The Annals of Applied Probability
2000, Vol. 10, No. 4, 1274–1301

WEIGHTED APPROXIMATIONS OF TAIL PROCESSES FOR β-MIXING RANDOM VARIABLES[1]

BY HOLGER DREES

University of Cologne

While the extreme value statistics for i.i.d data is well developed, much less is known about the asymptotic behavior of statistical procedures in the presence of dependence. We establish convergence of tail empirical processes to Gaussian limits for β-mixing stationary time series. As a consequence, one obtains weighted approximations of the tail empirical quantile function that is based on a random sequence with marginal distribution belonging to the domain of attraction of an extreme value distribution. Moreover, the asymptotic normality is concluded for a large class of estimators of the extreme value index. These results are applied to stationary solutions of a general stochastic difference equation.

1. Introduction. In many fields one must assess the potential risk that an extremal event may lead to serious damages or losses. For example, extremely high water levels may cause a dike to break and large negative stock returns can result in high financial losses or even bankruptcy. If this catastrophical event can be described by the occurrence of a large real observation, for that purpose one needs reliable estimators of the probability that a given high threshold is exceeded.

To be more concrete, assume that the observations can be modelled by random variables (r.v.'s) X_i, $1 \le i \le n$, with common distribution function (d.f.) F. If these r.v.'s may be assumed independent, then classical extreme value theory provides the necessary tools for estimating the upper tail of F. For this, one assumes that F belongs to the weak domain of attraction of some unknown extreme value d.f. G_γ [in short: $F \in D(G_\gamma)$], that is,

$$\mathscr{L}\left(a_n^{-1}\left(\max_{1\le i\le n} X_i - b_n\right)\right) \longrightarrow G_\gamma \quad \text{weakly} \tag{1.1}$$

for suitable normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ and

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

which is interpreted as $\exp(e^{-x})$ for $\gamma = 0$. It is well known that up to a scale and location parameter these are the only possible nondegenerate limiting d.f.'s under linear normalization. Note that, in case of i.i.d. data, (1.1) is

Received August 1999; revised April 2000.

[1]Supported in part by ONR Grant N00014-93-1-0043 and European Union TMR Grant ERB-FMRX-CT 960095.

AMS 1991 *subject classifications*. Primary 60F17, 62M10; secondary 60G70, 62G20.

Key words and phrases. ARCH-process, dependent, extreme value index, Hill estimator, invariance principle, statistical tail functional, stochastic difference equation, tail empirical distribution function, tail empirical quantile function, time series, β-mixing.

equivalent to

$$n(1 - F(a_n x + b_n)) \longrightarrow -\log G_\gamma(x), \qquad x \in \mathbb{R}.$$

Under this assumption the so-called extreme value index γ largely determines the behavior of $1 - F(x)$ as x tends to the right endpoint of the support of F, and so the estimation of γ is the crucial step in the analysis of the upper tail. Consequently, it has received a lot of attention since the seminal papers by Hill (1975) and Pickands (1975). Starting from an approximation of the tail empirical quantile function (q.f.), in foregoing papers [Drees (1998a,b)] we developed a general theory for estimators of the extreme value index that can be represented as a smooth functional of the tail empirical q.f.

However, often a certain dependency between the data can be observed, so that the application of the aforementioned theory for i.i.d.-models is not appropriate. For instance, returns on shares usually reveal a nonlinear dependence structure, which often is modelled by (generalized) ARCH-processes (see Section 4 for details). Further examples were discussed by Embrechts, Klüppelberg and Mikosch (1997) and Reiss and Thomas (1997). While the probabilistic extreme value theory for stationary time series is well established [see Leadbetter, Lindgren and Rootzén (1983)], unfortunately much less is known about the statistical side of the problem. In particular, by and large the mathematical results about the estimation of the extreme value index based on dependent r.v.'s are restricted to the popular Hill estimator in case of heavy tails, that is, for $\gamma > 0$ [see Hsing (1991), Resnick and Stărică (1997, 1998) and Stărică (1999)]. It is the main goal of the present paper to establish approximations of tail processes that render it possible to carry over the general theory of the estimation of γ from the i.i.d-case.

An important contribution to this goal was made by Rootzén (1995), who proved a limit theorem for the tail process of β-mixing time series. Let $(U_i)_{i\in\mathbb{N}}$ be a sequence of uniformly distributed r.v.'s. Recall that it is called (strictly) stationary if $\mathscr{L}((U_i)_{i\in\mathbb{N}}) = \mathscr{L}((U_{n+i})_{i\in\mathbb{N}})$ for all $n \in \mathbb{N}$, and it is β-mixing (or absolutely regular) if

$$\beta(k) := \sup_{l\in\mathbb{N}} E\Big(\sup_{A\in\mathscr{B}_{l+k+1}^{\infty}} \big|P(A|\mathscr{B}_1^l) - P(A)\big|\Big) \longrightarrow 0$$

as $k \to \infty$, where $\mathscr{B}_1^l$ and $\mathscr{B}_{l+k+1}^{\infty}$ denote the σ-fields generated by $(U_i)_{1\le i\le l}$ and $(U_i)_{l+k+1\le i<\infty}$, respectively. The class of β-mixing time series includes recurrent Markov chains under mild conditions and, more specifically, ARMA-, ARCH- and GARCH-models, where often the mixing coefficients $\beta(k)$ vanish with an exponential rate; see Doukhan [(1995), Section 2.4] for details.

The basic result of Rootzén (1995) gives conditions under which the tail empirical process

$$e_n := \Big((nv_n)^{-1/2} \sum_{i=1}^{n} \Big(1_{\{U_i > 1 - v_n x\}} - v_n x\Big)\Big)_{x\in[0,1]}$$

pertaining to $(U_i)_{1\le i<n}$ converges weakly to a Gaussian process e as $v_n \to 0$ and $nv_n \to \infty$.

However, since both processes e_n and e are close to 0 in the neighborhood of 0, the convergence $e_n \to e$ bears little information about the most extreme observations $U_i > 1 - v_n x$ for small x. In order to prove a counterpart of the limit Theorem 2.1 of Drees (1998a) for the tail empirical q.f. in case of β-mixing sequences, we need a stronger assertion about the asymptotic behavior of e_n nearby the origin, which can be provided by weighted approximations of the type

$$\frac{e_n}{q} \longrightarrow \frac{e}{q} \quad \text{weakly in } D[0,1], \tag{1.2}$$

where $D[0,1]$ denotes the Skorohod space on the unit interval and q: $[0,1] \to [0,\infty)$ is a suitable weight function tending to 0 as x tends to 0. Throughout the paper, we will restrict attention to functions $q \in D[0,1]$ satisfying

$$\inf_{x\in[\vartheta,1]} q(x) > 0 \ \forall\ \vartheta > 0 \quad \text{and} \quad x^{\nu}|\log x|^{\mu} = O(q(x)) \quad \text{as } x \downarrow 0 \tag{1.3}$$

for some $\nu \in [0,1/2]$ and $\mu \in \mathbb{R}$, although similar results can be proved if $x^{\nu}|\log x|^{\mu}$ is replaced by more general functions that are ν-varying at 0.

It turns out that the assumptions of Rootzén's result are sufficient to verify (1.2) if $\nu < 1/4$. Furthermore, under somewhat stronger conditions we obtain weighted approximations for all $\nu < 1/2$ and for $\nu = 1/2$ if $\mu > 1/4$ (Theorems 2.2 and 2.3). In Section 3 we employ these results to derive weighted approximations of the tail empirical q.f. of β-mixing stationary time series whose marginal d.f. belongs to the weak domain of attraction of an extreme value distribution. Then we conclude the asymptotic normality of a general class of estimators of the extreme value index. In Section 4, these results are applied to solutions of stochastic difference equations, including ARCH(1) time series. All proofs are collected in Section 5.

2. Approximations of uniform tail empirical processes. We start with a slight improvement of Rootzén's (1995) limit theorem for uniform tail empirical processes. For this, we assume that there exists a sequence $l_n \to \infty$ and a function r: $[0,1]^2 \to \mathbb{R}$ such that

(C1)

$$\lim_{n\to\infty} \frac{\beta(l_n)}{l_n} n + l_n (nv_n)^{-1/2} \log^2(nv_n) = 0.$$

(C2)

$$\lim_{n\to\infty} \frac{l}{l_n v_n} \operatorname{Cov}\left(\sum_{i=1}^{l_n} 1_{\{U_i > 1 - v_n x\}}, \sum_{i=1}^{l_n} 1_{\{U_i > 1 - v_n y\}} \right) = r(x,y) \quad \forall\ 0 \le x, y \le 1.$$

(C3) There exists a constant C such that

$$\frac{1}{l_n v_n}\operatorname{Var}\left(\sum_{i=1}^{l_n} 1_{\{1-v_n y<U_i\le 1-v_n x\}}\right) \le C(y-x) \qquad \forall\, 0\le x<y\le 1,\ n\in\mathbb{N}. \tag{2.1}$$

THEOREM 2.1. *If* $(U_i)_{i\in\mathbb{N}}$ *is a stationary, β-mixing sequence of uniformly distributed random variables that meets the conditions* (C1)–(C3)*, then*

$$e_n \longrightarrow e \quad \textit{weakly in } D[0,1], \tag{2.2}$$

where e is a centered continuous Gaussian process with covariance function r.

For the proof, the r.v.'s U_i, $1\le i\le n$, are divided into blocks of length l_n. Condition (C1), which is slightly weaker than the corresponding condition (5.1) in Rootzén (1995), ensures on the one hand that l_n is sufficiently large such that blocks that are not adjacent are asymptotically independent, and on the other hand that l_n does not grow too fast such that the contribution of a single block to the process e_n is negligible.

(C2) is a natural condition if one wants to prove convergence of e_n toward a Gaussian limit. Observe that from (C2) one may conclude that the left-hand side of (2.1) converges to $r(y,y)-2r(x,y)+r(x,x)$, which can be bounded by a multiple of $y-x$, provided r is Lipschitz continuous in both arguments. By condition (C3) we require that this boundedness holds true *uniformly* for $0\le x<y\le 1$. Although this assumption seems rather technical, it cannot be omitted without a substitute, for in Rootzén (1998) a β-mixing time series with exponentially decreasing β-coefficients is given that satisfies the conditions (C1) and (C2), but nevertheless the process e_n does not converge. On the other hand, if the sequence $(U_i)_{i\in\mathbb{N}}$ is ρ-mixing with mixing coefficients

$$\rho(k) := \sup_{X\in L_2(\mathscr{B}_1^l),\, Y\in L_2(\mathscr{B}_{i+k+l}^\infty)} \frac{|\operatorname{Cov}(X,Y)|}{(\operatorname{Var}(X)\operatorname{Var}(Y))^{1/2}}$$

(with $L_2(\mathscr{A})$ denoting the space of square integrable $\mathscr{A}$, $\mathbb{B}$-measurable functions) satisfying $\sum_{i=1}^\infty \rho(2^i)<\infty$ [e.g., $\rho(k)\le \text{const.}\log^{-2}k$], then (C3) is automatically fulfilled [Shao (1993), Lemma 2.3]. Moreover, it is satisfied if the size of each cluster of exceedances over a sufficiently high threshold has finite second moment. For details and further sufficient conditions for (C3) we refer to Rootzén (1995).

