*The Annals of Statistics*
1995, Vol. 23, No. 6, 2059–2080

# REFINED PICKANDS ESTIMATORS OF THE EXTREME VALUE INDEX

BY HOLGER DREES

*University of Cologne*

Consider a distribution function that belongs to the weak domain of attraction of an extreme value distribution. The extreme value index $\beta$ will be estimated by mixtures of Pickands estimators, where the weights are generated by a probability measure which satisfies a certain integrability condition. We prove a functional limit theorem for a process of Pickands estimators and asymptotic normality of the refined Pickands estimator. For negative $\beta$ the new estimator is asymptotically superior to previously defined estimators. A simulation study also demonstrates the good small-sample performance. In particular, the estimator proves to be robust against an inappropriate choice of the number of upper order statistics used for estimation.

**1. Introduction.** Let $X_i$, $i \in \mathbb{N}$, be a sequence of i.i.d. random variables (r.v.'s) with common distribution function (d.f.) $F$. Assume that $F$ belongs to the weak domain of attraction of an extreme value d.f. $G$ [in short, $F \in D(G)$], that is,

$$\mathscr{L}\Big(a_n^{-1}\Big(\max_{1\le i\le n} X_i - b_n\Big)\Big) \to G \quad \text{weakly} \tag{1.1}$$

for some normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$. Note that, subsequently, we do not distinguish between a d.f. and the pertaining distribution. Up to a location and scale parameter, $G$ has to be one of the following extreme value d.f.'s (given in von Mises representation).

$$G_\beta(x) := \begin{cases} \exp\big(-(1+\beta x)^{-1/\beta}\big), & \text{if } 1+\beta x > 0,\ \beta \neq 0, \\ \exp(-\exp(-x)), & \text{if } \beta = 0. \end{cases}$$

There is a rich literature about the estimation of the so-called extreme value index (or tail index) $\beta$ based on $X_i$, $1 \le i \le n$ [see, e.g., Hill (1975), Hosking and Wallis (1987), Smith (1987), Dekkers and de Haan (1989) and Dekkers, Einmahl and de Haan (1989); for an introduction see Reiss (1989)]. Pickands (1975) proposed the estimator

$$\hat\beta_n(i) := \frac{1}{\log(2)}\log\left(\frac{X_{n-i+1:n} - X_{n-2i+1:n}}{X_{n-2i+1:n} - X_{n-4i+1:n}}\right),$$

Received December 1992; revised March 1995.

*AMS* 1991 *subject classifications.* Primary 62G05; secondary 62G20, 62G30.

*Key words and phrases.* Extreme value index, tail index, refined Pickands estimator, Pickands process, asymptotic normality, robustness, moment estimator.

where $i \leq n/4$ and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ are the order statistics pertaining to $(X_i)_{1 \leq i \leq n}$. The main advantage of Pickands' estimator is the consistency of $\hat{\beta}_n(m_n)$ for any $\beta \in \mathbb{R}$ and any intermediate sequence $m_n \to \infty$, $m_n/n \to 0$, in contrast to the Hill estimator and the probability weighted moment estimator proposed by Hosking and Wallis. Moreover, the Pickands estimator is easy to compute and invariant under shift and scale transformations as used in (1.1).

On the other hand it has a rather poor asymptotic efficiency. Furthermore, the Pickands estimator $\hat{\beta}_n(i)$ is very sensitive to the choice of the intermediate order statistics which are used for estimation: even a small alteration of $i$ can yield a considerable change of the estimate.

To overcome these disadvantages, we consider mixtures

$$\hat{\beta}_{n,\nu} := \sum_{i=1}^{m_n} c_{ni} \hat{\beta}_n(i) \tag{1.2}$$

of Pickands estimators, where $m_n$ is an intermediate sequence and the scores $c_{ni}$, $1 \leq i \leq m_n$, are generated by a probability measure $\nu$ on $\mathbb{B}(0,1]$, that is, $c_{ni} := \nu((i-1)/m_n, i/m_n]$, $\hat{\beta}_{n,\nu}$ will be addressed as a *refined Pickands estimator* with score generating probability measure $\nu$. Observe that the refined Pickands estimator has the representation

$$\hat{\beta}_{n,\nu} = \int \hat{\beta}_n(]m_n x[)\nu(dx),$$

where $]x[$ denotes the smallest integer greater than or equal to $x$. The special case of a mixture of two different Pickands estimators was studied by Falk (1994). In contrast to that, we merely impose an integrability condition on $\nu$.

The paper is organized as follows. Section 2 contains the mathematical results. We establish the asymptotic distribution of a jump process based on $m_n^{1/2}(\hat{\beta}_n(i) - \beta)$, $1 \leq i \leq m_n$, and of the normalized error $m_n^{1/2}(\hat{\beta}_{n,\nu} - \beta)$ of the refined estimator. Furthermore, for each $\beta \neq -\frac{1}{2}$ a score generating probability measure $\nu^*(\beta)$ is determined that minimizes the asymptotic variance of $m_n^{1/2}(\hat{\beta}_{n,\nu} - \beta)$. Finally, we investigate the adaptive estimator $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$, where $\tilde{\beta}_n$ is a weakly consistent estimator of $\beta$ and $\nu(\beta)$ is an almost optimal score generating probability measure.

In Section 3 we compute the asymptotic relative efficiency (with respect to $\hat{\beta}_{n,\nu^*(\beta)}$) of well-known estimators that are based on the same number $k_n$ of upper order statistics. A simulation study demonstrates the good finite-sample behavior of the refined Pickands estimator. Particularly, the estimator proves to be less sensitive to an inappropriate choice of $k_n$ than the moment estimator of Dekkers, Einmahl and de Haan (1989). All proofs are collected in Section 4.

**2. Asymptotic normality of refined Pickands estimators.** Like the Pickands estimator many popular estimators of the extreme value index are based on a certain fraction of upper-order statistics $(X_{n-i+1:n})_{1 \leq i \leq k_n}$, where

$k_n$ denotes an intermediate sequence. It is well known that asymptotic normality of, for example, the Hill estimator or the moment estimator of Dekkers, Einmahl and de Haan (see Section 3 for the definitions) is obtained only if $k_n$ does not tend to infinity too fast. For a rigorous formulation of these conditions on $k_n$, one has to examine the so-called second-order behavior of the underlying d.f. $F$.

Recently, several different sets of second-order conditions on $F$ have been proposed [see, e.g., Smith (1987), Dekkers and de Haan (1989, 1993) and Dekkers, Einmahl and de Haan (1989)]. The most general approach was made by Pereira (1994), who assumed that there are functions $a, \Phi\colon (0,1) \to (0,\infty)$ and $\Psi\colon (0,\infty) \to \mathbb{R}$ such that

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

where $\Phi(t) = o(1)$ and $R(t,x) = o(\Phi(t))$ as $t \downarrow 0$. [By convention, $(x^{-\beta} - 1)/\beta := -\log(x)$ if $\beta = 0$.]

Observe that (2.1) implies

$$\frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)/(1+c\Phi(t))} = \frac{x^{-\beta}-1}{\beta} + \Phi(t)\left(\Psi(x) + c\frac{x^{-\beta}-1}{\beta}\right) + R(t,x).$$

Hence, w.l.o.g. one may assume that $\Psi \equiv 0$ or that $\Psi$ is not a multiple of $x \mapsto (x^{-\beta} - 1)/\beta$. This leads to the following condition.

CONDITION 2.1. Assume that the expansion (2.1) holds for some measurable, locally bounded functions $a$, $\Phi$ and $\Psi$, where (i) or (ii) holds:

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

Notice that Condition 2.1(i) is equivalent to $F \in D(G_\beta)$ [de Haan (1984), Lemma 1]. Essentially, the second-order Condition 2.1(ii) is the condition considered by de Haan and Stadtmüller (1993), who also examined the relationship to the conditions of Dekkers and de Haan (1993). They proved that, under Condition 2.1(ii), $\Phi$ is $\delta$-varying for some $\delta \geq 0$, that is,

$$\lim_{t\downarrow 0} \frac{\Phi(tx)}{\Phi(t)} = x^{\delta},$$

for all $x > 0$, and that $\Psi$ is of the form

$$\Psi(x) = c_1\int_1^x s^{-(\beta+1)}\int_1^s u^{\delta-1}\,du\,ds + c_2\int_1^x s^{\delta-(\beta+1)}\,ds,$$

for some real constants $c_1$ and $c_2$. In particular, $\Psi$ is differentiable. Moreover, the relations $R(t,x) = o(1)$ and $R(t,x) = o(\Phi(t))$, respectively, hold locally uniformly [Dekkers and de Haan (1989), Lemma 2.2; de Haan and Stadtmüller

(1993), Remark 4(ii)]. Many examples of d.f.'s satisfying Condition 2.1(ii) can be found in Pereira (1994).

