*The Annals of Statistics*
2003, Vol. 31, No. 6, 1956–1995


# KERNEL-TYPE ESTIMATORS FOR THE EXTREME VALUE INDEX

BY P. GROENEBOOM, H. P. LOPUHAÄ AND P. P. DE WOLF

*Delft University of Technology, Delft University of Technology and Statistics Netherlands*

A large part of the theory of extreme value index estimation is developed for positive extreme value indices. The best-known estimator of a positive extreme value index is probably the Hill estimator. This estimator belongs to the category of moment estimators, but can also be interpreted as a quasi-maximum likelihood estimator. It has been generalized to a kernel-type estimator, but this kernel-type estimator can, similarly to the Hill estimator, only be used for the estimation of positive extreme value indices. In the present paper, we introduce kernel-type estimators which can be used for estimating the extreme value index over the whole (positive and negative) range. We present a number of results on their distributional behavior and compare their performance with the performance of other estimators, such as moment-type estimators for the whole range and the quasi-maximum likelihood estimator, based on the generalized Pareto distribution. We also discuss an automatic bandwidth selection method and introduce a kernel estimator for a second-order parameter, controlling the speed of convergence.

**1. Introduction.** Let $X_1, \dots, X_n$ denote a sample from a distribution function $F$, which is assumed to be in the domain of attraction of an extreme value distribution with extreme value index $\gamma$, denoted by $F \in \mathcal{D}(G_\gamma)$. In the situation of estimating a positive extreme value index, one of the best-known estimators is the Hill estimator [Hill (1975)]. This estimator is consistent for all $\gamma > 0$, assuming only $F \in \mathcal{D}(G_\gamma)$. In the case that the tail of the underlying distribution function is Pareto shaped, that is, $1 - F(x) = Cx^{-1/\gamma}$ for all $x \geq u$ with $\gamma > 0$, $C > 0$ and $u > 0$, the Hill estimator can be interpreted as a maximum likelihood estimator. This "quasi" likelihood approach was extended in Smith (1987), where a generalized Pareto distribution was assumed to hold for the tail of the underlying distribution function. The resulting estimator is consistent for $\gamma > -1$. Pickands (1975) proposed an estimator that is invariant under shift and scale transformations and that is consistent for all $\gamma \in \mathbb{R}$. However, it has poor efficiency.

Dekkers, Einmahl and de Haan (1989) extended the Hill estimator to an estimator that is consistent for all $\gamma \in \mathbb{R}$. The resulting estimator, also called the moment estimator, consists of two terms. The first term is the Hill estimator, which converges to $\gamma \vee 0$. In order to have a consistent estimator for $\gamma < 0$, a second term was added that converges to $\gamma \wedge 0$. More recently, Beirlant, Vynckier and Teugels (1996) proposed an adaptive Hill estimator, which is also consistent for all $\gamma \in \mathbb{R}$.

Received April 2001; revised November 2002.
*AMS 2000 subject classifications.* Primary 60G70, 62E20, 62G05; secondary 62G20, 62G30.
*Key words and phrases.* Extreme value index, adaptive estimation, second order parameter estimation.

Drees (1995) investigated a multistage procedure that results in a refinement of the Pickands estimator, which is consistent for all $\gamma \in \mathbb{R}$ and improves the efficiency.

All the estimators mentioned above are based on the $k$ largest observations. A major drawback of the estimators is the discrete character of the behavior of these estimators: adding a single large-order statistic in the calculation of the estimator, that is, increasing $k$ by 1, can change the actual value of the estimate considerably. Plotting these estimators as a function of the order statistics used therefore often results in a zig-zag figure. In Csörgő, Deheuvels and Mason (1985), the Hill estimator is smoothed by a kernel. We call this estimator the CDM estimator. Incidentally, Hill's estimator reappears when substituting the uniform kernel in the CDM estimator. In the same paper, it was shown that it is possible to improve on the (asymptotic) variance of the estimator by choosing appropriate kernels. In this kernel-type estimator, the bandwidth $h$ plays a similar role as the number of order statistics $k$ in the aforementioned estimators: approximately $nh$ order statistics will be used to calculate the estimate. Consequently, the estimator now depends in a continuous way on the fraction of order statistics used. Hence, plotting the estimator as a function of the bandwidth $h$ then yields a smooth figure. Other attempts to construct smoothed versions of the Hill estimator can be found in Schultze and Steinebach (1996), Kratz and Resnick (1996) and Csörgő and Viharos (1997), which consider classical least squares estimators for the slope $\gamma > 0$ in a Pareto quantile plot.

Unfortunately, the least squares estimators and the CDM kernel estimator are only valid for $\gamma > 0$. In the present paper, we introduce a new class of kernel-type estimators that is consistent for all $\gamma \in \mathbb{R}$. It should be emphasized that our estimator is not a smoothed version of the moment estimator, but is based on the von Mises conditions

$$\lim_{t \uparrow x_F^\circ} \left( \frac{d}{dt} \frac{1 - F(t)}{F'(t)} \right) = \gamma, \tag{1.1}$$

where $x_F^\circ = \sup\{x : F(x) < 1\} \leq \infty$ is the upper endpoint of $F$. These conditions are sufficient but not necessary for $F \in \mathcal{D}(G_\gamma)$. Although this approach is different from the one that leads to the moment estimator, it will result in an estimator that also consists of two terms. We define the following estimator for $\gamma \in \mathbb{R}$:

$$\hat{\gamma}_{n,h}^{K} = \hat{\gamma}_{n,h}^{(\mathrm{pos})} - 1 + \frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}}, \tag{1.2}$$

where

$$\hat{\gamma}_{n,h}^{(\mathrm{pos})} = \sum_{i=1}^{n-1} \frac{i}{n} K_h\left(\frac{i}{n}\right) \left(\log X_{(n-i+1)} - \log X_{(n-i)}\right),$$

$$\hat{q}_{n,h}^{(1)} = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{\alpha} K_h\left(\frac{i}{n}\right) \left(\log X_{(n-i+1)} - \log X_{(n-i)}\right),$$

$$\hat{q}_{n,h}^{(2)} = \sum_{i=1}^{n-1} \frac{d}{du}[u^{\alpha+1}K_h(u)]_{u=i/n}\big(\log X_{(n-i+1)} - \log X_{(n-i)}\big),$$

with $K_h(u) = K(u/h)/h$ and $\alpha > 0$. Note that all three quantities have an integral representation involving the empirical quantile function. This is explained in Section 2 [see (2.4), (2.6) and (2.8)], where we also give the motivation for this estimator and specify the conditions for the kernel $K$. An example of a suitable kernel is the biweight $K(x) = \frac{15}{8}(1-x^2)^2$. The parameter $\alpha$ is needed to prevent singularities near 0 and must be greater than $1/2$ in order to have asymptotic normality. In our simulations, we took $\alpha = 0.6$. The first term in (1.2) is the kernel-type estimator of Csörgő, Deheuvels and Mason (1985) and is shown to converge to $\gamma \vee 0$. Similarly to the moment estimator, the second term will compensate the behavior of the CDM kernel-type estimator for $\gamma < 0$ and is shown to converge to $\gamma \wedge 0$. The resulting estimator will inherit the smooth behavior of the CDM kernel-type estimator as well as the general applicability of the moment estimator.

The content of the paper is organized as follows. In Section 2, we explain how estimator (1.2) is motivated by (1.1). In Section 3, consistency of the estimator will be derived under the single condition that the underlying distribution function is in the domain of attraction of an extreme value distribution. Under additional assumptions on the underlying distribution, asymptotic normality will be derived in Section 4, and sufficient conditions are provided in Section 5, under which the asymptotic bias vanishes. In Section 6, we compare our estimator with other estimators, such as the moment estimator and the maximum likelihood estimator, and the more recent proposals by Beirlant, Vynckier and Teugels (1996) and Drees (1995). Finally, in Section 7, we discuss automatic bandwidth selection methods, in the course of which we also introduce a kernel estimate for an important second-order parameter.

**2. Defining the estimator.** Let $X_1, \ldots, X_n$ denote a sample from a distribution function $F$, with support on $(0, \infty)$. Suppose that $F$ is in the domain of attraction of an extreme value distribution $G_\gamma$ for some $\gamma \in \mathbb{R}$, denoted by $F \in \mathcal{D}(G_\gamma)$; that is, there exist $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}$, with $a_n > 0$ and $b_n \in \mathbb{R}$, such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\big(-(1+\gamma x)^{-1/\gamma}\big)$$

for all $x$ with $1 + \gamma x > 0$. We will use the convention that $G_0(x) = \exp(-e^{-x})$ for $x \in \mathbb{R}$. Let $Q$ denote the quantile function corresponding to $F$. By replacing $t$ by $Q(1-s)$ in (1.1), the von Mises condition can be written as

$$\lim_{s\downarrow 0}\left(-1 - \frac{sF''(Q(1-s))}{(F'(Q(1-s)))^2}\right) = \gamma.$$

If $\log Q$ is well defined and differentiable, we can define the function $\phi$ by

$$\phi(s) = -s\frac{d}{ds}\log Q(1-s). \tag{2.1}$$

In this case, the limit relation (1.1) can be translated into

$$\lim_{s\downarrow 0}\left(-1+\phi(s)-\frac{s(d^2/ds^2)\log Q(1-s)}{(d/ds)\log Q(1-s)}\right)=\gamma. \tag{2.2}$$

The construction of our estimator is based on this relation. Basically, we have to estimate the value of $\phi$, the numerator and denominator in (2.2) at 0. To get some intuition on how to construct the estimator, it is useful to consider the generalized Pareto distribution (GPD). For the GPD, the function $\phi$ is given by

$$\phi_{\mathrm{GPD}}(s)=\begin{cases}\dfrac{\gamma}{1-s^{\gamma}}, & \gamma\neq 0,\\[2ex] \dfrac{1}{\log 1/s}, & \gamma=0.\end{cases}$$

Clearly, for the GPD one has that

$$\lim_{s\downarrow 0}\phi(s)=\gamma\vee 0. \tag{2.3}$$

Suppose for the moment that $\phi$ in (2.1) exists and also satisfies (2.3). Let the empirical quantile function be defined by $Q_n(u)=\inf\{x : F_n(x)\geq u\}$ and denote by $X_{(1)}\leq X_{(2)}\leq\cdots\leq X_{(n)}$ the order statistics corresponding to the sample $X_1, X_2, \ldots, X_n$. First, we estimate $\lim_{s\downarrow 0}\phi(s)$ by a kernel estimator

$$\begin{aligned}\hat{\gamma}_{n,h}^{(\mathrm{pos})}&=-\int_0^h uK_h(u)\,d\log Q_n(1-u)\\ &=\sum_{i=1}^{n-1}\frac{i}{n}K_h\left(\frac{i}{n}\right)\bigl(\log X_{(n-i+1)}-\log X_{(n-i)}\bigr),\end{aligned} \tag{2.4}$$

where $K_h(u)=K(u/h)/h$. Intuitively, using (2.3) and assuming that $K$ integrates to 1, for $h\downarrow 0$ this will behave as

$$-\int_0^h uK_h(u)\,d\log Q(1-u)=\int_0^1\phi(hu)K(u)\,du\to(\gamma\vee 0).$$

This is made rigorous for any $F\in\mathcal{D}(G_\gamma)$ in Lemma 3.3, without assuming the differentiability of $\log Q$. The numerator and the denominator on the left-hand side of (2.2) will be estimated separately at 0, using kernel-type estimators as well. In defining these estimators, we note that both numerator and denominator can be multiplied by any power of $s$, without changing the limit. Simulations show that this will lead to more stable estimators. For any $\alpha>0$, we have that

$$\lim_{s\downarrow 0}\frac{-s(d^2/ds^2)\log Q(1-s)}{(d/ds)\log Q(1-s)}=\lim_{s\downarrow 0}\frac{s^{\alpha+1}(d^2/ds^2)\log Q(1-s)}{-s^{\alpha}(d/ds)\log Q(1-s)}. \tag{2.5}$$

Note that, if (2.2) and (2.3) hold, the limit in (2.5) equals $1+(\gamma \wedge 0)$. For the denominator on the right-hand side of (2.5), we estimate $\lim_{s\downarrow 0} -s^{\alpha}(d/ds) \times \log Q(1-s)$ by a kernel estimator

$$
\begin{aligned}
\hat{q}_{n,h}^{(1)} &= -\int_0^h u^{\alpha} K_h(u)\, d\log Q_n(1-u) \\
&= \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{\alpha} K_h\left(\frac{i}{n}\right)\bigl(\log X_{(n-i+1)} - \log X_{(n-i)}\bigr).
\end{aligned}
\tag{2.6}
$$

If we were to treat the numerator on the right-hand side of (2.5) similarly and treat $Q_n$ as if it were differentiable, we could estimate $\lim_{s\downarrow 0} s^{\alpha+1}(d^2/ds^2) \times \log Q(1-s)$ by

$$
\int_0^h u^{\alpha+1} K_h(u) \frac{d^2}{du^2} \log Q_n(1-u). \tag{2.7}
$$

To overcome the difficulty that $Q_n$ is not differentiable, we use, as is customary in the literature on kernel estimation of derivatives of densities and regression functions, the derivative of the kernel instead of the derivative of a direct estimate of the unknown function. Hence, after using integration by parts in (2.7), we estimate $\lim_{s\downarrow 0} s^{\alpha+1}(d^2/ds^2)\log Q(1-s)$ by

$$
\begin{aligned}
\hat{q}_{n,h}^{(2)} &= -\int_0^h \frac{d}{du}[u^{\alpha+1} K_h(u)]\, d\log Q_n(1-u) \\
&= \sum_{i=1}^{n-1} \frac{d}{du}[u^{\alpha+1} K_h(u)]_{u=i/n}\bigl(\log X_{(n-i+1)} - \log X_{(n-i)}\bigr).
\end{aligned}
\tag{2.8}
$$

Intuitively, using (2.3), for $h \downarrow 0$, the term $\hat{q}_{n,h}^{(1)}$ as defined in (2.6) will behave as

$$
\begin{aligned}
-\int_0^h u^{\alpha} K_h(u)\, d\log Q(1-u) &= h^{\alpha-1}\int_0^1 \phi(hu) u^{\alpha-1} K(u)\, du \\
&\sim h^{\alpha-1}(\gamma \vee 0) \int_0^1 u^{\alpha-1} K(u)\, du.
\end{aligned}
$$

Similarly, $\hat{q}_{n,h}^{(2)}$ as defined in (2.8) will behave as

$$
\begin{aligned}
&-\int_0^h \frac{d}{du}[u^{\alpha+1} K_h(u)]\, d\log Q(1-u) \\
&\quad \sim h^{\alpha-1}\int_0^1 \phi(hu) u^{-1} \frac{d}{du}[u^{\alpha+1} K(u)]\, du \\
&\quad = h^{\alpha-1}(\gamma \vee 0)\int_0^1 u^{-1}\frac{d}{du}[u^{\alpha+1} K(u)]\, du \\
&\quad = h^{\alpha-1}(\gamma \vee 0)\int_0^1 u^{\alpha-1} K(u)\, du.
\end{aligned}
$$

In the case $\gamma > 0$, this would immediately suggest that $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)}$ tends to 1. Without assuming differentiability, it is shown in Lemma 3.4, for any $\gamma \in \mathbb{R}$ and for any $F \in \mathcal{D}(G_\gamma)$, that $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \to 1 + (\gamma \wedge 0)$.

The above discussion motivates the expression given in (1.2) as an estimator for $\gamma \in \mathbb{R}$. For the kernel $K$, we impose the following conditions. Let $K : [0, 1] \to \mathbb{R}^+$ be a fixed kernel function satisfying the following conditions:

(CK1) $K(x) = 0$, whenever $x \notin [0, 1)$ and $K(x) \geq 0$, whenever $x \in [0, 1)$;
(CK2) $K(1) = K'(1) = 0$;
(CK3) $\int_0^1 K(x)\,dx = 1$;
(CK4) $K$, $K'$ and $K''$ are bounded.