It turns out that the conditions (C1)–(C3) are sufficient to prove a weighted approximation for the uniform tail empirical process if the weight function $q(x)$ does not converge to 0 too fast as x tends to 0, or the sequence l_n grows sufficiently slowly.

THEOREM 2.2. *Suppose that the weight function q satisfies* (1.3) *with $\nu < 1/4$; or $\nu = 1/4$ and $\mu > 1/4$; or $\nu \le 1/2$, $\mu > 1/4$ if $\nu = 1/2$ and $s_n \to 0$, where*

$$(2.3)\quad s_n := \begin{cases} l_n^2(nv_n)^{-1}\log^4(nv_n)\log\log(nv_n), & \text{if } \nu=\mu=1/4,\\ l_n^2(nv_n)^{-1}\log^{5-4\mu}(nv_n), & \text{if } \nu=1/4, \mu<1/4,\\ l_n^2(nv_n)^{-3(1-2\nu)/(2(1-\nu))}\log^{-3\mu/(1-\nu)}(nv_n), & \text{if } \nu>1/4.\end{cases}$$

Then, under the conditions of Theorem 2.1, *convergence* (1.2) *holds true.*

It is well known that in the case of independent r.v.'s U_i convergence (1.2) holds if $(x\log\log x)^{1/2} = o(q(x))$ as $x \downarrow 0$. In the present situation, for $\nu = 1/2$, which again is the best achievable value, condition (2.3) reads as $l_n = o(\log^{3\mu}(nv_n))$; hence one may obtain an approximation of the tail empirical process for a β-mixing sequence that is almost as accurate as in the i.i.d.-case, provided the mixing coefficients converge to 0 sufficiently fast. If, for example, $(U_i)_{i\in\mathbb{N}}$ is geometrically β-mixing, that is, $\beta(k) = O(\eta^k)$ for some $\eta \in (0,1)$, and $v_n = n^{-b}$ for some $b \in (0,1)$, then one may choose $l_n = [-2\log n/\log\eta]$ and (1.2) holds for $q(x) \ge \text{const.}\, x^{1/2}|\log x|^{\mu}$ with $\mu > 1/3$.

In Theorem 2.2 we impose a stronger condition on the sequence l_n than in Theorem 2.1 to obtain a weighted approximation of e_n with $\nu > 1/4$. Alternatively, one may strengthen the boundedness condition (C3) by considering fourth instead of second moments.

(C3*) There exists a constant C such that

$$\frac{1}{l_n v_n} E\left(\sum_{i=1}^{l_n} 1_{\{1-v_n y < U_i \le 1-v_n x\}}\right)^4 \le C(y-x) \quad \forall\, 0 \le x < y \le 1, \quad n \in \mathbb{N}.$$

Note that condition (C3*) implies (C3) and, in addition, $l_n v_n \to 0$, that is, the average number of exceedances in each block vanishes asymptotically. This, however, is a natural assumption in the blocks approach [see Rootzén (1995), (2.5)]. Again in view of Shao [(1993), Lemma 2.3], (C3*) is automatically satisfied if the sequence of random variables is ρ-mixing with $\sum_{i=1}^{\infty}\rho^{1/2}(2^i) < \infty$, or if the fourth moment of the size of a cluster of exceedances is finite.

THEOREM 2.3. *Suppose $(U_i)_{i\in\mathbb{N}}$ is a stationary, β-mixing sequence of uniformly distributed random variables satisfying* (C1), (C2) *and* (C3*). *Then convergence* (1.2) *with q according to* (1.3) *holds true for $\nu < 1/2$ and for $\nu = 1/2$ and $\mu > 1/4$.*

REMARK 1. (i) A close inspection of the proof of Theorems 2.1–2.3 shows that the conditions (C3) and (C3*) can be weakened in that the inequalities are merely required for all $0 \le x < y \le 1$ and $n \in \mathbb{N}$ such that $y - x \ge \lambda_n$ for some sequence $\lambda_n = o(\min((nv_n)^{-1/2}, (nv_n)^{-1/(2(1-\nu))}\log^{\mu/(1-\nu)}(nv_n)))$. For, in view of (5.11), these weaker conditions are sufficient to ensure the moment inequalities (5.1) and (5.14), respectively.

(ii) Utilizing a standard quantile transformation technique, one can derive analogs to Theorems 2.1–2.3 for more general marginal d.f.'s F. However, in general this requires a nonlinear normalization of the argument of the tail empirical d.f., whereas a limit theorem for tail empirical processes of the type

$$\left(\left(n(1-F(d_n))\right)^{-1/2}\sum_{i=1}^{n}\left(1_{\{X_i>c_n x+d_n\}}-(1-F(c_n x+d_n))\right)\right)_{x\in[0,\,F^{-1}(1))}$$

require that F belongs to the domain of attraction of an extreme value distribution. For details we refer to Rootzén (1995).

3. Tail empirical quantile functions and statistical tail functionals. Consider a stationary, β-mixing sequence $(X_i)_{i\in\mathbb{N}}$ with marginal d.f. $F\in D(G_\gamma)$. In this section we investigate the asymptotic behavior of the pertaining tail empirical quantile function (q.f.),

$$Q_n(t):=F_n^{-1}\left(1-\frac{k_n}{n}t\right)=X_{n-[k_n t]:n},\qquad t\in[0,1],$$

which is based on the k_n+1 largest order statistics $\max_{1\le i\le n}X_i=X_{n:n}\ge X_{n-1:n}\ge\cdots\ge X_{n-k_n:n}$. Here F_n^{-1} denotes the empirical q.f. and $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence; that is, $k_n\to\infty$ and $k_n/n\to 0$.

These investigations are motivated by the fact that many estimators of the extreme value index can be represented as a *statistical tail functional*, that is, as $T(Q_n)$ for some functional T. Hence, under suitable regularity and smoothness conditions on T, one can deduce limit theorems for such estimators from approximations of Q_n using the well-known δ-method. For i.i.d. sequences this program was worked out in Drees (1998a,b).

Again we first examine the case of uniformly distributed r.v.'s U_i. More precisely, we want to derive weighted approximations for

$$V_n(t):=\frac{n}{k_n}(1-U_{n-[k_n t]:n}),\qquad t\in[0,1]$$

from Theorems 2.2 and 2.3. Here k_n corresponds to nv_n in Section 2, which is the average number of observations taken into account by e_n. However, note that we must consider $e_n(t)$, $t\in[0,1+\varepsilon]$ for some $\varepsilon>0$ to ensure that with probability tending to 1 at least k_n+1 observations are taken into account.

A key ingredient in the proof of weighted approximations for the uniform tail empirical q.f. V_n are the following "in-probability linear bounds." Analogous results for the whole empirical q.f. are well known for i.i.d. r.v.'s [see, e.g., Shorack and Wellner (1986), inequality (10.4.1)] and were established by Puri and Tran [(1980), Theorem 1.1] for φ-mixing sequences.

LEMMA 3.1. *Fix some $\varepsilon>0$. Let $(U_i)_{i\in\mathbb{N}}$ be a stationary, β-mixing sequence of uniformly distributed r.v.'s such that the conditions* (C1) *and* (C3) *hold for* *all* $0 \le x < y \le 1+\varepsilon$ *and* $v_n = k_n/n$. *Then*

$$\sup_{t\in[1/(2k_n),\,1]} \frac{V_n(t)}{t} = O_P(1), \tag{3.1}$$

$$\sup_{t\in(0,\,1]} \frac{t}{V_n(t)} = O_P(1). \tag{3.2}$$

Now we are ready to deduce a weighted approximation for V_n from Theorem 2.2 or Theorem 2.3 using Vervaat's (1972) lemma.

COROLLARY 3.1. *Suppose that, for* $v_n = k_n/n$ *and some fixed* $\varepsilon > 0$, *the conditions of Theorem* 2.2 *or Theorem* 2.3 *are fulfilled where* (C2), (C3) *and* (C3*), *respectively, hold for all* $0 \le x < y \le 1+\varepsilon$. *Then*

$$\frac{k_n^{1/2}(V_n - \mathrm{Id})}{q} 1_{[1/(2k_n),1]} \longrightarrow \frac{e}{q} \quad \textit{weakly in } D[0,1].$$

(*Here* Id *denotes the identity function* $[0,1] \ni t \mapsto t$ *on the unit interval.*)

Next we consider more general marginal d.f.'s F. Using quantile transformation techniques, from Corollary 3.1 one may deduce weighted approximations for the process

$$\left(\frac{n}{\sqrt{k_n}} f\left(F^{-1}\left(1-\frac{k_n}{n}t\right)\right)\left(Q_n(t) - F^{-1}\left(1-\frac{k_n}{n}t\right)\right)\right)_{t\in[0,\,1]} \tag{3.3}$$

if F possesses a sufficiently regular Lebesgue density f in the right tail. Compare Drees and de Haan (1999), where this program is carried out for the case of i.i.d. observations under very weak conditions on F^{-1}. However, since the normalizing factor in (3.3) depends on the tail of the density function, a different type of approximations for Q_n suits better for applications in extreme value statistics.

Recall that $F \in D(G_\gamma)$ if and only if

$$R(\lambda, t) := \frac{F^{-1}(1-\lambda t) - F^{-1}(1-\lambda)}{a(\lambda)} - F_\gamma^{-1}(t) \longrightarrow 0 \tag{3.4}$$

as $\lambda \downarrow 0$, where a: $(0,1) \to (0,\infty)$ is a normalizing function and

$$F_\gamma^{-1}(t) := (-\log G_\gamma)^{-1}(t) = \frac{t^{-\gamma}-1}{\gamma},$$

which is interpreted as $-\log t$ if $\gamma = 0$. In Drees (1998a), Lemma 2.1, it was shown that this convergence even holds uniformly:

$$\sup_{t\in(0,\,t_0]} t^{\gamma+\eta}|R(\lambda,t)| \longrightarrow 0$$

for all $t_0, \eta > 0$. Hence to any fixed $t_0 > 1$ there exists an intermediate sequence k_n such that

$$k_n^{1/2} \sup_{t \in (0, t_0]} \frac{t^{\gamma+1}}{q(t)} |R(k_n/n, t)| \longrightarrow 0, \tag{3.5}$$

provided q satisfies (1.3) for some $\nu \leq 1/2$.