EXAMPLE 2.1. Assume that the quantile function (q.f.) $F^{-1}$ is of the following type:

$$F^{-1}(1-t) = \mu + \sigma \begin{cases} t^{-\beta} + Lt^{\delta-\beta} + o(t^{\delta-\beta}), & \text{if } \beta \neq 0, \\ -\log(t) + Lt^{\delta} + o(t^{\delta}), & \text{if } \beta = 0, \end{cases} \tag{2.2}$$

as $t \downarrow 0$ for some $\delta > 0$, $\delta \neq \beta$. Then Condition 2.1(ii) holds with $\Phi(t) = t^{\delta}$ and $\Psi(x) = c(x^{\delta-\beta} - 1)$ for some constant $c \in \mathbb{R} \setminus \{0\}$. Examples of this type of d.f.'s are the extreme value d.f.'s $G_{\beta}$ ($\delta = 1$), the Cauchy distribution ($\beta = 1$, $\delta = 2$) and the logistic d.f. ($\beta = 0$, $\delta = 1$).

Note that the leading term $t^{-\beta}$ and $-\log(t)$, respectively, of the expansion (2.2) is the q.f. of a generalized Pareto distribution. Hence the underlying d.f. belongs to a certain neighborhood of a generalized Pareto d.f. with location and scale parameter. For $\beta > 0$ expansion (2.2) is equivalent to the well-known Hall condition [Hall (1982)]. Moreover, it is closely related to a condition which was partly introduced by Weiss (1971) and investigated in detail by Falk (1985), Falk and Reiss (1992), Falk and Marohn (1993) and Kaufmann (1994), among others.

Corresponding to the two sets of conditions on $F$, we assume two conditions that describe an upper bound on the rate at which the intermediate sequence $m_n$ tends to infinity. Recall that $m_n$ determines the fraction of upper order statistics that is involved in estimation of $\beta$.

CONDITION 2.2. The sequence $(m_n)_{n \in \mathbb{N}}$ is an intermediate sequence such that (i) or (ii) holds:

(i) $\lim_{n\to\infty} m_n^{1/2} \sup_{t \in (0, m_n/n],\, x \in [1-\eta, 4+\eta]} |R(t,x)| = 0$ for some $\eta > 0$;
(ii) $\lim_{n\to\infty} m_n^{1/2} \Phi(m_n/n) = \lambda \in [0, \infty)$.

If the second order Condition 2.1(ii) is satisfied, then Condition 2.2(i) applied to the "remainder term" $\Phi(t)\Psi(x) + R(t,x)$ reads as $\lim_{n\to\infty} m_n^{1/2}\Phi(m_n/n) = 0$. Thus in this case Condition 2.2(ii) (with $\lambda > 0$) allows $m_n$ to converge faster to infinity than Condition 2.2(i) does.

First we establish a functional limit theorem for the jump process

$$Z_{n,\beta}(t) := \sum_{i=1}^{m_n} m_n^{1/2}\big(\hat{\beta}_n(i) - \beta\big) 1_{((i-1)/m_n,\, i/m_n]}(t), \qquad t \in [0,1],$$

that may be addressed as a Pickands process. Observe that

$$m_n^{1/2}\big(\hat{\beta}_{n,\nu} - \beta\big) = \int Z_{n,\beta}(t)\,\nu(dt). \tag{2.3}$$

Define

$$c_\beta := \begin{cases} \dfrac{2\beta}{\log(2)(1-2^{-\beta})}, & \text{if } \beta \neq 0, \\ \dfrac{2}{\log^2(2)}, & \text{if } \beta = 0, \end{cases}$$

$$d_{\beta,\Psi} := \frac{1}{2}\left(-(1+2^{\beta})\Psi(2) + 2^{\beta}\Psi(4)\right)$$

and a Gaussian process

$$Z_\beta(t) := \frac{1}{t}\left(W\left(\frac{t}{4}\right) - \frac{1+2^{-\beta}}{2}W\left(\frac{t}{2}\right) + 2^{-(\beta+2)}W(t)\right), \qquad t \geq 0,$$

where $W$ denotes a standard Brownian motion.

THEOREM 2.1. *Let $h$: $[0,1] \to [0,\infty)$ be a continuous function such that*

$$\lim_{t\downarrow 0} h(t)\left(\frac{\log\log(3/t)}{t}\right)^{1/2} = 0. \tag{2.4}$$

*Assume that Conditions* 2.1(i) *and* 2.2(i) *or Conditions* 2.1(ii) *and* 2.2(ii) *are satisfied and set $\lambda = 0$ in the first case. Then*

$$\left(h(t)\cdot Z_{n,\beta}(t)\right)_{t\in[0,1]} \to \left(h(t)\cdot c_\beta\left(Z_\beta(t) + \lambda d_{\beta,\Psi}t^{\delta}\right)\right)_{t\in[0,1]} \quad \textit{weakly}$$

*in the Skorohod space $D[0,1]$.*

As an easy consequence one obtains asymptotic normality of the refined Pickands estimator.

COROLLARY 2.1. *If the score generating probability measure $\nu$ satisfies*

$$\int\left(\frac{\log\log(3/t)}{t}\right)^{1/2}\nu(dt) < \infty \tag{2.5}$$

*and the conditions of Theorem* 2.1 *hold, then*

$$\mathscr{L}\left(m_n^{1/2}\left(\hat{\beta}_{n,\nu} - \beta\right)\right) \to \mathscr{N}\left(\lambda c_\beta d_{\beta,\Psi}\int t^{\delta}\nu(dt), c_\beta^2\sigma_{\beta,\nu}^2\right) \quad \textit{weakly}, \tag{2.6}$$

*where $\sigma_{\beta,\nu}^2 := \int\sigma_\beta(s,t)\nu^2(ds,dt)$ and $\sigma_\beta(s,t) :=$* Cov$(Z_\beta(s), Z_\beta(t))$.

If the score generating probability measure has no mass in a neighborhood of 0, then an upper bound on the rate can be established at which asymptotic normality holds. For simplicity we restrict ourselves to the case where the estimator is asymptotically unbiased. (However, see the remark following the proof of Theorem 2.2.)

THEOREM 2.2. *If $\nu(0,\varepsilon] = 0$ for some $\varepsilon > 0$ and Conditions* 2.1(i) *and* 2.2(i) *are satisfied, then, for all $\eta > 0$*

$$\sup_{x\in\mathbb{R}}\left|P\left\{m_n^{1/2}\left(\hat{\beta}_{n,\nu}-\beta\right)\le x\right\}-\mathscr{N}\left(0,c_\beta^2\sigma_{\beta,\nu}^2\right)(-\infty,x]\right|$$
$$=O\left(\frac{m_n}{n}+\log(m_n)m_n^{-1/2}+m_n^{1/2}\sup_{t\in[\varepsilon,1],\,x\in[1-\eta,4+\eta]}\left|R\left(\frac{tm_n}{n},x\right)\right|\right).$$

Note that (2.6) includes the asymptotic normality of the Pickands estimator $\hat{\beta}_n(m_n)=\hat{\beta}_{n,\varepsilon_1}$ as a special case; this result was already established by Dekkers and de Haan (1989) under more restrictive conditions on $F$. In addition, we obtain from Theorem 2.2 the rate of convergence to $\mathscr{N}(0,c_\beta^2\sigma_\beta(1,1))$ in that case.

Observe that the asymptotic variance $c_\beta^2\sigma_{\beta,\nu}^2$ of $m_n^{1/2}(\hat{\beta}_{n,\nu}-\beta)$ depends on the score generating probability measure. Our next aim is to determine a probability measure $\nu^*(\beta)\in\mathscr{M}:=\{\nu:\ \nu$ satisfies (2.5)$\}$ that minimizes $\sigma_{\beta,\nu}^2$ and hence the asymptotic mean squared error if the estimator is asymptotically unbiased. In particular, this is true if Conditions 2.1(i) and 2.2(i) are satisfied.

Check that, for $s\le t$,

$$\sigma_\beta(s,t)=\frac{1}{st}\begin{cases}0, & \text{if } 0\le s\le \dfrac{t}{4},\\[2mm] -2^{-(\beta+2)}\left(s-\dfrac{t}{4}\right), & \text{if } \dfrac{t}{4}<s\le\dfrac{t}{2},\\[2mm] \dfrac{1+2^{-(\beta+1)}+2^{-(2\beta+1)}}{4s} & \\[2mm] \quad-\dfrac{(1+2^{-\beta}+2^{-(2\beta+1)})}{8t}, & \text{if } \dfrac{t}{2}<s\le t.\end{cases}$$

Since $\sigma_{-(\beta+1)}(s,t)=2^{2\beta+1}\sigma_\beta(s,t)$ for all $s,t\in(0,1]$, $\nu^*(\beta)$ minimizes $\sigma_{\beta,\nu}^2$ if and only if it minimizes $\sigma_{-(\beta+1),\nu}^2$. Hence we may restrict our attention to $\beta\ge-\frac{1}{2}$.

