

# Semiparametric Lower Bounds for Tail Index Estimation

JAN BEIRLANT <sup>\*</sup>  
CHRISTEL BOUQUIAUX <sup>†</sup>  
BAS J.M. WERKER <sup>‡§</sup>

Katholieke Universiteit Leuven  
Université Libre de Bruxelles  
Tilburg University

August 23, 2004

## Abstract

We consider estimation of the tail index parameter from i.i.d. observations in Pareto and Weibull type models, using a local and asymptotic approach. The slowly varying function describing the non-tail behavior of the distribution is considered as an infinite dimensional nuisance parameter. Without further regularity conditions, we derive a Local Asymptotic Normality (LAN) result for suitably chosen parametric submodels of the full semiparametric model. From this result, we immediately obtain the optimal rate of convergence of tail index parameter estimators for more specific models previously studied. On top of the optimal rate of convergence, our LAN result also gives the minimal limiting variance of estimators (regular for our parametric model) through the convolution theorem. We show that the classical Hill estimator is regular for the submodels introduced with limiting variance equal to the induced convolution theorem bound. We also discuss the Weibull model in this respect.

Running title: Semiparametric tail index estimation

Keywords: Extreme value theory, Local Asymptotic Normality, Pareto model, Weibull model

AMS-classification: 62G32

---

<sup>\*</sup>Department Wiskunde, Katholieke Universiteit Leuven, Celestijnenlaan 200 B, B-3001, Leuven, Belgium. E-mail: Jan.Beirlant@wis.kuleuven.ac.be

<sup>†</sup>Institut de Statistique, Université Libre de Bruxelles, Campus de la Plaine, CP210, B-1050 Bruxelles, Belgium. E-mail: Bouquiac@ulb.ac.be

<sup>‡</sup>Finance and Econometrics group, CentER, Tilburg University, P.O.Box 90153, 5000 LE, Tilburg, The Netherlands. E-mail: B.J.M.Werker@TilburgUniversity.nl

<sup>§</sup>The authors thank John Einmahl and two referees for their pertinent remarks.

# 1 Introduction

Consider an i.i.d. sequence of random variables  $X_1, \dots, X_n$  with common distribution function  $F$ . In this paper we assume that  $F$  is either of the Pareto type or of the Weibull type. More precisely,  $F$  is said to be of the Pareto type if

$$1 - F(x) = [xl(x)]^{-1/\gamma}, \quad x \geq 1, \quad (1.1)$$

where  $\gamma > 0$  is called the (Pareto) tail index parameter and  $l(\cdot)$  is some slowly varying function in the neighborhood of infinity. Similarly, we say that  $F$  is of the Weibull type if

$$-\log[1 - F(x)] = [xl(x)]^{-1/\tau}, \quad x \geq 1, \quad (1.2)$$

where  $\tau > 0$  is called the Weibull tail index parameter and, as before,  $l(\cdot)$  is some slowly varying function in the neighborhood of infinity. In this paper, we will be interested in the behavior of the distributions near infinity. We therefore, only require (1.1) and (1.2) for values of  $x \geq 1$ .

We analyze the Pareto and Weibull type models from a semiparametric point of view in which we take the tail index parameter ( $\gamma$  for the Pareto case and  $\tau$  for the Weibull model) as the parameter of interest and  $l(\cdot)$  as a (functional) nuisance parameter. A natural approach might be to use the tangent space arguments for semiparametric models with i.i.d. observations as set out in, e.g., Bickel, Klaassen, Ritov, and Wellner (1993). However, these results are not applicable in the model under study due to the non-smoothness of the parameter of interest as functional of the underlying distribution. The tangent space reasonings are based on pathwise differentiability of the parameter of interest with respect to the underlying distribution against the tangent spaces. This differentiability, however, does not hold for the extreme value index.

The present paper offers the following contributions. First, we *unify* several known results concerning the optimal rate of convergence for tail index estimators (notably, the results of Hall and Welsh (1984) and Drees (1998) for the Pareto model). Without imposing further restrictions to (1.1) or (1.2), we construct alternatives that are locally asymptotically normal with respect to some fixed distribution (which is not necessarily the strict Pareto) and that converge at an arbitrary rate. Subsequently, we show that the extra smoothness conditions imposed on the distribution in, e.g., Hall and Welsh (1984) or Drees (1998), induce immediately a bound on the rate of convergence any (uniformly consistent) estimator can achieve. Given a rate of convergence (we define precisely what we mean by this in Section 5), one may wonder what is the minimal limiting variance of estimators attaining this rate, i.e. a Cramer-Rao type bound. We introduce suitably chosen parametric submodels that are Locally Asymptotically Normal (LAN). The convolution theorem (see, e.g, Le Cam and Yang, 1990) then gives lower bounds for the (asymptotic) precision with which the tail index parameter can be estimated when using estimators that are regular with respect to these parametric submodels. For the Pareto model, we show that the widely-used Hill estimator has a limiting variance which equals the lower bound obtained from the convolution theorem. In these discussions we do not consider a possible adaptive choice of the rate of convergence, see, e.g., Hall and Welsh (1985).

We also consider *Weibull* type distributions. These distributions are much less studied than the Pareto type distributions. However, the Weibull model offers some properties that

are very useful in specific applications. We again give a LAN result (for suitably chosen local alternatives for the slowly varying nuisance function) and show that, under some conditions, an estimator provided in Beirlant et al. (1995) has a limiting variance which equals the lower bound induced by the convolution theorem in our parametric submodels.

Related work on lower bounds for the speed of convergence can be found in the papers of Hall and Welsh (1984) and Drees (1998). Hall and Welsh (1984) establish the optimal rate of convergence for a specific semiparametric model. Drees (1998) expands these results to a more general class of models and to other maximal domains of attraction (i.e., allowing  $\gamma \in \mathbf{R}$ ). We unify the aforementioned results for the positive  $\gamma$  case. Other papers using the LAN paradigm in the case of extreme value index estimation are Falk (1995), Wei (1995), and Marohn (1997). In these papers, a LAN condition is derived for the largest order statistics. We also find that inference can be based on the largest values observed, since only these observations appear in the central sequence of our parametric submodels. Both Wei (1995) and Marohn (1997) assume that the upper-tail of the distribution essentially belongs to a parametric family. Drees (2001) considers the estimation problem from the related point of view of convergence of experiments. While that paper is concerned with minimax bounds, we consider convolution theorem variance bounds. Compared to minimax results, results based on the convolution theorem are stronger, but only apply to estimators that are regular for the model under consideration. Proposition 2.1 of Drees (2001) can be used to obtain convolution theorem bounds in the vicinity of the strict Pareto distribution. We consider local alternatives to all distribution functions of the semiparametric model of interest.

The setup of the paper is as follows. In Section 2, we consider the Pareto model and obtain a LAN result for appropriately defined local alternatives. The LAN property yields lower bounds on the speed of convergence and on the asymptotic dispersion of estimators that are regular with respect to the parametric models introduced. This is detailed in Section 3. Applications of the general results to more specific Pareto type models are provided in Section 4. In Section 5, we show that the Hill estimator attains the variance lower bound induced by the convolution theorem applied to our parametric submodels. In Section 6 and 7 we prove similar results for the Weibull model. Finally, the appendix gathers some technical proofs.

## 2 Local Asymptotic Normality of the Pareto Model

Consider a fixed continuous distribution function  $F_0$  of the Pareto type (1.1) with parameters  $\gamma_0 > 0$  and  $l_0(\cdot)$ , i.e.,

$$1 - F_0(x) = [xl_0(x)]^{-1/\gamma_0}, \quad x \geq 1. \quad (2.1)$$

As mentioned in the introduction, in this paper we take a semiparametric point of view and are interested in the estimation of the Pareto tail index  $\gamma_0$ , while considering the slowly varying function  $l_0(\cdot)$  as nuisance. In this section, we derive a Local Asymptotic Normality (LAN) result for appropriately defined local alternatives of the distribution function  $F_0$ . This allows us not only to discuss optimal rates of convergence for semiparametric estimators, but also to discuss estimators in terms of their asymptotic variance. Formal results in this direction are discussed in general in Section 3 and in Section 4 in particular.

The LAN condition describes the asymptotic behavior of the likelihood ratio of local alternatives with respect to  $F_0$ . The rate of convergence is defined through an arbitrary positive sequence  $(\delta_n)$  with  $\delta_n \rightarrow 0$  and  $\sqrt{n}\delta_n \rightarrow \infty$ ,  $n \rightarrow \infty$ . As long as no further assumptions (like those discussed in Section 4) are made on the set of Pareto-type distributions, the sequence  $(\delta_n)$  is arbitrary.