In addition to (3.5), we need the following counterparts of the conditions (C1)–(C3) and (C3*): There exist a constant $\varepsilon > 0$ and a sequence $(l_n)_{n \in \mathbb{N}}$ such that

$(\widetilde{\text{C1}})$

$$\lim_{n \to \infty} \frac{\beta(l_n)}{l_n} n + l_n k_n^{-1/2} \log^2 k_n = 0.$$

$(\widetilde{\text{C2}})$

$$\lim_{n \to \infty} \frac{n}{l_n k_n} \text{Cov}\left(\sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_n/n)x)\}}, \sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_n/n)y)\}} \right) = r(x, y)$$

$$\forall\, 0 \leq x, y \leq 1 + \varepsilon.$$

$(\widetilde{\text{C3}})$ For some constant C,

$$\frac{n}{l_n k_n} \text{Var}\left(\sum_{i=1}^{l_n} 1_{\{F^{-1}(1-(k_n/n)y) < X_i \leq F^{-1}(1-(k_n/n)x)\}} \right) \leq C(y - x)$$

$$\forall\, 0 \leq x < y \leq 1 + \varepsilon, \; n \in \mathbb{N}.$$

$(\widetilde{\text{C3*}})$ For some constant C,

$$\frac{n}{l_n k_n} E\left(\sum_{i=1}^{l_n} 1_{\{F^{-1}(1-(k_n/n)y) < X_i \leq F^{-1}(1-(k_n/n)x)\}} \right)^4 \leq C(y - x)$$

$$\forall\, 0 \leq x < y \leq 1 + \varepsilon, \; n \in \mathbb{N}.$$

THEOREM 3.1. *Suppose $(X_i)_{i \in \mathbb{N}}$ is a stationary, β-mixing time series with continuous marginal d.f. $F \in D(G_\gamma)$ and let $(k_n)_{n \in \mathbb{N}}$ be an intermediate sequence satisfying* (3.5) *for some $t_0 > 1$. In addition, assume that either*

(i) $(\widetilde{\text{C1}})$–$(\widetilde{\text{C3}})$ *hold and q fulfills the conditions of Theorem* 2.2 *for $v_n = k_n/n$, or*

(ii) $(\widetilde{\text{C1}}), (\widetilde{\text{C2}})$ *and* $(\widetilde{\text{C3*}})$ *hold and q satisfies* (1.3) *with $\nu < 1/2$, or with $\nu = 1/2$ and $\mu > 1/4$.*

Then there exist versions of the tail empirical q.f. Q_n and a centered Gaussian process e with covariance function r such that

$$\sup_{t \in (0,1]} \frac{t^{\gamma+1}}{q(t)} \left| k_n^{1/2} \left(\frac{Q_n(t) - D_n}{a(k_n/n)} - F_\gamma^{-1}(t) \right) - t^{-(\gamma+1)} e(t) \right| \longrightarrow 0 \tag{3.6}$$

in probability, where

$$D_n := \begin{cases} F^{-1}(1-k_n/n), & \textit{if } \lim_{t\downarrow 0} t^{\gamma+1}/q(t) = 0,\\ X_{n:n} + a(k_n/n)/\gamma, & \textit{else.}\end{cases}$$

REMARK 2. Remark 1(i) applies analogously in the present situation. Furthermore, one may drop the continuity assumption for F, if one replaces $1_{\{F^{-1}(1-k_n/ny)<X_i\le F^{-1}(1-k_n/nx)\}}$ in the conditions $(\widetilde{\mathrm{C}3})$ and $(\widetilde{\mathrm{C}3^*})$, respectively, by $1_{\{F^{-1}(1-k_n/ny)\le X_i\le F^{-1}(1-k_n/nx)\}}$, provided the natural condition $l_n k_n/n \to 0$ holds.

To see this, we replace the definition of U_i in the proof of Theorem 3.1 by $U_i := \widetilde{U}_i F(X_i - 0) + (1-\widetilde{U}_i)F(X_i)$, where $\widetilde{U}_i$ are i.i.d. uniformly distributed r.v.'s that are independent of $(X_i)_{i\in\mathbb{N}}$. Then the r.v.'s U_i are uniformly distributed [Moore and Spruill (1975), Lemma 3.2] and condition (C1) is obtained from $(\widetilde{\mathrm{C}1})$ and Theorem 1.1.1 of Doukhan (1995). Moreover, it is easily seen that

$$\left\{1-\frac{k_n}{n}y < U_i \le 1-\frac{k_n}{n}x\right\} \subset \left\{F^{-1}\left(1-\frac{k_n}{n}y\right) \le X_i \le F^{-1}\left(1-\frac{k_n}{n}x\right)\right\},$$

so that the conditions (C3) and (C3*), respectively, are obvious. Finally, (C2) can be derived from $(\widetilde{\mathrm{C}2})$, for the modified condition $(\widetilde{\mathrm{C}3})$ implies

$$\frac{n}{l_n k_n}\mathrm{Var}\left(\sum_{i=1}^{l_n}\left(1_{\{U_i>1-(k_n/n)x\}} - 1_{\{X_i>F^{-1}(1-(k_n/n)x)\}}\right)\right) \longrightarrow 0.$$

Since the representation $X_i = F^{-1}(U_i)$ still holds true in the present more general situation, the proof can be completed in the same way as in the case of a continuous marginal d.f.

The assertion of Theorem 3.1 can be rewritten as a weak convergence result in a suitable normed space. Let

$$D_{\gamma,q} := \left\{z\colon [0,1]\to\mathbb{R} \,\middle|\, \lim_{t\downarrow 0}|z(t)|t^{\gamma+1}/q(t) = 0, (z(t)t^{\gamma+1}/q(t))_{t\in[0,1]} \in D[0,1]\right\}$$

be a function space equipped with the (semi)norm

$$\|z\|_{\gamma,q} := \sup_{t\in(0,1]} |z(t)|t^{\gamma+1}/q(t),$$

with the convention $0/0 := 0$. Because in general a constant function and hence F_γ^{-1} need not belong to $D_{\gamma,q}$, we define

$$\overline{F}_\gamma^{-1}(t) := \begin{cases} t^{-\gamma}/\gamma, & \text{if } \gamma\ne 0,\\ -\log t, & \text{if } \gamma = 0\end{cases}$$

and $\overline{D}_n := D_n - a(k_n/n)1_{\{\gamma\ne 0\}}/\gamma$. Furthermore, let $e_\gamma(t) := t^{-(\gamma+1)}e(t)$, $t\in[0,1]$. Then (3.6) is equivalent to

$$k_n^{1/2}\left(\frac{Q_n-\overline{D}_n}{a(k_n/n)} - \overline{F}_\gamma^{-1}\right) \longrightarrow e_\gamma \quad \text{weakly in } D_{\gamma,q}. \tag{3.7}$$

Next we turn to statistical tail functionals $T(Q_n)$ used as estimators of the extreme value index, where T: $\operatorname{span}(D_{\gamma,q}, 1) \to \mathbb{R}$ is a functional such that $T_{|D_{\gamma,q}}$ is $\mathscr{B}(D_{\gamma,q}), \mathscr{B}(\mathbb{R})$-measurable (with $\mathscr{B}$ denoting the respective Borel σ-field). In view of (3.7), the following conditions ensure the asymptotic normality of the estimator:

(T1)

$$T(az+b) = T(z) \qquad \forall\, a > 0, b \in \mathbb{R}, z \in D_{\gamma,q}.$$

(T2)

$$T(\overline{F}_\gamma^{-1}) = \gamma.$$

(T3) T is Hadamard differentiable at $\overline{F}_\gamma^{-1}$ tangentially to $C_{\gamma,q} := \{z \in D_{\gamma,q} | z_{|(0,1]} \in C(0,1]\}$ with derivative T'_γ, which means

For all sequences $\lambda_n \downarrow 0$ and $D_{\gamma,q} \in z_n \to z \in C_{\gamma,q}$ one has

$$\frac{T(\overline{F}_\gamma^{-1} + \lambda_n z_n) - T(\overline{F}_\gamma^{-1})}{\lambda_n} \longrightarrow T'_\gamma(z), \tag{3.8}$$

where T'_γ; $C_{\gamma,q} \to \mathbb{R}$ is a continuous linear functional [see Gill (1989), page 102].

The limit distribution of $T(Q_n)$ is more easily described in terms of the following alternative representation of the derivative T'_γ. According to the Riesz representation theorem there exists a unique signed measure $\nu_{T,\gamma}$ on $[0,1]$ satisfying

$$\int_{[0,1]} t^{-(\gamma+1)} q(t) |\nu_{T,\gamma}|(dt) < \infty \tag{3.9}$$

such that

$$T'_\gamma(z) = \int_{[0,1]} z(t) \nu_{T,\gamma}(dt) \qquad \forall\, z \in C_{\gamma,q}.$$

COROLLARY 3.2. *If convergence* (3.6) *holds and* T *meets the conditions* (T1)–(T3), *then*

$$\mathscr{L}\big(k_n^{1/2}(T(Q_n) - \gamma)\big) \longrightarrow \mathscr{N}(0, \sigma^2_{T,\gamma}) \quad \textit{weakly}$$

with $\sigma^2_{T,\gamma} := \int_{[0,1]^2} (st)^{-\gamma+1} r(s,t) \nu^2_{T,\gamma}(ds, dt)$.

The heuristic explanation for the asymptotic normality is given by the following calculations:

$$\begin{aligned} T(Q_n) &= T\left(\frac{Q_n - \overline{D}_n}{a(k_n/n)}\right) \approx T(\overline{F}_\gamma^{-1} + k_n^{-1/2} e_\gamma) \approx T(F_\gamma^{-1}) + k_n^{-1/2} T'_\gamma(e_\gamma) \\ &= \gamma + k_n^{-1/2} \int e_\gamma \, d\nu_{T,\gamma}, \end{aligned}$$

where the integral is $\mathcal{N}(0, \sigma^2_{T,\gamma})$-distributed. Likewise, one may prove consistency of the statistical tail functional $T(Q_n)$ if one merely assumes the continuity of T at $\overline{F}_\gamma^{-1}$ instead of (T3). For details we refer to Drees (1998a).

Condition (T3) can be weakened in so far as one requires convergence (3.8) only for sequence $\overline{F}_\gamma^{-1} + \lambda_n z_n \in \overline{D}_{\gamma,q}$ where $\overline{D}_{\gamma,q} \subset D_{\gamma,q}$ is some subset such that $P\{Q_n \in \overline{D}_{\gamma,q}\} \to 1$ [cf. Drees (1998a), Remark (i) following Theorem 3.2]. Then conditions (T1)–(T3) are fulfilled by many well-known estimators of the extreme value index, including the Pickands (1975) estimator and a certain class of generalized probability weighted moment estimators [see Drees (1998a), Example 3.1]. Perhaps the most prominent example is the maximum likelihood estimator $\hat{\gamma}^{(\mathrm{ML})}$ in a generalized Pareto model, obtained as the first component of a solution (γ, σ) of the likelihood equations

$$\frac{1}{k_n}\sum_{i=1}^{k_n}\Bigl(1 - \frac{\gamma}{\sigma}(X_{n-i+1:n} - X_{n-k_n:n})\Bigr)^{-1} = \frac{1}{\gamma+1},$$

$$\frac{1}{k_n}\sum_{i=1}^{k_n}\log\Bigl(1 - \frac{\gamma}{\sigma}(X_{n-i+1:n} - X_{n-k_n:n})\Bigr) = \gamma$$

with $\gamma > -1/2$ [see Smith (1987) for further details]. In Drees [(1998a), Example 4.1], it was shown that the functional pertaining to the maximum likelihood estimator meets the conditions (T1)–(T3) with $q(t) = t^\varepsilon$ for an arbitrary $\varepsilon > 0$ and measure

$$\nu_{\mathrm{ML},\gamma}(dt) = \frac{(\gamma+1)^2}{\gamma}(t^\gamma - (2\gamma+1)t^{2\gamma})\,dt + (\gamma+1)\varepsilon_1(dt) \tag{3.10}$$

pertaining to its Hadamard derivative at $\overline{F}_\gamma^{-1}$, where ε_1 denotes the Dirac measure concentrated at 1. Hence, under the conditions of Corollary 3.2, $\hat{\gamma}_n^{(\mathrm{ML})}$ is asymptotically normal with asymptotic variance $\sigma^2_{\mathrm{ML},\gamma} := \int_{[0,1]}(st)^{-(\gamma+1)} \times r(s,t)\nu^2_{\mathrm{ML},\gamma}(ds, dt)$.