LEMMA 2.1. *For $\beta\ge-\frac{1}{2}$,*

$$\inf_{\nu\in\mathscr{M}}\sigma_{\beta,\nu}^2=\frac{\left(1-2^{-(\beta+1)}\right)^2}{16}. \tag{2.7}$$

*For $\beta>-\frac{1}{2}$, the infimum is attained at*

$$\nu^*(\beta):=\sum_{i=0}^{\infty}a_{i,\beta}\varepsilon_{2^{-i}},$$

*where*

$$a_{i,\beta} := \begin{cases} \dfrac{2^{\beta+1}-1}{2^{\beta}-1}(1-2^{-(i+1)\beta})2^{-(i+2)}, & \text{if } \beta \neq 0, \\ (i+1)2^{-(i+2)}, & \text{if } \beta = 0. \end{cases}$$

Regrettably, we do not find any probability measure which minimizes $\sigma^2_{-1/2,\nu}$. Therefore we modify the asymptotically optimal probability measure for parameters $\beta$ on a small neighborhood of $-\frac{1}{2}$:

$$(2.8)\qquad \nu(\beta) := \begin{cases} \nu^*(-(\beta+1)), & \text{if } \beta < -\frac{1}{2}-\rho, \\ \nu^*(-\frac{1}{2}+\rho), & \text{if } |\beta+\frac{1}{2}| \leq \rho, \\ \nu^*(\beta), & \text{if } \beta > -\frac{1}{2}+\rho, \end{cases}$$

for some $\rho > 0$. Then for all $\varepsilon > 0$ there is a $\rho > 0$ such that

$$\sup_{\beta\in\mathbb{R}} \left| \inf_{\nu\in\mathscr{M}} \sigma^2_{\beta,\nu} - \sigma^2_{\beta,\nu(\beta)} \right| < \varepsilon.$$

Note that the (almost) optimal score generating probability measure $\nu(\beta)$ depends on the parameter which has to be estimated. Thus it suggests itself to introduce an adaptive estimator. It turns out that the estimator

$$\hat{\beta}_{\nu,\nu(\tilde{\beta}_n)} = \int \hat{\beta}_n(]m_n t[)\nu(\tilde{\beta}_n)(dt)$$

has the same asymptotic performance as the refined Pickands estimator whose score generating measure is based on the actual parameter $\beta$ if $\tilde{\beta}_n$ is a weakly consistent estimator of $\beta$.

THEOREM 2.3. *If Conditions* 2.1(i) *and* 2.2(i) *are satisfied and* $\tilde{\beta}_n$ *is a weakly consistent estimator of* $\beta$, *then*

$$(2.9)\qquad \mathscr{L}\Big(m_n^{1/2}\big(\hat{\beta}_{n,\nu(\tilde{\beta}_n)}-\beta\big)\Big) \to \mathscr{N}\big(0, c_\beta^2\sigma^2_{\beta,\nu(\beta)}\big) \quad \textit{weakly}.$$

REMARK. In a similar way one can determine an asymptotically optimal score generating probability measure $\nu^*$ (i.e., a measure which minimizes the asymptotic mean squared error) if Conditions 2.1(ii) and 2.2(ii) hold for some $\lambda > 0$. In this case, however, $\nu^*$ also depends on $\lambda$, $d_{\beta,\Psi}$ and $\delta$. Since these parameters are very difficult to estimate, an adaptive estimator would be of little practical use.

Besides the asymptotic normality of the refined Pickands estimator, one may investigate the asymptotic behavior of further functionals of the Pickands process, such as a weighted median or test statistics $\int Z^2_{n,\beta}(t)\nu(dt)$ of Cramér–von Mises type. For details we refer to Drees (1993).

**3. Comparison of estimators.** In this section we compare the adaptive estimator $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ with the Pickands estimator, the moment estimator proposed by Dekkers, Einmahl and de Haan (1989) and Hill's estimator [Hill (1975)]. Recall that the latter is defined by

$$\hat{\beta}_n^H := (k_n - 1)^{-1} \sum_{i=1}^{k_n - 1} \log\left( \frac{X_{n-i+1:n}}{X_{n-k_n+1:n}} \right)$$

for some intermediate sequence $k_n$. Because the Hill estimator is inconsistent if $\beta < 0$, Dekkers, Einmahl and de Haan (1989) proposed the following modification:

$$\hat{\beta}_n^D := \hat{\beta}_n^H + 1 - \left( 2\left( 1 - \frac{\left(\hat{\beta}_n^H\right)^2}{M_n^{(2)}} \right) \right)^{-1},$$

where

$$M_n^{(2)} := (k_n - 1)^{-1} \sum_{i=1}^{k_n - 1} \left( \log\left( \frac{X_{n-i+1:n}}{X_{n-k_n+1:n}} \right) \right)^2.$$

Utilizing Corollary 2.1, one can determine an asymptotically optimal choice of the number $k_n$ of upper order statistics used by the refined Pickands estimator, that is, a number such that the asymptotic mean squared error is minimized. Such an optimal $k_n^*$ was established under suitable second-order conditions by Dekkers and de Haan (1993) for the moment estimator and, for example, by Goldie and Smith (1987) for the Hill estimator [see also Hall and Welsh (1985)]. However, $k_n^*$ depends on certain parameters that describe the second-order behavior of the underlying d.f. $F$. Since these parameters (e.g., $\Phi$ and $\Psi$ in the case of the refined Pickands estimator) are very difficult to estimate, the theoretical results about $k_n^*$ are of little practical use. [This was already noticed by Smith (1987), page 1182.] Moreover, under the second-order Conditions 2.1(ii), the quotient of the minimal mean squared errors of two estimators usually depends on $\beta$, $\delta$ and $\Psi$ so that it is a cumbersome measure of the asymptotic relative efficiency in this case.

For these reasons we confine ourselves to calculating the asymptotic relative efficiency if $k_n$ is sufficiently small such that all estimators under consideration are asymptotically unbiased. It turns out that in most cases for all three estimators this restriction leads to the same upper bound on the rate at which $k_n$ may tend to infinity. For example, under the Hall condition $1 - F(x) = cx^{-1/\beta}(1 + dx^{-\delta/\beta} + o(x^{-\delta/\beta}))$ for some $\delta > 0$, $\delta \neq \beta$ and $d \neq 0$, the moment estimator, Hill's estimator and the refined Pickands estimator are asymptotically unbiased if and only if $k_n = o(n^{-2\delta/(2\delta+1)})$. Hall and Welsh [(1984), Theorem 1] proved that in some sense this is the best attainable rate.

It is well known that

$$\mathscr{L}\left(k_n^{1/2}\left(\hat{\beta}_n^H - \beta\right)\right) \to \mathscr{N}(0, \beta^2) \quad \text{weakly} \tag{3.1}$$

if $\beta > 0$ and Conditions 2.1(i) and 2.2(i) hold with $a(t) = \beta F^{-1}(1-t)$ [see Goldie and Smith (1987)]. Dekkers, Einmahl and de Haan (1989) proved that

$$\mathscr{L}\Big(k_n^{1/2}\big(\hat{\beta}_n^D - \beta\big)\Big) \to \mathscr{N}\Big(0, \big(\sigma_\beta^D\big)^2\Big) \quad \text{weakly} \tag{3.2}$$

for intermediate sequences $k_n$ which tend to $\infty$ sufficiently slowly, where

$$(\sigma_\beta^D)^2 := \begin{cases} 1+\beta^2, & \text{if } \beta \geq 0, \\ (1-\beta)^2(1-2\beta) \\ \quad \times\left(4 - 8\dfrac{1-2\beta}{1-3\beta} + \dfrac{(5-11\beta)(1-2\beta)}{(1-3\beta)(1-4\beta)}\right), & \text{if } \beta < 0 \end{cases} \tag{3.3}$$

if the right endpoint of the underlying d.f. is positive (which can be achieved by a simple shift operation).

It is easily seen that for all $\varepsilon > 0$ there is a constant $\rho > 0$ such that the variance of the limiting normal distribution of $k_n^{1/2}(\hat{\beta}_{n,\nu(\tilde{\beta}_n)} - \beta)$ is less than $4c_\beta^2 \inf_{\nu \in \mathscr{M}} \sigma_{\beta,\nu}^2 + \varepsilon$. So it is reasonable to compare

$$4c_\beta^2 \inf_{\nu \in \mathscr{M}} \sigma_{\beta,\nu}^2 = \begin{cases} \left(\dfrac{\beta(1-2^{-(\beta+1)})}{\log(2)(1-2^{-\beta})}\right)^2, & \text{if } \beta \geq -\dfrac{1}{2}, \beta \neq 0, \\ \dfrac{1}{4\log^4(2)}, & \text{if } \beta = 0, \\ \dfrac{\beta^2}{2\log^2(2)}, & \text{if } \beta < -\dfrac{1}{2}, \end{cases}$$

with the asymptotic variances of the other estimators.

