The Annals of Statistics
2001, Vol. 29, No. 1, 266–294

MINIMAX RISK BOUNDS IN EXTREME VALUE THEORY

BY HOLGER DREES

University of Heidelberg

Asymptotic minimax risk bounds for estimators of a positive extreme value index under zero-one loss are investigated in the classical i.i.d. setup. To this end, we prove the weak convergence of suitable local experiments with Pareto distributions as center of localization to a white noise model, which was previously studied in the context of nonparametric local density estimation and regression. From this result we derive upper and lower bounds on the asymptotic minimax risk in the local and in certain global models as well. Finally, the implications for fixed-length confidence intervals are discussed. In particular, asymptotic confidence intervals with almost minimal length are constructed, while the popular Hill estimator is shown to yield a little longer confidence intervals.

1. Introduction. Consider i.i.d. random variables X_i, $i \in \mathbb{N}$, whose distribution function (d.f.) F belongs to the weak domain of attraction of an extreme value d.f. G_γ, that is,

$$\mathscr{L}\left(a_n^{-1}\left(\max_{1\le i\le n} X_i - b_n\right)\right) \longrightarrow G_\gamma \quad \text{weakly}$$

for some constants $a_n > 0$ and $b_n \in \mathbb{R}$. Here $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$ for $1+\gamma x > 0$, which is interpreted as $G_0(x) = \exp(-e^{-x})$ if $\gamma = 0$. The shape of the upper tail of F is largely determined by the real parameter $\gamma = \gamma(F)$, the so-called extreme value index. Several estimators of γ have been discussed in literature [see, e.g., Hill (1975), Pickands (1975), Csörgő, Deheuvels and Mason (1985), Dekkers et al. (1989) and Drees (1998a,b)], yet much less is known about the best achievable performance of arbitrary estimators for γ. It is the main goal of the present paper to establish asymptotic risk bounds for arbitrary estimators of a positive extreme value index, and thereby to determine how much space is left for further improvements of known estimators.

Falk (1995) and Marohn (1997), among others, use LAN-theory to calculate asymptotically sharp risk bounds for sequences of estimators $\hat\gamma_n$ that depend only on the k_n largest order statistics. However, under the conditions they imposed on the sequence $(k_n)_{n\in\mathbb{N}}$, the distribution of $\hat\gamma_n$ under F can be approximated in the variational distance by its distribution under a suitable generalized Pareto distribution [Reiss (1989), Corollary 5.5.5], that is, essentially not semiparametric, but parametric models are considered. Moreover, estimators attaining these risk bounds are inefficient in the following sense:

Received August 1998; revised March 2000.
AMS 2000 *subject classifications*. Primary 62C20, 62G32; secondary 62G05, 62G15.
Key words and phrases. Confidence intervals, convergence of experiments, extreme value index, Gaussian shift, Hill estimator, local experiment, minimax affine estimator, white noise, zero-one loss.

For any family $\mathscr{F}$ of distributions considered in those papers one may find estimator sequences $\tilde{\gamma}_n$, $n \in \mathbb{N}$, which are based on a larger number $\tilde{k}_n > k_n$ of order statistics, such that for any sequence $(\hat{\gamma}_n)_{n\in\mathbb{N}}$ of the aforementioned type one has $\tilde{\gamma}_n - \gamma(F) = o_P(\hat{\gamma}_n - \gamma(F))$ uniformly over all $F \in \mathscr{F}$.

Here we consider genuinely semiparametric models of heavy–tailed distributions, that is, we assume $F \in D(G_\gamma)$ for some $\gamma > 0$. While for environmental data light–tailed distributions (with $\gamma < 0$) often cannot be excluded a priori, financial data sets and teletraffic data usually exhibit heavy tails [see Embrechts, Klüppelberg and Mikosch (1997), Section 6.2, particularly Example 6.2.9, and Adler, Feldman and Taqqu (1998), Chapter 1, for several examples]. Another important field of applications are non–life (re–)insurances, in particular fire insurances [Embrechts Klüppelberg and Mikosch (1997), page 332ff.], storm insurances [Rootzén and Tajvidi (1997)] and business interruption insurances [Zajdenweber (1995)], where indeed the very existence of heavy tails, which show up in the data sets of insurance claims, is the main reason why classical methods to calculate premiums based on the central limit theorem are supplemented (or even replaced) by extreme value procedures.

For the case $\gamma > 0$, Hall and Welsh (1984) established optimal uniform rates of convergence over families of d.f.'s with density of the type $f(x) = cx^{-(1/\gamma+1)}(1 + r(x))$, where $|r(x)| \le Ax^{-\rho/\gamma}$ for fixed constants $A, \rho > 0$. This result was generalized in Theorem 2.1 of Drees (1998c), which can be reformulated in the following way. Define

$$\mathscr{F}(\gamma_1, c, \varepsilon, u) := \left\{ F \text{ d.f.} \;\middle|\; F^{-1}(1-t) = ct^{-\gamma} \exp\left(\int_t^1 \eta(s)/s\, ds\right), \right. \qquad (1.1)$$
$$\left. |\gamma - \gamma_1| \le \varepsilon,\ |\eta(t)| \le u(t), \quad t \in (0, 1] \right\}$$

where $\gamma_1 > \varepsilon > 0$, $c > 0$ and u denotes a bounded function that is ρ–varying at 0 for some $\rho \ge 0$ and satisfies $\lim_{t\downarrow 0} u(t) = 0$ yet is bounded away from 0 on $[\delta, 1]$ for all $\delta > 0$. Observe that every d.f. $F \in \mathscr{F}(\gamma_1, c, \varepsilon, u)$ belongs to the weak domain of attraction of G_γ for some $\gamma > 0$. Conversely, if F satisfies the well-known von Mises condition $g(x) := (1 - F(x))/(xf(x)) \to \gamma$ as $x \to \infty$, then $F^{-1}(1-t) = ct^{-\gamma}\exp(\int_t^1 \eta(s)/s\, ds)$ for some $c > 0$ and $\eta(s) := g(F^{-1}(1-s)) - \gamma$ tends to 0 as s tends to 0; in particular, this holds true if F belongs to the weak domain of attraction of G_γ and if it has an eventually monotone density. Notice that $g(x) = \gamma$ (i.e. $\eta = 0$) corresponds to the Pareto d.f. $F_\gamma(x) = (x/c)^{-1/\gamma}$. Hence, essentially $\mathscr{F}(\gamma_1, c, \varepsilon, u)$ consists of smooth distributions in the domain of attraction of an extreme value distribution G_γ with $\gamma \in (\gamma_1 - \varepsilon, \gamma_1 + \varepsilon)$ such that the distance to the pertaining Pareto d.f. F_γ, measured in terms of the difference between the von Mises function g and its limit, is bounded by the function u. In view of the prominent role the (generalized) Pareto distributions play in extreme value theory [cf. Reiss (1989), Chapter 5], this choice of a semiparametric model seems natural. Furthermore, it can be shown that for any d.f. F satisfying the von Mises condition one can find a regularly varying function u that converges to 0 and dominates η, so that $F \in \mathscr{F}(\gamma_1, c, \varepsilon, u)$ for $\gamma_1 \in (\gamma(F) - \varepsilon, \gamma(F) + \varepsilon)$.

Let $(t_n)_{n\in\mathbb{N}}$ be a sequence satisfying

$$\lim_{n\to\infty} u(t_n)(nt_n)^{1/2} = 1; \tag{1.2}$$

the existence and asymptotic uniqueness of such a sequence follows from Theorem 1.5.12 of Bingham, Goldie and Teugels (1987).

Then, for arbitrary estimators $\hat{\gamma}_n$ based on a sample of size n, one has

$$\liminf_{n\to\infty} \sup_{F\in\mathscr{F}(\gamma_1,c,\varepsilon,u)} P_F^n\{|\hat{\gamma}_n - \gamma(F)| > au(t_n)\} > 0 \tag{1.3}$$

for all $a < \infty$ if $\rho > 0$, and all $a < 1$ if $\rho = 0$. Here P_F^n denotes the joint distribution of n i.i.d. random variables with d.f. F.

If $\rho = 0$, that is, u is slowly varying at 0, then $u(t_n)$ is an asymptotically *sharp lower bound* for the estimation error in the following sense: For any a_n, $\liminf_{n\to\infty} a_n/u(t_n) < 1$ implies $\limsup_{n\to\infty} \sup_{F\in\mathscr{F}(\gamma_1,c,\varepsilon,u)} P_F^n\{|\hat{\gamma}_n - \gamma(F)| > a_n\} > 0$ for all estimators $\hat{\gamma}_n$, whereas there are estimators $\hat{\gamma}_n$ (e.g., the well-known Hill estimator based on a suitable number of largest order statistics) such that $\lim_{n\to\infty} \sup_{F\in\mathscr{F}(\gamma_1,c,\varepsilon,u)} P_F^n\{|\hat{\gamma}_n - \gamma(F)| > a_n\} = 0$ if $\liminf_{n\to\infty} a_n/u(t_n) > 1$.

In contrast, such a sharp bound does not exist if $\rho > 0$; in this case inequality (1.3) merely describes the optimal *rate of convergence*. For example, if one chooses $u(t) = At^\rho$ (which essentially leads to the model considered by Hall and Welsh), then $u(t_n) \sim A^{1/(2\rho+1)} n^{-\rho/(2\rho+1)}$; that is, the optimal rate equals $n^{-\rho/(2\rho+1)}$. (Here $a_n \sim b_n$ means $a_n/b_n \to 1$.) Moreover, it was proved that the Hill estimator based on a suitable number of order statistics converges with the optimal rate [Drees (1998c), Theorem 2.2]. However, since the same holds true for a large class of estimators of γ, including Pickands' (1975) estimator and the maximum likelihood estimator examined by Smith (1987) [see Theorem 4.3 of Drees (1998d)], inequality (1.3) is too crude to yield a useful benchmark for the evaluation of the performance of estimators for the extreme value index. [A similar remark applies to a rather rough risk bound for the Hall–Welsh model that was derived by Donoho and Liu (1991a).]

In more general extreme value models allowing arbitrary $\gamma \in \mathbb{R}$ the situation is somewhat different, in that estimators of the extreme value index that are not location invariant [like, e.g, the moment estimator proposed by Dekkers, Einmahl and de Haan (1989)] usually do not converge with the optimal rate, provided the finite right endpoint is not known in advance in case of $\gamma < 0$, whereas a large class of location invariant estimators do attain the optimal rate [Drees (1998c), Section 3, and Drees (1998d), Theorem 4.3]. Hence, unlike in the case $\gamma > 0$, the result about the optimal rate of convergence renders it possible to sort out certain well-known estimators with unfavorable asymptotic properties. This new phenomenon is due to the qualitatively larger impact of a shift in the data if the underlying distribution has a finite right endpoint (compared with heavy–tailed distributions). However, since it was seen in Drees (1998c) that the incorporation of a location parameter leads to substantially more complicated local models, these general extreme value models will be examined elsewhere.

Here we will establish an asymptotically almost sharp lower bound for the left-hand side of (1.3); that is, a lower bound for the asymptotic minimax risk under the zero-one loss generated by the function $\mathbb{1}_{[-au(t_n),au(t_n)]^c}$. This new bound may serve as a benchmark to assess the efficiency of estimators in the present semiparametric setup. For the aforementioned reasons, we concentrate on the case $\rho > 0$.

First, it is proved in Section 2 that certain local experiments with Pareto distributions as center of localization converge to white noise models of a type that was previously studied in the context of nonparametric local density estimation and regression. Therefore, the powerful tools of LeCam's theory of the convergence of experiments show that the asymptotic minimax risk for the sequence of local experiments is bounded from below by the minimax risk in the limit experiment, which we investigate in Section 3 using methods introduced by Ibragimov and Khas'minskii (1985) and Donoho and Liu (1991b). In Section 4, implications for the minimax problem described above and for the construction of fixed-length confidence intervals are discussed. In particular, it is shown that in typical situations the Hill estimator leads to a confidence interval which is merely a few percent longer than the shortest possible fixed-size confidence interval. All proofs are collected in Section 5.

