The Annals of Statistics
1998, Vol. 26, No. 1, 434–448

OPTIMAL RATES OF CONVERGENCE FOR ESTIMATES OF THE EXTREME VALUE INDEX[1]

BY HOLGER DREES

University of Cologne

Hall and Welsh established the best attainable rate of convergence for estimates of a positive extreme value index γ under a certain second order condition implying that the distribution function of the maximum of n random variables converges at an algebraic rate to the pertaining extreme value distribution. As a first generalization, we obtain a surprisingly sharp bound on the estimation error if γ is still assumed to be positive, but the rate of convergence of the maximum may be nonalgebraic. This result allows a more accurate evaluation of the asymptotic performance of an estimator for γ than the Hall and Welsh theorem. For example, it is proved that the Hill and the Pickands estimators achieve the optimal rate, but only the Hill estimator attains the sharp bound. Finally, an analogous result is derived for a general, not necessarily positive, extreme value index. In this situation it turns out that location invariant estimators show the best performance.

1. Introduction. Consider n i.i.d. random variables (r.v.'s), whose distribution function (d.f.) F belongs to the weak domain of attraction of an extreme value d.f.:

$$G_\gamma(x) := \exp(-(1+\gamma x)^{-1/\gamma}), \qquad 1+\gamma x > 0, \ \gamma \in \mathbb{R},$$

which is interpreted as $\exp(-\mathrm{e}^{-x})$ if $\gamma = 0$. Then by definition there are sequences of normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) \to G_\gamma(x) \tag{1.1}$$

for all $x \in \mathbb{R}$. Whereas many estimators of the so-called extreme value index γ were proposed in literature [see, e.g., Hill (1975), Pickands (1975), Hall and Welsh (1985), Csörgő, Deheuvels and Mason (1985), Smith (1987), Dekkers, Einmahl and de Haan (1989) and Drees (1995a, 1997a, b)], little is known about optimality of these estimators.

The most important result in literature about what *can* be achieved is given in Hall and Welsh (1984), where for positive γ an upper bound on the rate is established at which any sequence of estimators $\hat{\alpha}_n$ for $\alpha = 1/\gamma$ converges towards the true parameter uniformly over certain neighborhoods of Pareto distributions. More precisely, they defined sets $\mathscr{D} = \mathscr{D}(\alpha_0, c_0, \varepsilon, \rho, A)$ consisting of all densities $f\colon (0,\infty) \to [0,\infty)$ which satisfy

$$f(x) = d\alpha x^{-(\alpha+1)}(1 + r(x)) \quad \text{where } |r(x)| \le A x^{-\alpha\rho} \tag{1.2}$$

Received January 1996; revised March 1997.

[1]Research partially supported by a DFG research grant.

AMS 1991 *subject classifications*. Primary 62G05; secondary 62G20, 62G35.

Key words and phrases. Extreme value index, rate of convergence, optimality, robustness, Hill estimator, moment estimator, Pickands estimator.

for all $x > 0$, $|\alpha - \alpha_0| \le \varepsilon$, $|d - d_0| \le \varepsilon$ and $\alpha_0, d_0, \varepsilon, \rho, A > 0$. Then it was proved that

$$\lim_{n\to\infty} \inf_{f\in\mathscr{D}} P_f^n\{|\hat{\alpha}_n - \alpha| \le a_n\} = 1$$

implies

$$\lim_{n\to\infty} a_n\, n^{\rho/(2\rho+1)} = \infty. \tag{1.3}$$

Here P_f^n denotes the distribution of n i.i.d. r.v.'s with density f. Moreover, Hall and Welsh showed that the reciprocal of the Hill estimator

$$\hat{\gamma}_{n,\,H} := \frac{1}{k_n}\sum_{i=1}^{k_n} \log\left(\frac{X_{n-i+1:n}}{X_{n-k_n:n}}\right) \tag{1.4}$$

actually attains this optimal bound if $k_n \sim n^{2\rho/(2\rho+1)}$. Here $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ denote the order statistics pertaining to $X_1, \ldots, X_n$.

Thus as far as rates of convergence are concerned, optimality in model (1.2) may be defined using (1.3). Furthermore, the rate of uniform convergence of an estimator for γ can be considered as a measure of its robustness against deviations of type (1.2) from the ideal Pareto distributions. However, note that model (1.2) is rather restrictive, since 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, for example, for loggamma d.f.'s with density $f(x) = cx^{-(1/\gamma+1)}\log^{\beta-1}(x)1_{(1,\,\infty)}(x)$, which are common in nonlife insurance mathematics [cf. Ramlau-Hansen (1988) and Hogg and Klugman (1984)], or the log-hyperbolic distribution often used in geology and other natural sciences [see Beirlant, Teugels and Vynckier (1996)] the rate of convergence is much slower, namely a power of the logarithm of the sample size. 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 the Hall and Welsh result to enlarged sets of distributions including quite arbitrary d.f.'s in the domain of attraction of an extreme value d.f.

Note that there is no obvious counterpart of condition (1.2) if the density is not tail equivalent to a Pareto density. This is partly because in (1.2) the bound on the remainder term $r(x)$ depends both on the first order parameter α and the second order parameter ρ. For that reason, in case $\gamma > 0$ we replace (1.2) by a suitable condition in terms of the function

$$U(t) := F^{-1}\left(1 - \frac{1}{t}\right), \qquad t > 1,$$

with F^{-1} denoting the quantile function (q.f.) pertaining to F.