On the other hand, some estimators of γ are merely scale invariant, but not location invariant, that is, the pertaining functional satisfies

$$T(az) = T(z) \qquad \forall\, a > 0, z \in \operatorname{span}(D_{\gamma,q}, 1) \tag{T1*}$$

instead of (T1); the most popular estimator of that type is the Hill estimator,

$$\hat{\gamma}_n^{(H)} := \frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{X_{n-i+1:n}}{X_{n-k_n:n}},$$

which is suitable only for $\gamma > 0$. Indeed, it was shown in Drees (1998b) that for estimating $\gamma \le 0$, location invariant estimators clearly outperform statistical tail functionals that merely fulfill (T1*). Therefore, here we restrict ourselves to positive extreme value indices.

For estimators that are not location invariant, an additional bias term related to

$$\Xi_n := \frac{\overline{D}_n}{a(k_n/n)} = \frac{F^{-1}(1-k_n/n)}{a(k_n/n)} - \frac{1}{\gamma}$$

has to be taken into account. For that reason, we need an extra condition on k_n describing the ratio between the asymptotic bias and standard deviation:

$$k_n^{1/2}\Xi_n \longrightarrow \lambda \in [0,\infty]. \tag{3.11}$$

COROLLARY 3.3. *Suppose convergence* (3.6) *holds with* $\gamma > 0$, *the functional* T *satisfies the conditions* (T1*), (T2) *and* (T3), *and the sequence* $(k_n)_{n\in\mathbb{N}}$ *fulfills* (3.11). *Then*

$$\lambda < \infty \quad\Longrightarrow\quad \mathscr{L}\Big(k_n^{1/2}(T(Q_n)-\gamma)\Big) \longrightarrow \mathscr{N}(\lambda\nu_{T,\gamma}[0,1], \sigma^2_{T,\gamma}) \quad \textit{weakly},$$

$$\lambda = \infty \quad\Longrightarrow\quad \frac{1}{\Xi_n}(T(Q_n)-\gamma) \longrightarrow \nu_{T,\gamma}[0,1] \quad \textit{in probability}.$$

The proof can be copied from Drees (1998b), Theorem 3.1 [see also Drees (1998c), Theorem 2.8]. Note that (3.9) guarantees that $\nu_{T,\gamma}[0,1]$ is well defined.

As an example, we consider Hill's estimator. The corresponding functional $T_H(z) := \int_0^1 \log(z(t)/z(1))\,dt$ satisfies (T1*), (T2) and the weakened version of (T3) with appertaining measure

$$\nu_{H,\gamma}(dt) = \gamma(t^\gamma\,dt - \varepsilon_1(dt))$$

[Drees (1998b), Example 3.1; see Drees (1998c), Example 2.6, for a formulation that comes closer to the present notation]. Thus, in case of $\lambda < \infty$, Corollary 3.3 yields the asymptotic normality of $\hat\gamma_n^{(H)}$, too.

It is worth mentioning that one may take more order statistics into account, that is, choose a larger number k_n, if a second order strengthening of the condition (3.4) can be obtained for the underlying d.f. $F \in D(G_\gamma)$. Then in general the limit distribution of $k_n^{1/2}(T(Q_n)-\gamma)$ has a nonvanishing bias even for location invariant functionals. See Drees (1998a–c) for analogous calculations in case of i.i.d. data.

As we have seen, once the assumptions of Theorem 3.1 are checked, Corollaries 3.2 and 3.3 immediately yield the limit distributions of many popular estimators for γ with virtually no expenditure, since the measure $\nu_{T,\gamma}$ pertaining to these estimators are determined in the foregoing papers, Drees (1990a–c). Moreover, there the classes of possible measures $\nu_{T,\nu}$ that can occur in the context of Corollaries 3.2 and 3.3, respectively, are characterized and for a given smooth family $(\nu_\gamma)_{\gamma\in I}$, with $I \subset \mathbb{R}$ denoting an interval, a functional T is constructed such that $\nu_{T,\gamma} = \nu_\gamma$ for all $\gamma \in I$. Hence all possible limit distributions of statistical tail functionals are known and for a given family of limit distributions one may construct an estimator with this prescribed asymptotic behavior. In the next section, the power of this approach is demonstrated by the example of times series that can be described by a stochastic difference equation.

Finally, it is worth mentioning that, in view of (C2), one may estimate the covariance function r of the Gaussian process e consistently by an empirical covariance function based on a division of the whole sample in $[1/l_n]$ blocks of length l_n. This estimate can be used to construct a consistent estimator of $\sigma^2_{T,\gamma}$ and thus of confidence intervals to a given asymptotic confidence coefficient.

4. Solutions of stochastic difference equations. In financial time series one often observes that the absolute values or the squares of the data show a much more pronounced autocorrelation than the original observations. Engle (1982) tried to capture this feature by modeling the data by an autoregressive conditional heteroskedastic (ARCH) time series. For example, an ARCH(1) time series is a stationary solution $(Y_i)_{i\in\mathbb{N}}$ of the stochastic difference equation

$$Y_i = (\alpha_0 + \alpha_1 Y_{i-1}^2)^{1/2} Z_i, \qquad i \in \mathbb{N},$$

where $\alpha_0, \alpha_1 > 0$ are real parameters and Z_i, $i \in \mathbb{N}$, are i.i.d. innovations with $EZ_i = 0$ and $\operatorname{Var} Z_i = 1$. More generally, we consider difference equations of the type

$$X_i = A_i X_{i-1} + B_i, \qquad i \in \mathbb{N} \tag{4.1}$$

where (A_i, B_i), $i \in \mathbb{N}$, are i.i.d. $[0,\infty)^2$-valued random vectors. Notice that $X_i = Y_i^2$ satisfies (4.1) with $A_i = \alpha_1 Z_i^2$ and $B_i = \alpha_0 Z_i^2$. A list of further applications of model (4.1) can be found in Vervaat (1979). For sake of simplicity, throughout this section we assume that the distribution of (A_1, B_1) is absolutely continuous, although this assumption can be weakened considerably [see Stărică (1999) for details].

The tail behavior of a stationary solution of (4.1) was first studied by Kesten (1973) under the following moment assumptions:

(S1) There exists $\kappa > 0$ such that

$$EA_1^\kappa = 1, \qquad E\big(A_1^\kappa \max(\log A_1, 0)\big) < \infty \quad \text{and} \quad EB_1^\kappa \in (0,\infty).$$

Then, according to Theorem 5 of Kesten (1973), there exists a distributionally unique stationary solution X_i, $i \in \mathbb{N}$, of (4.1) such that the d.f. F of X_i satisfied $1 - F(x) \sim cx^{-\kappa}$ as $x \to \infty$ for some constant $c > 0$; in particular, $F \in D(G_\gamma)$ with $\gamma = 1/\kappa$.

This result was refined by Goldie (1989), who gave a characterization of a constant $\rho > 0$ such that

$$1 - F(x) = cx^{-1/\gamma}\big(1 + O(x^{-\rho/\gamma})\big) \tag{4.2}$$

as $x \to \infty$ under the additional moment assumptions

(S2) There exists $\xi > 0$ such that $EA_1^{\kappa+\xi} < \infty$ and $EB_1^{\kappa+\xi} < \infty$.

By a simple inversion it can be concluded that $F^{-1}(1-t) = (t/c)^{-\gamma}(1+O(t^\rho))$ as $t \downarrow 0$, which in turn implies (3.4) with $a(\lambda) = \gamma F^{-1}(1-\lambda)$. Moreover, the left-hand side of (3.5) is of the order $k_n^{1/2}(k_n/n)^\rho$ and thus converges to 0 if

$$k_n = o(n^{2\rho/(2\rho+1)}). \tag{4.3}$$

Next we check the conditions $(\widetilde{\text{C1}})$–$(\widetilde{\text{C3}})$. According to Doukhan (1995), Corollary 2.4.1, the time series is geometrically β-mixing; that is, $\beta(k) = O(\eta^k)$ for some $\eta \in (0,1)$. Hence $(\widetilde{\text{C1}})$ holds for

$$l_n = [-2\log n/\log\eta] \tag{4.4}$$

and

$$\log^2 n \log^4(\log n) = o(k_n). \tag{4.5}$$

LEMMA 4.1. *If l_n is given by* (4.4) *and $k_n = o(n/l_n)$, then the conditions* $(\widetilde{\text{C2}})$ *and* $(\widetilde{\text{C3}})$ *hold with*

$$\begin{aligned} r(x,y) = \min(x,y) + \sum_{j=1}^{\infty}\biggl(x\int_0^{y/x} P\biggl\{\prod_{i=1}^{j} A_i > t^\gamma\biggr\}\,dt \\ + y\int_0^{x/y} P\biggl\{\prod_{i=1}^{j} A_i > t^\gamma\biggr\}\,dt\biggr). \end{aligned} \tag{4.6}$$

The proof of $(\widetilde{\text{C2}})$ is essentially due to Stărică (1999), Lemma 3.2.

Since the marginal d.f. F of a nondegenerate stationary solution of (4.1) is continuous [Vervaat (1979), Theorem 3.2], Theorem 3.1 applies. Here we restrict ourselves to a particular weight function where the extra condition $s_n \to 0$ is not needed.

COROLLARY 4.1. *If the conditions* (S1), (S2) *and* (4.2) *hold and k_n satisfies* (4.3) *and* (4.5), *then there exist versions of Q_n and e such that*

$$\sup_{t\in(0,1]} \frac{t^{\gamma+3/4}}{1-\log t}\left| k_n^{1/2}\left(\frac{Q_n(t)}{\gamma F^{-1}(1-k_n/n)} - \bar{F}_\gamma^{-1}(t)\right) - t^{-(\gamma+1)}e(t)\right| \longrightarrow 0$$

in probability.

Next, using Corollaries 3.2 and 3.3, we investigate the asymptotic behavior of the aforementioned maximum likelihood estimator $\hat{\gamma}_n^{(\mathrm{ML})}$ and the Hill estimator $\hat{\gamma}_n^{(\mathrm{H})}$ for the extreme value index γ. Recall that generally the latter only works for $\gamma > 0$, which holds true in the present situation, whereas the maximum likelihood estimator is suitable in case of $\gamma > -1/2$ (and a modification may be used for estimating arbitrary $\gamma \in \mathbb{R}$). Both estimators can be represented as a statistical tail functional, but only $\hat{\gamma}_n^{(\mathrm{ML})}$ is location invariant.