Figure 1 shows a plot of $4c_\beta^2 \inf_{\nu \in \mathscr{M}} \sigma_{\beta,\nu}^2$ divided by the asymptotic variances given in (3.1) (solid line) and (3.3) (broken line), respectively. The dotted line is the corresponding plot for the asymptotic variance of the

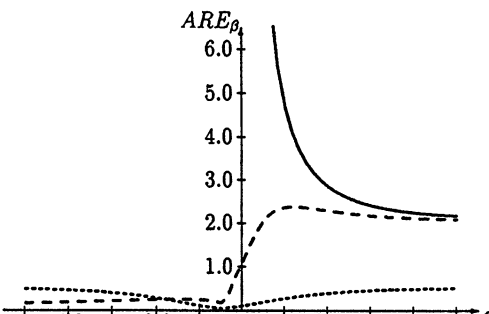


FIG. 1. *ARE of* $\hat{\beta}_n^H$ *(solid line),* $\hat{\beta}_n^D$ *(broken line) and* $\hat{\beta}_n([k_n/4])$ *(dotted line) w.r.t.* $\hat{\beta}_{n,\nu^*(\beta)}$.

Pickands estimator. Notice that, for $\beta \neq -\frac{1}{2}$, these ratios are the asymptotic relative efficiencies (ARE) with respect to $\hat{\beta}_{n,\nu^*(\beta)}$.

The plot clearly exhibits that $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ is the asymptotically best estimator for negative $\beta$, whereas for $\beta > 0$ the asymptotic variance of $\hat{\beta}_n^D$ is smaller. To be more precise, we have

$$\mathrm{ARE}_\beta\left(\hat{\beta}_n^D, \hat{\beta}_{n,\nu(\tilde{\beta}_n)}\right) \to 0 \qquad \text{as } \beta \to -\infty,$$

and

$$\mathrm{ARE}_\beta\left(\hat{\beta}_n^D, \hat{\beta}_{n,\nu(\tilde{\beta}_n)}\right) \to (\log(2))^{-2} \qquad \text{as } \beta \to \infty.$$

The ARE of the Pickands estimator w.r.t. $\hat{\beta}_{n,\nu^*(\beta)}$ converges to $\frac{1}{2}$ as $|\beta| \to \infty$. Furthermore, in the present situation the Hill estimator is the best estimator for $\beta > 0$ (yet not even consistent for $\beta < 0$). In view of Figure 1, it is natural to choose $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$, $\hat{\beta}_n^D$ or $\hat{\beta}_n^H$ for the estimation of $\beta$ according to an initial estimate of $\beta$. [For details see Drees (1993).]

Next, we want to examine the finite-sample behavior of $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$, $\hat{\beta}_n^D$ and $\hat{\beta}_n([k_n/4])$ by a simulation study. We choose $\tilde{\beta}_n = \hat{\beta}_{n,\nu(\bar{\beta}_n)}$ with $\bar{\beta}_n := \hat{\beta}_{n,\nu^*(0)}$ as the initial estimator. Furthermore, we put $\rho = 0.01$, but all "small" values for $\rho$ lead to almost the same results. (In practice one may even choose $\rho$ equal to 0.) In view of the definition of the moment estimator, it is obvious that this estimator works only if all order statistics used for estimation have the same sign. For this reason, in the case that the minimum of the observations is negative we shift the whole sample such that the minimum is equal to $\varepsilon = 0.001$. (The performance of the estimator depends on $\varepsilon$, particularly if a large fraction of the sample is utilized for estimation. We do not go into detail, because this is a specific feature of $\hat{\beta}_n^D$.)

The study is based on 10,000 Monte Carlo simulations. In each simulation we generate $n = 1000$ pseudo-r.v.'s according to a d.f. $F$ that satisfies Condition 2.1(ii). As already noted, the choice of the number $k_n$ of upper order statistics used for the estimation is crucial. We calculated the three estimators for $k_n = 50, 100, 200, \ldots, 900, 1000$. Then, for each underlying d.f. and each estimator, we determine the number $k_n$ for which the estimator shows

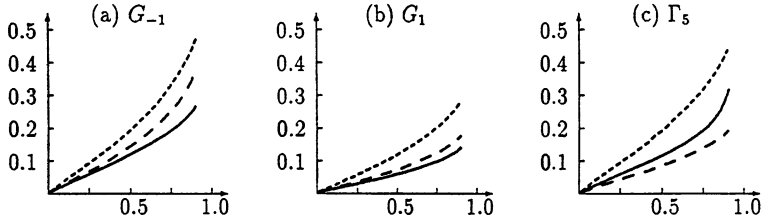


FIG. 2. *Empirical quantile functions of the absolute error of* $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ *(solid line),* $\hat{\beta}_n^D$ *(broken line) and* $\hat{\beta}_n([k_n/4])$ *(dotted line) with underlying d.f. F equal to* (a) $G_{-1}$, (b) $G_1$, (c) $\Gamma_5$, $n = 1000$ *and optimal* $k_n$.

the best performance. Figure 2 displays the empirical q.f.'s of the absolute error $|\bar{\beta}_n - \beta|$ for $\bar{\beta}_n = \hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ (solid line), $\hat{\beta}_n^D$ (broken line) and $\hat{\beta}_n([k_n/4])$ (dotted line).

In Figure 2a the underlying d.f. is the extreme value d.f. with parameter $-1$. In this case all three estimators under consideration perform best if they are based on the upper 200 order statistics. The plot demonstrates the superiority of the refined Pickands estimator over the moment estimator for negative $\beta$, which has already been indicated by Figure 1.

In contrast to this, at first glance it is surprising that the refined Pickands estimator yields slightly better results than the moment estimator in the case $F = G_1$. However, in this particular case where $\beta = \delta$ the second term of expansion (2.2) can be included in the location parameter $\mu$. Hence the location invariant estimator $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ may utilize more upper order statistics than $\hat{\beta}_n^D$. To be precise, the latter estimator shows the best performance for $k_n = 200$, whereas the best choice is $k_n = 800$ for the refined Pickands estimator and $k_n = 700$ for the Pickands estimator. (An asymptotically optimal sequence is $k_n \sim \lambda n^{2/3}$ for $\hat{\beta}_n^D$ and $k_n \sim \lambda^* n^{4/5}$ for $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$, where $\lambda$ and $\lambda^*$ are some positive constants.) It should be mentioned that the moment estimator outperforms the refined Pickands estimator for other underlying d.f.'s $F \in D(G_1)$ like the Cauchy d.f. (see below).

If the r.v.'s are distributed according to the gamma distribution with parameter 5, where $\beta = \delta = 0$ (Figure 2c), then for appropriately chosen $k_n$ [$k_n = 200$ for $\hat{\beta}_n^D$; $k_n = 300$ for $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ and $\hat{\beta}_n([k_n/4])$] the moment estimator shows the best behavior. However, the refined Pickands estimator performs much better than the original Pickands estimator, which proves to be the worst estimator under consideration for all three underlying d.f.'s.

Up to now, we have examined the behavior of the estimator if the number $k_n$ of upper order statistics used for estimation is taken appropriately. As already mentioned, in practice it is very difficult to determine such an optimal $k_n$ which depends on the unknown underlying d.f. For this reason, one important feature of an estimator of $\beta$ is the stability of its performance under changes of $k_n$.

Table 1 gives the median of the absolute error of the refined Pickands estimator $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ (upper lines) and the moment estimator (lower lines) for $k_n = 100, 200, 400, 600, 800$, and 1000 and several d.f.'s.

In all cases, due to the bias, the error of the moment estimator increases more rapidly than the error of the refined Pickands estimator as $k_n$ increases. In particular, for $F = G_\beta$, $\beta < 0$, the quality of the estimates given by $\hat{\beta}_n^D$ clearly deteriorates for $k_n > 500$. In contrast to this, $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ yields sensible results if one uses up to 800 order statistics or sometimes even the whole sample. (Note that for the generalized Pareto distribution $W_{-1/2}$, $k_n = 1000$ is the optimal choice!) Indeed, for almost all d.f.'s $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ is superior to $\hat{\beta}_n^D$ if one uses more than one-half of the sample. Thus the refined Pickands estimator is less sensitive to an inappropriate choice of $k_n$.