In the definition of $\hat{\gamma}_{n,h}^{K}$, the continuous parameter $h$ is used. This bandwidth determines the number of order statistics that is used in the computation of the estimator. The continuous nature of the bandwidth ensures that the estimator is a smooth function of the fraction of order statistics used, as opposed to the more discrete nature of, for example, the moment estimator.

## 3. Consistency.

By rearranging terms and using that $Q_n(1-u) = X_{n-k}$ for $k/n \leq u < (k+1)/n$, we can also write

$$\begin{aligned}
\hat{\gamma}_{n,h}^{(\mathrm{pos})} &= \int_0^1 \log Q_n(1-hu)\,d\big(uK(u)\big),\\
\hat{q}_{n,h}^{(1)} &= h^{\alpha-1}\int_0^1 \log Q_n(1-hu)\,d\big(u^{\alpha}K(u)\big),\\
\hat{q}_{n,h}^{(2)} &= h^{\alpha-1}\int_0^1 \log Q_n(1-hu)\,d\bigg(\frac{d}{du}[u^{1+\alpha}K(u)]\bigg).
\end{aligned}$$

Note that

$$Q_n(s) \stackrel{\mathcal{D}}{=} Q\big(\Gamma_n(s)\big) \quad \text{and} \quad \Gamma_n(1-s) \stackrel{\mathcal{D}}{=} 1 - \Gamma_n(s), \tag{3.1}$$

where $\Gamma_n$ is the empirical quantile function of a uniform (0, 1) sample $U_1, \dots, U_n$. Since conditions (CK2) and (CK4) yield that $\int d(uK(u)) = 0$, we have that

$$\hat{\gamma}_{n,h}^{(\mathrm{pos})} \stackrel{\mathcal{D}}{=} \int_0^1 \big(\log Q\big(1-\Gamma_n(hu)\big) - \log Q\big(1 - U_{(k+1)}\big)\big)\,d\big(uK(u)\big),$$

where $\Gamma_n$ is the empirical quantile function of a uniform (0, 1) sample $U_1, \dots, U_n$ and $k = \lfloor nh \rfloor$. To avoid differentiability of the quantile function, we use the following lemma.

LEMMA 3.1. *Suppose $F \in \mathcal{D}(G_\gamma)$ with $x_F^\circ > 0$. Denote the corresponding quantile function by $Q(s) = F^{-1}(s)$. Then, for some positive function $a(\cdot)$,*

$$\lim_{s\downarrow 0} \frac{\log Q(1-sy) - \log Q(1-s)}{a(s)/Q(1-s)} = \begin{cases} -\log y, & \gamma \geq 0, \\ \dfrac{y^{-\gamma}-1}{\gamma}, & \gamma < 0, \end{cases} \tag{3.2}$$

*for all $y > 0$. Moreover, for each $\varepsilon > 0$, there exists $s_0$ such that for $0 < s \leq s_0$ and $0 < y \leq 1$,*

$$\begin{aligned} (1-\varepsilon)\frac{1-y^\varepsilon}{\varepsilon} - \varepsilon &< \frac{\log Q(1-sy) - \log Q(1-s)}{a(s)/Q(1-s)} \\ &< (1+\varepsilon)\frac{y^{-\varepsilon}-1}{\varepsilon} + \varepsilon \end{aligned} \tag{3.3}$$

*provided $\gamma \geq 0$, and*

$$1-(1+\varepsilon)y^{-\gamma-\varepsilon} < \frac{\log Q(1-sy) - \log Q(1-s)}{\log Q(1) - \log Q(1-s)} < 1-(1-\varepsilon)y^{-\gamma+\varepsilon} \tag{3.4}$$

*provided $\gamma < 0$.*

PROOF. Rewrite Lemma 2.5 from Dekkers, Einmahl and de Haan (1989), using that $Q(1-s) = U(1/s)$, where $U$ is the inverse of $1/(1-F)$. Essentially, the inequalities are properties of regularly varying functions for $\gamma < 0$ and of $\Pi$-varying functions for $\gamma \geq 0$. $\square$

REMARK 3.1. From the properties of regularly varying functions, it follows that, in the case $\gamma > 0$, we can take $a(s)/Q(1-s) = \gamma$ in Lemma 3.1, whereas, in the case $\gamma < 0$, we can take $a(s)/Q(1-s) = -\gamma(\log Q(1) - \log Q(1-s))$. Moreover, as a consequence of Lemma 3.1 and the properties of $\Pi$-varying functions, we have that, in the case $\gamma = 0$, $a(s) = o(Q(1-s))$.

The idea now is to use (3.3) and (3.4) from Lemma 3.1 with $y = \Gamma_n(hu)/U_{(k+1)}$, where $k = \lfloor nh \rfloor$. Unfortunately, Lemma 3.1 cannot be applied directly. However, the next lemma shows that we may as well apply Lemma 3.1 with $y$ equal to $u$ instead of $\Gamma_n(hu)/U_{(k+1)}$.

LEMMA 3.2. *Let $\Gamma_n(\cdot)$ denote the empirical quantile function of $U_1, \dots, U_n$ with $U_i$ i.i.d. $\mathcal{U}(0,1)$, $h$ be a sequence of positive numbers with $h = h_n \to 0$ and $nh_n \to \infty$, as $n \to \infty$, and let $L(\cdot)$ be an integrable, bounded and positive function on $(0,1)$. Define $k = \lfloor nh \rfloor$ and $\bar{\lambda} = (\lambda \wedge 0)$ for $\lambda > -1$. Then, for each $\beta > (-1-\bar{\lambda})$,*

$$\int_0^1 \left[\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^\beta - u^\beta\right] u^\lambda L(u)\, du \xrightarrow{\mathcal{P}} 0 \qquad \text{as } n \to \infty. \tag{3.5}$$

PROOF. The case $\beta = 0$ is trivial; hence, we consider the case $\beta \neq 0$. Write the left-hand side of (3.5) as

$$\int_0^1 \left[\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^\beta - \left(\frac{\Gamma_n(ku/n)}{U_{(k+1)}}\right)^\beta\right] u^\lambda L(u)\,du \tag{3.6}$$

$$+\int_0^1 \left[\left(\frac{\Gamma_n(ku/n)}{U_{(k+1)}}\right)^\beta - u^\beta\right] u^\lambda L(u)\,du. \tag{3.7}$$

For (3.6), note that, for $j = 1, \ldots, k$, by definition,

$$\big(\Gamma_n(hu)\big)^\beta - \big(\Gamma_n(ku/n)\big)^\beta = \begin{cases} 0, & \dfrac{j-1}{k} < u \le \dfrac{j}{nh}, \\[2ex] U_{(j+1)}^\beta - U_{(j)}^\beta, & \dfrac{j}{nh} < u \le \dfrac{j}{k}. \end{cases}$$

Hence, (3.6) equals

$$U_{(k+1)}^{-\beta} \sum_{j=1}^{k} \int_{j/nh}^{j/k} (U_{(j+1)}^\beta - U_{(j)}^\beta) u^\lambda L(u)\,du.$$

Let $\|L\| = \sup_{s\in(0,1)} |L(s)|$ and $\bar{\lambda} = \lambda \wedge 0$. Using that $|x^{\lambda+1} - y^{\lambda+1}| \le (\lambda+1) \times (x-y) y^{\bar{\lambda}}$ for all $0 \le y \le x \le 1$, and $nh - k < 1$, we get

$$\begin{aligned} &\left| U_{(k+1)}^{-\beta} \sum_{j=1}^{k} \int_{j/nh}^{j/k} (U_{(j+1)}^\beta - U_{(j)}^\beta) u^\lambda L(u)\,du \right| \\ &\quad \le \frac{\|L\|}{\lambda+1} U_{(k+1)}^{-\beta} \sum_{j=1}^{k} |U_{(j+1)}^\beta - U_{(j)}^\beta| \left| \left(\frac{j}{k}\right)^{\lambda+1} - \left(\frac{j}{nh}\right)^{\lambda+1} \right| \\ &\quad < \|L\| U_{(k+1)}^{-\beta} \sum_{j=1}^{k} |U_{(j+1)}^\beta - U_{(j)}^\beta| \frac{j}{knh} \left(\frac{1}{nh}\right)^{\bar{\lambda}}. \end{aligned}$$

Note that the terms $U_{(j+1)}^\beta - U_{(j)}^\beta$ are either all positive (in the case $\beta > 0$) or all negative (in the case $\beta < 0$), which implies that the right-hand side is equal to

$$\begin{aligned} &\frac{\|L\|}{k(nh)^{1+\bar{\lambda}}} U_{(k+1)}^{-\beta} \left| \sum_{j=1}^{k} j (U_{(j+1)}^\beta - U_{(j)}^\beta) \right| \\ &\quad = \frac{\|L\|}{k(nh)^{1+\bar{\lambda}}} U_{(k+1)}^{-\beta} \left| (k+1) U_{(k+1)}^\beta - U_{(1)}^\beta - \sum_{j=1}^{k} U_{(j+1)}^\beta \right| \\ &\quad = \frac{\|L\|}{k(nh)^{1+\bar{\lambda}}} \left| \sum_{j=1}^{k+1} \left(1 - \left(\frac{U_j}{U_{(k+1)}}\right)^\beta\right) \right|. \end{aligned}$$

Note that, for $j = 1, \dots, k+1$, $U_{(k+1)} \geq U_{(j)}^{\beta} \geq U_{(1)}^{\beta}$ if $\beta > 0$, and $U_{(k+1)} \leq U_{(j)}^{\beta} \leq U_{(1)}^{\beta}$ if $\beta < 0$. This implies that, for all $\beta \neq 0$, the last expression is bounded by

$$\frac{\|L\|(k+1)}{k(nh)^{1+\bar{\lambda}}}\left|1 - \left(\frac{U_1}{U_{(k+1)}}\right)^{\beta}\right| \leq \frac{2\|L\|}{(nh)^{1+\bar{\lambda}}}\left|1 - \left(\frac{U_1}{U_{(k+1)}}\right)^{\beta}\right|. \tag{3.8}$$

In the case $\beta > 0$, we know that, with probability 1, $1 - (U_{(1)}/U_{(k+1)})^{\beta}$ is bounded between 0 and 1, hence, (3.8) tends to 0 as $n \to \infty$. In the case $\beta < 0$, first observe that, for any integer $1 \leq k \leq n-1$, we have that

$$\left(\frac{U_{(1)}}{U_{(k+1)}}, \dots, \frac{U_{(k)}}{U_{(k+1)}}\right) \stackrel{\mathcal{D}}{=} (V_{(1)}, \dots, V_{(k)}), \tag{3.9}$$

where $V_{(1)}, \dots, V_{(k)}$ are the order statistics of $k$ i.i.d. $\mathcal{U}(0, 1)$ variables. Therefore, we have that $U_{(1)}/U_{(k+1)} \stackrel{\mathcal{D}}{=} V_{(1)}$, so that, for any $\delta > 0$,

$$\mathbb{P}\left(\left[\left(\frac{U_{(1)}}{U_{(k+1)}}\right)^{\beta} - 1\right] > \delta(nh)^{1+\bar{\lambda}}\right) = 1 - \left(1 - \left(\delta(nh)^{1+\bar{\lambda}} + 1\right)^{1/\beta}\right)^{k}. \tag{3.10}$$

However, since $\bar{\lambda} > -1$ and $\beta < 0$, we have that

$$k \log\left(1 - \left(\delta(nh)^{1+\bar{\lambda}} + 1\right)^{1/\beta}\right) = -k\left(\delta(nh)^{1+\bar{\lambda}} + 1\right)^{1/\beta}\left(1 + o(1)\right) \qquad \text{as } n \to \infty.$$

Using that $k \sim nh$, we find that (3.10) tends to 0, whenever $1 + (1 + \bar{\lambda})/\beta < 0$. Hence, (3.8) tends to 0 in probability as $n \to \infty$, whenever $-1 - \bar{\lambda} < \beta < 0$.

Finally, consider the second term (3.7). Note that property (3.9) yields that all finite-dimensional projections of the process $u \mapsto \Gamma_n(hu)/U_{(k+1)}$ are equal in distribution to the finite-dimensional projections of the process $u \mapsto \Gamma_k(u)$, where $\Gamma_k(u)$ is the empirical quantile function of a $\mathcal{U}(0, 1)$ sample $V_1, \dots, V_k$. Hence, (3.7) is equal in distribution to

$$\int_0^1 \left(\Gamma_k(u)^{\beta} - u^{\beta}\right)u^{\lambda} L(u)\, du. \tag{3.11}$$

Moreover, for $0 < \nu_1 < 1 + \lambda$ and $0 < \nu_2 < 1$, we have

$$\begin{aligned} &\left|\int_0^1 \left(\Gamma_k(u)^{\beta} - u^{\beta}\right)u^{\lambda} L(u)\, du\right| \\ &\qquad \leq \sup_{0<u<1}\left[u^{\nu_1}(1-u)^{\nu_2}\left|\Gamma_k(u)^{\beta} - u^{\beta}\right|\right]\|L\| \int_0^1 u^{\lambda-\nu_1}(1-u)^{-\nu_2}\, du. \end{aligned}$$

For $\beta > 0$, according to (3.1), the right-hand side has the same distribution as

$$\sup_{0<u<1}\left[u^{\nu_1}(1-u)^{\nu_2}\left|F_{\beta,k}^{-1}(u) - F_{\beta}^{-1}(u)\right|\right]\|L\| \int_0^1 u^{\lambda-\nu_1}(1-u)^{-\nu_2}\, du, \tag{3.12}$$

where $F_\beta^{-1}$ is the quantile function corresponding to the distribution function $F_\beta(x) = x^{1/\beta}$ for $0 < x < 1$ and $F_{\beta,k}^{-1}$ denotes the empirical quantile function of a sample $Y_1, \dots, Y_k$ drawn from $F_\beta$. Note that, since $0 < |Y_1| < 1$, one has that $\mathbb{E}|Y_1 \wedge 0|^{1/\nu_1} = 0$ and $\mathbb{E}(Y_1 \vee 0)^{1/\nu_2} < \infty$ for $\nu_2 > 0$ and $\beta > 0$. Theorem 3 in Mason (1982) then yields that the supremum in (3.12) tends to 0 with probability 1 as $k \to \infty$. Since $\nu_1 < (1 + \lambda)$ and $\nu_2 < 1$, the integral in (3.12) is finite. We conclude that, in the case $\beta > 0$, (3.11) tends to 0 with probability 1 as $k \to \infty$. In the case $\beta < 0$, again using (3.1), note that

$$\sup_{0<u<1} u^{\nu_1}(1-u)^{\nu_2}|\Gamma_k(u)^\beta - u^\beta| \stackrel{\mathcal{D}}{=} \sup_{0<u<1} (1-u)^{\nu_1} u^{\nu_2} \big|G_{\beta,k}^{-1}(u) - G_\beta^{-1}(u)\big|,$$

with $G_\beta^{-1}$ the quantile function corresponding to the distribution function $G_\beta(x) = 1 - x^{1/\beta}$ for $x \ge 1$ and $G_{\beta,k}^{-1}$ denoting the empirical quantile function of a sample $Z_1, \dots, Z_k$ drawn from $G_\beta$. Again, use Theorem 3 in Mason (1982), with $\mathbb{E}\,|Z_1 \wedge 0|^{1/\nu_2} = 0$, whenever $\nu_2 > 0$ and $\beta < 0$, and

$$\mathbb{E}(Z_1 \vee 0)^{1/\nu_1} = -\frac{1}{\beta}\int_1^\infty z^{1/\nu_1 + 1/\beta - 1}\, dz < \infty,$$

whenever $\nu_1 > -\beta$. Hence, (3.11) tends to 0 almost surely as $k$ tends to $\infty$, taking $-\beta < \nu_1 < (1 + \lambda)$ and $0 < \nu_2 < 1$. $\square$