In view of (3.10), Corollary 3.2 yields

$$\mathscr{L}(k_n^{1/2}(\hat{\gamma}_n^{(\mathrm{ML})} - \gamma)) \longrightarrow \mathscr{N}(0, \sigma^2_{\mathrm{ML},\gamma}) \quad \text{weakly},$$

where straightforward calculations utilizing $r(tx, ty) = tr(x, y)$ show that

$$\begin{aligned}\sigma^2_{\mathrm{ML},\gamma} = {}& \frac{(\gamma+1)^4}{\gamma^2}\int_0^1\int_0^1 (st)^{-1}(1-(2\gamma+1)s^\gamma)(1-(2\gamma+1)t^\gamma)r(s,t)\,ds\,dt \\ & + 2\frac{(\gamma+1)^3}{\gamma}\int_0^1 s^{-1}(1-(2\gamma+1)s^\gamma)r(s,1)\,ds + (\gamma+1)^2 r(1,1) \\ = {}& (\gamma+1)^2 r(1,1).\end{aligned}$$

In case of the Hill estimator, first notice that $a(\lambda) = \gamma F^{-1}(1-\lambda)$ implies $\Xi_n = 0$. Therefore, under the conditions of Corollary 4.1, Corollary 3.3 yields

$$\mathscr{L}\Big(k_n^{1/2}\Big(\hat{\gamma}_n^{(H)} - \gamma\Big)\Big) \longrightarrow \mathscr{N}(0, \sigma^2_{H,\gamma}) \quad \text{weakly},$$

with

$$\begin{aligned}\sigma^2_{H,\gamma} = \gamma^2\Big(&\int_0^1\int_0^1 (st)^{-1} r(s,t)\,ds\,dt \\ & - 2\int_0^1 s^{-1} r(s,1)\,ds + r(1,1)\Big) = \gamma^2 r(1,1).\end{aligned}$$

This result was proved previously by Stărică (1999) under the additional assumption $k_n = o(n^{\gamma/(\gamma+1)})$, which is very restrictive if γ is close to 0.

It is remarkable that, compared with the i.i.d. case, the asymptotic variance is increased by the same factor $r(1,1) > 1$ for both estimators. This, however, does not hold true for other estimators of γ, like, for example, the Pickands estimator; for this and many other popular estimators of γ, the asymptotic normality can be established in a similar way as an immediate consequence of Corollaries 3.2 and 3.3.

The calculation of the covariance function r and the asymptotic variance as well as applications to ARCH- and GARCH-processes were discussed by Stărică (1999). There it was shown by the example of the foreign exchange rate between Japanese yen and U.S. dollar that in applications to real financial time series one should expect a much higher asymptotic variance and hence much wider confidence intervals than in the i.i.d. case. More precisely, in that case an estimate of 9.32 ± 0.46 for the factor $r(1,1)$ was obtained. Consequently, it is important *not* to use the formula for the asymptotic variance derived from the classical extreme statistics for i.i.d. r.v.'s, which would delude into assuming an estimation error much smaller than the actual one. For GARCH (1,1) time series, an alternative approach leading to more precise estimates of the extreme value index was introduced by Stărică (1999).

Here we restrict ourselves to mentioning that r can be estimated directly using (4.6) if the distribution of the innovations Z_i are known. Moreover, in a semiparametric ARCH- or GARCH-model this distribution can be estimated using kernel estimators (Drost and Klaassen (1997); see also Drost, Klaassen and Werker (1997)] and then, in a second step, one may calculate a plug-in estimator of the covariance function as an alternative to the nonparametric procedure suggested at the end of Section 3.

5. Proofs. In the sequel K and K_i, $i \in \mathbb{N}$, denote generic constants which may vary from line to line and may depend on μ, ν and ε but not on n, x, δ, ϑ or η.

PROOF OF THEOREM 2.1. Denote by $(Y_{n,j}(x))_{x\in[0,1]}$, $1 \le j \le m_n := [n/(2l_n)]$, i.i.d. stochastic processes with the same distribution as $((nv_n)^{-1/2} \times \sum_{i=1}^{l_n}(1_{\{U_i>1-v_n x\}} - v_n x))_{x\in[0,1]}$. Employing condition (C1) and a basic result by Eberlein (1984) about approximations of β-mixing sequences by i.i.d. sequences, it was shown by Rootzén [(1995), Lemma 5.1] that for the proof of (2.2) it suffices to verify the tightness of the process $S_n := \sum_{j=1}^{m_n} Y_{n,j}$.

For this, fix some sequence $\Delta := \Delta_n = o((nv_n)^{-1/2})$ such that $(nv_n)^{-1} = o(\Delta)$ and $1/\Delta \in \mathbb{N}$. Then, by (C3) and the proof of Lemma 5.4 of Rootzén (1995), one has

$$E(S_n(i\Delta) - S_n(j\Delta))^4 \le K\left(\frac{l_n^2}{nv_n}|i-j|\Delta + (i-j)^2\Delta^2\right) \tag{5.1}$$

for all $0 \le i, j \le 1/\Delta$. Next apply Móricz' (1982) theorem (with $\gamma = 4$, $f(b,m) = m\Delta$ and $\varphi(t,m) = K(l_n^2/(nv_n) + m\Delta)^{1/4}$) to obtain for all $\delta > 0$,

$$\begin{aligned}
&E\left(\max_{j\in\{i+1,\dots,\min([1+\delta/\Delta],\,1/\Delta)\}}(S_n(i\Delta) - S_n(j\Delta))^4\right)\\
&\quad\le K[\delta/\Delta]\Delta\left(\sum_{k=0}^{[\log[\delta/\Delta]/\log 2]}\left(\frac{l_n^2}{nv_n} + [[\delta/\Delta]2^{-k}]\Delta\right)^{1/4}\right)^4\\
&\quad\le K\delta\left(\sum_{k=0}^{[\log[\delta/\Delta]/\log 2]}\left(\left(\frac{l_n^2}{nv_n}\right)^{1/4} + (\delta 2^{-k})^{1/4}\right)\right)^4\\
&\quad\le K\delta\left(\frac{l_n^2}{nv_n}\log^4(\delta/\Delta) + \delta\right).
\end{aligned} \tag{5.2}$$

Hence, in view of (C1), for all $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that eventually

$$\begin{aligned}
&P\left\{\max_{j\in\{i+1,\dots,\min([i+\delta/\Delta],\,1/\Delta)\}}|S_n(i\Delta) - S_n(j\Delta)| > \varepsilon\right\}\\
&\quad\le K\delta\left(\frac{l_n^2}{nv_n}\log^4(\delta/\Delta) + \delta\right)\varepsilon^{-4} < \eta\delta
\end{aligned}$$

for all $i \in \{0, \dots, 1/\Delta - 1\}$. Using $(nv_n)^{1/2}\Delta \to 0$, one can conclude in a similar way as in the proof of Lemma 5.6 of Rootzén (1995) that for all $\varepsilon, \eta > 0$ there exists a $\delta > 0$ such that eventually for all $x \in [0,1]$,

$$P\left\{\sup_{y\in[x,\,\min(x+\delta,\,1)]}|S_n(x) - S_n(y)| > \varepsilon\right\} < \eta\delta$$

and thus the asserted tightness by Theorem 8.3 of Billingsley (1968). □

PROOF OF THEOREM 2.2. We take up the approach by Shao and Yu (1996). Since $e_n 1_{[\vartheta,1]}/q \to e1_{[\vartheta,1]}/q$ by Theorem 2.1 and $\inf_{x\in[\vartheta,1]} q(x) > 0$, it suffices to prove that

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Big\{\sup_{x\in(0,\vartheta]} |e_n(x)/q(x)| > \varepsilon\Big\} = 0, \tag{5.3}$$

$$\lim_{\vartheta\downarrow 0} P\Big\{\sup_{x\in(0,\vartheta]} |e(x)/q(x)| > \varepsilon\Big\} = 0 \tag{5.4}$$

for all $\varepsilon > 0$. In the sequel, we assume w.l.o.g. that $q(x) = x^\nu|\log x|^\mu$ for $x \in (0,\vartheta]$ and ϑ sufficiently small such that q is increasing and q/Id is decreasing on $(0,\vartheta]$.

It is readily seen that for the proof of (5.3) it suffices to verify that

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Big\{\sup_{x\in(0,\vartheta]} |S_n(x)/q(x)| > \varepsilon\Big\} = 0 \tag{5.5}$$

(cf. proof of Theorem 2.1). For this, fix some (small) $\eta > 0$ and let

$$i_n := \min\Big\{i \in \mathbb{N}\Big|(nv_n)^{1/2} \le \eta\frac{q}{\mathrm{Id}}(\vartheta e^{-(i+1)})\Big\}.$$

Observe that for $\widetilde{S}_n(x) := S_n(x) + (nv_n)^{-1/2} m_n l_n v_n x$ one has

$$\sup_{x\in[\vartheta e^{-(i+1)},\,\vartheta e^{-i}]} |S_n(x)| \le \widetilde{S}_n(\vartheta e^{-i}) + (nv_n)^{1/2}\frac{\vartheta}{2}e^{-i},$$

where by the monotonicity of q/Id and the definition of i_n,

$$(nv_n)^{1/2}\vartheta e^{-i} \le \eta\vartheta e^{-i}\frac{q}{\mathrm{Id}}(\vartheta e^{-(i+1)}) \le \varepsilon q(\vartheta e^{-(i+1)})$$

for $i \ge i_n$ and $\eta \le \varepsilon/e$. Therefore, we obtain by the Chebyshev inequality,

$$\begin{aligned}
&P\Big\{\sup_{x\in(0,\,\vartheta e^{-i_n}]} |S_n(x)/q(x)| > \varepsilon\Big\}\\
&\quad\le \sum_{i=i_n}^{\infty} P\big\{\widetilde{S}_n(\vartheta e^{-i}) > \varepsilon q(\vartheta e^{-(i+1)})/2\big\}\\
&\quad\le K\sum_{i=i_n}^{\infty} (nv_n)^{1/2}\frac{\mathrm{Id}}{\mathrm{q}}(\vartheta e^{-(i+1)}) \qquad (5.6)\\
&\quad\le K\eta\sum_{i=i_n}^{\infty} e^{(i_n-1)(1-\nu)}\left|\frac{\log(\vartheta e^{-(i_n+1)})}{\log(\vartheta e^{-(i+1)})}\right|^{\mu}\\
&\quad\le K\eta\sum_{j=0}^{\infty} e^{-j(1-\nu)}\left(1+\frac{j}{i_n+1-\log\vartheta}\right)^{-\mu}\\
&\quad\le K\eta \longrightarrow 0
\end{aligned}$$

as $\eta \downarrow 0$.