It is worth mentioning that although we focus on a lower bound on the minimax risk of estimators for γ under zero-one loss, the main result of Section 2 also makes it feasible to establish risk bounds under quadratic loss or even risk bounds for estimators of certain extreme quantiles (see Section 2).

For the sake of notational simplicity, in the sequel we assume that $c = 1$ and let $\mathscr{F}(\gamma_1, \varepsilon, u) := \mathscr{F}(\gamma_1, 1, \varepsilon, u)$.

2. Weak convergence of local experiments. As mentioned above, our investigation of the minimax risk is based on a result about the weak convergence of certain local experiments related to the model $\mathscr{F}(\gamma_1, \varepsilon, u)$. The basic idea of this approach is to consider sequences of local alternatives converging to a fixed center of localization in such a way that the increasing degree of difficulty to discriminate between these alternatives and the center of localization compensates the increase of information contained in the sample as n tends to infinity.

Relations (1.2) and (1.3) show that the function u, which describes the maximal deviation between the upper quantile function (q.f.) $F^{-1}(1-t)$ and the pertaining Pareto q.f. $t^{-\gamma}$, determines the optimal rate of convergence of estimators for γ. Hence it is natural to fix a Pareto q.f. $F_0^{-1}(1-t) = t^{-\gamma_0}$ as the center of localization and then to consider alternatives with extreme value index converging to γ_0 with the rate $u(t_n)$. More specifically, we define

$$\begin{aligned} F_{n,h}^{-1}(1-t) &:= F_{n,h,\gamma_0}^{-1}(1-t) \\ &:= t^{-\gamma_0} \exp\left(\int_t^1 \frac{d_n h(c_n s)}{s}\, ds\right) \\ &= t^{-(\gamma_0 + d_n h(0))} \exp\left(\int_t^1 \frac{d_n(h(c_n s) - h(0))}{s}\, ds\right), \qquad t \in (0,1], \end{aligned}$$

where $d_n \to 0$ and $c_n \to \infty$; in the applications below we choose $d_n = u(t_n)$ and $c_n = 1/t_n$, such that (1.2) reads as

$$\lim_{n\to\infty} \frac{n d_n^2}{c_n} = 1. \tag{2.1}$$

To discuss the effect of this localization, assume for the time being that the function h, which acts as the local parameter, has compact support, say [0,1]. Then $F_{n,h}^{-1} = F_0^{-1}$ on $[0, 1 - 1/c_n]$, that is, the deviation from the Pareto q.f. is more and more concentrated in the tail. A similar approach, which gave rise to the concept of extreme value tangents, was proposed by Janssen and Marohn (1994). Likewise, Low (1992) demonstrated that a sequence of local experiments, in which the difference between the densities of the center of localization and of the local alternatives, respectively, is supported by a compact interval shrinking towards a fixed point, is appropriate for local density estimation problems [cf. also Low (1997)]. Then it was proved that this sequence converges to essentially the same white noise model that occurs as the limit experiment in Theorem 2.1 below. The fact that local extreme value, density and regression models converge to the same type of white noise model underpins the close relation between the corresponding semiparametric estimation problems.

In order to ensure that $F_{n,h}^{-1}$ actually is a q.f., one has to assume that $\gamma_0 + d_n h(c_n s) \ge 0$ for all $s \in (0,1]$. Moreover, for the application to our minimax problem, we must assure that $F_{n,h} \in \mathscr{F}(\gamma_1, \varepsilon, u)$, and thus in particular $d_n(h(c_n s) - h(0)) = d_n h(c_n s) + o(1) \le u(s)$ which is bounded for $s \in (0,1]$. For the proof of the main result of this section we need uniform versions of these restrictions, that is, we suppose

$$\begin{aligned} h \in \mathscr{H}^{(n)} &:= \Big\{ h \in L_2[0,\infty) \;\Big|\; -\gamma_0 < \inf_{m\ge n} d_m \inf_{s\in(0,1]} h(c_m s) \\ &\qquad\qquad \le \sup_{m\ge n} d_m \sup_{s\in(0,1]} h(c_m s) < \infty \Big\}, \\ \uparrow \mathscr{H} &:= \Big\{ h \in L_2[0,\infty) \;\Big|\; -\gamma_0 < \liminf_{n\to\infty} d_n \inf_{s\in(0,1]} h(c_n s) \\ &\qquad\qquad \le \limsup_{n\to\infty} d_n \sup_{s\in(0,1]} h(c_n s) < \infty \Big\}. \end{aligned}$$

In particular, bounded square integrable functions belong to $\mathscr{H}$.

For $h \in \mathscr{H}^{(n)}$ denote by $P^n_{n,h} = P^n_{n,h,\gamma_0}$ the distribution of n i.i.d. random variables with q.f. $F^{-1}_{n,h}$ and let $P_0 := P_{1,0}$ be the Pareto distribution with q.f. F_0^{-1}. Moreover,

$$Q_h := Q_{h,\gamma_0} := \mathscr{L}\left(\frac{1}{\gamma_0}\int_0^{\cdot} h(s)\,ds + W\right)$$

where W denotes a standard Brownian motion (defined on some abstract probability space) with paths belonging to the space $C[0,\infty)$ of continuous functions on the positive real line, which is equipped with the σ–field $\mathscr{C}[0,\infty)$ generated by the canonical projections. (In the sequel, often the index γ_0 is omitted, when a fixed local model is considered.)

THEOREM 2.1. *Suppose that $d_n > 0$, $d_n \to 0$ and $c_n \to \infty$ are sequences satisfying* (2.1). *Then*

$$\left(\mathbb{R}^n, \mathbb{B}^n, (P^n_{n,h})_{h\in\mathscr{H}^{(n)}}\right) \longrightarrow \left(C[0,\infty), \mathscr{C}[0,\infty), (Q_h)_{h\in\mathscr{H}}\right)$$

weakly, that is, for all $h_1,\dots,h_m \in \mathscr{H}$ one has

$$\mathscr{L}\left(\left(\log\frac{P^n_{n,h_i}}{P^n_0}\right)_{1\le i\le m}\,\Big|\,P^n_0\right) \longrightarrow \mathscr{L}\left(\left(\log\frac{Q_{h_i}}{Q_0}\right)_{1\le i\le m}\,\Big|\,Q_0\right)$$
$$= \mathscr{N}\left(\left(-\frac{\|h_i\|^2}{2\gamma_0^2}\right)_{1\le i\le m}, \left(\frac{\langle h_i, h_j\rangle}{\gamma_0^2}\right)_{1\le i,j\le m}\right)$$

weakly, where $\langle h_i, h_j\rangle := \int_0^\infty h_i(s)h_j(s)\,ds$ and $\|h\| = \langle h, h\rangle^{1/2}$ are the inner product and the norm, respectively, of the Hilbert space $L_2[0,\infty)$.

The main step in the proof of this result is to verify a kind of L_2–differentiability in the tail, which implies an approximation of the loglikelihood:

PROPOSITION 2.1. *Under the assumptions of Theorem* 2.1, *one has*

$$\lim_{n\to\infty}\int_1^\infty \left[n^{1/2}\left(f^{1/2}_{n,h}(x) - f^{1/2}_0(x)\right) - \tfrac{1}{2}g_{n,h}(x)f^{1/2}_0(x)\right]^2 dx = 0$$

where $f_{n,h}$ and f_0 are the Lebesgue-densities of $P_{n,h}$ and P_0, respectively, and

$$g_{n,h}(x) := \frac{n^{1/2}d_n}{\gamma_0}\left(\int_{x^{-1/\gamma_0}}^1 \frac{h(c_n s)}{s}\,ds - h\left(c_n x^{-1/\gamma_0}\right)\right), \qquad x \ge 1.$$

Consequently,

$$\log\frac{dP^n_{n,h}}{dP^n_0}(x_1,\dots,x_n) = n^{-1/2}\sum_{i=1}^n g_{n,h}(x_i) - \frac{\|h\|^2}{2\gamma_0^2} + o_{P_0^n}(1). \tag{2.2}$$

By virtue of LeCam's asymptotic minimax theorem, Theorem 2.1 has implications for our estimation problem. Observe that in the local model the extreme value index $\gamma(F_{n,h}) = \gamma_0 + d_n h(0)$ is determined by $h(0)$, which hence is the local parameter of interest. An estimator $\widehat{h(0)}$ for this parameter defines an estimator for the extreme value index via $\hat{\gamma}_n = \gamma_0 + d_n \widehat{h(0)}$, so that $|\hat{\gamma}_n - \gamma(F_{n,h})|/d_n = |\widehat{h(0)} - h(0)|$. However, recall from the minimax theorem that in the limit experiment one has to consider the infimum over all *randomized* estimators for $h(0)$, that is, Markov kernels K from $(C[0,\infty), \mathscr{C}[0,\infty))$ to $(\mathbb{R}, \mathbb{B})$; the pertaining risk equals $\int K(x, [h(0)-a, h(0)+a]^c)\, Q_h(dx)$. Thus we have the following result:

COROLLARY 2.1. *Let $(\hat{\gamma}_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of estimators for γ. Then, under the assumptions of Theorem 2.1, for all $\tilde{\mathscr{H}} \subset \mathscr{H}$ and all $a > 0$*

$$\liminf_{n\to\infty} \sup_{h\in\tilde{\mathscr{H}}\cap\mathscr{H}^{(n)}} P^n_{n,h}\{|\hat{\gamma}_n - \gamma(F_{n,h})| > a d_n\} \geq \inf_K \sup_{h\in\tilde{\mathscr{H}}} \int K\Big(x, [h(0)-a, h(0)+a]^c\Big)\, Q_h(dx)$$

where on the right-hand side the infimum is taken over all randomized estimators K of $h(0)$.

Of course, the assertion holds true if we consider randomized estimators of the extreme value index, too.

Next we examine the implications of this result for the minimax estimation in the global model $\mathscr{F}(\gamma_1, \varepsilon, u)$. For the clarity of exposition, we concentrate on the case $u(t) = At^\rho$ for some $A, \rho > 0$, corresponding to the model considered by Hall and Welsh, yet in a sequence of remarks we indicate the necessary changes when dealing with more general ρ–varying boundary functions u with $\rho > 0$ satisfying the conditions stated below (1.1).

To employ Corollary 2.1 with $d_n = u(t_n) = A^{1/(2\rho+1)} n^{-\rho/(2\rho+1)}$ and $c_n = 1/t_n = A^{2/(2\rho+1)}\, n^{1/(2\rho+1)}$, one must ensure that eventually the local models are included in $\mathscr{F}(\gamma_1, \varepsilon, u)$, that is,

$$u(t_n)|h(s/t_n) - h(0)| \leq u(s),\ s \in (0,1] \iff |h(s) - h(0)| \leq s^\rho, \qquad s > 0.$$

Furthermore, we assume that $h \in L_2[0,\infty)$ is bounded. Thus, let

$$\mathscr{H}_\rho := \left\{ h \in L_2[0,\infty) \,\middle|\, |h(s) - h(0)| \leq s^\rho,\ s > 0,\ \sup_{s>0} |h(s)| < \infty \right\} \subset \mathscr{H}.$$

COROLLARY 2.2. *For $a, \rho > 0$, $u(t) = At^\rho$ for $t \in (0,1]$, and an arbitrary sequence of estimators $(\hat{\gamma}_n)_{n\in\mathbb{N}}$ for γ, we have*

$$\liminf_{n\to\infty} \sup_{F\in\mathscr{F}(\gamma_1,\varepsilon,u)} P^n_F\{|\hat{\gamma}_n - \gamma(F)| > a u(t_n)\} \geq \sup_{|\gamma_0-\gamma_1|<\varepsilon} \inf_K \sup_{h\in\mathscr{H}_\rho} \int K\Big(x, [h(0)-a, h(0)+a]^c\Big)\, Q_{h,\gamma_0}(dx).$$

Therefore, a lower bound on the minimax risk in the white noise experiment defines a lower asymptotic minimax risk bound in the model $\mathscr{F}(\gamma_1, \varepsilon, u)$, too.

