{U_{k_n+1:n} > a\}\Big). \qquad (3.33)
\end{aligned}$$

By (3.6) one has $|\log\max(F^{-1}(1-U_{1:n})/x_0, 1)| \le \text{const.}(1 + |\log U_{1:n}|)$ (where const. denotes a generic constant independent of n, which may vary from line to line) and hence by Reiss (1989), Lemma 1.7.2,

$$P\Big(\log^8 \max(F^{-1}(1-U_{1:n})/x_0, 1)\Big) \le \text{const.}(1 + nP(\log^8 U_1)) = O(n). \qquad (3.34)$$

Moreover, according to Reiss (1989), Problem 3.1, we have, for sufficiently large n, $P\{U_{k_n+1:n} > a\} \le \exp(-na^2/4)$, so that by (3.33) and (3.34) it follows that

$$\lim_{n\to\infty} P\Big(\big(k_n^{1/2}(\tilde\beta_{n,k_n} - \beta - E(k_n/n))\big)^4 1_{\{X_{n-k_n:n} < F^{-1}(1-a)\}}\Big) = 0.$$

Next we turn to

$$\begin{aligned}
&P\Big(\big(k_n^{1/2}(\tilde\beta_{n,k_n} - \beta - b_{n,k_n})\big)^4 1_{\{X_{n-k_n:n} \ge F^{-1}(1-a)\}}\Big)\\
&= k_n^2 \int_0^a P\Big(\Big(\frac{1}{k_n}\sum_{i=1}^{k_n} \log\frac{F^{-1}(1-U_{i:n})}{F^{-1}(1-U_{k_n+1:n})} - \beta - b_{n,k_n}\Big)^4 \,\Big|\, U_{k_n+1:n} = u\Big)\, \mathcal{L}(U_{k_n+1:n})(du)\\
&= k_n^2 \int_0^a P\Big(\Big(\frac{1}{k_n}\sum_{i=1}^{k_n} \log\frac{F^{-1}(1-V_{i,u})}{F^{-1}(1-u)} - \beta - E(k_n/n)\Big)^4\Big)\, \mathcal{L}(U_{k_n+1:n})(du)
\end{aligned}$$

with $V_{i,u}$, $i \in I\!N$, denoting i.i.d. r.v.'s that are uniformly distributed on $(0,u)$ (see Reiss (1989), Theorem 1.8.1).

Direct calculations show that

$$\begin{aligned}
P\Big(\log\frac{F^{-1}(1-V_{1,u})}{F^{-1}(1-u)}\Big) &= \frac{1}{u}\int_0^u -\beta\log\frac{v}{u}+\int_v^u\frac{\varepsilon(s)}{s}\,ds\,dv=\beta+E(u),\\
P\Big(\Big(\log\frac{F^{-1}(1-V_{1,u})}{F^{-1}(1-u)}-\beta-E(u)\Big)^2\Big) &= \int_0^1\Big(-\beta\log w+\int_w^1\frac{\varepsilon(su)}{s}\,ds-\beta-E(u)\Big)^2dw\\
&\le \text{const.}\Big(1+\sup_{0<t\le a}(\varepsilon^2(t)+E^2(t))\Big)<\infty,
\end{aligned}$$

and, likewise,

$$P\Big(\Big(\log\frac{F^{-1}(1-V_{1,u})}{F^{-1}(1-u)}-\beta-E(u)\Big)^4\Big) \le \text{const.}<\infty$$

uniformly for $0<u\le a$. Therefore, in the inequality

$$\begin{aligned}
&k_n^2P\Big(\Big(\frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{F^{-1}(1-V_{i,u})}{F^{-1}(1-u)}-\beta-E(k_n/n)\Big)^4\Big)\\
&\le 8k_n^2P\Big(\Big(\frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{F^{-1}(1-V_{i,u})}{F^{-1}(1-u)}-\beta-E(u)\Big)^4\Big)+8k_n^2\Big(E(u)-E(k_n/n)\Big)^4
\end{aligned}$$

the first term of the right-hand side equals

$$\begin{aligned}
&8k_n^{-2}\Big(k_nP\Big(\Big(\log\frac{F^{-1}(1-V_{1,u})}{F^{-1}(1-u)}-\beta-E(u)\Big)^4\Big)\\
&\qquad\qquad +3k_n(k_n-1)\Big(P\Big(\Big(\log\frac{F^{-1}(1-V_{1,u})}{F^{-1}(1-u)}-\beta-E(u)\Big)^2\Big)\Big)^2\Big),
\end{aligned}$$

which is uniformly bounded for $0<u\le a$.

It remains to prove that $\int_0^a(E(u)-E(k_n/n))^4\,\mathcal{L}(U_{k_n+1:n})(du)=O(k_n^{-2})$. For this end, first note that $E'(t)=(\varepsilon(t)-E(t))/t$ for almost all t and hence by the mean value theorem

$$\begin{aligned}
|E(u)-E(k_n/n)| &\le \sup_{0<t\le a}|\varepsilon(t)-E(t)|\frac{|u-k_n/n|}{\min(u,k_n/n)}\\
&\le \text{const.}\begin{cases} k_n/(nu)-1 & \text{if } u\le e^{-2}k_n/n,\\ n/k_n|u-k_n/n| & \text{if } u>e^{-2}k_n/n.\end{cases}
\end{aligned}$$

Now

$$\begin{aligned}
&\int_0^{\mathrm{e}^{-2}k_n/n}\Big(\frac{k_n}{nu}-1\Big)^4\,\mathcal{L}(U_{k_n+1:n})(du)\\
&= \int_0^\infty P\Big\{\Big(\frac{k_n}{nU_{k_n+1:n}}-1\Big)^4 1_{\{U_{k_n+1:n}\le \mathrm{e}^{-2}k_n/n\}}>x\Big\}\,dx\\
&= (\mathrm{e}^2-1)^4P\{U_{k_n+1:n}\le \mathrm{e}^{-2}k_n/n\}+\int_{(\mathrm{e}^2-1)^4}^\infty P\Big\{U_{k_n+1:n}<\frac{k_n}{n}\frac{1}{1+x^{1/4}}\Big\}\,dx\\
&= O\Big(\mathrm{e}^{-k_n}+\int_{(\mathrm{e}^2-1)^4}^\infty\Big(\frac{\mathrm{e}}{1+x^{1/4}}\Big)^{k_n+1}dx\Big)=o(k_n^{-2}),
\end{aligned}$$

where for the first equation Billingsley (1986), Problem 21.9, and in the third step Lemma 3.1.2 of Reiss (1989) have been utilized.

Furthermore, by Reiss (1989), Lemma 3.1.3, one has

$$\begin{aligned}
&\int_{e^{-2}k_n/n}^{\infty}\Big(\frac{n}{k_n}(u-k_n/n)\Big)^4\,\mathcal{L}(U_{k_n+1:n})(du)\\
&\le\ 8\Big(\frac{n}{k_n}\Big)^4\Big(P\Big(\Big(U_{k_n+1:n}-\frac{k_n+1}{n+1}\Big)^4\Big)+\Big(\frac{k_n+1}{n+1}-\frac{k_n}{n}\Big)^4\Big)=O(k_n^{-2}).
\end{aligned}$$

Hence the first assertion, from which the second one is immediate. □

Proof of Lemma 3.2. First assume that Condition 2.1(ii) holds with $\delta>0$. Recall from (2.37) that then

$$F^{-1}(1-t)=d_1t^{-\beta}\Big(1+(d_2+g(t))t^{\beta}/d_1\Big),$$

where the function g is $(\delta-\beta)$–varying at 0 if $\beta\neq\delta$, and it is Π–varying at 0 otherwise. A comparison with (3.6) shows that $c\exp(\int_0^1\varepsilon(s)/s\,ds)=d_1$ and

$$h(t):=-\int_0^t\frac{\varepsilon(s)}{s}\,ds\sim\frac{d_2+g(t)}{d_1}t^{\beta},$$

which is a ρ–varying function at 0, with $\rho=\beta$ under the conditions of Lemma 2.3(i), and $\rho=\delta$ else. Next note that integration by parts and Karamata's theorem (de Haan (1975), Theorem 1.2.1) yield

$$E(t)=\frac{1}{t}\int_0^t s\frac{\varepsilon(s)}{s}\,ds=-h(t)+\frac{1}{t}\int_0^t h(s)\,ds\sim-\frac{\rho}{\rho+1}h(t) \tag{3.35}$$

and hence assertion (i) for $\delta>0$.

In the case $\delta=0$ one has $(t\mapsto t^{\beta}F^{-1}(1-t))\in\Pi^0$ (cf. proof of Lemma 2.3), so that by Lemma 2.5 of Dekkers and de Haan (1993) the function $\log F^{-1}(1-t)+\beta\log t-\log c=\int_t^1\varepsilon(s)/s\,ds=:\tilde h(t)$ is Π–varying at 0 with some slowly varying auxiliary function ℓ (i.e., $(\tilde h(tx)-\tilde h(t))/\ell(t)\to-\log x$ as $t\downarrow 0$). Again integration by parts gives $E(t)=-\tilde h(t)+t^{-1}\int_0^t\tilde h(s)\,ds$, so that Bingham et al. (1987), Theorem 3.7.1, implies $E(t)\sim\ell(t)$, which shows that E is slowly varying, too.

For the converse implication, i.e., assertion (ii), note that by Bingham et al. (1987), Theorem 1.6.5, $\int_0^t\varepsilon(s)/s\,ds\sim E(t)(\rho+1)/\rho$ if E is ρ–varying at 0 for some $\rho>0$. Consequently, in view of (3.6), it follows for $a(t):=\beta c\exp(\int_0^1\varepsilon(s)/s\,ds)t^{-\beta}$

$$\begin{aligned}
&\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{a(t)}-F_{\beta}^{-1}(t)\\
&=\ \frac{x^{-\beta}}{\beta}\Big(\exp\Big(-\int_0^{tx}\frac{\varepsilon(s)}{s}\,ds\Big)-1\Big)-\frac{1}{\beta}\Big(\exp\Big(-\int_0^{t}\frac{\varepsilon(s)}{s}\,ds\Big)-1\Big)\\
&=\ -\frac{\rho+1}{\rho}E(t)\frac{x^{\rho-\beta}-1}{\beta}+o(E(t)),
\end{aligned}$$

where for the last step $\exp(-x)-1\sim -x$ as $x\to 0$ and the ρ–variation of E have been utilized. Now it is obvious that Condition 2.1(ii) holds with $\Phi(t)=|E(t)|$ if $\rho\neq\beta$.

Finally, consider the case $\rho = 0$. Check that

$$\tilde{h}(t) := \int_t^1 \frac{\varepsilon(s)}{s}\,ds = E(1) - E(t) + \int_t^1 \frac{E(s)}{s}\,ds. \tag{3.36}$$

Hence $\tilde{h}$ is $\Pi-$varying at 0 with auxiliary function E (de Haan (1975), Theorem 1.4.3) and

$$\frac{F^{-1}(1-tx) - F^{-1}(1-t)}{\beta F^{-1}(1-t)} - F_\beta^{-1}(t) = \frac{x^{-\beta}}{\beta}\Big(\exp(\tilde{h}(tx) - \tilde{h}(t)) - 1\Big) \sim -\frac{x^{-\beta}}{\beta}\log(x)E(t).$$

□

Proof of Lemma 3.3. Since the proof of assertion (ii) is virtually the same as the proof of Corollary 2.2, we restrict ourselves to part (i).

Check that the squared asymptotic bias and the asymptotic variance of $\hat{\beta}_{n,k_n^{(1)}}$ are of the same order. Because the asymptotic bias $E(k_n/n)$ is of larger order if $k_n^{(1)} = o(k_n)$ (de Haan (1975), Corollary 1.2.1 2.) and the asymptotic variance $\beta^2 k_n^{-1}$ is of larger order for $k_n = o(k_n^{(1)})$, it suffices to consider sequences of the type $k_n \sim \lambda k_n^{(1)}$ for some $\lambda \in (0,\infty)$ and minimize

$$E^2\Big(\frac{k_n}{n}\Big) + \beta^2 k_n^{-1} \sim (\lambda^{2\rho} + \beta^2\lambda^{-1})/k_n^{(1)}.$$

Now routine calculations show the optimality of the choice (3.13) and the stated equivalency of the two characterizations as well. The limiting distribution is immediate from (3.7). □

Proof of Lemma 3.4. W.l.o.g. one may assume that $X_i = F^{-1}(1-U_i)$ where U_i are i.i.d. uniformly on (0,1) distributed r.v.'s. Denote by q_n (the right continuous version of) the empirical q.f. pertaining to U_i:

$$q_n(t) = U_{i:n} \qquad \text{if} \qquad \frac{i-1}{n} \le t < \frac{i}{n}. \tag{3.37}$$

In case of $U_{i+1:n} \le t_0$, using integration by parts we obtain

$$\begin{aligned}
&\hat{\beta}_{n,i} && (3.38)\\
&= \int_0^1 \log\frac{F^{-1}(1-q_n(ti/n))}{F^{-1}(1-q_n(i/n))}\,dt\\
&= \beta + \int_0^1\int_{ti/n}^{i/n}\frac{\varepsilon(s)}{s}\,ds\,dt + \int_0^1 \log\frac{F^{-1}(1-q_n(ti/n))}{F^{-1}(1-ti/n)}\,dt - \log\frac{F^{-1}(1-q_n(i/n))}{F^{-1}(1-i/n)}\\
&= \beta + b_{n,i} + \int_0^1 \log\frac{F^{-1}(1-q_n(ti/n))}{F^{-1}(1-ti/n)}\,dt - \log\frac{F^{-1}(1-q_n(i/n))}{F^{-1}(1-i/n)}. && (3.39)
\end{aligned}$$

Hence

$$\begin{aligned}
&|\hat{\beta}_{n,i} - \beta - b_{n,i}|\\
&\le \Big|\int_0^1\Big(-\beta\log\frac{q_n(ti/n)}{ti/n} + \int_{q_n(ti/n)}^{ti/n}\frac{\varepsilon(s)}{s}\,ds\Big)dt + \beta\log\frac{q_n(i/n)}{i/n} - \int_{q_n(i/n)}^{i/n}\frac{\varepsilon(s)}{s}\,ds\Big|\\
&= O_P\Big(\int_0^1\Big|\log\frac{q_n(ti/n)}{ti/n}\Big|\,dt + \Big|\log\frac{q_n(i/n)}{i/n}\Big|\Big)
\end{aligned}$$

uniformly for $1 \leq i \leq k_n$, because $\varepsilon(s) \to 0$ as $s \downarrow 0$. Therefore, it suffices to prove that

$$\max_{1 \leq i \leq k_n} i^{1/2} \int_0^1 \Big| \log \frac{q_n(ti/n)}{ti/n} \Big| \, dt = O_P((\log\log n)^{1/2}) \tag{3.40}$$

and

$$\max_{1 \leq i \leq k_n} i^{1/2} \Big| \log \frac{q_n(i/n)}{i/n} \Big| = O_P((\log\log n)^{1/2}). \tag{3.41}$$

First note that

$$\max_{1 \leq i \leq k_n} i^{1/2} \int_0^{1/i} \Big| \log \frac{q_n(ti/n)}{ti/n} \Big| \, dt \leq \max_{1 \leq i \leq k_n} i^{-1/2} \Big(|\log(nU_{1:n})| - \int_0^1 \log t \, dt \Big) = O_P(1),$$

since $nU_{1:n}$ converges weakly to a standard exponential r.v. On the other hand,

$$\sup_{1/n \leq t \leq 1} \Big| \log \frac{q_n(t)}{t} \Big| = O_P(1)$$

(Shorack and Wellner (1986), (10.3.7) and (10.3.8)) and the mean value theorem imply

$$\sup_{1/n \leq t \leq 1} \Big| \frac{\log(q_n(t)/t)}{q_n(t)/t - 1} \Big| = O_P(1), \tag{3.42}$$

so that

$$\begin{aligned} \max_{1 \leq i \leq k_n} i^{1/2} \int_{1/i}^1 \Big| \log \frac{q_n(ti/n)}{ti/n} \Big| \, dt &= \max_{1 \leq i \leq k_n} \Big(\frac{n}{i}\Big)^{1/2} \int_{1/n}^{i/n} n^{1/2} \Big| \log \frac{q_n(t)}{t} \Big| \, dt \\ &= O_P\Big(\max_{1 \leq i \leq k_n} \Big(\frac{n}{i}\Big)^{1/2} \int_{1/n}^{i/n} n^{1/2} t^{-1/2} |q_n(t) - t| \, t^{-1/2} \, dt \Big). \end{aligned}$$

Now

$$\sup_{1/n \leq t \leq 1} n^{1/2} t^{-1/2} |q_n(t) - t| = O_P((\log\log n)^{1/2}) \tag{3.43}$$

(Csörgő and Horváth (1993), Theorem 4.2.1) gives (3.40). Likewise, (3.41) is an immediate consequence of (3.42) and (3.43). □

