Bernoulli **9**(1), 2003, 617–657

Extreme quantile estimation for dependent data, with applications to finance

HOLGER DREES

Fachrichtung Mathematik, Universität des Saarlandes, Postfach 151 150, 66041 Saarbrücken, Germany. E-mail: drees@num.uni-sb.de

The asymptotic normality of a class of estimators for extreme quantiles is established under mild structural conditions on the observed stationary β-mixing time series. Consistent estimators of the asymptotic variance are introduced, which render possible the construction of asymptotic confidence intervals for the extreme quantiles. Moreover, it is shown that many well-known time series models satisfy our conditions. The theory is then applied to a time series of stock index returns. Finally, the finite-sample behaviour of the proposed confidence intervals is examined in a simulation study. It turns out that for most time series models under consideration the actual coverage probability is pretty close to the nominal level if the sample fraction used for estimation is chosen appropriately.

Keywords: ARMA model; β-mixing; confidence interval; extreme quantiles; GARCH model; tail empirical quantile function; time series

1. Introduction

Let S_i, $0 \leqslant i \leqslant n$, be a sequence of consecutive share prices. In recent years the Value at Risk (VaR) – defined as a large quantile of the negative log-returns $X_i = -\log(S_i/S_{i-1})$, which are assumed stationary – has become a popular measure of the risk of an investment in these shares. It has long been known that the classical Gaussian models for log-returns (such as the famous Black–Scholes model) underestimate the risk of large losses and hence are not suitable as a basis for VaR estimation. As an alternative, it has been proposed to model series of log-returns by independent random variables with heavy tails (see, for example, Jansen and de Vries 1991; Longin 1996). To take into account the serial dependence which is usually observed in time series of log-returns, a large variety of more sophisticated autoregressive conditional heteroscedastic (ARCH) type models have been introduced since the seminal paper by Engle (1982).

Though some of these models describe real time series of log-returns reasonably well for specific purposes, none of them is able to capture all so-called 'stylized facts', that is, features common to most of these financial data sets; see Mikosch and Stărică (2002) for a comprehensive discussion. In particular, it is questionable whether such a model can well describe both the central part of the distribution and its tails. Therefore, it has recently been suggested that one should 'let the tails speak for themselves', that is, use merely the largest negative log-returns for the estimation of the VaR.

1350–7265

Statistical procedures of this type are provided by extreme value theory under rather mild structural assumptions on the tail of the marginal distributions of log-returns. Unfortunately, almost all results on the asymptotic behaviour of extreme quantile estimators available hitherto are restricted to independent observations. For financial time series, however, it is rarely realistic to assume independence of consecutive observations. Thus the main aim of the present paper is to investigate the asymptotic behaviour of quantile estimators based on large observations under mild assumptions on the serial dependence.

Of course, the results are relevant not only to VaR estimation but also to the tail analysis of any real-life time series if the assumption of independence seems inappropriate. For example, the interarrival times and package lengths in teletraffic networks often exhibit heavy tails and serial dependence as well.

Denote the common distribution function (df) of the stationary time series under consideration, X_i, $i \in \mathbb{N}$, by F. The basic assumption in the extreme value approach is

$$F^n(a_n x + b_n) \to G(x), \qquad x \in \mathbb{R}, \tag{1}$$

for some $a_n > 0$, $b_n \in \mathbb{R}$, where G is a non-degenerate limit df. It is well known that then G must then be one of the extreme value dfs (up to a scale and location parameter)

$$G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \qquad 1+\gamma x \geqslant 0,\ \gamma \in \mathbb{R},$$

which is interpreted as $\exp(\mathrm{e}^{-x})$ for $\gamma = 0$. For the sake of brevity we write $F \in D(G_\gamma)$. If the X_i are independent (or weakly dependent) then (1) is equivalent to the weak convergence of the df of the standardized maximum of n observations to G. In general, though, the maximum of a stationary time series is stochastically smaller than the maximum of an independent and identically distributed (i.i.d.) sequence with the same marginal df, since the serial dependence leads to a clustering of large values. Indeed, under mild conditions on the dependence structure, $F \in D(G_\gamma)$ implies

$$\mathcal{L}\left(a_n^{-1}\left(\max_{1\leqslant i\leqslant n} X_i - b_n\right)\right) \to G_\gamma^\theta \qquad \text{weakly} \tag{2}$$

for some $\theta \in [0, 1]$; see Leadbetter *et al*. (1983, Section 3.7), for details. Typically the so-called extremal index θ can be interpreted as the reciprocal value of the asymptotic mean cluster size.

Note that, by (1) and (2),

$$P\left\{a_n^{-1}\left(\max_{1\leqslant i\leqslant n} X_i - b_n\right) \leqslant x\right\} \sim P\left\{a_n^{-1}\left(\max_{1\leqslant i\leqslant [n\theta]} \tilde{X}_i - b_n\right) \leqslant x\right\}$$

as $n \to \infty$, where $\tilde{X}_i$, $i \in \mathbb{N}$, is an i.i.d. sequence with marginal df F. (Here $c_n \sim d_n$ means $c_n/d_n \to 1$, and $[x]$ denotes the largest integer smaller than or equal to x.) Hence, as far as the behaviour of the maximum is concerned, the serial dependence between large observations reduces the effective sample size by the factor θ. Since intuitively a cluster of large observations contains less information about F than the same number of independent large observations, one will also expect an influence of the serial dependence on the precision of statistical extreme value procedures. More precisely, the dependence will lead to an increase in the estimation error. Thus it is important *not* to use the classical confidence intervals developed for i.i.d. settings if the serial dependence is not negligible. If, for example, the VaR of a financial investment is to be estimated, then an upper confidence bound obtained from the i.i.d. theory will often indicate a risk much lower than the actual one.

To be more concrete, let $F^{-1}(1-p_n)$ be the extreme quantile that is to be estimated. We are mainly interested in the case $np_n = O(1)$, although our main result also holds if $np_n \to \infty$ not too fast.

Only estimators based, say, on the $k_n + 1$ largest order statistics $\max_{1 \leqslant i \leqslant n} X_i = X_{n:n} \geqslant X_{n-1:n} \geqslant \ldots \geqslant X_{n-k_n:n}$ are considered. In order to keep the paper to manageable proportions, we will focus on heavy-tailed distributions, $\gamma > 0$, which is the most important case in financial applications. However, we will also indicate how to construct and analyse similar estimators in the general case $\gamma \in \mathbb{R}$.

To construct extreme quantile estimators, recall that the basic assumption $F \in D(G_\gamma)$ with $\gamma > 0$ is equivalent to

$$R(\lambda, t) := \frac{F^{-1}(1-\lambda t)}{F^{-1}(1-\lambda)} - t^{-\gamma} \to 0, \qquad t > 0, \tag{3}$$

as $\lambda \downarrow 0$. Reading this convergence as an approximation for small λ, one obtains

$$\begin{aligned} x_{p_n} := F^{-1}(1-p_n) &\approx F^{-1}\left(1 - \frac{k_n}{n}\right)\left(\frac{np_n}{k_n}\right)^{-\gamma} \\ &\approx X_{n-k_n:n}\left(\frac{np_n}{k_n}\right)^{-\hat{\gamma}_n} \\ &=: \hat{x}_{p_n}^{(k_n)} = \hat{x}_{p_n}, \end{aligned} \tag{4}$$

where $\hat{\gamma}_n$ denotes a suitable estimator of the extreme value index γ depending only on the $k_n + 1$ largest order statistics. To justify the first approximation k_n/n has to be small, while on the other hand k_n should be sufficiently large that the empirical quantile $X_{n-k_n:n}$ estimates the intermediate quantile $F^{-1}(1-k_n/n)$ well. Thus in what follows we assume that the natural numbers k_n form an intermediate sequence, that is,

$$k_n \to \infty, \qquad k_n/n \to 0. \tag{5}$$

The extreme value index γ may be estimated, for example, by the Hill estimator

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

The consistency of the Hill estimator was proved in quite general time series models by Hsing (1991) and Resnick and Stărică (1998). Resnick and Stărică (1997) examined its asymptotic normality in specific models, while Novak (2002) proved the asymptotic normality of a closely related estimator under suitable mixing conditions. Rootzén *et al.* (1992) established asymptotic normality for the quantile estimator $\hat{x}_{p_n}$ based on the Hill estimator under a rather complex set of conditions; among other things, they assumed that the von Mises condition is met, $np_n \to 0$, and the time series is strongly mixing.

In contrast, Drees (2000) established the asymptotic normality of a much broader class of estimators for the extreme value index, including the maximum likelihood estimator, studied by Smith (1987) in an i.i.d. setting, the moment estimator proposed by Dekkers *et al*. (1989) and the Pickands (1975) estimator. The main mathematical tool underlying these asymptotic results is a weighted approximation of the tail empirical quantile function (qf). This functional limit theorem will also enable us to derive the asymptotic normality of the quantile estimators $\hat{x}_{p_n}$ based on the general class of estimators $\hat{\gamma}_n$.

In the general case $\gamma \in \mathbb{R}$, a necessary and sufficient condition for $F \in D(G_\gamma)$ is

$$\frac{F^{-1}(1-\lambda t) - F^{-1}(1-\lambda)}{a(\lambda)} \to \frac{t^{-\gamma}-1}{\gamma}, \qquad t > 0, \tag{6}$$

as $\lambda \downarrow 0$ for some normalizing function $a : (0, 1) \to (0, \infty)$; for $\gamma = 0$ the right-hand side is interpreted as $-\log t$. Hence the following extreme quantile estimator can be motivated in a similar fashion to $\hat{x}_{p_n}$ above:

$$\tilde{x}_{p_n} := X_{n-k_n:n} + \hat{a}(k_n/n)\frac{(np_n/k_n)^{-\hat{\gamma}_n} - 1}{\hat{\gamma}_n}. \tag{7}$$

Here $\hat{a}(k_n/n)$ denotes a suitable estimator for $a(k_n/n)$, for example

$$\hat{a}(k_n/n) := \frac{\hat{\gamma}_n}{2^{\hat{\gamma}_n} - 1}\left(X_{n-[k_n/2]:n} - X_{n-k_n:n}\right), \tag{8}$$

which is obtained by choosing $\lambda = k_n/n$ and $t = \frac{1}{2}$ in (6) and replacing the unknown quantiles by their respective empirical counterparts. In an i.i.d. setting, the limit distribution of particular estimators of this type was established by Dekkers *et al*. (1989) and de Haan and Rootzén (1993), among others. However, no general approach to constructing estimators of a and hence of extreme quantiles, comparable to the broad class of statistical tail functionals for γ introduced in Drees (1998a), has hitherto been proposed.

It should be emphasized that within a parametric model for the dependence structure, one may often construct more efficient estimators for extreme quantiles; see Section 3 for an example. However, these estimators will be very sensitive to deviations from the parametric model, while the estimators under consideration in the present paper yield reasonable results under mild structural assumptions.

The paper is organized as follows. In Section 2, first the approximation result for the tail empirical qf of absolutely regular time series established in Drees (2000) is specialized to the case $\gamma > 0$. Here we impose conditions which are more restrictive but often more easily checked. From this we derive the asymptotic normality of quantile estimators of type (4). Estimators of the asymptotic variance and resulting confidence intervals are also discussed. As examples of time series models satisfying our conditions, a particular class of nonlinear time series, including ARCH(1) models, and linear time series are considered in Section 3. Then the theory is applied to a time series of log-returns of the Nasdaq Composite index. It turns out that the classical i.i.d. theory leads to confidence intervals that are much shorter than the new confidence intervals that take into account the serial dependence.

In Section 5 the finite-sample performance of the statistical procedures is examined in a simulation study for several time series models with heavy tails. Again the confidence intervals proposed in the present paper usually have coverage probabilities that are much closer to the nominal level than those of classical confidence intervals.

Finally, we establish the asymptotic normality of a broad class of estimators for γ and a for general $\gamma \in \mathbb{R}$ and conclude the asymptotic normality of the resulting quantile estimators of type (7).