The LAN condition effectively gives the likelihood ratio for a model that contains a parameter  $u \in \mathbf{R}$  that is used to localize the parameter of interest  $\gamma_0$ . More precisely, for every  $u \in \mathbf{R}$ , we define, for all  $n \geq n_0 := \min\{m \in \mathbf{N} : \gamma_0 + u\delta_n > 0, \forall n \geq m\}$ ,

$$\gamma_n = \gamma_0 + u\delta_n. \quad (2.2)$$

We also define local alternatives for the nuisance function  $l_0(\cdot)$  as follows

$$l_n(x) = \begin{cases} x^{\gamma_n/\gamma_0-1}l_0(x)^{\gamma_n/\gamma_0}, & 1 \leq x \leq t_n \\ l_0(x)(n\delta_n^2)^{\gamma_n-\gamma_0}, & x > t_n \end{cases}, \quad (2.3)$$

where  $n \geq n_1 := n_0 \vee \min\{m \in \mathbf{N} : n\delta_n^2 > 1, \forall n \geq m\}$  and  $t_n := U_0(n\delta_n^2) \rightarrow \infty$ , as  $n \rightarrow \infty$ , with  $U_0(t) = F_0^{-1}(1 - 1/t) := \inf\{s \in \mathbf{R} : F_0(s) = 1 - 1/t\}$ . Since,  $F_0$  is continuous, we have

$$1 - F_0(t_n) = \frac{1}{n\delta_n^2}. \quad (2.4)$$

REMARK 2.1 The alternatives constructed through (2.2) and (2.3) are introduced here in an ad hoc way. However, they are specific in the sense that the Hill estimator is regular with respect to these alternatives and, at the same time, has a limiting variance which equals that of the lower bound induced by the convolution theorem for the alternatives. Details are discussed in Section 4.

REMARK 2.2 Drees (2001) introduces alternatives around the strict Pareto distribution (i.e., fixing  $l_0(\cdot) = 1$ ) of the form

$$F_n^{-1}(1 - t) = t^{-\gamma_0} \exp\left(u\delta_n \int_t^1 \frac{h(n\delta_n^2 s)}{s} ds\right),$$

where  $h$  is a function satisfying appropriate conditions. It remains an open question whether his results with the strict Pareto as center of localization can be extended to more general centers of localization as in (2.3).

The distribution function corresponding to  $\gamma_n$  and  $l_n(\cdot)$  is given by, for  $n \geq n_1$ ,

$$1 - F_n(x) = [xl_n(x)]^{-1/\gamma_n} = \begin{cases} 1 - F_0(x), & 1 \leq x \leq t_n \\ [1 - F_0(x)]^{\gamma_0/\gamma_n} [1 - F_0(t_n)]^{1-\gamma_0/\gamma_n}, & x > t_n \end{cases}. \quad (2.5)$$

It is obvious that, for each fixed  $n \geq n_1$ ,  $F_n$  defines a continuous distribution function such that  $1 - F_n$  is regularly varying at infinity with index  $-1/\gamma_n$ . Furthermore, note that  $F_n$  is absolutely continuous w.r.t.  $F_0$  and density

$$\frac{dF_n}{dF_0}(x) = \begin{cases} 1, & 1 \leq x \leq t_n \\ \frac{\gamma_0}{\gamma_n} \left[ \frac{1-F_0(x)}{1-F_0(t_n)} \right]^{\gamma_0/\gamma_n-1}, & x > t_n \end{cases}. \quad (2.6)$$

The following theorem gives the quadratic approximation of the likelihood ratio of  $F_n$  with respect to  $F_0$  for  $n$  i.i.d. copies  $X_1, \dots, X_n$  of  $X$  with cdf  $F_0$ . It proves that the alternatives constructed are, without further regularity conditions, LAN and identifies the so-called central sequence ( $\Delta^{(n)}$  below).

**Theorem 2.1** *The log-likelihood ratio*

$$\Lambda^{(n)} = \Lambda^{(n)}(X_1, \dots, X_n) = \sum_{i=1}^n \log \frac{dF_n(X_i)}{dF_0(X_i)}$$

of  $F_n$  with respect to  $F_0$  for  $n$  i.i.d. copies  $X_1, \dots, X_n$  of  $X$  with cdf  $F_0$ , satisfies

$$\Lambda^{(n)} = u\Delta^{(n)} - \frac{1}{2} \frac{u^2}{\gamma_0^2} + o_{\mathbf{P}}(1), \quad (2.7)$$

where

$$\Delta^{(n)} = -\frac{\delta_n}{\gamma_0} \sum_{i=1}^n \left( 1 + \log \frac{1 - F_0(X_i)}{1 - F_0(t_n)} \right) I\{X_i > t_n\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/\gamma_0^2). \quad (2.8)$$

Thus, the Fisher information is given by  $1/\gamma_0^2$ .

The proof of this LAN result relies on a simple lemma.

**Lemma 2.1** *Given  $F_0$  that is continuous, we have, for all  $k \in \mathbf{N}$ ,*

$$\int_{t_n}^{\infty} \left( \log \frac{1 - F_0(x)}{1 - F_0(t_n)} \right)^k dF_0(x) = (-1)^k k! [1 - F_0(t_n)].$$

PROOF: Using the transformation  $v = (1 - F_0(x))/(1 - F_0(t_n))$  the integral is reduced to a Gamma integral.  $\square$

PROOF OF THEOREM 2.1: Since  $n\delta_n^2[1 - F_0(t_n)] = 1$ , an application of Chebychev's inequality shows that, under  $F_0$ ,

$$\delta_n^2 \sum_{i=1}^n I\{X_i > t_n\} = 1 + o_{\mathbf{P}}(1),$$

and likewise, using Lemma 2.1 with  $k = 1$  and  $k = 2$ ,

$$\delta_n^2 \sum_{i=1}^n \log \left[ \frac{1 - F_0(X_i)}{1 - F_0(t_n)} \right] I\{X_i > t_n\} = -1 + o_{\mathbf{P}}(1).$$

The quadratic approximation (2.7) now follows immediately, since, under  $F_0$ , we have

$$\begin{aligned} \Lambda^{(n)} &= \sum_{i=1}^n \left( \log \frac{\gamma_0}{\gamma_n} + \left[ \frac{\gamma_0}{\gamma_n} - 1 \right] \log \frac{1 - F_0(X_i)}{1 - F_0(t_n)} \right) I\{X_i > t_n\} \\ &= -\frac{u\delta_n}{\gamma_0} \sum_{i=1}^n I\{X_i > t_n\} + \frac{u^2}{2\gamma_0^2} - \frac{u\delta_n}{\gamma_0} \sum_{i=1}^n \log \left[ \frac{1 - F_0(X_i)}{1 - F_0(t_n)} \right] I\{X_i > t_n\} - \frac{u^2}{\gamma_0^2} + o_{\mathbf{P}}(1) \\ &= u\Delta^{(n)} - \frac{1}{2} \frac{u^2}{\gamma_0^2} + o_{\mathbf{P}}(1) \end{aligned}$$

It remains to show the convergence in distribution of the central sequence  $\Delta^{(n)}$  in (2.8). To this extent, define, for  $1 \leq i \leq n$ ,

$$\xi_{ni} = -\delta_n \left( 1 + \log \frac{1 - F_0(X_i)}{1 - F_0(t_n)} \right) I\{X_i > t_n\} := -\delta_n(1 + a)I\{X_i > t_n\},$$

where  $a \leq 0$  when  $X_i > t_n$ . For fixed  $n$  sufficiently large, the  $\xi_{ni}$ ,  $i = 1, \dots, n$ , are independent random variables. Under  $F_0$ , using (2.4) and Lemma 2.1, we get

$$\begin{aligned} \mathbb{E}I\{X_i > t_n\} &= 1 - F_0(t_n) = 1/(n\delta_n), \\ \mathbb{E}|a|I\{X_i > t_n\} &= -\mathbb{E}(aI\{X_i > t_n\}) = 1 - F_0(t_n), \\ \mathbb{E}|a|^2I\{X_i > t_n\} &= 2(1 - F_0(t_n)), \\ \mathbb{E}|a|^3I\{X_i > t_n\} &= -\mathbb{E}(a^3I\{X_i > t_n\}) = 6(1 - F_0(t_n)). \end{aligned}$$

Therefore, making use of  $|1 + a|^3 \leq (1 + |a|)^3$ , we find

$$\begin{aligned} \mathbb{E}\xi_{ni} &= 0, \\ \text{Var } \xi_{ni} &= \mathbb{E}\xi_{ni}^2 = \delta_n^2(1 - 2 + 2)[1 - F_0(t_n)] = n^{-1}, \\ \mathbb{E}|\xi_{ni}|^3 &\leq \delta_n^3(1 + 3 + 6 + 6)[1 - F_0(t_n)] = 16n^{-1}\delta_n. \end{aligned}$$

Since, for  $n \rightarrow \infty$ ,

$$\frac{\sum_{i=1}^n \mathbb{E}|\xi_{ni}|^3}{(\sum_{i=1}^n \text{Var } \xi_{ni})^{3/2}} \leq 16\delta_n \rightarrow 0,$$

the Liapunov Central Limit Theorem implies

$$\Delta^{(n)} = \frac{1}{\gamma_0} \sum_{i=1}^n \xi_{ni} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/\gamma_0^2).$$

This completes the proof.  $\square$

The central sequence  $\Delta^{(n)}$  obtained in Theorem 2.1, is of the peak-over-threshold (POT) type. This means that we only look at observations that exceed the deterministic threshold  $t_n$ . We will later be interested in the behavior of Hill type estimators, where the threshold  $t_n$  is replaced by an appropriate empirical quantile of the observations. The following LAN result formalizes this. Let  $X_{i:n}$  denote the  $i$ -th order statistic of  $X_1, \dots, X_n$ . Moreover, define the Hill estimator for a sequence  $(k_n)$ , with  $k_n \rightarrow \infty$  and  $1 \leq k_n < n$ , as

$$H_k^{(n)} = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{X_{n-i+1:n}}{X_{n-k_n:n}}. \quad (2.9)$$