Proof of Corollary 3.1. Define

$$\tilde{k}_n(r_n) := \min\Big\{ k \in \{1, \ldots, n\} \mid \max_{1 \leq i \leq k} i^{1/2} |b_{n,i} - b_{n,k}| > r_n \Big\}.$$

By the uniform convergence theorem for regularly varying functions we have

$$\begin{aligned} \max_{1 \leq i \leq k} i^{1/2} |b_{n,i} - b_{n,k}| &= |E(k/n)| k^{1/2} \max_{1 \leq i \leq k} \Big(\frac{i}{k}\Big)^{1/2} \Big| 1 - \frac{E(i/n)}{E(k/n)} \Big| \\ &\sim |E(k/n)| k^{1/2} \sup_{0 \leq t \leq 1} t^{1/2} (1 - t^\rho) \\ &= |E(k/n)| k^{1/2} (2\rho + 1)^{-(1+1/(2\rho))} 2\rho, \end{aligned} \tag{3.44}$$

provided $k/n \to 0$. Since $|E(t)|t^{1/2}$ is $(\rho+1/2)$–varying at 0 and hence $\sup_{s\le t}|E(s)|s^{1/2} \sim |E(t)|t^{1/2}$, it follows for all $\iota \in (0,1)$

$$\begin{aligned}
&\max_{1\le k\le(1\pm\iota)\kappa_n}\max_{1\le i\le k} i^{1/2}|b_{n,i}-b_{n,k}| \\
&\quad\sim \sup_{t\le(1\pm\iota)\kappa_n/n}|E(t)|t^{1/2}n^{1/2}(2\rho+1)^{-(1+1/(2\rho))}2\rho \;\sim\; (1\pm\iota)^{\rho+1/2}r_n,
\end{aligned}$$

so that for any positive sequence $u_n = o(r_n)$

$$\tilde{k}_n(r_n \pm u_n) \sim \kappa_n. \tag{3.45}$$

On the other hand, note that by the triangle inequality and the definition of $\tilde{k}_n$ we have

$$\begin{aligned}
&\max_{1\le k<\tilde{k}_n(r_n-u_n)}\max_{1\le i\le k} i^{1/2}|\hat\beta_{n,i}-\hat\beta_{n,k}| \\
&= \max_{1\le k<\tilde{k}_n(r_n-u_n)}\max_{1\le i\le k} i^{1/2}\Big|b_{n,i}-b_{n,k}+(\hat\beta_{n,i}-\beta-b_{n,i})-(\hat\beta_{n,k}-\beta-b_{n,k})\Big| \\
&\le \max_{1\le k<\tilde{k}_n(r_n-u_n)}\max_{1\le i\le k} i^{1/2}|b_{n,i}-b_{n,k}| + 2\max_{2\le i\le\tilde{k}_n(r_n-u_n)} i^{1/2}|\hat\beta_{n,i}-\beta-b_{n,i}| \\
&\le r_n - u_n + 2\max_{1\le i\le\tilde{k}_n(r_n-u_n)} i^{1/2}|\hat\beta_{n,i}-\beta-b_{n,i}|.
\end{aligned}$$

Hence, for $(\log\log n)^{1/2} = o(u_n)$, Lemma 3.4 implies

$$\begin{aligned}
P\Big\{\tilde{k}_n(r_n-u_n)\le\bar{k}_n(r_n)\Big\} &= P\Big\{\max_{1\le k<\tilde{k}_n(r_n-u_n)}\max_{1\le i\le k} i^{1/2}|\hat\beta_{n,i}-\hat\beta_{n,k}|\le r_n\Big\} \\
&\ge P\Big\{r_n-u_n+2\max_{1\le i\le\tilde{k}_n(r_n-u_n)} i^{1/2}|\hat\beta_{n,i}-\beta-b_{n,i}|\le r_n\Big\} \\
&\longrightarrow 1,
\end{aligned}\tag{3.46}$$

since $(\tilde{k}_n(r_n-u_n))_{n\in I\!N}$ is an intermediate sequence. Likewise,

$$\begin{aligned}
&P\Big\{\bar{k}_n(r_n)\le\tilde{k}_n(r_n+u_n)\Big\} \\
&\quad\ge P\Big\{\max_{1\le i\le\tilde{k}_n(r_n+u_n)} i^{1/2}|\hat\beta_{n,i}-\hat\beta_{n,\tilde{k}_n(r_n+u_n)}|>r_n\Big\} \\
&\quad\ge P\Big\{r_n+u_n-2\max_{1\le i\le\tilde{k}_n(r_n+u_n)} i^{1/2}|\hat\beta_{n,i}-\beta-b_{n,i}|>r_n\Big\} \\
&\quad\longrightarrow 1.
\end{aligned}\tag{3.47}$$

Combining (3.45)–(3.47), we arrive at the assertion. □

Proof of Theorem 3.1. In view of the proof of Corollary 3.1, we have

$$\max_{1\le i\le k_n} i^{1/2}|\hat\beta_{n,i}-\hat\beta_{n,k_n}| \sim |E(k_n/n)|k_n^{1/2}(2\rho+1)^{-(1+1/(2\rho))}2\rho$$

in probability if $k_n \sim \tau\bar{k}_n(r_n)$ for some $\tau\in(0,\infty)$. Since E is ρ–varying at 0, it follows that

$$\hat\rho^{(1)}_{n,\lambda} = \log\frac{|E(\lambda\bar{k}_n(r_n)/n)|\lambda^{1/2}(1+o_P(1))}{|E(\bar{k}_n(r_n)/n)|(1+o_P(1))}\Big/\log\lambda - 1/2 \;\longrightarrow\; \rho$$

in probability.

By the ρ–variation of E (or representation (3.11)), $b_{n,[\lambda^i \bar{k}_n(r_n)]} - b_{n,[\lambda^{i-1}\bar{k}_n(r_n)]}$, $i = 1, 2$, is of the same order as $E(\bar{k}_n(r_n)/n)$, which, by Corollary 3.1 and $(\log\log n)^{1/2} = o(r_n)$, is of larger order than $(\log\log n/\bar{k}_n(r_n))^{1/2}$. Hence Lemma 3.4 yields

$$\hat{\rho}^{(2)}_{n,\lambda} = \log \frac{b_{n,[\lambda^2\bar{k}_n(r_n)]} - b_{n,[\lambda\bar{k}_n(r_n)]}}{b_{n,[\lambda\bar{k}_n(r_n)]} - b_{n,\bar{k}_n(r_n)}} \Big/ \log\lambda + o_p(1) = \rho + o_p(1),$$

i.e., the consistency of $\hat{\rho}^{(2)}_{n,\lambda}$.

Next note that, in view of (3.11), (3.17) reads as

$$\frac{\beta|D|\rho C^{-\rho}(\kappa_n/n)^\rho \kappa_n^{1/2}}{(\rho+1)r_n} \longrightarrow \frac{(2\rho+1)^{1+1/(2\rho)}}{2\rho},$$

and consequently

$$\bar{k}_n(r_n) = (2\rho+1)^{1/\rho}\Big(\frac{C^\rho(\rho+1)}{2\beta|D|\rho^2} r_n n^\rho\Big)^{2/(2\rho+1)}(1+o_p(1)).$$

Hence (3.2) implies that

$$\begin{aligned}\Big(\frac{\bar{k}_n(r_n^\xi)}{(\bar{k}_n(r_n))^\xi}\Big)^{1/(1-\xi)} &= (2\rho+1)^{1/\rho}\Big(\frac{C^\rho(\rho+1)}{2\beta|D|\rho^2} n^\rho\Big)^{2/(2\rho+1)}(1+o_p(1)) \\ &= (2\rho+1)^{1/\rho}(2\beta^2\rho)^{-1/(2\rho+1)} k_n^{opt}(1+o_p(1)),\end{aligned}$$

so that obviously $\hat{k}_n^{opt}/k_n^{opt} \to 1$ in probability and the last assertion follows from Hall and Welsh (1985), Theorem 4.1. □

Proof of Corollary 3.2. First suppose $\mathrm{sgn}(E) = 1$. Then one obtains as in the proof of Corollary 3.1 that

$$\begin{aligned}&\max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}\Big(b_{n,i} - b_{n,\bar{k}_n(r_n)}\Big) \\ &= E\Big(\frac{\bar{k}_n(r_n)}{n}\Big)(\bar{k}_n(r_n))^{1/2} \max_{1\le i\le \bar{k}_n(r_n)} \Big(\frac{i}{\bar{k}_n(r_n)}\Big)^{1/2}\Big(\frac{E(i/n)}{E(\bar{k}_n(r_n)/n)} - 1\Big),\end{aligned}$$

where by the ρ–variation of E the maximum on the right-hand side converges to 0. Hence, by Corollary 3.1 and Lemma 3.4

$$\max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}\Big(\hat{\beta}_{n,i} - \hat{\beta}_{n,\bar{k}_n(r_n)}\Big) = o_P(r_n) = o_P\Big(\max_{1\le i\le \bar{k}_n(r_n)} i^{1/2}|\hat{\beta}_{n,i} - \hat{\beta}_{n,\bar{k}_n(r_n)}|\Big).$$

Combining this with the analogous calculations in the case $\mathrm{sgn}(E) = -1$, we obtain (3.19). Now the second assertion follows immediately from (3.15) and Theorem 3.1. □

Proof of Theorem 3.2. The consistency of $\hat{\rho}^{(1)}_{n,\lambda}$ and $\hat{\rho}^{(2)}_{n,\lambda}$ can be deduced from the ρ–variation of E in the same way as in the proof of Theorem 3.1.

Denote by $(\kappa_n)_{n\in\mathbb{N}}$ and $(\kappa_{n,\xi})_{n\in\mathbb{N}}$ (deterministic) sequences such that $\kappa_n/\bar{k}_n(r_n) \to 1$ and $\kappa_{n,\xi}/\bar{k}_n(r_n^\xi) \to 1$ in probability. To finish the proof, it suffices to verify that

$$\Big|E\Big(\Big(\frac{\kappa_{n,\xi}}{\kappa_n^\xi}\Big)^{1/(1-\xi)} n^{-1}\Big)\Big|\Big(\frac{\kappa_{n,\xi}}{\kappa_n^\xi}\Big)^{1/(2(1-\xi))} \longrightarrow \frac{(2\rho+1)^{1+1/(2\rho)}}{2\rho},$$

because then again by the ρ–variation of E it follows that $\hat{k}_n^{opt}$ satisfies (3.14) in probability. By (3.17) and Bingham et al. (1987), Theorem 1.5.12, there exists a function $R \in RV^0_{1/(\rho+1/2)}$, an 'asymptotic inverse' of $t \mapsto |E(t)|t^{1/2}$, such that

$$\frac{\kappa_n}{n} = R\Big(\frac{(2\rho+1)^{1+1/(2\rho)}}{2\rho} r_n n^{-1/2}\Big) \tag{3.48}$$

(cf. Remark 2.5(i)). Using first Corollary 3.1 and then (3.20), direct calculations lead to

$$\begin{aligned}\Big|E\Big(\Big(\frac{\kappa_{n,\xi}}{\kappa_n^\xi}\Big)^{1/(1-\xi)} n^{-1}\Big)\Big|\Big(\frac{\kappa_{n,\xi}}{\kappa_n^\xi}\Big)^{1/(2(1-\xi))} &\sim \frac{(2\rho+1)^{1+1/(2\rho)}}{2\rho}\times \\ \times \exp\Big(\frac{1}{1-\xi}\Big((1-\xi)\int_{(\kappa_{n,\xi}/\kappa_n^\xi)^{1/(1-\xi)}n^{-1}}^1 \frac{\eta(s)}{s}\,ds - \int_{\kappa_{n,\xi}/n}^1 &\frac{\eta(s)}{s}\,ds + \xi\int_{\kappa_n/n}^1 \frac{\eta(s)}{s}\,ds\Big)\Big).\end{aligned}$$

Thus it remains to prove that

$$\begin{aligned}&\Big|(1-\xi)\int_{(\kappa_{n,\xi}/\kappa_n^\xi)^{1/(1-\xi)}n^{-1}}^1 \frac{\eta(s)}{s}\,ds - \int_{\kappa_{n,\xi}/n}^1 \frac{\eta(s)}{s}\,ds + \xi\int_{\kappa_n/n}^1 \frac{\eta(s)}{s}\,ds\Big| \\ = &\ \Big|\int_{(\kappa_{n,\xi}/\kappa_n^\xi)^{1/(1-\xi)}n^{-1}}^{\kappa_{n,\xi}/n} \frac{\eta(s)}{s}\,ds + \xi\int_{\kappa_n/n}^{(\kappa_{n,\xi}/\kappa_n^\xi)^{1/(1-\xi)}n^{-1}} \frac{\eta(s)}{s}\,ds\Big| \\ \le &\ \int_{(\kappa_{n,\xi}/\kappa_n)^{\xi/(1-\xi)}}^1 \frac{|\eta(s\kappa_{n,\xi}/n)|}{s}\,ds + \xi\int_{(\kappa_{n,\xi}/\kappa_n)^{1/(1-\xi)}}^1 \frac{|\eta(s\kappa_n/n)|}{s}\,ds \\ \le &\ \frac{2\xi}{1-\xi}\sup_{0<t\le\kappa_n/n}|\eta(t)|\Big|\log\frac{\kappa_{n,\xi}}{\kappa_n}\Big|\end{aligned}$$

converges to 0. This, however, is immediate from (3.21), as by (3.48), the analogous equation for $\kappa_{n,\xi}$ and the Potter bounds (Bingham et al. (1987), Theorem 1.5.6) one has

$$\frac{\kappa_n}{n} = o\Big(\Big(\frac{r_n^2}{n}\Big)^{1/(2\rho+1)-\iota}\Big) \qquad \text{and} \qquad \log\frac{\kappa_{n,\xi}}{\kappa_n} = O(\log r_n)$$

for all $\iota > 0$. □

Proof of Corollary 3.3. The assertion follows in the same way as Corollary 3.1, because for $k/n \to 0$ one has

$$\begin{aligned}\max_{1\le i\le k} i^{1/2}|b_{n,i} - b_{n,k}| &= |g(k/n)|k^{1/2}\max_{1\le i\le k}\Big(\frac{i}{k}\Big)^{1/2}\Big|\frac{E(i/n)-E(k/n)}{g(k/n)}\Big| \\ &\sim |g(k/n)|k^{1/2}\sup_{0<t\le 1} t^{1/2}|\log t| \\ &= \frac{2}{\mathrm{e}}|g(k/n)|k^{1/2}.\end{aligned} \tag{3.49}$$

□

Proof of Theorem 3.3. The consistency of $\hat\rho^{(1)}_{n,\lambda}$ follows from (3.49) and the slow variation of g by similar arguments as given in the proof of Theorem 3.1. Check that by Corollary 3.3 and Lemma 3.4

$$\begin{aligned}
&(\log\log n)^{1/2} = o(r_n) = o\Big(g\Big(\frac{\bar k_n(r_n)}{n}\Big)(\bar k_n(r_n))^{1/2}\Big)\\
\Longrightarrow\ & \hat\beta_{n,[\lambda^i \bar k_n(r_n)]} - \hat\beta_{n,[\lambda^{i-1}\bar k_n(r_n)]} = -g\Big(\frac{\bar k_n(r_n)}{n}\Big)\log\lambda(1+o_p(1)) \qquad (i=1,2)\\
\Longrightarrow\ & \hat\rho^{(2)}_{n,\lambda} \longrightarrow 0
\end{aligned}$$

in probability.

Denote by κ_n and $\kappa_{n,\xi}$ sequences satisfying (3.22) for $\bar k_n(r_n)$ and $\bar k_n(r_n^\xi)$, respectively. Similarly as in the proof of Theorem 3.2 one obtains

$$\frac{\kappa_n}{n} \sim R\Big(\Big(\frac{\mathrm{e}}{2}\Big)^2\frac{r_n^2}{n}\Big) \qquad \text{and} \qquad \frac{\kappa_{n,\xi}}{n} \sim R\Big(\Big(\frac{\mathrm{e}}{2}\Big)^2\frac{r_n^{2\xi}}{n}\Big) \tag{3.50}$$

for some $R\in RV_1^0$, which in combination with Corollary 1.2.1 1. of de Haan (1975) shows that $\kappa_n,\kappa_{n,\xi} = o(n^\iota)$ for all $\iota>0$ if $\log r_n = o(\log n)$. Moreover, the Potter bounds (Bingham et al. (1987), Theorem 1.5.6) imply $\log(\kappa_{n,\xi}/\kappa_n) = O(\log r_n)$. In view of (3.22), (3.24) and (3.23), straightforward calculations show that

$$\begin{aligned}
\Big|E\Big(\frac{\hat k^{opt}_{n,\vartheta}}{n}\Big)\Big|(\hat k^{opt}_{n,\vartheta})^{1/2} \ &\sim\ \frac{\mathrm{e}}{2}\vartheta^{1/2}\Big|\frac{(\eta(\kappa_n/n))^\xi}{\eta(\kappa_{n,\xi}/n)}\Big|^{1/(1-\xi)}\times\\
&\times\ \exp\Big(\frac{1}{1-\xi}\Big(\int_{(\kappa_{n,\xi}/\kappa_n)^{\xi/(1-\xi)}}^1 \frac{\eta(s\kappa_{n,\xi}/n)}{s}\,ds - \xi\int_{(\kappa_{n,\xi}/\kappa_n)^{1/(1-\xi)}}^1 \frac{\eta(s\kappa_n/n)}{s}\,ds\Big)\Big)
\end{aligned}$$

in probability, where, according to condition (i), the exponent is bounded. Since

$$\Big|\frac{\eta(\kappa_n/n)^\xi}{\eta(\kappa_{n,\xi}/n)}\Big|^{1/(1-\xi)} = \Big|\frac{\eta(\kappa_n/n)}{\eta(\kappa_{n,\xi}/n)}\Big|^{1/(1-\xi)}\Big/|\eta(\kappa_n/n)| \ \longrightarrow\ \infty$$

because of $\kappa_n > \kappa_{n,\xi}$ and $|\eta(t)|\downarrow 0$ as $t\downarrow 0$, it follows that

$$\Big|E\Big(\frac{\hat k^{opt}_{n,\vartheta}}{n}\Big)\Big|(\hat k^{opt}_{n,\vartheta})^{1/2} \longrightarrow \infty$$

and hence $k_n^{(1)}/\hat k^{opt}_{n,\vartheta}\to 0$ in probability.