LEMMA 3.3. *Assume that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in \mathbb{R}$. Let $K$ be a kernel satisfying conditions* (CK1)–(CK4) *and let $\hat{\gamma}_{n,h}^K$ be defined by* (1.2). *If $h = h_n$ is such that $h \downarrow 0$ and $nh \to \infty$ as $n \to \infty$, then $\hat{\gamma}_{n,h}^{(\mathrm{pos})} \to (\gamma \vee 0)$ in probability.*

PROOF. First, observe that, according to (3.1) and conditions (CK2) and (CK4), we can write

$$\hat{\gamma}_{n,h}^{(\mathrm{pos})} \stackrel{\mathcal{D}}{=} \int_0^1 \Big(\log Q\big(1 - \Gamma_n(hu)\big) - \log Q\big(1 - U_{(k+1)}\big)\Big)\, d\big(uK(u)\big),$$

where $\Gamma_n$ is the empirical quantile function of a uniform (0, 1) sample $U_1, \dots, U_n$ and $k = \lfloor nh \rfloor$. Consider the case $\gamma > 0$. By definition, $U_{(k+1)} \ge \Gamma_n(hu)$ with probability 1 for all $u \in (0, 1)$, and $U_{(k+1)} \to 0$ with probability 1 as $h \downarrow 0$. We can therefore apply Lemma 3.1, with $y = \Gamma_n(hu)/U_{(k+1)}$, $s = U_{(k+1)}$ and $a(s)/Q(1 - s) = \gamma$ (see Remark 3.1), to get that, with probability 1, for each $\varepsilon > 0$ there exists an $n_0$ such that, for all $n \ge n_0$,

$$\begin{aligned}(1-\varepsilon)\frac{1 - (\Gamma_n(hu)/U_{(k+1)})^\varepsilon}{\varepsilon} - \varepsilon &< \frac{\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})}{\gamma} \\ &< (1+\varepsilon)\frac{(\Gamma_n(hu)/U_{(k+1)})^{-\varepsilon} - 1}{\varepsilon} + \varepsilon\end{aligned}$$

for all $u \in (0,1)$. Defining $L(u) = d(uK(u))/du$, we get, by the boundedness of both $K$ and $K'$, that $L(u) = L^+(u) - L^-(u)$, where $L^\pm(u)$ are positive and bounded functions. Hence, for $\gamma > 0$,

$$\hat{\gamma}_{n,h}^{(\mathrm{pos})} < \gamma \int_0^1 \left[(1+\varepsilon)\frac{(\Gamma_n(hu)/U_{(k+1)})^{-\varepsilon} - 1}{\varepsilon} + \varepsilon\right] L^+(u)\,du$$
$$- \gamma \int_0^1 \left[(1-\varepsilon)\frac{1 - (\Gamma_n(hu)/U_{(k+1)})^{\varepsilon}}{\varepsilon} - \varepsilon\right] L^-(u)\,du.$$

Applying Lemma 3.2 twice (once with $\beta = -\varepsilon$, $\lambda = 0$ and $L^+$ and once with $\beta = \varepsilon$, $\lambda = 0$ and $L^-$) yields that, for any $0 < \varepsilon < 1$, this upper bound tends to

$$\gamma \int_0^1 \left[(1+\varepsilon)\frac{u^{-\varepsilon} - 1}{\varepsilon} + \varepsilon\right] L^+(u)\,du - \gamma \int_0^1 \left[(1-\varepsilon)\frac{1 - u^{\varepsilon}}{\varepsilon} - \varepsilon\right] L^-(u)\,du$$

in probability as $n \to \infty$. Letting $\varepsilon \downarrow 0$, by dominated convergence this tends to

$$\gamma \int_0^1 (-\log u) L^+(u)\,du - \gamma \int_0^1 (-\log u) L^-(u)\,du$$
$$= \gamma \int_0^1 (-\log u)\, d\big((uK(u)\big) = \gamma.$$

Similar arguments lead to a lower bound for $\hat{\gamma}_{n,h}^{(\mathrm{pos})}$ that tends to $\gamma$ in probability as well. This proves the lemma for the case $\gamma > 0$.

In the case $\gamma = 0$, first note that, as a consequence of Lemma 3.1 together with the properties of $\Pi$-varying functions, one has that $a(s) = o(Q(1-s))$ for $s \downarrow 0$. Since $U_{(k+1)} \to 0$ with probability 1, this means that $a(U_{(k+1)})/Q(1 - U_{(k+1)}) \to 0$ with probability 1. Similar to the case $\gamma > 0$, we can apply the inequalities of Lemma 3.1 to

$$\frac{\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})}{a(U_{(k+1)})/Q(1 - U_{(k+1)})}.$$

By similar arguments as above, we conclude that $\hat{\gamma}_{n,h}^{(\mathrm{pos})} \to 0$ in probability.

Finally, consider the case $\gamma < 0$. Lemma 3.1 now yields the inequalities

$$1 - (1+\varepsilon)\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^{-\gamma-\varepsilon} < \frac{\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})}{\log Q(1) - \log Q(1 - U_{(k+1)})}$$
$$< 1 - (1-\varepsilon)\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^{-\gamma+\varepsilon}.$$

Thus, with $L^\pm$ as before,

$$\frac{\hat{\gamma}_{n,h}^{(\mathrm{pos})}}{\log Q(1) - \log Q(1 - U_{(k+1)})} < \int_0^1 \left[1 - (1-\varepsilon)\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^{-\gamma+\varepsilon}\right] L^+(u)\,du$$
$$- \int_0^1 \left[1 - (1+\varepsilon)\left(\frac{\Gamma_n(hu)}{U_{(k+1)}}\right)^{-\gamma-\varepsilon}\right] L^-(u)\,du.$$

Again, by two applications of Lemma 3.2 (once with $\beta = -\gamma + \varepsilon$, $\lambda = 0$ and $L^+$ and once with $\beta = -\gamma - \varepsilon$, $\lambda = 0$ and $L^-$), we get that, for any $0 < \varepsilon < 1 - \gamma$, the upper bound tends to

$$\int_0^1 [1 - (1-\varepsilon)u^{-\gamma+\varepsilon}]L^+(u)\,du - \int_0^1 [1 - (1+\varepsilon)u^{-\gamma-\varepsilon}]L^-(u)\,du.$$

Since both integrals are bounded for $0 < \varepsilon < 1 - \gamma$ and $\log Q(1) - \log Q(1 - U_{(k+1)}) \to 0$ with probability 1, we get (with a similar lower bound) that $\hat{\gamma}_{n,h}^{(\mathrm{pos})} \to 0$. $\square$

LEMMA 3.4. *Assume that* $F \in \mathcal{D}(G_\gamma)$ *for some* $\gamma \in \mathbb{R}$. *Let* $K$ *be a kernel satisfying conditions* (CK1)–(CK4) *and, for arbitrary* $\alpha > 0$, *let* $\hat{\gamma}_{n,h}^K$ *be defined by* (1.2). *If* $h = h_n$ *is such that* $h \downarrow 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *then* $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \to 1 + (\gamma \wedge 0)$ *in probability.*

PROOF. Since we will consider $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)}$, we can scale both numerator and denominator by the same factor, without changing the ratio. Moreover, by conditions (CK2) and (CK4), we have that, for any $\alpha > 0$,

$$\int_0^1 d\big(u^\alpha K(u)\big) = [u^\alpha K(u)]_0^1 = 0$$

and

$$\int_0^1 d\left[\frac{d}{du}u^{\alpha+1}K(u)\right] = \big[(\alpha+1)u^\alpha K(u) + u^{\alpha+1}K'(u)\big]_0^1 = 0.$$

First, consider $\gamma \geq 0$. If we write $d(u^\alpha K(u)) = u^{\alpha-1}L_1(u)\,du$, then, by the previous remarks, we have that

$$\frac{h^{1-\alpha}\hat{q}_{n,h}^{(1)}}{a(U_{(k+1)})/Q(1-U_{(k+1)})} \stackrel{\mathcal{D}}{=} \int_0^1 \frac{\log Q(1-\Gamma_n(hu)) - \log Q(1-U_{(k+1)})}{a(U_{(k+1)})/Q(1-U_{(k+1)})} u^{\alpha-1}L_1(u)\,du.$$

Similarly to the argument used in the proof of Lemma 3.3, we can first apply the inequalities from Lemma 3.1. Then, with $\varepsilon > 0$ fixed, let $n \to \infty$ and apply Lemma 3.2 with $\lambda = \alpha - 1$, and finally let $\varepsilon \downarrow 0$. We conclude that

$$\frac{h^{1-\alpha}\hat{q}_{n,h}^{(1)}}{a(U_{(k+1)})/Q(1-U_{(k+1)})} \to \int_0^1 (-\log u)\,d\big(u^\alpha K(u)\big) = \int_0^1 u^{\alpha-1}K(u)\,du \tag{3.13}$$

in probability. On the other hand, if we write

$$d\left(\frac{d}{du}[u^{\alpha+1}K(u)]\right) = u^{\alpha-1}L_2(u)\,du,$$

we have

$$\frac{h^{1-\alpha}\hat{q}_{n,h}^{(2)}}{a(U_{(k+1)})/Q(1-U_{(k+1)})} \stackrel{\mathcal{D}}{=} \int_0^1 \frac{\log Q(1-\Gamma_n(hu)) - \log Q(1-U_{(k+1)})}{a(U_{(k+1)})/Q(1-U_{(k+1)})} u^{\alpha-1}L_2(u)\,du.$$

Similarly, by an application of Lemmas 3.1 and 3.2, this tends in probability to

$$\int_0^1 (-\log u)\,d\left(\frac{d}{du}u^{\alpha+1}K(u)\right) = \int_0^1 \frac{d}{du}\big(u^{\alpha+1}K(u)\big)u^{-1}\,du = \int_0^1 u^{\alpha-1}K(u)\,du.$$

Combining this with (3.13), we obtain that $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \xrightarrow{\mathcal{P}} 1$, whenever $\gamma \geq 0$.

In the case $\gamma < 0$, similar arguments yield that

$$\frac{h^{1-\alpha}\hat{q}_{n,h}^{(1)}}{\log Q(1) - \log Q(1-U_{(k+1)})} \to \int_0^1 (1-u^{-\gamma})\,d\big(u^{\alpha}K(u)\big) = -\gamma\int_0^1 u^{\alpha-\gamma-1}K(u)\,du$$

in probability, and that

$$\begin{aligned}\frac{h^{1-\alpha}\hat{q}_{n,h}^{(2)}}{\log Q(1) - \log Q(1-U_{(k+1)})} &\to \int_0^1 (1-u^{-\gamma})\,d\left(\frac{d}{du}u^{\alpha+1}K(u)\right)\\ &= -\gamma(1+\gamma)\int_0^1 u^{\alpha-\gamma-1}K(u)\,du\end{aligned}$$

in probability. Hence, $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \xrightarrow{\mathcal{P}} 1+\gamma$ as $n \to \infty$. □

The following theorem is now a direct corollary of Lemmas 3.3 and 3.4.

THEOREM 3.1 (Consistency). *Assume that* $F \in \mathcal{D}(G_\gamma)$ *for some* $\gamma \in \mathbb{R}$. *Let* $K$ *be a kernel satisfying conditions* (CK1)–(CK4) *and, for arbitrary* $\alpha > 0$, *let* $\hat{\gamma}_{n,h}^K$ *be defined by* (1.2). *If* $h = h_n$ *is such that* $h \downarrow 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *then* $\hat{\gamma}_{n,h}^K \to \gamma$ *in probability as* $n \to \infty$.

**4. Asymptotic normality.** In order to obtain asymptotic normality, we need additional assumptions on $F$. Suppose that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in \mathbb{R}$ and assume that $\phi$ from (2.1) exists and is well defined. Moreover, we assume that $F$ satisfies the following conditions:

(CP1) In the case $\gamma \geq 0$, assume that $\phi(s) \to \gamma$, as $s \downarrow 0$.
(CP2) In the case $\gamma < 0$, assume that, for some constant $c > 0$, $s^\gamma \phi(s) \to -c\gamma$, as $s \downarrow 0$.
(CP3) In the case $\gamma = 0$, for all $s > 0$ assume that $\phi(hs)/\phi(h) \to 1$, as $h \downarrow 0$.

Consider the deterministic equivalent of $\hat{\gamma}_{n,h}^K$:

$$\gamma_h = \gamma_h^{(\text{pos})} + \frac{q_h^{(2)}}{q_h^{(1)}} - 1, \tag{4.1}$$

with

$$\gamma_h^{(\text{pos})} = \int_0^1 \log Q(1-hu)\, d\big(uK(u)\big), \tag{4.2}$$

$$q_h^{(i)} = h^{\alpha-1} \int_0^1 \log Q(1-hu)\, dK^{(i)}(u), \qquad i = 1, 2, \tag{4.3}$$

where $K^{(1)}(u) = u^\alpha K(u)$ and $K^{(2)}(u) = d(u^{\alpha+1}K(u))/du$ for a kernel $K$. Also write

$$K_h^{(1)}(u) = u^\alpha K_h(u), \tag{4.4}$$

$$K_h^{(2)}(u) = \frac{d}{du}\big(u^{\alpha+1}K_h(u)\big). \tag{4.5}$$

LEMMA 4.1. *Let $X_1, \ldots, X_n$ be a sample from $F \in \mathcal{D}(G_\gamma)$ and suppose that $F$ satisfies conditions* (CP1)–(CP3). *Let $K$ be a kernel satisfying conditions* (CK1)–(CK4) *and let $\hat{\gamma}_{n,h}^K$ be defined as in* (1.2). *Then, for any $\alpha > \frac{1}{2}$ and $h = h_n$, with $h \downarrow 0$ and $(nh)^{-\alpha} \log n = O((nh)^{-1/2})$, as $n \to \infty$, we have, for $i = 1, 2$,*

$$\sqrt{nh}\, h^{1-\alpha}\big(\hat{q}_{n,h}^{(i)} - q_h^{(i)}\big) \overset{\mathcal{D}}{=} \int_0^1 \frac{W(u)}{u} \phi(hu)\, dK^{(i)}(u) + o_P(1) \tag{4.6}$$

*as $n \to \infty$, where $W$ denotes standard Brownian motion.*

PROOF. We will only present the proof for $\hat{q}_{n,h}^{(1)}$, since the proof for $\hat{q}_{n,h}^{(2)}$ is similar. The left-hand side of (4.6) can be decomposed into four parts:

$$\begin{aligned} &\hat{q}_{n,h}^{(1)} - q_h^{(1)} \\ &\quad = \int_0^{1/n} \log Q_n(1-u)\, dK_h^{(1)}(u) - \int_0^{1/n} \log Q(1-u)\, dK_h^{(1)}(u) \\ &\qquad + \int_{1/n}^{b_n} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) + \int_{b_n}^{h} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u), \end{aligned} \tag{4.7}$$

where $(b_n)$ is a sequence of positive real numbers that satisfies $1/n < b_n < h$.