Finally, we want to investigate the stability of the estimates against small alterations of $k_n$. Figure 3 displays, for a single sample, the refined Pickands

TABLE 1

*Median of the absolute error of the refined Pickands estimator (upper lines) and the moment estimator (lower lines) [$W_\beta := 1 + \log(G_\beta)$ generalized Pareto d.f.; $C^*$ is the Cauchy d.f. restricted to the positive half-axis, $Wei_3(x) = 1 - \exp(-x^3)$; $\mathcal{N}(0,1)^*$ is the standard normal d.f. restricted to the positive half-axis; and $L(x) = (1 + \exp(-x))^{-1}$ is the logistic distribution]*

| $F$ | 100 | 200 | 400 | 600 | 800 | 1000 |
|---|---|---|---|---|---|---|
| $G_{-1}$ | 0.162 | 0.119 | 0.156 | 0.249 | 0.414 | 1.350 |
| | 0.160 | 0.143 | 0.296 | 0.594 | 1.289 | > 10 |
| $G_{-1/2}$ | 0.172 | 0.123 | 0.104 | 0.144 | 0.228 | 0.735 |
| | 0.102 | 0.087 | 0.171 | 0.349 | 0.765 | > 10 |
| $G_0$ | 0.146 | 0.091 | 0.072 | 0.106 | 0.162 | 0.358 |
| | 0.070 | 0.052 | 0.056 | 0.123 | 0.309 | > 10 |
| $G_1$ | 0.174 | 0.117 | 0.076 | 0.062 | 0.057 | 0.132 |
| | 0.097 | 0.071 | 0.079 | 0.138 | 0.232 | 6.350 |
| $W_{-1/2}$ | 0.170 | 0.122 | 0.091 | 0.070 | 0.071 | 0.069 |
| | 0.100 | 0.077 | 0.066 | 0.066 | 0.076 | 9.900 |
| $C^*$ | 0.174 | 0.113 | 0.079 | 0.081 | 0.131 | 0.211 |
| | 0.095 | 0.068 | 0.053 | 0.076 | 0.161 | 3.301 |
| $\mathrm{Wei}_3$ | 0.186 | 0.210 | 0.262 | 0.319 | 0.386 | 0.548 |
| | 0.189 | 0.231 | 0.330 | 0.479 | 0.796 | > 10 |
| $\Gamma_5$ | 0.146 | 0.110 | 0.129 | 0.168 | 0.220 | 0.381 |
| | 0.084 | 0.084 | 0.128 | 0.208 | 0.392 | 5.851 |
| $\mathcal{N}(0,1)^*$ | 0.161 | 0.157 | 0.189 | 0.217 | 0.243 | 0.269 |
| | 0.128 | 0.140 | 0.180 | 0.234 | 0.358 | > 10 |
| $L$ | 0.143 | 0.093 | 0.116 | 0.185 | 0.277 | 0.683 |
| | 0.072 | 0.068 | 0.150 | 0.304 | 0.668 | > 10 |

estimator (solid line), the moment estimator (broken line) and the Pickands estimator (dotted line) as a function of the number $k_n$ of upper order statistics used for estimation. The underlying d.f.'s are $F = G_1$ (Figure 3a) and $F = \Gamma_5$ (Figure 3b), respectively. Of course, these plots are of little statistical significance since they show the behavior of the estimators for just one realization. However, for practical purposes it is hard to estimate $\beta$ if the estimate changes a lot under small alterations of $k_n$. Hence some statisticians will prefer an estimator whose dependence on $k_n$ can be described by a smooth function. In particular, such a behavior may be helpful for an interactive data-driven choice of $k_n$ using plots like Figure 3, as has been suggested by, for example, Dekkers, Einmahl and de Haan (1989) or Falk, Hüsler and Reiss [(1994), Chapter 6].

The plots displayed here are in some sense typical. Usually the function pertaining to the moment estimator is the smoothest, in particular for $100 \le k_n \le 400$. In contrast to this, using the Pickands estimator, even small

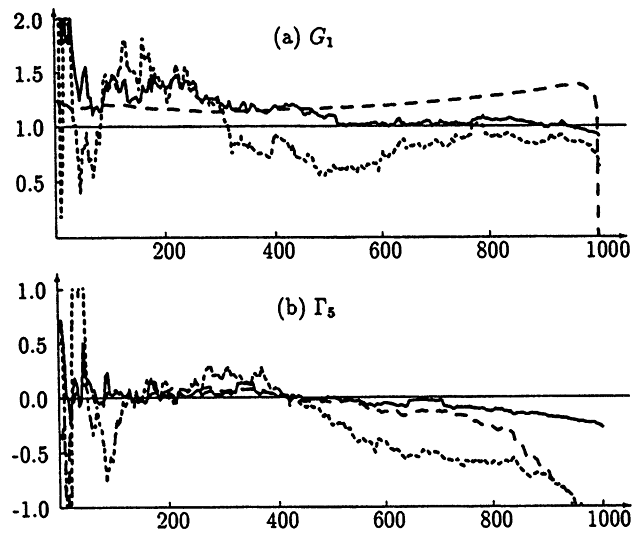


FIG. 3. *Estimates pertaining to $\hat{\beta}_{n,\nu(\hat{\beta}_n)}$ (solid line), $\hat{\beta}_n^D$ (broken line) and $\hat{\beta}_n([k_n/4])$ (dotted line) as a function of $k_n$, one sample of size $n = 1000$*: (a) $F = G_1$; (b) $F = \Gamma_5$.

alterations of $k_n$ can cause large changes of the estimates. The smoothness of the curve corresponding to $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ lies somewhere in between. Particularly for $k_n > 500$, the curve is almost as smooth as the curve pertaining to $\hat{\beta}_n^D$.

Another striking feature of the plots is the large systematic deviation of the moment estimate from the true value in Figure 3a (respectively, the rapidly increasing error in Figure 3b), which is apparently due to the bias. In accordance with the results given in Table 1, this effect is much less evident in the case of the refined Pickands estimator.

To sum up, it can be said that the refined Pickands estimator is the best estimator under consideration for negative parameters, whereas usually the moment estimator is superior if $\beta > 0$ and $k_n$ is chosen appropriately. Moreover, $\hat{\beta}_{n,\nu(\tilde{\beta}_n)}$ is more robust against an unsuitable choice of $k_n$ for all values of $\beta$.

**4. Proofs.** One main tool for proving the functional limit Theorem 2.1 is the "Hungarian construction" of Komlós, Major and Tusnády (1975, 1976) [see also Shorack and Wellner (1986), (2.7.7)]. Given a standard Brownian motion $W$, this construction yields a sequence of standard exponential r.v.'s $\xi_j$ such that

$$P\left\{ \max_{1 \le i \le k} \left| \sum_{j=1}^{i} (\xi_j - 1) - W(i) \right| > d_1 \log(k) + x \right\} < d_2 \exp(-d_3 x), \tag{4.1}$$

for some constants $d_1, d_2, d_3 > 0$ and all $x \in \mathbb{R}$ and $k \in \mathbb{N}$.

Define

$$A_{n,i} := \frac{F^{-1}\left(1 - n^{-1}\Sigma_{j=1}^{i}\xi_j\right) - F^{-1}\left(1 - n^{-1}\Sigma_{j=1}^{2i}\xi_j\right)}{F^{-1}\left(1 - n^{-1}\Sigma_{j=1}^{2i}\xi_j\right) - F^{-1}\left(1 - n^{-1}\Sigma_{j=1}^{4i}\xi_j\right)} - 2^{\beta}$$

if $n^{-1}\Sigma_{j=1}^{l}\xi_j \in (0,1)$ for $l \in \{i, 2i, 4i\}$ and $A_{n,i} := 0$ otherwise. Moreover, let

$$B_{n,i} := \frac{\log\left(A_{n,i} + 2^{\beta}\right)}{\log(2)}.$$

By Lemma 5.4.2 of Reiss (1989) and a quantile transformation, one obtains the following bound on the variational distance between the distribution of the vector of Pickands estimators $\hat{\beta}_n(i)$ and of the vector of $B_{n,i}$:

$$\left\|\mathscr{L}\left(\left(\hat{\beta}_n(i)\right)_{1\le i\le m_n}\right) - \mathscr{L}\left((B_{n,i})_{1\le i\le m_n}\right)\right\| = O\left(\frac{m_n}{n}\right). \tag{4.2}$$

The following technical lemmas are basic for the proof of Theorems 2.1 and 2.2.

LEMMA 4.1. *Suppose that Condition* 2.1(i) *or* 2.1(ii) *is satisfied. Let*

$$\Delta_{n,m}^{(1)} := \sup_{t\in[\varepsilon,1]}\left|\Phi\left(\frac{tm_n}{n}\right)\right| \cdot 1_{\{\Psi\ne 0\}},$$

$$\Delta_{n,m}^{(2)} := \sup_{t\in[\varepsilon,1],\, l\in\{1,2,4\},\, |x|\le d(\log(m)/m)^{1/2}}\left|R\left(\frac{tm}{n}, l(1+x)\right)\right|,$$

*where* $d := (d_1 + 2d_3^{-1} + 4)/\varepsilon$.