To verify the second part of (3.16), choose a sequence $(\tilde k_n)_{n\in I\!N}$ such that

$$\Big|g\Big(\frac{\tilde k_n}{n}\Big)\Big|\tilde k_n^{1/2} \ \longrightarrow\ 1. \tag{3.51}$$

Then, on the analogy of (3.50), $\tilde k_n/n\sim R(1/n)$ and again the Potter bounds yield

$$\log\frac{\hat k^{opt}_{n,\vartheta}}{\tilde k_n} = \log\Big(\Big(\frac{R(r_n^{2\xi}/n)}{R(1/n)}\Big)^{1/(1-\xi)}\Big(\frac{R(r_n^2/n)}{R(1/n)}\Big)^{-\xi/(1-\xi)}\Big) + O_P(1) = o_P(\log r_n). \tag{3.52}$$

On the other hand,
$$\Big|g\Big(\frac{k_n^{(1)}}{n}\Big)\Big|(k_n^{(1)})^{1/2} \ \sim\ \Big|\eta\Big(\frac{k_n^{(1)}}{n}\Big)\Big|,$$
and thus by (3.51) and similar arguments as above
$$\frac{k_n^{(1)}}{\tilde{k}_n} \sim \frac{R(|\eta(k_n^{(1)}/n)|^2/n)}{R(1/n)} = o\Big(\Big|\eta\Big(\frac{k_n^{(1)}}{n}\Big)\Big|^{2-\iota}\Big) \tag{3.53}$$
for all $\iota > 0$. Finally, a combination of (3.52) and (3.53) leads to
$$\begin{aligned}\Big|\log\frac{E(\hat{k}^{opt}_{n,\vartheta}/n)}{E(k_n^{(1)}/n)}\Big| &= \Big|\int_{\hat{k}^{opt}_{n,\vartheta}/n}^{k_n^{(1)}/n}\frac{\eta(s)}{s}\,ds\Big| + o_P(1)\\ &\le \sup_{0<t\le n^{\iota-1}}|\eta(t)|\Big|\log\frac{k_n^{(1)}}{\hat{k}^{opt}_{n,\vartheta}}\Big| + o_P(1)\\ &= |\eta(n^{\iota-1})|\Big(o_P(\log r_n) + O_P\Big(\log|\eta(n^{-1})|\Big)\Big)\\ &\longrightarrow 0\end{aligned}$$
in probability by the conditions (i) and (ii). Here the monotonicity of $|\eta|$ has been utilized for the last equation. Now the second convergence condition of (3.16) is obvious and the last assertion of the theorem follows from Lemma 3.3. □

Proof of Lemma 3.5. The arguments given in the proof of Lemma 3.3 show that it is sufficient to consider sequences $k_n \sim \lambda k_n^{(1)}$ and then to minimize
$$\Big(\Phi(k_n/n)\mu_{T,\beta,\Psi}\Big)^2 + k_n^{-1}\sigma^2_{T,\beta} \ \sim\ (\lambda^{2\delta}\mu^2_{T,\beta,\Psi} + \lambda^{-1}\sigma^2_{T,\beta})/k_n^{(1)},$$
so that the proof of the first two assertions can be completed by standard calculations. The weak convergence result (3.27) is immediate from Theorem 2.3. □

Proof of Lemma 3.6. Following the lines of the proof of Theorem 2.1, it is not difficult to see that almost surely
$$\sup_{t\in(0,1]} t^\beta h(t)\Big|\frac{\tilde{Q}_{n,i}(t)-\tilde{D}_{n,i}}{a(i/n)} - \frac{t^{-\beta}-1}{\beta} - \Phi(i/n)\Psi(t)\Big| = O\Big((\log\log(3i)/i)^{1/2}\Big) + o(\Phi(i/n)) \tag{3.54}$$
uniformly for $1 \le i \le k_n$, where
$$\tilde{D}_{n,i} = \begin{cases} F^{-1}(1-i/n) & \text{if } \beta \ge -1/2,\\ F^{-1}(1-S_1/n) + a(i/n)\Big(\dfrac{1}{\beta} + \dfrac{c_\Psi}{\beta-\delta}\Phi(i/n)\cdot 1_{\{\delta>0\}}\Big) & \text{if } \beta < -1/2.\end{cases}$$
For this end, first note that by Lemma 2.1 for $\eta := 0$ if $\beta \ge -1/2$ and $\eta \in (0,1)$ if $\beta < -1/2$
$$\sup_{t\in(\eta/i,1]} t^{\beta+1/2}\Big|\frac{\tilde{Q}_{n,i}(t)-F^{-1}(1-i/n)}{a(i/n)} - \frac{(S_{[it]+1}/i)^{-\beta}-1}{\beta} - \Phi(i/n)\Psi(S_{[it]+1}/i)\Big| = o(\Phi(i/n))$$
almost surely uniformly for $1 \le i \le k_n$. Then one can use the following bound instead of Proposition 2.1.

Proposition 3.1

$$\sup_{i\in I\!N}\Big(i/\log\log(3i)\Big)^{1/2}\sup_{t\in(\eta/i,1]}h(t)t^{\beta}\Big|\frac{(S_{[it]+1}/i)^{-\beta}-1}{\beta}-\frac{t^{-\beta}-1}{\beta}\Big|<\infty. \qquad (3.55)$$

Proof. We restrict ourselves to the case $\beta\neq 0$, in which the left-hand side of (3.55) is bounded by

$$\begin{aligned}
&\frac{1}{\beta}\Big(\sup_{i\in I\!N}\Big(i/\log\log(3i)\Big)^{1/2}\max_{1\le j\le i}\sup_{t\in I(j,i)}h(t)\Big(\frac{it}{j}\Big)^{\beta}\Big|\Big(\frac{S_j}{j}\Big)^{-\beta}-1\Big|\\
&\qquad+\sup_{i\in I\!N}\Big(i/\log\log(3i)\Big)^{1/2}\max_{1\le j\le i}\sup_{t\in I(j,i)}h(t)t^{\beta}\Big|\Big(\frac{j}{i}\Big)^{-\beta}-t^{-\beta}\Big|\Big)\\
=:&\ \frac{1}{\beta}(T_1+T_2),
\end{aligned}$$

where as in the proof of Theorem 2.1 we define $I(j,i):=[(j-1)/i,j/i)\cap(\eta/i,1]$.

By the law of iterated logarithm and a Taylor expansion of the function $x\mapsto x^{-\beta}$ one obtains that $(j/\log\log(3j))^{1/2}|(S_j/j)^{-\beta}-1|$ is bounded almost surely. Hence by the definition of η one has (cf. (2.44))

$$T_1\le \text{const.}\sup_{i\in I\!N}\max_{1\le j\le i}\sup_{t\in I(j,i)}h(t)t^{-1/2}(it/j)^{\beta+1/2}<\infty.$$

The finiteness of T_2 was already verified in the proof of Proposition 2.1. □

Now Proposition 2.2 (with k replaced by i) completes the proof in the case $\beta\ge -1/2$. In the case of $\beta<-1/2$ the arguments used in the proof of Proposition 2.3 give the desired result.

Next note that (3.54) can be rewritten as

$$\Big\|\frac{\tilde Q_{n,i}-D^*_{n,i}}{a(i/n)}-\bar F_\beta^{-1}-\Phi(i/n)\bar\Psi\Big\|_{\beta,h}=O\Big((\log\log(3i)/i)^{1/2}\Big)+o(\Phi(i/n))\qquad\text{a.s.}$$

uniformly for $1\le i\le k_n$ where $D^*_{n,i}=\tilde D_{n,i}-a(i/n)(1_{\{\beta\neq 0\}}/\beta+c_\Psi/(\beta-\delta)\Phi(i/n)1_{\{\beta\neq\delta>0\}})$. In particular, for every $i_0\in I\!N$ the left-hand side is bounded uniformly for $1\le i\le i_0$. Hence Condition 2.3 and the boundedness assumption of Condition 3.1 imply

$$\max_{1\le i\le i_0} i^{1/2}\Big|\tilde\beta_{n,i}-\beta-\Phi(i/n)T'_\beta(\bar\Psi)\Big|=O(1)\qquad\text{a.s.,}$$

so that a standard diagonal proof (cf. Remark 2.5(ii)) gives the existence of a sequence $i_n\to\infty$, $1\le i_n\le k_n$, such that

$$\max_{1\le i\le i_n} i^{1/2}\Big|\tilde\beta_{n,i}-\beta-\Phi(i/n)T'_\beta(\bar\Psi)\Big|=o((\log\log k_n)^{1/2})\qquad\text{a.s.}\qquad(3.56)$$

On the other hand, Condition 2.3 and the Fréchet differentiability of $T_{|D_{\beta,h}}$ at $\bar F_\beta^{-1}$ yield

$$\Big|\tilde\beta_{n,i}-\beta-\Phi(i/n)T'_\beta(\bar\Psi)\Big|=O\Big((\log\log(3i)/i)^{1/2}\Big)+o(\Phi(i/n))\qquad\text{a.s.}\qquad(3.57)$$

uniformly for $i_n \le i \le k_n$. A combination of (3.56) and (3.57) proves the assertion. $\square$

Proof of Corollary 3.4. Define

$$\tilde{k}_n(r_n) := \min\Big\{k \in \{1,\dots,n\} \mid \max_{1\le i\le k} i^{1/2}|b_{n,i} - b_{n,k}| > r_n\Big\}$$

and check in the same way as in the proof of Corollary 3.1 that

$$\max_{1\le i\le k} i^{1/2}|b_{n,i} - b_{n,k}| \sim (2\delta+1)^{-(1+1/(2\delta))} 2\delta |\mu_{T,\beta,\Psi}| \Phi(k/n) k^{1/2}. \tag{3.58}$$

Because of $\max_{1\le i\le k} i^{1/2}|rem_{n,i}| = o(\max_{1\le i\le k} i^{1/2}\Phi(i/n)) = o(\max_{1\le i\le k} i^{1/2}|b_{n,i} - b_{n,k}|)$, Lemma 3.6 implies

$$\max_{1\le i\le k} i^{1/2}|\tilde{\beta}_{n,i} - \tilde{\beta}_{n,k}| = \max_{1\le i\le k} i^{1/2}|b_{n,i} - b_{n,k}|(1 + o_P(1)) + O\Big((\log\log k)^{1/2}\Big). \tag{3.59}$$

Consequently, in view of $(\log\log \bar{k}_n(r_n))^{1/2} = o(r_n)$ and (3.29) one has $\bar{k}_n(r_n) = \tilde{k}_n(r_n)(1 + o_P(1))$ and the assertion is obvious from (3.58). $\square$

Proof of Lemma 3.7. According to de Haan and Stadtmüller (1996), Theorem 2, Condition 2.1(ii) is satisfied for some function Φ that is δ–varying at 0, $\delta > 0$, if and only if there are constants $b \in I\!R$ and $c > 0$ such that $g(t) := F^{-1}(1-t) - (b + c\bar{F}_\beta^{-1}(t))$ is $(\delta - \beta)$–varying at 0 if $\delta \ne \beta$ and Π–varying at 0 with auxiliary function L, say, if $\delta = \beta$.

Direct calculations show that then one may choose

$$a(t) = ct^{-\beta}, \qquad \Phi(t) = t^\beta \begin{cases} |g(t)| & \text{if } \delta \ne \beta, \\ |L(t)| & \text{if } \delta = \beta, \end{cases}$$

and Ψ according to (2.4) (see proof of Lemma 2.1). Hence $\Phi(t) \sim \text{const.}t^\delta$ is equivalent to $g(t) \sim \text{const.}t^{\delta-\beta}$ in case of $\beta \ne \delta$, and thus the assertion in this case.

If $\beta = \delta$, then $\Phi(t) \sim \text{const.}t^\delta$ is equivalent to the convergence of $|L|$ to a positive constant, which, according to the representation theorem 3.6.6 of Bingham et al. (1987), in turn is equivalent to $g(t) \sim -d\log t$ for some $d \in I\!R$, i.e., the assertion. $\square$

Proof of Theorem 3.4. By Lemma 3.7 one may assume w.l.o.g. that $\Phi(t) \sim t^\delta$. Hence, in the same way as in the proof of Theorem 3.1 assertion (i) follows from (3.58) and (3.59).

To verify (ii), check that according to Corollary 3.4

$$\bar{k}_n(r_n) = (2\delta+1)^{1/\delta}(r_n n^\delta/(2\delta|\mu_{T,\beta,\Psi}|))^{2/(2\delta+1)}(1 + o_P(1)),$$

so that

$$\begin{aligned} \left(\frac{\bar{k}_n(r_n^\xi)}{(\bar{k}_n(r_n))^\xi}\right)^{1/(1-\xi)} &= (2\delta+1)^{1/\delta}(n^\delta/(2\delta|\mu_{T,\beta,\Psi}|))^{2/(2\delta+1)}(1 + o_P(1)) \\ &= (2\delta+1)^{1/\delta}(2\delta\sigma^2_{T,\beta})^{-1/(2\delta+1)} k_n^{opt}(1 + o_P(1)). \end{aligned}$$

Now the continuity of $\beta \mapsto \sigma^2_{T,\beta}$ in combination with Theorem 2.3 yields the assertion. $\square$

The *proofs of Corollary 3.5* and *Theorem 3.5* follow in exactly the same way as the proofs of Theorem 3.2 and Corollary 3.2, respectively, and are therefore omitted.

4 Best attainable rates of convergence

In Chapter 2 we have determined estimators of the extreme value index that are optimal in a certain sense within given classes of statistical tail functionals. A natural extension of this problem is the question how accurate the extreme value index can be estimated by an *arbitrary* sequence of estimators, and which estimator performs best.

A first partial answer to this question was given by Hall and Welsh (1984), whose main theorem is reformulated here in terms of the index β of regular variation of the upper tail of the distribution (instead of the lower tail as in the original paper). Define sets $\mathcal{D} = \mathcal{D}(\beta_0, d_0, \vartheta, \rho, A)$ consisting of all densities $f : (0, \infty) \to [0, \infty)$ that satisfy

$$f(x) = \frac{d}{\beta} x^{-(1/\beta+1)}(1 + r(x)) \quad \text{where} \quad |r(x)| \le A x^{-\rho/\beta} \tag{4.1}$$

for all $x > 0$, $|\beta - \beta_0| \le \vartheta$, $|d - d_0| \le \vartheta$, $\beta_0, d_0, \rho, A > 0$ and $0 < \vartheta < \min(\beta_0, d_0)$. Then for all sequences of estimators $\hat\beta_n$ based on an i.i.d. sample X_i, $1 \le i \le n$, of size n and all real sequences $(a_n)_{n\in\mathbb{N}}$

$$\lim_{n\to\infty} \inf_{f\in\mathcal{D}} P_f\{|\hat\beta_n - \beta| \le a_n\} = 1$$

implies

$$\lim_{n\to\infty} a_n\, n^{\rho/(2\rho+1)} = \infty. \tag{4.2}$$

Here P_f indicates that the random variables X_i have density f. For notational convenience, the dependence of β on f is suppressed.

By a simple indirect proof it is easily seen that this implication can be restated in the following way: For all $a \in (0, \infty)$ and any sequence of estimators $\hat\beta_n$ one has

$$\limsup_{n\to\infty} \inf_{f\in\mathcal{D}} P_f\{|\hat\beta_n - \beta| \le a n^{-\rho/(2\rho+1)}\} < 1. \tag{4.3}$$

Donoho and Liu (1987, 1991) established upper bounds for the left-hand side of (4.3).