For the first term of (4.7), note that $Q_n(1-u)$ is constant for $0 \le u < 1/n$. Together with property (3.1), we get that

$$h^{1-\alpha}\int_0^{1/n} \log Q_n(1-u)\,dK_h^{(1)}(u) \stackrel{\mathcal{D}}{=} \log Q(1-U_{(1)})\int_0^{1/nh} dK^{(1)}(u),$$

where $U_{(1)}$ is the first order statistic from a sample $U_1,\dots,U_n$ from a uniform $(0,1)$ distribution. Note that, from the properties of slowly varying functions, it follows that

$$\frac{\log Q(1-s)}{-\log s} \to 0 \tag{4.8}$$

[see de Wolf (1999) for a formal proof]. Therefore, since $U_{(1)} \to 0$ almost surely, we have that

$$h^{1-\alpha}\int_0^{1/n} \log Q_n(1-u)\,dK_h^{(1)}(u) \stackrel{\mathcal{D}}{=} o_P\left(\frac{-\log U_{(1)}}{(nh)^\alpha}\right) = o_P\big((nh)^{-1/2}\big). \tag{4.9}$$

The last equality follows from the fact that, for any $\varepsilon > 0$,

$$\begin{aligned}&\mathbb{P}\big(-(nh)^{1/2-\alpha}\log U_{(1)} \ge \varepsilon\big)\\ &\quad = 1 - \big(1-\exp(-(nh)^{\alpha-1/2}\varepsilon)\big)^n \le n\exp\big(-(nh)^{\alpha-1/2}\varepsilon\big),\end{aligned}$$

which tends to 0 according to the conditions on $h$. For the second part of (4.7), observe that, by integration by parts and application of (4.8),

$$\begin{aligned}&h^{1-\alpha}\int_0^{1/n} \log Q(1-u)\,dK_h^{(1)}(u)\\ &\quad = h^{1-\alpha}\log Q\left(1-\frac{1}{n}\right)K_h^{(1)}\left(\frac{1}{n}\right) + \int_0^{1/nh} \phi(hu)K^{(1)}(u)\,du,\end{aligned}$$

where $\phi$ is defined in (2.1). Conditions (CP1)–(CP3) yield that $\phi(s) \to (\gamma \vee 0)$ as $s \downarrow 0$. From the conditions on $h$, together with another application of (4.8), we conclude

$$\begin{aligned}&h^{1-\alpha}\int_0^{1/n} \log Q(1-u)\,dK_h^{(1)}(u)\\ &\quad = o\left(\frac{\log n}{(nh)^\alpha}\right) + O\big((nh)^{-1}\big) = o\big((nh)^{-1/2}\big).\end{aligned} \tag{4.10}$$

For the third part of (4.7), first observe that

$$\begin{aligned}&\int_{1/n}^{b_n} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u)\\ &\quad \stackrel{\mathcal{D}}{=} \int_{1/n}^{b_n} \big[\log Q\big(1-\Gamma_n(u)\big) - \log Q(1-u)\big]\,dK_h^{(1)}(u),\end{aligned}$$

where $\Gamma_n$ is the empirical quantile function of a uniform (0, 1) sample of size $n$. By the mean value theorem, we then get that

$$\int_{1/n}^{b_n} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) \overset{\mathcal{D}}{=} \int_{1/n}^{b_n} \frac{\phi(u+\xi_{n,u})}{u+\xi_{n,u}}\big(u-\Gamma_n(u)\big)\, dK_h^{(1)}(u),$$

with $|\xi_{n,u}| \le |\Gamma_n(u)-u|$. We have that $\sup_{0<u<1}|\Gamma_n(u)-u| \to 0$ with probability 1 as $n \to \infty$, and from Wellner (1978),

$$\sup_{1/n\le u\le 1}\left|\frac{\Gamma_n(u)}{u}\right| = O_P(1) \quad \text{and} \quad \sup_{1/n\le u\le 1}\left|\frac{u}{\Gamma_n(u)}\right| = O_P(1). \tag{4.11}$$

From the conditions on $F$, it follows that $\phi$ is uniformly bounded in a neighborhood of 0. Furthermore, note that $u/(u+\xi_{n,u})$ lies between $u/\Gamma_n(u)$ and 1. Hence,

$$\sup_{1/n\le u\le b_n}\left|\phi(u+\xi_{n,u})\frac{u}{u+\xi_{n,u}}\frac{\Gamma_n(u)-u}{u}\right| = O_P(1).$$

Writing $dK^{(1)}(u)/du = u^{\alpha-1}L_1(u)$, we therefore obtain that

$$\begin{aligned}\left|\int_{1/n}^{b_n} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u)\right| &\le O_P(1)\int_{1/n}^{b_n}\left|\frac{dK_h^{(1)}(u)}{du}\right| du\\ &= h^{\alpha-1}O_P(1)\int_{1/nh}^{b_n/h} u^{\alpha-1}|L_1(u)|\, du\\ &= h^{\alpha-1}O_P\big((b_n/h)^{\alpha}\big).\end{aligned}$$

Taking $b_n = h(nh)^{-(1/2+\lambda)/\alpha}$ for some $0<\lambda<\alpha-1/2$, we get that

$$h^{1-\alpha}\int_{1/n}^{b_n} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) = o_P\big((nh)^{-1/2}\big). \tag{4.12}$$

Finally, consider the fourth part of the decomposition (4.7). Following the same arguments as for the third part, we arrive at

$$\int_{b_n}^{h} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) \overset{\mathcal{D}}{=} \int_{b_n}^{h} \frac{\phi(u+\xi_{n,u})}{u+\xi_{n,u}}\big(\Gamma_n(u)-u\big)\, dK_h^{(1)}(u)$$

for some $|\xi_{n,u}| \le |\Gamma_n(u)-u|$. Since now $b_n \le u \le h$, we have that

$$\sup_{b_n\le u\le 1}\left|\frac{\Gamma_n(u)-u}{u}\right| = o_P(1) \tag{4.13}$$

for any sequence $(b_n)$ of positive numbers satisfying $nb_n \to \infty$ as $n \to \infty$ [see Wellner (1978)]. Condition (CP3) states that $\phi$ is slowly varying. This implies that $\phi(hs)/\phi(h) \to 1$ as $h \downarrow 0$ uniformly for $s \in [a,b]$ for any $0<a<b<\infty$. By means of (4.13), we have that, for $n$ sufficiently large, $1/2 < 1 + \xi_{n,u}/u < 3/2$, which implies that

$$\frac{\phi(u+\xi_{n,u})}{\phi(u)} = \frac{\phi(u(1+\xi_{n,u}/u))}{\phi(u)} \to 1$$

uniformly for $b_n \le u \le h$. It follows that, for all $\gamma \in \mathbb{R}$,

$$\sup_{b_n \le u \le h} \left| \frac{\phi(u+\xi_{n,u})}{\phi(u)} \frac{u}{u+\xi_{n,u}} \right| = 1 + o_P(1).$$

This implies that

$$\int_{b_n}^{h} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) = \bigl(1+o_P(1)\bigr) \int_{b_n}^{h} \phi(u) \frac{\Gamma_n(u)-u}{u}\, dK_h^{(1)}(u).$$

Note that, from Theorem 2.1 in Csörgő, Csörgő, Horváth and Mason (1986), there exists a sequence $(B_n)$ of Brownian bridges such that, for $0 \le \nu < 1/2$,

$$\sup_{1/n \le u \le 1-1/n} \frac{|\sqrt{n}(\Gamma_n(u)-u) - B_n(u)|}{u^{1/2-\nu}} = O_P(n^{-\nu}) \tag{4.14}$$

as $n \to \infty$, where $\Gamma_n$ is the quantile function of $U_1, \dots, U_n$. Applying (4.14), we get that

$$\int_{b_n}^{h} \phi(u) \frac{\Gamma_n(u)-u}{u}\, dK_h^{(1)}(u) = n^{-1/2} \int_{b_n}^{h} \phi(u) \frac{B_n(u)}{u}\, dK_h^{(1)}(u) + R_{n,h},$$

where, for arbitrary $0 \le \nu < 1/2$,

$$\begin{aligned}
|R_{n,h}| &\le O_P(n^{-1/2-\nu}) \int_{b_n}^{h} u^{-1/2-\nu} |\phi(u)| \left| \frac{dK_h^{(1)}(u)}{du} \right| du \\
&\le h^{\alpha-1} O_P\bigl((nh)^{-1/2-\nu}\bigr) \int_{b_n/h}^{1} u^{-1/2-\nu} |\phi(hu)| \left| \frac{dK^{(1)}(u)}{du} \right| du \\
&= h^{\alpha-1} O_P\bigl((nh)^{-1/2-\nu}\bigr).
\end{aligned}$$

Using that $B_n(u) \stackrel{\mathcal{D}}{=} W_n(u) + \zeta_n u$, where $W_n$ is distributed as standard Brownian motion and $\zeta_n$ is a standard normal variable, independent of $W_n$, we obtain, for $h \downarrow 0$ and $nh \to \infty$,

$$\begin{aligned}
&\int_{b_n}^{h} \phi(u) \frac{B_n(u)}{u}\, dK_h^{(1)}(u) \\
&\quad \stackrel{\mathcal{D}}{=} \int_{b_n}^{h} \phi(u) \frac{W_n(u)}{u}\, dK_h^{(1)}(u) + \zeta_n \int_{b_n}^{h} \phi(u)\, dK_h^{(1)}(u) \\
&\quad = \int_{b_n}^{h} \phi(u) \frac{W_n(u)}{u}\, dK_h^{(1)}(u) + h^{\alpha-1} \zeta_n \int_{b_n/h}^{1} \phi(hu)\, dK^{(1)}(u) \\
&\quad \stackrel{\mathcal{D}}{=} h^{\alpha-1} h^{-1/2} \int_{b_n/h}^{1} \phi(hu) \frac{W_n(u)}{u}\, dK^{(1)}(u) + h^{\alpha-1} O_P(1),
\end{aligned}$$

where in the last equality we used that $W_n(hu) \stackrel{\mathcal{D}}{=} \sqrt{h} W_n(u)$. Finally, since $\mathbb{E}|W_n(u)| \le \sqrt{\mathbb{E}\, W_n(u)^2} = \sqrt{u}$, we find that

$$\int_0^{b_n/h} \phi(hu) \frac{W(u)}{u} \, dK^{(1)}(u) = O_P\big((b_n/h)^{\alpha-1/2}\big) = o_P(1).$$

Therefore, by taking $b_n = h(nh)^{-(1/2+\lambda)/\alpha}$ for some $0 < \lambda < \alpha - 1/2$, we obtain that

$$\begin{aligned}
h^{1-\alpha} & \int_{b_n}^{h} \log\left(\frac{Q_n(1-u)}{Q(1-u)}\right) dK_h^{(1)}(u) \\
& \stackrel{\mathcal{D}}{=} (nh)^{-1/2}\big(1 + o_P(1)\big) \int_0^1 \phi(hu) \frac{W(u)}{u} \, dK^{(1)}(u) + o_P\big((nh)^{-1/2}\big) \\
& = (nh)^{-1/2} \int_0^1 \phi(hu) \frac{W(u)}{u} \, dK^{(1)}(u) + o_P\big((nh)^{-1/2}\big).
\end{aligned}$$

Together with decomposition (4.7), (4.9), (4.10) and (4.12), we obtain the assertion of the lemma for $q_{n,h}^{(1)}$. The argument for $q_{n,h}^{(2)}$ runs similarly. □

THEOREM 4.1 (Asymptotic normality). *Let $X_1, \ldots, X_n$ be a sample from $F$ with $F$ satisfying* (CP1)–(CP3). *Let $K$ be a kernel satisfying conditions* (CK1)–(CK4) *and let $\hat{\gamma}_{n,h}^K$ be defined as in* (1.2). *Then, for any $\alpha > 1/2$ and $h = h_n$ with $h \downarrow 0$ and $(nh)^{-\alpha} \log n = O((nh)^{-1/2})$ as $n \to \infty$,*

$$\sqrt{nh}(\hat{\gamma}_{n,h}^K - \gamma_h) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \sigma_K^2),$$

*where $\gamma_h$ is defined in* (4.1) *and*

$$\sigma_K^2 = \int_0^1 \big(a_0 \tilde{K}(u) + a_1 \tilde{K}^{(2)}(u) - a_2 \tilde{K}^{(1)}(u)\big)^2 \, du,$$

*with*

$$\begin{aligned}
\tilde{K}(u) &= \int_u^1 x^{-1} \, d\big(xK(x)\big), \qquad & u \in (0,1], \\
\tilde{K}^{(i)}(u) &= \int_u^1 x^{-1-(\gamma \wedge 0)} \, dK^{(i)}(x), \qquad & u \in (0,1],
\end{aligned}$$

*and*

$$\begin{aligned}
a_0 &= \gamma \vee 0, \\
a_1 &= 1 \Big/ \int_0^1 x^{-1-(\gamma \wedge 0)} K^{(1)}(x) \, dx, \\
a_2 &= \big(1 + (\gamma \wedge 0)\big) a_1.
\end{aligned}$$

PROOF. First, note that by partial integration and application of (4.8), we have, for $i = 1, 2$,

$$h^{1-\alpha} q_h^{(i)} = \int_0^1 \frac{\phi(hu)}{u} K^{(i)}(u)\, du.$$

Note that, from the conditions on $F$, it follows that, in the case $\gamma \neq 0$, we have

$$\sup_{0<u<1} \left| \frac{\phi(hu)}{\phi(h)} - u^{-\bar{\gamma}} \right| \to 0 \tag{4.15}$$

as $h \downarrow 0$, where $\bar{\gamma} = \gamma \wedge 0$. This implies that

$$\left| \int_0^1 \left( \frac{\phi(hu)}{\phi(h)} - u^{-\bar{\gamma}} \right) \frac{K^{(i)}(u)}{u}\, du \right| = o(1). \tag{4.16}$$

In the case $\gamma = 0$, the function $\phi$ is slowly varying. This means we can apply the following inequality, taken from the proof of the proposition in the Appendix of de Haan and Pereira (1999): for each $\varepsilon, \varepsilon_1 > 0$, there exists an $h_0$ such that, for all $h \le h_0$ and all $hu \le h_0$,

$$\left| \frac{\phi(hu)}{\phi(h)} - 1 \right| \le \varepsilon e^{\varepsilon_1 |\log u|} = \varepsilon u^{-\varepsilon_1}, \tag{4.17}$$

where in the last equality we used that $u \in (0, 1)$. This implies that (4.16) also holds in the case $\gamma = 0$. Hence, for all $\gamma \in \mathbb{R}$, we have, for $i = 1, 2$,

$$h^{1-\alpha} q_h^{(i)} = \phi(h) \left[ \int_0^1 u^{-1-\bar{\gamma}} K^{(i)}(u)\, du + o(1) \right]. \tag{4.18}$$

Since this is $O(1)$, we have from Lemma 4.1 that

$$\frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}} = \frac{q_h^{(2)}}{q_h^{(1)}} + \frac{(nh)^{-1/2} A_n^{(2)}}{h^{1-\alpha} q_h^{(1)}} - \frac{(nh)^{-1/2} A_n^{(1)} h^{1-\alpha} q_h^{(2)}}{(h^{1-\alpha} q_h^{(1)})^2} + o_P\big((nh)^{-1/2}\big), \tag{4.19}$$

where, for $i = 1, 2$,

$$A_n^{(i)} = \int_0^1 \phi(hu) \frac{W(u)}{u}\, dK^{(i)}(u). \tag{4.20}$$

Because $\hat{\gamma}_{n,h}^{(\mathrm{pos})} - \gamma_h^{(\mathrm{pos})}$ is a special case of $\hat{q}_{n,h}^{(1)} - q_h^{(1)}$ for $\alpha = 1$, another consequence of Lemma 4.1 is that

$$\begin{aligned} \sqrt{nh}\big(\hat{\gamma}_{n,h}^{(\mathrm{pos})} - \gamma_h^{(\mathrm{pos})}\big) &= (\gamma \vee 0) \int_0^1 \frac{W(u)}{u}\, d\big(uK(u)\big) + o_P(1) \\ &= -a_0 \int_0^1 W(u)\, d\tilde{K}(u) + o_P(1). \end{aligned}$$