Next let

$$\Delta_i := \Delta_{i,n,\vartheta,\varepsilon} := \varepsilon q(\vartheta e^{-(i+1)})(nv_n)^{-1/2}. \tag{5.7}$$

Since $S_n((j-1)\Delta_i) - \Delta_i(nv_n)^{1/2}/2 \le S_n(x) \le S_n(j\Delta_i) + \Delta_i(nv_n)^{1/2}/2$ for all $x \in [(j-1)\Delta_i, j\Delta_i]$, we may conclude in the same way as in the proof of (5.2), using the Chebyshev inequality and Móricz' theorem, that

$$\begin{aligned}
&P\Big\{\sup_{x\in(\vartheta e^{-i_n},\,\vartheta]} |S_n(x)/q(x)| > \varepsilon\Big\}\\
&\quad\le \sum_{i=0}^{i_n-1} P\Big\{\max_{[\vartheta e^{-(i+1)}/\Delta_i]\le j\le[\vartheta e^{-i}/\Delta_i]+1} |S_n(j\Delta_i)|\\
&\qquad\qquad + \Delta_i(nv_n)^{1/2}/2 > \varepsilon q(\vartheta e^{-(i+1)})\Big\}\\
&\quad\le \sum_{i=0}^{i_n-1} P\Big\{\max_{[\vartheta e^{-(i+1)}/\Delta_i]\le j\le[\vartheta e^{-i}/\Delta_i]+1} |S_n(j\Delta_i)| > \varepsilon q(\vartheta e^{-(i+1)})/2\Big\}\\
&\quad\le K\sum_{i=0}^{i_n-1} \vartheta e^{-i}\bigg(\frac{l_n^2}{nv_n}\log^4([\vartheta e^{-i}/\Delta_i]+2) + \vartheta e^{-i}\bigg) q^{-4}(\vartheta e^{-(i+1)}).
\end{aligned} \tag{5.8}$$

Moreover,

$$\begin{aligned}
\sum_{i=0}^{i_n=1} (\vartheta e^{-i})^2 q^{-4}(\vartheta e^{-(i+1)}) &\le K\int_1^\infty (\vartheta e^{-t})^{2-4\nu}|\log(\vartheta e^{-t})|^{-4\mu}\,dt\\
&\le K\int_0^{\vartheta/e} u^{1-4\nu}|\log u|^{-4\mu} du \longrightarrow 0
\end{aligned} \tag{5.9}$$

as $\vartheta \downarrow 0$ if $\nu < 1/2$, or $\nu = 1/2$ and $\mu = 1/4$. In view of (5.6)–(5.9), for (5.5) it remains to prove that for all sufficiently small η, $\vartheta > 0$,

$$\frac{l_n^2}{nv_n}\sum_{i=0}^{i_n-1} e^{-i}\log^4([\vartheta e^{-i}/\Delta_i]+2)q^{-4}(\vartheta e^{-(i+1)}) \longrightarrow 0 \tag{5.10}$$

as $n \to \infty$.

From the definition of i_n it is easily see that

$$e^{-i_n} \asymp (nv_n)^{-1/(2(1-\nu))}\log^{\mu/(1-\nu)}(nv_n) \tag{5.11}$$

and $i_n \asymp \log(nv_n)$. [Here, $a_n \asymp b_n$ means $a_n = O(b_n)$ and $b_n = O(a_n)$.] Because of

$$\begin{aligned}
\log(\vartheta e^{-i}/\Delta_i) &\sim \log\Big((nv_n)^{1/2}\frac{\mathrm{Id}}{\mathrm{q}}(\vartheta e^{-(i+1)})\Big)\\
&\sim \log\frac{(\mathrm{Id}/q)(\vartheta e^{-(i+1)})}{(\mathrm{Id}/q)(\vartheta e^{-(i_n+1)})} \sim (1-\nu)(i_n - i),
\end{aligned}$$

the left-hand side of (5.10) is of the order

$$\frac{l_n^2}{nv_n}\sum_{i=0}^{i_n-1} e^{(4\nu-1)i}(i_n-i)^4(i+1-\log\vartheta)^{-4\mu} = O\left(\frac{l_n^2}{nv_n}\log^4(nv_n)\sum_{i=0}^{i_n-1} e^{(4\nu-1)i}(i+1-\log\vartheta)^{-4\mu}\right). \tag{5.12}$$

Thus, for $\nu < 1/4$, and for $\nu = 1/4$ and $\mu > 1/4$, convergence (5.10) is immediate from (C1).

If $\nu = 1/4$ and $\mu = 1/4$ or $\mu > 1/4$, then the sum on the right-hand side of (5.12) is of the order $\log i_n$ and $i_n^{1-4\mu}$, respectively, and (5.10) is a consequence of $s_n \to 0$.

Finally, for $1/4 < \nu \le 1/2$, the sum on the left-hand side of (5.12) can be rewritten as

$$\begin{aligned} &e^{(4\nu-1)i_n} i_n^{-4\mu}\sum_{j=1}^{i_n} e^{-(4\nu-1)j} j^{-4}\left(1-\frac{j-1+\log\vartheta}{i_n}\right)^{-4\mu} \\ &\quad = O\big(e^{(4\nu-1)i_n} i_n^{-4\mu}\sum_{j=1}^{i_n} e^{-(4\nu-1)j} j^{4+4\max(\mu,0)}\big) = O\big(e^{(4\nu-1)i_n} i_n^{-4\mu}\big) \\ &\quad = O\big((nv_n)^{(4\nu-1)/(2(1-\nu))}\log^{-3\mu/(1-\nu)}(nv_n)\big), \end{aligned}$$

where in the second step the inequality $i_n j - j(j-1+\log\vartheta) \ge i_n$ has been used, which holds for $j \in \{1, i_n\}$ and hence for $1 \le j \le i_n$. Now (5.10) can be derived from (5.12) and $s_n \to 0$.

To prove (5.4), first note that by (C2) and (C3),

$$E(e(y)-e(x))^2 = r(y,y) - 2r(x,y) + r(x,x) \le C|y-x|.$$

Hence Lemma 4.1.3 of Fernique (1975) yields for sufficiently small ϑ,

$$\begin{aligned} &P\Big\{\sup_{x\in(0,\vartheta]} |e(x)/q(x)| > \varepsilon\Big\} \\ &\quad \le \sum_{i=0}^{\infty} P\Big\{\sup_{x\in(\vartheta e^{-(i+1)},\,\vartheta e^{-i})} |e(x)| > \varepsilon(\vartheta e^{-(i+1)})^{\nu}|\log(\vartheta e^{-(i+1)})|^{\mu}\Big\} \\ &\quad \le K\sum_{i=0}^{\infty}\big(1-\Phi(K_1(\vartheta e^{-(i+1)})^{\nu-1/2}|\log(\vartheta e^{-(i+1)})|^{\mu})\big) \\ &\quad \le K\sum_{i=0}^{\infty}(\vartheta e^{-(i+1)})^{1/2-\nu}|\log(\vartheta e^{-(i+1)})|^{-\mu} \\ &\qquad\qquad \times \exp\big(-K_1(\vartheta e^{-i})^{(2\nu-1)}|\log(\vartheta e^{-(i+1)})|^{2\mu}\big) \\ &\quad \le K\vartheta^{1/2-\nu}|\log\vartheta|^{-\mu}\sum_{i=0}^{\infty} e^{-(1/2-\nu)i}\exp\big(-K_1 i^{2\mu}(\vartheta e^{-i})^{2\nu-1}\big), \end{aligned}$$

where Φ denotes the standard normal d.f. Now (5.4) follows from the fact that the last sum is bounded, provided $\nu < 1/2$, or $\nu = 1/2$ and $\mu > 0$. □

PROOF OF THEOREM 2.3. It suffices to verify that

$$\lim_{\vartheta \downarrow 0} \limsup_{n\to\infty} P\left\{ \sup_{x\in(\vartheta e^{-i_n},\, \vartheta]} |S_n(x)/q(x)| > \varepsilon \right\} = 0 \tag{5.13}$$

[cf. (5.8)], because the rest of the proof of Theorem 2.2 can be copied literally. By condition (C3*), $E(Y_{n,1}(y) - Y_{n,1}(x))^j \le C l_n v_n (n v_n)^{-j/2}(y-x)$ for all $0 \le x \le y \le 1$ and $j \in \{2, 4\}$. Hence, using Burkholder's inequality we obtain

$$\begin{aligned} E(S_n(y) - S_n(x))^4 &\le K(m_n E(Y_{n,1}(y) - Y_{n,1}(x))^4 \\ &\qquad + m_n(m_n - 1)(E(Y_{n,1}(y) - Y_{n,1}(x))^2)^2) \\ &\le K\left(\frac{y-x}{n v_n} + (y-x)^2\right). \end{aligned}$$

Since, according to the definition of i_n and (5.11), Δ_i defined by (5.7) is of larger order than $(n v_n)^{-1}$, in particular

$$E(S_n(j\Delta_i) - S_n(l\Delta_i))^4 \le K(l-j)^2 \Delta_i^2. \tag{5.14}$$

By the same arguments as used for proving (5.8) one may conclude that

$$P\left\{ \sup_{x\in(\vartheta e^{-i_n},\, \vartheta]} |S_n(x)/q(x)| > \varepsilon \right\} \le K \sum_{i=0}^{i_n - 1} (\vartheta e^{-i})^2 q^{-4}(\vartheta e^{-(i+1)})$$

and thus (5.9) implies (5.13). □

PROOF OF LEMMA 3.1. Denote by E_n the tail empirical d.f.,

$$E_n(x) := \frac{1}{k_n} \sum_{i=1}^{n} 1_{\{U_i > 1 - (k_n/n)x\}}. \tag{5.15}$$

Check that

$$V_n(t) = \inf\{x \mid E_n(x) > t\} = E_n^{-1}(t+0). \tag{5.16}$$

Hence (3.1) is equivalent to

$$\begin{aligned} &\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\left\{ \sup_{t\in[1/(2k_n),\, 1]} \frac{E_n^{-1}(t)}{t} > \lambda \right\} \\ &\quad = \lim_{\lambda\to\infty} \limsup_{n\to\infty} P\{E_n(\lambda t) < t \text{ for some } t \in [1/(2k_n),\, 1]\} = 0. \end{aligned}$$

By the same arguments as in the proof of Theorem 2.1 [see Rootzén (1995), proof of Lemma 5.1], it suffices to prove that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\{\widetilde{E}_n(\lambda t) < t/2 \text{ for some } t \in [1/(2k_n), 1]\} = 0,$$

where $\widetilde{E}_n = \sum_{j=1}^{m_n} Z_{nj}$ and $Z_{n,j}$, $1 \le j \le m_n := [n/(2l_n)]$, are i.i.d. copies of the process $(k_n^{-1} \sum_{i=1}^{l_n} 1_{\{U_i > 1 - xk_n/n\}})_{x>0}$, that is, $(E_n(x))_{x\in[0,1+\varepsilon]} =^d (k_n^{-1/2} S_n(x) + x m_n l_n/n)_{x\in[0,1+\varepsilon]}$ with S_n defined in the proof of Theorem 2.1.