Recall that $F \in D(G_\gamma)$, $\gamma > 0$, if and only if U is regularly varying with exponent γ. Here we assume the slightly stronger condition that U is normalized regularly varying, that is,

$$U(t) = ct^\gamma \exp\left(\int_1^t \eta(s)/s\,ds\right) \tag{1.5}$$

with $\eta(t) \to 0$ as $t \to \infty$. Representation (1.5) is equivalent to the well-known von Mises condition $h_1(x) := (1 - F(x))/(xf(x)) \to \gamma$ as $x \to \infty$ where f denotes a suitable Lebesgue-density of F; in particular, (1.5) is necessary for $F \in D(G_\gamma)$ if f is eventually monotone [Bingham, Goldie and Teugels (1987), Theorem 1.7.2]. Observe that $\eta(t) = tU'(t)/U(t) - \gamma = h_1(U(t)) - \gamma$ measures the speed of convergence in Karamata's theorem applied to the regularly varying functions U' and f.

In Theorem 2.1 we establish an upper bound on the rate of convergence of an arbitrary sequence of estimators for γ uniformly over all d.f.'s satisfying (1.5) with $|\eta(t)| \le g(t)$ where g is $(-\rho)$–varying for some $\rho \ge 0$. In the special case of $g(t) = \text{const.}\, t^{-\rho}$, $\rho > 0$, this is essentially equivalent to the Hall and Welsh result (see Lemma 2.1); whereas for slowly varying function g, we do not only obtain an upper bound on the *rate* of convergence but an asymptotically sharp bound on the estimation error itself. Hence, the latter case leads to a criterion to discriminate the robustness of different estimators for γ against deviations from a Pareto tail, which is more accurate than the one following from the Hall and Welsh theorem: While most prominent estimators of γ converge at the optimal rate if a suitable fraction of the observations is used for estimation, for example the Pickands estimator defined by

$$\hat\gamma_{n,P} := \log\left(\frac{X_{n-k_n+1:n} - X_{n-2k_n+1:n}}{X_{n-2k_n+1:n} - X_{n-4k_n+1:n}}\right)\Big/\log(2) \tag{1.6}$$

does not attain the asymptotic bound in case of $\rho = 0$ but the Hill estimator does.

In the general case, where the sign of γ is not known in advance, there is no simple unifying representation of the function U comparable to (1.5), but the following analogue for its derivative U' is sufficient for $F \in D(G_\gamma)$ if $\eta(t) \to 0$ as $t \to \infty$:

$$U'(t) = ct^{\gamma-1} \exp\left(\int_1^t \eta(s)/s\,ds\right). \tag{1.7}$$

Note that (1.7) is equivalent to the von Mises type condition $h_2(x) := ((1 - F)/f)'(x) \to \gamma$ as $x \uparrow F^{-1}(1)$ [see Reiss (1989), (5.1.25)]. Here $\eta(t) = tU''(t)/U'(t) - (\gamma - 1) = h_2(U(t)) - \gamma$ quantifies the speed of Karamata's convergence for U'' and of the convergence of h_2.

Theorem 3.1 is the counterpart of Theorem 2.1 in this general setup where representation (1.5) is replaced by (1.7). While the bounds are literally the same, in contrast to the case $\gamma > 0$, estimators of the extreme value index that are not location invariant, as, for example, the moment estimator introduced by Dekkers, Einmahl and de Haan (1989):

$$\hat{\gamma}_{n,M} := \hat{\gamma}_{n,H} + 1 - \frac{1}{2(1 - \hat{\gamma}_{n,H}^2 / M_n^{(2)})} \quad \text{with}$$
$$M_n^{(2)} := \frac{1}{k_n} \sum_{i=1}^{k_n} \log^2\left(\frac{X_{n-i+1:n}}{X_{n-k_n:n}}\right), \tag{1.8}$$

in general do *not* converge at the optimal rate in model (1.7). Therefore, in the general situation, we recommend the use of location invariant estimators, like Pickands-type estimators or (in the case $\gamma > -1/2$) the maximum likelihood estimator examined by Smith (1987), which attain the optimal rates in both models (1.5) and (1.7). All proofs are given in Section 4.

Further bounds on the estimation error as well as modifications of the results presented here and more detailed proofs can be found in a technical report [Drees (1995b)].

2. Optimal rates of convergence in the case $\gamma > 0$. First we establish an upper bound on the rate of convergence in model (1.5) uniformly over all distributions for which the approximation error in the Karamata convergence for U' is bounded by a given regularly varying function g.

THEOREM 2.1. *Fix constants c, $\gamma_0 > 0$ and $0 < \varepsilon < \gamma_0$. Suppose that the function g: $(1, \infty) \to (0, \infty)$ is $(-\rho)$-varying for some $\rho \geq 0$, bounded away from 0 and finitely integrable on compact intervals, and eventually nonincreasing with $\lim_{t\to\infty} g(t) = 0$. Moreover, let $(\hat{\gamma}_n)$ be an arbitrary sequence of estimators for γ and (a_n), a sequence of positive real numbers.*

If

$$\lim_{n\to\infty} P_U^n\{|\hat{\gamma}_n - \gamma| \leq a_n\} = 1$$

uniformly for all functions U satisfying (1.5) *with $|\eta(t)| \leq g(t)$ for all $t > 1$ and $|\gamma - \gamma_0| \leq \varepsilon$, then*

$$\liminf_{n\to\infty} \frac{a_n}{g(t_n)} \begin{cases} = \infty, & \text{if } \rho > 0, \\ \geq 1, & \text{if } \rho = 0 \end{cases} \tag{2.1}$$

for any sequence (t_n) satisfying

$$\lim_{n\to\infty} g(t_n)\left(\frac{n}{t_n}\right)^{1/2} = 1. \tag{2.2}$$

(*Here P_U^n denotes the distribution of n i.i.d. r.v.'s with d.f. F pertaining to U.*)

REMARKS. (i) In contrast to (1.3), in general the bound $g(t_n)$ cannot be described explicitly. However, note that for a $(-\rho)$-varying function g there always exists a sequence (t_n) satisfying (2.2) and that (2.1) is equivalent for all such sequences.

(ii) By the same methods as used in the proof of Theorem 2.1, in the case $\rho > 0$ one can establish very crude asymptotic lower bounds on the probability that the estimation error exceeds $lg(t_n)$ where $l > 0$ is an arbitrary fixed constant [see Drees (1995b)].