Moreover, Hall and Welsh showed that the rate $n^{-\rho/(2\rho+1)}$ cannot be improved, since, for a suitable choice of the number of order statistics involved in estimation, the Hill estimator $\hat\beta^H_{n,k_n}$ defined in (3.1) (see also Example 2.1(ii)) actually converges at this rate in the following sense:

$$\lim_{a\to\infty} \liminf_{n\to\infty} \inf_{f\in\mathcal{D}} P_f\{|\hat\beta^H_{n,k_n} - \beta| \le a n^{-\rho/(2\rho+1)}\} = 1. \tag{4.4}$$

From a different point of view, the uniform rate of convergence of an estimator can be regarded as a rough measure of its robustness against deviations of type (4.1) from the ideal Pareto distributions. However, model (4.1) is rather restrictive, as it includes only distributions with positive extreme value index whose upper tail is fitted very well by a Pareto tail. In particular, for those d.f.'s the rate of convergence in (1.1) w.r.t. the variational distance is algebraic (Reiss (1989), Corollary 5.2.7), whereas, e.g., for loggamma d.f.'s with density $f(x) = cx^{-(1/\beta+1)}\log^{\gamma-1}(x)1_{(1,\infty)}(x)$, which are common in non-life insurance mathematics (cf. Ramlau-Hansen (1988) and Hogg and Klugman

(1984)), or the loghyperbolic distribution often used in geology and other natural sciences (see Beirlant et al. (1996a)) the rate of convergence is much slower, namely, a power of the logarithm of the sample size. Such a behavior is typical if (3.6) holds with ε being slowly varying at 0 (cf. de Haan and Resnick (1996), Theorem 3.1). Consequently, one should expect a poor rate of convergence for estimates of the extreme value index if the underlying d.f. is of this type. In order to safeguard oneself against large estimation errors in this "worst case", it is particularly important that an estimator converges at the best possible rate for such d.f.'s. To this end, we aim at generalizing Hall's and Welsh's result to enlarged sets of distributions including quite arbitrary d.f.'s in the domain of attraction of an extreme value d.f., where our attention is mainly directed to d.f.'s with a poor rate of convergence in (1.1).

Note that there is no obvious counterpart of condition (4.1) if the density is not tail equivalent to a Pareto density. This is partially due to the fact that in (4.1) the bound on the remainder term $r(x)$ depends both on the first-order parameter β and the second-order parameter ρ. For that reason, in case $\beta > 0$ we replace (4.1) by a suitable condition in terms of the q.f. F^{-1}.

Starting from Condition 2.1 or Lemma 2.1, one may consider conditions of the type

$$\sup_{x\in[x_0,x_1]} \left| \frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} - F_\beta^{-1}(x) \right| \le g(t)$$

or

$$\sup_{x\in(0,x_1]} x^{\beta+\iota} \left| \frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} - F_\beta^{-1}(x) \right| \le g(t),$$

yet for this one must arbitrarily choose the range of x-values the supremum is taken over and, in the latter case, the constant $\iota > 0$, too. Therefore, instead we require that the q.f. is normalized regularly varying, i.e., it admits representation (3.6):

$$F^{-1}(1-t) = ct^{-\beta} \exp\Big(\int_t^1 \varepsilon(s)/s \, ds \Big), \qquad \varepsilon(t) \to 0 \text{ as } t \downarrow 0. \tag{3.6}$$

Recall that (3.6) is equivalent to the von Mises condition $h_1(x) := (1-F(x))/(xf(x)) \to \beta$ as $x \to \infty$. Moreover, for sufficiently small t, $\varepsilon(t) = t(F^{-1})'(1-t)/F^{-1}(1-t) - \beta = h_1(F^{-1}(1-t)) - \beta$ measures the speed of convergence in Karamata's theorem applied to the regularly varying functions $(F^{-1})'(1-\cdot)$ or f. (Here $(F^{-1})'$ denotes a suitable Lebesgue density of the absolutely continuous function F^{-1}.)

In Theorem 4.1 we establish a lower bound on the rate of convergence of an arbitrary sequence of estimators for positive β uniformly over all d.f.'s satisfying (3.6) with $|\varepsilon(t)| \le g(t)$ where g is ρ–varying at 0 for some $\rho \ge 0$. In the special case of $g(t) = \text{const.}t^{-\rho}$, $\rho > 0$, this is essentially equivalent to Hall's and Welsh's result (see Lemma 4.1), whereas for slowly varying functions g we do not obtain a bound merely on the *rate* of convergence but an asymptotically sharp "in probability bound" on the estimation error (cf. the discussion below Example 4.1). Hence, the latter case leads to a criterion to discriminate the robustness of different estimators for β against deviations from a Pareto tail which is more accurate than the one following from Hall's and Welsh's theorem: While most prominent estimators of β converge at the optimal rate (analogously to (4.4)) if a suitable fraction of the observations is used for estimation, for example, the Pickands estimator defined in Example 2.1(i) does not attain the asymptotic bound in case of $\rho = 0$, but the Hill estimator does.

In the general case considered in Section 2, where the sign of β is not known in advance, there is no simple unifying representation of the q.f. comparable with (3.6), but the following analog for its derivative is sufficient for $F \in D(G_\beta)$ if $\varepsilon(t) \to 0$ as $t \downarrow 0$:

$$(F^{-1})'(1-t) = ct^{-(\beta+1)} \exp\Big(\int_t^1 \varepsilon(s)/s\, ds \Big) \tag{4.5}$$

or, equivalently,

$$f(F^{-1}(1-t)) = \frac{1}{c} t^{\beta+1} \exp\Big(-\int_t^1 \varepsilon(s)/s\, ds \Big). \tag{4.6}$$

Note that (4.5) is equivalent to the von Mises type condition $h_2(x) := ((1-F)/f)'(x) \to \beta$ as $x \uparrow F^{-1}(1)$ (cf. Reiss (1989), (5.1.25)). Here, for sufficiently small t, $\varepsilon(t) = t(F^{-1})''(1-t)/(F^{-1})'(1-t) - (\beta+1) = h_2(F^{-1}(1-t)) - \beta$ quantifies the speed of Karamata's convergence for $(F^{-1})''(1-\cdot)$ and of the convergence of h_2.

Theorem 4.4 is the counterpart of Theorem 4.1 in this general setup where representation (3.6) is replaced by (4.5). While the bounds are literally the same, in contrast to the case $\beta > 0$ estimators of the extreme value index that are not location invariant, like, e.g., the moment estimator considered in Example 2.1(iii), in general do *not* converge at the optimal rate in model (4.5). Therefore, again in turns out that in the general situation location-invariant estimators are preferable. All proofs are given in Section 3.

4.1 Optimal rates of convergence in the case $\beta > 0$

First we establish a lower bound on the rate of convergence in model (3.6) uniformly on sets of d.f.'s of the type

$$\mathcal{F}_1 := \mathcal{F}_1(\beta_0, c, \vartheta, g) := \Big\{ F \text{ d.f.} \mid F^{-1} \text{ satisfies (3.6) with } |\beta - \beta_0| \le \vartheta, \\ |\varepsilon(t)| \le g(t) \text{ for all } t \in (0,1) \Big\} \tag{4.7}$$

where $\beta_0 > \vartheta > 0$, $c > 0$ and g is a suitable function that is ρ–varying at 0 for some $\rho \ge 0$. The bound will be given in terms of a sequence $(t_n)_{n \in I\!N}$ satisfying

$$\lim_{n\to\infty} g(t_n)(nt_n)^{1/2} = 1. \tag{4.8}$$

Recall that under the conditions of Theorem 4.1 the arguments of Remark 2.5(i) ensure the existence and the asymptotic uniqueness of such a sequence.

Theorem 4.1 [Drees (1998c), Theorem 2.1] *Fix constants $c, \beta_0 > 0$ and $\vartheta \in (0, \beta_0)$. Suppose that the function $g : (0,1) \to (0,\infty)$ is ρ–varying at 0 for some $\rho \ge 0$, bounded away from 0 and finitely integrable on compact subintervals of $(0,1]$, and non-decreasing* *in a neighborhood of 0 with* $\lim_{t\downarrow 0} g(t) = 0$. *Then for an arbitrary sequence* $(\hat\beta_n)_{n\in\mathbb{N}}$ *of estimators for* β *one has*

$$\limsup_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta_n - \beta| \le a g(t_n)\} < 1 \tag{4.9}$$

for all

$$a < a^* := \begin{cases} \infty & \\ & \textit{if} \\ 1 & \end{cases} \begin{matrix} \rho > 0, \\ \\ \rho = 0. \end{matrix} \tag{4.10}$$

(Here P_F *indicates that* F *is the underlying d.f.)*

Remarks 4.1 (i) An equivalent formulation of the assertion of Theorem 4.1 is

$$\begin{aligned} &\lim_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta_n - \beta| \le a_n\} = 1 \\ &\implies \liminf_{n\to\infty} \frac{a_n}{g(t_n)} \ge a^* \end{aligned} \tag{4.11}$$

for all real sequences $(a_n)_{n\in\mathbb{N}}$.

(ii) Note that the parameter ρ has a meaning different from the one in Chapter 3, since it describes the exponent of regular variation of the upper bound on ε, while the functions ε or E, defined by (3.9), need not be regularly varying.

(iii) The proof of Theorem 4.1 shows that the left-hand side of (4.9) is bounded by $1 - \exp(-2a^{2+1/\rho}(\beta_0+\vartheta)^{-2})/4$ for $\rho > 0$ and by $3/4$ if $\rho = 0$. We will not consider this very crude bound further. □

Example 4.1 Let $g(t) = A/(1-\log t)$ for some sufficiently large constant A. Then (3.6) with $|\varepsilon(t)| \le g(t)$ is satisfied for large t by loggamma distributions. Since $(nt_n)^{1/2}/|\log t_n| \to 1$ implies $|\log t_n| \sim \log n$, the optimal uniform bound $g(t_n)$ asymptotically behaves as $A/\log n$. In particular, one should not expect an estimator of β to converge at a faster rate than $1/\log n$ if the observations are loggamma distributed.

Note that it is not necessary to determine the asymptotic behavior of the sequence $(t_n)_{n\in\mathbb{N}}$ explicitly, because only the sequence $(g(t_n))_{n\in\mathbb{N}}$ enters the bound in (4.9). □

While, analogously to Hall's and Welsh's result, for $\rho > 0$ Theorem 4.1 merely gives a lower bound for the *rate* of convergence, in case of $\rho = 0$ for any sequence $(a_n)_{n\in\mathbb{N}}$ that possesses a subsequence asymptotically smaller than the "boundary sequence" $b_n = g(t_n)$, i.e., $\liminf_{n\to\infty} a_n/b_n < 1$, one has $\liminf_{n\to\infty} \inf_{F\in\mathcal{F}_1} P_F\{|\hat\beta_n - \beta| \le a_n\} < 1$. Next we show that this "in probability bound" is asymptotically sharp, as it is attained by the Hill estimator in the following sense: We say that a sequence of estimators $\hat\beta_n$ *attains the optimal rate of convergence for* $\rho > 0$ if $\lim_{a\to\infty} \liminf_{n\to\infty} \inf_{F\in\mathcal{F}_1} P_F\{|\hat\beta_n - \beta| \le a g(t_n)\} = 1$; and it *attains the optimal asymptotic bound for* $\rho = 0$ if $\lim_{n\to\infty} \inf_{F\in\mathcal{F}_1} P_F\{|\hat\beta_n - \beta| \le a g(t_n)\} = 1$ for all $a > 1$. Note that in the latter case a standard diagonal procedure proves the existence of a sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n/g(t_n) \to 1$ and $\lim_{n\to\infty} \inf_{F\in\mathcal{F}_1} P_F\{|\hat\beta_n - \beta| \le a_n\} = 1$ (cf. Remark 2.5 (ii)).

Notice that, according to the Theorems 4.1 and 4.2 (see below), in the case $\rho > 0$ there is no sharp boundary sequence $(b_n)_{n\in I\!N}$, such that $\liminf_{n\to\infty} a_n/b_n < 1$ implies $\liminf_{n\to\infty} \inf_{F\in\mathcal{F}_1} P_F\{|\hat\beta_n - \beta| \le a_n\} < 1$ for all sequence of estimators $\hat\beta_n$, and for all $(a_n)_{n\in I\!N}$ with $\liminf_{n\to\infty} a_n/b_n > 1$ there exist estimators $\hat\beta_n$ with $\lim_{n\to\infty} \inf_{F\in\mathcal{F}_1} P_F\{|\hat\beta_n - \beta| \le a_n\} = 1$. This difference can be explained by the fact that for $\rho > 0$ it is optimal to balance the asymptotic variance and the asymptotic bias, while for $\rho = 0$ Corollary 2.2 shows that under certain second-order conditions it is better to use a larger number of observations for estimation, such that the asymptotic squared bias is of larger order than the asymptotic variance. Thus in this case the random error part is negligible and the limit distribution is degenerated (if it exists).

It turns out that this holds true in the present situation, too. To be more concrete, let $(t_n^*)_{n\in I\!N}$ be a sequence such that

$$\lim_{n\to\infty} \frac{t_n^*}{t_n} = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{g(t_n^*)}{g(t_n)} = 1 \tag{4.12}$$

(which exists according to Remark 2.5(ii)) and define

$$k_n := \begin{cases} [nt_n] & \quad \rho > 0, \\ & \text{if} \\ [nt_n^*] & \quad \rho = 0, \end{cases} \tag{4.13}$$

where again $[x]$ denotes the integral part of x. Then we have the following result for the Hill estimator $\hat\beta^H_{n,k_n}$, which is a direct consequence of a uniform version of (3.7).

Theorem 4.2 [Drees (1998c), Theorem 2.2 and its proof] *Suppose that the conditions of Theorem 4.1 are fulfilled. Then for $a, \rho > 0$ and k_n defined by (4.13)*

$$\begin{aligned} \lim_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta^H_{n,k_n} - \beta| \le a g(t_n)\} \quad &= \quad \inf_{|\beta-\beta_0|\le\vartheta} \mathcal{N}(1/(\rho+1), \beta^2)\,[-a,a] \\ &\longrightarrow \quad 1 \qquad \text{as } a \to \infty, \end{aligned} \tag{4.14}$$

and in case of $\rho = 0$

$$\lim_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta^H_{n,k_n} - \beta| \le a_n\} = 1 \tag{4.15}$$

for all sequences $(a_n)_{n\in I\!N}$ satisfying

$$\lim_{n\to\infty} (nt_n^*)^{1/2}(a_n - g(t_n^*)) = \infty. \tag{4.16}$$

In particular, (4.15) holds true for $a_n = a g(t_n)$ with $a > 1$ and hence also for a suitable sequence $(a_n)_{n\in I\!N}$ such that $a_n/g(t_n) \to 1$.

An analogous result can be proven for all kernel type estimators considered by Csörgő et al. (1985) if the support of the kernel K is a subset of the unit interval; then the normal distribution in (4.14) must be replaced by $\mathcal{N}(\int_0^1 t^\rho K(t)\,dt, \beta^2 \int_0^1 K^2(t)\,dt)$. On the other hand, while most usual estimators of the extreme value index converge at the optimal rate if the sample fraction involved in estimation is chosen appropriately, often the asymptotic bound in the case $\rho = 0$ is not attained. In particular, this holds true for scale- and location-invariant statistical tail functionals satisfying the following differentiability condition.

Condition 4.1 *The restriction $T_{|D_{\beta,h}}$ of the functional $T : span(1, \cup_{|\beta-\beta_0|\le\vartheta} D_{\beta,h}) \to I\!R$ is Fréchet differentiable at $\bar{F}_\beta^{-1}$ with derivative T'_β uniformly for all $\beta \in [\beta_0 - \vartheta, \beta_0 + \vartheta]$. Moreover, the pertaining measures $\nu_{T,\beta}$ (cf. (2.22)) fulfill*

$$\sup_{|\beta-\beta_0|\le\vartheta} \int_{[0,1]} \frac{1}{t^\beta h(t)} |\nu_{T,\beta}|(dt) < \infty. \tag{4.17}$$

Theorem 4.3 [cf. Drees (1998c), Theorem 2.3 and the following discussion] *Denote by Q_{n,k_n} the empirical tail q.f. based on the largest $k_n + 1$ order statistics (see (1.4)) with k_n defined by (4.13). Assume that the functional T satisfies Condition 4.1 and for all $\beta \in [\beta_0 - \vartheta, \beta_0 + \vartheta]$ the restriction $T_{|span(D_{\beta,h},1)}$ fulfills Condition 2.3. Then under the conditions of Theorem 4.1*

$$\begin{aligned}&\liminf_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|T(Q_{n,k_n}) - \beta| \le ag(t_n)\}\\ &= \inf_{|\beta-\beta_0|\le\vartheta} \mathcal{N}\Big(\int_0^1 s^{\rho-1}\Big|\int_{[0,s]} \bar{F}_\beta^{-1} d\nu_{T,\beta}\Big|\, ds, \sigma^2_{T,\beta}\Big)\,[-a,a] \longrightarrow 1 \quad \textit{as } a\to\infty\end{aligned} \tag{4.18}$$

if $\rho > 0$ with $\sigma^2_{T,\beta}$ defined by (2.23), and for $\rho = 0$

$$\lim_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|T(Q_{n,k_n}) - \beta| \le ag(t_n)\} = \left\{\begin{array}{ll} 1 & \\ & \textit{if} \\ 0 & \end{array}\begin{array}{l} a > \sup_{|\beta-\beta_0|\le\vartheta} a^*_{T,\beta}, \\ \\ a < \sup_{|\beta-\beta_0|\le\vartheta} a^*_{T,\beta} \end{array}\right. \tag{4.19}$$

where $a^*_{T,\beta} := \int_0^1 s^{-1}|\int_{[0,s]} \bar{F}_\beta^{-1}\, d\nu_{T,\beta}|\, ds > 1$ and $\sup_{|\beta-\beta_0|\le\vartheta} a^*_{T,\beta} < \infty$.