*Then, for all* $\varepsilon > 0$ *and* $\eta \in (0, \frac{1}{4})$, *there are positive constants* $d_4$ *and* $d_5$ *such that, for all* $m \le ]\eta n[$,

$$P\left\{\max_{]\varepsilon m[\ \le i\le m}\left|B_{n,i} - \beta - c_\beta\left(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi\left(\frac{i}{n}\right)\right)\right| > d_4\left(\frac{\log(m)}{m} + \left(\Delta_{n,m}^{(1)}\right)^2 + \Delta_{n,m}^{(2)}\right)\right\} \le d_5 m^{-2}. \tag{4.3}$$

*In particular*,

$$\max_{]\varepsilon m_n[\ \le i\le m_n} i^{1/2}\left|B_{n,i} - \beta - c_\beta\left(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi\left(\frac{i}{n}\right)\right)\right| \to 0 \quad a.s., \tag{4.4}$$

*if Condition* 2.2(i) [*respectively*, 2.2(ii)] *is satisfied.*

PROOF. We restrict ourselves to the case where $\beta \ne 0$ is satisfied because the proof runs along the same lines in the other case. Subsequently, $a(i,m,n) = O(b(m,n))$ is interpreted as $|a(i,m,n)| \le c|b(m,n)|$ for all $]\varepsilon m[\le i \le m \le ]\eta n[$ and some constant $c$ depending only on $\varepsilon$ and $\eta$.

For $r_i := \Sigma_{j=1}^{i}(\xi_j - 1) - W(i)$, we obtain, from (4.1),

$$P\left\{\max_{1\le i\le 4m}|r_i| > \left(d_1 + 2d_3^{-1}\right)\log(4m)\right\} = O(m^{-2}). \tag{4.5}$$

The reflection principle yields

$$(4.6)\qquad \begin{aligned} &P\Big\{\max_{1\le i\le 4m}|W(i)| > (16m\log(m))^{1/2}\Big\}\\ &\quad\le 4P\big\{W(4m) > (16m\log(m))^{1/2}\big\} = O(m^{-2}).\end{aligned}$$

Now assume that

$$(4.7)\qquad \begin{aligned} &\max_{1\le i\le 4m}|r_i| \le \big(d_1 + 2d_3^{-1}\big)\log(4m) \quad\text{and}\\ &\max_{1\le i\le 4m}|W(i)| \le (16m\log(m))^{1/2}.\end{aligned}$$

Then $n^{-1}\sum_{j=1}^{4m}\xi_j = n^{-1}(4m + W(4m) + r_{4m}) < 1$ for sufficiently large $n$. Thus, for $l \in \{1,2,4\}$ and $]\varepsilon m[\, \le i \le m$, we have

$$\begin{aligned} &F^{-1}\Bigg(1 - n^{-1}\sum_{j=1}^{li}\xi_j\Bigg)\\ &\quad= F^{-1}\bigg(1 - \frac{i}{n}l\bigg(1 + \frac{W(li)+r_{li}}{li}\bigg)\bigg)\\ &\quad= F^{-1}\bigg(1-\frac{i}{n}\bigg) + a\bigg(\frac{i}{n}\bigg)\Bigg(\frac{l^{-\beta}\big(1+(W(li)+r_{li})/(li)\big)^{-\beta}-1}{\beta}\\ &\qquad\qquad +\Phi\bigg(\frac{i}{n}\bigg)\Psi\bigg(l\bigg(1+\frac{W(li)+r_{li}}{li}\bigg)\bigg) + R\bigg(\frac{i}{n}, l\bigg(1+\frac{W(li)+r_{li}}{li}\bigg)\bigg)\Bigg)\\ &\quad= F^{-1}\bigg(1-\frac{i}{n}\bigg) + a\bigg(\frac{i}{n}\bigg)\Bigg(\frac{l^{-\beta}-1}{\beta} - l^{-\beta}\frac{W(li)}{li} + \Phi\bigg(\frac{i}{n}\bigg)\Psi(l)\\ &\qquad\qquad +O\bigg(\frac{\log(m)}{m} + \Delta_{n,m}^{(1)}\bigg(\frac{\log(m)}{m}\bigg)^{1/2} + \Delta_{n,m}^{(2)}\bigg)\Bigg),\end{aligned}$$

where the last equation follows from the Taylor expansion $(1+x)^{-\beta} = 1 - \beta x + O(x^2)$ and the differentiability of $\Psi$. By (4.7) and $(1+O(x))^{-1} = 1 + O(x)$, it follows that

$$\begin{aligned} A_{n,i} &= \beta\Big(-(1/i)\big(W(i) - (1+2^{-\beta})/2W(2i) + 2^{-(\beta+2)}W(4i)\big)\\ &\qquad +2d_{\beta,\Psi}\Phi(i/n)\\ &\qquad +O\Big(\log(m)/m + \Delta_{n,m}^{(1)}(\log(m)/m)^{1/2} + \Delta_{n,m}^{(2)}\Big)\Big)\\ &\qquad \times\Big[2^{-\beta} - 4^{-\beta} + O\Big((\log(m)/m)^{1/2} + \Delta_{n,m}^{(1)} + \Delta_{n,m}^{(2)}\Big)\Big]^{-1}\\ &= 2^{\beta-1}\log(2)c_\beta\bigg(-4Z_\beta(4i) + 2d_{\beta,\Psi}\Phi\bigg(\frac{i}{n}\bigg)\bigg)\\ &\qquad + O\bigg(\frac{\log(m)}{m} + \big(\Delta_{n,m}^{(1)}\big)^2 + \Delta_{n,m}^{(2)}\bigg).\end{aligned}$$

Utilizing a Taylor expansion of the logarithm one obtains

$$B_{n,i} - \beta = \frac{2^{-\beta} A_{n,i}}{\log(2)} + O(A_{ni}^2)$$
$$= c_\beta\left(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi\left(\frac{i}{n}\right)\right) + O\left(\frac{\log(m)}{m} + \left(\Delta_{n,m}^{(1)}\right)^2 + \Delta_{n,m}^{(2)}\right).$$

In view of (4.5) and (4.6), assertion (4.3) is proven.

Under Condition 2.2(i) [respectively, 2.2(ii)], $m_n^{1/2}(\log(m_n)/m_n + (\Delta_{n,m_n}^{(1)})^2 + \Delta_{n,m_n}^{(2)}) \to 0$. Therefore (4.4) follows by the Borel–Cantelli lemma. □

LEMMA 4.2. *Under the conditions of Theorem* 2.1, *one has*

$$\max_{1 \le i \le m_n} i^{1/2}\left|B_{n,i} - \beta - c_\beta\left(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi\left(\frac{i}{n}\right)\right)\right| = O_p(1).$$

PROOF. Let $C_{n,i} := i^{1/2}|B_{n,i} - \beta - c_\beta(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi(i/n))|$. Assume that the assertion is false. Then there is an $\varepsilon > 0$ such that, for all $k, N \in \mathbb{N}$, there exists $n > N$ with

$$\varepsilon < P\left\{\max_{1 \le i \le m_n} C_{n,i} > k\right\} \le \sum_{i=1}^{m_n} P\{C_{n,i} > k\}.$$

Thus one can find sequences $(n(k))_{k\in\mathbb{N}}$ and $(i(k))_{k\in\mathbb{N}}$ with $n(k) \to \infty$, $1 \le i(k) \le m_{n(k)}$, such that

$$P\{C_{n(k),i(k)} > k\} > \varepsilon\left(\sum_{j=1}^{\infty} j^{-3/2}\right)^{-1} (i(k))^{-3/2}. \tag{4.8}$$

If $(i(k))_{k\in\mathbb{N}}$ is unbounded, then it has an intermediate subsequence which by (4.8) is a contradiction to (4.3). Hence there is a constant subsequence of $(i(k))_{k\in\mathbb{N}}$. Consequently,

$$P\{C_{n(k(l)),i} > k(l)\} > \tilde{\varepsilon}, \tag{4.9}$$

for some sequence $k(l) \to \infty$ and some fixed $i \in \mathbb{N}$ and $\tilde{\varepsilon} > 0$.

On the other hand, $F \in D(G_\beta)$ implies that, for some normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$, $(a_n^{-1}(X_{n-j+1:n} - b_n))_{1\le j\le i}$ converges weakly. Because of (4.2) and $\Phi(i/n) \to 0$, it follows that $(C_{n,i})_{n\in\mathbb{N}}$ is stochastically bounded, which is a contradiction to (4.9). Hence the assertion is proven. □

PROOF OF THEOREM 2.1. Let

$$I(n,i) := \left(\frac{i-1}{m_n}, \frac{i}{m_n}\right], \quad \tilde{Z}_{n,\beta} := \sum_{i=1}^{m_n} m_n^{1/2}(B_{n,i} - \beta)1_{I(n,i)} \quad \text{and}$$

$$\bar{Z}_{n,\beta}(t) := -(4m_n)^{1/2} Z_\beta(4m_n t).$$

Because of $\mathscr{L}(\bar{Z}_{n,\beta}) = \mathscr{L}(Z_\beta)$ and (4.2), it suffices to prove that

$$\begin{aligned}
&\sup_{t\in[0,1]} h(t)\left|\tilde{Z}_{n,\beta}(t) - c_\beta\left(\bar{Z}_{n,\beta}(t) + \lambda d_{\beta,\Psi} t^\delta\right)\right| \\
&\quad \le \max_{1\le i\le m_n} \sup_{t\in I(n,i)} h(t)\left|\tilde{Z}_{n,\beta}\left(\frac{i}{m_n}\right) - c_\beta\left(\bar{Z}_{n,\beta}\left(\frac{i}{m_n}\right) + d_{\beta,\Psi} m_n^{1/2}\Phi\left(\frac{i}{n}\right)\right)\right| \\
&\qquad + |c_\beta| \max_{1\le i\le m_n} \sup_{t\in I(n,i)} \left|\left(h(t) - h\left(\frac{i}{m_n}\right)\right)\bar{Z}_{n,\beta}\left(\frac{i}{m_n}\right)\right| \\
&\qquad + |c_\beta| \max_{1\le i\le m_n} \sup_{t\in I(n,i)} \left|h\left(\frac{i}{m_n}\right)\bar{Z}_{n,\beta}\left(\frac{i}{m_n}\right) - h(t)\bar{Z}_{n,\beta}(t)\right| \\
&\qquad + |c_\beta d_{\beta,\Psi}| \max_{1\le i\le m_n} \sup_{t\in I(n,i)} h(t)\left|m_n^{1/2}\Phi\left(\frac{i}{n}\right) - \lambda t^\delta\right| \\
&\quad =: T_1 + T_2 + T_3 + T_4 \to 0
\end{aligned}$$

in probability.