REMARK. For more general ρ–varying functions u ($\rho > 0$), one may use the Potter bounds [Bingham, Goldie and Deheuvels (1987), Theorem 1.5.6] to show that for all

$$h \in \tilde{\mathscr{H}}_\rho := \Big\{ h \in L_2[0,\infty) \,\Big|\, \sup_{s>0} |h(s)| < \infty, \text{ for some } \delta > 0$$
$$|h(s) - h(0)| \le (1-\delta)\min(s^{\rho-\delta}, s^{\rho+\delta}),\ s > 0 \Big\}$$

one has $F_{n,h} \in \mathscr{F}(\gamma_1, \varepsilon, u)$ for sufficiently large n. Therefore, the analog to Corollary 2.2 with $\mathscr{H}_\rho$ replaced by $\tilde{\mathscr{H}}_\rho$ holds true in this situation [where, of course, $d_n = u(t_n)$ and $c_n = 1/t_n$ with t_n satisfying (1.2) differ from the constants in the Hall–Welsh model]. □

Likewise, one gets asymptotic lower risk bounds under arbitrary lower semicontinuous, level compact loss functions, for example, under quadratic loss. Furthermore, one may also obtain asymptotic lower risk bounds for estimators of extreme quantiles

$$F_{n,h}^{-1}(1 - q/c_n) = \left(\frac{q}{c_n}\right)^{-\gamma_0} \exp\left(d_n \int_q^{c_n} \frac{h(s)}{s}\, ds\right), \qquad q > 0,$$

from minimax risk bounds for estimators of the linear functional $\int_q^{t_0} h(s)/s\, ds$ if h is assumed to have a compact support $\subset [0, t_0]$.

3. Minimax risk bounds in the white noise model. In this section we analyze the minimax risk in the limiting white noise experiment $(C[0,\infty), \mathscr{C}[0,\infty), (Q_h)_{h\in\mathscr{H}_\rho})$. While unfortunately that risk is not known exactly, one may establish suitable bounds, taking up the approach by Donoho and Liu (1991b).

To derive an *upper* bound on the minimax risk, we first look for affine minimax estimators in one-dimensional linear submodels. Because $\mathscr{H}_\rho$ is symmetric about 0, one may restrict oneself to families of the type $(Q_{\lambda h})_{\lambda\in[-1,1]}$, where $h \in \mathscr{H}_\rho$ with $h(0) > 0$.

Let

$$Y_h := \frac{\int h\, dx}{\|h\|},$$

where again $\int h\, dx$ denotes a stochastic integral. Hence

$$\mathscr{L}(Y_h \mid Q_{\lambda h}) = \mathscr{N}(\lambda\|h\|/\gamma_0, 1) = \mathscr{N}(\vartheta, 1), \tag{3.1}$$

where $\vartheta := \lambda\|h\|/\gamma_0$ is bounded in absolute value by $\tau_h := \|h\|/\gamma_0$. An estimator $\hat\vartheta$ for the bounded normal mean ϑ defines an estimator $\hat\lambda := \gamma_0\hat\vartheta/\|h\|$ and hence an estimator for the parameter of interest $\lambda h(0)$, too, such that $|\hat\lambda h(0) - \lambda h(0)| > a$ is equivalent to $|\hat\vartheta - \vartheta| > a\|h\|/(\gamma_0 h(0)) =: a_h$.

The minimax affine estimator for ϑ under the zero-one loss generated by $\mathbb{1}_{[-a_h,a_h]^c}$ equals $c_h Y_h$ with

$$c_h := \left(\frac{1}{2} + \left(\frac{1}{4} + \frac{1}{2a_h\tau_h}\log\frac{\tau_h + a_h}{\tau_h - a_h}\right)^{1/2}\right)^{-1}$$
$$= \left(\frac{1}{2} + \left(\frac{1}{4} + \frac{h(0)\gamma_0^2}{2a\|h\|^2}\log\frac{h(0)+a}{h(0)-a}\right)^{1/2}\right)^{-1}$$

if $h(0) > a$, and $c_h = 0$ if $h(0) \le a$; the corresponding minimax risk is equal to

$$R_{\text{aff}}(h) := \Phi\left(-\frac{\tau_h + a_h}{c_h} + \tau_h\right) + \Phi\left(\frac{\tau_h - a_h}{c_h} - \tau_h\right)$$

if $h(0) > a$, and $R_{\text{aff}}(h) = 0$ else [Drees (1999), Theorem 1]. Here Φ denotes the standard normal d.f. The arguments of Donoho and Liu (1991b) and Donoho (1994) show that the maximum of these minimax affine risks among all one-dimensional submodels equals the minimax risk among all affine estimators in the full model and thus is an upper bound for the minimax risk among all estimators for $h(0)$ in the full model. (An estimator in the full model is called affine if it can be represented as $d + \int h\,dx$ for some $d \in \mathbb{R}$ and $h \in L_2[0,\infty)$.) Furthermore, the minimax affine estimator in the most difficult one-dimensional submodel defines a minimax affine estimator in the full model. This leads to:

THEOREM 3.1. *Let ξ be the unique positive solution of the equation*

$$\frac{\xi^{1+1/\rho}}{\log\frac{\xi+a}{\xi-a}} = \frac{\gamma_0^2(\rho+1)}{a} \tag{3.2}$$

and define

$$h_\xi(t) := (\xi - t^\rho)\,\mathbb{1}_{[0,\xi^{1/\rho}]}(t). \tag{3.3}$$

Then

$$\widehat{h(0)} := \gamma_0(1+1/\rho)\xi^{-(1+1/\rho)}\int h_\xi\,dx \tag{3.4}$$

is minimax among all affine estimators for $h(0)$ in the full model $(Q_h)_{h\in\mathscr{H}_\rho}$. The corresponding minimax risk is

$$\begin{aligned}&\Phi\left(-\frac{\xi^{1+1/(2\rho)}}{\gamma_0(2(\rho+1)(2\rho+1))^{1/2}}\left(1+\frac{2\rho+1}{\xi}a\right)\right)\\&\quad+\Phi\left(\frac{\xi^{1+1/(2\rho)}}{\gamma_0(2(\rho+1)(2\rho+1))^{1/2}}\left(1-\frac{2\rho+1}{\xi}a\right)\right).\end{aligned} \tag{3.5}$$

Observe that the least favorable direction h must be of type (3.3) for some $\xi > 0$, since this way $h(0) = \xi$ is maximized among all $h \in \mathscr{H}_\rho$ with fixed norm $\|h\|$ (and thus fixed τ_h), so that a_h is minimized and the risk pertaining to $\hat{\vartheta}$ and hence also to $\hat{\lambda} h(0)$ is maximized.

Next we turn to the *lower* bound on the minimax risk, which is obtained as the supremum of all minimax risks in one-dimensional linear submodels. Notice that now the restriction to affine estimators has been dropped. Zeytinoglu and Mintz (1984) described the minimax estimator for ϑ in model (3.1) under zero-one loss. It turned out that the minimax risk depends on τ_h only through the smallest integer l that is greater than or equal to $\tau_h/a_h = h(0)/a$. The arguments given in the last paragraph show that it is sufficient to consider functions h of type (3.3) when looking for the supremum of these minimax risks among all one-dimensional submodels. Moreover, for fixed l, $a_{h_\xi} = a\|h_\xi\|/(\gamma_0 \xi)$ is minimized and thus the minimax risk maximized if $\xi \downarrow (l-1)a$, because $\|h_\xi\|/\xi$ is a strictly increasing function of ξ.

The corresponding minimax risk $R_{\mathrm{mm}}(h_\xi)$ can be derived from the solution of a system of nonlinear equations; for details refer to Zeytinoglu and Mintz (1984) or Drees (1999). To find the supremum of all minimax risks in one-dimensional submodels, one must compare $\lim_{\xi \downarrow (l-1)a} R_{\mathrm{mm}}(h_\xi)$ for all $l \geq 2$ numerically (see Section 4). Note that here the supremum is not attained, that is, there is no most difficult one-dimensional submodel. A detailed discussion of the relationship between the minimax risks in the full model and its one-dimensional submodels, respectively, can be found in the paper by Donoho and Liu (1991b).

REMARK. Since in the definition of $\tilde{\mathscr{H}}_\rho$ the constant $\delta > 0$ may be chosen arbitrarily small, it is easily seen that the minimax affine risk and the minimax risk are both the same in the models $(Q_h)_{h \in \mathscr{H}_\rho}$ and $(Q_h)_{h \in \tilde{\mathscr{H}}_\rho}$.

To sum up, we have found an upper and a lower bound for the minimax risk in the limiting white noise model. In the next section it is seen that these bounds carry over to the sequence of local models, and that they are almost identical for those values a which arise from the construction of confidence intervals.

4. Minimax risk bounds for the extreme value index. According to Corollary 2.1, the lower risk bound in the limiting white noise model obtained in the preceding section, namely, the supremum of minimax risks among all one-dimensional linear subfamilies, is a lower bound for the asymptotic minimax risk in the sequence of local models. In contrast to that, it is not obvious from the LeCam–theory that the upper bound on the minimax risk in the limiting model carries over in a similar manner. To prove that this holds true indeed, we construct a sequence of estimators for the local parameter of interest $h(0)$ that asymptotically attains the upper bound (3.5) on the minimax risk.

To this end, recall that the minimax affine estimator (3.4) in the limit experiment is a multiple of the stochastic integral of h_ξ defined in (3.3). In view of Theorem 2.1, approximation (2.2) and the Girsanov formula (5.10) (see the proof of Theorem 2.1 below), the estimator

$$\begin{aligned}\gamma_0^2(1+1/\rho)\xi^{-(1+1/\rho)}n^{-1/2}\sum_{i=1}^n g_{n,h_\xi}(X_i)\\ &= \xi^{-1/\rho}u(t_n)\frac{\rho+1}{\rho}\sum_{i=1}^n\Biggl(\log\frac{X_i}{(\xi^{1/\rho}t_n)^{-\gamma_0}}\\ &\qquad +\gamma_0\frac{\rho+1}{\rho}\Biggl(\Biggl(\frac{X_i}{(\xi^{1/\rho}t_n)^{-\gamma_0}}\Biggr)^{-\rho/\gamma_0}-1\Biggr)\Biggr)\\ &\qquad \times\mathbb{1}_{[1,\infty)}\Biggl(\frac{X_i}{(\xi^{1/\rho}t_n)^{-\gamma_0}}\Biggr),\end{aligned}$$

with $d_n = u(t_n)$ and $t_n = 1/c_n$ satisfying (1.2), is a natural candidate. [Here and in the sequel we use the canonical model, that is, $X_i := X_{i,n} : \mathbb{R}^n \to \mathbb{R}$, $(x_1,\dots,x_n)\mapsto x_i$; hence, under $P^n_{n,h}$ the random variables X_i are i.i.d. with d.f. $F_{n,h}$.]