The proof of this theorem shows that it is the location invariance which precludes the estimator $T(Q_{n,k_n})$ from achieving the optimal asymptotic bound if $\rho = 0$. (Indeed, the proof shows that this bound is not attained for any choice of the intermediate sequence $(k_n)_{n\in I\!N}$.) On the other hand, a combination of Corollary 2.2 with the *uniform* approximation of the estimation error $T(Q_{n,k_n}) - \beta$, which is the main step in the proof of Theorem 4.3, shows that the bound is attained if one takes the infimum merely over those d.f.'s for which not only the dominating function g but also the function ε itself is regularly varying.

Examples 4.2 (i) For the Pickands functional T_P defined in Example 2.1(i), Condition 4.1 is fulfilled (cf. Example 2.2(i)) and the definition of $a^*_{T,\beta}$ reads as

$$a^*_{T_P,\beta} = \frac{1}{\log 2(2^\beta - 1)}\Big(\int_{1/4}^{1/2} \frac{2^\beta}{s}\, ds + \int_{1/2}^1 \frac{1}{s}\, ds\Big) = \frac{2^\beta + 1}{2^\beta - 1} = 1 + \frac{2}{2^\beta - 1},$$

so that the boundary value in (4.19) equals

$$\sup_{|\beta-\beta_0|\le\vartheta} a^*_{T_P,\beta} = 1 + \frac{2}{2^{\beta_0-\vartheta} - 1}.$$

(ii) Recall from Example 2.3 and the discussion following Remark 3.4 that the ML-estimator is a scale- and location-invariant statistical tail functional, whose Fréchet derivative at $\bar{F}_\beta^{-1}$ corresponds to the variance minimizing measure ν_β^* defined by (2.25). By direct calculations we obtain

$$a^*_{T_{ML},\beta} = \Big(\frac{\beta+1}{\beta}\Big)^2 \int_0^1 s^{-1} \Big| \int_0^s 1-(2\beta+1)t^\beta \, dt \Big| \, ds = 2\frac{\beta+1}{\beta}\Big(\frac{\beta+1}{2\beta+1}\Big)^{1/\beta} - 1,$$

which is a decreasing function of β, so that $\sup_{|\beta-\beta_0|\le\vartheta} a^*_{T_{ML},\beta} = a^*_{T_{ML},\beta_0-\vartheta}$. Numerical computations show that $a^*_{T_{ML},\beta} < a^*_{T_P,\beta}$ if $\beta < 3.635$, which is satisfied in most applications.

To justify the application of Theorem 4.3, we briefly indicate how to prove the Fréchet differentiability of the maximum likelihood fuctional T_{ML} at $\bar{F}_\beta^{-1}$ *uniformly* for $\beta \in [\beta_0-\vartheta, \beta_0+\vartheta]$, $0<\vartheta\le\beta_0$, by utilizing the notation and methods introduced in Example 2.3. First check that η is locally uniformly Fréchet differentiable considered as a map $\eta : \bar{D}_{\beta,h} \to C_1^2(\Theta)$, $|\beta-\beta_0|\le\vartheta$, into the space of continuously differentiable functions on Θ with values in $I\!R^2$ equipped with the norm defined on p. 14 of Rieder (1994). Moreover, according to Theorem 1.4.7 of that book, on a suitable open subset $U \subset C_1^2(\Theta)$ containing $\eta(\bar{F}_\beta^{-1},\cdot,\cdot)$, $|\beta-\beta_0|\le\vartheta$, there exists a unique Fréchet differentiable map $\tilde{T} : U \to \Theta$ satisfying $\xi(\tilde{T}(\xi)) = 0$ for all $\xi \in U$ and the Fréchet derivative $d_F\tilde{T} : C_1^2(\Theta) \to I\!R^2$ is continuous w.r.t. the usual operator norm $\|\cdot\|_{L(C_1^2(\Theta),I\!R^2)}$ on the space of continuous linear maps from $C_1^2(\Theta)$ into $I\!R^2$ equipped with the supremum norm. Hence, for all $\delta>0$ and all β with $|\beta-\beta_0|\le\vartheta$, there exists $r_\beta>0$ such that

$$\sup_{\|\xi\|_{C_1^2(\Theta)}\le r_\beta} \Big\| d_F\tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot)) - d_F\tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot)+\xi)\Big\|_{L(C_1^2(\Theta),I\!R^2)} \le \delta/2.$$

Due to the compactness of $[\beta_0-\vartheta,\beta_0+\vartheta]$ and the continuity of $\beta\mapsto\eta(\bar{F}_\beta^{-1},\cdot,\cdot)$, there exists a finite set $\{\beta_1,\ldots,\beta_m\}\subset[\beta_0-\vartheta,\beta_0+\vartheta]$ such that for all $\beta\in[\beta_0-\vartheta,\beta_0+\vartheta]$ there exists β_i with

$$\Big\|\eta(\bar{F}_\beta^{-1},\cdot,\cdot) - \eta(\bar{F}_{\beta_i}^{-1},\cdot,\cdot)\Big\|_{C_1^2(\Theta)} \le r_{\beta_i}/2;$$

define $r := \min_{1\le i\le m} r_{\beta_i}/2$. Then the mean value theorem and the definition of r_β yield

$$\begin{aligned}
&\Big\|\tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot)+\xi) - \tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot)) - d_F\tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot))(\xi)\Big\|_\Theta \\
\le\ & \Big\| d_F\tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot)+\lambda\xi) - d_F\tilde{T}(\eta(\bar{F}_\beta^{-1},\cdot,\cdot))\Big\|_{L(C_1^2(\Theta),I\!R^2)} \cdot \|\xi\|_{C_1^2(\Theta)} \\
\le\ & \delta r
\end{aligned}$$

for some $\lambda\in(0,1)$ if $\|\xi\|_{C_1^2(\Theta)}\le r$. Combining these uniform Fréchet approximations of η and $\tilde{T}$ with $T_{ML} = pr_1\circ\tilde{T}\circ\eta$, one may prove the asserted uniform Fréchet differentiability. $\square$

For the sake of completeness, it should be mentioned that some estimators of the extreme value index attain the optimal asymptotic bound $g(t_n)$ in case of $\rho = 0$ only if β is not too small.

Example 4.3 Following the lines of the proof of Theorem 3.4 in Dekkers and de Haan (1993), it can be proven by tedious calculations that (4.19) holds true for the moment estimator $\hat\beta^M_{n,k_n} = T_M(Q_{n,k_n})$ defined in Example 2.1(iii) if $a^*_{T,\beta}$ is replaced by $a^*_{T_M,\beta} := 1_{\{\beta\ge 1\}} + (2e^{\beta-1}/\beta - 1)\cdot 1_{\{\beta<1\}}$, which is strictly larger than 1 if $\beta < 1$. Hence the optimal bound is attained if and only if $\beta_0 - \vartheta \ge 1$. Check that $a^*_{T_M,\beta} < \min(a^*_{T_P,\beta}, a^*_{T_{ML},\beta})$ for all $\beta > 0$, so that for $\rho = 0$ the performance of the moment estimator lies somewhere in between the performance of the Hill and of the two estimators considered in Example 4.2. □

We close this section with a comparison between Hall's and Welsh's result and Theorem 4.1.

Lemma 4.1 [Drees (1998c), Lemma 2.1] *Consider the two conditions:*

(C1) The density f fulfills (4.1) for all $x \ge x_0$ with $|\beta - \beta_0| \le \vartheta$, $|d - d_0| \le \vartheta$.

(C2) The q.f. F^{-1} fulfills

$$F^{-1}(1-t) = d^\beta t^{-\beta} \exp\Big(- \int_0^t \frac{\varepsilon(s)}{s}\, ds\Big) \tag{4.20}$$

with $|\varepsilon(t)| \le Bt^\rho$ for all $t \le t_0$, $|\beta - \beta_0| \le \vartheta$ and $|d - d_0| \le \vartheta$.

Fix constants $\beta_0, d_0, \rho > 0$ and $0 < \vartheta < \min(\beta_0, d_0)$. Then the following relations hold:

(i) For all constants $A, x_0 > 0$ there exist constants $B > 0$ and $t_0 < 1$ such that if f satisfies (C1), then the corresponding q.f. F^{-1} satisfies (C2).

(ii) Conversely, for all $B > 0$ and $t_0 > 1$ there exist constants $A, x_0 > 0$ such that if F^{-1} satisfies (C2), then f satisfies (C1).

Observe that (4.20) implies (3.6) with $c = d^\beta \exp(-\int_0^1 \varepsilon(s)/s\, ds)$. Conversely, if (3.6) is fulfilled and $\int_0^1 g(s)/s\, ds < \infty$, then (4.20) holds with $d = c^{1/\beta}\exp(\int_0^1 \varepsilon(s)/(\beta s)\, ds) \in [c^{1/\beta}\exp(-\int_0^1 g(s)/(\beta s)\, ds), c^{1/\beta}\exp(\int_0^1 g(s)/(\beta s)\, ds)]$. Therefore, essentially condition (C2) is equivalent to the conditions on F^{-1} imposed in the definition of $\mathcal{F}_1(\beta_0, c, \vartheta, g)$ with $g(t) = Bt^\rho$ if one is merely interested in the upper tail of F^{-1}. In this sense, Hall's and Welsh's result is just a special case of Theorem 4.1.

If $\int_0^1 g(s)/s\, ds$ is infinite, then a similar reformulation of the conditions on F^{-1} in terms of the density is not possible. Furthermore, if $\int_0^1 g(s)/s\, ds < \infty$, yet g is slowly varying, i.e., $\rho = 0$, then an analog to the implication $(C2) \Rightarrow (C1)$ with $Ax^{-\rho/\beta}$ in (4.1) replaced by $A\int_0^{x^{-1/\beta}} g(s)/s\, ds$ holds true, but in general the converse implication is false. This demonstrates that it is indeed appropriate to define neighborhoods of Pareto tails in terms of the q.f. instead of the density.

4.2 Best attainable rates in the general case $\beta \in I\!R$

As already announced in the introduction, one obtains literally the same bounds on the rate of convergence of estimators for arbitrary extreme value indices $\beta \in I\!R$ if the conditions on the q.f. F^{-1} imposed in Theorem 4.1 are replaced by analogous conditions on its derivative. For this let

$$\begin{aligned}\mathcal{F}_2 := \mathcal{F}_2(\beta_0, c, \vartheta, g) \quad &:= \quad \Big\{F \text{ d.f.} \mid (F^{-1})' \text{ satisfies (4.5) with } |\beta - \beta_0| \leq \vartheta, \\ &\qquad |\varepsilon(t)| \leq g(t) \text{ for all } t \in (0,1)\Big\}. \qquad (4.21)\end{aligned}$$

Theorem 4.4 [Drees (1998c), Theorem 3.1] *Fix constants $c, \vartheta > 0$, $\beta_0 \in I\!R$, let $(t_n)_{n \in I\!N}$ fulfill (4.8), assume that the function $g : (0,1) \to (0, \infty)$ satisfies the conditions of Theorem 4.1 and define a^* as in Theorem 4.1. Then we have for an arbitrary sequence $(\hat{\beta}_n)_{n \in I\!N}$ of estimators for β and all $a < a^*$*

$$\limsup_{n \to \infty} \inf_{F \in \mathcal{F}_2(\beta_0, c, \vartheta, g)} P_F\{|\hat{\beta}_n - \beta| \leq a g(t_n)\} < 1. \qquad (4.22)$$

One important difference between the Theorems 4.1 and 4.4 is the fact that (4.5) allows of an arbitrary location parameter. As a consequence, estimators for β that are not (at least asymptotically) location invariant, like, for example, the moment estimator $\hat{\beta}^M_{n,k_n}$, in general do *not* converge uniformly at the optimal rate; this holds true even if the location parameter is restricted to a compact interval by the additional condition $|F^{-1}(1-t_0) - b_0| \leq \vartheta$, where $t_0 < 1$ and $b_0 \in I\!R$ are fixed constants. To see this, consider, e.g., the function $g(t) = t^\rho$ and assume $\rho > \beta_0 > 0$. Then the shifted Pareto d.f. with pertaining q.f.

$$F^{-1}(1-t) = b + \frac{c}{\beta_0} t^{-\beta_0} = \Big(b + \frac{c}{\beta_0}\Big) t^{-\beta_0} \exp\Big(\int_t^1 \frac{-\beta_0^2 b s^{\beta_0}}{c + \beta_0 b s^{\beta_0}} s^{-1}\, ds\Big) \qquad (4.23)$$

belongs to $\mathcal{F}_2(\beta_0, c, \vartheta, g)$, and b may be chosen such that $F^{-1}(1-t_0) = b_0$. On the other hand, using representation (4.23), it follows by Dekkers and de Haan (1993), Theorem 3.4, that for the optimal choice of k_n the estimation error $|\hat{\beta}^M_{n,k_n} - \beta|$ is of the stochastic order $n^{-\beta_0/(2\beta_0+1)}$, while the lower bound on the rate of convergence given in Theorem 4.4 is of the smaller order $n^{-\rho/(2\rho+1)}$. The slower convergence of $\hat{\beta}^M_{n,k_n}$ is caused by the additional location parameter b, which, according to (4.23), acts as a function $\varepsilon(s) \sim \text{const.} s^{\beta_0}$ in the situation of Section 2, thus leading to a rate of convergence $n^{-\beta_0/(2\beta_0+1)}$.

Likewise it can be proven that the rate of convergence is not optimal if $-\rho < \beta < 0$. Hence again it turns out that it is not advisable to use estimators that are not location invariant.

In contrast, the scale- and location-invariant statistical tail functionals considered in Theorem 4.3 converge with the optimal rate for all $\rho \geq 0$. Moreover, they even attain the optimal bound in case of $\rho = 0$ if an additional condition on the measures $\nu_{T,\beta}$ pertaining to the Fréchet derivatives of T at $\bar{F}_\beta^{-1}$, $|\beta - \beta_0| \leq \vartheta$, is fulfilled.

Theorem 4.5 [cf. Drees (1998c), Theorem 3.2 and the following remarks] *If the assumptions of Theorem 4.3 hold, then for $\rho > 0$*

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{F}_2(\beta_0,c,\vartheta,g)} P_F\{|T(Q_{n,k_n})-\beta| \le ag(t_n)\}$$
$$= \inf_{|\beta-\beta_0|\le\vartheta} \mathcal{N}\Big(\int_0^1 s^{\rho-1}\Big|\int_{[0,1]} \bar F_\beta^{-1}(\min(s,t))\nu_{T,\beta}(dt)\Big|\,ds, \sigma^2_{T,\beta}\Big)[-a,a] \longrightarrow 1 \quad \text{as } a\to\infty,$$

and for $\rho = 0$

$$\lim_{n\to\infty} \inf_{F\in\mathcal{F}_2(\beta_0,c,\vartheta,g)} P_F\{|T(Q_{n,k_n})-\beta| \le ag(t_n)\} = \begin{cases} 1 & \\ & \text{if} \\ 0 & \end{cases} \begin{array}{l} a > \sup_{|\beta-\beta_0|\le\vartheta} \tilde a_{T,\beta}, \\ a < \sup_{|\beta-\beta_0|\le\vartheta} \tilde a_{T,\beta}, \end{array}$$

where $\tilde a_{T,\beta} := \int_0^1 s^{-1}|\int_{[0,1]} \bar F_\beta^{-1}(\min(s,t))\,\nu_{T,\beta}(dt)|\,ds \ge 1$ and equality holds if and only if $\int_{[0,1]} \bar F_\beta^{-1}(\min(s,t))\,\nu_{T,\beta}(dt)$ has constant sign for almost all $s\in[0,1]$.

The following examples demonstrate that some scale- and location-invariant statistical tail functionals, including the Pickands and the ML-estimator, attain the optimal bound $g(t_n)$ asymptotically if $\rho = 0$, while others do not.