2. Asymptotics for $\gamma > 0$

In the following we assume that the sequence X_i, $i \in \mathbb{N}$, is strictly stationary, that is, $\mathcal{L}((X_i)_{i\in\mathbb{N}}) = \mathcal{L}((X_{i+n})_{i\in\mathbb{N}})$ for all $n \in \mathbb{N}$. Since the quantile estimator of type (4) depends only on the $k_n + 1$ largest order statistics, it is essential to analyse the asymptotic behaviour of the pertaining tail empirical qf

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

Drees (2000) gave a weighted approximation of this stochastic process for stationary β-mixing time series with a continuous marginal df $F \in D(G_\gamma)$, $\gamma \in \mathbb{R}$. (In fact, the continuity assumption may be dropped; see Remark 2 of that paper.) Recall that X_i, $i \in \mathbb{N}$, is called β-mixing (or absolutely regular) if

$$\beta(l) := \sup_{m\in\mathbb{N}} \mathrm{E}\left(\sup_{A\in\mathcal{B}_{m+l+1}^{\infty}} |P(A|\mathcal{B}_1^m) - P(A)| \right) \to 0$$

as $l \to \infty$, where $\mathcal{B}_1^m$ and $\mathcal{B}_{m+l+1}^{\infty}$ denote the σ-fields generated by $(X_i)_{1\leqslant i\leqslant m}$ and $(X_i)_{m+l+1\leqslant i}$, respectively. More precisely, it is assumed that there exists a sequence l_n, $n \in \mathbb{N}$, such that

(C1) $\displaystyle \lim_{n\to\infty} \frac{\beta(l_n)}{l_n} n + l_n k_n^{-1/2} \log^2 k_n = 0.$

Since the β-coefficients measure the influence of the past on future events, condition (C1) states that this influence vanishes sufficiently fast as past and future are separated by a time interval of increasing length. Typical examples are Harris recurrent Markov chains, for which the β-coefficients decrease geometrically; see Doukhan (1994, Section 2.4), for details. More specific, autoregressive moving average (ARMA), ARCH and generalized ARCH (GARCH) time series are geometrically β-mixing under rather mild conditions (Doukhan, 1994, Section 2.3). In these cases, condition (C1) is satisfied with $l_n = [C \log n]$ for a sufficiently large constant $C > 0$ and k_n satisfying

$$\log^2 n \log^4(\log n) = o(k_n). \tag{9}$$

Furthermore, we assume a regularity condition for the joint tail of (X_1, X_{1+m}):

(C2) There exist $\varepsilon > 0$ and functions c_m, $m \in \mathbb{N}$, such that

$$\lim_{n\to\infty} \frac{n}{k_n} P\left\{X_1 > F^{-1}\left(1-\frac{k_n}{n}x\right),\, X_{1+m} > F^{-1}\left(1-\frac{k_n}{n}y\right)\right\} \to c_m(x, y)$$
$$\forall m \in \mathbb{N},\, 0 < x,\, y \leqslant 1+\varepsilon.$$

In addition, we need a uniform bound on the probability that both X_1 and X_{1+m} belong to an extreme interval:

(C3) There exist $D_1 \geqslant 0$ and a sequence $\tilde{\rho}(m)$, $m \in \mathbb{N}$, satisfying $\sum_{m=1}^{\infty}\tilde{\rho}(m) < \infty$ such that

$$\frac{n}{k_n} P\{X_1 \in I_n(x, y),\, X_{1+m} \in I_n(x, y)\} \leqslant (y-x)\left(\tilde{\rho}(m) + D_1 \frac{k_n}{n}\right)$$
$$\forall m \in \mathbb{N},\, 0 < x,\, y \leqslant 1+\varepsilon$$

with $I_n(x, y) = (F^{-1}(1 - yk_n/n),\, F^{-1}(1 - xk_n/n)]$.

Remark 2.1. Condition (C2) is satisfied if all vectors (X_1, X_{1+m}) belong to the domain of attraction of a bivariate extreme value distribution, that is, if the suitably standardized coordinatewise maxima of n i.i.d. copies of (X_1, X_{1+m}) converge to a non-trivial limiting distribution as n tends to ∞. If the marginals of the limiting vector are independent, then $c_m(x, y) = 0$ for all $m \in \mathbb{N}$ and all $0 < x,\, y \leqslant 1+\varepsilon$.

Remark 2.2. It is readily seen that condition (C3) is met if the ρ-mixing coefficients of the time series are finitely summable, that is, $\sum_{l=1}^{\infty}\rho(l) < \infty$ with

$$\rho(l) := \sup_{m\in\mathbb{N}} \sup_{U\in L_2(\mathcal{B}_1^m), V\in L_2(\mathcal{B}_{m+l+1}^{\infty})} \frac{|\mathrm{cov}(U, V)|}{(\mathrm{var}(U)\mathrm{var}(V))^{1/2}} \tag{10}$$

and $L_2(\mathcal{A})$ denoting the space of square-integrable $(\mathcal{A}, \mathbb{B})$-measurable functions.

Conditions (C2) and (C3) ensure that the suitably standardized covariance of the numbers of exceedances over different high quantiles of F converges to a limit covariance function as the sample size increases. Moreover, they imply a bound on the second moment of the number of observations in an extreme interval.

Proposition 2.1. *Suppose that* $l_n = o(n/k_n)$ *and that conditions* (C2) *and* (C3) *are satisfied. Then, for all* $0 < x,\, y \leqslant 1+\varepsilon$,

$$\lim_{n\to\infty} \frac{n}{l_n k_n} \mathrm{cov}\left(\sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_n/n)x)\}},\, \sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_n/n)y)\}}\right) = c(x, y) \tag{11}$$

with

$$c(x, y) := x \wedge y + \sum_{m=1}^{\infty}(c_m(x, y) + c_m(y, x)) \in \mathbb{R} \tag{12}$$

and $x \wedge y := \min(x, y)$.

Moreover, there exists $D > 0$ *such that, for all* $0 < x, y \leqslant 1 + \varepsilon$ *and all* $n \in \mathbb{N}$,

$$\frac{n}{l_n k_n} \mathrm{E}\left(\sum_{i=1}^{l_n} 1_{\{X_i \in I_n(x,y)\}} \right)^2 \leqslant D(y - x) \tag{13}$$

Proof. In (C3) choose $y = 1 + \varepsilon$ and let x tend to 0 to obtain

$$\frac{n}{k_n} P\left\{ X_1 > F^{-1}\left(1 - \frac{k_n}{n}(1+\varepsilon)\right), X_{1+m} > F^{-1}\left(1 - \frac{k_n}{n}(1+\varepsilon)\right) \right\}$$
$$\leqslant (1+\varepsilon)\left(\tilde{\rho}(m) + D_1 \frac{k_n}{n} \right).$$

Because of (C2), $l_n k_n / n \to 0$ and $\lim_{n\to\infty} \sum_{m=1}^{l_n} (\tilde{\rho}(m) + (D_1 + (1+\varepsilon)^2) k_n/n) = \sum_{m=1}^{\infty} \tilde{\rho}(m) < \infty$, Pratt's (1960) lemma yields

$$\lim_{n\to\infty} \frac{n}{k_n} \sum_{m=1}^{l_n} \operatorname{cov}\left(1_{\{X_1 > F^{-1}(1-(k_n/n)x)\}}, 1_{\{X_{1+m} > F^{-1}(1-(k_n/n)y)\}}\right) = \sum_{m=1}^{\infty} c_m(x, y) \in \mathbb{R}.$$

Hence, by the stationarity of the time series,

$$\frac{n}{k_n l_n} \sum_{1 \leqslant i < j \leqslant l_n} \operatorname{cov}\left(1_{\{X_i > F^{-1}(1-(k_n/n)x)\}}, 1_{\{X_j > F^{-1}(1-(k_n/n)y)\}}\right)$$
$$= \frac{1}{l_n} \sum_{i=1}^{l_n} \frac{n}{k_n} \sum_{j=i+1}^{i+l_n-1} \operatorname{cov}\left(1_{\{X_i > F^{-1}(1-(k_n/n)x)\}}, 1_{\{X_j > F^{-1}(1-(k_n/n)y)\}}\right)$$
$$- \frac{1}{l_n} \sum_{i=2}^{l_n} \frac{n}{k_n} \sum_{j=l_n+1}^{i+l_n-1} \operatorname{cov}\left(1_{\{X_i > F^{-1}(1-(k_n/n)x)\}}, 1_{\{X_j > F^{-1}(1-(k_n/n)y)\}}\right)$$
$$\to \sum_{m=1}^{\infty} c_m(x, y)$$

since the second term can be bounded by $\sum_{m=1}^{l_n-1} m(\tilde{\rho}(m) + (D_1 + (1+\varepsilon)^2) k_n/n)/l_n$ which tends to 0. Now (11) is obvious.

Likewise, one obtains

$$\frac{n}{l_n k_n} \mathrm{E}\left(\sum_{i=1}^{l_n} 1_{\{X_i \in I_n(x,y)\}} \right)^2 \leqslant (y - x)\left(1 + 2 \sum_{m=1}^{l_n - 1} \left(\tilde{\rho}(m) + D_1 \frac{k_n}{n} \right) \right)$$

so that (13) follows from the summability of $\tilde{\rho}(m)$ and $l_n k_n / n \to 0$. □

Remark 2.3. Using Theorem 1.1 of Shao (1995), in (13) one may even replace the second moment with the fourth moment. These moment conditions can be interpreted in terms of moments of cluster sizes of exceedances; see Drees (2000) for details.

Finally, we need a condition on the rate of convergence of $k_n \to \infty$ to ensure that the extreme value approximation used in (4) is sufficiently accurate. For the sake of simplicity, we assume that the qf admits the following representation:

(C4) $F^{-1}(1-t) = dt^{-\gamma}(1+r(t))$, with $|r(t)| \leqslant \Phi(t)$, for some constant $d > 0$ and a function Φ which is τ-varying at 0 for some $\tau > 0$, or $\tau = 0$ and Φ is non-decreasing with $\lim_{t\downarrow 0}\Phi(t) = 0$.

Then we assume that k_n is an intermediate sequence such that

(C5) $\lim_{n\to\infty} k_n^{1/2}\Phi(k_n/n) = 0.$

(However, see Remark 2.6 below for more general conditions.)

Theorem 2.1. *Under conditions* (C1)–(C5) *with* $l_n = o(n/k_n)$ *there exist versions of the tail empirical qf* Q_n *and a centred Gaussian process* e *with covariance function* c *defined by* (12) *such that*

$$\sup_{t\in(0,1]} t^{\gamma+1/2}(1+|\log t|)^{-1/2}\left| k_n^{1/2}\left(\frac{Q_n(t)}{F^{-1}(1-k_n/n)} - t^{-\gamma}\right) - \gamma t^{-(\gamma+1)}e(t)\right| \to 0 \quad (14)$$

in probability.

Proof. In view of (C4), the remainder term defined in (3) equals $R(\lambda, t) = t^{-\gamma}\, O(\Phi(\lambda t) + \Phi(\lambda))$ uniformly for bounded t. Thus, because of either the τ-variation of Φ with $\tau > 0$ or the monotonicity of Φ, $k_n^{1/2}\Phi(k_n/n) \to 0$ implies

$$\lim_{n\to\infty} k_n^{1/2} \sup_{0<t\leqslant 1+\varepsilon} t^{\gamma+1/2}(1+|\log t|)^{-1/2}|R(k_n/n, t)| = 0.$$

Combining this with Proposition 2.1 and Remark 2.3, we see that the conditions of Theorem 3.1 of Drees (2000) are satisfied, from which the assertion is obvious. □

Remark 2.4. Conditions (C2) and (C3) are only needed to verify (11) and the analogue of (13) for the fourth moment. Hence Theorem 2.1 holds under these considerably weaker (but somewhat more complex) conditions. If (C2) and (C3) are replaced with (11) and (13), then (14) holds with weight function $t^{\gamma+1/2}$ replaced with $t^{\gamma+3/4}$. See Drees (2002) for a more detailed discussion of these conditions.

It is worth mentioning that for independent observations the Gaussian process e is a standard Brownian motion. Hence, in that case, Theorem 2.1 is essentially equivalent to Theorem 2.1(i) of Drees (1998b).

In the next step, we deduce the asymptotic normality of estimators of the extreme value index that use only the $k_n + 1$ largest order statistics, and of the pertaining quantile estimators of type (4). Drees (1998a; 1998b) observed that almost every estimator $\hat{\gamma}_n$ of this type can be represented as a so-called statistical tail functional, that is, as a smooth functional applied to the tail empirical qf: $\hat{\gamma}_n = T(Q_n)$.

To establish asymptotic normality for this class of estimators we impose the following regularity conditions on T:

(T0) T is a Borel-measurable real-valued functional on the set of functions $z \in D(0, 1]$ satisfying $t^{\gamma+1/2}|\log t|^{-1/2} z(t) \to 0$ as $t \downarrow 0$.

(T1) T is scale-invariant: $T(az) = T(z)$ for all $a > 0$.

(T2) $T((t^{-\gamma})_{0<t\leqslant 1}) = \gamma$.

(T3) There exists a signed measure $\nu_{T,\gamma}$ on $(0, 1]$ with $\int_{(0,1]} t^{-\gamma-1/2}(1 + |\log t|)^{1/2} |\nu_{T,\gamma}|(\mathrm{d}t) < \infty$ such that

$$\varepsilon_n^{-1}\big(T((t^{-\gamma} + \varepsilon_n z_n(t))_{0<t\leqslant 1}) - T((t^{-\gamma})_{0<t\leqslant 1})\big) \to \int_{(0,1]} z(t)\nu_{T,\gamma}(\mathrm{d}t)$$

for all $\varepsilon_n \downarrow 0$ and z_n satisfying

$$\sup_{0<t\leqslant 1} t^{\gamma+1/2}(1 + |\log t|)^{-1/2}|z_n(t) - z(t)| \to 0$$

for some continuous function z as described in (T0).

Condition (T2) means that precisely the true extreme value index is obtained if one plugs in the limiting Pareto qf instead of the tail empirical qf. Condition (T3) can be interpreted as T being Hadamard differentiable at $(t^{-\gamma})_{0<t\leqslant 1}$ in a suitable function space. Refer to Drees (1998a; 1998b) for a thorough discussion of these regularity conditions. In particular, there it is shown that the Hill estimator and the maximum likelihood estimator in a generalized Pareto model satisfy these conditions with signed measures

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

and

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

where ε_1 denotes the Dirac measure with mass 1 at 1. Other examples are the Pickands estimator (Pickands 1975), the moment estimator proposed by Dekkers *et al*. (1989) and generalized probability-weighted moment estimators.

In addition to (C5) we need the following assumption about the relationship between the number of order statistics used for estimation and the expected number of exceedances over the extreme quantile to be estimated:

$$\lim_{n\to\infty} k_n^{-1/2}\log(np_n) = 0, \qquad \lim_{n\to\infty} np_n/k_n = 0. \tag{15}$$

The first assumption is very weak; it is satisfied if, for example, $n^{-m} = o(p_n)$ for some $m > 0$ and $\log^2 n = o(k_n)$. Notice that we allow np_n to tend to infinity, but the whole extreme value approach only makes sense if $np_n = o(k_n)$.