Using the tightness of $(S_n(x))_{x\in[0,1+\varepsilon]}$, which can be verified in the same way as in the proof of Theorem 2.1, the Chebyshev inequality and (C3), one obtains

$$\begin{aligned}
&P\big\{\widetilde{E}_n(\lambda t) < t/2 \text{ for some } t \in [1/(2k_n), 1]\big\}\\
&\quad\le P\{\widetilde{E}_n(1+\varepsilon) < 1/2\}\\
&\qquad + P\big\{\widetilde{E}_n(\lambda t) < t/2 \text{ for some } t \in [1/(2k_n), (1+\varepsilon)/\lambda]\big\}\\
&\quad\le P\left\{S_n(1+\varepsilon) < k_n^{1/2}\left(\frac{1}{2} - \frac{1+\varepsilon}{2+\varepsilon}\right)\right\}\\
&\qquad + P\Big\{\widetilde{E}_n((1+\varepsilon)e^{-(i+1)})\frac{1+\varepsilon}{2\lambda}e^{-i}\\
&\qquad\qquad \text{for some } 0 \le i \le i_{n,\lambda} := [\log(2(1+\varepsilon)k_n/\lambda)]\Big\}\\
&\quad= \sum_{i=1}^{i_{n,\lambda}} P\left\{S_n((1+\varepsilon)e^{-(i+1)}) < k_n^{1/2}(1+\varepsilon)\left(\frac{e^{-i}}{2\lambda} - \frac{e^{-(i+1)}}{2+\varepsilon}\right)\right\} + o(1)\\
&\quad\le \sum_{i=0}^{i_{n,\lambda}} m_n C \frac{l_n}{n}(1+\varepsilon)e^{-(i+1)} k_n^{-1}(1+\varepsilon)^{-2}\left|\frac{e^{-i}}{2\lambda} - \frac{e^{-(i+1)}}{2+\varepsilon}\right|^{-2} + o(1)\\
&\quad\le K\frac{e^{i_{n,\lambda}}}{k_n}\left|\frac{1}{\lambda} - \frac{1}{(2+\varepsilon)e}\right|^{-2} + o(1)\\
&\quad\le \frac{K}{\lambda}\left|\frac{1}{\lambda} - \frac{1}{(2+\varepsilon)e}\right|^{-2} + o(1)\\
&\quad\to 0 \quad \text{as } \lambda \to \infty
\end{aligned}$$

and thus (3.1).

Likewise, for the proof of (3.2) it suffices to verify that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\big\{\widetilde{E}_n(t/\lambda) \ge t/(2+\delta) \text{ for some } t \in (0,1]\big\} = 0$$

for some $\delta > 0$, which in turn is a consequence of

$$\begin{aligned}
&P\big\{\widetilde{E}_n(t/\lambda) \ge t/(2+\delta) \text{ for some } t \in (0,1]\big\}\\
&\qquad\le P\left\{\widetilde{E}_n\Big(\frac{2+\delta}{\lambda k_n}\Big) \ge \frac{1}{k_n}\right\}
\end{aligned}$$

$$
\begin{aligned}
&+P\Big\{\widetilde{E}_n(e^{-i}) \ge \frac{\lambda}{2+\delta}e^{-(i+1)} \\
&\qquad \text{for some } 0 \le i \le i_{n,\lambda} := [\log(\lambda k_n/(2+\delta))]\Big\} \\
&\le m_n l_n \frac{2+\delta}{\lambda n} + \sum_{i=0}^{i_{n,\lambda}} P\Big\{S_n(e^{-i}) \ge k_n^{1/2}\Big(\frac{\lambda e^{-(i+1)}}{2+\delta} - \frac{e^{-i}}{2}\Big)\Big\} \\
&\le \frac{2+\delta}{2\lambda} + \sum_{i=0}^{i_{n,\lambda}} \frac{C}{2} e^{-i} k_n^{-1} e^{2i} \Big(\frac{\lambda}{(2+\delta)e} - \frac{1}{2}\Big)^{-2} \\
&\le \frac{2+\delta}{2\lambda} + K\lambda \Big(\frac{\lambda}{(2+\delta)e} - \frac{1}{2}\Big)^{-2} \\
&\to 0 \quad \text{as } \lambda \to \infty. \qquad \square
\end{aligned}
$$

PROOF OF COROLLARY 3.1. According to Theorem 2.1 the assumptions ensure that $(e_n(x))_{x\in[0,1+\varepsilon]} \to (e(x))_{x\in[0,1+\varepsilon]}$ weakly. Because of $e_n = k_n^{1/2}(E_n - \mathrm{Id})$ with E_n defined by (5.15), $-e =^d e$ and (5.16), an obvious modification of Vervaat's (1972) Theorem 1 leads to $k_n^{1/2}(V_n - \mathrm{Id}) \to e$ weakly in $D[0,1]$. It remains to prove that

$$
\lim_{\vartheta \downarrow o} \limsup_{n\to\infty} P\Big\{\sup_{t\in[1/(2k_n),\,\vartheta]} k_n^{1/2} \frac{|V_n(t)-t|}{q(t)} > \eta\Big\} = 0
$$

for all $\eta > 0$ (cf. proof of Theorem 2.2). For that purpose, we may assume w.l.o.g. that $q(t) = t^\nu |\log t|^\mu$, $t \in (0, \vartheta]$ for sufficiently small $\vartheta > 0$ so that q is increasing and q/Id is decreasing.

According to Lemma 3.1, for fixed $\delta > 0$ there exists $\lambda > 0$ such that

$$
\limsup_{n\to\infty} P\Big\{\sup_{t\in[1/(2k_n),\,1]} \frac{|V_n(t)-t|}{t} > \lambda\Big\} < \delta.
$$

Therefore, with $t_n := \sup\{t \in (0, \vartheta] \,|\, \lambda k_n^{1/2} t/q(t) \le \eta\}$ we have

$$
\limsup_{n\to\infty} P\Big\{\sup_{t\in[1/(2k_n),\,t_n]} k_n^{1/2} \frac{|V_n(t)-t|}{q(t)} > \eta\Big\} < \delta.
$$

Furthermore, (5.16) implies

$$
\begin{aligned}
&P\Big\{\sup_{t\in[t_n,\,\vartheta]} k_n^{1/2} \frac{V_n(t)-t}{q(t)} > \eta\Big\} \\
&\quad = P\Big\{E_n(t + \eta q(t) k_n^{-1/2}) < t \text{ for some } t \in [t_n, \vartheta]\Big\} \qquad (5.17) \\
&\quad = P\Big\{e_n(t + \eta q(t) k_n^{-1/2}) < -\eta q(t) \text{ for some } t \in [t_n, \vartheta]\Big\}.
\end{aligned}
$$

Observe that $\eta q(t)k_n^{-1/2} \le \lambda t$ for all $t \in [t_n, \vartheta]$ which implies

$$\inf_{t\in[t_n,\vartheta]} \frac{q(t)}{q(t+\eta q(t)k_n^{-1/2})} \ge \inf_{t\in[t_n,\vartheta]} \frac{q(t)}{q((\lambda+1)t)} \ge \frac{1}{2}(\lambda+1)^{-(\nu+1)}$$

for all sufficiently small $\vartheta > 0$. Hence, the convergence of the right-hand side of (5.17) to 0 as ϑ tends to 0 is an immediate consequence of (5.3).

Likewise one obtains

$$\lim_{\vartheta\downarrow 0} \limsup_{n\to\infty} P\left\{ \sup_{t\in[t_n,\,\vartheta]} -k_n^{1/2}\frac{V_n(t)-t}{q(t)} > \eta \right\} = 0$$

and thus the assertion. □

PROOF OF THEOREM 3.1. Essentially we follow the lines of the proof of Theorem 2.1 in Drees (1998a). Again we may assume that $q(t) = t^\nu|\log t|^\mu$ on some neighborhood of 0. Let $\tau := 0$ if $t^{\gamma+1}/q(t) \to 0$ as $t \downarrow 0$ (i.e., if $\nu < \gamma+1$, or $\nu = \gamma+1$ and $\mu > 0$), and $\tau := 1/2$ else.

Since F is continuous, the r.v.'s $U_i := F(X_i)$ are uniformly distributed. Moreover, $X_i = F^{-1}(U_i)$ and hence $Q_n = F^{-1}(1 - V_n k_n/n)$ almost surely. Consequently,

$$\begin{aligned} &\sup_{t\in[\tau/k_n,1]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2} \left| \frac{Q_n(t) - F^{-1}(1-k_n/n)}{a(k_n/n)} - F_\gamma^{-1}(V_n(t)) \right| \\ &\quad\le \sup_{t\in(0,\,V_n(1)]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2} |R(k_n/n,t)| \sup_{t\in[\tau/k_n,1]} \left(\frac{t}{V_n(t)}\right)^{\gamma+1} \frac{q(V_n(t))}{q(t)} \longrightarrow 0 \end{aligned} \tag{5.18}$$

in probability, where in the last step (3.5) and Lemma 3.1 were utilized.

Next we check that the assumptions of Corollary 3.1 are satisfied. Condition (C1) follows from ($\widetilde{C}1$) and the fact that the mixing coefficients $\beta(k)$ for the sequence $(U_i)_{i\in\mathbb{N}}$ are less than or equal to the coefficients pertaining to $(X_i)_{i\in\mathbb{N}}$ [see, e.g., Doukhan (1995), equality 1.1(2′)]. ($\widetilde{\mathrm{C}2}$), ($\widetilde{\mathrm{C}3}$) and ($\widetilde{\mathrm{C}3^*}$), respectively, combined with the equivalence

$$X_i > F^{-1}\left(1 - \frac{k_n}{n}x\right) \iff U_i > 1 - \frac{k_n}{n}x$$

imply the corresponding conditions for the sequence $(U_i)_{i\in\mathbb{N}}$. Hence, according to Corollary 3.1, $k_n^{1/2}(V_n - \mathrm{Id})/q1_{[1/(2k_n),\,1]} \to e/q$ weakly in $D[0,1]$. By virtue of Skorohod's representation theorem, we may assume that the convergence holds a.s. Thus, by the mean value theorem applied to $t \mapsto t^{-\gamma}$, there exist

$\xi_t \in (0,1)$ such that

$$\begin{aligned}
&\sup_{t\in[1/(2k_n),1]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2}\Big|F_\gamma^{-1}(V_n(t)) - F_\gamma^{-1}(t) - k_n^{-1/2} t^{-(\gamma+1)} e(t)\Big| \\
(5.19)\quad &\le \sup_{t\in[1/(2k_n),1]} \frac{1}{q(t)}\Big|\big(1+\xi_t(V_n(t)/t-1)\big)^{-(\gamma+1)} k_n^{1/2}(V_n(t)-t) - e(t)\Big| \\
&\le \sup_{t\in[1/(2k_n),1]} \frac{k_n^{1/2}(V_n(t)-t)}{q(t)}\Big|\big(1+\xi_t(V_n(t)/t-1)\big)^{-(\gamma+1)} - 1\Big| + o(1).
\end{aligned}$$

Now recall that for all $\eta > 0$,

$$\lim_{\vartheta\downarrow 0} \limsup_{n\to\infty} P\bigg\{\sup_{t\in(1/(2k_n),\vartheta]} \frac{k_n^{1/2}|V_n(t)-t|}{q(t)} > \eta\bigg\} = 0$$

and for all $\vartheta \in (0,1]$,

$$\limsup_{n\to\infty} P\bigg\{\sup_{t\in[\vartheta,1]} \Big|\big(1+\xi_t(V_n(t)/t-1)\big)^{-(\gamma+1)} - 1\Big| > \eta\bigg\} = 0.$$