Note that this estimator is based only on the exceedances over the threshold $(\xi^{1/\rho}t_n)^{-\gamma_0}$. As usual, instead one may use all exceedances over the random threshold $X_{n-k_n:n}$; that is, the (k_n+1)th largest order statistic, where k_n is chosen such that $F_0^{-1}(1-k_n/n) = (k_n/n)^{-\gamma_0} \approx (\xi^{1/\rho}t_n)^{-\gamma_0}$. By virtue of (1.2) and $\gamma(F_{n,h}) = \gamma_0 + u(t_n)h(0)$, this modification leads to the following estimator for the extreme value index in the local model:

$$\begin{aligned}\hat\gamma_n^{\text{loc}} &:= \gamma_0 + \frac{\rho+1}{\rho}\cdot\frac{1}{k_n}\sum_{i=1}^{k_n}\Biggl(\log\frac{X_{n-i+1:n}}{X_{n-k_n:n}}\\ &\qquad +\gamma_0\frac{\rho+1}{\rho}\Biggl(\Biggl(\frac{X_{n-i+1:n}}{X_{n-k_n:n}}\Biggr)^{-\rho/\gamma_0}-1\Biggr)\Biggr)\\ &= \hat\gamma_n^{\text{Hill}} + \frac{1}{\rho}\Bigl(\hat\gamma_n^{\text{Hill}}-\gamma_0\Bigr)\\ &\qquad +\gamma_0\Bigl(\frac{\rho+1}{\rho}\Bigr)^2\frac{1}{k_n}\sum_{i=1}^{k_n}\Biggl(\Biggl(\frac{X_{n-i+1:n}}{X_{n-k_n:n}}\Biggr)^{-\rho/\gamma_0}-\frac{1}{\rho+1}\Biggr),\end{aligned}\tag{4.1}$$

where $k_n := [\xi^{1/\rho}A^{-2/(2\rho+1)}n^{2\rho/(2\rho+1)}]$, ξ is defined by (3.2) and

$$\hat\gamma_n^{\text{Hill}} := \frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{X_{n-i+1:n}}{X_{n-k_n:n}}$$

is the popular Hill estimator. Notice that $\hat\gamma_n^{\text{loc}}$ can be interpreted as a modification of the Hill estimator, depending on A and a only through the number

$k_n + 1$ of order statistics on which it is based. Moreover, it can be represented as the functional

$$(4.2)\quad T_{\text{loc}}(z) := \gamma_0 + \frac{\rho+1}{\rho}\int_0^1 \log\frac{z(t)}{z(1)} + \gamma_0\frac{\rho+1}{\rho}\left(\left(\frac{z(t)}{z(1)}\right)^{-\rho/\gamma_0} - 1\right) dt$$

applied to the tail empirical quantile function

$$Q_{n,k_n}(t) := X_{n-[k_n t]:n}, \qquad 0 \le t \le 1,$$

that is, $\hat{\gamma}_n^{\text{loc}}$ is a statistical tail functional in the sense of Drees (1998a,b). Following the approach of those papers, it is not difficult to prove that indeed this estimator attains the asymptotic minimax affine risk bound:

THEOREM 4.1. *Let* $d_n = A^{1/(2\rho+1)} n^{-\rho/(2\rho+1)}$ *and* $c_n = A^{2/(2\rho+1)} n^{1/(2\rho+1)}$. *Then, for all* $a > 0$, *the maximal risk*

$$\sup_{h \in \mathscr{H}_\rho \cap \mathscr{H}^{(n)}} P_{n,h}^n \left\{ |\hat{\gamma}_n^{\text{loc}} - \gamma(F_{n,h})| > a d_n \right\}$$

converges to the minimax affine risk (3.5) *in the white noise model*, *which thus is an upper bound on the asymptotic minimax risk in the sequence of local models*.

REMARK. Check that the parameter space $\mathscr{H}_\rho$ enters the proof of Theorem 4.1 only via the bias term (5.18), and that for k_n it is merely essential that $k_n^{-1/2} \sim \xi^{-1/(2\rho)} d_n$. So if in case of a more general boundary function u we choose $k_n = [\xi^{1/\rho}(u(t_n))^{-2}]$, then we arrive at the same conclusion as in Theorem 4.1, because obviously the limit in (5.19) is not altered if $\mathscr{H}_\rho$ is replaced with $\tilde{\mathscr{H}}_\rho$.

So far we have shown that the bounds on the minimax risk in the limiting white noise model are also bounds on the asymptotic minimax risk in the sequence of local models. To obtain an analogous result for the global model $\mathscr{F}(\gamma_1, \varepsilon, u)$, in the next step we construct an estimator of the extreme value index in $\mathscr{F}(\gamma_1, \varepsilon, u)$ by replacing the parameter γ_0 of the center of localization by a suitable initial estimator. Although any estimator that converge to the true value with the optimal rate $n^{-\rho/(2\rho+1)}$ works, the calculations become particularly simple if one uses the Hill estimator in the initial estimation step. For the asymptotic behavior of this estimator has been studied thoroughly, and, in addition, the second term in the representation (4.1) vanishes. Thus, we define $\bar{k}_n := [n^{2\rho/(2\rho+1)}]$ and $\bar{\gamma}_n := \bar{k}_n^{-1}\sum_{i=1}^{\bar{k}_n} \log(X_{n-i+1:n}/X_{n-\bar{k}_n:n})$. Let $\bar{\xi}_n$ be the solution of (3.2) if γ_0 is replaced by $\bar{\gamma}_n$ and define $\gamma_n^* := \hat{k}_n^{-1}\sum_{i=1}^{\hat{k}_n} \log(X_{n-i+1:n}/X_{n-\hat{k}_n:n})$ with $\hat{k}_n := [\bar{\xi}_n A^{-2/(2\rho+1)} n^{2\rho/(2\rho+1)}]$. Then our adaptive estimator for γ in the model $\mathscr{F}(\gamma_1, \varepsilon, u)$ is

$$\hat{\gamma}_n := \gamma_n^* + \gamma_n^* \left(\frac{\rho+1}{\rho}\right)^2 \frac{1}{\hat{k}_n} \sum_{i=1}^{\hat{k}_n} \left(\left(\frac{X_{n-i+1:n}}{X_{n-\hat{k}_n:n}} \right)^{-\rho/\gamma_n^*} - \frac{1}{\rho+1} \right).$$

In the following theorem it is shown that the asymptotic maximal risk of $\hat{\gamma}_n$ equals the supremum of the asymptotic maximal risks pertaining to the estimators $\hat{\gamma}_n^{\text{loc}}$ among all local models under consideration. To determine this supremum, note that the solution $\xi = \xi(\gamma_0, a, \rho)$ of (3.2) satisfies

$$\xi\left(c\gamma_0, c^{2\rho/(2\rho+1)}a, \rho\right) = c^{2\rho/(2\rho+1)}\xi(\gamma_0, a, \rho) \tag{4.3}$$

for all $c > 0$. Consequently, the minimax affine risk (3.5) remains the same if (γ_0, a) is replaced with $(c\gamma_0, c^{2\rho/(2\rho+1)}a)$. Since obviously the risk is a decreasing function of a for fixed γ_0, it follows that (3.5) is an increasing function of γ_0 for fixed a, and thus the aforementioned supremum of maximal risk of $\hat{\gamma}_n^{\text{loc}}$ among all local models is obtained by replacing γ_0 with $\gamma_1 + \varepsilon$ in (3.5).

THEOREM 4.2. *Under the assumptions of Theorem* 4.1, $\hat{\gamma}_n$ *attains the asymptotic minimax affine risk*, *that is*,

$$\lim_{n\to\infty} \sup_{F\in\mathscr{F}(\gamma_1,\varepsilon,u)} P_F^n\{|\hat{\gamma}_n - \gamma(F)| > ad_n\}$$
$$= \Phi\left(-\frac{\xi^{1+1/(2\rho)}}{(\gamma_1+\varepsilon)(2(\rho+1)(2\rho+1))^{1/2}}\left(1 + \frac{2\rho+1}{\xi}a\right)\right)$$
$$+\Phi\left(\frac{\xi^{1+1/(2\rho)}}{(\gamma_1+\varepsilon)(2(\rho+1)(2\rho+1))^{1/2}}\left(1 - \frac{2\rho+1}{\xi}a\right)\right),$$

where $\xi = \xi(\gamma_1 + \varepsilon, a, \rho)$ *is the solution of* (3.2) *with* γ_0 *replaced by* $\gamma_1 + \varepsilon$.

It should be emphasized that the estimator $\hat{\gamma}_n$ in the global model is adaptive only in the sense that the parameter γ_0 of the center of localization used in the construction of $\hat{\gamma}_n^{\text{loc}}$ is estimated from the data. In particular, $\hat{\gamma}_n$ still depends on the choice of A and ρ, which determine the global model under consideration. However, because A and ρ are parameters of the model and *not* of the underlying d.f. F, in general, it does not make sense to "estimate" them. On the other hand, one may assume that F itself is of the form

$$F^{-1}(1-t) = ct^{-\gamma}\exp\left(\int_t^1 \frac{\eta(s)}{s}\,ds\right) \quad \text{with} \quad \eta(s) = As^\rho(1+o(1)) \tag{4.4}$$

(and not merely $|\eta(s)| \le As^\rho$ as supposed up to now). In this situation, Drees and Kaufmann [(1998), Theorem 1] and Danielsson et al. [(1998), (3.9)] proposed consistent estimators $\hat{\rho}_n$ of ρ and a data–driven choice $\hat{k}_n^{opt}$ of the number of order statistics that minimizes the asymptotic mean squared error of the Hill estimator. Since the number k_n used in (4.1) is related to $\hat{k}_n^{opt}$ via

$$k_n = \Big(\xi(\bar{\gamma}_n, a, \hat{\rho}_n)\Big)^{1/\hat{\rho}_n} \left(\frac{2\hat{\rho}_n}{(\hat{\rho}_n+1)^2\bar{\gamma}_n^2}\right)^{1/(2\hat{\rho}_n+1)} \hat{k}_n^{opt}(1+o_P(1)),$$

one may construct a consistent estimator $\hat{k}_n$ for k_n (in the sense that $\hat{k}_n/k_n \to 1$ in probability) based on $\hat{\rho}_n$ and $\hat{k}_n^{opt}$. Then the estimator $\hat{\gamma}_n$ depending on the unknown values of k_n and ρ and the estimator where these are replaced with their respective estimators have the same limit distribution under F [although the latter need not be adaptive in the sense used, e.g., by Lepskii (1991)]. In fact, in (4.4) one may even replace $As^\rho(1+o(1))$ by a more general ρ–varying function; see Drees and Kaufmann [(1998), Theorem 3] for details.

In Figure 1, the solid lines represent the asymptotic maximal risk (3.5) of $\hat{\gamma}_n^{\text{loc}}$, regarded as a function of the constant a defining the loss function, for $\rho = 0.1$ (left) and $\rho = 1$ (right), where $\gamma_0 = 1$ in both cases. The broken line is the lower bound for the asymptotic minimax risk, which was computed numerically as $\sup_{l\geq 2} \lim_{\xi\downarrow(l-1)a} R_{\text{mm}}(h_\xi)$ (see Section 3).

Notice that for the calculation of the lower bound it is crucial not to use the one-dimensional submodel that is most difficult for affine estimators. For example, for $\rho = 1$ and $a = 0.82$, the direction h_ξ is least favorable for the minimax affine estimator if $\xi = \xi_{\text{aff}} \approx 1.638$ and for the general minimax estimator if $\xi \downarrow 1.64$; despite this small difference, the minimax risk $\lim_{\xi\downarrow 1.64} R_{\text{mm}}(h_\xi)$ is about 12% greater than $R_{\text{mm}}(h_{\xi_{\text{aff}}})$. Hence, the use of $h_{\xi_{\text{aff}}}$ would yield a substantially less accurate lower bound. In fact, the minimax risk $R_{\text{mm}}(h_\xi)$ is a discontinuous function of ξ as the minimax risk in model (3.1) is a discontinuous function of τ_h. Thus, even minor changes of ξ may lead to large changes of the risk. On the other hand, for some a, the minimax risks $R_{\text{mm}}(h_{\xi_{\text{aff}}})$ and

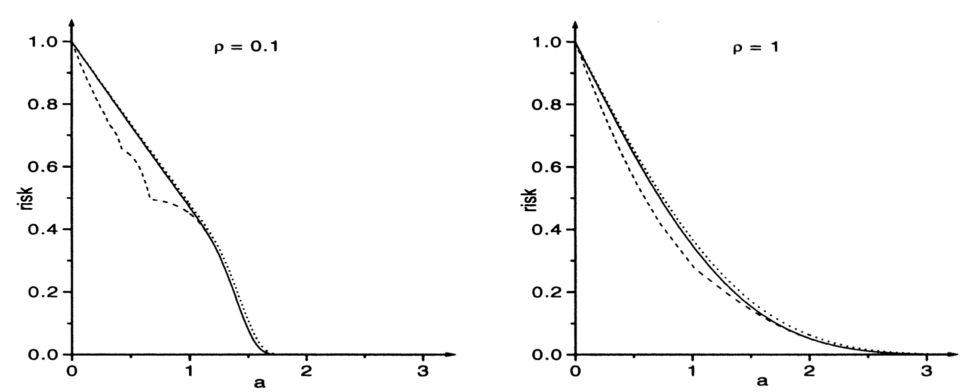

FIG. 1. *Asymptotic maximal risk of* $\hat{\gamma}_n^{\text{loc}}$ (*solid line*), *of* $\hat{\gamma}_n^{\text{Hill}}$ (*dotted line*) *and asymptotic lower risk bound* (*dashed line*)

$\sup_{l\in\mathbb{N}} \lim_{\xi\downarrow(l-1)a} R_{\mathrm{mm}}(h_\xi) =: \lim_{\xi\downarrow(l^*-1)a} R_{\mathrm{mm}}(h_\xi)$ are almost equal, although $(l^*-1)a$ is significantly larger than $h_{\xi_{\mathrm{aff}}}$; for instance, $\rho=1$ and $a=0.01$ lead to $(l^*-1)a = 2.62$ and $\xi_{\mathrm{aff}} \approx 1.587$, yet the relative deviation of the pertaining risks is less than 0.1%.