Theorem 2.2. *Suppose that the conditions of Theorem 2.1 and condition* (15) *are satisfied. If* $\hat{\gamma}_n = T(Q_n)$, *with* T *fulfilling conditions* (T0)–(T3), *then*

$$\frac{k_n^{1/2}}{\log(k_n/(np_n))}\log\frac{\hat{x}_{p_n}}{x_{p_n}} \sim \frac{k_n^{1/2}}{\log(k_n/(np_n))}\left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right) \sim k_n^{1/2}(\hat{\gamma}_n - \gamma) \to \mathcal{N}(0, \sigma^2_{\mathrm{T},\gamma}) \tag{16}$$

weakly with

$$\sigma^2_{T,\gamma} = \gamma^2 \int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} c(s,\, t)\nu_{T,\gamma}(\mathrm{d}s)\nu_{T,\gamma}(\mathrm{d}t).$$

Proof. The weak convergence of $k_n^{1/2}(\hat{\gamma}_n - \gamma)$ follows from the following calculation:

$$\begin{aligned}
\hat{\gamma}_n &\overset{(\mathrm{T1})}{=} T\left(\frac{Q_n}{F^{-1}(1-k_n/n)}\right) \\
&=^d T\left(\left(t^{-\gamma} + k_n^{-1/2}\gamma t^{-(\gamma+1)}e(t) + o_P(k_n^{-1/2})\right)_{0<t\leqslant 1}\right) \\
&\overset{(\mathrm{T3})}{=} T((t^{-\gamma})_{0<t\leqslant 1}) + k_n^{-1/2}\gamma\int_{(0,1]} t^{-(\gamma+1)}e(t)\nu_{T,\gamma}(\mathrm{d}t) + o_P(k_n^{-1/2}).
\end{aligned}$$

Hence, by (T2),

$$k_n^{1/2}(\hat{\gamma}_n - \gamma) \to \gamma\int_{(0,1]} t^{-(\gamma+1)}e(t)\nu_{T,\gamma}(\mathrm{d}t),$$

which proves the assertion; we refer to Drees (1998a) for technical details.

Because $\log(1+x) \sim x$ as $x \to 0$, it remains to verify that

$$\frac{1}{\log(k_n/(np_n))}\left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right) = \hat{\gamma}_n - \gamma + o_P(k_n^{-1/2}). \tag{17}$$

To this end, check that

$$\hat{x}_{p_n} - x_{p_n} = \left(Q_n(1) - F^{-1}\left(1 - \frac{k_n}{n}\right)\right)\left(\frac{np_n}{k_n}\right)^{-\hat{\gamma}_n} + F^{-1}\left(1 - \frac{k_n}{n}\right)\left(\left(\frac{np_n}{k_n}\right)^{-\hat{\gamma}_n} - \left(\frac{np_n}{k_n}\right)^{-\gamma}\right)$$
$$+ \left(F^{-1}\left(1 - \frac{k_n}{n}\right)\left(\frac{np_n}{k_n}\right)^{-\gamma} - F^{-1}(1 - p_n)\right)$$
$$=: I + II + III.$$

As in the proof of Theorem 2.1, (C4) and (C5) imply

$$\left|\frac{F^{-1}(1 - p_n)}{F^{-1}(1 - k_n/n)}\left(\frac{np_n}{k_n}\right)^{\gamma} - 1\right| = \left|R\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right)\right|\left(\frac{np_n}{k_n}\right)^{\gamma} = O\left(\Phi\left(\frac{k_n}{n}\right)\right) = o(k_n^{-1/2}).$$

Hence Theorem 2.1, condition (15) and $\hat{\gamma}_n - \gamma = O_P(k_n^{-1/2})$ yield

$$\frac{I}{x_{p_n}\log(k_n/(np_n))} = k_n^{-1/2}(e(1) + o_P(1))\frac{F^{-1}(1 - k_n/n)}{F^{-1}(1 - p_n)}\frac{1}{\log(k_n/(np_n))}\left(\frac{np_n}{k_n}\right)^{-\hat{\gamma}_n}$$
$$= o_P(k_n^{-1/2}).$$

Likewise

$$\frac{III}{x_{p_n}\log(k_n/(np_n))} = \left(\frac{F^{-1}(1 - k_n/n)}{F^{-1}(1 - p_n)}\left(\frac{np_n}{k_n}\right)^{-\gamma} - 1\right)\Big/\log\frac{k_n}{np_n} = o(k_n^{-1/2}).$$

Finally, because $\partial/(\partial\tau)x^{\tau} = x^{\tau}\log x$, using the mean value theorem and (15) one obtains

$$\frac{II}{x_{p_n}\log(k_n/(np_n))} = \frac{F^{-1}(1 - k_n/n)}{F^{-1}(1 - p_n)}\left(\frac{np_n}{k_n}\right)^{-\gamma}\left(\frac{np_n}{k_n}\right)^{\vartheta(\gamma - \hat{\gamma}_n)}(\hat{\gamma}_n - \gamma) = (\hat{\gamma}_n - \gamma)(1 + o_P(1))$$

for some $\vartheta \in (0, 1)$. Adding the expressions for *I*, *II* and *III*, one arrives at (17). □

Remark 2.5. If $np_n \to \infty$ one may also estimate x_{p_n} consistently by the empirical quantile $X_{n-[np_n]:n}$, that is, $X_{n-[np_n]:n}/x_{p_n} \to 1$ in probability. However, typically the relative estimation error will be of order $(np_n)^{-1/2}$. In particular this holds if conditions (C1)–(C5) are fulfilled with k_n replaced by $[np_n] + 1$. So the quantile estimator $\hat{x}_{p_n}$ is asymptotically more efficient provided $np_n = o(k_n)$. Of course, this higher efficiency is achieved only under considerably stronger model assumptions than necessary to ensure consistency of the empirical quantile.

Remark 2.6. Time series models are often described implicitly as stationary solutions of certain equations involving innovations with a given distribution (see Section 3 for examples). Then no analytical expression is usually available for the distribution function F of the time series at any time t. In this situation it might be difficult to verify condition (C4), but for $F \in D(G_\gamma)$ the following milder condition replacing (C4) and (C5) is *always* satisfied for some intermediate sequence k_n tending to infinity not too fast:

$$\lim_{n\to\infty} k_n^{1/2} \sup_{0<t\leqslant 1+\varepsilon} t^{\gamma+1/2}(1+|\log t|)^{-1/2}|R(k_n/n,\ t)| = 0 \tag{18}$$

with R defined by (3); cf. Drees (1998a; 2000). Under the conditions (C1)–(C3) and (18) the assertion of Theorem 2.1 holds. To prove asymptotic normality of the quantile estimators, in addition to (15) we need

$$\lim_{n\to\infty} \frac{k_n^{1/2}}{\log(k_n/(np_n))} \left(\frac{np_n}{k_n}\right)^{\gamma} R\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right) = 0. \tag{19}$$

This convergence implies

$$\frac{F^{-1}(1-p_n)}{F^{-1}(1-k_n/n)} \left(\frac{np_n}{k_n}\right)^{\gamma} - 1 = \left(\frac{np_n}{k_n}\right)^{\gamma} R\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right) = o\left(k_n^{-1/2} \log \frac{k_n}{np_n}\right).$$

Hence the proof of Theorem 2.2 shows that convergence (16) holds under these milder conditions.

In simulations we found that in most cases the normal approximation is more accurate for $\log(\hat{x}_{p_n}/x_{p_n})$ than for $\hat{x}_{p_n}/x_{p_n} - 1$. Heuristically, this may be explained by the fact that $\log \hat{x}_{p_n}$ is a linear function of $\hat{\gamma}_n$, whose estimation error determines the dominating part of the error of the quantile estimator. So if the distribution of $\hat{\gamma}_n$ is well approximated by a normal distribution (which is usually true for the Hill estimator), this often also holds for $\log \hat{x}_{p_n}$ but not necessarily for $\hat{x}_{p_n}$ which, according to the δ-method, is only locally linear in $\hat{\gamma}_n$.

In order to construct confidence intervals based on (16), one has to estimate the asymptotic variance $\sigma^2_{T,\gamma}$, which depends not only on γ but also on the unknown limiting covariance function c. Instead of trying to estimate this function nonparametrically, it seems more reasonable to employ (16) for the estimation of $\sigma^2_{T,\gamma}$.

In a blocks approach, one would split the time series into blocks of constant length m_n, say, and estimate an extreme quantile or γ for each block separately. If m_n is not too small, by condition (C1) these estimates are almost independent. Hence one may estimate $\sigma^2_{T,\gamma}$ by the suitably standardized sample variance of the block estimates. In practice, however, this procedure is rather cumbersome, because one must find not only a suitable block length m_n, but also a number $\tilde{k}_n < m_n$ such that in every block it is reasonable to use the $\tilde{k}_n + 1$ largest order statistics for estimation. Given that it is often not easy to choose k_n for one fixed sample, this may be a delicate task.

As an alternative, we propose an approach which uses a process version of convergence (16). For this, note that under very weak conditions the covariance function c is homogeneous:

$$c(\lambda x, \lambda y) = \lambda c(x,\ y), \qquad \lambda,\ x,\ y \in [0,\ 1], \tag{20}$$

and hence that the Gaussian process e is self-similar, that is,

$$e(\lambda\cdot) =^d \lambda^{1/2} e(\cdot), \qquad \lambda \in [0,\ 1]. \tag{21}$$

For example, if condition (11) holds for k_n and two sequences $k_{n,\lambda_j} \sim \lambda_j k_n$ with $\lambda_j \in (0, 1)$, $j = 1, 2$, then, by the continuity of c,

$$c(x, y) \leftarrow \frac{n}{l_n k_{n,\lambda_j}} \operatorname{cov}\left(\sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_{n,\lambda_j}/n)x)\}}, \sum_{i=1}^{l_n} 1_{\{X_i > \mathrm{F}^{-1}(1-(k_{n,\lambda_j}/n)y)\}} \right)$$

$$\sim \frac{n}{l_n k_n \lambda_j} \operatorname{cov}\left(\sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_n/n)(k_{n,\lambda_j}/k_n)x)\}}, \sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-(k_n/n)(k_{n,\lambda_j}/k_n)y)\}} \right)$$

$$\to \frac{1}{\lambda_j} c(\lambda_j x, \lambda_j y).$$

If $\log \lambda_1 / \log \lambda_2$ is irrational then, by Theorem 1.4.3 of Bingham *et al*. (1987), this in turn implies (20). Likewise, if convergence (14) holds for k_n and k_{n,λ_j} then from the regular variation of $F^{-1}(1 - \cdot)$ it follows that $e(\cdot) =^d \lambda_j^{-1/2} e(\lambda_j \cdot)$ and thus (21) and (20).

Now one may argue heuristically as follows. Denote by $\hat{\gamma}_n^{(i)}$ the estimator for γ that uses the $i + 1$ largest order statistics: $\hat{\gamma}_n^{(i)} = T(Q_n(i/k_n \cdot))$. Under a slightly stronger differentiability condition than (T3), one obtains as in Theorem 2.2 the approximation

$$k_n^{1/2}(\hat{\gamma}_n^{([k_n s])} - \gamma) \approx \frac{\gamma}{s} \int_{(0,1]} t^{-(\gamma+1)} e(st) \nu_{T,\gamma}(\mathrm{d}t) =: Z_{T,\gamma}(s) \tag{22}$$

for $1/k_n \leqslant s \leqslant 1$. Notice that, by the homogeneity property (20) of c,

$$\tilde{Z}_{T,\gamma}(u) := \mathrm{e}^{u/2} Z_{T,\gamma}(\mathrm{e}^u), \qquad u \in (-\infty, 0], \tag{23}$$

defines a strictly stationary centred Gaussian process with covariance function

$$\operatorname{cov}(\tilde{Z}_{T,\gamma}(u), \tilde{Z}_{T,\gamma}(v)) = \gamma^2 \exp\left(-\frac{u+v}{2}\right) \int_{(0,1]} \int_{(0,1]} (st)^{-(\gamma+1)} c(\mathrm{e}^u s, \mathrm{e}^v t) \nu_{T,\gamma}(\mathrm{d}s) \nu_{T,\gamma}(\mathrm{d}t)$$

$$= \gamma^2 \exp\left(\frac{u-v}{2}\right) \int_{(0,1]} \int_{(0,1]} (st)^{-(\gamma+1)} c(s, \mathrm{e}^{v-u} t) \nu_{T,\gamma}(\mathrm{d}s) \nu_{\mathrm{T},\gamma}(\mathrm{d}t)$$

depending only on $u - v$. Using the ergodic theorem, one can show that

$$\left(\log \frac{k_n}{j_n}\right)^{-1} \sum_{i=j_n}^{k_n} (\hat{\gamma}_n^{(i)} - \hat{\gamma}_n^{(k_n)})^2 \approx \left(\log \frac{k_n}{j_n}\right)^{-1} \int_{j_n/k_n}^{1} (Z_{T,\gamma}(s) - Z_{T,\gamma}(1))^2 \,\mathrm{d}s \tag{24}$$

$$\sim \left(\log \frac{k_n}{j_n}\right)^{-1} \int_{\log(j_n/k_n)}^{0} \tilde{Z}_{T,\gamma}^2(u) \,\mathrm{d}u$$

$$\to \mathrm{E}(\tilde{Z}_{T,\gamma}^2(0)) = \sigma_{T,\gamma}^2,$$

provided $k_n / j_n \to \infty$ (refer to the proof of Theorem 2.3 for details). Unfortunately, from Theorem 2.1 it can only be shown that approximation (24) is sufficiently accurate for some sequence $j_n = o(k_n)$; for a more precise assertion about j_n one would need the rate of convergence in (14).