**Theorem 2.2** *Let  $(k_n)$  be a sequence of integers tending to infinity. Consider the sequence  $\delta_n = 1/\sqrt{k_n}$  and the corresponding central sequence  $\Delta^{(n)}$  as defined in (2.8). Then, still assuming  $\sqrt{n}\delta_n \rightarrow \infty$  (i.e.,  $k_n/n \rightarrow 0$ ), we have, under  $F_0$ ,*

$$\gamma_0^2 \Delta^{(n)} = -\frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( 1 + \log \frac{1 - F_0(X_{n-i+1:n})}{1 - F_0(X_{n-k_n:n})} \right) + o_{\mathbb{P}}(1), \quad (2.10)$$

or, equivalently,

$$\gamma_0^2 \Delta^{(n)} - \sqrt{k_n} (H_{k_n}^{(n)} - \gamma_0) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log \frac{l_0(X_{n-i+1:n})}{l_0(X_{n-k_n:n})} + o_{\mathbf{P}}(1). \quad (2.11)$$

The proof being more technical, it is left for the appendix.

### 3 LAN, optimal rates of convergence, and the convolution theorem

A LAN condition as in Theorem 2.1 or 2.2 allows for the derivation of bounds on the optimal rate of convergence of “reasonable” estimators for the tail-index parameter  $\gamma$ . For various specific models (see, e.g., Hall and Welsh, 1984, and Drees, 1998) such optimal rates of convergence are already known and Section 4 discusses in detail how these known results can easily be obtained in the present framework. But the LAN condition allows for more precise lower bounds on the asymptotic behavior of estimators regular in the parametric model than the *rate* of convergence alone. Through the so-called convolution theorem, one obtains lower bounds for the asymptotic distribution of these estimators whose rate of convergence is optimal. In particular, this gives a lower bound for the variance of the asymptotic distribution. All general consequences of the LAN condition discussed in this section are well known, but repeated for the reader’s convenience. A proof of all results can be found in, e.g., Le Cam and Yang (1990) or Bickel et al. (1993).

Optimal rates of convergence follow from the fact that sequences of probability measures that are LAN, are automatically contiguous.

**Lemma 3.1** *If the product measures based on i.i.d. copies of  $F_n$  and  $F_0$  are LAN (as in Theorem 2.1 and 2.2), then they are contiguous.*

We use contiguity in this paper in the sense of Theorem 3.1.1.b of Le Cam and Yang (1990), i.e. for any sequence of random variables  $r_n = r_n(X_1, \dots, X_n)$ , we have  $r_n = O_{\mathbf{P}}(1)$ , under  $F_n$ , if and only if  $r_n = O_{\mathbf{P}}(1)$ , under  $F_0$ .

Let  $\mathcal{P}$  denote an arbitrary class of distributions of the Pareto type (2.1). More specific examples for the Pareto case will be considered in Section 4. Fix a distribution  $F_0 \in \mathcal{P}$  and a sequence  $(\delta_n)$  such that  $\sqrt{n}\delta_n \rightarrow \infty$ . The sequence  $(\delta_n)$  provides an upper bound on the rate of convergence of an estimator, provided that the local alternatives  $F_n$  constructed from  $\gamma_n$  in (2.2) and  $l_n(\cdot)$  in (2.3) belong to the model  $\mathcal{P}$  and provided that we require the estimator to be uniformly consistent over  $\mathcal{P}$ .

**Theorem 3.1** *Suppose that the local alternatives  $F_n$  constructed in (2.2) and (2.3) are such that  $F_n \in \mathcal{P}$ . Let  $\hat{\gamma}_n$  be an estimator of  $\gamma$  for which*

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{F \in \mathcal{P}} \mathbf{P}_F \{ \alpha_n |\hat{\gamma}_n - \gamma| > M \} = 0, \quad (3.1)$$

then

$$\alpha_n = O(\delta_n^{-1}). \quad (3.2)$$

PROOF: The consistency condition (3.1) implies in particular that  $\alpha_n(\hat{\gamma}_n - \gamma_n) = O_{\mathbf{P}}(1)$ , under  $F_n$ . By the contiguity following from Lemma 3.1, this implies that  $\alpha_n(\hat{\gamma}_n - \gamma_n) = O_{\mathbf{P}}(1)$  under  $F_0$ . Since we obviously also have from (3.1) that  $\alpha_n(\hat{\gamma}_n - \gamma_0) = O_{\mathbf{P}}(1)$  under  $F_0$ , we obtain immediately  $\alpha_n(\gamma_n - \gamma_0) = O(1)$ . Using (2.2), this completes the proof.  $\square$

If the model  $\mathcal{P}$  is taken as all distribution functions of the form (2.1), then, as we have seen in Section 2, the alternatives  $F_n$  belong to  $\mathcal{P}$  whatever the sequence  $(\delta_n)$ . Thus, given a possible sequence  $(\alpha_n)$ , one can always find a sequence  $(\delta_n)$ , converging to zero very slowly, such that (3.2) does not hold, i.e., such that  $\limsup_{n \rightarrow \infty} \alpha_n \delta_n = \infty$ . This implies that there cannot exist a uniformly consistent estimator of  $\gamma$  in the full Pareto model, no matter how weak the consistency requirement in (3.1), i.e., no matter how slowly  $(\alpha_n)$  converges to infinity. Even if  $\mathcal{P}$  is taken as a subset of the full semiparametric model consisting of all distribution functions of the form (2.1), uniformly consistent estimation is not possible if the interior of  $\mathcal{P}$  (with respect to the variational distance) is not empty. This follows along the same lines as the proof of Theorem 3.1 upon noting that the variational distance between  $F_n$  and  $F_0$  is bounded by  $2[1 - F_0(t_n)]$  and, hence, converges to zero. The same result can easily be obtained by direct methods, but it is also an immediate consequence of our general LAN result. Concluding, if meaningful optimal rates of convergence are to be found, one must restrict the model by imposing extra regularity on the slowly varying function  $l(\cdot)$  in (2.1). This will be considered for previously studied models in Section 4.

Another important consequence of the LAN property is the so-called convolution theorem (see, e.g., Le Cam and Yang (1990), page 85). This theorem gives a lower bound for the asymptotic variance of regular estimators, given a fixed rate of convergence  $\alpha_n = \delta_n^{-1}$ .

**Theorem 3.2** *Suppose that the product measures based on i.i.d. copies of  $F_n$  and  $F_0$  are LAN (as in Theorem 2.1 and 2.2). Suppose, moreover, that  $\hat{\gamma}_n$  is a regular estimator for  $\gamma$  in the sense, for  $n \rightarrow \infty$ ,*

$$\begin{aligned} \delta_n^{-1}(\hat{\gamma}_n - \gamma_0) &\xrightarrow{\mathcal{L}} U, \text{ under } F_0, \text{ and} \\ \delta_n^{-1}(\hat{\gamma}_n - \gamma_n) &\xrightarrow{\mathcal{L}} U, \text{ under } F_n, \end{aligned} \quad (3.3)$$

where  $U$  denotes an arbitrary random variable. Then, we have, under  $F_0$ ,

$$\begin{pmatrix} \gamma_0^2 \Delta^{(n)} \\ \delta_n^{-1}(\hat{\gamma}_n - \gamma_0) - \gamma_0^2 \Delta^{(n)} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} V \\ Z \end{pmatrix}, \quad (3.4)$$

where  $V \sim N(0, \gamma_0^2)$  and  $Z$  are independently distributed. Under  $F_n$ , the same convergence of the sequence of vectors in (3.4) holds, but with  $V \sim N(u, \gamma_0^2)$ .

The convolution theorem states that, given regularity of the estimator as defined above, the most concentrated limiting distribution possible for estimating  $\gamma$ , is a  $N(0, \gamma_0^2)$  distribution. All regular estimators have a limiting distribution that is the convolution of this  $N(0, \gamma_0^2)$  and some other distribution. If this other distribution is not degenerated, the limiting distribution is more spread out than the  $N(0, \gamma_0^2)$  distribution, in the sense that it gives rise to a larger asymptotic variance. In Section 5, we show that the Hill estimator with  $k_n = \delta_n^{-1/2}$  is, under some conditions, regular for the alternatives introduced and has a limiting variance equal to  $\gamma_0^2$ . Section 7 shows the analogous result for an estimator introduced in Beirlant et al. (1995) for the Weibull model.

## 4 More specific Pareto type models

We illustrate the general theory of the previous sections by reviewing two examples from the literature. In these examples, more specific assumptions are made on the slowly varying function  $l(\cdot)$ . We will consider in this section the models introduced in Hall and Welsh (1984) and Drees (1998).