Examples 4.4 (i) For the Pickands functional T_P the constants $\tilde a_{T_P,\beta}$ equal 1 and hence the bound in the case $\rho = 0$ is attained, since

$$\int_0^1 \bar F_\beta^{-1}(\min(s,t))\,\nu_{T,\beta}(dt)$$
$$= \frac{\beta}{(2^\beta-1)\log 2}\begin{cases} 0 & 0\le s<1/4 \text{ or } s=1, \\ 2^{-\beta}(\bar F_\beta^{-1}(1/4)-\bar F_\beta^{-1}(s)) \quad \text{if} & 1/4\le s<1/2, \\ \bar F_\beta^{-1}(s)-\bar F_\beta^{-1}(1) & 1/2\le s<1 \end{cases} \ge 0.$$

(ii) For the variance minimizing measure ν^*_β defined by (2.25) one has

$$\int_0^1 \bar F_\beta^{-1}(\min(s,t))\,\nu^*_\beta(dt) = \begin{cases} (\beta+1)s(1-s^\beta)/\beta & \\ & \text{if} \\ \beta s(1-s^{-(\beta+1)})/(\beta+1) & \end{cases} \begin{array}{l} \beta\ge -1/2, \\ \beta < -1/2, \end{array}$$

which is of constant sign. In particular, the maximum likelihood estimator considered in Example 2.3 attains the optimal bound, at least, for $\beta > 0$.

(iii) Check that the measures

$$\nu_\beta := -\frac{\beta}{(2^\beta-1)(4^\beta a-1)\log 2}\times$$
$$\times\Big(\varepsilon_1 - (2^\beta a + 1 + 2^{-\beta})\varepsilon_{1/2} + ((2^\beta+1)a + 2^{-\beta})\varepsilon_{1/4} - a\varepsilon_{1/8}\Big)$$

satisfy the conditions of Lemma 2.2 for $|\beta-\beta_0|\le\vartheta$ if $a > 4^{\vartheta-\beta_0}$. Because the map $\beta\mapsto\nu_\beta$ is continuous in the sense of (2.34), the construction of Theorem 2.6 gives a functional T with Hadamard derivatives at $\bar{F}_\beta^{-1}$ pertaining to ν_β. Furthermore, it is easily seen that the functional is even uniformly Fréchet differentiable. Since

$$\begin{aligned}
&\int_0^1 \bar{F}_\beta^{-1}(\min(s,t))\,\nu_\beta(dt)\\
&= \; -\frac{\beta}{(2^\beta-1)(4^\beta a-1)\log 2}\begin{cases} -a(8^\beta - s^{-\beta})/\beta & \text{if } \quad s\in[1/8,1/4),\\ (s^{-\beta}-1)/\beta & \phantom{\text{if}} \quad s\in[1/2,1)\end{cases}
\end{aligned}$$

has different sign for $s\in[1/8,1/4)$ and $s\in[1/2,1)$, it follows that $\tilde{a}_{T,\beta}>1$ and thus the optimal bound is not attained. □

4.3 Proofs

Proof of Theorem 4.1. The basic idea of the proof is similar to the one used by Hall and Welsh (1984) and Farrell (1972). Fix some $\lambda>0$ and let $s_n := \lambda t_n$. Define

$$(F_n^+)^{-1}(1-t) := ct^{-\beta_n^+}\exp\Big(\int_t^1 \frac{\varepsilon_n^+(s)}{s}\,ds\Big) \text{ and } (F_n^-)^{-1}(1-t) := ct^{-\beta_n^-}\exp\Big(\int_t^1 \frac{\varepsilon_n^-(s)}{s}\,ds\Big)$$

for $t\in(0,1)$ where

$$\beta_n^\pm := \beta_0 \pm g(s_n) \quad \text{and} \quad \varepsilon_n^\pm := (\beta_0-\beta_n^\pm)1_{[s_n,1]} = \mp g(s_n)1_{[s_n,1]}.$$

Obviously, for the pertaining d.f.'s one has $F_n^\pm \in \mathcal{F}_1(\beta_0,c,\vartheta,g)$ for sufficiently large n, since $g(t)$ is eventually non-increasing and converges to 0 as $t\downarrow 0$.

The corresponding densities are

$$f_n^\pm(x) = \begin{cases} \dfrac{1}{c\beta_0}\Big(\dfrac{x}{c}\Big)^{-(1/\beta_0+1)} & \text{if} \quad c\le x\le cs_n^{-\beta_0},\\ \dfrac{s_n^{\beta_0+1}}{c\beta_n^\pm}\Big(\dfrac{s_n^{\beta_0}x}{c}\Big)^{-(1/\beta_n^\pm+1)} & \phantom{\text{if}} \quad cs_n^{-\beta_0}<x. \end{cases}$$

Thus

$$\begin{aligned}
&\int_c^\infty \Big(1-\frac{f_n^+(x)}{f_n^-(x)}\Big)^2 f_n^-(x)\,dx\\
&= \int_{cs_n^{-\beta_0}}^\infty \bigg(1-\frac{\beta_n^-}{\beta_n^+}\,\Big(\frac{s_n^{\beta_0}x}{c}\Big)^{1/\beta_n^- - 1/\beta_n^+}\bigg)^2 \frac{s_n^{\beta_0+1}}{c\beta_n^-}\Big(\frac{s_n^{\beta_0}x}{c}\Big)^{-(1/\beta_n^-+1)}\,dx\\
&= s_n\int_1^\infty \Big(1-\frac{\beta_n^-}{\beta_n^+}x^{1/\beta_n^- - 1/\beta_n^+}\Big)^2 \frac{1}{\beta_n^-}x^{-(1/\beta_n^-+1)}\,dx\\
&= s_n\Big(1-2+\frac{\beta_n^-}{\beta_n^{+2}(2/\beta_n^+ - 1/\beta_n^-)}\Big)\\
&= s_n\frac{(\beta_n^+-\beta_n^-)^2}{\beta_n^+(2\beta_n^- - \beta_n^+)}\\
&\sim 4\lambda^{2\rho+1}\beta_0^{-2}n^{-1}
\end{aligned}$$

by the definitions of s_n and t_n and the ρ-variation of g. Thus

$$\int \prod_{i=1}^{n} \Big(\frac{f_n^+(x_i)}{f_n^-(x_i)}\Big)^2 f_n^-(x_i)\, \boldsymbol{\lambda}^n(d\mathbf{x}) = \Big(1+\int_1^\infty \Big(1-\frac{f_n^+(x)}{f_n^-(x)}\Big)^2 f_n^-(x)\,dx\Big)^{n/2} \longrightarrow \exp\Big(2\lambda^{2\rho+1}\beta_0^{-2}\Big)$$

as $n \to \infty$, where $\boldsymbol{\lambda}^n(d\mathbf{x})$ denotes integration w.r.t. the Lebesgue-measure on $I\!R^n$. By Theorem 3.1 of Hall (1989) it follows that

$$\limsup_{n\to\infty} \inf_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta_n-\beta| \le (\beta_n^+-\beta_n^-)/2\} \le 1-\exp\Big(-2\lambda^{2\rho+1}\beta_0^{-2}\Big)/4 \qquad (4.24)$$

so that, in view of $(\beta_n^+-\beta_n^-)/2 = g(s_n) \sim \lambda^\rho g(t_n) > ag(t_n)$ for $a < a^*$, $\lambda > a^{1/\rho}$ if $\rho > 0$, and λ arbitrary else, the assertion follows readily. □

Proof of Theorem 4.2. The main step is to verify that convergence (3.7) holds uniformly in $F \in \mathcal{F}_1(\beta_0, c, \vartheta, g)$, i.e.,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} \Big|P_F\{k_n^{1/2}(\hat\beta^H_{n,k_n}-\beta-b^H_{n,k_n,F}) \le x\} - \mathcal{N}(0,\beta^2)\,(-\infty,x]\Big| = 0, \qquad (4.25)$$

where $b^H_{n,k_n,F} := \int_0^1 \varepsilon(tk_n/n)\,dt$. For this, we are following the lines of the proof presented in Csörgő et al. (1985), where k_n/n is denoted by λ and $K = 1_{[0,1]}$ is the kernel pertaining to the Hill estimator. Moreover, the process U_n used by Csörgő et al. (1985) and the process q_n introduced in the proof of Lemma 3.4 are linked via $q_n(s) = 1-U_n(1-s)$, since the uniform r.v.'s considered by Csörgő et al. (1985) correspond to the r.v.'s $1-U_i$ in the definition of q_n. Recall from (3.39) that w.l.o.g.

$$\hat\beta^H_{n,k_n} - \beta - b^H_{n,k_n,F} = \int_{[0,1]} \log\frac{F^{-1}(1-q_n(tk_n/n))}{F^{-1}(1-tk_n/n)}\,(\lambda-\varepsilon_1)(dt),$$

where again λ denotes the Lebesgue measure and ε_1 the Dirac measure at 1. According to Csörgő et al. (1985), the right-hand side can be approximated by an integral of a certain Brownian bridge B_n. More precisely, consider

$$\begin{aligned}
&k_n^{1/2}\Big(\hat\beta^H_{n,k_n}-\beta-b^H_{n,k_n,F}\Big) - \beta\int_{[0,1]}\frac{B_n(1-tk_n/n)}{t(k_n/n)^{1/2}}\,(\lambda-\varepsilon_1)(dt)\\
&= \; k_n^{1/2}\int_0^{n/((n+1)k_n)}\log\frac{F^{-1}(1-q_n(tk_n/n))}{F^{-1}(1-tk_n/n)}\,dt - \beta\int_0^{n/((n+1)k_n)}\frac{B_n(1-tk_n/n)}{t(k_n/n)^{1/2}}\,dt\\
&\quad -\beta k_n^{1/2}\int_{[n/((n+1)k_n),1]}\log\frac{q_n(tk_n/n)}{tk_n/n}+\frac{B_n(1-tk_n/n)}{tk_n n^{-1/2}}\,(\lambda-\varepsilon_1)(dt)\\
&\quad +k_n^{1/2}\int_{[n/((n+1)k_n),1]}\int_{q_n(tk_n/n)}^{tk_n/n}\frac{\varepsilon(s)}{s}\,ds\,(\lambda-\varepsilon_1)(dt)\\
&=: \; T_1+T_2+T_3+T_4, \qquad \text{say.}
\end{aligned}$$

First note that, as $q_n(tk_n/n) = U_{1:n}$ for $0 < t \le n/((n+1)k_n)$, using integration by parts one obtains

$$\begin{aligned}
T_1 &= -\beta k_n^{1/2}\int_0^{n/((n+1)k_n)}\log\frac{U_{1:n}}{tk_n/n}\,dt + \frac{n}{n+1}k_n^{-1/2}\int_{U_{1:n}}^{1/(n+1)}\frac{\varepsilon(s)}{s}\,ds\\
&\quad -k_n^{1/2}\int_0^{n/((n+1)k_n)}\varepsilon(tk_n/n)\,dt \;=:\; -\beta T_{1,1}+T_{1,2}-T_{1,3}.
\end{aligned}$$

According to the proof of Lemma 11 of Csörgő et al. (1985), the term $T_{1,1}$, which does not depend on the underlying d.f. F, converges to 0 in probability. Moreover, by the monotonicity of g one has for sufficiently large n

$$|T_{1,2}| \leq k_n^{-1/2} \max\Big(g(1/(n+1)), g(U_{1:n})\Big) |\log((n+1)U_{1:n})| = o_P(1)$$

and

$$|T_{1,3}| \leq k_n^{-1/2} g(1/(n+1)) = o_P(1),$$

so that $|T_1| = o_P(1)$ uniformly for $F \in \mathcal{F}_1(\beta_0, c, \vartheta, g)$.

By the Lemmas 11 and 12 of Csörgő et al. (1985), for suitable Brownian bridges B_n independent of F, the terms T_2/β and T_3/β converge to 0 in probability, and hence $|T_2| + |T_3| = o_P(1)$ uniformly for $F \in \mathcal{F}_1(\beta_0, c, \vartheta, g)$.

Finally,

$$\begin{aligned} |T_4| &\leq k_n^{1/2} \int_{n/((n+1)k_n)}^1 \Big| \int_{q_n(tk_n/n)}^{tk_n/n} \frac{g(s)}{s} ds \Big| dt + k_n^{1/2} \int_{[n/((n+1)k_n),1]} \Big| \int_{q_n(tk_n/n)}^{tk_n/n} \frac{g(s)}{s} ds \Big| \varepsilon_1(dt) \\ &=: L_{n,1} + L_{n,2}, \end{aligned}$$

where according to Csörgő et al. (1985, p. 1071) $L_{n,i} = o_P(1)$, $i = 1, 2$, and hence $|T_4| = o_P(1)$ uniformly for $F \in \mathcal{F}_1(\beta_0, c, \vartheta, g)$.

To sum up, we have proven that

$$k_n^{1/2}\Big(\hat\beta^H_{n,k_n} - \beta - b^H_{n,k_n,F}\Big) = \beta \int_{[0,1]} \frac{B_n(1 - tk_n/n)}{t(k_n/n)^{1/2}} (\lambda - \varepsilon_1)(dt) + o_P(1)$$

uniformly for $F \in \mathcal{F}_1(\beta_0, c, \vartheta, g)$. Now Csörgő et al. (1985), Lemma 9, implies (4.25).

Next note that $\sup_{F \in \mathcal{F}_1(\beta_0,c,\vartheta,g)} |b^H_{n,k_n,F}| = \int_0^1 g(tk_n/n)\, dt$. For $\rho > 0$, Karamata's theorem and the definition of k_n yield $\int_0^1 g(tk_n/n)\, dt \sim g(t_n)/(\rho+1)$, so that

$$\lim_{n\to\infty} \inf_{F \in \mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta_n - \beta| \leq a g(t_n)\} = \inf_{|\mu| \leq 1/(\rho+1), |\beta - \beta_0| \leq \vartheta} \mathcal{N}(\mu, \beta^2)\,[-a, a],$$

from which (4.14) is immediate.

In the case $\rho = 0$, (4.16) and the monotonicity of g imply $k_n^{1/2}(a_n - \int_0^1 g(tk_n/n)\, dt) \to \infty$, so that $k_n^{1/2}(a_n \pm b^H_{n,k_n,F}) \to \infty$ uniformly on $\mathcal{F}_1(\beta_0, c, \vartheta, g)$. Therefore,

$$\begin{aligned} &\liminf_{n\to\infty} \inf_{F \in \mathcal{F}_1(\beta_0,c,\vartheta,g)} P_F\{|\hat\beta_n - \beta| \leq a_n\} \\ &\quad \geq \liminf_{n\to\infty} \inf_{F \in \mathcal{F}_1(\beta_0,c,\vartheta,g)} \mathcal{N}(0, \beta^2)\,[-k_n^{1/2}(a_n + b^H_{n,k_n,F}), k_n^{1/2}(a_n - b^H_{n,k_n,F})] = 1. \end{aligned}$$

Now the last assertion is obvious, since by the definition of t_n^* and t_n we have $(nt_n^*)^{1/2}(ag(t_n) - g(t_n^*)) \sim (t_n^*/t_n)^{1/2}(a - g(t_n^*)/g(t_n)) \to \infty$ if $a > 1$. $\square$

For the *proof of Theorem 4.3* we need analogs to the Theorems 2.1 and 2.3. First, we derive a uniform approximation similar to Lemma 2.1. For this purpose, let

$$\mu_F(t, x) := \frac{x^{-\beta}}{\beta} \int_x^1 \frac{\varepsilon(st)}{s}\, ds.$$

Proposition 4.1 *For all $x_0, \iota > 0$ one has*

$$\sup_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} \sup_{x\in(0,x_0]} x^{\beta+\iota}\Big|\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{\beta F^{-1}(1-t)} - F_\beta^{-1}(x) - \mu_F(t,x)\Big| = O(g^2(t)). \tag{4.26}$$

Proof. The left-hand side of (4.26) equals

$$\begin{aligned}
&\sup_{F\in\mathcal{F}_1} \sup_{x\in(0,x_0]} \frac{x^\iota}{\beta}\Big|\exp\Big(\int_x^1 \frac{\varepsilon(st)}{s}\,ds\Big) - 1 - \int_x^1 \frac{\varepsilon(st)}{s}\,ds\Big| \\
\le\ & \sup_{F\in\mathcal{F}_1} \sup_{x\in(0,x_0]} \frac{x^\iota}{2\beta}\Big(\int_x^1 \frac{\varepsilon(st)}{s}\,ds\Big)^2 \exp\Big|\int_x^1 \frac{\varepsilon(st)}{s}\,ds\Big| \\
\le\ & \sup_{x\in(0,x_0]} \frac{x^\iota}{2(\beta_0-\vartheta)} \cdot g^2(\max(1,x_0)t)\cdot \log^2 x \cdot \max(x^{\iota/2}, x^{-\iota/2}) \\
=\ & O(g^2(t))
\end{aligned}$$

for all sufficiently small t, where for the first inequality a Taylor expansion of the exponential function has been used and then the eventually monotone convergence of the ρ–varying function g to 0. $\square$

Next, we establish a uniform analog to Theorem 2.1 for $\tilde{Q}_{n,k_n,F} = \tilde{Q}_n$ defined by (2.7).