Theorem 2.3. *Suppose that* (20) *(or, equivalently,* (21)*) holds and that* T *is Fréchet differentiable at* $(t^{-\gamma})_{0<t\leqslant 1}$*:*

$$\varepsilon^{-1}(T((t^{-\gamma}+\varepsilon z(t))_{0<t\leqslant 1}) - T((t^{-\gamma})_{0<t\leqslant 1})) \to \int_{(0,1]} z(t)\nu_{T,\gamma}(\mathrm{d}t), \tag{25}$$

as $\varepsilon \downarrow 0$*, uniformly for all* z *satisfying*

$$\sup_{0<t\leqslant 1} t^{\gamma+1/2}(1+|\log t|)^{-1/2}|z(t)| \leqslant 1.$$

Then, under the conditions of Theorem 2.2, there exists a sequence $j_n = o(k_n)$ *such that*

$$\hat\sigma^2_{T,\gamma,1} := \left(\log\frac{k_n}{j_n}\right)^{-1}\sum_{i=j_n}^{k_n}(\hat\gamma_n^{(i)} - \hat\gamma_n^{(k_n)})^2 \to \sigma^2_{T,\gamma} \tag{26}$$

$$\hat\sigma^2_{T,\gamma,2} := \left(\log\frac{k_n}{j_n}\right)^{-1}\sum_{i=j_n}^{k_n}\left(\frac{\log/(\hat x_{p_n}^{(i)}/\hat x_{p_n}^{(k_n)})}{\log(i/(np_n))}\right)^2 \to \sigma^2_{T,\gamma} \tag{27}$$

in probability. Here $\hat\gamma_n^{(i)} = T(Q_n(i/k_n\cdot))$ *and* $\hat x_{p_n}^{(i)}$ *is defined as in* (4) *with* $\hat\gamma_n$ *replaced by* $\hat\gamma_n^{(i)}$ *and* k_n *by* i*.*

Proof. By (T1), (T2) and the differentiability assumption (25)

$$\begin{aligned}
\hat\gamma_n^{(i)} &= T\left(Q_n\left(\frac{i}{k_n}\cdot\right)/F^{-1}(1-k_n/n)\right)\\
&=^d T\Bigg(\Bigg(\left(\frac{i}{k_n}t\right)^{-\gamma} + \gamma k_n^{-1/2}\left(\frac{i}{k_n}t\right)^{-(\gamma+1)} e\left(\frac{i}{k_n}t\right)\\
&\qquad + o_P\left(k_n^{-1/2}\left(\frac{i}{k_n}t\right)^{-(\gamma+1/2)}\left(1+\left|\log\left(\frac{i}{k_n}t\right)\right|\right)^{1/2}\right)\Bigg)_{0<t\leqslant 1}\Bigg)\\
&= T\left(\left(t^{-\gamma} + \gamma\frac{k_n^{1/2}}{i}t^{-(\gamma+1)}e\left(\frac{i}{k_n}t\right) + o_P\left(k_n^{-1/2}t^{-(\gamma+1/2)}(1+|\log t|)^{1/2}\right)\right)_{0<t\leqslant 1}\right)\\
&= \gamma + \gamma\frac{k_n^{1/2}}{i}\int_{(0,1]} t^{-(\gamma+1)}e\left(\frac{i}{k_n}t\right)\nu_{T,\gamma}(\mathrm{d}t) + o_P(k_n^{-1/2})
\end{aligned}$$

uniformly for $s_0 k_n \leqslant i \leqslant k_n$ and all $s_0 > 0$. Hence, by the continuity of e and a standard diagonal argument, there exists a sequence $s_n \downarrow 0$ such that

$$\sup_{s_n \leqslant s \leqslant 1} \left| k_n^{1/2}(\hat{\gamma}_n^{([k_n s])} - \gamma) - Z_{T,\gamma}(s) \right| \to 0$$

in probability, with $Z_{T,\gamma}$ defined in (22).

Therefore, for $j_n := [s_n k_n] + 1$ and $\tilde{Z}_{T,\gamma}$ defined by (23),

$$\sum_{i=j_n}^{k_n} (\hat{\gamma}_n^{(i)} - \hat{\gamma}_n^{(k_n)})^2 = \int_{j_n/k_n}^{1} (Z_{T,\gamma}(s) - Z_{T,\gamma}(1))^2 \, \mathrm{d}s \cdot (1 + o_P(1)) \tag{28}$$

$$= \int_{\log(j_n/k_n)}^{0} \left(\tilde{Z}_{T,\gamma}(u) - \mathrm{e}^{u/2} \tilde{Z}_{T,\gamma}(0) \right)^2 \mathrm{d}u \cdot (1 + o_P(1)).$$

By the stationarity of $\tilde{Z}_{T,\gamma}$ and the ergodic theorem (see Cramér and Leadbetter 1967, p. 151), one has

$$\left(\log \frac{k_n}{j_n} \right)^{-1} \int_{\log(j_n/k_n)}^{0} \tilde{Z}_{T,\gamma}^2(\mathrm{u}) \, \mathrm{d}u \to \mathrm{E} \tilde{Z}_{T,\gamma}^2(0) = \sigma_{T,\gamma}^2 \quad \text{almost surely.}$$

Hence assertion (26) follows from $\int_{\log(j_n/k_n)}^{0} (\mathrm{e}^{u/2} \tilde{Z}_{T,\gamma}(0))^2 \, \mathrm{d}u = O(1)$.

Similarly, one can show that, for some $s_n \downarrow 0$,

$$\sup_{s_n \leqslant s \leqslant 1} \left| \frac{k_n^{1/2}}{\log([k_n s]/(np_n))} \log \frac{\hat{x}_{p_n}^{([k_n s])}}{x_{p_n}} - Z_{T,\gamma}(s) \right| \to 0$$

in probability. Without loss of generality one may assume that $\log(s_n) = o(\log(k_n/(np_n)))$. Hence

$$\sum_{i=j_n}^{k_n} \left(\frac{\log(\hat{x}_{p_n}^{(i)} / \hat{x}_{p_n}^{(k_n)})}{\log(i/(np_n))} \right)^2$$

$$= \int_{j_n/k_n}^{1} \left(Z_{T,\gamma}(s) - \frac{\log(k_n/(np_n))}{\log([k_n s]/(np_n))} Z_{T,\gamma}(1) \right)^2 \mathrm{d}s \cdot (1 + o_P(1)) \tag{29}$$

$$= \int_{\log(j_n/k_n)}^{0} \left(\tilde{Z}_{T,\gamma}(u) - \frac{\log(k_n/(np_n))}{\log([k_n \mathrm{e}^u]/(np_n))} \mathrm{e}^{u/2} \tilde{Z}_{T,\gamma}(0) \right)^2 \mathrm{d}u \cdot (1 + o_P(1)).$$

Now assertion (27) follows by the above arguments and

$$\frac{\log(k_n/(np_n))}{\log([k_n \mathrm{e}^u]/(np_n))} \leqslant \frac{1}{1 + \log s_n / \log(k_n/(np_n))} \to 1$$

for all $\log(j_n/k_n) \leqslant u \leqslant 0$. □

The proof shows that the left-hand sides of (26) and (27) are consistent estimators of $\sigma_{T,\gamma}^2$ for all sequences $(j_n)_{n \in \mathbb{N}}$ such that j_n/k_n converges to 0 not too fast. In practice usually one may choose j_n to be rather small. Indeed, even the smallest number for which the estimator is defined will often do the job; cf. Sections 4 and 5.

In the proof it was also shown that in (28) and in (29) the terms pertaining to $Z_{T,\gamma}(1)$ are asymptotically negligible. This suggests the approximation

$$\mathrm{E}\sum_{i=j_n}^{k_n}\left(\frac{\log(\hat{x}_{p_n}^{(i)}/\hat{x}_{p_n}^{(k_n)})}{\log(i/(np_n))}\right)^2 \approx \mathrm{E}\int_{j_n/k_n}^{1} Z_{T,\gamma}^2(s)\,\mathrm{d}s = \int_{j_n/k_n}^{1}\frac{\sigma_{T,\gamma}^2}{s}\,\mathrm{d}s = \sigma_{T,\gamma}^2 \log\frac{k_n}{j_n},$$

and likewise for the estimator $\hat{\sigma}_{T,\gamma,1}^2$, which leads to the normalizing factor $\log(k_n/j_n)$ in the definition of $\hat{\sigma}_{T,\gamma,1}^2$ and $\hat{\sigma}_{T,\gamma,2}^2$. For moderate sample sizes, however, this approximation is too crude, that is, it overestimates the left-hand side considerably and hence yields too short confidence intervals. More appropriate would be the normalizing factor

$$\begin{aligned}&\frac{1}{k_n\sigma_{T,\gamma}^2}\mathrm{E}\sum_{i=j_n}^{k_n}\left(Z_{T,\gamma}\left(\frac{i}{k_n}\right)-\frac{\log(k_n/(np_n))}{\log(i/(np_n))}Z_{T,\gamma}(1)\right)^2\\ &\quad=\sum_{i=j_n}^{k_n}\frac{1}{i}-\frac{2}{\sigma_{T,\gamma}^2}\frac{\log(k_n/(np_n))}{\log(i/(np_n))}\frac{\operatorname{cov}\big(Z_{T,\gamma}(i/k_n),\,Z_{T,\gamma}(1)\big)}{k_n}+\left(\frac{\log(k_n/(np_n))}{\log(i/(np_n))}\right)^2\frac{1}{k_n}. \qquad (30)\end{aligned}$$

Unfortunately, the covariance of $Z_{T,\gamma}(i/k_n)$ and $Z_{T,\gamma}(1)$ depends on the unknown limiting covariance function c, so (30) cannot be used directly for the estimation of the asymptotic variance. Instead we propose to use the lower bound

$$\begin{aligned}&\sum_{i=j_n}^{k_n}\frac{1}{i}-\frac{2}{\sigma_{T,\gamma}^2}\cdot\frac{\log(k_n/(np_n))}{\log(i/(np_n))}\frac{\big(\operatorname{var}Z_{T,\gamma}(i/k_n)\cdot\operatorname{var}Z_{T,\gamma}(1)\big)^{1/2}}{k_n}+\left(\frac{\log(k_n/(np_n))}{\log(i/(np_n))}\right)^2\frac{1}{k_n}\\ &\quad=\sum_{i=j_n}^{k_n}\left(i^{-1/2}-\frac{\log(k_n/(np_n))}{\log(i/(np_n))}k_n^{-1/2}\right)^2 \qquad (31)\end{aligned}$$

as the normalizing factor replacing $\log(k_n/j_n)$ in (27). Note that (31) is asymptotically equivalent to $\log(k_n/j_n)$. Thus the resulting modified estimator

$$\hat{\sigma}_{T,\gamma,3}^2 := \left(\sum_{i=j_n}^{k_n}\left(i^{-1/2}-\frac{\log(k_n/(np_n))}{\log(i/(np_n))}k_n^{-1/2}\right)^2\right)^{-1}\sum_{i=j_n}^{k_n}\left(\frac{\log(\hat{x}_{p_n}^{(i)}/\hat{x}_{p_n}^{(k_n)})}{\log(i/(np_n))}\right)^2 \qquad (32)$$

is also consistent for the asymptotic variance, yet for finite sample sizes it yields substantially more conservative confidence intervals. For example, the two-sided asymptotic confidence interval for the nominal coverage probability $1-\alpha\in(0,1)$ is given by

$$\left[\hat{x}_{p_n}\exp\left(-z_{\alpha/2}\hat{\sigma}_{T,\gamma,3}k_n^{-1/2}\log\frac{k_n}{np_n}\right),\ \hat{x}_{p_n}\exp\left(z_{\alpha/2}\hat{\sigma}_{T,\gamma,3}k_n^{-1/2}\log\frac{k_n}{np_n}\right)\right], \qquad (33)$$

with $z_{\alpha/2}$ denoting the $(1-\alpha/2)$ quantile of the standard normal distribution. For that reason, we will mainly use $\hat{\sigma}_{T,\gamma,3}^2$ to construct confidence intervals in our application and the simulation study.

Note that one may also modify the estimator $\hat{\sigma}_{T,\gamma,1}^2$ in a similar way, but it seems more natural to use a variance estimator that is based on quantile estimators if one is interested in confidence intervals for extreme quantiles.

3. Time series models

Here we demonstrate the applicability of the theory outlined in the previous section to specific time series models. First we consider solutions of certain stochastic recurrence equations, including ARCH(1) time series, and then linear time series.

3.1. Solutions of a stochastic difference equation

Consider the stochastic recursion

$$X_i = A_i X_{i-1} + B_i, \qquad i \in \mathbb{N}, \tag{34}$$

where (A_i, B_i), $i \in \mathbb{N}$, denote i.i.d. $\mathbb{R}^2$-valued random vectors. Such stochastic difference equations occur in many contexts. For example, X_i describes the balance of an account at time i if A_i denotes the inverse of the stochastic discount factor for the time interval from $i-1$ to i and B_i a random deposit made just before time i; see Embrechts *et al*. (1997, Section 8.4.1) for details.

Closely related is the ARCH(1) time series, which is a popular simple model for returns on a risky investment:

$$Y_i = \left(\alpha_0 + \alpha_1 Y_{i-1}^2\right)^{1/2} Z_i, \qquad i \in \mathbb{N}, \tag{35}$$

where Z_i are i.i.d. innovations with mean 0 and variance 1. Then $X_i = Y_i^2$ satisfies equation (34) with $A_i = \alpha_1 Z_i^2$ and $B_i = \alpha_0 Z_i^2$. Further applications of model (34) were discussed by Vervaat (1979).

In what follows, we assume that A_1 and B_1 have an absolute continuous df. Kesten (1973) proved that a stationary solution of (34) with heavy-tailed marginals exists if

(D1) $A_1, B_1 > 0$ and there exists $\kappa > 0$ such that

$$\mathrm{E}A_1^\kappa = 1, \quad \mathrm{E}\left(A_1^\kappa \max(\log A_1, 0)\right) < \infty \quad \text{and} \quad \mathrm{E}B_1^\kappa \in (0, \infty).$$

Then the df F of X_1 belongs to the domain of attraction of G_γ with extreme value index $\gamma = 1/\kappa$. Indeed, F satisfies $1 - F(x) \sim cx^{-\kappa}$ as $x \to \infty$ for some $c > 0$. Note that one obtains heavy tails for X_i even if the 'random coefficients' A_i and B_i have light tails.