**Example 4.1** Hall and Welsh (1984) consider the model described by all densities of the form

$$f(x) = C \frac{x^{1/\gamma-1}}{\gamma} [1 + r(x)], \quad \gamma > 0, C > 0. \quad (4.1)$$

The model  $\mathcal{P}$  considered in Hall and Welsh (1984) is defined starting from fixed  $\gamma_0 > 0$ ,  $\rho > 0$ ,  $C_0 > 0$ , and  $\varepsilon > 0$ , as the set of distribution functions, satisfying (4.1), for which  $|\gamma - \gamma_0| \leq \varepsilon$ ,  $|C - C_0| \leq \varepsilon$ , and

$$\sup_x |x^{\rho/\gamma} r(x)|, \quad (4.2)$$

is bounded over  $\mathcal{P}$ . For this model, estimators which are uniformly consistent in the sense of (3.1) can be constructed, provided that  $\alpha_n$  converges not too quickly to infinity (i.e., if  $\delta_n$  converges not too slowly to zero).

To be precise, consider the alternatives  $F_n$  constructed around the strict Pareto distribution, i.e.,

$$1 - F_0(x) = x^{-1/\gamma_0}, \quad x \geq 1, \quad (4.3)$$

for some  $\gamma_0 > 0$ . In the notation of (4.1), we have  $C_0 = 1$  and  $r_0(x) = 0$ . One easily verifies that the alternatives  $F_n$  as constructed in (2.2) and (2.3) are such that (4.1) holds with

$$r_n(x) = \begin{cases} \frac{\gamma_n}{\gamma_0} (x/t_n)^{1/\gamma_n-1/\gamma_0} - 1, & 1 \leq x \leq t_n \\ 0 & x > t_n \end{cases}. \quad (4.4)$$

Since

$$\sup_x |x^{\rho/\gamma_n} r_n(x)| = O\left(t_n^{\rho/\gamma_n} \left| \frac{1}{\gamma_n} - \frac{1}{\gamma_0} \right| \right)$$

we find that  $\sup_x |x^{\rho/\gamma_n} r_n(x)|$  remains bounded (as  $n \rightarrow \infty$ ) if and only if  $t_n^{\rho/\gamma_n} \delta_n = O(1)$ , i.e., if and only if

$$\delta_n = O(n^{-\frac{\rho}{2\rho+1}}), \quad n \rightarrow \infty.$$

From Theorem 3.1 we now obtain that  $\alpha_n(\hat{\gamma}_n - \gamma) = O_{\mathbf{P}}(1)$  uniformly over the Hall and Welsh (1984) model implies

$$\alpha_n = O(n^{\frac{\rho}{2\rho+1}}), \quad n \rightarrow \infty. \quad (4.5)$$

In this example, we assumed that  $l_0(x) = C_0 = 1$ , but it can easily be extended to cover the case  $l_0(x) = C_0 \neq 1$ .

**Example 4.2** Drees (1998) imposes that the slowly varying function  $l(\cdot)$  is normalized, i.e., for some  $\eta : [1, \infty) \rightarrow \mathbf{R}$ ,

$$l(x) = C \exp\left(\int_1^x \eta(z)/z dz\right). \quad (4.6)$$

The model  $\mathcal{P}$  considered in Drees (1998) is now defined as all distributions satisfying (2.1) and (4.6) such that

$$\sup_{z \geq 1} |\eta(z)|/h(z) \quad (4.7)$$

is bounded over  $\mathcal{P}$  for some given continuous, positive, and decreasing function  $h$ . As in the Hall and Welsh (1984) model, this model does allow for uniformly consistent estimators in the sense of (3.1).

Fix  $F_0, C_0 > 0$ ,  $\gamma_0 > 0$ , and  $\eta_0$  according to (4.6). The alternatives  $F_n$  constructed in (2.2) and (2.3) now also satisfy (4.6) with  $C_n = C_0$  and

$$\eta_n(z) = \left[ \frac{\gamma_0}{\gamma_n} + \left(1 - \frac{\gamma_0}{\gamma_n}\right) I\{z \leq t_n\} \right] \eta_0(z) + \left( \frac{1}{\gamma_n} - \frac{1}{\gamma_0} \right) I\{z \leq t_n\}. \quad (4.8)$$

Since  $h$  is decreasing, we find

$$\sup_z \frac{|\eta_n(z)|}{h(z)} \leq \max\{1, \frac{\gamma_0}{\gamma_n}\} \sup_z \frac{|\eta_0(z)|}{h(z)} + \left| \frac{1}{\gamma_n} - \frac{1}{\gamma_0} \right| \frac{1}{h(t_n)}$$

The first term on the right-hand side is bounded as  $n \rightarrow \infty$ . In order that the second term is bounded as  $n \rightarrow \infty$ , we need

$$\delta_n/h(t_n) = O(1), \quad n \rightarrow \infty. \quad (4.9)$$

In the special case that  $h(z) = z^{-\rho/\gamma_0}$  and  $\eta_0(z) = 0$ , the condition (4.9) translates to the requirement that  $\delta_n/[n\delta_n^2]^{-\rho}$  is bounded, i.e.,

$$\delta_n = O(n^{-\frac{\rho}{2\rho+1}}) \quad (4.10)$$

The present example is in fact a variation of the Drees (1998) model. Drees (1998) imposes the conditions (4.6) and (4.7) on the slowly varying part of the function  $U$  as defined in Section 2. It is possible to consider exactly Drees' (1998) model in our framework in the neighborhood of the strict Pareto distribution. More precisely, consider  $U_0(t) = F_0^{-1}(1 - 1/t) = t^{\gamma_0}$ ,  $t \geq 1$ . The function  $U_n$  defined by  $U_n(t) = F_n^{-1}(1 - 1/t)$ ,  $t \geq 1$ , with  $F_n$  defined in (2.5), for given sequence  $(\delta_n)$ ,  $\gamma_n = \gamma_0 + u\delta_n$ , and  $t_n = (n\delta_n^2)^{\gamma_0}$ , is then easily seen to be given by

$$U_n(t) = t^{\gamma_n} \exp\left(\int_1^t \frac{\eta_n(z)}{z} dz\right),$$

with

$$\eta_n(z) = \begin{cases} \gamma_0 - \gamma_n & , \quad 1 \leq z \leq n\delta_n^2 \\ 0 & , \quad z > n\delta_n^2 \end{cases}.$$

In this case, the condition that  $\sup_{z \geq 1} |\eta_n(z)|/h(z)$  remains bounded (as  $n \rightarrow \infty$ ) implies that

$$\frac{\delta_n}{h(n\delta_n^2)} = O(1), \quad n \rightarrow \infty.$$

For  $h(z) = z^{-\rho}$  we find the same condition (4.10). Note that Drees (1998) considers the non-Pareto case, i.e., where the tail-index  $\gamma$  may be zero or negative. This is a non-trivial extension that is not covered by our present results.

## 5 The Hill estimator

Section 2 provides a LAN result for suitably chosen parametric families of the semiparametric Pareto type model. In the previous section, we have seen how this result immediately yields the optimal rates of convergence in more specific Pareto type models, like those of Hall and Welsh (1984) and a model inspired by Drees (1998). Furthermore, the LAN result, via the convolution theorem, gives a lower bound on the asymptotic variance of estimators which are regular for the alternatives introduced. In this section, we show that, given a fixed rate of convergence, and apart from a well-known asymptotic bias, the Hill estimator, under a regularity condition, attains this lower bound. Thus, throughout this section, we fix a sequence of nonnegative integers  $(k_n)$  with  $k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$  as  $n \rightarrow \infty$  and the corresponding sequence  $\delta_n = 1/\sqrt{k_n}$ .

Let  $\mathcal{P}$  denote an arbitrary class of distributions of the Pareto type (2.1). Consider a Pareto type distribution  $F \in \mathcal{P}$ . We may decompose the inverse of  $1/(1 - F)$  as follows:

$$\left(\frac{1}{1 - F}\right)^{-1}(t) := \inf\{s : F(s) = 1 - 1/t\} = t^\gamma L(t), \quad t > 1, \quad (5.1)$$

with  $L(\cdot)$  slowly varying at infinity. In order to study the asymptotic behavior of the Hill estimator, we have to impose (like Smith (1982)) a second order condition which specifies the rate of convergence of  $L(tx)/L(t)$  to 1. More precisely, let  $c$  be some constant and  $g : (0, \infty) \rightarrow (0, \infty)$  a  $\rho$ -varying function with  $\rho \leq 0$ . Consider the following asymptotic condition

$$(SR2) \quad \forall x > 1 : \frac{L(tx)}{L(t)} = 1 + cg(t) \int_1^x v^{\rho-1} dv + o(g(t)), \quad \text{as } t \rightarrow \infty. \quad (5.2)$$

The SR2-condition is widely accepted as an appropriate condition to specify the slowly varying part of the model (1.1) in a semi-parametric way. Under the SR2-condition, we have the following result.