Proposition 4.2 *For all $h \in \mathcal{H}$ (see (2.9)) one has almost surely*

$$\begin{aligned}
\sup_{F\in\mathcal{F}_1(\beta_0,c,\vartheta,g)} \Big\| & \frac{\tilde{Q}_{n,k_n,F} - F^{-1}(1-k_n/n)}{\beta F^{-1}(1-k_n/n)} \\
& -\Big(F_\beta^{-1}(t) - t^{-(\beta+1)}k_n^{-1}W(k_n t) + \mu_F(k_n/n,t)\Big)_{t\in[0,1]}\Big\|_{\beta,h} = o\Big(k_n^{-1/2} + g(k_n/n)\Big).
\end{aligned}$$

Proof. Like in the proof of Theorem 2.1, it follows from Proposition 4.1 that

$$\begin{aligned}
\sup_{F\in\mathcal{F}_1} \sup_{t\in(0,1]} t^{\beta+1/2}\Big| & \frac{\tilde{Q}_{n,k_n,F}(t) - F^{-1}(1-k_n/n)}{\beta F^{-1}(1-k_n/n)} - \frac{(S_{[k_n t]+1}/k_n)^{-\beta}-1}{\beta} \\
& -\mu_F(k_n/n, S_{[k_n t]+1}/k_n)\Big| = O(g^2(k_n/n))
\end{aligned}$$

almost surely. Moreover, in the same way as in the proof of Proposition 3.1, (2.44), the strong law of large numbers, the law of iterated logarithms and the definition of $\mathcal{H}$ yield

$$
\begin{aligned}
&\sup_{F\in\mathcal{F}_1}\sup_{t\in(0,1]} t^\beta h(t)\Big|\mu_F(k_n/n, S_{[k_nt]+1}/k_n)-\mu_F(k_n/n,t)\Big| \\
&= \frac{1}{\beta}\sup_{F\in\mathcal{F}_1}\sup_{t\in(0,1]} h(t)\Big|\Big(\frac{S_{[k_nt]+1}}{k_nt}\Big)^{-\beta}\int_{S_{[k_nt]+1}/k_n}^{t}\frac{\varepsilon(sk_n/n)}{s}\,ds \\
&\qquad\qquad + \Big(\Big(\frac{S_{[k_nt]+1}}{k_nt}\Big)^{-\beta}-1\Big)\int_t^1\frac{\varepsilon(sk_n/n)}{s}\,ds\Big| \\
&= O\Big(g(k_n/n)\sup_{t\in(0,1]} h(t)\Big(\Big|\log\frac{S_{[k_nt]+1}}{k_nt}\Big| + ([k_nt]+1)^{-1/2}(\log\log(3([k_nt]+1)))^{1/2}|\log t|\Big)\Big) \\
&= o\big(g(k_n/n)\big)
\end{aligned}
$$

almost surely. A combination with a uniform version of Proposition 2.1 proves the assertion. □

In view of Proposition 4.2, the assumptions on the functional T imply

$$
\sup_{F\in\mathcal{F}_1}\Big|T(\tilde{Q}_{n,k_n,F})-\beta+T'_\beta\Big((t^{-(\beta+1)}k_n^{-1}W(k_nt))_{t\in[0,1]}\Big)-T'_\beta(\mu_F(k_n/n,\cdot))\Big| = o(k_n^{-1/2}+g(k_n/n)) \tag{4.27}
$$

almost surely, as $\|(t^{-(\beta+1)}k_n^{-1}W(k_nt))_{t\in[0,1]}\|_{\beta,h} = O(k_n^{-1/2})$ and $\sup_{F\in\mathcal{F}_1}\|\mu_F(k_n/n,\cdot)\|_{\beta,h} = O(g(k_n/n))$. Next note that

$$
\begin{aligned}
\sup_{F\in\mathcal{F}_1}|T'_\beta(\mu_F(k_n/n,\cdot))| &= \sup_{F\in\mathcal{F}_1}\Big|\int_{[0,1]}\bar{F}_\beta^{-1}(t)\int_t^1\frac{\varepsilon(sk_n/n)}{s}\,ds\,\nu_{T,\beta}(dt)\Big| \\
&= \sup_{F\in\mathcal{F}_1}\Big|\int_0^1\frac{\varepsilon(sk_n/n)}{s}\int_{[0,s]}\bar{F}_\beta^{-1}(t)\nu_{T,\beta}(dt)\,ds\Big| \\
&= \sup_{|\beta-\beta_0|\le\vartheta}\int_0^1\frac{g(sk_n/n)}{s}\Big|\int_{[0,s]}\bar{F}_\beta^{-1}(t)\,\nu_{T,\beta}(dt)\Big|\,ds \\
&\sim g(t_n)\sup_{|\beta-\beta_0|\le\vartheta}\int_0^1 s^{\rho-1}\Big|\int_{[0,s]}\bar{F}_\beta^{-1}(t)\,\nu_{T,\beta}(dt)\Big|\,ds
\end{aligned} \tag{4.28}
$$

by the ρ–variation of g and the uniform convergence theorem for regularly varying functions, since, according to (4.17), $\int_{[0,s]}\bar{F}_\beta^{-1}(t)|\nu_{T,\beta}|(dt) = o(s^{-1/2})$ uniformly.

In case of $\rho>0$ it follows from (4.27) and (2.6) that

$$
\sup_{F\in\mathcal{F}_1}\sup_{x\in I\!R}\Big|P_F\{k_n^{1/2}(T(Q_{n,k_n})-\beta)\le x\} - \mathcal{N}\Big(k_n^{1/2}T'_\beta(\mu_F(k_n/n,\cdot)),\sigma^2_{T,\beta}\Big)(-\infty,x]\Big| \to 0,
$$

so that (4.18) can be concluded as in the proof of Theorem 4.2. For, according to (4.17), $\int_0^1 s^{\rho-1}\Big|\int_{[0,s]}\bar{F}_\beta^{-1}(t)\,\nu_{T,\beta}(dt)|\,ds$ and $\sigma^2_{T,\beta}$ are bounded for $|\beta-\beta_0|\le\vartheta$.

If $\rho=0$, then (4.27) and the definition of k_n lead to

$$
\sup_{F\in\mathcal{F}_1}\Big|T(\tilde{Q}_{n,k_n,F})-\beta-T'_\beta(\mu_F(k_n/n,\cdot))\Big| = o(g(t_n))
$$

almost surely, and hence for all $\iota>0$

$$
\sup_{F\in\mathcal{F}_1}P_F\{|T(Q_{n,k_n})-\beta-T'_\beta(\mu_F(k_n/n,\cdot))|>\iota g(t_n)\}\longrightarrow 0.
$$

Thus, in view of (4.28), assertion (4.19) is immediate.

The boundedness of $a^*_{T,\beta}$ for $|\beta-\beta_0|\le\vartheta$ follows again from (4.17), so that it remains to prove that $a^*_{T,\beta}>1$. For this end, note that by Lemma 2.2(iii)

$$\begin{aligned}\int_0^1 s^{-1}\int_{[0,s]}\bar F_\beta^{-1}(t)\,\nu_{T,\beta}(dt)\,ds &= \int_{[0,1]}\int_t^1 s^{-1}\,ds\,\bar F_\beta^{-1}(t)\,\nu_{T,\beta}(dt)\\ &= -\int_{[0,1]}\bar F_\beta^{-1}(t)\log t\,\nu_{T,\beta}(dt)=1.\end{aligned}$$

Therefore, $a^*_{T,\beta}\ge 1$ and equality holds if and only if $\int_{[0,s]}\bar F_\beta^{-1}d\nu_{T,\beta}$ has constant sign for almost all $s\in[0,1]$. On the other hand, in view of Lemma 2.2, using integration by parts one obtains $0=\nu_{T,\beta}[0,1]=\int_{[0,1]}s^\beta s^{-\beta}\nu_{T,\beta}(ds)=\int_0^1\beta s^{\beta-1}\int_{[0,s]}t^{-\beta}\nu_{T,\beta}(dt)\,ds$, which clearly implies that $s\mapsto\int_{[0,s]}t^{-\beta}\nu_{T,\beta}(dt)$ changes its sign, since $\nu_{T,\beta}$ cannot be the zero measure. □

Proof of Lemma 4.1. First assume that condition (C1) is fulfilled. Then we have $1-F(x)=dx^{-1/\beta}(1+\tilde r(x))$ with $|\tilde r(x)|\le Ax^{-\rho/\beta}/(\rho+1)$ for $x\ge x_0$, so that

$$F^{-1}(1-t)=d^\beta t^{-\beta}(1+h(t)),\tag{4.29}$$

where the tail of h is given by the condition

$$1-F(F^{-1}(1-t))=t\quad\Longrightarrow\quad(1+h(t))^{1/\beta}=1+\tilde r(F^{-1}(1-t)).\tag{4.30}$$

Now consider

$$\begin{aligned}\varepsilon(t)&:=\frac{t(F^{-1})'(1-t)}{F^{-1}(1-t)}-\beta=\frac{t}{f(F^{-1}(1-t))F^{-1}(1-t)}-\beta\\ &=\frac{\beta(1+h(t))^{1/\beta}}{1+r(F^{-1}(1-t))}-\beta=\beta\frac{\tilde r(F^{-1}(1-t))-r(F^{-1}(1-t))}{1+r(F^{-1}(1-t))}.\end{aligned}$$

Since by (4.29) and (4.30) one has $(d/(2t))^\beta<F^{-1}(1-t)<(2d/t)^\beta$ for sufficiently small t, it follows that

$$|\varepsilon(t)|\le\beta A\Big(\frac{1}{\rho+1}+1\Big)\frac{(F^{-1}(1-t))^{-\rho/\beta}}{1-A(F^{-1}(1-t))^{-\rho/\beta}}\le\beta A\frac{\rho+2}{\rho+1}2(d/2)^{-\rho}t^\rho,\quad t\le t_0,$$

for some $t_0<1$. Thus, in view of

$$F^{-1}(1-t)=t_0^\beta F^{-1}(1-t_0)t^{-\beta}\exp\Big(\int_t^{t_0}\frac{\varepsilon(s)}{s}\,ds\Big)=\tilde d t^{-\beta}\exp\Big(-\int_0^t\frac{\varepsilon(s)}{s}\,ds\Big)$$

and (4.29), assertion (C2) is immediate.

For the proof of (ii) assume that F^{-1} satisfies (C2). Then

$$1-F(x)=dx^{-1/\beta}(1+h(x))\qquad\text{with}\qquad 1+h(x)=\exp\Big(-\int_0^{1-F(x)}\frac{\varepsilon(s)}{\beta s}\,ds\Big)$$

for sufficiently large x. Therefore

$$f(x) = \frac{1}{(F^{-1})'(1-(1-F(x)))} = \frac{d}{\beta} x^{-(1/\beta+1)}(1+r(x))$$

where

$$r(x) = \Big(\exp\Big(-\int_0^{1-F(x)} \frac{\varepsilon(s)}{\beta s}\,ds\Big) - 1 - \frac{\varepsilon(1-F(x))}{\beta}\Big)\Big/\Big(1+\frac{\varepsilon(1-F(x))}{\beta}\Big)$$

for almost all sufficiently large x. In the same way as above, one can conclude that $1-F(x) \le 2dx^{-1/\beta}$ and thus by the mean value theorem

$$|r(x)| \;\le\; 2^{\rho+1}\Big(\frac{2}{\rho}+1\Big)\frac{B}{\beta_0-\vartheta}(d_0+\vartheta)^\rho x^{-\rho/\beta} =: Ax^{-\rho/\beta}$$

for almost all sufficiently large x. Thus there exists a constant $x_0 > 0$, such that a suitable version of the density f satisfies (C1). $\square$

Proof of Theorem 4.4. Since every neighborhood of 0 includes subintervals of $(0,\infty)$ (and $(-\infty,0)$) and the bound in (4.22) is independent of ϑ, w.l.o.g. one may assume that either $\beta_0 > 0$ or $\beta_0 < 0$. We concentrate on the latter case, because the first one can be treated in a similar way as Theorem 4.1. Moreover, utilizing a scale transformation it is easily seen that one may assume $c = 1$.

Fix some $\lambda > 0$, $\tau > 0$ and $0 < \iota < (1+1/\tau)^{1/\beta_0}$. Let $s_n := \lambda t_n$ and

$$b_n := \frac{1-\iota}{1+\tau} g\Big(((1+1/\tau)^{1/\beta_0} - \iota)s_n\Big).$$

Define

$$\begin{aligned} \beta_n^\pm &:= \beta_0 \pm b_n, \\ \tau_n^+ := \tau, \qquad \tau_n^- &:= \Big(\frac{\beta_n^+}{\beta_n^-}\Big(1+\frac{1}{\tau}\Big)-1\Big)^{-1}, \\ \gamma_n^\pm &:= \Big(\frac{1}{\beta_0} + \tau_n^\pm\Big(\frac{1}{\beta_0} - \frac{1}{\beta_n^\pm}\Big)\Big)^{-1} \end{aligned}$$

and

$$\chi_n^\pm := \Big(1+\frac{1}{\tau_n^\pm}\Big)^{1/\gamma_n^\pm},$$

so that $\beta_n^\pm, \gamma_n^\pm < 0$ and $0 < \chi_n^\pm < 1$ for sufficiently large n. Consider

$$(F_n^\pm)^{-1}(1-t) := \begin{cases} t^{-\beta_0}/\beta_0 & s_n \le t \le 1, \\ s_n^{-\beta_0}(1/\beta_0 - 1/\gamma_n^\pm) + s_n^{\gamma_n^\pm-\beta_0} t^{-\gamma_n^\pm}/\gamma_n^\pm \quad \text{if} & \chi_n^\pm s_n \le t < s_n, \\ s_n^{\beta_n^\pm-\beta_0}(\chi_n^\pm)^{\beta_n^\pm-\gamma_n^\pm} t^{-\beta_n^\pm}/\beta_n^\pm & 0 \le t < \chi_n^\pm s_n. \end{cases}$$

Check that $(F_n^+)^{-1}$ is continuously differentiable with derivative

$$\begin{aligned}((F_n^+)^{-1})'(1-t) &= \begin{cases} t^{-(\beta_0+1)} & s_n \le t \le 1, \\ s_n^{\gamma_n^+ - \beta_0} t^{-(\gamma_n^+ +1)} & \text{if} \quad \chi_n^+ s_n \le t < s_n, \\ s_n^{\beta_n^+ - \beta_0} (\chi_n^+)^{\beta_n^+ - \gamma_n^+} t^{-(\beta_n^+ +1)} & 0 < t < \chi_n^+ s_n \end{cases} \\ &= t^{-(\beta_n^+ +1)} \exp\Big(\int_t^1 \frac{(\beta_0 - \beta_n^+) 1_{[s_n,1]}(s) + (\gamma_n^+ - \beta_n^+) 1_{(\chi_n^+ s_n, s_n]}(s)}{s} \, ds \Big).\end{aligned}$$

Since $\beta_n^+ - \gamma_n^+ \sim (1+\tau) b_n$ and hence by the monotonicity of g

$$g(s_n) \ge g(\chi_n^+ s_n) \ge g\Big(\big((1+\frac{1}{\tau})^{1/\beta_0} - \iota\big) s_n \Big) \ge |\gamma_n^+ - \beta_n^+|$$

for sufficiently large n, eventually $F_n^+ \in \mathcal{F}_2(\beta_0, 1, \vartheta, g)$. Straightforward calculations show that the density pertaining to $(F_n^+)^{-1}$ equals

$$f_n^+(x) = \begin{cases} (\beta_0 x)^{-(1/\beta_0 + 1)} & \frac{1}{\beta_0} \le x \le \frac{s_n^{-\beta_0}}{\beta_0}, \\ s_n^{\beta_0+1} \Big(\gamma_n^+ s_n^{\beta_0} (x - s_n^{-\beta_0}/\beta_0) + 1 \Big)^{-(1/\gamma_n^+ + 1)} & \text{if} \quad \frac{s_n^{-\beta_0}}{\beta_0} < x \le \frac{s_n^{-\beta_0} (\chi_n^+)^{-\gamma_n^+}}{\beta_n^+}, \\ s_n^{\beta_0+1} (\chi_n^+)^{\gamma_n^+ + 1} \Big(\beta_n^+ s_n^{\beta_0} (\chi_n^+)^{\gamma_n^+} x \Big)^{-(1/\beta_n^+ + 1)} & \frac{s_n^{\beta_0} (\chi_n^+)^{-\gamma_n^+}}{\beta_n^+} < x < 0, \end{cases}$$

and $f_n^+(x) = 0$ else. Likewise, the analogous properties can be proven for $(F_n^-)^{-1}$. Notice that

$$\frac{(\chi_n^-)^{-\gamma_n^-}}{\beta_n^-} = \frac{(\chi_n^+)^{-\gamma_n^+}}{\beta_n^+} = \frac{1}{\beta_n^+(1+1/\tau)} =: r_n,$$

so that the partition of the support is the same for f_n^+ and f_n^-.