Using (2.4) and Lemmas 4.1 and 4.2, we obtain $T_1 \to 0$ in probability, since, for all $\varepsilon > 0$,

$$\begin{aligned}
T_1 \le &\sup_{t\in(0,1]} h(t)t^{-1/2} \max_{]\varepsilon m_n[\le i\le m_n} i^{1/2} \\
&\times\left|B_{n,i} - \beta - c_\beta\left(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi\left(\frac{i}{n}\right)\right)\right| \\
&+ \sup_{t\in(0,\varepsilon]} h(t)t^{-1/2} \max_{1\le i\le m_n} i^{1/2} \\
&\times\left|B_{n,i} - \beta - c_\beta\left(-2Z_\beta(4i) + d_{\beta,\Psi}\Phi\left(\frac{i}{n}\right)\right)\right|.
\end{aligned}$$

Let $\tilde{h}(t) := (\log\log(3/t)/t)^{-1/2}$. Check that the continuity of $h/\tilde{h}$ and the monotonicity of $\tilde{h}$ in combination with (2.4) yield

$$\begin{aligned}
&\max_{1\le i\le m_n} \sup_{t\in I(n,i)} \frac{|h(t) - h(i/m_n)|}{\tilde{h}(i/m_n)} \\
&\quad \le \max_{1\le i\le m_n} \sup_{t\in I(n,i)} \left|\frac{h(t)}{\tilde{h}(t)} - \frac{h(i/m_n)}{\tilde{h}(i/m_n)}\right| \\
&\qquad + \max_{1\le i\le m_n} \sup_{t\in I(n,i)} h(t)\left|\frac{1}{\tilde{h}(t)} - \frac{1}{\tilde{h}(i/m_n)}\right| \to 0.
\end{aligned}$$

Hence, by the law of the iterated logarithm,

$$T_2 \le |c_\beta| \max_{1\le i\le m_n} \tilde{h}\left(\frac{i}{m_n}\right)\left|\bar{Z}_{n,\beta}\left(\frac{i}{m_n}\right)\right| \cdot \max_{1\le i\le m_n} \sup_{t\in I(n,i)} \frac{|h(t) - h(i/m_n)|}{\tilde{h}(i/m_n)} \to 0.$$

Next observe that the continuity of $h \cdot Z_\beta$ implies $T_3 \to 0$ in probability.

Finally, by the uniform convergence theorem for regularly varying functions [see, e.g., de Haan (1970), Corollary 1.2.1.4], one has

$$T_4 \le |c_\beta d_{\beta,\Psi}| \Bigg( \sup_{t\in(0,1]} h(t) t^{-1/2} m_n^{1/2} \Phi\Big(\frac{m_n}{n}\Big) \max_{1\le i\le m_n} \Big(\frac{i}{m_n}\Big)^{1/2} \times \left| \frac{\Phi(i/n)}{\Phi(m_n/n)} - \Big(\frac{i}{m_n}\Big)^{\delta} \right| + \max_{1\le i\le m_n} \sup_{t\in I(n,i)} h(t) \left| m_n^{1/2} \Phi\Big(\frac{m_n}{n}\Big)\Big(\frac{i}{m_n}\Big)^{\delta} - \lambda t^{\delta} \right| \Bigg) \to 0. \qquad \square$$

REMARK. Theorem 2.1 can also be proven by utilizing strong or stochastic approximations of the quantile or tail quantile process as given by Csörgő, (Csörgő, Horváth and Mason (1986), Einmahl and Mason (1988) or Einmahl (1992) instead of the Hungarian construction and the approximation (4.2). The latter reference gives a nice survey of such approximations [see also Csörgő and Horváth (1993)]. This approach was fruitfully utilized by Csörgő, Deheuvels and Mason (1985), among others. However, some of these approximations require additional conditions (like the existence of a density of $F$) and, moreover, by this means it may be difficult to obtain rates of convergence as in Theorem 2.2.

PROOF OF COROLLARY 2.1. Choose a strictly decreasing sequence $t_n \to 0$ such that $\int_{(0,t_n]} (\log\log(3/t)/t)^{1/2} \nu(dt) < n^{-3}$. Then

$$h(t) := \left( \frac{t}{\log\log(3/t)} \right)^{1/2} \Bigg/ \begin{cases} n - (t - t_{n+1})/(t_n - t_{n+1}), & \text{if } t \in (t_{n+1}, t_n], \\ 1, & \text{if } t \in (t_2, 1], \end{cases}$$

defines a positive, continuous function $h$ which satisfies (2.4). Since $\int 1/h(t)\nu(dt) < \infty$, the map $D[0,1] \to \mathbb{R}$, $z \mapsto \int z(t)/h(t)\nu(dt)$, is continuous at any $z \in C[0,1]$. Therefore the assertion follows from Theorem 2.1, (2.3) and the continuous mapping theorem, because $\mathscr{L}(\int Z_\beta(t)\nu(dt)) = \mathscr{N}(0, \sigma_{\beta,\nu}^2)$ [Shorack and Wellner (1986), Proposition 2.2.1]. $\square$

PROOF OF THEOREM 2.2. We will prove the stronger result

$$\sup_{x\in\mathbb{R}} \left| P\left\{ m_n^{1/2}\big(\hat\beta_{n,\nu} - \beta\big) \le x \right\} - \mathscr{N}\big(0, c_\beta^2 \sigma_{\beta,\nu}^2\big)(-\infty, x] \right| = O\left( \frac{m_n}{n} + \log(m_n) m_n^{-1/2} + m_n^{1/2} \Delta_{n,m_n}^{(2)} \right),$$

where $\Delta^{(2)}_{n,m_n}$ is defined in Lemma 4.1. Define $I(n,i)$, $\tilde{Z}_{n,\beta}$ and $\bar{Z}_{n,\beta}$ as in the proof of Theorem 2.1. Let $\nu_n := \sum_{i=1}^{m_n} \nu(I(n,i))\varepsilon_{i/m_n}$. Lemma 4.1 implies that

$$P\left\{\left|\int \tilde{Z}_{n,\beta}(t)\nu(dt) - c_\beta \int \bar{Z}_{n,\beta}(t)\nu_n(dt)\right| > d_4\left(\log(m_n)m_n^{-1/2} + m_n^{1/2}\Delta^{(2)}_{n,m_n}\right)\right\}$$

$$\leq P\left\{\max_{]\varepsilon m_n[\leq i \leq m_n} |B_{n,i} - \beta + 2c_\beta Z_\beta(4i)| > d_4\left(\frac{\log(m_n)}{m_n} + \Delta^{(2)}_{n,m_n}\right)\right\}$$

$$\leq d_5 m_n^{-2}.$$

Observe that

$$\sup_{x\in\mathbb{R}} \left|\mathcal{N}\left(0, c_\beta^2\sigma^2_{\beta,\nu_n}\right)(-\infty, x+y] - \mathcal{N}\left(0, c_\beta^2\sigma^2_{\beta,\nu_n}\right)(-\infty, x]\right| \leq \left(2\pi\sigma^2_{\beta,\nu_n}\right)^{-1/2}|y|,$$

for all $y \in \mathbb{R}$ and $\mathcal{L}(c_\beta \int \bar{Z}_{n,\beta}(t)\nu_n(dt)) = \mathcal{N}(0, c_\beta^2\sigma^2_{\beta,\nu_n})$. Hence

$$\sup_{x\in\mathbb{R}}\left|P\left\{\int \tilde{Z}_{n,\beta}(t)\nu(dt) \leq x\right\} - \mathcal{N}\left(0, c_\beta^2\sigma^2_{\beta,\nu_n}\right)(-\infty, x]\right| \tag{4.10}$$
$$= O\left(\log(m_n)m_n^{-1/2} + m_n^{1/2}\Delta^{(2)}_{n,m_n}\right).$$