**Theorem 5.1** *Suppose that  $F$  is of the Pareto type (1.1) and satisfies the SR2-condition with*

$$\sqrt{k_n}g(n/k_n) \rightarrow A, \quad (5.3)$$

*for some  $A \in \mathbf{R}$ . Then, under the local alternatives defined by  $\gamma_n = \gamma_0 + u\delta_n$  and (2.3), with  $\delta_n = 1/\sqrt{k_n}$ , we have*

$$\sqrt{k_n}(H_{k_n}^{(n)} - \gamma_0) \xrightarrow{\mathcal{L}} N(cA/(1 - \rho) + u, \gamma_0^2),$$

*and*

$$\sqrt{k_n}(H_{k_n}^{(n)} - \gamma_n) \xrightarrow{\mathcal{L}} N(cA/(1 - \rho), \gamma_0^2).$$

The limiting behavior of the Hill estimator for  $u = 0$ , i.e., under  $F_0$  in Theorem 5.1 is well-known (see, e.g., Hall, 1982, Haeusler and Teugels, 1985, or the more recent papers Csörgő and Viharos, 1998, de Haan and Resnick, 1998, and de Haan and Peng, 1998). However, we describe its behavior under our local alternatives as well. We provide a proof in the appendix that is effectively based on Theorem 2.2. Note that Theorem 5.1 is not at odds

with Theorem 2.2 of Drees (1998) which considers the estimator  $H_{\tilde{k}_n}^{(n)}$  with  $\tilde{k}_n/k_n \rightarrow \infty$ . Such an estimator is not regular at the rate  $\delta_n = 1/\sqrt{k_n}$  that we consider.

Observe that, if the SR2-condition is satisfied, then it is also satisfied by the local alternatives constructed in Section 2. More precisely, if the inverse of  $1/(1 - F_0)$  evaluated in  $t > 1$  can be written as  $t^\gamma L_0(t)$  where  $L_0(\cdot)$  satisfies the SR2-condition, say

$$L_0(tx)/L_0(t) = 1 + c_0 g_0(t) \int_1^x u^{\rho_0-1} du + o(g_0(t)),$$

then the same is true for the alternatives  $F_n$ , i.e. the corresponding slowly varying function  $L_n(\cdot)$  can be constructed such that

$$L_n(tx)/L_n(t) = 1 + c_n g_n(t) \int_1^x u^{\rho_n-1} du + o(g_n(t)),$$

with

$$\begin{aligned} c_n &= c_0 \frac{\gamma_n}{\gamma_0}, \\ \rho_n &= \rho_0 \frac{\gamma_n}{\gamma_0}, \\ g_n(t) &= g_0 \left( t^{\gamma_n/\gamma_0} (n\delta_n^2)^{\gamma_n/\gamma_0-1} \right). \end{aligned}$$

Note that  $g_n(n\delta_n^2) = g_0(n\delta_n^2)$ . The above can be proven by noting that the inverse of  $1/(1 - F_n)$  is given by (see (2.5))

$$U_n(t) = \begin{cases} U_0(t) & \text{for } t \leq n\delta_n^2, \\ U_0(t^{\gamma_n/\gamma_0} (1 - F_0(t_n))^{\gamma_n/\gamma_0-1}) & \text{for } t > n\delta_n^2, \end{cases}$$

where, as before,  $U_0(t) = F_0^{-1}(1 - 1/t)$ . Thus, for  $t > n\delta_n^2$ ,

$$L_n(t) = (n\delta_n^2)^{\gamma_0-\gamma_n} L_0 \left( t^{\gamma_n/\gamma_0} (n\delta_n^2)^{1-\gamma_n/\gamma_0} \right).$$

Note, however, that condition (5.3) is not necessarily satisfied by the alternatives  $F_n$ .

## 6 Local Asymptotic Normality of the Weibull Model

The Pareto model, while popular in practice, is not always the best choice in some applications, see, e.g., Keller and Klüppelberg (1991) or Klüppelberg and Villaseñor (1993). See furthermore Chapter 4 in the Beirlant, Teugels, Vynckier (1996) monograph.

Fix  $\tau_0 > 0$  and a slowly varying function  $l_0(\cdot)$  and consider the distribution  $F_0$  given by

$$-\log[1 - F_0(x)] = [xl_0(x)]^{1/\tau_0}, \quad x \geq 1. \quad (6.1)$$

As for the Pareto type model, we consider local alternatives based on an arbitrary positive sequence  $(\delta_n)$  with  $\delta_n \rightarrow 0$  and  $\delta_n^{-1} = o(\log n)$  as  $n \rightarrow \infty$ . For every  $u \in \mathbb{R}$ , we define the local alternatives  $F_n$  through (1.2) with

$$\tau_n = \tau_0 + u\delta_n, \quad (6.2)$$

$$l_n(x) = \begin{cases} x^{\tau_n/\tau_0-1} [l_0(x)]^{\tau_n/\tau_0}, & 1 \leq x \leq t_n \\ l_0(x) [\log(n\delta_n^2)]^{\tau_n-\tau_0}, & x > t_n \end{cases}, \quad (6.3)$$

where  $t_n$  is given by  $-\log(1 - F_0(t_n)) = \log(n\delta_n^2)$ . Elementary calculations show that  $F_n$  is absolutely continuous with respect to  $F_0$ , where  $F_n$  and  $F_0$  coincide for  $1 \leq x \leq t_n$  and for  $x > t_n$  we have

$$\begin{aligned} \log \frac{dF_n}{dF_0}(x) &= \log \frac{\tau_0}{\tau_n} + \left( \frac{\tau_0}{\tau_n} - 1 \right) \log \frac{-\log[1 - F_0(x)]}{-\log[1 - F_0(t_n)]} \\ &\quad - \log(1 - F_0(x)) \left\{ 1 - \left[ \frac{-\log[1 - F_0(x)]}{-\log[1 - F_0(t_n)]} \right]^{\tau_0/\tau_n - 1} \right\}. \end{aligned} \quad (6.4)$$

To state the LAN result for the Weibull model, we define the log-likelihood ratio of the  $n$  i.i.d. variables  $X_1, \dots, X_n$  of  $F_n$  with respect to  $F_0$ :

$$\Lambda^{(n)} = \Lambda^{(n)}(X_1, \dots, X_n) = \sum_{i=1}^n \log \left[ \frac{dF_n}{dF_0}(X_i) \right].$$

**Theorem 6.1** *The log-likelihood ratio  $\Lambda^{(n)}$  satisfies, under  $F_0$ ,*

$$\Lambda^{(n)} = u\Delta^{(n)} - \frac{1}{2} \frac{u^2}{\tau_0^2} + o_{\mathbf{P}}(1), \quad (6.5)$$

where

$$\begin{aligned} \Delta^{(n)} &= \frac{\delta_n}{\tau_0} \sum_{i=1}^n \left( -\log \frac{1 - F_0(X_i)}{1 - F_0(t_n)} - 1 \right) I \{X_i > t_n\} \\ &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/\tau_0^2). \end{aligned} \quad (6.6)$$

The proof of this LAN result is similar to that for the Pareto case. Observe that, from Lemma 2.1, we obtain

$$\int_{t_n}^{\infty} \left( \frac{-\log[1 - F_0(x)]}{-\log[1 - F_0(t_n)]} - 1 \right)^k dF_0(x) = k! \frac{1 - F_0(t_n)}{(-\log[1 - F_0(t_n)])^k}, \quad (6.7)$$

by dividing by  $(\log(1 - F_0(t_n)))^k$ .

**PROOF OF THEOREM 6.1** From (6.7), with  $k = 1$  and  $k = 2$ , we get

$$\delta_n \sum_{i=1}^n \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right) I \{X_i > t_n\} = o_{\mathbf{P}}(1).$$

This implies

$$\sum_{i=1}^n \left( \frac{\tau_0}{\tau_n} - 1 \right) \log \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} I \{X_i > t_n\} = o_{\mathbf{P}}(1).$$

Moreover, combining the inequality

$$\forall t > 1, \forall a < 2: |1 - t^a + a(t - 1)| \leq |a(a - 1)| (t - 1)^2$$

with (6.7) for  $k = 2$  and  $k = 3$  gives

$$\begin{aligned} & \sum_{i=1}^n -\log[1 - F_0(X_i)] \left\{ 1 - \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} \right)^{\tau_0/\tau_n - 1} \right\} I\{X_i > t_n\} \\ &= \sum_{i=1}^n -\log[1 - F_0(X_i)] \left( 1 - \frac{\tau_0}{\tau_n} \right) \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right) I\{X_i > t_n\} + o_{\mathbf{P}}(1). \end{aligned}$$

This last expression can be written as the sum of

$$-(1 - \frac{\tau_0}{\tau_n}) \log[1 - F_0(t_n)] \sum_{i=1}^n \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right)^2 I\{X_i > t_n\},$$

and

$$-(1 - \frac{\tau_0}{\tau_n}) \log[1 - F_0(t_n)] \sum_{i=1}^n \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right) I\{X_i > t_n\},$$

of which the first part vanishes asymptotically in view of (6.7).

The above results imply that we may write

$$\begin{aligned} \Lambda^{(n)} &= -(1 - \frac{\tau_0}{\tau_n}) \log[1 - F_0(t_n)] \sum_{i=1}^n \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right) I\{X_i > t_n\} \\ &+ \log \frac{\tau_0}{\tau_n} \sum_{i=1}^n I\{X_i > t_n\} + o_{\mathbf{P}}(1). \end{aligned}$$

Note

$$\delta_n^2 \sum_{i=1}^n I\{X_i > t_n\} = 1 + o_{\mathbf{P}}(1)$$

and, in virtue of (6.7),

$$-\delta_n^2 \log[1 - F_0(t_n)] \sum_{i=1}^n \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right) I\{X_i > t_n\} = 1 + o_{\mathbf{P}}(1),$$

which proves the quadratic expansion for the log-likelihood ratio.

