Semiparametric Lower Bounds for Tail Index Estimation

Beirlant, J.; Bouquiaux, C.; Werker, B.J.M.

Publication date:
2001

CentER

No. 2001-65

SEMIPARAMETRIC LOWER BOUNDS FOR TAIL INDEX ESTIMATION

By Jan Beirlant, Christel Bouquiaux and Bas J.M. Werker

September 2001

ISSN 0924-7815

Discussion paper

Semiparametric Lower Bounds for Tail Index Estimation

Jan Beirlant[*]
Christel Bouquiaux[†]
and Bas J.M. Werker[‡]

Katholieke Universiteit Leuven
Université Libre de Bruxelles
Tilburg University

September 20, 2001

Abstract

We consider semiparametric 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 infinite dimensional nuisance parameter. Without further regularity conditions, we derive a Local Asymptotic Normality (LAN) result that describes essentially the least favorable submodel for the tail index parameter. 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 (regular) semiparametric estimators through the convolution theorem. We show that the Hill estimator is also semiparametrically efficient in the Pareto case in this much stronger sense. We also discuss the Weibull model in this respect.

JEL-code: C14

[*]Department Wiskunde, Katholieke Universiteit Leuven, Celestijnenlaan 200 B, B-3001, Leuven, Belgium. E-mail: Jan.Beirlant@wis.kuleuven.ac.be

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

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

1 Introduction

Consider an i.i.d. sequence of random variables $X_1, \ldots, X_n$ 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

$$1 - F(x) = [xl(x)]^{-1/\gamma}, \quad x > 0, \tag{1.1}$$

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

$$-\log[1 - F(x)] = [xl(x)]^{-1/\tau}, \quad x > 0, \tag{1.2}$$

where $\tau > 0$ is called the Weibull tail index parameter and, as before, l is some slowly varying function.

We analyze Pareto and Weibull type models from a semiparametric point of view in which we take the tail index parameter (γ for the Pareto case and τ for the Weibull model) as the parameter of interest and l as the nuisance parameter. We are interested in *efficient* estimation of the tail index parameter. We derive lower bounds for the (asymptotic) precision with which the tail index parameter can be estimated. We consider not only the optimal rate of convergence for the tail index parameter, but also the optimal asymptotic variance of (regular) estimators in the Hájek-Le Cam convolution theorem sense. For the Pareto model, we show that the widely-used Hill estimator is efficient in this strong sense, but for a well documented bias. For the Weibull model, we assess the efficiency of an estimator proposed in Beirlant et al. (1995).

From a technical point of view, our results are based on the Local Asymptotic Normality condition and the Hájek-Le Cam Convolution Theorem. We essentially find the most difficult parametric submodel of the semiparametric models that we consider. The advantage of adopting these techniques is that they allow for the determination of the "minimal" asymptotic distribution of arbitrary regular estimators given a (possibly optimal) rate of convergence. "Minimal" should here be understood in a convolution sense. Consequently, we get more precise results on the optimal behavior of estimators, yielding bounds on the width of (asymptotic) confidence intervals. It is known that the Hill estimator attains the optimal rate of convergence for the model in Hall and Welsh (1984) [see, Hall and Welsh (1985)]. We use the convolution theorem to assess the semiparametric efficiency of the Hill estimator in Section 4. Using a new type of proof for the asymptotic behavior of the Hill estimator, this furthermore allows to obtain the behavior of the Hill estimator under the local alternatives that we consider. Finally, we obtain similar results for the Weibull model.

In the current paper, we only consider $\gamma > 0$. 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 three 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 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.

Our second contribution is that we *extend* the efficiency results beyond that of rate efficiency. Given the optimal rate of convergence, one may wonder what the minimal limiting variance is of estimators attaining this rate, i.e. a Cramer-Rao type bound. We show that, for Pareto type distributions, the Hill estimator not only attains the optimal rate of convergence, but is also essentially efficient given this rate. That is, there do not exist (regular) estimators that attain a smaller limiting variance than the Hill estimator. From this we may also conclude that the local alternatives that we construct describe the least favorable parametric submodel of the full semiparametric model.