In Drees (2000, Corollary 4.1) it is shown that the conditions of Theorem 2.1 are satisfied and hence (14) holds with covariance function

$$c(x, y) = x \wedge y + \sum_{j=1}^{\infty} \left(x \int_0^{y/x} P\left\{ \prod_{i=1}^{j} A_i > t^\gamma \right\} \mathrm{d}t + y \int_0^{x/y} P\left\{ \prod_{i=1}^{j} A_i > t^\gamma \right\} \mathrm{d}t \right) \tag{36}$$

if, in addition, the following conditions hold:

(D2) There exists $\xi > 0$ such that $\mathrm{E}A_1^{\kappa+\xi} < \infty$ and $\mathrm{E}B_1^{\kappa+\xi} < \infty$.
(D3) $\log^2 n \log^4(\log n) = o(k_n)$ and $k_n = o(n^{2\tau/(2\tau+1)})$ where $\tau > 0$ is such that

$$1 - F(x) = dx^{-1/\gamma}\big(1 + O(x^{-\tau/\gamma})\big). \tag{37}$$

Goldie (1989) proved that, under conditions (D1) and (D2), there does indeed always exist a $\tau > 0$ satisfying (37), which is a special case of (C4), while the upper bound on k_n required in (D3) is equivalent to (C5). Therefore, under the additional condition (15), we obtain the asymptotic normality of the statistical tail functionals and the pertaining quantile estimators as well.

Likewise, one may check the conditions of Theorem 2.1 for the ARCH(1) model (35). However, if the distribution of the innovations Z_i is symmetric, then (C1)–(C5) follow immediately from the corresponding conditions for Y_i^2 and thus from the aforementioned result established in Drees (2000). For example, (C2) for Y_i^2, combined with the relationship

$$P\left\{ Y_i > F_Y^{-1}\left(1 - \frac{k_n}{n}x\right),\ Y_j > F_Y^{-1}\left(1 - \frac{k_n}{n}y\right)\right\}$$
$$= \frac{1}{4} P\left\{ Y_i^2 > F_{Y^2}^{-1}\left(1 - \frac{k_n}{n}2x\right),\ Y_j^2 > F_{Y^2}^{-1}\left(1 - \frac{k_n}{n}2y\right)\right\},$$

implies (C2) for the ARCH(1) time series Y_i with

$$c_{m,Y}(x,\ y) = \frac{1}{4} c_{m,Y^2}(2x,\ 2y) = \frac{1}{2} c_{m,Y^2}(x,\ y)$$

and F_Y^{-1} and $F_{Y^2}^{-1}$ denoting the qf of Y_i and Y_i^2, respectively. Hence the analogous relation also holds for the limiting covariance functions c_Y and c_{Y^2} given by (36). Note that in general this covariance function cannot be calculated analytically, but Stărică (1998) proposed a method to compute it by simulation.

3.2. Linear time series

Here we examine classical linear time series

$$X_i = \sum_{j=0}^{\infty} \psi_j Z_{i-j}, \qquad i \in \mathbb{N}, \tag{38}$$

with i.i.d. innovations Z_i. Without loss of generality, we assume $\psi_0 = 1$. For simplicity, we confine ourselves to geometrically decreasing coefficients, that is,

$$|\psi_j| = O(\tau^j) \tag{39}$$

as $j \to \infty$ for some $\tau \in (0, 1)$; in particular, finite-order ARMA models are included. However, the results given below hold true under much weaker summability conditions on the coefficients (cf. Datta and McCormick 1998, Lemma 5.2; or Mikosch and Samorodnitsky 2000, Lemma A.3).

In model (38) the variables X_i are heavy-tailed if and only if the innovations have heavy tails. Hence the stochastic behaviour of the linear time series (38) is very different from that of the nonlinear time series considered in Section 3.1.

In what follows we assume that the df F_Z of Z_1 has balanced heavy tails, that is,

$$F_Z \in D(G_\gamma), \qquad \lim_{x\to\infty} \frac{1-F_Z(x)}{F_Z(-x)} = \frac{p}{q} \quad \text{for some } p = 1-q \in (0, 1) \tag{40}$$

(or, equivalently, $1 - F_Z(x) \sim px^{-1/\gamma} l(x)$ and $F_Z(-x) \sim qx^{-1/\gamma} l(x)$ as $x \to \infty$ for some slowly varying function l). Then, by Lemma 5.2 of Datta and McCormick (1998), the df F of X_1 satisfies

$$\frac{1-F(x)}{1-F_Z(x)} \to \sum_{j=0}^{\infty} \Big(p\psi_j^{1/\gamma} 1_{\{\psi_j > 0\}} + q|\psi_j|^{1/\gamma} 1_{\{\psi_j < 0\}} \Big) =: d_\psi \tag{41}$$

as $x \to \infty$. In particular, $F \in D(G_\gamma)$, too.

If, in addition, F_Z has a Lebesgue density f_Z which is L_1-Lipshitz continuous, that is,

$$\int |f_Z(z+u) - f_Z(z)| \, \mathrm{d}z = O(u) \tag{42}$$

as $u \downarrow 0$, then the time series X_i, $i \in \mathbb{N}$, is geometrically β-mixing (Doukhan 1994, Theorem 2.3.2). (For a finite-order ARMA process the mere existence of a Lebesgue density is sufficient; see Doukhan 1994, Theorem 2.4.6.) Hence condition (C1) is satisfied with $l_n = [\text{const.} \cdot \log n]$ provided k_n satisfies (9).

In the same way as in Lemma 5.1 of Datta and McCormick (1998), one can show that

$$\frac{P\{X_1 > u,\ X_{1+m} > uv\}}{1-F_Z(u)} \to \sum_{j=0}^{\infty} \Big(|\psi_j|^{1/\gamma} \wedge (v^{-1/\gamma} |\psi_{j+m}|^{1/\gamma}) \Big)$$

as $u \to \infty$, for $v > 0$ and $m > 1$. Combining this with (41) and $F^{-1}(1 - yk_n/n)/F^{-1}(1 - xk_n/n) \to (y/x)^{-\gamma}$, one obtains (C2) with

$$c_m(x, y) = \frac{1}{pd_\psi} \sum_{j=0}^{\infty} \Big((x|\psi_j|^{1/\gamma}) \wedge (y|\psi_{j+m}|^{1/\gamma}) \Big) \big(p1_{\{\psi_j \wedge \psi_{j+m} > 0\}} + q1_{\{\psi_j \vee \psi_{j+m} < 0\}} \big). \tag{43}$$

It is more complicated to check (C3) for general linear time series. Of course, any finite-order moving average meets this condition. More interesting is the example given by Bosq (1998, p. 18): if the innovations Z_i have finite variance (which is ensured by $\gamma < \frac{1}{2}$), then the time series X_i, $i \in \mathbb{N}$, is geometrically ρ-mixing and hence Remark 2.2 applies.

Though in practice it is often realistic to assume a finite variance, this condition is a little disturbing in an extreme value setting. As a simple example of a time series that is neither m-dependent for a finite m nor necessarily of finite variance, we consider a first-order autoregressive (AR(1)) process

$$X_i = \theta X_{i-1} + Z_i$$

for some $\theta \in (-1, 1)$. This time series has representation (38) with $\psi_j = \theta^j$, so that (C2) holds if the df of the innovations has a Lebesgue density and satisfies (40).

Next we verify condition (C3). We restrict ourselves to the case $\theta \geqslant 0$; the other case can be treated in the same way. Then the representation $X_{1+m} = \theta^m X_1 + \sum_{k=2}^{1+m} \theta^{1+m-k} Z_k$ shows that

$$\begin{aligned} P\{X_1 \in I_n(x, y),\ X_{1+m} \in I_n(x, y)\} &\leqslant P\left\{X_1 \in I_n(x, y), \theta^m X_1 > \theta^{1/2} F^{-1}\left(1 - \frac{k_n}{n} y\right)\right\} \\ &\quad + P\left\{X_1 \in I_n(x, y), \sum_{k=2}^{1+m} \theta^{1+m-k} Z_k > (1 - \theta^{1/2}) F^{-1}\left(1 - \frac{k_n}{n} y\right)\right\} \qquad (44) \\ &\leqslant \left(1 - F\left(\theta^{1/2-m} F^{-1}\left(1 - \frac{k_n}{n} y\right)\right) - \frac{k_n}{n} x\right)^+ \\ &\quad + \frac{k_n}{n}(y - x) \cdot P\left\{\sum_{j=0}^{\infty} \theta^j |Z_j| > (1 - \theta^{1/2}) F^{-1}\left(1 - \frac{k_n}{n}(1 + \varepsilon)\right)\right\}. \qquad (45) \end{aligned}$$

Here the second term is of order $(k_n/n)^2(y - x)$. By the Potter bounds (Bingham *et al.* 1987, Theorem 1.5.6), we have

$$\begin{aligned} \left(1 - F\left(\theta^{1/2-m} F^{-1}\left(1 - \frac{k_n}{n} y\right)\right) - \frac{k_n}{n} x\right)^+ &\leqslant \left(\theta^{(m-1)/(2\gamma)} \frac{k_n}{n} y - \frac{k_n}{n} x\right)^+ \\ &\leqslant \frac{k_n}{n} \theta^{(m-1)/(2\gamma)} (y - x). \end{aligned}$$

Combine this with (45) to obtain (C3).

To sum up, if X_i allows representation (38) with coefficients satisfying (39) and F_Z satisfying (40) and (42), and if k_n meets conditions (18), (9) and $k_n = O(n/\log n)$, then the approximation (14) of the tail empirical qf Q_n holds with limiting covariance function given by (12) and (43), provided that $\gamma < \frac{1}{2}$, or $\psi_j = 0$ for all but finitely many j, or $\psi_j = \theta^j$ for some $\theta \in (-1, 1)$. Hence, under the additional conditions (19) and (15) on k_n, the asymptotic normality of the quantile estimator $\hat{x}_{p_n}$ follows.

Notice that for the AR(1) model the asymptotic variance is particularly simple if one uses the Hill or the maximum likelihood estimator for the estimation of the extreme value index, since

$$c(1, 1) = 1 + 2 \cdot \begin{cases} \theta^{1/\gamma}/(1 - \theta^{1/\gamma}), & \text{if } \theta \geqslant 0, \\ |\theta|^{1/\gamma}/(1 - |\theta|^{2/\gamma}), & \text{if } \theta < 0. \end{cases}$$

Hence, within this model, one may construct confidence intervals without using the variance estimators discussed in Section 2. Instead one may define an estimator of $c(1, 1)$ using, for example, the same estimator for γ as for the quantile estimation and the sample autocorrelation function at lag 1 as an estimator of θ.

Resnick and Stărică (1997) demonstrated that, if one trusts in the simple AR(1) model, one obtains more accurate estimates of the extreme value index by first estimating θ and then the extreme value index based on the resulting residuals $X_i - \hat{\theta} X_{i-1}$. By fitting a Pareto distribution to the tails of the residuals and then using relation (41), one might also obtain an accurate estimate of extreme quantiles of F. The main advantage of the approach presented here is its robustness, as it does not rely on a specific model but yields reasonable estimates under mild structural assumptions.

4. The Nasdaq Composite index: a case study

In this section we analyse the 'risk' of a large hike in the Nasdaq Composite index. (In fact, it is a risk for investors betting on a fall in the index, which may seem a reasonable strategy given the huge losses observed in the last months of the period considered here.) More precisely, we examine the (log-)returns $X_i = \log(S_i/S_{i-1})$, $1 \leqslant i \leqslant n$, with S_i denoting the index calculated at the end of the ith trading day in the years 1997 to 2000, amounting to a sample size $n = 1007$. We do not consider negative returns, that is to say, the left tail of the returns, because, somewhat surprisingly, there is only very weak evidence for a positive extreme value index there. Hence for the analysis of the left tail one must apply estimators for general $\gamma \in \mathbb{R}$, which we discuss in less detail in Section 6.

A scatterplot of these returns is given in Figure 1. To some extent, there is an increasing trend in the volatility, which seems to contradict stationarity of the time series. On the other hand, bursts of volatility have also been observed in the first half of the observation period (for i about 200 and 400). Moreover, the observed increase may also be due to a persistence

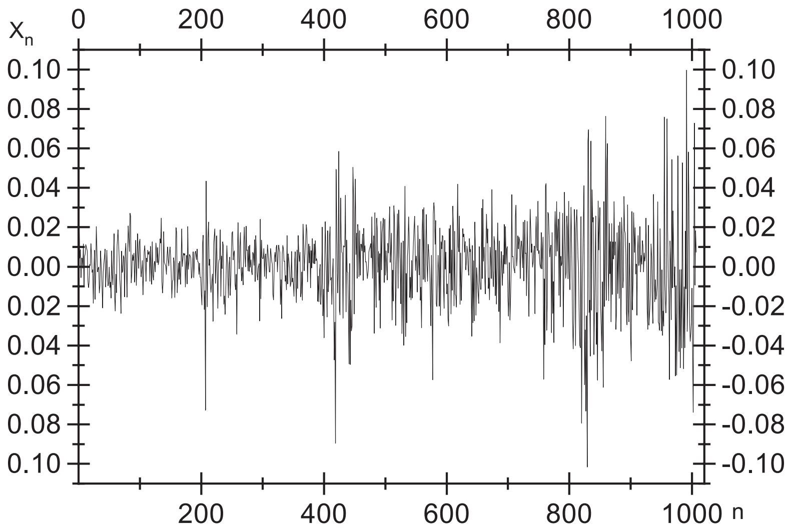

Figure 1. Log-returns of the Nasdaq Compostie index, 1997–2000.

in the volatility after these random shocks. All in all, given the moderate length of the period under consideration and the fact that there was no obvious structural change in the economic environment during this period, stationarity may be regarded as a reasonable approximation to reality. (Contrary to that judgement, Stărică and Granger (2001) argue for shorter periods of stationarity of the S&P500 index in the second half of the 1990s.)