As already mentioned, the case $\rho = 0$ is most interesting since then Theorem 2.1 is substantially stronger than the Hall and Welsh theorem, in that it gives a bound on the estimation error itself.

EXAMPLE. Let $g(t) = A/(1+\log(t))$ for some sufficiently large constant A. Then (1.5) with $|\eta(t)| \le g(t)$ is satisfied for large t by loggamma distributions. Since $(n/t_n)^{1/2}/(\log(t_n)) \to 1$ implies $\log(t_n) \sim \log(n)$, the optimal uniform bound a_n on the estimation error asymptotically behaves as $A/\log(n)$. In particular, one should not expect an estimator of γ to converge at a faster rate than $1/\log(n)$ if the observations are loggamma distributed.

Next we demonstrate by means of the examples introduced in Section 1 how Theorem 2.1 can be used to evaluate the robustness of estimators of the extreme value index against deviations of the tail of the underlying d.f. from the pertaining Pareto tail, particularly against large deviations in the case $\rho = 0$. To this end it is crucial to choose the number of order statistics used for estimation suitably depending on the dominating function g. In the case $\rho = 0$, the following proposition, which follows immediately from the uniform convergence theorem for slowly varying functions, plays a central role.

PROPOSITION 2.1. *If g is slowly varying and $\lim_{n\to\infty} t_n = \infty$, then there exists a sequence (t_n^*) such that*

$$\lim_{n\to\infty} \frac{t_n^*}{t_n} = 0 \quad \textit{yet} \quad \lim_{n\to\infty} \frac{g(t_n^*)}{g(t_n)} = 1. \tag{2.3}$$

Now fix some function g satisfying the assumptions of Theorem 2.1 and choose sequences (t_n) according to (2.2) and, for $\rho = 0$, (t_n^*) according to (2.3). In what follows we will always assume that the sequence (k_n), which determines the order statistics used by the estimators, is defined by

$$k_n := \begin{cases} \left[\dfrac{n}{t_n}\right], & \text{if } \rho > 0, \\[2ex] \left[\dfrac{n}{t_n^*}\right], & \text{if } \rho = 0, \end{cases} \tag{2.4}$$

where $[x]$ denotes the integral part of x.

First we examine the Hill estimator defined by (1.4). It turns out that $\hat{\gamma}_{n,H}$ is optimal in the sense that it attains the bounds given in Theorem 2.1. In particular, these bounds cannot be improved.

THEOREM 2.2. *Suppose the conditions of Theorem* 2.1 *are satisfied and*

$$\lim_{n\to\infty} \frac{a_n}{g(t_n)} = \infty \quad \text{if } \rho > 0, \tag{2.5}$$
$$\lim_{n\to\infty} \left(\frac{n}{t_n^*}\right)^{1/2} (a_n - g(t_n^*)) = \infty \quad \text{if } \rho = 0.$$

Then with k_n *according to* (2.4) *one has*

$$\lim_{n\to\infty} P_U^n\{|\hat{\gamma}_{n,H} - \gamma| \le a_n\} = 1$$

uniformly for all functions U considered in Theorem 2.1.

Observe that in case of $\rho = 0$ by (2.2) and (2.3) one has $(n/t_n^*)^{1/2} g(t_n^*) \to \infty$ and $g(t_n)/g(t_n^*) \to 1$, so that $(n/t_n^*)^{1/2}(a_n - g(t_n^*)) = (n/t_n^*)^{1/2} g(t_n^*)(a_n/g(t_n^*) - 1) \to \infty$, that is, (2.5), if $a_n/g(t_n) \downarrow 1$ sufficiently slowly. This proves that the optimal bound is actually attained by the Hill estimator.

REMARKS. (i) By obvious changes in the proof of Theorem 2.2, we see that the assertion holds true if $\hat{\gamma}_{n,H}$ is replaced by any kernel estimator defined by Csörgő, Deheuvels and Mason (1985) with support $\subset [0, 1]$.

(ii) Observe that the choice $k_n = [n/t_n]$ guarantees that the maximal bias and the standard deviation of the Hill estimator are of the same order. Now Proposition 2.1 enables us to reduce the variance to a smaller order without changing the bias asymptotically if one chooses $k_n = [n/t_n^*]$ in the case $\rho = 0$. Therefore the limiting distribution of $\hat{\gamma}_{n,H}$ (if it exists) will be degenerate if $\rho = 0$ and k_n is chosen optimally. In Drees (1997a), Corollary 4.1, a similar result is proved for a quite general class of estimators of γ. In fact, this degeneracy property explains why it is possible to obtain a sharp bound on the estimation error instead of merely a bound on the rate of convergence.

Many estimators of γ converge at the optimal rate but do not attain the asymptotically optimal bound in the case $\rho = 0$. A typical example is the Pickands estimator $\hat{\gamma}_{n,P}$ defined by (1.6).

THEOREM 2.3. *Suppose that the conditions of Theorem* 2.1 *hold and* k_n *fulfills* (2.4). *Moreover, assume that* a_n *satisfies*

$$\liminf_{n\to\infty} \frac{a_n}{g(t_n)} \begin{cases} = \infty, & \text{if } \rho > 0, \\ > \dfrac{1 + 2^{\varepsilon - \gamma_0}}{1 - 2^{\varepsilon - \gamma_0}}, & \text{if } \rho = 0. \end{cases} \tag{2.6}$$

Then

$$\lim_{n\to\infty} P_U^n\{|\hat{\gamma}_{n,P} - \gamma| \le a_n\} = 1 \tag{2.7}$$

uniformly for all U considered in Theorem 2.1.