Our third contribution is that 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 (including local alternatives for the slowly varying nuisance function) and consider efficient estimation of the tail-index parameter. Again, we obtain essentially the least favorable parametric submodel. We show that, under some conditions, an estimator provided in Beirlant et al. (1995) is semiparametrically efficient.

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 γ real). We unify the aforementioned results for the positive γ 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 *deduce* that optimal inference can be based on the largest values observed, since it is only these observations that appear in the central sequence of the least favorable parametric submodel. Another difference with Wei (1995) and Marohn (1997) is that we consider a fully semiparametric model. 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 efficiency in the convolution theorem sense. Compared to minimax results, our results are stronger, but only apply to so-called regular estimators.

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 (regular) estimators. 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 is (but for a well-known asymptotic bias) efficient, both with respect to the rate of convergence as with respect to the asymptotic variance. In Section 6, we prove a similar LAN result for the Weibull model and in Section 7 we consider the efficiency of the estimator proposed in Beirlant et al. (1995). 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 γ_0 and l_0, i.e.

$$1 - F_0(x) = [x l_0(x)]^{-1/\gamma_0}, \quad x > 0. \tag{2.1}$$

As mentioned in the introduction, in this paper we take a semiparametric point of view in this paper and are interested in the estimation of the Pareto tail index γ_0, while considering the slowly varying function l_0 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 optimality of these estimators in terms of the asymptotic variance given this optimal rate of convergence. 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 (δ_n) with $\delta_n \to 0$ and $\sqrt{n}\delta_n \to \infty$, $n \to \infty$. As long as no further assumptions (like those discussed in Section 4) are made on the set of Pareto-type distributions, the sequence (δ_n) is arbitrary.

The LAN condition effectively gives the likelihood ratio for a model that contains a parameter $u \in \mathbb{R}$ that is used to localize the parameter of interest γ_0. More precisely, for every $u \in \mathbb{R}$, we define

$$\gamma_n = \gamma_0 + u\delta_n. \tag{2.2}$$

We also define local alternatives for the nuisance function l_0 as follows

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

where $t_n = U_0(n\delta_n^2)$ with $U_0(t) = F_0^{-1}(1 - 1/t)$. Since, F_0 is continuous, we have

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

Remark 2.1 The alternatives constructed through (2.2) and (2.3) will turn out later to be optimal in the sense that they essentially define the "least favorable parametric submodel" for various more specific models. They turn out to be least favorable as far as the rate of convergence is concerned, but also, up to a deterministic bias, with respect to the limiting variance of estimators. Details are discussed in Section 4.

The distribution function corresponding to γ_n and l_n is given by

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

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

$$\frac{\mathrm{d}F_n}{\mathrm{d}F_0}(x) = \begin{cases} 1, & x \le t_n \\ \frac{\gamma_0}{\gamma_n}\left[\frac{1-F_0(x)}{1-F_0(t_n)}\right]^{\gamma_0/\gamma_n - 1}, & x > t_n \end{cases}. \tag{2.6}$$

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 $X_1, \ldots, X_n$. 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*

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

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

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

where

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

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

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

Lemma 2.2 *Given F_0 that satisfies (2.1), we have, for all $k \in \mathbb{N}$,*

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

Proof: Using the transformation $v = (1 - F_0(x))/(1 - F_0(t_n))$ the integral is reduced to a Gamma integral. 2

Proof of Theorem 2.1: Since $n\delta_n^2[1 - F_0(t_n)] = 1$, we find, under F_0,

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

and also, using Lemma 2.2 with $k = 1$ and $k = 2$,

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

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

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

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

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

For fixed n, the ξ_{ni} are independent random variables. From Lemma 2.2, the definition of t_n in (2.4), and $|1+a|^3\le 1-3a+3a^2-a^3$ for $a\le 0$, we get