We find that

$$\sqrt{nh}(\hat{\gamma}_{n,h}^{K} - \gamma_h) = \sqrt{nh}(\hat{\gamma}_{n,h}^{(\mathrm{pos})} - \gamma_h^{(\mathrm{pos})}) + \sqrt{nh}\left(\frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}} - \frac{q_h^{(2)}}{q_h^{(1)}}\right)$$
$$= -a_0 \int_0^1 W(u)\, d\tilde{K}(u) + \frac{A_n^{(2)}}{h^{1-\alpha} q_h^{(1)}} - \frac{A_n^{(1)} h^{1-\alpha} q_h^{(2)}}{(h^{1-\alpha} q_h^{(1)})^2} + o_P(1).$$

To deal with the $A_n^{(i)}$, $i = 1, 2$, we again make use of (4.15) and (4.17). Together with $\mathbb{E}|W(u)| \le \sqrt{\mathbb{E}\, W(u)^2} = \sqrt{u}$, Markov's inequality, the conditions on $K$ and the fact that $\alpha > 1/2$, this implies that, for $\gamma \in \mathbb{R}$, we have, for $i = 1, 2$,

$$\left| \int_0^1 \left( \frac{\phi(hu)}{\phi(h)} - u^{\bar{\gamma}} \right) \frac{W(u)}{u}\, dK^{(i)}(u) \right| = o_P(1).$$

Hence, for all $\gamma \in \mathbb{R}$, we have, for $i = 1, 2$,

$$A_n^{(i)} = \phi(h) \left[ \int_0^1 u^{-1-\bar{\gamma}} W(u)\, dK^{(i)}(u) + o_P(1) \right]. \tag{4.21}$$

By using (4.18) and (4.21), it follows that, for $h \downarrow 0$,

$$\frac{A_n^{(2)}}{h^{1-\alpha} q_h^{(1)}} = \frac{\int_0^1 u^{-1-\bar{\gamma}} W(u)\, dK^{(2)}(u)}{\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u)\, du} + o_P(1)$$
$$= -a_1 \int_0^1 W(u)\, d\tilde{K}^{(2)}(u) + o_P(1)$$

and

$$\frac{A_n^{(1)} h^{1-\alpha} q_h^{(2)}}{(h^{1-\alpha} q_h^{(1)})^2} = \frac{\int_0^1 u^{-1-\bar{\gamma}} W(u)\, dK^{(1)}(u)}{\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u)\, du} \frac{\int_0^1 u^{-1-\bar{\gamma}} K^{(2)}(u)\, du}{\int_0^1 u^{-1-\bar{\gamma}} K^{(1)}(u)\, du} + o_P(1)$$
$$= a_2 \int_0^1 W(u)\, d\tilde{K}^{(1)}(u) + o_P(1),$$

because

$$a_2 = a_1^2 \int_0^1 u^{-1-\bar{\gamma}} K^{(2)}(u)\, du = \big(1 + (\gamma \wedge 0)\big) a_1.$$

Hence, by integration by parts,

$$\sqrt{nh}(\hat{\gamma}_{n,h}^{K} - \gamma_h) \xrightarrow{\mathcal{D}} \int_0^1 \big[a_0 \tilde{K}(u) + a_1 \tilde{K}^{(2)}(u) - a_2 \tilde{K}^{(1)}(u)\big]\, dW(u). \tag{4.22}$$

The assertion of the theorem follows. $\square$

The asymptotic variance depends on $\gamma$ and the choice of the kernel $K$. We tried the following three different kernels: the biweight $K(x) = \frac{15}{8}(1 - x^2)^2$,

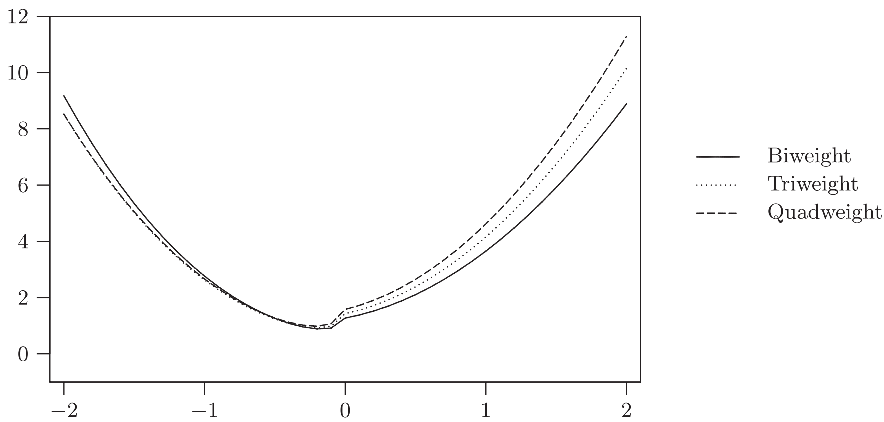


FIG. 1. *Asymptotic variances as a function of $\gamma$ for three different kernels.*

the triweight $K(x) = \frac{35}{16}(1 - x^2)^3$ and the quadweight $K(x) = \frac{315}{128}(1 - x^2)^4$. The asymptotic variances of the corresponding estimators as a function of $\gamma$ are displayed in Figure 1. It can be seen that one can reduce the variance for $\gamma > 0$ by taking higher powers of $(1 - x^2)$, but then the variance for $\gamma < 0$ increases. It seems that, among the above three estimators, the one constructed with the biweight kernel $K(x) = \frac{15}{8}(1 - x^2)^2$ has the best overall performance.

**5. Exploring the bias.** The formulation of Theorem 4.1 implies that $\hat{\gamma}_{n,h}^K$ might have asymptotic bias of the form $\sqrt{nh}(\gamma_h - \gamma)$. In Dekkers and de Haan (1993), conditions are stated that cover all possible second-order behavior of quantile functions corresponding to distribution functions that are in the domain of attraction of an extreme value distribution. Under these additional conditions, we will derive asymptotic expressions for the bias. The conditions can be formulated in the following way:

(RV1) In the case $\gamma > 0$, let $U_1(s) = \log Q(1 - s) + \gamma \log s - \log c$. Suppose that either $U_1$ or $-U_1$ eventually remains positive, as $s \downarrow 0$, and there exist $\rho > 0$ and $c > 0$ such that, for all $x > 0$,

$$\lim_{s \downarrow 0} \frac{U_1(sx)}{U_1(s)} = x^{\gamma\rho}. \tag{5.1}$$

(RV2) In the case $\gamma < 0$, let $U_2(s) = s^{\gamma}(\log Q(1) - \log Q(1 - s)) - c/Q(1)$. Suppose that either $U_2$ or $-U_2$ eventually remains positive, as $s \downarrow 0$, and there exist $\rho > 0$ and $c > 0$ such that, for all $x > 0$,

$$\lim_{s \downarrow 0} \frac{U_2(sx)}{U_2(s)} = x^{-\gamma\rho}. \tag{5.2}$$

Note that condition (RV1) states that either $U_1$ or $-U_1$ is regularly varying at 0 with index $\gamma\rho$, whereas condition (RV2) states that either $U_2$ or $-U_2$ is regularly varying at 0 with index $-\gamma\rho$. The generalized Pareto distribution satisfies conditions (RV1) and (RV2) for suitable choices of the parameters, and similarly this holds for the generalized extreme value distribution and the model considered in Hall and Welsh (1984). Other examples are the Cauchy distribution, which satisfies (RV1), and the uniform distribution, which satisfies (RV2).

The second set of conditions concerns the second-order $\Pi$-varying behavior of the quantile function.

(PV1) In the case $\gamma > 0$, suppose there exists a positive function $b_1(\cdot)$ such that, for all $x > 0$,

$$\lim_{s \downarrow 0} \frac{V_1(sx) - V_1(s)}{a_1(s)} = -\log x, \tag{5.3}$$

where $V_1(s) = \pm(\log Q(1-s) + \gamma \log s)$ and $a_1(s) = s^{-\gamma} b_1(s)/Q(1-s)$.

(PV2) In the case $\gamma = 0$, suppose there exist positive functions $b_2(\cdot)$ and $b_3(\cdot)$, with $b_2(s) \to 0$, as $s \downarrow 0$, such that for all $x > 0$,

$$\lim_{s \downarrow 0} \frac{V_2(sx) - V_2(s) + b_2(s)\log x}{b_3(s)} = -\frac{(\log x)^2}{2},$$

where $V_2(s) = \log Q(1-s)$.

(PV3) In the case $\gamma < 0$, suppose there exists a positive function $b_4(\cdot)$ such that for all $x > 0$,

$$\lim_{s \downarrow 0} \frac{V_3(sx) - V_3(s)}{a_3(s)} = -\log x, \tag{5.4}$$

where $V_3(s) = \pm s^{\gamma} (\log Q(1) - \log Q(1-s))$ and $a_3(s) = b_4(s)/Q(1)$.

Note that condition (PV1) states that either $\log Q(1-s) + \gamma \log s$ or $-(\log Q(1-s) + \gamma \log s)$ is $\Pi$-varying at 0 with auxiliary function $s^{-\gamma} b_1(s)/Q(1-s)$ and that condition (PV3) states either $s^{\gamma}(\log Q(1) - \log Q(1-s))$ or $-s^{\gamma}(\log Q(1) - \log Q(1-s))$ is $\Pi$-varying at 0 with auxiliary function $b_4(s)/Q(1)$. The generalized Pareto distribution and the generalized extreme value distribution, both with $\gamma = 0$, are examples that satisfy condition (PV2).

The following lemmas are analogous to Lemma 3.1 and will be needed to apply dominated convergence to integrals such as $\int U_i(su)/U_i(s)\, dK^{(j)}(u)$ as $s \downarrow 0$ for $i = 1, 2$ and $j = 1, 2$.

LEMMA 5.1. *Assume that conditions* (RV1) *and* (RV2) *hold. Then, for any* $\varepsilon > 0$, *there exists* $s_0 > 0$ *such that for all* $0 < s < s_0$ *and* $0 < y < 1$, *for* $\gamma > 0$,

$$(1-\varepsilon) y^{\gamma\rho + \varepsilon} < \frac{U_1(sy)}{U_1(s)} < (1+\varepsilon) y^{\gamma\rho - \varepsilon}$$

*and, for* $\gamma < 0$,

$$(1-\varepsilon)y^{-\gamma\rho+\varepsilon} < \frac{U_2(sy)}{U_2(s)} < (1+\varepsilon)y^{-\gamma\rho-\varepsilon},$$

*where* $U_1$ *and* $U_2$ *are defined in conditions* (RV1) *and* (RV2).

PROOF. The inequalities are the well-known inequalities of regularly varying functions [see, e.g., Geluk and de Haan (1987)]. □

Similar inequalities can be derived in the case of second-order Π-variation. They are stated in the next lemma, which is a reformulation of Lemma 3.5 in Dekkers, Einmahl and de Haan (1989) in terms of the quantile function.

LEMMA 5.2. *In the case* $\gamma > 0$, *assume that* (5.3) *holds for* $V_1$. *Then, for any* $\varepsilon > 0$, *there exists* $s_0 > 0$ *such that for all* $0 < s < s_0$ *and* $0 < y < 1$,

$$(1-\varepsilon)\frac{1-y^{\varepsilon}}{\varepsilon} - \varepsilon < \frac{V_1(sy) - V_1(y)}{a_1(s)} < (1+\varepsilon)\frac{y^{-\varepsilon}-1}{\varepsilon} + \varepsilon.$$

*In the case* $\gamma = 0$, *for any* $\varepsilon > 0$, *there exists* $s_0 > 0$ *such that for all* $0 < s < s_0$ *and* $0 < y < 1$,

$$\frac{(1-\varepsilon)^2 y^{\varepsilon}(\log y)^2}{2} + 2\varepsilon\log y - \varepsilon < \frac{V_2(sy) - V_2(s) + b_2(s)\log y}{b_3(s)} < \frac{(1+\varepsilon)^2 y^{-\varepsilon}(\log y)^2}{2} - 2\varepsilon\log y + \varepsilon.$$

*In the case* $\gamma < 0$, *assume that* (5.4) *holds for* $V_3$. *Then, for any* $\varepsilon > 0$, *there exists* $s_0 > 0$ *such that for all* $0 < s < s_0$ *and* $0 < y < 1$,

$$(1-\varepsilon)\frac{1-y^{\varepsilon}}{\varepsilon} - \varepsilon < \frac{V_3(sy) - V_3(s)}{a_3(s)} < (1+\varepsilon)\frac{y^{-\varepsilon}-1}{\varepsilon} + \varepsilon.$$

PROOF. In the case $\gamma \neq 0$, the inequalities are just the well-known inequalities for Π-varying functions [see, e.g., Geluk and de Haan (1987), page 27]. In the case $\gamma = 0$, the inequalities follow using Omey and Willekens (1987) to obtain an asymptotic expression for $b_2(\cdot)$ and applying the inequalities for Π-varying functions to that expression [see the proof of Lemma 3.5 in Dekkers, Einmahl and de Haan (1989)]. □

Defining

$$\lambda_{st} = \int_0^1 u^s(\log u)^t K(u)\,du, \qquad s,t \geq 0, \tag{5.5}$$

the results concerning the asymptotic bias can be formulated in the following way.

THEOREM 5.1. *Let $\gamma_h$ be given by* (4.1) *for some $\alpha > 0$. Assume that $K$ satisfies conditions* (CK1)–(CK4). *Suppose that $Q$ satisfies conditions* (RV1) *and* (RV2) *and that $h = h_n$ is such that $h_n \downarrow 0$ as $n \to \infty$. Then, as $n \to \infty$, in the case $\gamma > 0$,*

$$\gamma_h - \gamma = \mu_1 U_1(h) + \frac{\mu_2 U_1(h) + o(U_1(h))}{\mu_3 + \mu_4 U_1(h) + o(U_1(h))} + o\big(U_1(h)\big),$$

*and, in the case $\gamma < 0$,*

$$\gamma_h - \gamma = \mu_5 h^{-\gamma} U_2(h) + \mu_6 h^{-\gamma} + \frac{\mu_7 U_2(h) + o(U_2(h))}{\mu_8 + \mu_9 U_2(h) + o(U_2(h))} + o\big(h^{-\gamma} U_2(h)\big),$$

*where the functions $U_1$ and $U_2$ are defined in* (RV1) *and* (RV2) *where, using the notation for the coefficients $\lambda_{st}$ introduced in* (5.5),

$$\begin{aligned}
\mu_1 &= -\gamma\rho\lambda_{\gamma\rho,0}, & \mu_5 &= -\gamma(1+\rho)\lambda_{-\gamma(1+\rho),0},\\
\mu_2 &= \gamma\rho^2\lambda_{\gamma\rho+\alpha-1,0}, & \mu_6 &= -\gamma c\lambda_{-\gamma,0}/Q(1),\\
\mu_3 &= \lambda_{\alpha-1,0}, & \mu_7 &= \gamma\rho(1+\rho)\lambda_{\alpha-\gamma(1+\rho)-1,0},\\
\mu_4 &= -\rho\lambda_{\gamma\rho+\alpha-1,0}, & \mu_8 &= c\lambda_{\alpha-\gamma-1,0}/Q(1),\\
& & \mu_9 &= (1+\rho)\lambda_{\alpha-\gamma(1+\rho)-1,0}.
\end{aligned}$$

*Here $c$ and $\rho$ are defined as in* (RV1) *and* (RV2).