It can be proved that for $\rho = 0$ the bound (2.6) is sharp in the sense that no sequence (a_n) with $\liminf_{n\to\infty} a_n/g(t_n) < (1+2^{\varepsilon-\gamma_0})/(1-2^{\varepsilon-\gamma_0})$ satisfies (2.7). Hence, asymptotically the uniform estimation error is at least three times as high as the minimal error if the extreme value index is less than 1 (which is satisfied in most applications), so that Hill's estimator is clearly superior in this case. However, it should be mentioned that the Pickands estimator attains the optimal bound, too, if not only the dominating function g but also η is regularly varying, which indeed is the case for the usual textbook distributions.

As it is shown in Drees (1997a, b), such a behavior is typical for a large class of estimators of the extreme value index, including the moment estimator $\hat{\gamma}_{n,M}$ defined in (1.8), whose performance lies somewhere in between the Hill and the Pickands estimators; it attains the optimal bound in the case $\rho = 0$ for $\gamma \geq 1$ but not for smaller extreme value indices. More precisely, one can show by lengthy calculations along the lines of Dekkers and de Haan (1993), proof of Theorem 3.4, that $\hat{\gamma}_{n,M}$ fulfills the assertions of Theorem 2.3 if for $\rho = 0$, (2.6) is replaced by

$$\liminf_{n\to\infty} \frac{a_n}{g(t_n)} > \left(\frac{2\exp(\gamma_0 - \varepsilon - 1)}{\gamma_0 - \varepsilon} - 1\right) 1_{\{\gamma_0 - \varepsilon < 1\}} + 1_{\{\gamma_0 - \varepsilon \geq 1\}}$$

where the right-hand side is strictly greater than 1 if $\gamma_0 - \varepsilon < 1$. On the other hand, asymptotically the uniform estimation error is always less than the error of the Pickands estimator and it is less than 1/3 of this error if the extreme value index is less than 1. Hence, the Pickands estimator shows the worst performance of the three estimators under consideration.

We close this section with a comparison between the Hall and Welsh result and Theorem 2.1.

LEMMA 2.1. *Consider the following two conditions.*

(C1) *The density* f *fulfills* (1.2) *for all* $x \geq x_0$ *with* $|1/\alpha - 1/\alpha_0| \leq \varepsilon$, $|d - d_0| \leq \varepsilon$.

(C2) *The function* U *fulfills*

$$U(t) = d^\gamma t^\gamma \exp\left(-\int_t^\infty \frac{\eta(s)}{s}\,ds\right) \tag{2.8}$$

with $|\eta(t)| \leq Bt^{-\rho}$ *for all* $t \geq t_0$, $|\gamma - \gamma_0| \leq \varepsilon$ *and* $|d - d_0| \leq \varepsilon$.

Then we have the following relations.

(i) *For all constants* $\alpha_0, d_0, \rho, A, x_0 > 0$ *and* $0 < \varepsilon < \min(d_0, 1/\alpha_0)$ *there exist constants* $B > 0$ *and* $t_0 > 1$ *such that if* f *satisfies* (C1), *then the corresponding function* U *satisfies* (C2) *with* $\gamma_0 = 1/\alpha_0$.

(ii) *Conversely, for all* $\gamma_0, d_0, \rho, B > 0$, $t_0 > 1$ *and* $0 < \varepsilon < \min(d_0, \gamma_0)$, *there exist constants* $A, x_0 > 0$ *such that if* U *satisfies* (C2), *then* f *satisfies* (C1) *with* $\alpha_0 = 1/\gamma_0$.

Observe that (2.8) implies (1.5) with

$$c = d^{\gamma} \exp\left(-\int_1^{\infty} \eta(s)/s\, ds\right) \quad \text{if } \int_1^{\infty} g(s)/s\, ds < \infty.$$

Conversely, if (1.5) is fulfilled, then (2.8) holds with

$$\begin{aligned} d &= c^{1/\gamma} \exp\left(\int_1^{\infty} \eta(s)/(\gamma s)\, ds\right) \\ &\in \left[c^{1/\gamma} \exp\left(-\int_1^{\infty} g(s)/(\gamma s)\, ds\right), c^{1/\gamma} \exp\left(\int_1^{\infty} g(s)/(\gamma s)\, ds\right)\right]. \end{aligned}$$

Therefore, essentially condition (C2) is equivalent to the conditions on U imposed in Theorem 2.1 with $g(t) = Bt^{-\rho}$ if one is merely interested in the upper tail of U. In this sense, the Hall and Welsh result is just a special case of Theorem 2.1.

If $\int_1^{\infty} g(s)/s\, ds$ is infinite, then a similar reformulation of the conditions on U in terms of the density is not available. Furthermore, if $\int_1^{\infty} g(s)/s\, ds < \infty$ but g is slowly varying, that is, $\rho = 0$, then an analogue to the implication (C2) $\Rightarrow$ (C1) with $Ax^{-\rho/\gamma}$ in (1.2) replaced by $A \int_{x^{1/\gamma}}^{\infty} g(s)/s\, ds$ 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.

3. Best attainable rates in the general case. As explained in the introduction, one obtains literally the same bounds for the rate of convergence of estimators for arbitrary extreme value indices $\gamma \in \mathbb{R}$ if the conditions on the function U imposed in Theorem 2.1 are replaced by analogous conditions on its derivative.