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

Since, for $n\to\infty$,

$$\frac{\sum_{i=1}^n \mathrm{E}|\xi_{ni}|^3}{(\sum_{i=1}^n \mathrm{Var}\,\xi_{ni})^{3/2}}\le 16\delta_n\to 0,$$

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. 2

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 efficiency 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 $k_n\to\infty$, as

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

Theorem 2.3 *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, we have, under F_0,*

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

or, equivalently,

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

The proof being more technical, it is left for the appendix. Relation (2.11) is essential in assessing the semiparametric efficiency of the Hill estimator. The right-hand side of (2.11) measures the inefficiency of the Hill estimator. We will study this inefficiency in more detail in Section 5.

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

A LAN condition as in Theorem 2.1 or 2.3 allows for the derivation of bounds on the optimal rate of convergence of "reasonable" estimators for the tail-index parameter γ. 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 under the present general framework. But the LAN condition allows for much more precise lower bounds on the asymptotic behavior of regular (in a sense to be made precise later) estimators than the *rate* of convergence alone. Through the so-called convolution theorem, we also obtain lower bounds for the asymptotic distribution of 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.

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

Theorem 3.2 *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 γ for which*

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

then

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

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 Corollary 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. 2

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 (δ_n). Thus, given a possible sequence (α_n), one can always find a sequence (δ_n), converging to zero very slowly, such that (3.2) does not hold, i.e., such that $\limsup_{n\to\infty} \alpha_n \delta_n = \infty$. This implies that there cannot exist a uniformly consistent estimator of γ in the full Pareto model, no matter how weak the consistency requirement in (3.1), i.e., no matter how slowly (α_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 3.2 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. If meaningful optimal rates of convergence are to be found, we must restrict the model by imposing extra regularity on the slowly varying function l in (2.1). This will be consider 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 3.3 *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 2.3). Suppose, moreover, that $\hat{\gamma}_n$ is a regular estimator for γ in the sense, for $n \to \infty$,*

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

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

$$\begin{pmatrix} \gamma_0^2 \Delta^{(n)} \\ \delta_n^{-1}(\hat{\gamma}_n - \gamma_0) - \gamma_0^2 \Delta^{(n)} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} V \\ Z \end{pmatrix}, \tag{3.3}$$

where $V \sim N(0, \gamma_0^2)$ and Z are independently distributed.

The convolution theorem states that, given the rate of convergence δ_n^{-1}, the most concentrated limiting distribution possible for estimating γ (and, hence, the "best" limiting distribution possible), is a $N(0, \gamma_0^2)$ distribution. All 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 larger confidence bands. This provides a much stronger efficiency concept than the rate of convergence alone. In Section 5, we will show that the Hill estimator is also efficient (up to a *deterministic* bias) in the convolution sense, for appropriately defined submodels à la Hall and Welsh (1984) or Drees (1998). Section 7 shows the efficiency of 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. 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

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

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 $|\gamma - \gamma_0| \le \varepsilon$, $|C - C_0| \le \varepsilon$, and

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

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

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

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

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

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

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 (if $n \to \infty$) if and only if $t_n^{\rho/\gamma_n}\delta_n = O(1)$, i.e., if and only if

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

From Theorem 3.2 we now obtain that $\alpha_n(\hat{\gamma}_n - \gamma) = O_{\mathbf{P}}(1)$ uniformly over the Hall and Welsh (1984) model implies

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

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 is normalized, i.e., for some $\eta : [1, \infty) \to \mathbb{R}$,

$$l(x) = C \exp\left(\int_1^x \eta(z)/z \mathrm{d}z\right). \tag{4.6}$$

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

$$\sup_z |\eta(z)|/h(z) \tag{4.7}$$

is bounded over $\mathcal{P}$ for some given 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 η_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