PROOF. It is sufficient to consider only the case where $U_1$ eventually remains positive and satisfies (5.1). For $i = 1, 2$, consider

$$\begin{aligned}
h^{1-\alpha} q_h^{(i)} &= \int_0^1 \log Q(1-hu)\, dK^{(i)}(u)\\
&= U_1(h) \int_0^1 \frac{U_1(hu)}{U_1(h)}\, dK^{(i)}(u) - \int_0^1 \big(\gamma \log(hu) - \log c\big)\, dK^{(i)}(u),
\end{aligned}$$

where the function $U_{(1)}$ is defined in condition (RV1). For any $\alpha > 0$ and $i = 1, 2$, we have that

$$K^{(i)}(0) = K^{(i)}(1) = 0, \tag{5.6}$$

and, for any $s, t \geq 0$ and $i = 1, 2$, we have that

$$\begin{aligned}
&\int_0^1 u^s (\log u)^t \, dK^{(i)}(u)\\
&\quad = -s\lambda_{s+\alpha-1,t} - t\lambda_{s+\alpha-1,t-1}\\
&\qquad + (i-1)\big\{s^2\lambda_{s+\alpha-1,t} + 2st\lambda_{s+\alpha-1,t-1} + t(t-1)\lambda_{s+\alpha-1,t-2}\big\}.
\end{aligned} \tag{5.7}$$

For $i = 1, 2$, write $L_i(u) = d(K^{(i)}(u))/du$. From condition (CK4), it follows that $L_i(u) = L_i^+(u) - L_i^-(u)$, where $L_i^{\pm}$ are positive and bounded. Hence, similar to the proof of Lemma 3.3, using the inequalities of Lemma 5.1 and dominated convergence, condition (RV1) yields that

$$\int_0^1 \frac{U_1(hu)}{U_1(h)}\, dK^{(i)}(u) \to \int_0^1 u^{\gamma\rho}\, dK^{(i)}(u) \qquad \text{for } i = 1, 2. \tag{5.8}$$

From (5.6)–(5.8), it follows that, for $i = 1, 2$,

$$h^{1-\alpha} q_h^{(i)} = \gamma\lambda_{\alpha-1,0} + U_1(h) \int_0^1 u^{\gamma\rho}\, dK^{(i)}(u) + o\bigl(U_1(h)\bigr).$$

The integral on the right-hand side can be evaluated by means of (5.7). Note that $\gamma_h^{(\mathrm{pos})}$ equals $q_h^{(1)}$ with $\alpha = 1$. Putting things together proves the theorem for the case $\gamma > 0$.

For the case $\gamma < 0$, it is sufficient to consider only the case where $U_2$ eventually remains positive and satisfies (5.2). Similarly, using (RV2) and (5.6), for $i = 1, 2$ we can write

$$\begin{aligned} h^{1-\alpha} q_h^{(i)} &= -h^{-\gamma} U_2(h) \int_0^1 u^{-\gamma} \frac{U_2(hu)}{U_2(h)}\, dK^{(i)}(u) - \frac{h^{-\gamma} c}{Q(1)} \int_0^1 u^{-\gamma}\, dK^{(i)}(u) \\ &= -h^{-\gamma} U_2(h) \int_0^1 u^{-\gamma(1+\rho)}\, dK^{(i)}(u) - \frac{h^{-\gamma} c}{Q(1)} \int_0^1 u^{-\gamma}\, dK^{(i)}(u) \\ &\quad + o\bigl(h^{-\gamma} U_2(h)\bigr), \end{aligned}$$

where the function $U_2$ is defined in condition (RV2) and where we have used the inequalities of Lemma 5.1, together with dominated convergence and condition (RV2). Again, the integrals on the right-hand side can be evaluated with (5.7). Hence, by putting things together this proves the theorem for the case $\gamma < 0$. $\square$

REMARK 5.1. According to condition (RV1), $|U_1|$ is regularly varying with index $\gamma\rho > 0$, so that, by Proposition 1.7.1 in Geluk and de Haan (1987), it follows that $U_1(s) \to 0$ and, similarly, $U_2(s) \to 0$. This means that, for the case $\gamma > 0$, one can write

$$\gamma_h - \gamma = c_1 U_1(h) + o\bigl(U_1(h)\bigr),$$

where $c_1 = (\mu_1\mu_3 + \mu_2)/\mu_3$, and for the case $\gamma < 0$,

$$\gamma_h - \gamma = c_2 U_2(h) + \mu_6 h^{-\gamma} + O\bigl(h^{-\gamma} U_2(h)\bigr) + o\bigl(U_2(h)\bigr),$$

where $c_2 = \mu_7/\mu_8$.

COROLLARY 5.1. *Assume the conditions of Theorem* 4.1 *and suppose that conditions* (RV1) *and* (RV2) *are satisfied. Suppose that* $h = h_n$ *is such that, as* $n \to \infty$, $h \downarrow 0$ *and*:

(i) *in the case* $\gamma > 0$, $nhU_1(h)^2 \to 0$;
(ii) *in the case* $\gamma < 0$, $nhU_2(h)^2 \to 0$ *and* $nh^{1-2\gamma} \to 0$.

*Then, as* $n \to \infty$,

$$\sqrt{nh}(\hat{\gamma}^K_{n,h} - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_K^2),$$

*with* $\sigma_K^2$ *as defined in Theorem* 4.1.

In the derivation of the asymptotic expansion of the bias under condition (PV1), we have to distinguish between the case that either $V_1(s) = \log Q(1-s) + \log s$ or $V_1(s) = -(\log Q(1-s) + \log s)$ satisfies (5.3), and similarly for asymptotic expansion of the bias under condition (PV3).

THEOREM 5.2. *Let* $\gamma_h$ *be given by* (4.1) *for some* $\alpha > 0$. *Assume that* $K$ *satisfies conditions* (CK1) *and* (CK2) *and that* $Q$ *satisfies conditions* (PV1)–(PV3). *Suppose that* $h = h_n$ *is such that, when* $n \to \infty$, $h \downarrow 0$. *Then, in the case* $\gamma > 0$,

$$\gamma_h - \gamma = \pm a_1(h) + \frac{o(a_1(h))}{\nu_1(\gamma \pm a_1(h)) + o(a_1(h))} + o\big(a_1(h)\big),$$

*where one should read* $a_1$ (*or* $-a_1$) *whenever* $V_1(s) = \log Q(1-s) + \log s$ [*or* $V_1(s) = -(\log Q(1-s) + \log s)$] *satisfies* (5.3). *In the case* $\gamma = 0$,

$$\gamma_h = b_2(h) + \nu_2 b_3(h) + \frac{\nu_3 b_3(h) + o(b_3(h))}{\nu_4 b_3(h) + \nu_1 b_2(h) + o(b_3(h))} + o\big(b_3(h)\big).$$

*In the case* $\gamma < 0$,

$$\begin{aligned}\gamma_h - \gamma = {}& \pm\nu_5 h^{-\gamma} a_3(h) + \nu_6 h^{-\gamma} V_3(h) \\ & + \frac{\pm\nu_7 a_3(h) + o(a_3(h))}{\pm\nu_8 a_3(h) + \nu_7 V_3(h) + o(a_3(h))} + o\big(h^{-\gamma} a_3(h)\big),\end{aligned}$$

*where one should read* $a_3$ (*or* $-a_3$) *whenever* $V_3(s) = s^\gamma(\log Q(1) - \log Q(1-s))$ [*or* $V_3(s) = -s^\gamma(\log Q(1) - \log Q(1-s))$] *satisfies* (5.4). *The functions* $a_1, b_2, b_3, a_3$ *and* $V_3$ *are defined in* (PV1)–(PV3) *and, using the notation for the coefficients* $\lambda_{st}$ *introduced in* (5.5),

$$\begin{aligned} \nu_1 &= \lambda_{\alpha-1,0}, & \nu_5 &= \gamma\lambda_{-\gamma,1} - \lambda_{-\gamma,0}, \\ \nu_2 &= \lambda_{0,1}, & \nu_6 &= -\gamma\lambda_{-\gamma,0}, \\ \nu_3 &= -\lambda_{\alpha-1,0}, & \nu_7 &= -\gamma\lambda_{\alpha-\gamma-1,0}, \\ \nu_4 &= \lambda_{\alpha-1,1}, & \nu_8 &= \gamma\lambda_{\alpha-\gamma-1,1} - \lambda_{\alpha-\gamma-1,0}. \end{aligned}$$

PROOF. For the case $\gamma > 0$, we only consider the case where $V_1(s) = \log Q(1-s) + \log s$ satisfies (5.3). The case $V_1(s) = -(\log Q(1-s) + \log s)$

can be handled by a similar argument. The proof is similar to that of Theorem 5.1. Using (5.6) and (5.7), we have, for $i = 1, 2$,

$$h^{1-\alpha} q_h^{(i)} = a_1(h) \int_0^1 \frac{V_1(hu) - V_1(h)}{a_1(h)} \, dK^{(i)}(u) + \gamma \lambda_{\alpha-1,0},$$

where the function $a_1$ is defined in condition (PV1). Again, writing $d(K^{(i)}(u))/du = L_i(u) = L_i^+(u) - L_i^-(u)$, with $L_i^{\pm}$ positive and bounded, we use similar arguments as in the proofs of Lemma 3.3 and Theorem 5.1. Using the inequalities of Lemma 5.2 and dominated convergence, from condition (PV1) it follows that, for $i = 1, 2$,

$$\int_0^1 \frac{V_1(hu) - V_1(h)}{a_1(h)} \, dK^{(i)}(u) \to \int_0^1 (-\log u) \, dK^{(i)}(u) = \lambda_{\alpha-1,0}$$

as $h \downarrow 0$, where we again used (5.7). Combining things proves the theorem for $\gamma > 0$.

In the case $\gamma = 0$, using (5.6), for $i = 1, 2$ we can write

$$\begin{aligned} h^{1-\alpha} q_h^{(i)} &= b_3(h) \int_0^1 \frac{V_2(hu) - V_2(h) + b_2(h) \log u}{b_3(h)} \, dK^{(i)}(u) \\ &\quad - b_2(h) \int_0^1 \log u \, dK^{(i)}(u), \end{aligned}$$

where the functions $V_2, b_2$ and $b_3$ are defined in condition (PV2). By a similar argument, using the inequalities of Lemma 5.2 and dominated convergence, we have from condition (PV2) that, for $i = 1, 2$,

$$\int_0^1 \frac{V_2(hu) - V_2(h) + b_2(h) \log u}{b_3(h)} \, dK^{(i)}(u) \to -\frac{1}{2} \int_0^1 (\log u)^2 \, dK^{(i)}(u).$$

By means of (5.7), we find that, for $i = 1, 2$,

$$h^{1-\alpha} q_h^{(i)} = b_3(h) \{ \lambda_{\alpha-1,1} - (i-1) \lambda_{\alpha-1,0} \} + b_2(h) \lambda_{\alpha-1,0} + o(b_3(h)).$$

Putting things together proves the theorem for $\gamma = 0$.

For the case $\gamma < 0$, we only consider the case where $V_3(s) = s^{\gamma} (\log Q(1) - \log Q(1-s))$ satisfies (5.4). The case $V_3(s) = -s^{\gamma} (\log Q(1) - \log Q(1-s))$ can be handled by a similar argument. Using (5.6), for $i = 1, 2$, write

$$\begin{aligned} h^{1-\alpha} q_h^{(i)} &= -h^{-\gamma} a_3(h) \int_0^1 u^{-\gamma} \frac{V_3(hu) - V_3(h)}{a_3(h)} \, dK^{(i)}(u) \\ &\quad - h^{-\gamma} V_3(h) \int_0^1 u^{-\gamma} \, dK^{(i)}(u), \end{aligned}$$

where the function $a_3$ is defined in condition (PV3). As before, using the inequalities of Lemma 5.2 together with dominated convergence, from condition (PV3) we obtain

$$\int_0^1 u^{-\gamma} \frac{V_3(hu) - V_3(h)}{a_3(h)} \, dK^{(i)}(u) \to -\int_0^1 u^{-\gamma} \log u \, dK^{(i)}(u)$$

as $h \downarrow 0$. If we evaluate the integrals by means of (5.7), we find that, for $i = 1, 2$,

$$\begin{aligned} h^{1-\alpha} q_h^{(i)} = h^{-\gamma} a_3(h)\{&\gamma \lambda_{\alpha-\gamma-1,1} - \lambda_{\alpha-\gamma-1,0} \\ &+ (i-1)(\gamma^2 \lambda_{\alpha-\gamma-1,1} - 2\gamma \lambda_{\alpha-\gamma-1,0})\} \\ &- h^{-\gamma} V_3(h)\{\gamma \lambda_{\alpha-\gamma-1,0} + (i-1)\gamma^2 \lambda_{\alpha-\gamma-1,0}\} + o(h^{-\gamma} a_3(h)). \end{aligned}$$

Putting things together proves the theorem for $\gamma < 0$. □

COROLLARY 5.2. *Assume the conditions of Theorem* 5.2 *and suppose that* (PV1)–(PV3) *are satisfied. Suppose that* $h = h_n$ *is such that, as* $n \to \infty$, $h \downarrow 0$ *and in the case* $\gamma > 0$,

$$nha_1(h)^2 \to 0,$$

*in the case* $\gamma = 0$,

$$nhb_2(h)^2 \to 0 \quad \text{and} \quad nh\left(\frac{b_3(h)}{b_2(h)}\right)^2 \to 0,$$

*and in the case* $\gamma < 0$,

$$nh^{1-2\gamma} V_3(h)^2 \to 0 \quad \text{and} \quad nh\left(\frac{a_3(h)}{V_3(h)}\right)^2 \to 0.$$

*Then* $\sqrt{nh}(\gamma_h - \gamma) \to 0$ *as* $n \to \infty$.

Note that the condition for the case $\gamma > 0$ and the second condition for the case $\gamma < 0$ resemble the conditions on the parameter $k$ in the case of the moment estimator as defined in Dekkers, Einmahl and de Haan (1989).

**6. Comparison with other estimators.** To illustrate the finite-sample behavior of our estimator, we present some results from a small simulation study. We will compare our estimator to the moment estimator of Dekkers, Einmahl and de Haan (1989), the (quasi) MLE of Smith (1987) and the more recent proposals of Beirlant, Vynckier and Teugels (1996) and Drees (1995). For easy reference, we restate their definitions. The moment estimator is given by

$$\hat{\gamma}_{n,k}^{\mathrm{M}} = M_{n,k}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_{n,k}^{(1)})^2}{M_{n,k}^{(2)}}\right)^{-1},$$

where, for $r = 1, 2$,

$$M_{n,k}^{(r)} = \frac{1}{k}\sum_{i=1}^{k}\bigl(\log X_{(n-i+1)} - \log X_{(n-k)}\bigr)^r.$$

Note that $k$ is the number of largest order statistics from the sample used to calculate the moment estimator. The (quasi) maximum likelihood estimator $\hat{\gamma}_{n,k}^{\mathrm{ML}}$ is defined using the excesses $Y_i = X_j - u_n$, where $X_j$ is the $i$th exceedance over the threshold $u_n$ tending the upper endpoint of the distribution that generated the sample. Assuming that these excesses are distributed as a sample of a generalized Pareto distribution with parameters $\gamma$ and $\sigma(u_n)$, the estimator is defined by maximizing the likelihood of $Y_1, \dots, Y_N$, where $N$ is the number of excesses over the threshold $u_n$. In our simulations, we took $u_n = X_{(n-k)}$. Note that, again, $k$ is the number of upper order statistics used to calculate the estimator. The adjusted Hill estimator from Beirlant, Vynckier and Teugels (1996) is defined as

$$\hat{\gamma}_{n,h}^{\mathrm{AH}} = \frac{1}{k}\sum_{i=1}^{k}\log UH_{i,k} - \log UH_{k+1,n},$$

where

$$UH_{l,n} = X_{(n-l)}\left(\frac{1}{l}\sum_{j=1}^{l}\log X_{(n-j+1)} - \log X_{(n-l)}\right).$$

For the multistage procedure that leads to the refined Pickands estimator $\hat{\gamma}_{n,h}^{\mathrm{RP}}$, we refer to Drees (1995). For our kernel estimator, we took $\alpha = 0.6$ (to ensure asymptotic normality) and the biweight kernel $K$ defined by $K(x) = \frac{15}{8}(1 - x^2)^2$ for $0 \le x \le 1$.