Let

$$\xi_{ni} = \delta_n \left( -\log[1 - F_0(t_n)] \left( \frac{-\log[1 - F_0(X_i)]}{-\log[1 - F_0(t_n)]} - 1 \right) - 1 \right) I\{X_i > t_n\}.$$

The limiting distribution of the central sequence  $\Delta^{(n)}$  follows from the Liapunov Central Limit Theorem, using (6.7) to obtain

$$\begin{aligned} E\xi_{ni} &= 0, \\ \text{Var } \xi_{ni} &= n^{-1}, \\ E|\xi_{ni}|^3 &\leq 16|\delta_n|/n. \end{aligned}$$

□

The LAN result of Theorem 6.1 is based on a central sequence of the POT-type, i.e. the central sequence consists only of those observations that exceed a given deterministic threshold  $t_n$ . As in the Pareto case, we can also for the Weibull model provide a central sequence based on order statistics.

**Theorem 6.2** *Let  $(k_n)$  be a sequence of integers tending to infinity with  $\sqrt{k_n} = o(\log(n))$ . Consider the sequence  $\delta_n = 1/\sqrt{k_n}$  and the corresponding central sequence  $\Delta^{(n)}$ . Then, we may write*

$$\tau_0^2 \Delta^{(n)} = \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( -\log \frac{1 - F_0(X_{n-i+1:n})}{1 - F_0(X_{n-k_n:n})} - 1 \right) + o_{\mathbf{P}}(1),$$

and

$$\tau_0^2 \Delta^{(n)} - \sqrt{k_n}(\hat{\tau}_{k_n}^{(n)} - \tau_0) = \frac{\log(n/k_n)}{\sqrt{k_n}} \sum_{i=1}^{k_n} \frac{X_{n-i+1:n}}{X_{n-k_n:n}} \left( \frac{l_0(X_{n-i+1:n})}{l_0(X_{n-k_n:n})} - 1 \right) + o_{\mathbf{P}}(1),$$

where the estimator  $\hat{\tau}_{k_n}^{(n)}$  is defined by

$$\hat{\tau}_{k_n}^{(n)} = \frac{\log(n/k_n)}{k_n} \sum_{i=1}^{k_n} \left( \frac{X_{n-i+1:n}}{X_{n-k_n:n}} - 1 \right).$$

The proof is again left for the appendix.

## 7 Estimation in the Weibull model

Beirlant et al. (1995) provide the limiting distribution of the estimator

$$\hat{\tau}_{k_n}^{(n)} = \frac{\log(n/k_n)}{k_n} \sum_{i=1}^{k_n} \left( \frac{X_{n-i+1:n}}{X_{n-k_n:n}} - 1 \right).$$

Our results again allow us to study the behavior of this estimator under the local alternatives constructed. We introduce the following notation. Let  $K_0$  denote the generalized inverse of  $-\log(1 - F_0)$ . Then, we may write  $K_0(t) = t^{\tau_0} L_0(t)$  with  $L_0(\cdot)$  slowly varying.

**Theorem 7.1** *Suppose  $L_0(\cdot)$  defined above satisfies SR2. Let  $(k_n)$  be a sequence of integers tending to infinity with  $\sqrt{k_n} = o(\log(n))$  and  $\sqrt{k_n} g(\log(n/k_n)) \rightarrow A$ . Now, under the local alternatives defined by  $\tau_n = \tau_0 + u\delta_n$  and (6.3), with  $\delta_n = 1/\sqrt{k_n}$ , we find*

$$\sqrt{k_n}(\hat{\tau}_{k_n}^{(n)} - \tau_0) \xrightarrow{\mathcal{L}} \mathcal{N}(-cA + u, 1/\tau_0^2),$$

and

$$\sqrt{k_n}(\hat{\tau}_{k_n}^{(n)} - \tau_n) \xrightarrow{\mathcal{L}} \mathcal{N}(-cA, 1/\tau_0^2).$$

Theorem 6.1 and Theorem 7.1 impose  $1/\delta_n = o(\log n)$ . This implies that the  $\delta_n$  are relatively large and the alternatives  $F_n$  are 'far' from  $F_0$ . We conjecture, but were unable to prove formally, that, e.g., geometric rates of convergence can not be obtained in the Weibull model. This conjecture is based on two considerations. First, a small change in

the parameter  $\tau$  in (6.1) leads to a much larger change in the distribution  $F$ , than a similar change in  $\gamma$  in (2.1). As a consequence, inference about  $\tau$  in the Weibull model is much more difficult than inference about  $\gamma$  in the Pareto model. Formally, for geometric rates  $\delta_n = n^{-\alpha}$  with  $\alpha > 0$ , we expect the log-likelihood ratio in (6.5) to converge to zero. The second consideration regards the estimator discussed in Theorem 7.1 above. In case  $k_n = 1/\delta_n^2$  is chosen too large, the bias  $A$  tends to infinity. This suggest that there is no convergence in distribution of  $\sqrt{k_n}(\hat{\tau}_{k_n}^{(n)} - \tau_0)$ .

## A Some proofs

This appendix contains three proofs that were omitted from the main text in order to improve readability.

PROOF OF THEOREM 2.2: Let  $U_{1:n} \leq \dots \leq U_{n:n}$  be the order statistics of  $n$  i.i.d. uniformly over the interval  $[0, 1]$  distributed r.v.'s  $U_1, \dots, U_n$ . Using the quantile transformation, we obtain

$$\begin{aligned} & \gamma_0^2 \Delta^{(n)} + \frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( 1 + \log \frac{1 - F_0(X_{n-i+1:n})}{1 - F_0(X_{n-k_n:n})} \right) \\ & \stackrel{d}{=} -\frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^n \left( 1 + \log \frac{1 - U_i}{k_n/n} \right) I\{U_i > 1 - k_n/n\} \\ & \quad + \frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( 1 + \log \frac{1 - U_{n-i+1:n}}{1 - U_{n-k_n:n}} \right). \end{aligned}$$

We decompose the latter expression into  $T_1^{(n)} + T_2^{(n)} + T_3^{(n)}$ , with

$$\begin{aligned} T_1^{(n)} &= -\frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^n \log \left[ \frac{1 - U_i}{k_n/n} \right] I\{U_i > 1 - k_n/n\} \\ & \quad + \frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log \left[ \frac{1 - U_{n-i+1:n}}{k_n/n} \right], \\ T_2^{(n)} &= \frac{\gamma_0}{\sqrt{k_n}} \left( k_n - \sum_{i=1}^n I\{U_i > 1 - k_n/n\} \right), \\ T_3^{(n)} &= -\gamma_0 \sqrt{k_n} \log \left[ \frac{1 - U_{n-k_n:n}}{k_n/n} \right]. \end{aligned}$$

Since  $k_n = o(n)$ , we have, by Chebyshev's inequality,

$$U_{n-k_n:n} = 1 - k_n/n + O_P(\sqrt{k_n}/n). \quad (\text{A.1})$$

Since  $\mathbf{P}\{U_i = U_j; i \neq j\} = 0$ , we have

$$\begin{aligned} T_1^{(n)} & \stackrel{d}{=} -\frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^n \log \left[ \frac{1 - U_i}{k_n/n} \right] I\{U_i > 1 - k_n/n\} \\ & \quad + \frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^n \log \left[ \frac{1 - U_i}{k_n/n} \right] I\{U_i > U_{n-k_n:n}\}. \end{aligned}$$

Now, for any  $d \in (0, \infty)$ , put

$$T_1^{(n)}(d) = \frac{\gamma_0}{\sqrt{k_n}} \sum_{i=1}^n \left| \log \frac{1 - U_i}{k_n/n} \right| I \left\{ 1 - k_n/n - d\sqrt{k_n}/n \leq U_i \leq 1 - k_n/n + d\sqrt{k_n}/n \right\}.$$

Note that we have, using (A.1),

$$\lim_{d \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P} \left\{ \left| T_1^{(n)} \right| > T_1^{(n)}(d) \right\} \leq \lim_{d \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P} \left\{ |U_{n-k_n:n} - (1 - k_n/n)| > d\sqrt{k_n}/n \right\} = 0.$$

Hence, in order to prove

$$T_1^{(n)} = o_{\mathbf{P}}(1),$$

it is sufficient to show, for each  $d \in (0, \infty)$ ,

$$T_1^{(n)}(d) = o_{\mathbf{P}}(1). \quad (\text{A.2})$$

But, (A.2) follows easily from the Markov inequality, since

$$E \left( T_1^{(n)}(d) \right) \leq 2d\gamma_0 \max \left( \log \left[ 1 + d/\sqrt{k_n} \right], -\log \left[ 1 - d/\sqrt{k_n} \right] \right) \rightarrow 0.$$