In the following, we aim to estimate the upper $p_n = 0.001$ quantile $x_{p_n} = F^{-1}(0.999)$ under the assumption of stationarity. Note that np_n is about 1, so that we are actually looking for an extreme quantile.

Figure 2(a) displays the graphs of the Hill estimator, the maximum likelihood estimator and the moment estimator proposed by Dekkers *et al*. (1989) as a function of k, the number of largest order statistics reduced by 1. All estimates are positive for k from about 50 to 460, so that we may assume a heavy-tailed distribution. However, the values obtained by the different estimation methods differ quite a lot. In particular, the Hill estimator shows a clear upward trend starting from $k = 100$, whereas the curve pertaining to the maximum likelihood estimator is much more stable. This may indicate that the Pareto approximation (3) becomes sufficiently accurate only after a suitable shift of the data, that is, $F^{-1}(1-t) \approx dt^{-\gamma} + \mu$ for some $\mu \neq 0$, because it is well known that a non-vanishing location parameter leads to a large bias of the Hill estimator, whereas the maximum likelihood estimator is invariant under a shift transformation and the moment estimator is less sensitive to shifts than the Hill estimator.

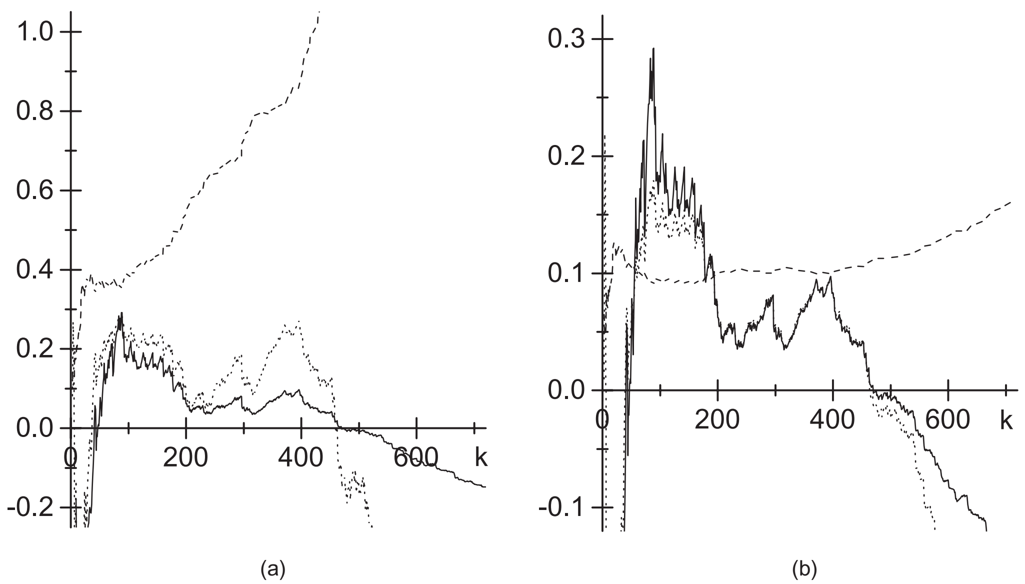

Figure 2. Hill (dashed curve), moment (dotted) and maximum likelihood estimator (solid) for (a) original returns and (b) returns shifted into positive real halfline.

This hypothesis can be checked by subtracting a suitable constant from the data. A choice suggesting itself is the smallest observation, since after this shift all transformed data points are non-negative, thus allowing us to use (almost) up to the full sample for the Hill and the moment estimator. Figure 2(b) shows the resulting estimates for the extreme value index based on these shifted data. Now the behaviour of the Hill estimator has changed completely, yielding an almost flat line for k ranging from 100 to 400. (Note that the scale of the y-axis has been changed to magnify the relevant range of y-values.) Even more strikingly, the moment and the maximum likelihood estimator (which is not influenced by the transformation) are now almost identical for k between 180 and 470.

From our experience with several data sets, as a rule of thumb this similarity indicates that the pure Pareto approximation without a location parameter is particularly accurate, that is, $F_s^{-1}(1-t) \approx dt^{-\gamma}$ with F_s denoting the df of the shifted random variables. To check this for the shifted data set under consideration, in Figure 3 we have plotted a linearly interpolated version of the tail empirical qf $Q_{n,k}$ for $k = 400$ based on the transformed data together with the estimated Pareto approximation $X_{n-k:n}t^{-\hat{\gamma}_n^{(H)}}$ using the Hill estimator $\hat{\gamma}_n^{(H)} \approx 0.10$. The fit is convincing for the whole unit interval and almost perfect for its upper half.

Encouraged by this fit, we carry on with the statistical analysis of the shifted returns. Figure 4 displays the estimator $\hat{\sigma}_{T,\gamma,3}$ whose square is defined in (32). Here we have used the Hill estimator and $j_n = 2$, the smallest integer exceeding np_n, so that $\hat{\sigma}_{T,\gamma,3}$ is well defined. (Different small values for $j_n > np_n$ lead to similar results, but if j_n is chosen too large then the performance of the variance estimator deteriorates.)

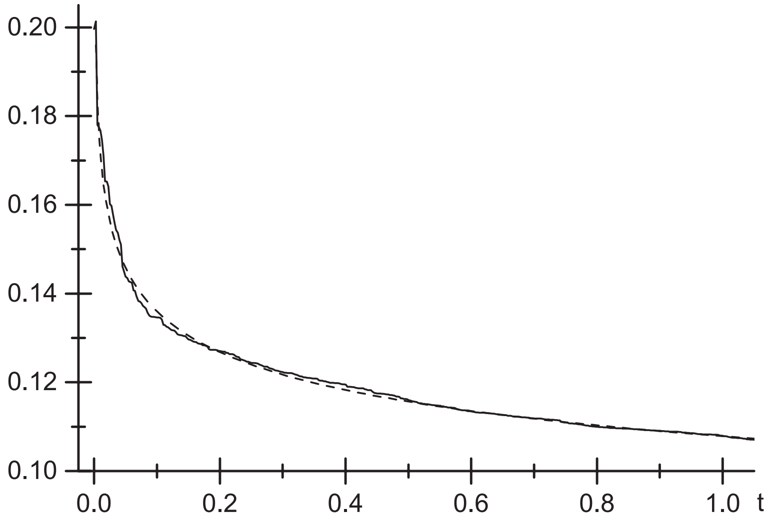

Figure 3. Continuous version of $Q_{n,k}$ for shifted log-returns with $k = 400$ (solid line) and estimated Pareto approximation (dashed).

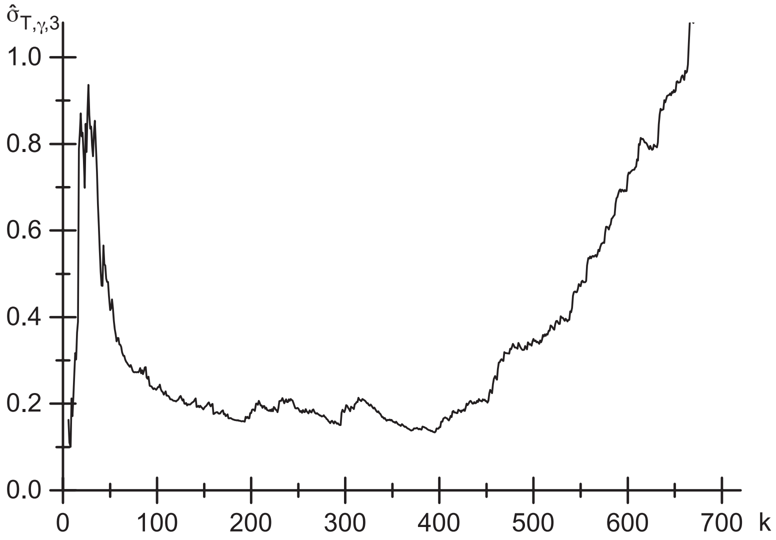

Figure 4. Estimated asymptotic standard deviation $\hat{\sigma}_{T,\gamma,3}$.

After large fluctuations when only few order statistics are used for estimation, the curve stabilizes at a value slightly below 0.2. Then, starting at about $k = 400$, there is a strong upward trend in the curve, suggesting that a non-negligible bias has appeared. Note that the kink in the curve at $k = 400$ is much more pronounced in this plot than in the graph of the estimators for the extreme value index or the extreme quantile (Figure 5). Hence to plot $\hat{\sigma}_{T,\gamma,3}$ against k might be a useful data-analytic tool for choosing a suitable sample fraction, even in the case of i.i.d. data where such an estimate of the variance is not needed for the construction of confidence intervals.

After these preparations, we arrive at our final plot in Figure 5. Here the quantile estimator $\hat{x}_{p_n}$ is plotted against k, together with 99% confidence intervals (33) and the confidence intervals

$$[\hat{x}_{p_n} \exp(-z_{\alpha/2}\hat{\gamma}_n^{(\mathrm{H})} k_n^{-1/2} \log(k_n/(np_n))), \hat{x}_{p_n} \exp(z_{\alpha/2}\hat{\gamma}_n^{(\mathrm{H})} k_n^{-1/2} \log(k_n/(np_n)))], \tag{46}$$

suggested by the theory for i.i.d. data. The estimators are calculated from the shifted data and then the shift has been corrected, so that the graphs show the estimates for the original distribution of the returns. For $k = 400$ one obtains a quantile estimate of about 0.096 with confidence interval (33) equal to [0.075, 0.119]. As expected, the intervals ignoring the serial dependence are much shorter than the intervals obtained by the new approach presented here, indicating that perhaps the former claim a much higher estimation accuracy than is actually achieved. Despite this fact, in the literature about the statistical analysis of financial series with a clear serial dependence often confidence intervals are displayed which are motivated

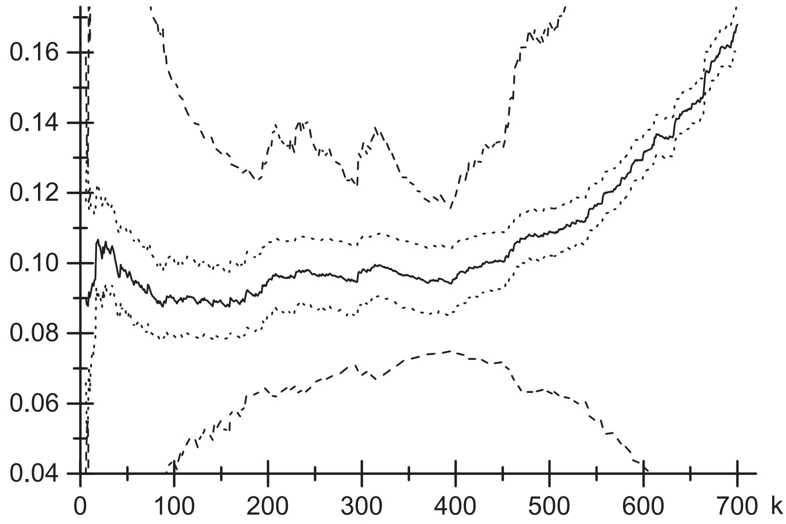

Figure 5. Estimated 0.999 quantile of the original return distribution (solid curve) with 99% confidence intervals (33) (dashed) and (46) (dotted).

by the classical extreme value theory for independent data; see Longin (1996), Caserta *et al*. (1998) and Müller *et al*. (1998). This, of course, does not mean that the standard confidence intervals are necessarily too short because in some cases the dependence may be negligible for large observations, but the theoretical justification for these confidence intervals is very weak.

It is also worth noting that, unlike the intervals (46), the confidence intervals based on the estimator $\hat{\sigma}_{T,\gamma,3}$ automatically widen for large k where the bias kicks in. Hence they actually reflect not only the variance of the quantile estimator but also the bias, thus avoiding an empty intersection of confidence intervals based on different sample fractions. In contrast, the standard confidence intervals are completely misleading if too many order statistics are used for estimation. In fact, in our simulation study it turned out that in some cases the actual coverage probability of the new intervals comes quite close to the nominal probability even for large k, though we do not offer any theoretical explanation for this effect.

5. Simulations

In this section we study the actual coverage probabilities of the two-sided confidence interval (33) derived in the present paper in comparison with those of the confidence interval (46) suggested by the theory for i.i.d. samples. Here both types of confidence interval are calculated for the nominal coverage probability of 95%, and they are based on the Hill estimator for γ. In the definition of $\hat{\sigma}_{T,\gamma,3}$ used in (33) we choose j_n equal to 2 for $p_n \leqslant 1/n$ and $j_n = 3$ for $p_n = 2/n$ so that $\log(j_n/(np_n))$ is strictly positive. All simulations were carried out using the programming language StatPascal which is part of the software package XTREMES (see Reiss and Thomas 2001).

As examples of linear time series we consider four ARMA(1, 1) models

$$X_i - \phi X_{i-1} = Z_i + \theta Z_{i-1}. \tag{47}$$

Here the i.i.d. innovations Z_i have a two-sided Pareto df with extreme value index $\gamma = 1/3$, that is,

$$1 - F_Z(x) = F_Z(-x) = \frac{1}{2} x^{-3}, \qquad x \geqslant 1,$$

and

(i) $\phi = 0.95, \quad \theta = 0.9,$
(ii) $\phi = 0.95, \quad \theta = -0.6,$
(iii) $\phi = 0.95, \quad \theta = -0.9,$
(iv) $\phi = 0.3, \quad \theta = 0.9,$

respectively. Observe that the innovations have finite variance and thus, according to Section 3.2, these models satisfy the conditions of Theorem 2.2.