We start by presenting a plot of the above methods used to estimate the extreme value index of a real-life data set. The data concerned were obtained from Lobith, the village where the first inhabitants of the Netherlands (the “Bataviers”) are supposed to have entered on rafts along the Rhine River. They represent the peaks in the water discharges at that particular place along the Rhine. During the period 1901–1991, the maximum water discharge was measured on a daily basis. These maxima were plotted against time and only those maxima above a certain threshold and at least a fortnight apart were recorded. Whenever several values appeared above the threshold but within a fortnight of each other, the maximum of these values was recorded. This resulted in a data set of 155 measurements. To be able to compare the estimators, we will plot each estimator as a function of the fraction of order statistics used to calculate the estimator. That is, we will use $k = \lfloor nh \rfloor$ and plot each estimator as a function of $h \in (0, 1)$. The plots are given in Figure 2. All estimators have a kind of dip near 0.15. This is caused by a gap between the largest order statistics and the other sample values. The refined Pickands estimator $\hat{\gamma}_{n,k}^{\mathrm{RP}}$ reduces the jumpy behavior of the original Pickands estimator, but is still

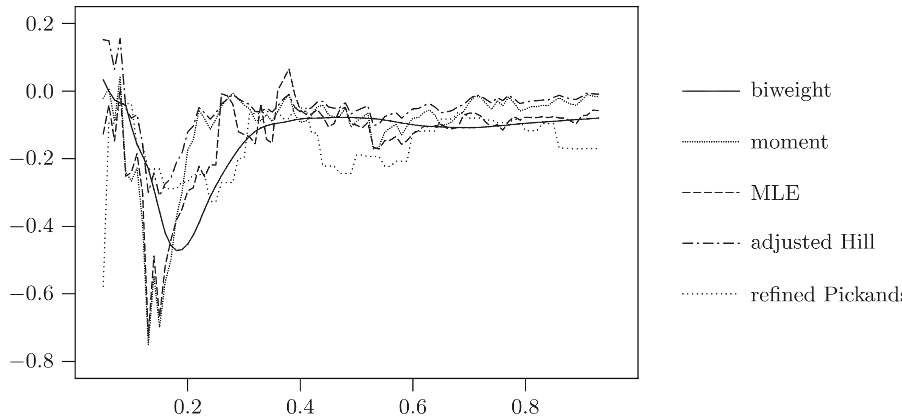


FIG. 2. *Estimates of the extreme value index for the Lobith data.*

less stable than all other methods. One striking feature of the kernel estimator is its smoothness: whereas the other estimators behave rather erratically as a function of $h$, the kernel estimator behaves very smoothly. A major advantage of this feature is that the exact choice of the bandwidth $h$ to be used is not as crucial as the exact choice of the $k$ in the other estimators: increasing $k$ by 1 can seriously change the value of the estimator. Changing $h$ by $1/n$, however, does not change the kernel estimator too much. Indeed, only an approximately optimal bandwidth would produce an estimate almost as good as the estimate using the exact optimal bandwidth.

We also compared the stability of the estimators as a function of $h$ for a single sample of size $n = 100$ from three distributions corresponding to $\gamma$ negative, zero and positive: a uniform distribution on the interval $(2, 5)$, the exponential distribution with mean 1 and a distribution derived from the Hall model with extreme value index $\gamma = 1/3$ [see, e.g., Hall and Welsh (1984)]. The Hall model that we use corresponds to the distribution function

$$F(x) = 1 - \frac{2}{x^3}\left(1 - \frac{1}{2x^3}\right), \qquad x \geq 1.$$

The results are shown in Figure 3. First of all, each estimator is quite close to the true value of the extreme value index, considering the small sample size. The behavior of the estimators is similar to that in Figure 2.

Finally, we compared the simulated mean squared error of the $\hat{\gamma}_{n,k}^{\mathrm{AH}}$, $\hat{\gamma}_{n,k}^{\mathrm{M}}$ and $\hat{\gamma}_{n,k}^{\mathrm{ML}}$ with our kernel estimator, for a sample of size $n = 100$ from the same three distributions mentioned above. The results of 1000 samples of size $n = 100$ are displayed in Figure 4. The kernel estimator and the adjusted Hill estimator outperform the other estimators for $\gamma < 0$. For $\gamma > 0$, all estimators, except the refined Pickands estimator, behave similarly. For $\gamma = 0$, the kernel

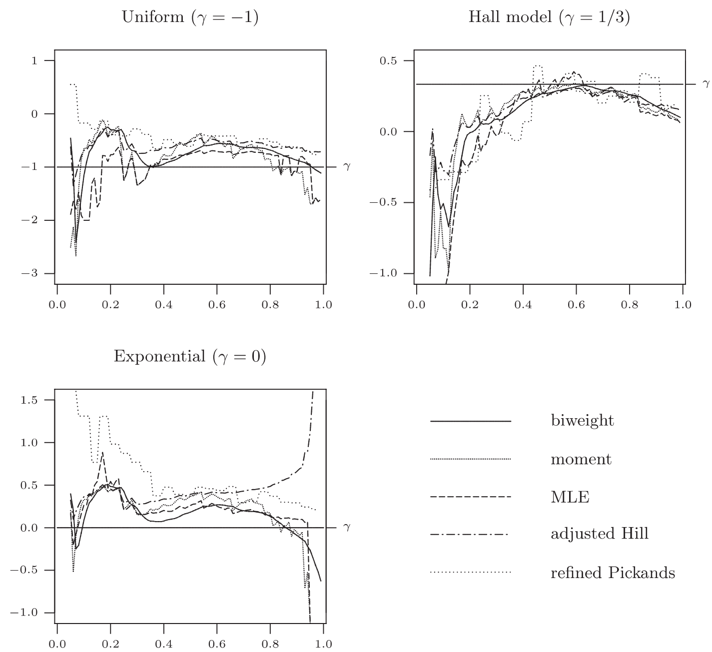


FIG. 3. *Estimates of $\gamma$ for a sample of size $n = 100$.*

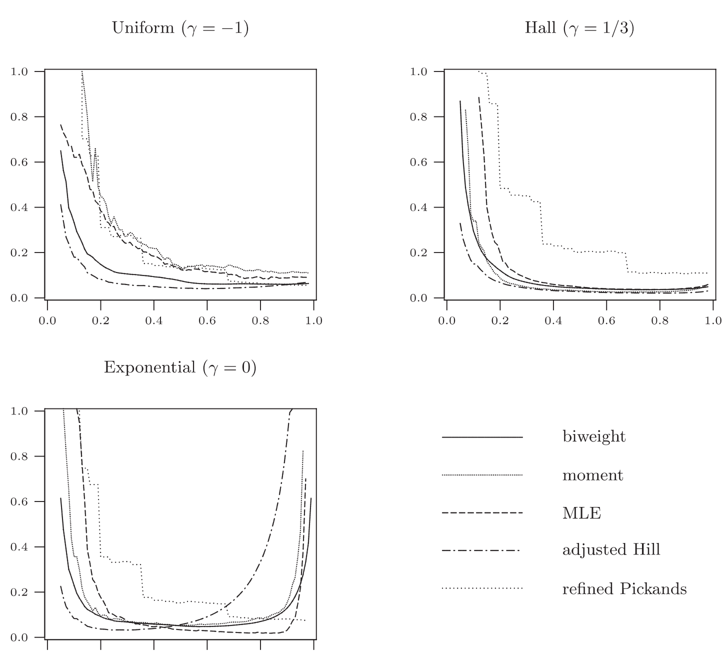


FIG. 4. *Simulated mean squared error for $n = 100$.*

estimator behaves similarly to $\hat{\gamma}_{n,k}^{\mathrm{ML}}$. Apart from $\hat{\gamma}_{n,k}^{\mathrm{RP}}$, all estimators reach the similar minimum values for the mean squared error.

We conclude that the kernel estimator behaves more smoothly as a function of the bandwidth than the moment-type estimators and quasi-maximum likelihood estimator as a function of the fraction of order statistics and that its mean squared error attains values of the same order as the mean squared error of the other estimators.

**7. Automatic bandwidth choice.** One of the things that remains to be discussed is the (automatic) choice of the bandwidth. This topic is the subject of a manuscript in preparation. But because of the importance of this issue, we still want to discuss the matter here.

A bootstrap-based approach to the choice of number of largest order statistics in moment-type estimators is presented in, for example, Draisma, de Haan, Peng and Pereira (1999) and Danielsson, de Haan, Peng and de Vries (2001). The basic difficulty in a bootstrap-based approach is the fact that in the empirical (nonparametric) bootstrap the bias is not adequately estimated in the evaluation of the bootstrap mean squared error, unless one performs bootstrapping with vanishing sample fractions. This fact has been clearly pointed out in, for example, Hall (1990), where the idea of bootstrapping with vanishing sample fractions was introduced.

Now, with our kernel-type estimators, we can follow a similar approach as in Draisma, de Haan, Peng and Pereira (1999) and Danielsson, de Haan, Peng and de Vries (2001). In these papers, the difference of two moment-type estimators is used for dealing with the difficulty of estimating the bias. Instead, we can use two estimators $\hat{\gamma}_{n,h}^{K_1}$ and $\hat{\gamma}_{n,h}^{K_2}$ based on two different kernels, say the biweight kernel

$$K_1(u) = \tfrac{15}{8}(1-u^2)^2 \mathbb{1}_{[0,1]}(u),$$

and the triweight kernel

$$K_2(u) = \tfrac{35}{16}(1-u^2)^3 \mathbb{1}_{[0,1]}(u).$$

We first present the method, outlined in Draisma, de Haan, Peng and Pereira [(1999, page 368], as it would apply to our kernel-type estimators. Let $X_1, \dots, X_n$ be a sample from a distribution for which we want to estimate the extreme value index.

STEP 1. For a sample size $n_1 \ll n$, select a bootstrap sample $X_1^*, \dots, X_{n_1}^*$ from the original sample and compute the estimates $(\hat{\gamma}_{n_1,h}^{K_1})^*$ and $(\hat{\gamma}_{n_1,h}^{K_2})^*$ defined as in (1.2), with the order statistics $X_{(i)}$ replaced by the order statistics $X_{(i)}^*$ of the bootstrap sample. Next, compute

$$\delta_{n_1,h}^* = \big(\hat{\gamma}_{n_1,h}^{K_1}\big)^* - \big(\hat{\gamma}_{n_1,h}^{K_2}\big)^*. \tag{7.1}$$

STEP 2. Repeat this procedure $r$ times independently, yielding a sequence $\delta^*_{n_1,h,1},\dots,\delta^*_{n_1,h,r}$. Then compute

$$\widehat{\mathrm{MSE}}^*(\delta^*_{n_1,h}) = \frac{1}{r}\sum_{i=1}^{r}(\delta^*_{n_1,h,i})^2,$$

which is an estimate of the bootstrap mean squared error of $\delta^*_{n_1,h}$.

STEP 3. Compute

$$h^*(n_1) \stackrel{\mathrm{def}}{=} \arg\min_h \widehat{\mathrm{MSE}}^*(\delta^*_{n_1,h}). \tag{7.2}$$

In practice, one would compute $\widehat{\mathrm{MSE}}^*(\delta^*_{n_1,h})$ on a grid of values of $h_i$, say, with distance 0.01 between successive values (the exact distance might be chosen to be dependent on the sample size $n_1$), and then take for $h^*(n_1)$ the minimizer of the values $\widehat{\mathrm{MSE}}^*(\delta^*_{n_1,h_i})$.

STEP 4. Repeat steps 1 to 3 independently with $n_1$ replaced by $n_2 = \lfloor n_1^2/n \rfloor$. This yields a value $h^*(n_2)$, defined by

$$h^*(n_2) \stackrel{\mathrm{def}}{=} \arg\min_h \widehat{\mathrm{MSE}}^*(\delta^*_{n_2,h}).$$

STEP 5. Estimate the optimal bandwidth $\hat{h}_{n,\mathrm{opt}}$ by

$$\hat{h}_{n,\mathrm{opt}} = c\big(h^*(n_1), h^*(n_2)\big)\frac{h^*(n_1)^2}{h^*(n_2)}, \tag{7.3}$$

where $c(h_1, h_2)$ is a function of $h_1$ and $h_2$, depending on the kernels $K_1$ and $K_2$ and the sample sizes $n_1$ and $n_2$.

Next, we discuss why this procedure would "work" for our kernel-type estimator, for example, under the second-order condition used in Danielsson, de Haan, Peng and de Vries (2001). Note that this is our (RV1) condition of Section 5. If $n_1 = O(n^{1-\varepsilon})$, for some $\varepsilon \in (0,1)$, then, using Theorem 4.1, we have

$$\hat{\gamma}^{K_1}_{n_1,h} - \hat{\gamma}^{K_2}_{n_1,h} = \gamma^{K_1}_h - \gamma^{K_2}_h + \frac{D_{n_1,h}}{\sqrt{n_1 h}}, \tag{7.4}$$

where $D_{n_1,h}$ has a limiting normal distribution with mean 0, and where, according to (4.1)–(4.3), for $i = 1, 2$,

$$\gamma^{K_i}_h = \int_0^1 \log Q(1-hu)\, d\big(uK_i(u)\big) + \frac{\int_0^1 \log Q(1-hu)\, dK_i^{(2)}(u)}{\int_0^1 \log Q(1-hu)\, dK_i^{(1)}(u)} - 1.$$

This means that the random variable $\delta^*_{n_1,h}$, defined by (7.1), has an expansion of the form

$$\delta^*_{n_1,h} = \hat{\gamma}^{K_1}_{n,h} - \hat{\gamma}^{K_2}_{n,h} + \frac{D^*_{n_1,h}}{\sqrt{n_1 h}} + O_p\left(\frac{1}{\sqrt{nh}}\right), \tag{7.5}$$

where the conditional distribution of $D^*_{n_1,h}$, given the original sample $X_1, \ldots, X_n$, is again asymptotically normal as $n_1 \to \infty$ under the conditions on $n_1$ and $h$ given in Theorem 4.1 (with $n$ replaced by $n_1$).

Now, as an example, consider the model in Hall and Welsh (1984) with $\gamma > 0$. This corresponds to a function $\phi$, as defined in (2.1), with expansion

$$\phi(s) = \gamma + cs^\tau + o(s^\tau), \qquad s \downarrow 0, \tag{7.6}$$

for some $\tau > 0$. Then we get

$$\gamma^{K_1}_h - \gamma^{K_2}_h = c_{K_1,K_2} h^\tau + o(h^\tau), \tag{7.7}$$

where $c_{K_1,K_2}$ only depends on the kernels $K_1$ and $K_2$, the constant $c$ in (7.6) and possibly the parameters $\gamma$ and $\tau$. So we get the following expansion for $\delta^*_{n_1,h}$ in (7.5):

$$\begin{aligned} \delta^*_{n_1,h} &= c_{K_1,K_2} h^\tau + \frac{D^*_{n_1,h}}{\sqrt{n_1 h}} + o(h^\tau) + O_p\big(1/\sqrt{nh}\big) \\ &= c_{K_1,K_2} h^\tau + \frac{D^*_{n_1,h}}{\sqrt{n_1 h}} + o(h^\tau) + o_p\big(1/\sqrt{n_1 h}\big), \end{aligned} \tag{7.8}$$

using $n_1/n \to 0$ in the last step. Comparing (7.8) with (7.4) and (7.7) means that the bootstrap mean squared error of $\delta^*_{n_1,h}$ has the same asymptotic behavior as the real mean squared error $\mathrm{MSE}(\hat{\gamma}^{K_1}_{n_1,h} - \hat{\gamma}^{K_2}_{n_1,h})$, implying that the minimizer $h^*(n_1)$, as defined in (7.2), will (in probability) be asymptotically equivalent to the minimizer $h^{K_1-K_2}_{n_1,\mathrm{opt}}$ of $\mathrm{MSE}(\hat{\gamma}^{K_1}_{n_1,h} - \hat{\gamma}^{K_2}_{n_1,h})$.