It remains to consider  $T_2^{(n)} + T_3^{(n)}$ . We start by rewriting  $T_3^{(n)}$ . Applying a Taylor series expansion, we find, for  $\theta_n$  between  $k_n/n$  and  $1 - U_{n-k_n:n}$  and using (A.1),

$$\begin{aligned} T_3^{(n)} &= -\gamma_0 \frac{n}{\sqrt{k_n}} (1 - U_{n-k_n:n} - k_n/n) + \frac{\gamma_0 \sqrt{k_n}}{2 \theta_n^2} (1 - U_{n-k_n:n} - k_n/n)^2 \\ &= -\gamma_0 \frac{n}{\sqrt{k_n}} (1 - U_{n-k_n:n} - k_n/n) + o_{\mathbf{P}}(1). \end{aligned} \quad (\text{A.3})$$

To complete the proof, we define the uniform empirical process

$$\alpha_n(s) = \sqrt{n} (G_n(s) - s), \text{ for } 0 \leq s \leq 1,$$

and the uniform quantile process

$$\beta_n(s) = \sqrt{n} (s - U_n(s)), \text{ for } 0 \leq s \leq 1,$$

where

$$G_n(s) = \frac{1}{n} \#\{k : 1 \leq k \leq n, U_k \leq s\},$$

and

$$U_n(s) = \begin{cases} U_{k:n} & \text{if } (k-1)/n < s \leq k/n, \\ U_{1:n} & \text{if } s = 0. \end{cases}$$

Using (A.3), the sum of  $T_2^{(n)}$  and  $T_3^{(n)}$  can now be written as

$$\gamma_0 \sqrt{\frac{n}{k_n}} \left( \alpha_n \left( 1 - \frac{k_n}{n} \right) - \beta_n \left( 1 - \frac{k_n}{n} \right) \right) + o_{\mathbf{P}}(1)$$

From Corollary 2.3 in Csörgő et al. (1986), with  $\lambda = 1$ , one finds

$$\sqrt{\frac{n}{k_n}} \left( \alpha_n \left( 1 - \frac{k_n}{n} \right) - \beta_n \left( 1 - \frac{k_n}{n} \right) \right) = o_{\mathbf{P}}(1).$$

This completes the proof of Theorem 2.2.  $\square$

In order to prove Theorem 5.1, we need two technical lemma's.

**Lemma A.1** *Let  $Y_1, \dots, Y_n$  be independent random variables with common distribution function  $G(y) = 1 - 1/y$ ,  $y \geq 1$ . Let  $Y_{1:n}, \dots, Y_{n:n}$  denote the order statistics of  $Y_1, \dots, Y_n$ . Let  $(k_n)$  be a sequence of integers with  $k_n \leq n$  and  $k_n \rightarrow \infty$ ,  $n \rightarrow \infty$ . Then, as  $n \rightarrow \infty$  and for all  $\beta < 1$ ,*

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left( \frac{Y_{n-i+1:n}}{Y_{n-k_n:n}} \right)^\beta \xrightarrow{\mathbf{P}} \frac{1}{1-\beta},$$

and

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{Y_{n-i+1:n}}{Y_{n-k_n:n}} \xrightarrow{\mathbf{P}} 1.$$

PROOF: The first result is Lemma 2.4 of Dekkers et al. (1989). The second result follows easily from the law of large numbers upon noting that  $(\log [Y_{n-i+1:n}/Y_{n-k_n:n}])_{i=1}^{k_n}$  is distributed as the order statistics of a standard exponential sample of size  $k_n$ . Hence, the result follows from the consistency of the Hill estimator for the strict Pareto case.  $\square$

The second lemma we need can be found in Smith (1982).

**Lemma A.2** *Suppose  $L(\cdot)$  satisfies the SR2-condition with  $\rho \leq 0$ . If  $\rho < 0$ , then for all  $\varepsilon > 0$  there exists a  $t_\varepsilon$  such that we have*

$$\left| \log \frac{L(tx)}{L(t)} - cg(t) \int_1^x u^{\rho-1} du \right| \leq \varepsilon g(t), \quad (\text{A.4})$$

whenever  $t \geq t_\varepsilon$  and  $x > 1$ . If  $\rho = 0$ , then the same result holds with the right-hand side replaced by  $\varepsilon g(t)x^\varepsilon$ .

We now may prove Theorem 5.1.

PROOF OF THEOREM 5.1: We first consider the behavior of the Hill estimator under the null hypothesis  $F_0$ . In the literature, many proofs exist of the asymptotic behavior of the Hill estimator under the null. We present the proof for completeness only. In virtue of Theorem 2.2 and using the quantile transformation, we need to prove that

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log \frac{l(F^{-1}(U_{n-i+1:n}))}{l(F^{-1}(U_{n-k_n:n}))} \quad (\text{A.5})$$

tends to  $cA/(\rho - 1)$  in probability, where  $U_{1:n}, \dots, U_{n:n}$  denote the order statistics of a uniform sample of size  $n$ . Now

$$\begin{aligned} 1-t &= 1 - F(F^{-1}(t)) \\ &= \left[ F^{-1}(t)l(F^{-1}(t)) \right]^{-1/\gamma} \\ &= \left[ (1/(1-t))^\gamma L(1/(1-t))l(F^{-1}(t)) \right]^{-1/\gamma}, \end{aligned}$$

implies  $l(F^{-1}(t)) = 1/L(1/(1-t))$ . Since  $(1/(1-U_{i:n}))_{i=1}^n \stackrel{d}{=} (Y_{i:n})_{i=1}^n$ , we have

$$(A.5) \stackrel{d}{=} -\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log \frac{L(Y_{n-i+1:n})}{L(Y_{n-k_n:n})}.$$

Moreover,  $Y_{n-k_n:n} = (n/k_n)(1 + o_{\mathbf{P}}(1))$  and, since  $g$  is regularly varying, this implies

$$g(Y_{n-k_n:n})/g(n/k_n) = 1 + o_{\mathbf{P}}(1).$$

Thus, using Condition (5.3),

$$\sqrt{k_n}g(Y_{n-k_n:n}) = A + o_{\mathbf{P}}(1). \quad (A.6)$$

Provided that

$$\frac{-1}{k_n g(Y_{n-k_n:n})} \sum_{i=1}^{k_n} \log \frac{L(Y_{n-i+1:n})}{L(Y_{n-k_n:n})} \quad (A.7)$$

tends to  $c/(\rho-1)$  in probability, the desired result follows.

In case  $\rho < 0$ , since  $\epsilon$  is arbitrary in (A.4), it follows that

$$(A.7) = -\frac{1}{k_n} \sum_{i=1}^{k_n} c \int_1^{Y_{n-i+1:n}/Y_{n-k_n:n}} u^{\rho-1} du + o_{\mathbf{P}}(1).$$

Applying Lemma A.1 for  $\rho < 0$ , we indeed find

$$(A.7) = c/(\rho-1) + o_{\mathbf{P}}(1).$$

The case  $\rho = 0$  follows similarly using again Lemma A.1 and noting that the extra factor  $x^\epsilon$  in the right-hand side of (A.4) doesn't affect the conclusion.

The behavior of the Hill estimator under the local alternatives (2.2) and (2.3) now follows immediately from Le Cam's third lemma (see, e.g., Bickel et al. (1993), p. 503).  $\square$

Before proving Theorem 6.2, we first establish the following lemma.

**Lemma A.3** *Let  $0 < k_n \leq n$  with  $k_n \rightarrow \infty$  and  $\sqrt{k_n} = o(\log(n/k_n))$ . Let  $\omega_{1:n}, \dots, \omega_{n:n}$  be order statistics of a sample of size  $n$  from the standard exponential distribution. Then, for  $m \in \mathbf{N}$ ,*

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( \frac{\omega_{n-i+1:n}}{\omega_{n-k_n:n}} - 1 \right)^m = o_{\mathbf{P}}(1/(\log(n/k_n))^{m-1})$$

PROOF: Note that  $(\omega_{n-i+1:n} - \omega_{n-k_n:n}, i = 1, \dots, k_n)$  are distributed as the order statistics of a standard exponential sample of size  $k_n$ . Hence, from the law of large numbers we get

$$\frac{1}{k_n} \sum_{i=1}^{k_n} (\omega_{n-i+1:n} - \omega_{n-k_n:n})^m \xrightarrow{\mathbf{P}} m!.$$

Now,

$$\begin{aligned}
& (\log(n/k_n))^{m-1} \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( \frac{\omega_{n-i+1:n}}{\omega_{n-k_n:n}} - 1 \right)^m \\
&= \frac{\sqrt{k_n}}{\log(n/k_n)} \left( \frac{\log(n/k_n)}{\omega_{n-k_n:n}} \right)^m \frac{1}{k_n} \sum_{i=1}^{k_n} (\omega_{n-i+1:n} - \omega_{n-k_n:n})^m,
\end{aligned}$$

and the desired result follows from

$$\omega_{n-k_n:n} = \log(n/k_n) + o_{\mathbf{P}}(1).$$

□

PROOF OF THEOREM 6.2: The proof will follow the same lines as that of Theorem 2.2. Using the quantile transformation, we obtain

$$\begin{aligned}
\tau_0^2 \Delta^{(n)} &= \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( -\log \frac{1 - F_0(X_{n-i+1:n})}{1 - F_0(X_{n-k:n})} - 1 \right) \\
&\stackrel{d}{=} \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^n (-\log[1 - U_i] + \log[k_n/n] - 1) I\{U_i > 1 - k_n/n\} \\
&\quad - \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (-\log[1 - U_{n-i+1:n}] + \log[1 - U_{n-k:n}] - 1).
\end{aligned}$$