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

Assuming that h is decreasing, we find

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

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

$$\delta_n/h(t_n) = O(1), \quad n \to \infty. \tag{4.9}$$

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.,

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

The present example is in fact a slight variation of the Drees (1998) model. Drees (1998) imposes the conditions (4.6) and (4.7) on the slowly varying part of the quantile function corresponding to the distributions in $\mathcal{P}$. Clearly, this does not affect the optimal rates of convergence. It is possible to consider exactly Drees' (1998) model in our framework by considering slightly more general alternatives than those presented in (2.2) and (2.3) with $1 - F_0(t_n) = 1/(n\delta_n^2)$. More precisely, t_n should be chosen such that $n\delta_n^2(1 - F_0(t_n)) = 1/k$ with $k > 0$ arbitrary.

Note that Drees (1998) considers the non-Pareto case, i.e., where the tail-index γ may be zero or negative. This is a non-trivial extension that is not covered by our present results.

5 Efficiency of the Hill estimator

Section 2 provides a general LAN result for Pareto type models. In the previous section, we have seen how this general result immediately yields the optimal rates of convergence in more specific Pareto type models, like those of Hall and Welsh (1984) and Drees (1998). Furthermore, the LAN result, via the convolution theorem, gives a lower bound on the asymptotic variance of regular estimators. In this section, we show that the Hill estimator, under a regularity condition and apart from a well-known asymptotic bias, attains this lower bound. This proves that the Hill estimator is efficient in the sense of the convolution theorem and also proves that the local alternatives that we introduced essentially describe the least favorable parametric submodel for estimating the Pareto tail index.

Consider a Pareto type distribution F of the form (1.1). We may decompose the inverse of $1/(1-F)$ as follows:

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

with L slowly varying at infinity. In order to derive the asymptotic distribution 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 $g : (0,\infty) \to (0,\infty)$ a ρ-varying function with $\rho \leq 0$. Consider the following asymptotic condition

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

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, the Hill-estimator is essentially efficient (but for a deterministic bias), i.e. the right-hand side of (2.11) is asymptotically constant.

Theorem 5.1 *Suppose that F is of the Pareto type (1.1) and satisfies the SR2-condition. Fix a sequence of integers (k_n) with $k_n \to \infty$, $k_n/n \to 0$, and*

$$\sqrt{k_n}g(n/k_n) \to A, \tag{5.3}$$

for some $A \in \mathbb{R}$. Then, under the local alternatives defined by $\gamma_n = \gamma_0 + u\delta_n$ together with (2.3), we have

$$\sqrt{k_n}(H_{k_n}^{(n)} - \gamma) \xrightarrow{\mathcal{L}} N(cA/(1-\rho) + u, \gamma^2). \tag{5.4}$$

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., Csörgo and Viharos, 1998). However, our contribution is to relate this result to the efficiency of the Hill estimator and that we get its behavior under our local alternatives as well. We provide a proof in the appendix that is effectively based on our Theorem 2.3.

The SR2 condition is sufficient to show that the Hill estimator is efficient, given its rate of convergence, in the convolution theorem sense apart from a deterministic bias. 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 satisfies the SR2-condition, say

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

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

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

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))

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

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

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

6 Local Asymptotic Normality of the Weibull Model

The Pareto model, while popular in practice, is not always the best choice in some applications. Robert (2001), for instance, shows that the stationary distribution of the alpha-ARCH model, used for modeling financial returns, is of the Weibull type (1.2). See also Chapter 4 in the Beirlant, Teugels, Vynckier (1996) monograph.

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

$$-\log[1-F_0(x)] = [x l_0(x)]^{1/\tau_0}, \quad x > 0. \tag{6.1}$$

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

$$\tau_n = \tau_0 + u\delta_n, \tag{6.2}$$

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

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 $x \le t_n$ and for $x > t_n$ we have

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

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

$$\Lambda^{(n)} = \Lambda^{(n)}(X_1, ..., X_n) = \sum_{i=1}^{n} \log \left[\frac{\mathrm{d}F_n}{\mathrm{d}F_0}(X_i)\right].$$

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

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

where

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

The proof of this LAN result is similar to that for the Pareto case. We first prove the equivalent of Lemma 2.2.