The plots clearly demonstrate that the difference between the asymptotic lower risk bound and the asymptotic risk of the "minimax affine" estimator $\hat\gamma_n^{\mathrm{loc}}$ is moderate; particularly for large values of a, which are decisive for the construction of confidence intervals (see below), it is negligible. Hence $\hat\gamma_n^{\mathrm{loc}}$ is almost optimal.

Since $\hat\gamma_n^{\mathrm{loc}}$ is a modification of $\hat\gamma_n^{\mathrm{Hill}}$, it is interesting to compare its risk with that of the much simpler Hill estimator. Taking up the approach by Csörgő, Deheuvels and Mason (1985) (or imitating the proof of Theorem 4.1), one can prove that for $k_n = [\lambda n^{2\rho/(2\rho+1)}]$

$$\begin{aligned}
&\sup_{h\in\mathscr{H}_\rho\cap\mathscr{H}^{(n)}} P_{n,h}^n\{|\hat\gamma_n^{\mathrm{Hill}} - \gamma(F_{n,h})| > a d_n\} \\
&\quad\longrightarrow \mathscr{N}\left(\lambda^\rho A^{2\rho/(2\rho+1)}/(\rho+1), \lambda^{-1}\gamma_0^2 A^{-2/(2\rho+1)}\right)[-a,a]^c \\
&\quad= \Phi\left(-\frac{1}{\gamma_0}\left(\lambda^{\rho+1/2}\frac{A}{\rho+1} + \lambda^{1/2}A^{1/(2\rho+1)}a\right)\right) \\
&\qquad+\Phi\left(\frac{1}{\gamma_0}\left(\lambda^{\rho+1/2}\frac{A}{\rho+1} - \lambda^{1/2}A^{1/(2\rho+1)}a\right)\right)
\end{aligned}$$

[cf. Drees (1998c), proof of Theorem 2.2]. Straighforward calculations show that an asymptotically optimal number of order statistics is given by

$$k_n^{\mathrm{Hill}} := \left[\left(\frac{\rho+1}{2\rho+1}\xi_{\mathrm{Hill}}\right)^{1/\rho} A^{-2/(2\rho+1)} n^{2\rho/(2\rho+1)}\right],$$

leading to the maximal risk

$$\begin{aligned}
&\Phi\left(-\frac{(\rho+1)^{1/(2\rho)}}{\gamma_0(2\rho+1)^{1+1/(2\rho)}}\xi_{\mathrm{Hill}}^{1+1/(2\rho)}\left(1+\frac{2\rho+1}{\xi_{\mathrm{Hill}}}a\right)\right) \\
&\quad+\Phi\left(\frac{(\rho+1)^{1/(2\rho)}}{\gamma_0(2\rho+1)^{1+1/(2\rho)}}\xi_{\mathrm{Hill}}^{1+1/(2\rho)}\left(1-\frac{2\rho+1}{\xi_{\mathrm{Hill}}}a\right)\right)
\end{aligned} \tag{4.5}$$

where ξ_{Hill} is the unique solution of

$$\frac{\xi_{\mathrm{Hill}}^{1+1/\rho}}{\log\frac{\xi_{\mathrm{Hill}}+a}{\xi_{\mathrm{Hill}}-a}} = \frac{\gamma_0^2(2\rho+1)^{1+1/\rho}}{2a(\rho+1)^{1/\rho}}. \tag{4.6}$$

Observe that the structure of these formulas is similar to that of (3.2) and (3.5).

TABLE 1
Length of α–confidence intervals based on $\hat{\gamma}_n^{\mathrm{loc}}$ and $\hat{\gamma}_n^{\mathrm{Hill}}$ divided by lower bound for confidence interval length

ρ	α	**0.9**	**0.95**	**0.99**	**0.995**
0.1	$\hat{\gamma}_n^{\mathrm{loc}}$	1.00	1.00	1.00	1.00
	$\hat{\gamma}_n^{\mathrm{Hill}}$	1.02	1.02	1.02	1.02
1	$\hat{\gamma}_n^{\mathrm{loc}}$	1.01	1.00	1.00	1.00
	$\hat{\gamma}_n^{\mathrm{Hill}}$	1.05	1.04	1.04	1.04

The graphs of (4.5) are represented by the dotted lines in Figure 1. By and large, the difference between the asymptotic risk of $\hat{\gamma}_n^{\mathrm{loc}}$ and $\hat{\gamma}_n^{\mathrm{Hill}}$ is even smaller than the distance between the risk for $\hat{\gamma}_n^{\mathrm{loc}}$ and the lower bound, yet now the difference is more distinct in the upper tail. Notice that the behavior of the risk for large a is particularly important for the construction of asymptotic confidence intervals with fixed (deterministic) length that are symmetric about the estimator under consideration if, as usual, the confidence coefficient α is close to 1. For the half-length of such a confidence interval equals the value a where the graph of the pertaining risk intersects the line "risk $= 1 - \alpha$."

In Table 1, the ratio of the length of these confidence intervals to the lower bound on the length of arbitrary fixed-size confidence intervals, which is obtained from the lower risk bound plotted in Figure 1, is given for $\alpha = 0.9$, 0.95, 0.99 and 0.995 and $\rho = 0.1$ and 1. Note that these ratios can be regarded as measures of efficiency for the estimators $\hat{\gamma}_n^{\mathrm{loc}}$ and $\hat{\gamma}_n^{\mathrm{Hill}}$. The confidence intervals based on $\hat{\gamma}_n^{\mathrm{loc}}$ have almost minimal length in all cases. In contrast, if one uses the Hill estimator, then the confidence intervals are a few percent longer. Nevertheless, in practice the slight loss of efficiency hardly justifies the use of the substantially more complicated adaptive minimax estimator.

Finally, it is worth mentioning that one obtains a similar picture for different parameters γ_0 of the center of localization. For it is easily seen from (4.3) and the subsequent discussion, in combination with analogous arguments for the lower risk bound and the risk of the Hill estimator, that the plots for general $\gamma_0 > 0$ can be obtained from Figure 1 by stretching the a-axis by the factor $\gamma_0^{2\rho/(2\rho+1)}$. Consequently, the change of the center of localization does not affect the ratios given in Table 1, which hence also equal the corresponding ratios of the length of fixed-size confidence intervals in the global model $\mathscr{F}(\gamma_1, \varepsilon, u)$.

5. Proofs.

PROOF OF PROPOSITION 2.2. For the sake of notational simplicity, the index h, which indicates the dependence of $g_{n,h}$ on h, is omitted. Throughout the proof, *const*. denotes a generic constant which depends on h but not on n, s or t and which may vary from line to line.

First note that g_n is well defined since for $h \in L_2[0,\infty)$ and $t > 0$ the Cauchy-Schwarz inequality yields

$$\int_t^{c_n} \frac{|h(s)|}{s}\,ds \le \left(\int_t^\infty h^2(s)\,ds \int_t^\infty s^{-2}\,ds\right)^{1/2} \le \|h\| t^{-1/2}. \tag{5.1}$$

Furthermore, the boundedness of $h \in \mathscr{H}$ on compact intervals, which is immediate from the definition of $\mathscr{H}$, implies

$$\int_t^\infty \frac{|h(s)|}{s}\,ds \le \|h\| + \sup_{0<s\le 1} |h(s)|\,|\log t| < \infty \qquad \text{for } t \in (0,1]. \tag{5.2}$$

By the definition of $F_{n,h}^{-1}$, one has

$$\begin{aligned} f_{n,h}(F_{n,h}^{-1}(1-t)) &= -\frac{1}{d/dt(F_{n,h}^{-1}(1-t))} \\ &= \frac{t}{(\gamma_0 + d_n h(c_n t))F_{n,h}^{-1}(1-t)}, \qquad t \in (0,1], \end{aligned}$$

and $f_0(x) = x^{-(1/\gamma_0+1)}/\gamma_0$ for $x \ge 1$. Hence the change of variables $x = F_{n,h}^{-1}(1-t/c_n)$ leads to

$$\begin{aligned} &\int_1^\infty \left[n^{1/2}\left(f_{n,h}^{1/2}(x) - f_0^{1/2}(x)\right) - \frac{1}{2} g_n(x) f_0^{1/2}(x)\right]^2 dx \\ &\quad = \frac{n}{c_n}\int_0^{c_n} \left[\left(\frac{f_{n,h}(F_{n,h}^{-1}(1-t/c_n))}{f_0(F_{n,h}^{-1}(1-t/c_n))}\right)^{1/2} - 1 \right. \\ &\qquad\qquad \left. - \frac{n^{-1/2}}{2} g_n(F_{n,h}^{-1}(1-t/c_n))\right]^2 \frac{f_0(F_{n,h}^{-1}(1-t/c_n))}{f_{n,h}(F_{n,h}^{-1}(1-t/c_n))}\,dt \\ &\quad = \frac{n d_n^2}{c_n}\int_0^{c_n} \left[d_n^{-1}\left(\left(1 + d_n\frac{h(t)}{\gamma_0}\right)^{-1/2} \exp\left(\frac{d_n}{2\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right) - 1\right)\right. \\ &\qquad\qquad \left. - \frac{n^{-1/2}}{2 d_n} g_n(F_{n,h}^{-1}(1-t/c_n))\right]^2 \\ &\qquad\qquad \times \left(1 + d_n\frac{h(t)}{\gamma_0}\right)\exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)dt. \end{aligned} \tag{5.3}$$

To prove that this expression tends to 0, first we consider

$$\int_0^{c_n}\Bigg[d_n^{-1}\left(\left(1+d_n\frac{h(t)}{\gamma_0}\right)^{-1/2}\exp\left(\frac{d_n}{2\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)-1\right)$$
$$-\frac{n^{-1/2}}{2d_n}g_n((t/c_n)^{-\gamma_0})\Bigg]^2$$
$$\times\left(1+d_n\frac{h(t)}{\gamma_0}\right)\exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)dt$$

(5.4)

$$\begin{aligned}
&\le 3\Bigg(\int_0^{c_n}\left[d_n^{-1}\left(\exp\left(\frac{d_n}{2\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)-1\right)-\frac{1}{2\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right]^2\\
&\qquad\times\exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)dt\\
&\quad+\int_0^{c_n}\left[d_n^{-1}\left(1-\left(1+d_n\frac{h(t)}{\gamma_0}\right)^{1/2}\right)+\frac{1}{2\gamma_0}h(t)\right]^2\\
&\qquad\times\exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)dt\\
&\quad+\frac{1}{4\gamma_0^2}\int_0^{c_n}\left(\int_t^{c_n}\frac{h(s)}{s}\,ds-h(t)\right)^2\left(1-\left(1+d_n\frac{h(t)}{\gamma_0}\right)^{1/2}\right)^2\\
&\qquad\times\exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n}\frac{h(s)}{s}\,ds\right)dt\Bigg)\\
&=: 3\left(I_1+I_2+\frac{1}{4\gamma_0^2}I_3\right).
\end{aligned}$$

By (5.1), (5.2) and a Taylor expansion of the exponential function, one can show that, for large n, the integrand of I_1 is bounded in absolute value by

$$\begin{aligned}
&\left[\frac{d_n}{8\gamma_0^2}\left(\int_t^{c_n}\frac{h(s)}{s}\,ds\right)^2\right]^2\exp\left(\frac{d_n}{\gamma_0}\int_t^{c_n}\frac{|h(s)|}{s}\,ds\right)\\
&\quad\le\frac{d_n^2}{64\gamma_0^4}\left[\|h\|^4t^{-2}\,\mathbb{1}_{[1,\infty)}(t)+\left(\|h\|+\sup_{0<s\le 1}|h(s)||\log t|\right)^4\mathbb{1}_{(0,1]}(t)\right]\\
&\qquad\times\exp\left(1+\frac{|\log t|}{2}\,\mathbb{1}_{(0,1]}(t)\right).
\end{aligned}$$

Hence,

$$I_1 \le d_n^2 \left(\text{const.} \int_1^\infty t^{-2}\, dt + \text{const.} \int_0^1 (1+|\log t|)^4 t^{-1/2}\, dt \right) \longrightarrow 0.$$

Using the fact that, according to the definition of $\mathscr{H}$, $1 + d_n h(t)/\gamma_0$ is uniformly bounded and bounded away from 0 on $(0, c_n]$, one can prove in a similar way that the integrand of I_2 converges to 0 and that it is bounded by an integrable function of the form const.$h^2(t)\,\mathbb{1}_{[1,\infty)}(t)$ +const.$t^{-1/2}\,\mathbb{1}_{(0,1)}(t)$. Hence, by the dominated convergence theorem, I_2 vanishes asymptotically, too.