We proceed as in the proof of Theorem 4.1 by computing the χ^2–'distance' between f_n^+ and f_n^-:

$$\begin{aligned}&\int_{-\infty}^{\infty} \Big(1 - \frac{f_n^+(x)}{f_n^-(x)}\Big)^2 f_n^-(x)\, dx \\ = \quad & \int_{s_n^{-\beta_0}/\beta_0}^{s_n^{-\beta_0} r_n} \Big(1 - \frac{(\gamma_n^- s_n^{\beta_0}(x - s_n^{-\beta_0}/\beta_0) + 1)^{1/\gamma_n^- + 1}}{(\gamma_n^+ s_n^{\beta_0}(x - s_n^{-\beta_0}/\beta_0) + 1)^{1/\gamma_n^+ + 1}} \Big)^2 \times \\ & \qquad\qquad \times s_n^{\beta_0+1} (\gamma_n^- s_n^{\beta_0} (x - s_n^{-\beta_0}/\beta_0) + 1)^{-(1/\gamma_n^- + 1)}\, dx \\ & + \int_{s_n^{-\beta_0} r_n}^{0} \Big(1 - \frac{(\chi_n^+)^{\gamma_n^+ + 1}}{(\chi_n^-)^{\gamma_n^- + 1}} (s_n^{\beta_0} x / r_n)^{1/\beta_n^- - 1/\beta_n^+} \Big)^2 \times \\ & \qquad\qquad \times s_n^{\beta_0+1} (\chi_n^-)^{\gamma_n^- + 1} (s_n^{\beta_0} x/r_n)^{-(1/\beta_n^- + 1)}\, dx \\ = \quad & s_n \int_0^{r_n - 1/\beta_0} \Big(1 - \frac{(1+\gamma_n^- x)^{1/\gamma_n^- + 1}}{(1+\gamma_n^+ x)^{1/\gamma_n^+ + 1}} \Big)^2 (1+\gamma_n^- x)^{-(1/\gamma_n^- + 1)}\, dx \\ & + \frac{s_n \chi_n^-}{-\beta_n^-} \int_0^1 \Big(1 - \frac{\beta_n^- \chi_n^+}{\beta_n^+ \chi_n^-} x^{1/\beta_n^- - 1/\beta_n^+} \Big)^2 x^{-(1/\beta_n^- + 1)}\, dx \\ =: \quad & I_1 + I_2.\end{aligned}$$

Observe that $\gamma_n^\pm(r_n - 1/\beta_0) \to -1/(1+\tau) > -1$. Hence a Taylor expansion yields

$$\begin{aligned}\frac{(1+\gamma_n^- x)^{1/\gamma_n^-+1}}{(1+\gamma_n^+ x)^{1/\gamma_n^++1}} &= \Big(1 + \frac{(\gamma_n^- - \gamma_n^+)x}{1+\gamma_n^+ x}\Big)^{1/\gamma_n^++1}(1+\gamma_n^- x)^{1/\gamma_n^- - 1/\gamma_n^+} \\ &= 1 + O(|\gamma_n^+ - \gamma_n^-|) \;=\; 1 + O(b_n)\end{aligned}$$

uniformly for $x \in [0, r_n - 1/\beta_0]$. As a consequence we obtain $I_1 = O(s_n b_n^2)$.

Furthermore, direct calculations show that

$$I_2 = \frac{s_n}{\chi_n^-}\Big((\chi_n^- - \chi_n^+)^2 - (\chi_n^+)^2 \frac{(\beta_n^- - \beta_n^+)^2}{\beta_n^+(\beta_n^+ - 2\beta_n^-)}\Big) = O(s_n b_n^2),$$

where for the last identity

$$\begin{aligned}\chi_n^- - \chi_n^+ &= (1+1/\tau)^{1/\gamma_n^-}\Big(\Big(\frac{\beta_n^+}{\beta_n^-}\Big)^{1/\gamma_n^-} - (1+1/\tau)^{1/\gamma_n^+ - 1/\gamma_n^-}\Big) \\ &= O(|\beta_n^+ - \beta_n^-| + |\gamma_n^+ - \gamma_n^-|) \;=\; O(b_n)\end{aligned}$$

has been utilized.

To sum up, we have proven that

$$\int_{-\infty}^{\infty}\Big(1 - \frac{f_n^+(x)}{f_n^-(x)}\Big)^2 f_n^-(x)\,dx = O(s_n b_n^2) = O(n^{-1}).$$

Now one can conclude the asssertion as in the proof of Theorem 4.1, since

$$\frac{\beta_n^+ - \beta_n^-}{2g(t_n)} = \frac{b_n}{g(t_n)} \longrightarrow \frac{1-\iota}{1+\tau}\Big(\Big((1+1/\tau)^{1/\beta_0} - \iota\Big)\lambda\Big)^\rho,$$

which can be made arbitrarily large in the case $\rho > 0$ by letting $\lambda \to \infty$, and arbitrarily close to 1 in case of $\rho = 0$ by choosing ι and τ sufficiently small. □

Remark 4.2 In view of the proof of Theorem 4.1, one is tempted to make the simple approach

$$((F_n^\pm)^{-1})'(1-t) := t^{-(\beta_n^\pm+1)}\exp\Big(\int_t^1 \frac{(\beta_0 - \beta_n^\pm)1_{[s_n,1]}(s)}{s}\,ds\Big).$$

However, since then $\int_0^1((F_n^+)^{-1})'(t)\,dt \neq \int_0^1((F_n^-)^{-1})'(t)\,dt$ for $\beta_0 < 0$, the d.f.'s F_n^+ and F_n^- have different supports. Therefore, the χ^2–'distance' between f_n^+ and f_n^- is of larger order than n^{-1} if $\beta_0 < -1/2$. □

Proof of Theorem 4.5. Since the arguments are very similar to those used to verify Theorem 4.3, we just give an outline of the proof. Let

$$\mu_F(t,x) := \begin{cases} \int_x^1 s^{-(\beta+1)}\int_s^1 \varepsilon(ut)/u\,du\,ds & \text{if } \beta \geq -1/2, \\ -\int_0^x s^{-(\beta+1)}\int_s^1 \varepsilon(ut)/u\,du\,ds & \phantom{\text{if }} \beta < -1/2.\end{cases}$$

Then

$$\sup_{F\in\mathcal{F}_2}\sup_{x\in(0,x_0]} x^{\beta+\iota}\Big|\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{t(F^{-1})'(1-t)} - \Big(\int_0^1 s^{-(\beta+1)}\exp\Big(\int_s^1\frac{\varepsilon(ut)}{u}\,du\Big)\,ds - \frac{1}{\beta}\Big)\cdot 1_{\{\beta<-1/2\}} - F_\beta^{-1}(x) - \mu_F(t,x)\Big| = O(g^2(t))$$

for all $x_0,\iota>0$, from which one can deduce that

$$\sup_{F\in\mathcal{F}_2}\Big\|\frac{\tilde{Q}_{n,k_n,F}-\tilde{D}_{n,k_n,F}}{k_n/n(F^{-1})'(1-k_n/n)} - \Big(F_\beta^{-1} - t^{-(\beta+1)}k_n^{-1}W(k_nt)+\mu_F(k_n/n,t)\Big)_{t\in[0,1]}\Big\|_{\beta,h} = o\Big(k_n^{-1/2}+g(k_n/n)\Big),$$

where $\tilde{D}_{n,k_n,F}$ denotes a suitable real constant if $\beta\ge -1/2$, and it equals a sum of a constant and $\tilde{Q}_{n,k_n,F}(0+)$ for $\beta<-1/2$.

Next note that for $\beta\ge -1/2$

$$\begin{aligned}
T_\beta'(\mu_F(k_n/n,\cdot)) &= \int_{[0,1]}\int_t^1 s^{-(\beta+1)}\int_s^1\frac{\varepsilon(uk_n/n)}{u}\,du\,ds\,\nu_{T,\beta}(dt)\\
&= \int_0^1\frac{\varepsilon(uk_n/n)}{u}\int_{[0,u]}\int_t^u s^{-(\beta+1)}\,ds\,\nu_{T,\beta}(dt)\,du\\
&= \int_0^1\frac{\varepsilon(uk_n/n)}{u}\int_{[0,u]}\bar{F}_\beta^{-1}(t)-\bar{F}_\beta^{-1}(u)\,\nu_{T,\beta}(dt)\,du\\
&= \int_0^1\frac{\varepsilon(uk_n/n)}{u}\int_{[0,1]}\bar{F}_\beta^{-1}(\min(t,u))\,\nu_{T,\beta}(dt)\,du,
\end{aligned}$$

where in the last step $\nu_{T,\beta}[0,1]=0$ has been used. Likewise, if $\beta<-1/2$, then

$$\begin{aligned}
T_\beta'(\mu_F(k_n/n,\cdot)) &= -\int_0^1\frac{\varepsilon(uk_n/n)}{u}\int_{[0,1]}\int_0^{\min(t,u)} s^{-(\beta+1)}\,ds\,\nu_{T,\beta}(dt)\,du\\
&= \int_0^1\frac{\varepsilon(uk_n/n)}{u}\int_{[0,1]}\bar{F}_\beta^{-1}(\min(t,u))\,\nu_{T,\beta}(dt)\,du.
\end{aligned}$$

Now the proof can be finished in the same way as before, since for $\beta\ge -1/2$

$$\begin{aligned}
&\int_0^1 u^{-1}\int_{[0,1]}\bar{F}_\beta^{-1}(\min(t,u))\,\nu_{T,\beta}(dt)\,du\\
&= \int_{[0,1]}\int_t^1 s^{-(\beta+1)}\int_s^1 u^{-1}\,du\,ds\,\nu_{T,\beta}(dt)\\
&= \left\{\begin{array}{ll} -\int_{[0,1]}\bar{F}_\beta^{-1}\cdot\log + F_\beta^{-1}/\beta\;d\nu_{T,\beta} & \\ & \text{if} \\ -\int_{[0,1]}\bar{F}_0^{-1}\cdot\log/2\;d\nu_{T,0} & \end{array}\right.\begin{array}{l}\beta\ne 0,\\ \\ \beta=0\end{array} = 1,
\end{aligned}$$

according to Lemma 2.2, and, likewise, for $\beta<-1/2$

$$\int_0^1 u^{-1}\int_{[0,1]}\bar{F}_\beta^{-1}(\min(t,u))\,\nu_{T,\beta}(dt)\,du = 1.$$

□

References

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

BALKEMA, A.A. AND DE HAAN, L. (1974). Residual life time at great age. *Ann. Probab.* **2**, 792–804.

BEIRLANT, J., TEUGELS, J.L. AND VYNCKIER, P. (1996a). *Practical Analysis of Extreme Values.* Leuven University Press, Leuven.

BEIRLANT, J., TEUGELS, J.L. AND VYNCKIER, P. (1996b). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.* **91**, 1659–1667.

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

BINGHAM, N.H., GOLDIE, C.M. AND TEUGELS, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

CASTILLO, E. (1988). *Extreme Value Theory in Engineering.* Academic Press, San Diego.

CASTILLO, E. (1994). Extremes in engineering applications. In: *Extreme Value Theory and Applications I* (J. Galambos, J. Lechner and E. Simiu, eds.), 15–42. Kluwer, Dordrecht.

CSÖRGŐ, M. AND HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley, Chichester.

CSÖRGŐ, S., DEHEUVELS, P. AND MASON, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050–1077.

CSÖRGŐ, S. AND VIHAROS, L. (1995). On the asymptotic normality of Hill's estimator. *Math. Proc. Cambridge Philos. Soc.* **118**, 375–382.

DANIELSSON, J., DE HAAN, L., PENG, L. AND DE VRIES, C.G. (1996). Using a bootstrap method to choose the sample fraction in tail index estimation. Preprint, Econometric Institute, Erasmus University Rotterdam.

DAVIS, R.A. AND RESNICK, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12**, 1467–1487.

DEKKERS, A.L.M., EINMAHL, J.H.J. AND DE HAAN, L. (1989). A moment estimator for the index of an extreme value distribution. *Ann. Statist.* **17**, 1833–1855.

DEKKERS, A.L.M. AND DE HAAN, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17**, 1795–1832.

DEKKERS, A.L.M. AND DE HAAN, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal.* **47**, 173–195.

DIEUDONNÉ, J. (1970). *Treatise on Analysis II.* Academic Press, New York.

DONOHO, D.L. AND LIU, R.C. (1987). Geometrizing rates of convergence, I. Technical Report 137, Dept. Statist., University of California, Berkeley.

DONOHO, D.L. AND LIU, R.C. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19**, 633–667.

DREES, H. (1995). Refined Pickands estimators of the extreme value index. *Ann. Statist.* **23**, 2059–2080.

DREES, H. (1996). Refined Pickands estimators with bias correction. *Comm. Statist. Theory Methods* **25**, 837–851.

DREES, H. (1998a). On smooth statistical tail functionals. To appear in *Scand. J. Statist.*

DREES, H. (1998b). A general class of estimators of the extreme value index. *J. Statist. Plann. Inference* **66**, 95–112.

DREES, H. (1998c). Optimal rates of convergence for estimates of the extreme value index. To appear in *Ann. Statist.* **26** (1).

DREES, H. AND KAUFMANN, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. To appear in *Stochastic Processes Appl.*

EINMAHL, J.H.J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *J. Statist. Plann. Inference* **32**, 137–145.

EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997). *Modelling Extremal Events.* Springer, Berlin.

FALK, M. (1995). LAN of extreme order statistics. *Ann. Inst. Statist. Math.* **47**, 693–717.

FALK, M., HÜSLER, J. AND REISS, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser, Basel.

FARRELL, R.H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43**, 170–180.

FERNHOLZ, L.T. (1983). *Von Mises Calculus for Statistical Functionals.* Lecture Notes in Statistics **19**. Springer, New York.

FISHER, R.A. AND TIPPETT, L.H.C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Cambridge Phil. Soc.* **24**, 180–190.

GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

GELUK, J. AND DE HAAN, L. (1987) *Regular Variation, Extensions and Tauberian Theorems.* CWI tracts **40**. CWI, Amsterdam.

GELUK, J., DE HAAN, L., RESNICK, S. AND STĂRICĂ, C. (1996). Second order regular variation, convolution and the central limit theorem. Preprint, Erasmus University Rotterdam and Cornell University.

GILL, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1). *Scand. J. Statist.* **16**, 97–128.

GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423–453.

DE HAAN, L. (1975). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes.* Math. Centre Tracts **32**. Mathematisch Centrum, Amsterdam.

DE HAAN, L. (1984). Slow variation and characterization of domains of attraction. In: *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.), 31–48. Reidel Publishing, Dordrecht.

DE HAAN, L., PENG, L. AND PEREIRA, T.T. (1997). A bootstrap-based method to achieve optimality in estimating the extreme-value index. Preprint, Erasmus University Rotterdam and University of Lisbon.

DE HAAN, L. AND RESNICK, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Probab.* **24**, 97–124.

DE HAAN, L. AND ROOTZÉN, H. (1993). On the estimation of high quantiles. *J. Statist. Plann. Inference* **35**, 1–13.

DE HAAN, L. AND STADTMÜLLER, U. (1996). Generalized regular variation of second order. *J. Austr. Math. Soc. Ser. A* **61**, 381–395.

HAEUSLER, E. AND TEUGELS, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.* **13**, 743–756.

HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37–42.

HALL, P. (1989). On convergence rates in nonparametric problems. *Internat. Statist. Rev.* **57**, 45–58.

HALL, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177–203.

HALL, P. AND WELSH, A.H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12**, 1079–1084.

HALL, P. AND WELSH, A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331–341.

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

HOGG, R.V. AND KLUGMAN, S.A. (1984). *Loss Distributions.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

HOSKING, J.R.M. AND WALLIS, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.

KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1975). An approximation of partial sums of independent rv-'s, and the sample df. I. *Probab. Theory Related Fields* **32**, 111–131.

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent rv-'s, and the sample df. II. *Probab. Theory Related Fields* **34**, 33–58.

Lepskii, O.V. (1991). Asymptotic minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36**, 682–697.

Lepskii, O.V., Mammen, E. and Spokoiny, V.G. (1997). Optimal spatial adaption to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25**, 929–947.

Marohn, F. (1995). *Contributions to a Local Approach in Extreme Value Statistics.* Habilitation thesis, University of Eichstätt.

Marohn, F. (1997). Local asymptotic normality in extreme value index estimation. To appear in *Ann. Inst. Statist. Math.*

Mason, D.M. and Turova, T.S. (1994). Weak convergence of the Hill estimator process. In: *Extreme Value Theory and Applications I* (J. Galambos, J. Lechner and E. Simiu, eds.), 419–431. Kluwer, Dordrecht.

Omey, E. and Willekens, E. (1988). Π-variation with remainder. *J. London Math. Soc.* **37**, 105–118.

Pereira, T.T. (1994). Second order behaviour of domains of attraction and the bias of generalized Pickands' estimator. In: *Extreme Value Theory and Applications III* (J. Galambos, J. Lechner and E. Simiu, eds.), 165–177. NIST special publication 866.

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

Ramlau-Hansen, H. (1988). A solvency study in non-life insurance. *Scand. Actuar. J.*, 3–34.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics.* Springer Series in Statistics. Springer, New York.

Reiss, R.-D. (1993). *A Course on Point Processes.* Springer Series in Statistics. Springer, New York.

Reiss, R.-D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values.* Birkhäuser, Basel.

Resnick, S. and Stărică, C. (1997). Smoothing the Hill estimator. *Adv. Appl. Probab.* **29**, 271–293.

Rieder, H. (1994). *Robust Asymptotic Statistics.* Springer Series in Statistics. Springer, New York.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

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