THEOREM 3.1. *Fix constants $c, \varepsilon > 0$, $\gamma_0 \in \mathbb{R}$ and assume that the function g: $(1, \infty) \to (0, \infty)$ satisfies the conditions of Theorem* 2.1. *Denote by $(\hat{\gamma}_n)$ an arbitrary sequence of estimators for γ and (a_n) a sequence of positive real numbers.*

Then (2.1) *holds true if*

$$\lim_{n\to\infty} P_U^n\{|\hat{\gamma}_n - \gamma| \le a_n\} = 1$$

uniformly for all functions U whose derivative satisfies (1.7) *with $|\eta(t)| \le g(t)$ for all $t > 1$ and $|\gamma - \gamma_0| \le \varepsilon$.*

One important difference between Theorems 2.1 and 3.1 is the fact that (1.7) allows of an arbitrary location parameter. Note that in many applications a good fit of a generalized Pareto tail to extreme data can only be achieved if a location parameter is taken into account [see Reis and Thomas (1997), pages 187 and 214]. As a consequence, estimators for γ that are not (at least asymptotically) location invariant, as, for example, the moment estimator $\hat{\gamma}_{n,M}$, 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 $|U(t_0) - b_0| \le \varepsilon$ where $t_0 > 1$ and $b_0 \in \mathbb{R}$ are fixed constants. To see this, consider the example

$$U(t) = b + \frac{c}{\gamma} t^\gamma = \left(\frac{c}{\gamma} + b\right) t^\gamma \exp\left(\int_1^t \frac{-\gamma b s^{-\gamma}}{c/\gamma + b s^{-\gamma}} s^{-1}\, ds\right)$$

with $\gamma > 0$, that is, in the notation of Theorem 2.1, the additional location parameter corresponds to a function η that is $(-\gamma)$-varying. It follows by Dekkers and de Haan (1993), Theorem 3.4, that for the optimal choice of k_n, the estimation error $|\hat{\gamma}_{n,M} - \gamma|$ is of the stochastic order $n^{-\gamma/(2\gamma+1)}$, so that in the case of $\rho > \gamma$, the optimal rate given in Theorem 3.1 is not attained. Likewise it can be proved that the rate of convergence is not optimal if $-\rho < \gamma < 0$. Hence (at least if the asymptotic behavior is regarded as decisive), it is not advisable to use estimators that are not location invariant in the present context.

As opposed to this, a natural candidate for an estimator that does converge at the optimal rate is the location invariant Pickands estimator $\hat{\gamma}_{n,P}$.

THEOREM 3.2. *Suppose that the assumptions of Theorem* 3.1 *are satisfied*, k_n *is defined by* (2.4) *and*

$$\liminf_{n\to\infty} \frac{a_n}{g(t_n)} \begin{cases} = \infty, & \text{if } \rho > 0, \\ > 1, & \text{if } \rho = 0. \end{cases}$$

Then

$$\lim_{n\to\infty} P_U^n\{|\hat{\gamma}_{n,P} - \gamma| \le a_n\} = 1$$

uniformly for all functions U considered in Theorem 3.1.

In contrast to the situation for $\gamma > 0$, here $\hat{\gamma}_{n,P}$ even attains the optimal bound for $\rho = 0$. On the other hand, one should not forget that this optimality only reflects the fact that the bias is asymptotically minimal [see Remark (ii) following Theorem 2.3], while the variance, which is of smaller order than the squared bias, may be large (and in fact *is* large compared with other estimator) for moderate sample sizes.

Therefore we recommend using more advanced location invariant estimators, for example, mixtures of Pickands estimators as introduced by Drees (1995a), certain generalizations of probability weighted moment estimators [see Hosking and Wallis (1987) and Drees (1997a)] or, if γ is assumed to be greater than $-1/2$, the maximum likelihood estimator [cf. Smith (1987)]. Notice that all these estimators belong to the class of statistical tail functionals as introduced in Drees (1997a), based on a Frechét differentiable functional. Therefore it can be proven by the methods used in that paper that these estimators converge uniformly at the optimal rate.

4. Proofs.

PROOF OF THEOREM 2.1. The basic idea of the proof is similar to the one used by Hall and Welsh (1984) and Farrell (1972). Fix some $l > 0$ and let

$s_n := lt_n$. Define

$$U_n^+(t) := ct^{\gamma_n^+} \exp\left(\int_1^t \frac{\eta_n^+(s)}{s}\,ds\right) \quad \text{and} \quad U_n^-(t) := ct^{\gamma_n^-} \exp\left(\int_1^t \frac{\eta_n^-(s)}{s}\,ds\right)$$

for $t \geq 1$ where

$$\gamma_n^{\pm} := \gamma_0 \pm g(s_n) \quad \text{and} \quad \eta_n^{\pm} := (\gamma_0 - \gamma_n^{\pm})1_{[1,\, s_n]}.$$

Obviously $U_n^{\pm}$ satisfy the conditions imposed in the theorem for sufficiently large n, since g is an eventually nonincreasing function converging to 0.

The densities corresponding to U_n^+ and U_n^- are

$$f_n^{\pm}(x) = \begin{cases} \dfrac{1}{c\gamma_0}\left(\dfrac{x}{c}\right)^{-(1/\gamma_0+1)}, & \text{if } c \leq x \leq cs_n^{\gamma_0}, \\[2ex] \dfrac{s_n^{-(\gamma_0+1)}}{c\gamma_n^{\pm}}\left(\dfrac{x}{cs_n^{\gamma_0}}\right)^{-(1/\gamma_n^{\pm}+1)}, & \text{if } cs_n^{\gamma_0} < x. \end{cases}$$

Thus

$$\begin{aligned}
&\int_1^{\infty}\left(1 - \frac{f_n^+(x)}{f_n^-(x)}\right)^2 f_n^-(x)\,dx \\
&\quad = \int_{cs_n^{\gamma_0}}^{\infty}\left(1 - \frac{\gamma_n^-}{\gamma_n^+}\left(\frac{x}{cs_n^{\gamma_0}}\right)^{1/\gamma_n^- - 1/\gamma_n^+}\right)^2 \frac{s_n^{-(\gamma_0+1)}}{c\gamma_n^-}\left(\frac{x}{cs_n^{\gamma_0}}\right)^{-(1/\gamma_n^-+1)} dx \\
&\quad = s_n^{-1}\int_1^{\infty}\left(1 - \frac{\gamma_n^-}{\gamma_n^+}x^{1/\gamma_n^- - 1/\gamma_n^+}\right)^2 \frac{1}{\gamma_n^-}x^{-(1/\gamma_n^-+1)}\,dx \\
&\quad = s_n^{-1}\left(1 - 2 + \frac{\gamma_n^-}{\gamma_n^{+2}}\frac{1}{2/\gamma_n^+ - 1/\gamma_n^-}\right) \\
&\quad = s_n^{-1}\frac{(\gamma_n^+ - \gamma_n^-)^2}{\gamma_n^+(2\gamma_n^- - \gamma_n^+)} \\
&\quad \sim 4l^{-(2\rho+1)}\gamma_0^{-2}n^{-1}
\end{aligned}$$