Because of (5.1), (5.2) and the definition of $\mathscr{H}$, on $(0, 1]$ the integrand of I_3 is eventually bounded by const.$(1+|\log t|)^2 t^{-1/2}$. On the other hand, using the definition of $\mathscr{H}$ and a Taylor expansion, it is readily seen that $(1-(1+d_n h(t)/\gamma_0)^{1/2})^2 \le \min(\text{const.}, \text{const.} h^2(t))$ for $t \in (0, c_n]$, so on account of (5.1) and (5.2), the integrand is bounded by the integrable function

$$\text{const.} \left(\|h\|^2 t^{-1} + h^2(t) \right) \min(1, h^2(t)) \le \text{const.} \left(t^{-1} h^2(t) + h^2(t) \right)$$

on $[1, c_n]$. Since obviously the integrand tends to 0 for every $t > 0$, it follows by the dominated convergence theorem that $I_3 \to 0$.

In view of (2.1), (5.3) and (5.4), it remains to verify that

$$\int_0^{c_n} \left[\frac{n^{-1/2}}{2d_n} g_n\left(F_{n,h}^{-1}(1-t/c_n)\right) - \frac{n^{-1/2}}{2d_n} g_n\left((t/c_n)^{-\gamma_0}\right) \right]^2 \times \left(1 + d_n \frac{h(t)}{\gamma_0}\right) \exp\left(-\frac{d_n}{\gamma_0} \int_t^{c_n} \frac{h(s)}{s}\, ds\right) dt$$

vanishes asymptotically. Because the factor $1+d_n h(t)/\gamma_0$ is uniformly bounded on $(0, c_n]$, it suffices to prove that

$$\begin{aligned} &\int_0^{c_n} \left(\int_{c_n(F_{n,h}^{-1}(1-t/c_n))^{-1/\gamma_0}}^{t} \frac{h(s)}{s}\, ds \right)^2 \exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n} \frac{h(s)}{s}\, ds\right) dt \\ &\quad + \int_0^{c_n} \left(h\left(c_n \left(F_{n,h}^{-1}(1-t/c_n)\right)^{-1/\gamma_0}\right) - h(t) \right)^2 \\ &\qquad \times \exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n} \frac{h(s)}{s}\, ds\right) dt \\ &\quad =: I_4 + I_5 \longrightarrow 0. \end{aligned} \tag{5.5}$$

For this end, check that $c_n(F_{n,h}^{-1}(1-t/c_n))^{-1/\gamma_0} = t\exp(-d_n/\gamma_0 \int_t^{c_n} h(s)/s\, ds)$ implies

$$I_4 \le \sup_{0<s\le 2} |h(s)| \int_0^1 \left(\frac{d_n}{\gamma_0}\int_t^{c_n} \frac{h(s)}{s}\, ds\right)^2 \exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n} \frac{h(s)}{s}\, ds\right) dt$$

$$+ \|h\|^2 \int_1^{c_n} \left| t^{-1} - c_n^{-1}(F_{n,h}^{-1}(1-t/c_n))^{1/\gamma_0} \right| \exp\left(-\frac{d_n}{\gamma_0}\int_t^{c_n} \frac{h(s)}{s}\, ds\right) dt$$

$$\leq o(1) + \text{const.} \int_1^{c_n} t^{-1} \left| 1 - \exp\left(\frac{d_n}{\gamma_0} \int_t^{c_n} \frac{h(s)}{s}\, ds \right) \right| dt$$

$$\leq o(1) + \text{const.}\, d_n \int_1^{c_n} t^{-1} \int_t^{c_n} \frac{|h(s)|}{s}\, ds\, dt.$$

Hence integration by parts shows that $I_4 \to 0$:

$$\int_1^{c_n} t^{-1} \int_t^{c_n} \frac{|h(s)|}{s}\, ds\, dt = \int_1^{c_n} \log t \frac{|h(t)|}{t}\, dt \leq \|h\| \left(\int_1^{\infty} t^{-2} \log^2 t\, dt \right)^{1/2} < \infty.$$

Now we turn to the last integral. From $c_n(F_{n,h}^{-1}(1-t/c_n))^{-1/\gamma_0} \to t$ and (5.2) we conclude that for $m \in (0,1]$ and sufficiently large n,

$$\int_0^m \left(h\left(c_n \left(F_{n,h}^{-1}(1-t/c_n) \right)^{-1/\gamma_0} \right) - h(t) \right)^2 \exp\left(-\frac{d_n}{\gamma_0} \int_t^{c_n} \frac{h(s)}{s}\, ds \right) dt$$
$$\leq 8 \sup_{t \leq 2} h^2(t) \int_0^m t^{-1/2}\, dt,$$

which can be made arbitrarily small by choosing m small. Moreover,

$$\frac{d}{dt} c_n \left(F_{n,h}^{-1} (1-t/c_n) \right)^{-1/\gamma_0} = \left(1 + d_n \frac{h(t)}{\gamma_0} \right) \exp\left(-\frac{d_n}{\gamma_0} \int_t^{c_n} \frac{h(s)}{s}\, ds \right) \tag{5.6}$$

is uniformly bounded away from 0 on $[m, c_n]$, so that eventually

$$\int_M^{c_n} \left(h\left(c_n \left(F_{n,h}^{-1} (1-t/c_n) \right)^{-1/\gamma_0} \right) - h(t) \right)^2 \exp\left(-\frac{d_n}{\gamma_0} \int_t^{c_n} \frac{h(s)}{s}\, ds \right) dt$$
$$\leq 4 \int_M^{c_n} h^2 \left(c_n \left(F_{n,h}^{-1}(1-t/c_n) \right)^{-1/\gamma_0} \right) + h^2(t)\, dt$$
$$\leq \text{const.} \int_{c_n(F_{n,h}^{-1}(1-M/c_n))^{-1/\gamma_0}}^{c_n} h^2(s)\, ds + 4 \int_M^{c_n} h^2(t)\, dt$$
$$\leq \text{const.} \int_{M/2}^{c_n} h^2(s)\, ds + 4 \int_M^{c_n} h^2(t)\, dt$$

which is arbitrarily small for large M because $h \in L_2[0,\infty)$.

For every $\delta > 0$, one may approximate h on $[m/2, 2M]$ by a Lebesgue-a.e. continuous function $\tilde{h}$ (e.g., a step function), such that

$$\int_{m/2}^{2M} \left(\tilde{h}(t) - h(t) \right)^2 dt \leq \delta \quad \text{and} \quad \sup_{m/2 \leq t \leq 2M} |\tilde{h}(t)| \leq \sup_{m/2 \leq t \leq 2M} |h(t)| + 1.$$

Hence, using the boundedness of $\exp(-d_n/\gamma_0 \int_t^{c_n} h(s)/s\, ds)$ on $[m, M]$ and the boundedness of the derivative (5.6) away from 0, we obtain for sufficiently large n,

$$\int_m^M \left(h\left(c_n \left(F_{n,h}^{-1}(1-t/c_n) \right)^{-1/\gamma_0} \right) - h(t) \right)^2 \exp\left(-\frac{d_n}{\gamma_0} \int_t^{c_n} \frac{h(s)}{s}\, ds \right) dt$$

$$
\begin{aligned}
&\le 4\Bigg(\int_m^M \Big(h\Big(c_n\left(F_{n,h}^{-1}(1-t/c_n)\right)^{-1/\gamma_0}\Big) \\
&\qquad\qquad -\tilde h\Big(c_n\left(F_{n,h}^{-1}(1-t/c_n)\right)^{-1/\gamma_0}\Big)\Big)^2\,dt \\
&\qquad + \int_m^M \Big(\tilde h\Big(c_n\left(F_{n,h}^{-1}(1-t/c_n)\right)^{-1/\gamma_0}\Big) - \tilde h(t)\Big)^2\,dt \\
&\qquad\qquad + \int_m^M \Big(\tilde h(t) - h(t)\Big)^2\,dt\Bigg) \\
&\le \text{const.} \int_{c_n(F_{n,h}^{-1}(1-m/c_n))^{-1/\gamma_0}}^{c_n(F_{n,h}^{-1}(1-M/c_n))^{-1/\gamma_0}} \Big(h(t)-\tilde h(t)\Big)^2\,dt \\
&\quad + 4\int_m^M \Big(\tilde h\Big(c_n\left(F_{n,h}^{-1}(1-t/c_n)\right)^{-1/\gamma_0}\Big) - \tilde h(t)\Big)^2\,dt + \delta \\
&\le \text{const.}\,\delta + o(1) + \delta
\end{aligned}
$$

where the dominated convergence theorem has been used in the last step.

Summing up, we have shown that I_5 becomes arbitrarily small as $n \to \infty$. Thus (5.5) is verified and so is the first assertion.

Applying first the change of variables $x = (t/c_n)^{-\gamma_0}$ and then Fubini's theorem, one gets

$$
\begin{aligned}
\int_1^\infty g_n(x) f_0(x)\,dx &= \frac{n^{1/2} d_n}{c_n \gamma_0} \int_0^{c_n} \int_{t/c_n}^1 \frac{h(c_n s)}{s}\,ds - h(t)\,dt \\
&= \frac{n^{1/2} d_n}{c_n \gamma_0} \left(\int_0^{c_n} \int_0^s dt\, \frac{h(s)}{s}\,ds - \int_0^{c_n} h(t)\,dt \right) = 0.
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
&\int_1^\infty g_n^2(x) f_0(x)\,dx \\
&\quad = \frac{n d_n^2}{c_n \gamma_0^2} \int_0^{c_n} \left(\int_t^{c_n} \frac{h(s)}{s}\,ds - h(t) \right)^2 dt \\
&\quad = \frac{1+o(1)}{\gamma_0^2} \Bigg(\int_0^{c_n} h^2(t)\,dt - 2\int_0^{c_n} h(t) \int_t^{c_n} \frac{h(s)}{s}\,ds\,dt \\
&\qquad\qquad + \int_0^{c_n}\int_0^{c_n} \min(s,r) \frac{h(s)}{s}\,ds\, \frac{h(r)}{r}\,dr \Bigg) \qquad (5.7) \\
&\quad = \frac{1+o(1)}{\gamma_0^2} \int_0^{c_n} h^2(t)\,dt \longrightarrow \frac{\|h\|^2}{\gamma_0^2}.
\end{aligned}
$$

Next we check that the Lindeberg condition

$$\begin{aligned}\int_1^\infty g_n^2(x)\mathbb{1}_{\{|g_n(x)|\geq \delta n^{1/2}\}}f_0(x)\,dx \\ = \frac{nd_n^2}{c_n\gamma_0^2}\int_0^{c_n}\left(\int_t^{c_n}\frac{h(s)}{s}\,ds - h(t)\right)^2 \\ \times \mathbb{1}_{\{|\int_t^{c_n} h(s)/s\,ds - h(t)|\geq \delta\gamma_0/d_n\}}\,dt \longrightarrow 0\end{aligned} \tag{5.8}$$

is satisfied for all $\delta > 0$. Since $|\int_t^{c_n} h(s)/s\,ds - h(t)|$ is bounded on (0,1] by the square integrable function $\|h\|^2 + \sup_{0<t\leq 1}|h(t)|(|\log t| + 1)$, one has

$$\int_0^1\left(\int_t^{c_n}\frac{h(s)}{s}\,ds - h(t)\right)^2\mathbb{1}_{\{|\int_t^{c_n} h(s)/s\,ds - h(t)|\geq \delta\gamma_0/d_n\}}\,dt \longrightarrow 0.$$