Elementary calculations show that $|\sigma^2_{\beta,\nu_n} - \sigma^2_{\beta,\nu}| = O(m_n^{-1})$, and so

$$\sup_{x\in\mathbb{R}}\left|\mathcal{N}\left(0, c_\beta^2\sigma^2_{\beta,\nu_n}\right)(-\infty, x] - \mathcal{N}\left(0, c_\beta^2\sigma^2_{\beta,\nu}\right)(-\infty, x]\right| = O\left(\sigma^{-1}_{\beta,\nu_n} - \sigma^{-1}_{\beta,\nu}\right)$$
$$= O(m_n^{-1})$$

[cf. Reiss (1989), Problem 4.2(ii)]. Now the assertion follows readily from (4.2) and (4.10). □

REMARK. In a similar way one can prove that

$$\sup_{x\in\mathbb{R}}\left|P\left\{m_n^{1/2}\left(\hat{\beta}_{n,\nu} - \beta\right) \leq x\right\} - \mathcal{N}\left(\lambda c_\beta d_{\beta,\Psi}\int t^\delta \nu(dt), c_\beta^2\sigma^2_{\beta,\nu}\right)(-\infty, x]\right|$$
$$= O\Bigg(\frac{m_n}{n} + \log(m_n)m_n^{-1/2} + m_n^{1/2}\left(\left(\Delta^{(1)}_{n,m_n}\right)^2 + \Delta^{(2)}_{n,m_n}\right)$$
$$+ \left|m_n^{1/2}\Phi\left(\frac{m_n}{n}\right) - \lambda\right| + \sup_{t\in[\varepsilon,1]}\left|\frac{\Phi(tm_n/n)}{\Phi(m_n/n)} - t^\delta\right|\Bigg)$$

if Conditions 2.1(ii) and 2.2(ii) are satisfied and $\nu(0,\varepsilon] = 0$.

PROOF OF LEMMA 2.1. The map $\nu \mapsto \sigma^2_{\beta,\nu}$ has a local minimum at $\nu^*$ if and only if, for all finite signed measures $\tilde{\nu}$ on $\mathbb{B}(0,1]$ with $\tilde{\nu}(0,1] = 0$ and $\int(\log\log(3/t)/t)^{1/2}|\tilde{\nu}|(dt) < \infty$,

$$\iint \sigma_\beta(s,t)(\nu^* + \rho\tilde{\nu})^2(ds,dt) \geq \iint \sigma_\beta(s,t)(\nu^*)^2(ds,dt) \quad \forall \rho \in \mathbb{R}$$

$$\Leftrightarrow \quad \iint \sigma_\beta(s,t)\nu^*(ds)\tilde{\nu}(dt) = 0.$$

This means that $\int\sigma_\beta(s,t)\nu^*(ds)$ is constant for all $t \in (0,1]$. Furthermore, if two probability measures $\nu_1^*$ and $\nu_2^*$ have this property in common, then

$$\iint \sigma_\beta(s,t)(\nu_1^*)^2(ds,dt) = \iint \sigma_\beta(s,t)\big((\nu_2^*)^2 + (\nu_1^* - \nu_2^*)^2\big)(ds,dt)$$

$$\geq \iint \sigma_\beta(s,t)(\nu_2^*)^2(ds,dt)$$

and vice versa. Hence $\sigma^2_{\beta,\nu_1^*} = \sigma^2_{\beta,\nu_2^*}$, which shows that every local minimum of the map $\nu \mapsto \sigma^2_{\beta,\nu}$ is a global minimum. By some direct calculations, one obtains that $\int\sigma_\beta(s,t)\nu^*(\beta)(ds)$ is independent of $t$. Thus the infimum is attained at $\nu^*(\beta)$, and (2.7) can be proven by elementary calculations. □

PROOF OF THEOREM 2.3. Let $h(t) := t^{1/2+\xi}$, $t \in [0,1]$, for some $\xi \in (0, \min(\rho, \frac{1}{2}))$. According to Theorem 2.1 there are versions $\tilde{Z}_{n,\beta}$ of $Z_{n,\beta}$ and a version $\tilde{Z}_\beta$ of $Z_\beta$ such that $\sup_{t\in[0,1]} h(t)|\tilde{Z}_{n,\beta}(t) - c_\beta\tilde{Z}_\beta(t)| \to 0$ a.s. The triangle inequality yields

$$\begin{aligned}
&\left|\int \tilde{Z}_{n,\beta}(t)\nu(\tilde{\beta}_n)(dt) - \int c_\beta\tilde{Z}_\beta(t)\nu(\beta)(dt)\right| \\
&\quad\leq \int |\tilde{Z}_{n,\beta}(t) - c_\beta\tilde{Z}_\beta(t)|\nu(\tilde{\beta}_n)(dt) \\
&\qquad + |c_\beta|\left|\int \tilde{Z}_\beta(t)\nu(\tilde{\beta}_n)(dt) - \int \tilde{Z}_\beta(t)\nu(\beta)(dt)\right| \\
&\quad\leq \sup_{t\in[0,1]} h(t)|\tilde{Z}_{n,\beta}(t) - c_\beta\tilde{Z}_\beta(t)| \int \frac{1}{h(t)}\nu(\tilde{\beta}_n)(dt) \\
&\qquad + |c_\beta| \sup_{t\in(0,1]} h(t)|\tilde{Z}_\beta(t)| \sum_{i=0}^\infty \frac{|\nu(\tilde{\beta}_n)\{2^{-i}\} - \nu(\beta)\{2^{-i}\}|}{h(2^{-i})}.
\end{aligned}$$

Since $\sup_{\bar{\beta}\in\mathbb{R}} \int 1/h(t)\nu(\bar{\beta})(dt) < \infty$, the first term converges to 0 a.s. Using a generalized version of the Scheffé lemma [Reiss (1989), Lemma 3.3.4] and the weak consistency of $\tilde{\beta}_n$, it is easily seen that $\sum_{i=0}^\infty|\nu(\tilde{\beta}_n)\{2^{-i}\} - \nu(\beta)\{2^{-i}\}|/h(2^{-i}) \to 0$ in probability. Thus, the convergence of the second term to 0 in probability is an immediate consequence of $\sup_{t\in(0,1]}h(t)|\tilde{Z}_\beta(t)| < \infty$ a.s. Now the assertion follows from $\mathscr{L}(\int c_\beta\tilde{Z}_\beta(t)\nu(\beta)(dt)) = \mathscr{N}(0, c_\beta^2\sigma^2_{\beta,\nu(\beta)})$ and (2.3). □

**Acknowledgments.** I am grateful to R.-D. Reiss for stimulating discussions. I also want to thank the referees for their valuable suggestions and remarks, which led to a substantial improvement.

## REFERENCES

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

CSÖRGŐ, S., DEHEUVELS, P. and MASON, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13** 1050–1077.

CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley, New York.

DE HAAN, L. (1970). *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.) 31–38. Reidel, Dordrecht.

DE HAAN, L. and STADTMÜLLER, U. (1993). Generalized regular variation of second order. Preprint, Tinbergen Institute, Rotterdam.

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.

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.

DREES, H. (1993). Refined estimation of the extreme value index. Dissertation, Universität-Gesamthochschule Siegen.

EINMAHL, J. H. J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *J. Statist. Plann. Inference* **32** 137–145.

EINMAHL, J. H. J. and MASON, D. M. (1988). Strong limit theorems for weighted quantile processes. *Ann. Probab.* **16** 1623–1643.

FALK, M. (1985). Uniform convergence of extreme order statistics. Habilitationsschrift, Universität-Gesamthochschule Siegen.

FALK, M. (1994). Efficiency of convex combinations of Pickands estimator of the extreme value index. *Journal of Nonparametric Statistics* **4** 133–147.

FALK, M., HÜSLER, J. and REISS, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser, Basel.

FALK, M. and MAROHN, F. (1993). Von Mises conditions revisited. *Ann. Probab.* **21** 1310–1328.

FALK, M. and REISS, R.-D. (1992). Poisson approximation of empirical processes. *Statist. Probab. Lett.* **14** 39–48.

GOLDIE, C. M. and SMITH, R. L. (1987). Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford Ser.* (2) **38** 45–71.

HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44** 37–42.

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.

HOSKING, J. R. M. and WALLIS, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29** 339–349.

KAUFMANN, E. (1994). Von Mises conditions, $\delta$-neighborhoods and rates of convergence for maxima. Preprint, Universität-Gesamthochschule Siegen.

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.

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. Klumar, Dordrecht.

PICKANDS, III J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131.

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer, New York.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174–1207.

WEISS, L. (1971). Asymptotic inference about a density function at the end of its range. *Naval Res. Logist. Quart.* **18** 111–114.

MATHEMATISCHES INSTITUT
UNIVERSITÄT ZU KÖLN
WEYERTAL 86-90
50931 KÖLN
GERMANY