by the definitions of s_n and t_n. The Cauchy–Schwarz inequality yields

$$\begin{aligned}
&P_{U_n^+}^n\{|\hat{\gamma}_n - \gamma_n^+| \leq a_n\} \\
&\quad = \int 1_{\{|\hat{\gamma}_n - \gamma_n^+| \leq a_n\}} \prod_{i=1}^n (f_n^-(x_i))^{1/2} \prod_{i=1}^n \frac{f_n^+(x_i)}{(f_n^-(x_i))^{1/2}}\, \boldsymbol{\lambda}^n(d\mathbf{x}) \\
&\quad \leq P_{U_n^-}^n\{|\hat{\gamma}_n - \gamma_n^+| \leq a_n\}^{1/2}\left(1 + \int_1^{\infty}\left(1 - \frac{f_n^+(x)}{f_n^-(x)}\right)^2 f_n^-(x)\,dx\right)^{n/2} \\
&\quad \leq KP_{U_n^-}^n\{|\hat{\gamma}_n - \gamma_n^+| \leq a_n\}^{1/2}
\end{aligned}$$

for some $K \in (0, \infty)$, where $\boldsymbol{\lambda}^n(d\mathbf{x})$ denotes integration w.r.t. the Lebesgue measure on $\mathbb{R}^n$. Hence

$$\lim_{n\to\infty} P^n_{U_n^-}\{|\hat{\gamma}_n - \gamma_n^-| \le a_n\} = 1 \quad \text{and} \quad \liminf_{n\to\infty} P^n_{U_n^-}\{|\hat{\gamma}_n - \gamma_n^+| \le a_n\} \ge K^{-2},$$

which in turn implies

$$a_n \ge \tfrac{1}{2}(\gamma_n^+ - \gamma_n^-) = g(s_n) \sim l^{-\rho} g(t_n).$$

Because this holds for all $l > 0$, (2.1) is proved. $\square$

PROOF OF THEOREM 2.2. In Csörgő, Deuheuvels and Mason (1985), Theorem 1, (see also page 1073, line 2), it was proved that

$$\mathscr{L}\left(\frac{k_n^{1/2}}{\gamma}(\hat{\gamma}_{n,H} - \gamma - \beta_{n,\eta})\right) \to \mathscr{N}(0,1)$$

weakly where

$$\beta_{n,\eta} := \int_0^1 \eta\left(\frac{n}{k_n v}\right) dv.$$

A close inspection of the proof presented in that paper shows that this result holds uniformly for all U satisfying the conditions of Theorem 2.1 and hence

$$\left| P^n_U\{|\hat{\gamma}_{n,H} - \gamma| \le a_n\} - \mathscr{N}(0,1)\left[\frac{k_n^{1/2}}{\gamma}(-a_n - \beta_{n,\eta}), \frac{k_n^{1/2}}{\gamma}(a_n - \beta_{n,\eta})\right]\right| \to 0$$

uniformly. Since eventually $|\beta_{n,\eta}| < g(t_n)$ if $\rho > 0$ and $|\beta_{n,\eta}| < g(t_n^*)$ if $\rho = 0$, the assumptions imply that $k_n^{1/2}(-a_n - \beta_{n,\eta}) \to -\infty$ and $k_n^{1/2}(a_n - \beta_{n,\eta}) \to \infty$ uniformly for all U under consideration. Now the assertion is obvious. $\square$

PROOF OF THEOREM 2.3. First check that a Taylor expansion of the exponential function yields, for all $\delta \in (0,1)$,

$$\sup_{x\in[1-\delta,\,4+\delta]}\left|\frac{U(tx) - U(t)}{\gamma U(t)} - \frac{x^\gamma - 1}{\gamma} - \frac{x^\gamma}{\gamma}\int_1^x \frac{\eta(st)}{s}\,ds\right| = O(g^2(t))$$

uniformly for all functions U satisfying the conditions of Theorem 2.1. Therefore, following the lines of Drees (1995a), proof of Theorem 2.1, one can define a sequence of r.v.'s $\tilde{\gamma}_n$ and a Brownian motion W such that the variational distance between the distributions of $\tilde{\gamma}_n$ and $\hat{\gamma}_{n,P}$ vanishes asymptotically $\|\mathscr{L}(\tilde{\gamma}_n) - \mathscr{L}(\hat{\gamma}_{n,P})\| \to 0$ and

$$\begin{aligned}
&\Bigg|\tilde{\gamma}_n - \gamma - \frac{1}{(1-2^{-\gamma})\log(2)}\Bigg(\frac{\gamma}{k_n}\Big(W(k_n) - \frac{1+2^{-\gamma}}{2}W(2k_n) + 2^{-(\gamma+2)}W(4k_n)\Big)\\
&\qquad\qquad - 2^{-\gamma}\int_{1/4}^{1/2}\frac{\eta(sn/k_n)}{s}\,ds + \int_{1/2}^{1}\frac{\eta(sn/k_n)}{s}\,ds\Bigg)\Bigg| \qquad (4.1)\\
&\quad = O\left(\frac{\log(k_n)}{k_n} + g^2\left(\frac{n}{k_n}\right)\right)
\end{aligned}$$

almost surely uniformly for all U under consideration.