In models (i)–(iii) the dependence is mainly due to the autoregressive part and, roughly speaking, the degree of dependence decreases from model (i) to model (iii) as the effect of the large autoregressive parameter $\phi = 0.95$ is partly compensated by the negative moving average parameter θ. (Note that for $\theta = -\phi$ one has $X_i = Z_i$, that is, independent random variables are observed.) In model (iv) the dependence is locally strong due to the large moving average parameter θ, but it has a very short memory because ϕ is small.

In addition, we consider two nonlinear (G)ARCH time series

$$X_i = \sigma_i Z_i$$

with i.i.d. standard normal innovations Z_i and

(v) $\sigma_i^2 = 0.0001 + 0.9 X_{i-1}^2,$
(vi) $\sigma_i^2 = 0.0001 + 0.4 X_{i-1}^2 + 0.5 \sigma_{i-1}^2,$

respectively. For the ARCH(1) model (v) our conditions have been checked in Section 3.1. GARCH(1, 1) time series like (vi) are widely used in finance to model returns of risky assets. It is known that such time series are geometrically β-mixing (Doukhan 1994, Section 2.4.2.3), but conditions (C2) and (C3) have not yet been verified. The choice of the parameters describing the influence of X_{i-1}^2 and σ_{i-1}^2 on σ_i^2 is motivated by the observation that in financial applications typically the sum of these parameters is close to but less than 1.

Finally, we also simulate i.i.d. sequences of Fréchet random variables with df

(vii) $F(x) = \exp(-x^{-3}),\ x > 0,$

in order to examine the performance of the confidence interval (33) in a situation where the interval (46) is appropriate.

The quantiles $x_{p_n} = F^{-1}(1 - p_n)$ are to be estimated for $p_n = 1/n$ and $p_n = 1/(5n)$. Since the quantiles are not known exactly for models (i)–(vi), they are determined by simulation. For this purpose, recall from Theorem 2.1 that an empirical intermediate quantile is asymptotically normal with median equal to the pertaining true quantile. Thus we simulate $m = 1000$ time series of length 5×10^6 and estimate x_{p_n} by the median of the empirical $(1 - p_n)$ quantiles. Table 1 gives the resulting estimates and 95% confidence intervals $[Y_{[(1-z_{0.025}m^{-1/2})m/2]:m}, Y_{[(1+z_{0.025}m^{-1/2})m/2]:m}]$ with Y_i, $1 \leqslant i \leqslant m$, denoting the observed empirical quantiles.

Next, $m = 10\,000$ time series of length $n = 2000$ are simulated from each of the above models and the relative frequency of samples is determined for which the true quantile lies outside the confidence intervals (33) and (46), respectively. In Figure 6 the resulting empirical non-coverage probabilities of (33) and (46) are plotted against k, the number of order statistics reduced by 1, for models (i)–(vi) and $p_n = 1/n = 0.0005$. The nominal level 5% is indicated by the dotted line. The maximal k-values are chosen such that in (almost) all samples the kth largest order statistic is still positive, so that the Hill estimator is well defined.

The confidence interval (46) derived from the theory for i.i.d. samples yields an acceptable level of non-coverage only for the ARMA(1, 1) model (iii), which is close to an i.i.d. model. In all other cases, the non-coverage probability is always larger than 13% and typically larger than 20%. Moreover, if k is taken too large such that a non-negligible bias appears, then the probability of non-coverage increases rapidly, as the confidence interval (46) does not take into account any bias.

In contrast to that behaviour, the confidence interval (33) is unbiased in all cases if one uses at least 40 order statistics for estimation, except in the ARCH(1) model (v) with k between 240 and 410 when the nominal level is exceeded by less than 2%. At first glance, it is somewhat surprising that the confidence interval is most conservative for the ARMA(1, 1) model (i) which exhibits the strongest dependence. This, however, is due to the rather poor fit of the tail of the stationary distribution by a Pareto distribution if one uses more than 200 order statistics. As a result, the quantile estimator has a large bias if k is much bigger than 200 (as can be seen from the quickly increasing actual non-coverage probability

Table 1. Estimated quantiles x_{p_n} for models (i)–(vi) with 95% confidence intervals

Model	$p_n = 0.0005$		$p_n = 0.0001$	
(i)	41.88,	[41.75, 41.96]	63.77,	[63.35, 64.28]
(ii)	11.74,	[11.71, 11.76]	19.03,	[18.94, 19.13]
(iii)	10.02,	[10.00, 10.03]	17.13,	[17.08, 17.17]
(iv)	14.59,	[14.56, 14.61]	24.38,	[24.32, 24.47]
(v)	0.2479,	[0.2452, 0.2494]	0.4940,	[0.4854, 0.5029]
(vi)	0.2114,	[0.2109, 0.2117]	0.3450,	[0.3435, 0.3466]

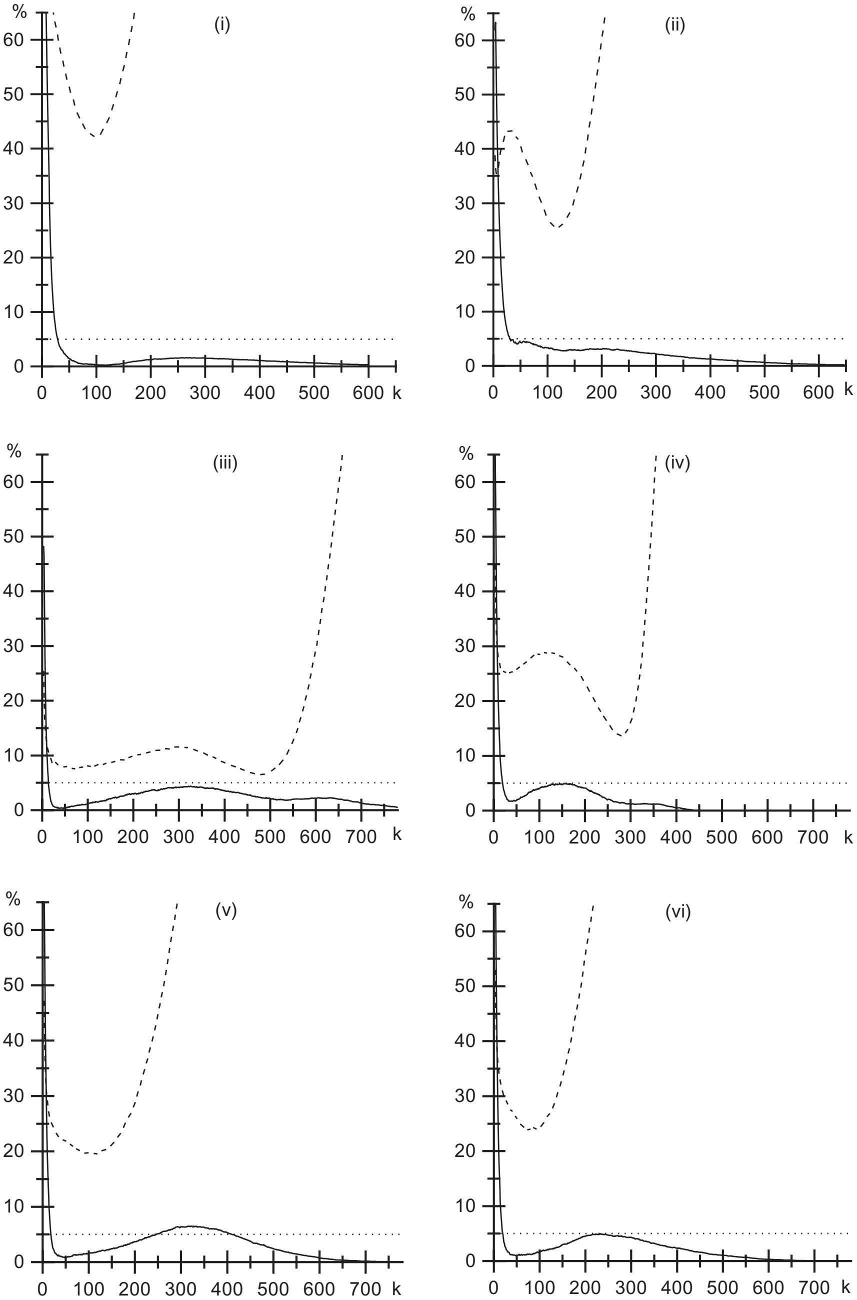

Figure 6. Empirical non-coverage probabilities of (33) (solid line) and (46) (dashed line) for $p_n = 1/n$; the nominal probability 5% is indicated by the dotted line.

of the i.i.d. confidence interval (46)). This in turn leads to an overestimation of the asymptotic variance by $\hat{\sigma}^2_{T,\gamma,3}$ and hence to too wide a confidence interval (33).

Figure 7 is the analogue of Figure 6 for $p_n = 1/(5n) = 0.0001$. By and large, the performance of the confidence interval assuming i.i.d. observations is the same as for $p_n = 1/n$. In contrast, the non-coverage probabilities of the confidence interval (33) are considerably higher than in Figure 6. This is particularly true for the nonlinear time series models (v) and (vi), where the actual probability is much larger than the nominal level for most k. This problem is mainly due to the large estimation error of the estimator $\hat{\sigma}^2_{T,\gamma,3}$ for the asymptotic variance, which is based on the quantile estimates $\hat{x}^{(i)}_{1/(5n)}$. The dash-dotted line in Figure 7 shows the empirical non-coverage probability when this variance estimator is replaced with the one based on the quantile estimates $\hat{x}^{(i)}_{1/n}$, that is, the same estimator as used in Figure 6. Indeed, now the nominal 5% level is exceeded only for very small k and, for models (iv) and (v) with k between 90 and 240 and between 230 and 430 respectively, by merely a few percentage points. So apparently the estimates for $x_{1/(5n)}$ are not reliable enough to be used for the estimation of the asymptotic variance. This, of course, is not completely surprising, since it is much more delicate to estimate the quantile $x_{1/(5n)}$ which lies far outside the range of observations than to estimate the quantile $x_{1/n}$ on the boundary of that range.

Next, we consider our model (vii) of i.i.d. Fréchet observations (Figure 8). Not surprisingly, the confidence interval (46) derived from the theory for i.i.d. observations does a very good job if one uses an appropriate number of order statistics, while the confidence interval (33) is often too conservative. On the other hand, the latter is less sensitive to a misspecification of the sample fraction used for estimation, albeit that the nominal level is exceeded if k is chosen much too large.

As usual in extreme value theory, the choice of the sample fraction used for estimation is crucial for the performance of the quantile estimators and the pertaining confidence intervals. Given a fixed level $1-\alpha$, one often aims at a confidence interval as short as possible such that the coverage probability is at least $1-\alpha$. Hence it seems natural to choose k such that the estimate of the asymptotic variance is minimized. Obviously, this approach relies on good variance estimates. Therefore, as mentioned above, the estimator $\hat{\sigma}^2_{T,\gamma,3}$ should be based on quantile estimators $\hat{x}^{(i)}_{\tilde{p}_n}$ for a quantile $x_{\tilde{p}_n}$ that lies inside the range of observations. On the other hand, $\tilde{p}_n$ must be sufficiently small to justify the use of extreme value theory. As a compromise between these conditions, we choose $\tilde{p}_n = 2/n$. (Taking $\tilde{p}_n = 1/n$ as in Figure 7 leads to slightly worse results, with non-coverage probabilities about 1–2% higher than reported in Table 2.) In addition, one has to rule out k being too small, since then the variance estimates are not reliable. Here we restrict k to values larger than or equal to 80, that is, at least 4% of the sample is used for estimation. In addition, we exclude unrealistic small-variance estimates by requiring that the estimate is at least $(\hat{\gamma}^{(k)}_n)^2$, the estimated variance in the case of independent observations. To sum up, we choose

$$\hat{k} := \arg\min\left\{ \hat{\sigma}^{(k)}_{T,\gamma,3} \,\middle|\, k \geqslant 80,\, \hat{\sigma}^{(k)}_{T,\gamma,3} \geqslant \hat{\gamma}^{(k)}_n \right\} \tag{48}$$