From (5.1) and $d_n \to \infty$ it follows that eventually $\{|\int_t^{c_n} h(s)/s\,ds - h(t)| \geq \delta\gamma_0/d_n\}\cap[1,\infty) \subset \{|h(t)| \geq \delta\gamma_0/(2d_n)\}\cap[1,\infty)$. Consequently,

$$\begin{aligned}&\int_1^{c_n}\left(\int_t^{c_n}\frac{h(s)}{s}\,ds - h(t)\right)^2\mathbb{1}_{\{|\int_t^{c_n} h(s)/s\,ds - h(t)|\geq \delta\gamma_0/d_n\}}dt \\ &\leq 2\int_1^{c_n}\left(\int_t^{c_n}\frac{h(s)}{s}\,ds\right)^2\mathbb{1}_{\{|h(t)|\geq \delta\gamma_0/(2d_n)\}}\,dt \\ &\quad +2\int_1^{c_n} h^2(t)\mathbb{1}_{\{|h(t)|\geq \delta\gamma_0/(2d_n)\}}\,dt \\ &\leq 2\left(\int_1^{c_n}\left(\int_t^{c_n}\frac{h(s)}{s}\,ds\right)^4 dt\int_1^{c_n}\mathbb{1}_{\{|h(t)|\geq \delta\gamma_0/(2d_n)\}}\,dt\right)^{1/2} \\ &\quad +o(1) \\ &\leq 2\|h\|^2\left(\int_1^{c_n} s^{-2}\,ds\right)^{1/2}\|h\|\frac{2d_n}{\delta\gamma_0} + o(1) \longrightarrow 0,\end{aligned}$$

where in the last two steps we have used the Cauchy-Schwarz and the Chebychev inequality. Therefore, the conditions of Proposition A.8 of van der Vaart (1988) are satisfied, which gives the approximation (2.2). □

PROOF OF THEOREM 2.1. Check that, for $h_1, \ldots, h_m \in \mathscr{H}$, analogously to (5.7) and (5.8),

$$\int_1^\infty g_{n,h_j}(x)g_{n,h_k}(x)f_0(x)\,dx \longrightarrow \frac{\langle h_j, h_k\rangle}{\gamma_0^2}$$

and the multivariate Lindeberg condition

$$\int_1^\infty \sum_{j=1}^m g_{n,h_j}^2(x) \mathbb{1}_{\{\sum_{j=1}^m g_{n,h_j}^2(x) \ge \delta^2 n\}} f_0(x)\,dx$$

$$\le \sum_{j,k=1}^m \int_1^\infty g_{n,h_j}^2(x) \mathbb{1}_{\{|g_{n,h_k}(x)| \ge \delta n^{1/2}/m\}} f_0(x)\,dx \longrightarrow 0$$

hold true for all $\delta > 0$. Hence the multivariate central limit theorem yields

$$\mathscr{L}\left(\left(n^{-1/2}\sum_{i=1}^n g_{n,h_j}(pr_i)\right)_{1\le j\le m}\middle| P_0^n\right) \longrightarrow \mathscr{N}\left(0, \left(\frac{\langle h_j, h_k\rangle}{\gamma_0^2}\right)_{1\le j,k\le m}\right) \tag{5.9}$$

where pr_i denotes the projection on the ith coordinate. According to the famous Girsanov formula, one has

$$\log \frac{dQ_h}{dQ_0}(x) = \frac{1}{\gamma_0}\int h\,dx - \frac{\|h\|^2}{2\gamma_0^2}, \tag{5.10}$$

where under Q_0 the stochastic integral $(\int h\,dx)_{h\in\mathscr{H}}$ is a centered Gaussian process with covariance function $(\langle h, \bar{h}\rangle)_{h,\bar{h}\in\mathscr{H}}$. A combination of (2.2), (5.9) and (5.10) proves the assertion. □

PROOF OF COROLLARY 2.1. Apply Strasser [(1985), Corollary 62.6 and Theorem 43.5] to the convergence established in Theorem 2.1. □

PROOF OF COROLLARY 2.2. The assertion is an immediate consequence of Corollary 2.1, since $\mathscr{H}_\rho \subset \mathscr{H}$ and, for all $\gamma_0 \in (\gamma_1 - \varepsilon, \gamma_1 + \varepsilon)$ and $h \in \mathscr{H}_\rho$, eventually $F_{n,h,\gamma_0} \in \mathscr{F}(\gamma_1, \varepsilon, u)$. □

PROOF OF THEOREM 3.1. Let φ denote the standard normal density.

First recall that it suffices to consider functions h_ξ as defined in (3.3) such that $\tau_{h_\xi} > a_{h_\xi}$ (see discussion below Theorem 3.1). Using

$$\|h_\xi\|^2 = \frac{2\rho^2}{(\rho+1)(2\rho+1)}\xi^{2+1/\rho} \tag{5.11}$$

and

$$(\tau_h + a_h)\varphi\Big(\frac{\tau_h + a_h}{c_h} - \tau_h\Big) = (\tau_h - a_h)\varphi\Big(\frac{\tau_h - a_h}{c_h} - \tau_h\Big)$$

[Drees (1999), proof of Theorem 1], one obtains by straightforward calculations that

$$\frac{\partial}{\partial\xi}R_{\text{aff}}(h_\xi) = \frac{2\tau_{h_\xi}a_{h_\xi}}{\xi(\tau_{h_\xi} - a_{h_\xi})}\left(\frac{1}{c_{h_\xi}} - \frac{2\rho+1}{2\rho}\right)\varphi\left(\frac{\tau_{h_\xi} + a_{h_\xi}}{c_{h_\xi}} - \tau_{h_\xi}\right).$$

Consequently, the maximum is attained if $c_{h_\xi} = 2\rho/(2\rho+1)$, which in turn is equivalent to (3.2). This equation has a unique solution as its left-hand side is a strictly increasing function of ξ which converges to 0 as ξ tends to a from above and to infinity as ξ tends to infinity.

According to Donoho [(1994), Theorem 1], the minimax affine estimator for $h(0)$ in the full model equals $\hat{\lambda} h_\xi(0) = \gamma_0 c_{h_\xi} Y_{h_\xi}/\|h_\xi\|$; that is, $\widehat{h(0)}$ defined in (3.4). (In fact, it is not difficult to verify the minimaxity of this estimator directly using Anderson's lemma.) Plugging (5.11) and $c_{h_\xi} = 2\rho/(2\rho+1)$ into the formula for the risk $R_{\text{aff}}(h_\xi)$, one arrives at the last assertion. □

PROOF OF THEOREM 4.1. Since the arguments follow the lines of Drees (1998a,b), we just give a sketch of the proof. Let η_i, $i \in \mathbb{N}$, be i.i.d. standard exponential random variables (defined on some abstract probability space), $S_i := \sum_{j=1}^{i} \eta_j$, $i \in \mathbb{N}$, the pertaining partial sums and $\tilde{Q}_{n,k_n,F} := F^{-1}(1 - S_{[k_n t]+1}/n)$, $0 \le t \le 1$. According to Theorem 5.4.3 of Reiss (1989), the variational distance between the distribution of Q_{n,k_n} under $P_{n,h}^n$ and the distribution of $\tilde{Q}_{n,k_n,F_{n,h}}$ vanishes uniformly as n tends to infinity:

$$
\begin{aligned}
&\|\mathscr{L}(Q_{n,k_n}|P_{n,h}^n) - \mathscr{L}(\tilde{Q}_{n,k_n,F_{n,h}})\| \longrightarrow 0 \\
&\quad\Longrightarrow \left|\mathscr{L}\left(\hat{\gamma}_n^{\text{loc}}|P_{n,h}^n\right) - \mathscr{L}\left(T_{\text{loc}}(\tilde{Q}_{n,k_n,F_{n,h}})\right)\right| \longrightarrow 0
\end{aligned}
\tag{5.12}
$$

uniformly for $h \in \mathscr{H}_\rho \cap \mathscr{H}^{(n)}$ with T_{loc} defined in (4.2).

The arguments of Drees [(1998a), proof of Theorem 2.1] show that there exists a standard Brownian motion W such that for all $\delta \in (0, 1/2)$,

$$
\begin{aligned}
\sup_{h \in \mathscr{H}_\rho \cap \mathscr{H}^{(n)}} \sup_{t \in (0,1]} t^{\gamma_0 + 1/2 + \delta} \Bigg| & \frac{\tilde{Q}_{n,k_n,F_{n,h}}}{F_{n,h}^{-1}(1 - k_n/n)} \\
& - \Bigg(t^{-\gamma(F_{n,h})} - \gamma(F_{n,h}) t^{-(\gamma(F_{n,h})+1)} \frac{W(k_n t)}{k_n} \\
& + t^{-\gamma(F_{n,h})} \int_t^1 \frac{d_n(h(s c_n k_n/n) - h(0))}{s}\, ds \Bigg) \Bigg| \\
= o(k_n^{-1/2}) \quad \text{a.s.}
\end{aligned}
$$

and thus a.s.,

$$
\begin{aligned}
\sup_{h \in \mathscr{H}_\rho \cap \mathscr{H}^{(n)}} \sup_{t \in (0,1]} t^{\gamma_0 + 1/2 + \delta} \Bigg| & \frac{\tilde{Q}_{n,k_n,F_{n,h}}}{F_{n,h}^{-1}(1 - k_n/n)} \\
& - \Bigg(t^{-\gamma_0} - \gamma_0 t^{-(\gamma_0+1)} \frac{W(k_n t)}{k_n} \\
& + d_n t^{-\gamma_0} \int_t^1 \frac{h(\xi^{1/\rho} s)}{s}\, ds \Bigg) \Bigg| = o(k_n^{-1/2}).
\end{aligned}
\tag{5.13}
$$

Next check that the functional T_{loc} is differentiable at $z_{\gamma_0}: t \mapsto t^{-\gamma_0}$ in the following sense. As $\lambda_n \to 0$ the convergence

$$\begin{aligned} &\frac{T_{\mathrm{loc}}(z_{\gamma_0} + \lambda_n y_n) - \gamma_0}{\lambda_n} \\ &\qquad \longrightarrow T'_{\mathrm{loc}}(y) := \frac{\rho+1}{\rho} \int_0^1 t^{\gamma_0}(1-(\rho+1)t^\rho) y(t)\,dt \end{aligned} \tag{5.14}$$

holds uniformly for all functions y_n with $\sup_{t\in(0,1]} t^{\gamma_0+1/2+\delta}|y_n(t)| \le 1$ such that $z_{\gamma_0} + \lambda_n y_n$ is a nondecreasing function. This can be verified in a similar way as the differentiability of the Hill functional in Example 3.1 of Drees (1998b) using the representation

$$\frac{z_{\gamma_0}(t) + \lambda_n y_n(t)}{z_{\gamma_0}(1) + \lambda_n y_n(1)} = t^{-\gamma_0}\left(1 + \lambda_n \frac{t^{\gamma_0} y_n(t) - y_n(1)}{1 + \lambda_n y_n(1)}\right),$$

which implies

$$\log \frac{z_{\gamma_0}(t) + \lambda_n y_n(t)}{z_{\gamma_0}(1) + \lambda_n y_n(1)} = -\gamma_0 \log t + \lambda_n(t^{\gamma_0} y(t) - y(1)) + o(\lambda_n)$$

and

$$\left(\frac{z_{\gamma_0}(t) + \lambda_n y_n(t)}{z_{\gamma_0}(1) + \lambda_n y_n(1)}\right)^{-\rho/\gamma_0} = t^\rho\left(1 - \lambda_n \frac{\rho}{\gamma_0}(t^{\gamma_0} y(t) - y(1))\right) + o(\lambda_n). \tag{5.15}$$