To illustrate the procedure for finding the optimal $h$ in the model (7.6), we present only the computations for the positive part $\hat{\gamma}^{(\mathrm{pos})}_{n,h}$ of our estimator which is the CDM estimator proposed in Csörgő, Deheuvels and Mason (1985). The procedure for the full estimator is similar, but just involves more constants. It turns out that in the model (7.6) we only have to perform the bootstrap samples of size $n_1$ and we do *not* need to perform the second experiment with the smaller sample size $n_2 = \lfloor n_1^2/n \rfloor$.

If we write $\tau = \gamma\rho$, then, for the model (7.6), similar to the expressions obtained in Theorems 4.1 and 5.1, the asymptotic bias of $\hat{\gamma}^{(\mathrm{pos})}_{n,h}$ is given by $\mu_1(\tau) c h^\tau$, where

$$\mu_1(\tau) = -\tau\lambda_{\tau,0} = -\tau \int_0^1 u^\tau K(u)\, du,$$

and the limiting variance of $\hat{\gamma}_{n,h}^{(\text{pos})}$ is given by

$$\sigma_{\widetilde{K}}^2 = \gamma^2 \int_0^1 \widetilde{K}(u)^2\, du,$$

using the notation introduced in Theorems 4.1 and 5.1. Minimizing the expression

$$\text{MSE}(\hat{\gamma}_{n,h}^K) = \frac{\sigma_{\widetilde{K}}^2}{nh} + \mu_1(\tau)^2 c^2 h^{2\tau}$$

as a function of $h$ yields the theoretically (asymptotically) optimal $h$ for sample size $n$,

$$h_{n,\text{opt}}^K = \left\{ \frac{\sigma_{\widetilde{K}}^2}{2c^2\tau\, \mu_1(\tau)^2 n} \right\}^{1/(1+2\tau)}. \tag{7.9}$$

For the biweight kernel $K_1$, we get $\sigma_{\widetilde{K}_1}^2 = 10\gamma^2/7$ and $\mu_1(\tau) = -15\tau/((1+\tau)(3+\tau)(5+\tau))$, so that (7.9) becomes

$$h_{n,\text{opt}}^{K_1} = \left\{ \frac{\gamma^2(1+\tau)^2(3+\tau)^2(5+\tau)^2}{315c^2\tau^3 n} \right\}^{1/(1+2\tau)}. \tag{7.10}$$

Now, if we do the same computation for the difference of two kernels $K_1$ and $K_2$, minimizing $\text{MSE}(\hat{\gamma}_{n_1,h}^{K_1} - \hat{\gamma}_{n_1,h}^{K_2})$ as a function of $h$, we get, for the asymptotically optimal $h_{n,\text{opt}}^{K_1-K_2}$,

$$h_{n,\text{opt}}^{K_1-K_2} = \left\{ \frac{\sigma_{\widetilde{K}_1-\widetilde{K}_2}^2}{2c^2\tau\overline{\mu}_1(\tau)^2 n} \right\}^{1/(1+2\tau)},$$

where

$$\sigma_{\widetilde{K}_1-\widetilde{K}_2}^2 = \gamma^2 \int_0^1 \{\widetilde{K}_1(u) - \widetilde{K}_2(u)\}^2\, du$$

and

$$\overline{\mu}_1(\tau) = -\tau \int_0^1 u^\tau \{K_1(u) - K_2(u)\}\, du.$$

For the biweight kernel $K_1$ and the triweight kernel $K_2$, we get $\sigma_{\widetilde{K}_1-\widetilde{K}_2}^2 = 30\gamma^2/1001$ and $\overline{\mu}_1(\tau) = -15\tau^2/((1+\tau)(3+\tau)(5+\tau)(7+\tau))$, implying

$$h_{n,\text{opt}}^{K_1-K_2} = \left\{ \frac{\gamma^2(1+\tau)^2(3+\tau)^2(5+\tau)^2(7+\tau)^2}{15015c^2\tau^5 n} \right\}^{1/(1+2\tau)}. \tag{7.11}$$

Combining (7.10) and (7.11) yields

$$h_{n,\mathrm{opt}}^{K_1} = \left\{ \frac{143\tau^2}{3(7+\tau)^2} \right\}^{1/(1+2\tau)} h_{n,\mathrm{opt}}^{K_1-K_2}. \tag{7.12}$$

Applying (7.11) to sample sizes $n$ and $n_1$ gives

$$h_{n,\mathrm{opt}}^{K_1-K_2} = \left( \frac{n_1}{n} \right)^{1/(1+2\tau)} h_{n_1,\mathrm{opt}}^{K_1-K_2}. \tag{7.13}$$

Combining this with (7.12), we find

$$h_{n,\mathrm{opt}}^{K_1} = \left\{ \frac{143 n_1 \tau^2}{3n(7+\tau)^2} \right\}^{1/(1+2\tau)} h_{n_1,\mathrm{opt}}^{K_1-K_2}. \tag{7.14}$$

Hence, if we have a bootstrap estimate of $h_{n_1,\mathrm{opt}}^{K_1-K_2}$, the last step is the estimation of $\tau$. Draisma, de Haan, Peng and Pereira (1999) propose the following estimator (here interpreted for our situation):

$$\hat{\tau} = -\frac{\log n_1 + \log h^*(n_1)}{2 \log h^*(n_1)}, \tag{7.15}$$

where $h^*(n_1)$ is the bootstrap estimate of $h_{n_1,\mathrm{opt}}^{K_1-K_2}$, as defined in (7.2). Since, indeed,

$$\begin{aligned} \frac{\log n_1 + \log h^*(n_1)}{-2 \log h^*(n_1)} &= \frac{\log n_1 - \{1+2\tau\}^{-1} \log n_1 + O_p(1)}{2\{1+2\tau\}^{-1} \log n_1 + O_p(1)} \\ &= \tau + O_p\left( \frac{1}{\log n_1} \right), \end{aligned}$$

this is also a consistent estimate of $\tau$ in our situation.

According to (7.14), in the model (7.6) we only need a bootstrap estimate for $h_{n_1,\mathrm{opt}}^{K_1-K_2}$. Nevertheless, if we apply (7.11) to sample sizes $n_1$ and $n_2 = n_1^2/n$, similar to (7.13) we find that

$$\left( \frac{n_1}{n} \right)^{1/(1+2\tau)} = \left( \frac{n_2}{n_1} \right)^{1/(1+2\tau)} = \frac{h_{n_1,\mathrm{opt}}^{K_1-K_2}}{h_{n_2,\mathrm{opt}}^{K_1-K_2}},$$

so that from (7.14) we find

$$h_{n,\mathrm{opt}}^{K_1} = \left\{ \frac{143\tau^2}{3(7+\tau)^2} \right\}^{1/(1+2\tau)} \frac{\{h_{n_1,\mathrm{opt}}^{K_1-K_2}\}^2}{h_{n_2,\mathrm{opt}}^{K_1-K_2}}.$$

Plugging in (7.15) results in the following expression for the bootstrap estimate for $h_{n,\mathrm{opt}}^{K_1}$:

$$\left\{\frac{143\{\log n_1 + \log h^*(n_1)\}^2}{3\{\log n_1 - 13\log h^*(n_1)\}^2}\right\}^{-\log h^*(n_1)/\log n_1} \frac{(h^*(n_1))^2}{h^*(n_2)}, \tag{7.16}$$

which is of the form (7.3) and is similar to the expressions in Draisma, de Haan, Peng and Pereira (1999) and Danielsson, de Haan, Peng and de Vries (2001).

Note, however, that if we assume the model (7.6), we do not have to do the second bootstrap experiment with bootstrap sample size $n_2$, and that, using (7.14), we can estimate the asymptotically optimal bandwidth by

$$\left\{\frac{143 n_1\{\log n_1 + \log h^*(n_1)\}^2}{3n\{\log n_1 - 13\log h^*(n_1)\}^2}\right\}^{-\log h^*(n_1)/\log n_1} h^*(n_1), \tag{7.17}$$

only involving the bandwidth $h^*(n_1)$.

Our simulation experiments showed that the bootstrap method worked rather well for determining the asymptotically optimal bandwidth $h_{n_1,\mathrm{opt}}^{K_1-K_2}$ and that the bottleneck of the whole procedure is the estimation of $\tau$ (which is also the case for the approach using moment estimators, although our impression is that there the bootstrap method seems to work somewhat less well, possibly as a result of the nonsmooth dependence on the sample fraction). The estimator (7.15) may have a large bias, because the estimate can be rather far from its target value, even when evaluated at the theoretically optimal bandwidth. Moreover, it only converges at logarithmic speed.

We therefore propose another estimate, which is more in line with the methods of the present paper. An estimate of the parameter $\tau$ can be based on the following relation which holds, at least in a (Schwarz) distributional sense, in the model (7.6):

$$\tau = 1 + \lim_{s\downarrow 0} \frac{s^{1+\alpha}\phi''(s)}{s^{\alpha}\phi'(s)}, \qquad \alpha \ge 0. \tag{7.18}$$

Here we introduce only the differentiability of $\phi$ for the motivation of our estimator; the proof of its consistency does not require the differentiability of $\phi$, just as in our estimate of $\gamma$. So we can estimate $\tau$ in a similar way as we estimated the possibly negative $\gamma$ in the general case, that is, by a ratio of kernel estimators. The proposed estimator for $\tau$ is

$$\hat{\tau}_{n,h} = 1 + \frac{\hat{p}_{n,h}^{(2)}}{\hat{p}_{n,h}^{(1)}}, \tag{7.19}$$

where $\hat{p}_{n,h}^{(i)}$, $i = 1, 2$, are defined for some $\alpha > 0$ by

$$\begin{aligned}\hat{p}_{n,h}^{(1)} &= \int_0^h u \frac{d}{du}\big(u^\alpha K_h(u)\big)\, d\log Q_n(1-u)\\ &= -\sum_{i=1}^{n-1}\Big[u\frac{d}{du}\big(u^\alpha K_h(u)\big)\Big]_{u=i/n}\big\{\log X_{(n-i+1)} - \log X_{(n-i)}\big\}\end{aligned} \tag{7.20}$$

and

$$\begin{aligned}\hat{p}_{n,h}^{(2)} &= -\int_0^h u \frac{d^2}{du^2}\big(u^{1+\alpha} K_h(u)\big)\, d\log Q_n(1-u)\\ &= \sum_{i=1}^{n-1}\Big[u\frac{d^2}{du^2}\big(u^{1+\alpha} K_h(u)\big)\Big]_{u=i/n}\big\{\log X_{(n-i+1)} - \log X_{(n-i)}\big\},\end{aligned} \tag{7.21}$$

respectively. Note the similarity to the definitions of $\hat{q}_{n,h}^{(1)}$ and $\hat{q}_{n,h}^{(2)}$ by (2.6) and (2.8), but also note that we have to take one extra derivative to get hold of the second-order parameter $\tau$. As in the definition of $\hat{q}_{n,h}^{(i)}$, we have some freedom in the choice of the parameter $\alpha$ in these expressions.

Estimator (7.19) will be asymptotically normal and will have polynomial rate of convergence under conditions that are similar to conditions proposed in the recent literature on moment estimators of $\tau$ in, for example, Gomes, de Haan and Peng (2003) and Fraga Alves, de Haan and Lin (2003) (our $\tau$ is $-\rho$ in their notation), in contrast with the estimator (7.15), which only has a logarithmic speed of convergence. Simulations show that the difference in smoothness of the dependence on the bandwidth of the estimator (7.19) with respect to the dependence on the sample fraction of the moment estimators, proposed in these papers, is even more striking than the corresponding difference in smoothness in the estimation of $\gamma$ between moment estimators and the kernel estimators, discussed above. Since the estimator (7.19) in a sense deals with a third derivative of the logarithm of the quantile function (although, as noted above, we do not need to assume differentiability), it comes as no surprise that the optimal bandwidths for $\hat{\tau}_{n,h}$ are larger than those for $\hat{\gamma}_{n,h}$. This is in accordance with the findings reported in Gomes, de Haan and Peng (2003) and Fraga Alves, de Haan and Lin (2003) for their moment estimators of $\tau$.

As mentioned in the beginning of this section, a more detailed treatment of the automatic bandwidth choice for the full kernel estimators, introduced in the present paper, will be given in a sequel to the present paper. The research on automatic selection of sample fractions for moment-type estimators is rather intensive at present. We have followed the bootstrap approach for our kernel estimators, but we should mention that for moment-type estimators other (more or less) automatic methods have also been suggested [see, e.g., Drees and Kaufmann (1998) and Matthys and Beirlant (2000)].

## REFERENCES

BEIRLANT, J., VYNCKIER, P. and TEUGELS, J. L. (1996). Excess functions and estimation of the extreme-value index. *Bernoulli* **2** 293–318.

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

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

CSÖRGŐ, S. and VIHAROS, L. (1997). Asymptotic normality of least-squares estimators of tail indices. *Bernoulli* **3** 351–370.

DANIELSSON, J., DE HAAN, L., PENG, L. and DE VRIES, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.* **76** 226–248.

DEKKERS, A. L. M. and DE HAAN, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal.* **47** 173–195.

DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855.

DRAISMA, G., DE HAAN, L., PENG, L. and PEREIRA, T. T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2** 367–404.

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

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

FRAGA ALVES, M. I., DE HAAN, L. and LIN, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods Statist.* To appear.

GELUK, J. L. and DE HAAN, L. (1987). *Regular Variation*, *Extensions and Tauberian Theorems*. Mathematical Centre Tracts **40**. Centre for Mathematics and Computer Science, Amsterdam.

GOMES, M. I., DE HAAN, L. and PENG, L. (2003). Semi-parametric estimation of the second order parameter—asymptotic and finite sample behaviour. *Extremes*. To appear.

DE HAAN, L. and PEREIRA, T. (1999). Estimating the index of a stable distribution. *Statist. Probab. Lett.* **41** 39–55.

HALL, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32** 177–203.

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

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

KRATZ, M. and RESNICK, S. I. (1996). The QQ-estimator and heavy tails. *Comm. Statist. Stochastic Models* **12** 699–724.

MASON, D. M. (1982). Some characterizations of almost sure bounds for weighted multidimensional empirical distributions and a Glivenko–Cantelli theorem for sample quantiles. *Z. Wahrsch. Verw. Gebiete* **59** 505–513.

MATTHYS, G. and BEIRLANT, J. (2000). Adaptive threshold selection in tail index estimation. In *Extremes and Integrated Risk Management* (P. Embrechts, ed.) 37–49. Risk Books, London.

OMEY, E. and WILLEKENS, E. (1987). $\pi$-variation with remainder. *J. London Math. Soc. (2)* **37** 105–118.

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

SCHULTZE, J. and STEINEBACH, J. (1996). On least squares estimates of an exponential tail coefficient. *Statist. Decisions* **14** 353–372.

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

WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45** 73–88.

DE WOLF, P. P. (1999). Estimating the extreme value index–Tale of tails. Ph.D. thesis, Delft Univ. Technology.

P. GROENEBOOM
H. P. LOPUHAÄ
DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SYSTEMS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS
E-MAIL: p.groeneboom@its.tudelft.nl
h.p.lopuhaa@its.tudelft.nl

P. P. DE WOLF
STATISTICS NETHERLANDS
THE NETHERLANDS
E-MAIL: pwof@cbs.nl