Together with the stochastic boundedness of $k_n^{1/2}|V_n(t)-t|/q(t)$ and $(1+\xi_t(V_n(t)/t-1))^{-(\gamma+1)}$ uniformly for $t \in [1/(2k_n),1]$, this shows the convergence of (5.19) to 0 in probability. If $t^{\gamma+1}/q(t) \to 0$ as $t \downarrow 0$, then one even gets convergence uniformly for $t \in (0,1]$, because $V_n(t) = V_n(1/(2k_n))$ for $t \in (0, 1/(2k_n)]$ and

$$\begin{aligned}
&\sup_{t\in(0,1/(2k_n)]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2}\Big|F_\gamma^{-1}(t) - F_\gamma^{-1}(1/(2k_n)) \\
&\qquad\qquad + k_n^{-1/2}\left(t^{-(\gamma+1)}e(t) - (2k_n)^{\gamma+1} e(1/(2k_n))\right)\Big| \\
&\quad = O_P\bigg(\sup_{t\in(0,1/(2k_n)]} \frac{t}{q(t)} k_n^{1/2} + \sup_{t\in(0,1/(2k_n)]} \frac{|e(t)|}{q(t)}\bigg) = o_P(1).
\end{aligned}$$

A combination with (5.18) yields the assertion in case of $t^{\gamma+1}/q(t) \to 0$. Otherwise we have $\gamma < -1/2$ and hence the convergence follows from

$$\begin{aligned}
&\sup_{t\in[1/(2k_n),1]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2}\left|\frac{D_n - F(1-k_n/n)}{a(k_n/n)}\right| \\
&\quad = \sup_{t\in[1/(2k_n),1]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2}\left|\frac{Q_n(1/(2k_n)) - F(1-k_n/n)}{a(k_n/n)} + \frac{1}{\gamma}\right| \\
&\quad = \sup_{t\in[1/(2k_n),1]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2}\left|\frac{(2k_n)^\gamma}{\gamma} + k_n^{-1/2}(2k_n)^{\gamma+1} e(1/(2k_n))\right| + o_P(1) = o_P(1)
\end{aligned}$$

and

$$\sup_{t\in[0,\,1/(2k_n)]} \frac{t^{\gamma+1}}{q(t)} k_n^{1/2} \left| \frac{Q_n(t)-D_n}{a(k_n/n)} - F_\gamma^{-1}(t) - k_n^{-1/2} t^{-(\gamma+1)} e(t) \right|$$

$$\leq \sup_{t\in(0,\,1/(2k_n)]} \frac{k_n^{1/2} t}{\gamma q(t)} + \frac{|e(t)|}{q(t)} = o_P(1). \qquad \square$$

PROOF OF LEMMA 4.1. According to Stărică [(1999), Lemma 3.2(e)] and Rootzén (1995), (2.4) $\Longleftrightarrow$ (2.6),

$$\frac{n}{l_n k_n} \mathrm{Cov}\left(\sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-k_n/l_n)u\}}, \sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-k_n/n)v\}} \right)$$

$$\longrightarrow (\max(u,v))^{-1/\gamma} + \sum_{j=1}^{\infty} \left(v^{-1/\gamma} \int_{u/v}^{\infty} P\left\{ \prod_{i=1}^{j} A_i > t^{-1} \right\} \gamma t^{-(1/\gamma+1)}\, dt \right.$$

$$\left. + u^{-1/\gamma} \int_{v/u}^{\infty} P\left\{ \prod_{i=1}^{j} A_i > t^{-1} \right\} \gamma t^{-(1/\gamma+1)}\, dt \right)$$

for all $u, v > 0$. Since the left-hand side is a monotone function of u and v for all n, the convergence holds locally uniformly. Hence, in view of (3.4) with $a(\lambda) = \gamma F^{-1}(1-\lambda)$, we may replace $F^{-1}(1-k_n/n)u$ by $F^{-1}(1-(k_n/n)u^{-1/\gamma})$. A change of variables $x = u^{-1/\gamma}$ and $y = v^{-1/\gamma}$ leads to ($\widetilde{\mathrm{C}2}$).

For the proof of ($\widetilde{\mathrm{C}3}$), let $\prod_{i+1,\,j} := \prod_{k=i+1}^{j} A_k$ and $Y_{i+1,\,j} := \sum_{k=i+1}^{j} \prod_{k+1,\,j} \times B_k$. Then an iteration of the stochastic difference equation (4.1) yields $X_j = Y_{i+1,\,j} + \prod_{i+1,\,j} X_i$, $1 \leq i < j$, where $(Y_{i+1,\,j}, \prod_{i+1,\,j})$ and X_i are independent. Next observe that $EA_1^{\xi} < 1$ for $\xi \in (0, \kappa)$ [see, e.g., Stărică (1999), (2.8)]. Hence, with $I_n(x,y) := (F^{-1}(1-k_n/ny), F^{-1}(1-k_n/nx)]$, one has

$$P\{X_i \in I_n(x,y), X_j \in I_n(x,y)\}$$

$$\leq P\left\{ X_i \in I_n(x,y), Y_{i+1,\,j} > F^{-1}\left(1 - \frac{k_n}{n} y\right) \Big/ 2 \right\}$$

$$+ P\left\{ X_i \in I_n(x,y), \textstyle\prod_{i+1,\,j} X_i > F^{-1}\left(1 - \frac{k_n}{n} y\right) \Big/ 2 \right\}$$

$$\leq P\{X_i \in I_n(x,y)\} P\left\{ X_j > F^{-1}\left(1 - \frac{k_n}{n} y\right) \Big/ 2 \right\}$$

$$+\int_{I_n(x,y)} P\Big\{\prod\nolimits_{i+1,\,j} u > F^{-1}\Big(1-\frac{k_n}{n}y\Big)\Big/2\Big\} F(du)$$

$$\le \frac{k_n}{n}(y-x)\bigg(1-F\bigg(F^{-1}\Big(1-\frac{k_n}{n}y\Big)\Big/2\bigg)\bigg)$$

$$+\int_{I_n(x,y)} E\Big(\prod\nolimits^{\xi}_{i+1,\,j}\Big)\bigg(\frac{F^{-1}(1-(k_n/n)y)}{2u}\bigg)^{-\xi} F(du)$$

$$\le \frac{k_n}{n}(y-x)2^{1+1/\gamma}\frac{k_n}{n}y + (EA_1^{\xi})^{j-i}2^{\xi}\frac{k_n}{n}\int_x^y\bigg(\frac{F^{-1}(1-(k_n/n)y)}{F^{-1}(1-(k_n/n)v)}\bigg)^{-\xi} dv,$$

where in the last step (4.2) and a change of variables were used. The Potter bounds [Bingham, Goldie and Teugels (1987), Theorem 1.5.6] yield

$$\int_x^y\bigg(\frac{F^{-1}(1-(k_n/n)y)}{F^{-1}(1-(k_n/n)v)}\bigg)^{-\xi} dv \le 2\int_x^y (y/v)^{\xi\gamma+\tau}\,dv$$

$$= \frac{2y}{1-\xi\gamma-\tau}\Big(1-(x/y)^{1-\xi\gamma-\tau}\Big)$$

$$\le \frac{2}{1-\xi\gamma-\tau}(y-x)$$

for $\tau\in(0,1-\xi\gamma)$. Thus, for generic constants K_i,

$$E\bigg(\sum_{i=1}^{l_n} 1_{\{X_i\in I_n(x,y)\}}\bigg)^2$$

$$= \sum_{i=1}^{l_n} P\{X_i\in I_n(x,y)\}$$

$$+2\sum_{1\le i<j\le l_n} P\{X_i\in I_n(x,y),\, X_j\in I_n(x,y)\}$$

$$\le \frac{l_nk_n}{n}(y-x)+K_1\bigg(\frac{l_nk_n}{n}\bigg)^2(y-x)+K_2\frac{l_nk_n}{n}\sum_{j=1}^{\infty}(EA_1^{\xi})^j(y-x)$$

$$\le K_3\frac{l_nk_n}{n}(y-x),$$

and thus condition ($\widetilde{\text{C3}}$). □

Acknowledgments. I started the research presented in the paper during a visit to the Center for Stochastic Processes of the University of North Carolina at Chapel Hill and continued it while visiting the Department of Mathematical Statistics of Chalmers University. Grateful acknowledgment is made for their hospitality, particularly Ross Leadbetter and Holger Rootzén.

Moreover, I thank Catalin Stărică for helpful and inspiring discussions about extreme value statistics for financial time series.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

DOUKHAN, P. (1995). *Mixing. Properties and Examples*. Springer, New York.

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

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

DREES, H. (1998c). *Estimating the Extreme Value Index*. Habilitation thesis, Univ. Cologne. Available at `euklid.mi.uni-koeln.de/~hdrees/habil.ps`

DREES, H. and DE HAAN, L. (1999). Conditions for quantile process approximations. *Comm. Statist. Stochastic Models* **15** 485–502.

DROST, F. C. and KLAASSEN, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *J. Econometrics* **81** 193–221.

DROST, F. C., KLAASSEN, C. A. J. and WERKER, B. J. M. (1997). Adaptive estimation in time-series models. *Ann. Statist.* **25** 786–817.

EBERLEIN, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.* **2** 291–293.

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

ENGLE, R. (1982). Autoregressive conditional heteroskedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007.

FERNIQUE, X. M. (1975). Regularite des trajectoires des fonctions aleatoires gaussiennes. *Ecole d'Eté de Probabilités des Saint-Flour IV. Lecture Notes in Math.* 1–96. Springer, Berlin.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method I. *Scand. J. Statist.* **16** 97–128.

GOLDIE, C. M. (1989). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1** 126–166.

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

HSING, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569.

KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

MOORE, D. S. and SPRUILL, M. C. (1975). Unified large-sample theory for general chi-squared statistics for tests of fit. *Ann. Statist.* **3** 599–616.

MÓRICZ, F. (1982). A general moment inequality for the maximum of partial sums of single series. *Acta Sci. Math.* (*Szeged*) **44** 67–75.

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

PURI, M. L. and TRAN, L. T. (1980). Empirical distribution functions and functions of order statistics for mixing random variables. *J. Multivariate Anal.* **10** 402–425.

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

RESNICK, S. and STĂRICĂ, C. (1997). Asymptotic behavior of Hill's estimator for autoregressive data. *Comm. Statist. Stochastic Models* **13** 703–721.

RESNICK, S. and STĂRICĂ, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Probab.* **8** 1156–1183.

ROOTZÉN, H. (1995). The tail empirical process for stationary sequences. Preprint, Chalmers Univ. Gothenburg. Available at `www.math.chalmers.se/Stat/Research/Preprints/ms95-hr.ps`

ROOTZÉN, H. (1998). Example which satisfies (1.1), (2.2), and (2.4) of "The tail empirical process", but not the conclusion of Theorem 2.1. Unpublished manuscript.

SHAO, Q. M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stochastic Process. Appl.* **48** 319–334.

SHAO, Q. M. and YU, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* **24** 2098–2127.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

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

STĂRICĂ, C. (1999). On the tail empirical process of solutions of stochastic difference equations. Preprint, Chalmers Univ. Gothenburg. Available at `www.math.chalmers.se/~starica/resume/aarch.ps.gz`.

VERVAAT, W. (1972). Functional centeral limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. Verw. Gebiete* **23** 245–253.

VERVAAT, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. Appl. Probab.* **11** 750–783.

INSTITUTE FOR APPLIED MATHEMATICS
UNIVERSITY OF HEIDELBERG
IM NEUENHEIMER FELD 294
69120 HEIDELBERG
GERMANY
E-mail: hdrees@statLab.uni-heidelberg.de