Since $W_n: t \mapsto -k_n^{-1/2} W(k_n t)$ is a standard Brownian motion, too, and T_{loc} is scale invariant [i.e., $T_{\mathrm{loc}}(az) = T_{\mathrm{loc}}(z)$ for $a > 0$], it follows from (5.13) and (5.14) by the well-known δ–method that

$$\begin{aligned} &\sup_{h \in \mathscr{H}_\rho \cap \mathscr{H}^{(n)}} \Bigg| T_{\mathrm{loc}}(\tilde{Q}_{n,k_n,F_{n,h}}) \\ &\qquad - \Bigg(\gamma_0 + k_n^{-1/2}\gamma_0 T'_{\mathrm{loc}}(z_{\gamma_0+1} W_n) \\ &\qquad\qquad + d_n T'_{\mathrm{loc}}\left(\left(t^{-\gamma_0}\int_t^1 \frac{h(\xi^{1/\rho}s)}{s}\,ds\right)_{t\in(0,1]}\right)\Bigg)\Bigg| \\ &= o(k_n^{-1/2}) \qquad \text{a.s.} \end{aligned} \tag{5.16}$$

with

$$\begin{aligned}
&\mathscr{L}(T'_{\mathrm{loc}}(z_{\gamma_0+1}W_n)) \\
&\quad = \mathscr{L}\left(\frac{\rho+1}{\rho}\int_0^1 t^{-1}\,(1+(\rho+1)t^{\rho})\,W_n(t)\,dt\right) \\
&\quad = \mathscr{N}\Bigg(0, \left(\frac{\rho+1}{\rho}\right)^2 \int_0^1 (st)^{-1}(1+(\rho+1)s^{\rho}) \\
&\qquad\qquad \times (1+(\rho+1)t^{\rho})\min(s,t)\,ds\,dt\Bigg) \qquad (5.17)\\
&\quad = \mathscr{N}\,(0, 2(\rho+1)/(2\rho+1))
\end{aligned}$$

[cf. Drees (1998b), Theorem 3.1] and

$$T'_{\mathrm{loc}}\left(\left(t^{-\gamma_0}\int_t^1 \frac{h(\xi^{1/\rho}s)}{s}\,ds\right)_{t\in(0,1]}\right) = \frac{\rho+1}{\rho}\int_0^1 (1-t^{\rho})\,h\left(\xi^{1/\rho}t\right)\,dt, \qquad (5.18)$$

where for the last equation we used integration by parts.

In view of $k_n^{-1/2} \sim \xi^{-1/(2\rho)}d_n$ and $\gamma(F_{n,h}) = \gamma_0 + d_n h(0)$, we conclude from (5.12) and (5.16)–(5.18) that

$$\sup_{h\in\mathscr{H}_\rho\cap\mathscr{H}^{(n)}}\ \sup_{x\in\mathbb{R}} \left|P^n_{n,h}\left\{\hat{\gamma}^{\mathrm{loc}}_n - \gamma(F_{n,h}) \le d_n x\right\} - \mathscr{N}\,(\mu_h, \sigma_h^2)\,(-\infty, x]\right| \longrightarrow 0,$$

so

$$\sup_{h\in\mathscr{H}_\rho\cap\mathscr{H}^{(n)}} P^n_{n,h}\left\{\left|\hat{\gamma}^{\mathrm{loc}}_n - \gamma(F_{n,h})\right| > d_n a\right\} \longrightarrow \sup_{h\in\mathscr{H}_\rho} \mathscr{N}\left(\mu_h, \sigma_h^2\right)\,[-a,a]^c, \qquad (5.19)$$

where

$$\mu_h = \frac{\rho+1}{\rho}\int_0^1 (1-t^{\rho})(h(\xi^{1/\rho}t) - h(0))\,dt \quad \text{and} \quad \sigma_h^2 = \frac{2(\rho+1)\gamma_0^2}{(2\rho+1)\xi^{1/\rho}}.$$

According to Anderson's Lemma, the supremum on the right-hand side is attained at that $h \in \mathscr{H}_\rho$ which maximizes $|\int_0^1 (1-t^{\rho})(h(\xi^{1/\rho}t) - h(0))\,dt|$. This leads to the maximal risk

$$\mathscr{N}\left(\frac{\rho+1}{\rho}\int_0^1 (1-t^{\rho})(\xi^{1/\rho}t)^{\rho}\,dt, \sigma_h^2\right)\,[-a,a]^c = \mathscr{N}\left(\frac{\xi}{2\rho+1}, \sigma_h^2\right)\,[-a,a]^c,$$

which equals (3.5). □

PROOF OF THEOREM 4.2. First note that there exist constants $0 < m < M < \infty$ such that

$$\sup_{F\in\mathscr{F}(\gamma_1,\varepsilon,u)} P^n_F\left\{m < \hat{k}_n n^{-2\rho/(2\rho+1)} < M\right\} \longrightarrow 1, \qquad (5.20)$$

because $\bar{\gamma}_n$ is uniformly consistent on $\mathscr{F}(\gamma_1, \varepsilon, u)$ and ξ defined in (3.2) is a continuous function of $\gamma_0 \in (\gamma_1 - \varepsilon, \gamma_1 + \varepsilon)$. Moreover, it is easily seen that the following uniform analog to (5.13) holds for $\delta \in (0, 1/2)$:

$$\begin{aligned}
&\sup_{m<k_n n^{-2\rho/(2\rho+1)}<M} \; \sup_{F\in\mathscr{F}(\gamma_1,\varepsilon,u)} \; \sup_{t\in(0,1]} t^{\gamma_0+1/2+\delta}\Bigg| \frac{\tilde{Q}_{n,k_n,F}}{F^{-1}(1-k_n/n)} \\
&\qquad - \left(t^{-\gamma(F)} - \gamma(F) t^{-(\gamma(F)+1)} \frac{W(k_n t)}{k_n} + t^{-\gamma(F)} \int_t^1 \frac{\eta(s)}{s}\, ds \right) \Bigg| \\
&\qquad = o(n^{-\rho/(2\rho+1)}) \qquad \text{a.s.}
\end{aligned} \tag{5.21}$$

Next observe that $\hat{\gamma}_n = T_{\text{glob}}(Q_{n,\hat{k}_n})$ where

$$T_{\text{glob}}(z) := T_{\text{Hill}}(z) + T_{\text{Hill}}(z) \left(\frac{\rho+1}{\rho} \right)^2 \left(\int_0^1 \left(\frac{z(t)}{z(1)} \right)^{-\rho/T_{\text{Hill}}(z)} dt - \frac{1}{\rho+1} \right),$$

$$T_{\text{Hill}}(z) := \int_0^1 \log \frac{z(t)}{z(1)}\, dt.$$

Define $\gamma_n := T_{\text{Hill}}(z_{\gamma_0} + \lambda_n y_n)$ with z_{γ_0}, λ_n and y_n as in the proof of Theorem 4.1. Analogously to (5.15), one obtains

$$\left(\frac{z_{\gamma_0}(t) + \lambda_n y_n(t)}{z_{\gamma_0}(1) + \lambda_n y_n(1)} \right)^{-\rho/\gamma_n} = t^{\rho\gamma_0/\gamma_n} - \lambda_n \frac{\rho}{\gamma_0} t^{\rho} \left(t^{\gamma_0} y(t) - y(1) \right) + o(\lambda_n)$$

uniformly for $\gamma_0 \in (\gamma_1 - \varepsilon, \gamma_1 + \varepsilon)$. Therefore, again one may verify by the methods used in Drees [(1998b), Example 3.1] that $(\gamma_n - \gamma_0)/\lambda_n \to \int_0^1 (t^{\gamma_0} y(t) - y(1))\, dt$ and

$$\begin{aligned}
&T_{\text{glob}}(z_{\gamma_0} + \lambda_n y_n) \\
&\quad = \gamma_0 + \lambda_n \int_0^1 t^{\gamma_0} y(t) - y(1)\, dt \\
&\qquad + \gamma_n \left(\frac{\rho+1}{\rho} \right)^2 \left(\frac{1}{\rho\gamma_0/\gamma_n + 1} - \frac{1}{\rho+1} - \lambda_n \frac{\rho}{\gamma_0} \int_0^1 t^{\rho} (t^{\gamma_0} y(t) - y(1))\, dt \right) \\
&\qquad + o(\lambda_n) \\
&\quad = \gamma_0 + \lambda_n \frac{\rho+1}{\rho} \int_0^1 t^{\gamma_0} (1 - (\rho+1) t^{\rho}) y(t)\, dt + o(\lambda_n) \\
&\quad = \lambda_n T'_{\text{loc}}(y) + o(\lambda_n)
\end{aligned}$$

uniformly for $\gamma_0 \in (\gamma_1 - \varepsilon, \gamma_1 + \varepsilon)$, that is, T_{glob} is uniformly Hadamard differentiable at z_{γ_0} with the same derivative as T_{loc}.

Due to (5.20) and the uniform approximation (5.21) of $\tilde{Q}_{n,k_n,F}$, one can conclude the proof in the same way as the proof of Theorem 4.1. □

Acknowledgments The criticism by an Associate Editor and the remarks by the referees led to an improvement of the presentation.

REFERENCES

ADLER, R. J., FELDMAN, R. E. and TAQQU, M. S. (1998). *A Practical Guide to Heavy Tails.* Birkhäuser, Boston.

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.

DANIELSSON, J., DE HAAN, L., PENG, L. and DE VRIES, C. G. (1998). Using a bootstrap method to choose the sample fraction in tail index estimation. Preprint, Erasmus Univ., Rotterdam. (Available at http://www.few.eur.nl/few/people/cdevries/workingpapers/hilboot.pdf.)

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.

DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270.

DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633–667.

DONOHO, D. L. and LIU, R. C. (1991b). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

DREES, H. (1998a). On smooth statistical tail functionals. *Scand. J. Statist* **25** 187–210.

DREES, H. (1998b). A general class of estimators of the extreme value index. *J. Statist. Plann. Inference* **66** 95–112.

DREES, H. (1998c). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.* **26** 434–448.

DREES, H. (1998d). Estimating the extreme value index. Habilitation thesis, Univ. Cologne. (Available at http://euklid.mi.uni-koeln.de/ ~hdrees/habil.ps.)

DREES, H. (1999). On fixed-length confidence intervals for a bounded normal mean. *Statist. Probab. Lett.* **44** 399–404.

DREES, H. and KAUFMANN, E. (1998). Selecting the optimal sample fraction in univariate extreme value statistics. *Stochastic Process. Appl.* **75** 149–172.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, TH. (1997). *Modelling Extremal Events.* Springer, Berlin.

FALK, M. (1995). LAN of extreme order statistics. *Ann. Inst. Statist. Math.* **47** 693–717.

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084.

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1985). On nonparametric estimation of values of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 19–32.

JANSSEN, A. and MAROHN, F. (1994). On statistical information of extreme order statistics, local extreme value alternatives, and Poisson point processes. *J. Multivariate Anal.* **48** 1–30.

LEPSKII, O. V. (1991). Asymptotic minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LOW, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545–554.

LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554.

MAROHN, F. (1997). Local asymptotic normality in extreme value index estimation. *Ann. Inst. Statist. Math.* **49** 645–666.

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.

ROOTZÉN, H. and TAJVIDI, N. (1997). Extreme value statistics and wind storm losses: A case study. *Scand. Actuar. J.* **97** 70–94.

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

STRASSER, H. (1985). *Mathematical Theory of Statistics.* de Gruyter, Berlin.

VAN DER VAART, A. W. (1988) *Statistical Estimation in Large Parameter Spaces. CWI Tract* **44**. CWI, Amsterdam.

ZAJDENWEBER, D. (1995). Business interruption insurance, a risky business. A study on some Paretian risk phenomena. *Fractals* **3** 601–608.

ZEYTINOGLU, M. and MINTZ, M. (1984). Optimal fixed size confidence procedures for a restricted normal mean. *Ann. Statist.* **2** 945–957.

INSTITUTE FOR APPLIED MATHEMATICS
RUPRECHT-KARLS-UNIVERSITY
HEIDELBERG IM NEUENHEIMER FELD 294
69120 HEIDELBERG
GERMANY
E-MAIL: hdrees@statlab.uni-heidelberg.de