Under the conditions of the theorem, one has $k_n^{-1}\max(|W(k_n)|, |W(2k_n)|, |W(4k_n)|) = O_p(k_n^{-1/2}) = o_P(a_n)$ and $\log(k_n)/k_n = o(a_n)$. Hence the assertion follows from

$$\left|-2^{-\gamma}\int_{1/4}^{1/2}\frac{\eta(sn/k_n)}{s}\,ds + \int_{1/2}^{1}\frac{\eta(sn/k_n)}{s}\,ds\right|$$
$$\leq g(n/k_n)\log(2)(2^{-\gamma}4^{\rho}+2^{\rho})(1+o(1))$$

in combination with (2.3), (2.4) and (2.6). □

Proof of Lemma 2.1. First assume that condition (C1) is fulfilled. Then we have $1-F(x) = dx^{-1/\gamma}(1+\tilde{r}(x))$ with $|\tilde{r}(x)| \leq Ax^{-\rho/\gamma}/(\rho+1)$ for $x \geq x_0$, so that

$$U(t) = (dt)^{\gamma}(1+h(t)), \tag{4.2}$$

where the tail of h is given by the condition

$$1-F(U(t)) = t^{-1} \quad \Rightarrow \quad (1+h(t))^{1/\gamma} = 1+\tilde{r}(U(t)). \tag{4.3}$$

Now consider

$$\eta(t) := \frac{tU'(t)}{U(t)} - \gamma = \frac{1}{tf(U(t))U(t)} - \gamma$$
$$= \frac{\gamma(1+h(t))^{1/\gamma}}{1+r(U(t))} - \gamma = \gamma\frac{\tilde{r}(U(t)) - r(U(t))}{1+r(U(t))}.$$

Since by (4.2) and (4.3) one has $(dt/2)^{\gamma} < U(t) < (2dt)^{\gamma}$ for sufficiently large t, it follows that

$$|\eta(t)| \leq \gamma A\left(\frac{1}{\rho+1}+1\right)\frac{(U(t))^{-\rho/\gamma}}{1-A(U(t))^{-\rho/\gamma}} \leq \gamma A\frac{\rho+2}{\rho+1}2(d/2)^{-\rho}t^{-\rho}, \qquad t \geq t_0,$$

for some $t_0 > 1$. Thus in view of

$$U(t) = \frac{U(t_0)}{t_0^{\gamma}}t^{\gamma}\exp\left(\int_{t_0}^{t}\frac{\eta(s)}{s}\,ds\right) = \tilde{d}t^{\gamma}\exp\left(-\int_{t}^{\infty}\frac{\eta(s)}{s}\,ds\right),$$

and (4.2), assertion (C2) is immediate.

For the proof of (ii), assume that U satisfies (C2). Then

$$\frac{1}{1-F(x)} = \frac{1}{d}x^{1/\gamma}(1+h(x)) \quad \text{with } 1+h(x) = \exp\left(\int_{1/(1-F(x))}^{\infty}\frac{\eta(s)}{\gamma s}\,ds\right)$$

for sufficiently large x. Therefore

$$f(x) = \frac{(1-F(x))^2}{U'(1/(1-F(x)))} = \frac{d}{\gamma}x^{-(1/\gamma+1)}(1+r(x)),$$

where

$$r(x) = \left(\exp\left(-\int_{1/(1-F(x))}^{\infty} \frac{\eta(s)}{\gamma s}\, ds \right) - 1 - \frac{\eta(1/(1-F(x)))}{\gamma} \right) \\ \times \left(1 + \frac{\eta(1/(1-F(x)))}{\gamma} \right)^{-1}$$

for almost all sufficiently large x. In the same way as above one can conclude that $1/(1-F(x)) \geq x^{1/\gamma}/(2d)$ and thus by the mean value theorem

$$|r(x)| \leq 2^{\rho+1}\left(\frac{2}{\rho}+1\right)\frac{B}{\gamma_0-\varepsilon}(d_0+\varepsilon)^{\rho}x^{-\rho/\gamma} =: Ax^{-\rho/\gamma}$$

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). □

PROOF OF THEOREM 3.1. Since every neighborhood of 0 includes subintervals of $(0,\infty)$ [and $(-\infty,0)$] and (2.1) is independent of ε, w.l.o.g. one may assume that either $\gamma_0 > 0$ or $\gamma_0 < 0$. We concentrate on the latter case, because the former can be treated in a similar way to Theorem 2.1.

Fix some $l > 0$, $\tau > 0$ and $\delta \in (0,1)$. Let $s_n := lt_n$ and

$$b_n := \frac{1-\delta}{1+\tau} g\Big(((1+1/\tau)^{-1/\gamma_0}+\delta)s_n\Big).$$

Define

$$\gamma_n^{\pm} := \gamma_0 \pm b_n, \qquad \tau_n^{+} := \tau, \qquad \tau_n^{-} := \left(\frac{\gamma_n^{+}}{\gamma_n^{-}}\left(1+\frac{1}{\tau}\right)-1\right)^{-1},$$

$$\beta_n^{\pm} := \left(\frac{1}{\gamma_0}+\tau_n^{\pm}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma_n^{\pm}}\right)\right)^{-1} \quad \text{and} \quad \chi_n^{\pm} := \left(1+\frac{1}{\tau_n^{\pm}}\right)^{-1/\beta_n^{\pm}}.$$

Then

$$U_n^{\pm}(t) := c \cdot \begin{cases} t^{\gamma_0}/\gamma_0, & \text{if } 1 \leq t \leq s_n, \\ s_n^{\gamma_0}(1/\gamma_0 - 1/\beta_n^{\pm}) + s_n^{\gamma_0-\beta_n^{\pm}} t^{\beta_N^{\pm}}/\beta_n^{\pm}, & \text{if } s_n < t \leq \chi_n^{\pm}s_n, \\ s_n^{\gamma_0-\gamma_n^{\pm}}\chi_n^{\beta_n^{\pm}-\gamma_n^{\pm}} t^{\gamma_n^{\pm}}/\gamma_n^{\pm}, & \text{if } \chi_n^{\pm}s_n < t \end{cases}$$

satisfy the conditions imposed in the theorem. By somewhat lengthy computations, one can prove that the χ^2-distance between the corresponding densities f_n^{+} and f_n^{-} is of the order $1/n$. [For details we refer to Drees (1995b).] As in the proof of Theorem 2.1, it follows that

$$a_n \geq b_n \sim \frac{1-\delta}{1+\tau}\Big(\Big((1+1/\tau)^{-1/\gamma_0}+\delta\Big)l\Big)^{-\rho} g(t_n)$$

for all $l, \tau > 0$ and $\delta \in (0,1)$. Now let l tend to 0 for $\rho > 0$ and let δ and τ converge to 0 in the case $\rho = 0$ to obtain the assertion. □