We decompose this expression into three terms

$$\begin{aligned}
T_1^{(n)} &= \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^n (-\log[1 - U_i] + \log[k_n/n]) I\{U_i > 1 - k_n/n\} \\
&\quad - \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (-\log[1 - U_{n-i+1:n}] + \log[k_n/n]), \\
T_2^{(n)} &= \frac{\tau_0}{\sqrt{k_n}} \left( k_n - \sum_{i=1}^n I\{U_i > 1 - k_n/n\} \right), \\
T_3^{(n)} &= -\tau_0 \sqrt{k_n} (\log[1 - U_{n-k_n:n}] - \log[k_n/n]).
\end{aligned}$$

The terms  $T_2^{(n)}$  and  $T_3^{(n)}$  are equal to the terms appearing in the proof of Theorem 2.2. The term  $T_1^{(n)}$  is somewhat different, but can be handled analogously. More precisely, for any  $d \in (0, \infty)$ , we define

$$T_1^{(n)}(d) = \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^n |\log[n/k_n(1 - U_i)]| I\left\{1 - k_n/n - d\sqrt{k_n}/n \leq U_i \leq 1 - k_n/n + d\sqrt{k_n}/n\right\}$$

Since for each  $d \in (0, \infty)$ ,

$$T_1^{(n)}(d) = o_{\mathbf{P}}(1),$$

we get

$$T_1^{(n)} = o_{\mathbf{P}}(1).$$

Furthermore,

$$\begin{aligned} & \tau_0^2 \Delta^{(n)} - \sqrt{k_n}(\hat{\tau}_{k_n}^{(n)} - \tau_0) - \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log(n/k_n) \frac{X_{n-i+1:n}}{X_{n-k_n:n}} \left( \frac{l_0(X_{n-i+1:n})}{l_0(X_{n-k_n:n})} - 1 \right) \\ &= \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (-\log[1 - F_0(X_{n-i+1:n})] + \log[1 - F_0(X_{n-k_n:n})] - 1) - \sqrt{k_n}(\hat{\tau}_{k_n}^{(n)} - \tau_0) \\ & \quad - \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log(n/k_n) \frac{X_{n-i+1:n}}{X_{n-k_n:n}} \left( \frac{l_0(X_{n-i+1:n})}{l_0(X_{n-k_n:n})} - 1 \right) + o_{\mathbf{P}}(1) \end{aligned}$$

which is distributed as

$$\begin{aligned} & \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (\omega_{n-i+1:n} - \omega_{n-k_n:n} - 1) - \sqrt{k_n} \left( \log(n/k_n) \frac{1}{k_n} \sum_{i=1}^{k_n} \left( \frac{K_0(\omega_{n-i+1:n})}{K_0(\omega_{n-k_n:n})} - 1 \right) - \tau_0 \right) \\ & \quad - \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log(n/k_n) \frac{K_0(\omega_{n-i+1:n})}{K_0(\omega_{n-k_n:n})} \left( \frac{l_0(K_0(\omega_{n-i+1:n}))}{l_0(K_0(\omega_{n-k_n:n}))} - 1 \right) \\ &= \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (\omega_{n-i+1:n} - \omega_{n-k_n:n}) - \log(n/k_n) \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( \frac{K_0(\omega_{n-i+1:n}) l_0(K_0(\omega_{n-i+1:n}))}{K_0(\omega_{n-k_n:n}) l_0(K_0(\omega_{n-k_n:n}))} - 1 \right). \end{aligned}$$

Since  $l_0(K_0(t)) = 1/L_0(t)$ , this expression can be simplified into

$$\frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (\omega_{n-i+1:n} - \omega_{n-k_n:n}) - \log(n/k_n) \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( \left( \frac{\omega_{n-i+1:n}}{\omega_{n-k_n:n}} \right)^{\tau_0} - 1 \right). \quad (\text{A.8})$$

Applying a Taylor expansion of  $t^\tau - 1$ , for  $t > 1$  and around 1 of order  $\max(\lfloor \tau \rfloor, 1)$  and using Lemma A.3, we get

$$\begin{aligned} (\text{A.8}) &= \frac{\tau_0}{\sqrt{k_n}} \sum_{i=1}^{k_n} (\omega_{n-i+1:n} - \omega_{n-k_n:n}) - \tau_0 \log(n/k_n) \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( \frac{\omega_{n-i+1:n}}{\omega_{n-k_n:n}} - 1 \right) \\ &= \tau_0 (\omega_{n-k_n:n} - \log(n/k_n)) \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left( \frac{\omega_{n-i+1:n}}{\omega_{n-k_n:n}} - 1 \right) \\ &= o_{\mathbf{P}}(1). \end{aligned}$$

□

PROOF OF THEOREM 7.1: Again, we start by considering the asymptotic behavior of  $\hat{\tau}_n$  under the null. Under  $F_0$ , we need to establish that

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log[n/k_n] \frac{X_{n-i+1:n}}{X_{n-k_n:n}} \left( \frac{l_0(X_{n-i+1:n})}{l_0(X_{n-k_n:n})} - 1 \right)$$

converges to  $-cA$ , in probability. As before, we use the quantile transformation. Let  $\omega_{1:n}, \dots, \omega_{n:n}$  be the order statistics of a sample of size  $n$  from the standard exponential distribution. Now,

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log[n/k_n] \left( \frac{\omega_{n-i+1:n}}{\omega_{n-k_n:n}} \right)^\tau \left( 1 - \frac{L_0(\omega_{n-i+1:n})}{L_0(\omega_{n-k_n:n})} \right)$$

converges to  $-cA$ , in probability, in view of the results in the proof of Theorem 3.2(i) of Beirlant et al. (1995). An application of Le Cam's third lemma then completes the proof.  $\square$

### References

- BEIRLANT, J., M. BRONIATOWSKI, J.L. TEUGELS, and P. VYNCKIER (1995), The mean residual life function at great age: Applications to tail estimation, *Journal of Statistical Planning and Inference*, **45**, 21–48.
- BEIRLANT, J., J.L. TEUGELS, and P. VYNCKIER (1996), *Practical Analysis of Extreme Values*, Leuven University Press, Leuven.
- BICKEL, P.J., C.A.J. KLAASSEN, Y. RITOV, and J.A. WELLNER (1993), *Efficient and Adaptive Statistical Inference for Semiparametric Models*, John Hopkins University Press, Baltimore.
- CSÖRGÖ, M., S. CSÖRGÖ, L. HORVATH, and D.M. MASON (1986), Weighted empirical and quantile processes, *Annals of Probability*, **14**, 31–85.
- CSÖRGÖ, S. and L. VIHAROS (1998), Estimating the tail index, *Asymptotic Methods in Probability and Statistics*, 833–881, Elsevier.
- DE HAAN, L. and L. PENG (1998), Comparison of tail index estimators, *Statistica Neerlandica*, **52**, 60–70.
- DE HAAN, L. and S. RESNICK (1998), On asymptotic normality of the Hill estimator, *Stochastic Models*, **14**, 849–867.
- DEKKERS, A.L.M., J.H.J. EINMAHL, and L. DE HAAN (1989), A moment estimator for the index of an extreme-value distribution, *Annals of Statistics*, **17**, 1833–1855.
- DREES, H. (1998), Optimal rates of convergence for estimates of the extreme value index, *Annals of Statistics*, **26**, 434–448.
- DREES, H. (2001), Minimax Risk Bounds in Extreme Value Theory, *Annals of Statistics*, **29**, 266–294.
- FALK, M.(1995), On testing the extreme value index via the POT-method, *Annals of Statistics*, **23**, 2013–2025.
- HAEUSLER, E. and J.L. TEUGELS (1985) On asymptotic normality of Hill's estimator for the exponent of regular variation, *Annals of Statistics*, **13**, 743–756.
- HALL, P. (1982), On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society Series B*, **44**, 37–42.

- HALL, P. and A.H. WELSH, A.H. (1984), Best attainable rates of convergence for estimates of parameters of regular variation, *Annals of Statistics*, **12**, 1079-1084.
- HALL, P. and WELSH, A.H. (1985), Adaptive estimates of parameters of regular variation, *Annals of Statistics*, **13**, 331-341.
- KELLER, B. and C. KLÜPPELBERG (1991), Statistical estimation of large claim distributions, *Mitteilungen SVVM*, 203–216.
- KLÜPPELBERG, C. and J.A. VILLASEÑOR (1993), Estimation of distribution tails – a semiparametric approach, *Blätter der DGVM*, **21**, 213–235.
- MAROHN, F. (1997), Local asymptotic normality in extreme value index estimation, *Annals of the Institute of Statistical Mathematics*, **49**, 645–666.
- LE CAM, L. and G. YANG (1990), *Asymptotics in Statistics: Some Basic Concepts*, Springer-Verlag, New York.
- SMITH, R.L. (1982), Uniform rates of convergence in extreme-value theory, *Advances in Applied Probability*, **14**, 600-622.
- WEI, X. (1995), Asymptotically efficient estimation of the index of regular variation, *Annals of Statistics*, **23**, 2036–2058.