Lemma 6.2 *For F_0 defined in (6.1), we have, for all $k \in \mathbb{N}$,*

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

Proof: Using the transformation $v = -\log(1-F_0(x)) + \log(1-F_0(t_n))$, one is again lead to a Gamma integral. 2

Proof of Theorem 6.1 Applying Lemma 6.2 with $k=1$ and $k=2$ gives

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

This implies

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

Moreover, combining the inequality

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

with Lemma 6.2 for $k = 2$ and $k = 3$ gives

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

This last expression can be written as the sum of

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

and

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

of which the first part vanishes asymptotically in view of Lemma 6.2.

The above results imply that we may write

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

Note

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

and, in virtue of Lemma 6.2,

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

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

Let

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

The limiting distribution of the central sequence $\Delta^{(n)}$ follows from the Liapunov Central Limit Theorem, using Lemma 6.2 to obtain

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

2

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 6.3 *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 Efficient 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 slowly varying.

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

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

A Some proofs

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

Proof of Theorem 2.3: Let $U_{1:n} \le \ldots \le U_{n:n}$ be the order statistics of n i.i.d. uniformly over the interval $[0,1]$ distributed r.v.'s $U_1, \ldots, U_n$. Using the quantile transformation, we obtain

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

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,

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

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

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

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

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

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

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

Hence, in order to prove

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

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

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

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

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

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 θ_n between k_n/n and $1 - U_{n-k_n:n}$ and using (A.1),

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

To complete the proof, we define the uniform empirical process

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

and the uniform quantile process

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

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 Csörgo 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 2.3. 2

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

Lemma A.1 *Let $Y_1, \ldots, Y_n$ be independent random variables with common distribution function $G(y) = 1 - 1/y$, $y \geq 1$. Let $Y_{1:n}, \ldots, Y_{n:n}$ denote the order statistics of $Y_1, \ldots, Y_n$. Let (k_n) be a sequence of integers with $k_n \leq n$ and $k_n \to \infty$, $n \to \infty$. Then, as $n \to \infty$ 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. 2

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

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

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

whenever $t \ge t_\varepsilon$ 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 2.3 and using the quantile transformation, we need to prove that

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

tends to $cA/(\rho-1)$ in probability, where $U_{1:n},\ldots,U_{n:n}$ 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),

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

Provided that

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

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

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

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

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^{ε} 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). 2

Before proving Theorem 6.3, we first establish the following lemma.

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

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

PROOF: Note that $(\omega_{n-i+1:n} - \omega_{n-k_n:n},\ i = 1, \ldots, k_n)$ 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

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

2

Proof of Theorem 6.3: The proof will follow the same lines as that of Theorem 2.3. Using the quantile transformation, we obtain

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

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 2.3. The term $T_1^{(n)}$ is somewhat different, but can be handled analogously. More precisely, for any $d\in(0,\infty)$, we define

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

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,

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

which is distributed as

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

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

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

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

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

2

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 $\omega_{1:n},\ldots,\omega_{n:n}$ 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. 2

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.

Csörgo, M., S. Csörgo, L. Horvath, and D.M. Mason (1986), Weighted empirical and quantile processes, *Annals of Probability*, **14**, 31–85.

Csörgo, S. and L. Viharos (1998), Estimating the tail index, *Asymptotic Methods in Probability and Statistics*, 833–881, Elsevier.

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*, forthcoming.

Falk, M.(1995), On testing the extreme value index via the POT-method, *Annals of Statistics*, **23**, 2013–2025.

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.

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.

Robert, C. (2001), Extremes of alpha-ARCH models, *Lecture Notes in Statistics*, **147**, 223–257, Springer, 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.