REMARK. In view of the proof of Theorem 2.1, one is tempted to make the simple approach

$$(U_n^{\pm})'(t) := ct^{\gamma_n^{\pm}-1}\exp\biggl(\int_1^t \frac{(\gamma_0-\gamma_n^{\pm})1_{[1,\,s_n]}(s)}{s}\,ds\biggr).$$

However, since then $\int_1^\infty (U_n^+)'(t)\,dt \neq \int_1^\infty (U_n^-)'(t)\,dt$ for $\gamma_0 < 0$, the distributions pertaining to U_n^+ and U_n^- have different supports. Therefore the χ^2–distance between f_n^+ and f_n^- is of larger order than $1/n$ if $\gamma_0 < -1/2$. □

PROOF OF THEOREM 3.2. By a Taylor expansion of the exponential function it is easily seen that

$$\begin{aligned}
&\biggl|\frac{U(tx)-U(t)}{tU'(t)} - \frac{x^\gamma-1}{\gamma} - \int_1^x u^{\gamma-1}\int_1^u \frac{\eta(st)}{s}\,ds\,du\biggr|\\
&\qquad = \biggl|\int_1^x u^{\gamma-1}\biggl(\exp\biggl(\int_1^u \frac{\eta(st)}{s}\,ds\biggr) - 1 - \int_1^u \frac{\eta(st)}{s}\,ds\biggr)\,du\biggr| = O(g^2(t))
\end{aligned}$$

uniformly for all $x \in [1-\delta, 4+\delta]$ and all U satisfying the assumptions of the theorem. Therefore we get the following counterpart of (4.1):

$$\begin{aligned}
&\biggl|\tilde{\gamma}_n - \gamma - \frac{\gamma}{(1-2^{-\gamma})\log(2)}\biggl(\frac{1}{k_n}\biggl(W(k_n) - \frac{1+2^{-\gamma}}{2}W(2k_n) + 2^{-(\gamma+2)}W(4k_n)\biggr)\\
&\qquad\qquad + 2^\gamma\int_{1/4}^{1/2} u^{\gamma-1}\int_u^1 \frac{\eta(sn/k_n)}{s}\,ds\,du\\
&\qquad\qquad\qquad - \int_{1/2}^1 u^{\gamma-1}\int_u^1 \frac{\eta(sn/k_n)}{s}\,ds\,du\biggr)\biggr|\\
&\quad = O\biggl(\frac{\log(k_n)}{k_n} + g^2\biggl(\frac{n}{k_n}\biggr)\biggr)
\end{aligned}$$

uniformly where by convention $(1-t^{-\gamma})/\gamma := \log(t)$ if $\gamma = 0$.

Integration by parts and the ρ-variation of g yield

$$\begin{aligned}
&\biggl|2^\gamma\int_{1/4}^{1/2} u^{\gamma-1}\int_u^1 \frac{\eta(sn/k_n)}{s}\,ds\,du - \int_{1/2}^1 u^{\gamma-1}\int_u^1 \frac{\eta(sn/k_n)}{s}\,ds\,du\biggr|\\
&\quad = \biggl|2^\gamma\int_{1/4}^{1/2} \frac{\eta(sn/k_n)}{s}\,\frac{s^\gamma-4^{-\gamma}}{\gamma}\,ds + \int_{1/2}^1 \frac{\eta(sn/k_n)}{s}\,\frac{1-s^\gamma}{\gamma}\,ds\biggr|\\
&\quad = O(g(n/k_n)) \quad \text{if } \rho > 0\\
&\quad \le g(n/k_n)\log(2)(1-2^{-\gamma})/\gamma(1+o(1)) \quad \text{if } \rho = 0.
\end{aligned}$$

Now the assertion follows in the same way as in the proof of Theorem 2.3. □

Acknowledgments. The first version of the paper was written while I was visiting Erasmus University, Rotterdam. Thanks are due to L. de Haan for his kind hospitality. His valuable suggestions and remarks are gratefully acknowledged. Moreover, I would like to thank an Associate Editor and the referees for their remarks, which led to an improvement of the presentation of the results.

REFERENCES

BEIRLANT, J., TEUGELS, J. L. and VYNCKIER, P. (1996). *Practical Analysis of Extreme Values.* Leuven Univ. Press.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation.* Cambridge Univ. Press.

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

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. (1995a). Refined Pickands estimators of the extreme value index. *Ann. Statist.* **23** 2059–2080.

DREES, H. (1995b). Optimal rates of convergence for estimates of the extreme value index. Technical report, Univ. Cologne.

DREES, H. (1997a). On smooth statistical tail functionals. *Scand. J. Statist.* To appear.

DREES, H. (1997b). A general class of estimators of the extreme value index. *J. Statist. Plann. Inference*. To appear.

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.

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, New York.

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

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.* **80** 3–34.

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

REISS, R.-D. and THOMAS, M. (1997). *Statistical Analysis of Extreme Values.* Birkhäuser, Basel.

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

MATHEMATISCHES INSTITUT
UNIVERSITÄT ZU KÖLN
WEYERTAL 86–90
50931 KÖLN
GERMANY
E-MAIL: hdrees@mi.uni-koeln.de