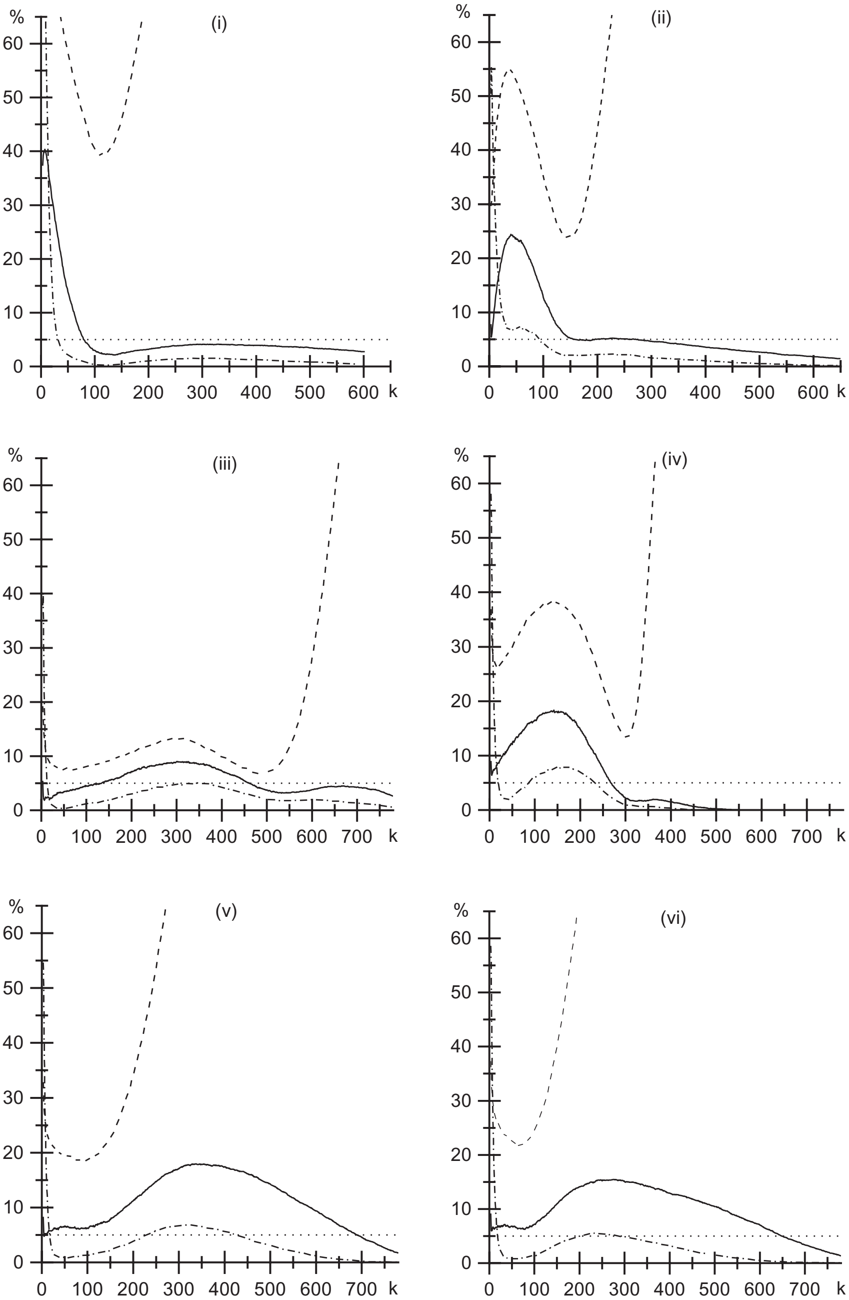

Figure 7. Empirical non-coverage probabilities of (33) (solid line), (33) with variance estimator $\hat{\sigma}^2_{T,\gamma,3}$ based on $\hat{x}^{(i)}_{1/n}$ (dash-dotted line) and (46) (dashed line) for $p_n = 1/(5n)$; the nominal probability 5% is indicated by the dotted line.

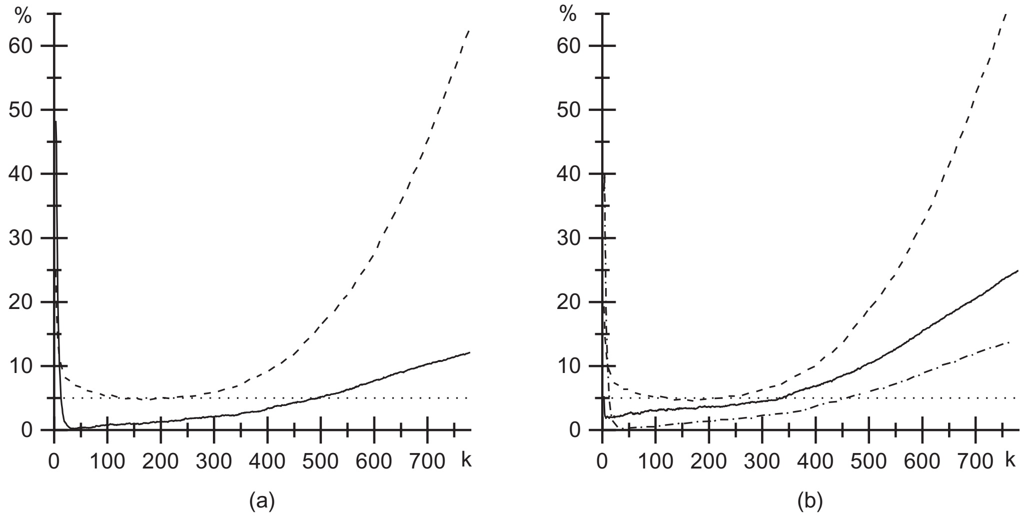

Figure 8. Empirical non-coverage probabilities of (33) (solid line), (33) with variance estimator $\hat{\sigma}^2_{T,\gamma,3}$ based on $\hat{x}^{(i)}_{1/n}$ (dash-dotted line) and (46) (dashed line) for the Fréchet model (vii) and (a) $p_n = 1/n$, (b) $p_n = 1/(5n)$; the nominal probability 5% is indicated by the dotted line.

Table 2. Empirical non-coverage probabilities for models (i)–(vi) with k chosen according to (48)

Model	$p_n = 0.0005$	$p_n = 0.0001$
(i)	2.5%	2.2%
(ii)	5.3%	6.6%
(iii)	6.1%	6.7%
(iv)	10.1%	14.1%
(v)	7.7%	8.6%
(vi)	5.5%	6.3%
(vii)	5.4%	6.0%

where $\hat{\sigma}^{(k)}_{T,\gamma,3}$ is the estimator of the asymptotic standard deviation defined analogously to (32), but based on the estimators $\hat{x}^{(i)}_{2/n}$, $3 \leqslant i \leqslant k$, instead of $\hat{x}^{(i)}_{p_n}$.

The resulting empirical probabilities of non-coverage are reported in Table 2. With the exception of the ARMA(1, 1) model (iv), the method works pretty well: the nominal level is at most exceeded by just a narrow margin in models (i)–(iii), (vi) and (vii), and by about 2.5–3.5% for the ARCH(1) model (v). In contrast, the actual probability of non-coverage is 2–3 times as large as the nominal one in model (iv), which exhibits a strong local but very short-ranged dependence. Nevertheless, the approach to minimizing the estimated asymptotic variance seems very promising if reliable variance estimates are at hand.

Finally, we discuss the effect, observed in the analysis of the Nasdaq Composite index, that a shift in the data can considerably improve the fit of the extremes by a Pareto distribution and consequently also the estimation accuracy. In the present study this applies particularly to the nonlinear ARCH(1) and GARCH(1, 1) time series used to model the returns of risky financial assets. For example, Figure 9 shows the Pareto quantile–quantile plot $(\log((n+1)/i), \log X_{n-i+1:n})_{1\leqslant i\leqslant n^+}$ for a GARCH time series of size $n = 50\,000$ drawn from model (vi) and for the sample shifted by 0.035. (Here n^+ denotes the number of positive observations; we use simulated data to obtain an estimate for the unknown dfs.) Clearly the quantile–quantile plot for the shifted data set can be well approximated by a line over a much wider range than the plot for the original data, thus indicating that a larger sample fraction of extremes can be fitted well by a Pareto distribution. Here the amount by which the data set is shifted does not depend on the particular sample (but, of course, on the model). Indeed, the value 0.035 was chosen so that the Hill plot for a different sample from model (vi) appeared flat, and the moment estimator and the maximum likelihood estimator yield similar results over a wide range of k-values; cf. the discussion in Section 4.

Figure 10 displays the empirical non-coverage probabilities of the confidence intervals (33) and (46) for the GARCH(1, 1) model (vi) shifted by 0.035. From the curves corresponding to the interval (46) one can see that a significant bias occurs only if k is taken to be larger than 400, whereas for the original model this happens for about $k \geqslant 150$. Hence on the average one obtains much shorter confidence intervals by choosing k between

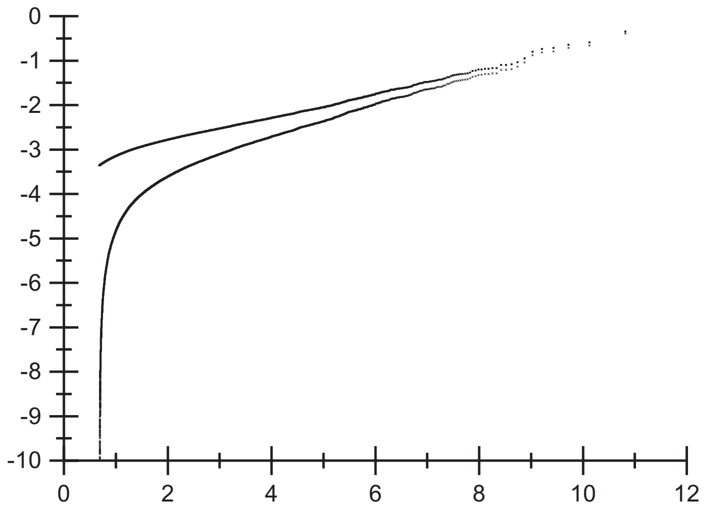

Figure 9. Pareto quantile–quantile plot for GARCH(1, 1) model (vi) (lower plot) and for model shifted by 0.035 (upper plot).

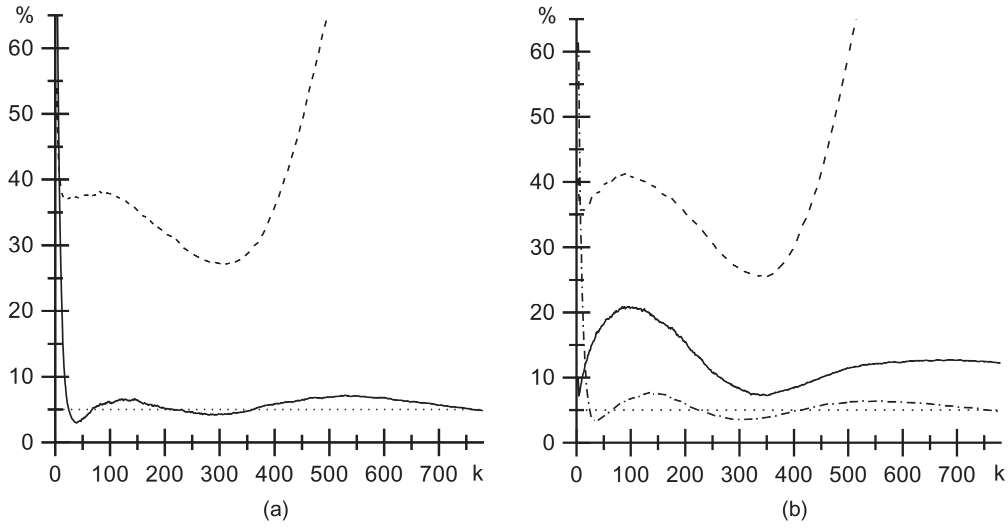

Figure 10. Empirical non-coverage probabilities of (33) (solid line), (33) with variance estimator $\hat{\sigma}^2_{T,\gamma,3}$ based on $\hat{x}^{(i)}_{1/n}$ (dash-dotted line) and (46) (dashed line) for the GARCH(1, 1) model (vi) shifted by 0.035 and (a) $p_n = 1/n$ (left), (b) $p_n = 1/(5n)$; the nominal probability 5% is indicated by the dotted line.

250 and 350, say, leading to non-coverage probabilities of about 4–5% for $p_n = 1/n$ and about 3.5–5% for $p_n = 1/(5n)$. More precisely, although these probabilities are smaller than those reported in Table 2 for the original GARCH model when almost shortest confidence intervals are used, there the average length of the confidence intervals is more than 3 times as large as in the shifted model for $p_n = 1/n$ and about 6 times as large for $p_n = 1/(5n)$. This demonstrates the huge improvement in the estimation accuracy achieved by an appropriate shift of the data. In addition, the confidence interval becomes less sensitive to a misspecification of the sample fraction used for estimation, as far as the coverage probability is concerned.

6. Asymptotics: the general case

In this section we analyse the asymptotic behaviour of quantile estimators of type (7) when $F \in D(G_\gamma)$ for some $\gamma \in \mathbb{R}$. Unlike in the special case $\gamma > 0$, there is no simple unifying representation of the quantile function that is sufficient for $F \in D(G_\gamma)$ for all $\gamma \in \mathbb{R}$. Therefore we replace (C4) and (C5) with an analogue of condition (18) based on convergence (6):

(C4)′ $\lim_{n\to\infty} k_n^{1/2} \sup_{0<t\leq 1+\varepsilon} t^{\gamma+1/2}(1+|\log t|)^{-1/2}|\tilde{R}(k_n/n,\, t)| = 0$

with

$$\tilde{R}(\lambda,\, t) := \frac{F^{-1}(1-\lambda t) - F^{-1}(1-\lambda)}{a(\lambda)} - \frac{t^{-\gamma}-1}{\gamma}.$$

Then we have the following counterpart to Theorem 2.1 (see Drees 2000):

Theorem 6.1. *Under conditions* (C1)–(C3) *and* (C4)′ *for some* $l_n = o(n/k_n)$ *there exist versions of the tail empirical qf* Q_n, *random variables* D_n *and a centred Gaussian process* e *with covariance function* c *defined by* (12) *such that*

$$\sup_{t\in(0,1]} t^{\gamma+1/2}(1+|\log t|)^{-1/2}\left| k_n^{1/2}\left(\frac{Q_n(t)-D_n}{a(k_n/n)} - \frac{t^{-\gamma}-1}{\gamma}\right) - t^{-(\gamma+1)}e(t)\right| \to 0 \qquad (49)$$

in probability.

Remark 6.1. For $\gamma \geqslant -\frac{1}{2}$ the random variables D_n may be replaced with $F^{-1}(1-k_n/n)$, while for $\gamma < -\frac{1}{2}$ one merely has $D_n - F^{-1}(1-k_n/n) = o_P(k_n^{-1/2})$; see Drees (1998a) for more about D_n.

Remark 6.2. Remark 2.4 also applies in the present case.

As in the case $\gamma > 0$, the extreme value index may be estimated by a statistical tail functional $T(Q_n)$. However, in conditions (T1)–(T3) $t^{-\gamma}$ must be replaced with $(t^{-\gamma}-1)/\gamma$ and, in addition to the scale invariance of T, we need location invariance to deal with the random shift by D_n in (49). This leads to the following modified conditions:

(T1)′ $T(az+b) = T(z)$ for all $a>0$ and $b\in\mathbb{R}$.

(T2)′ $T\left(\left(\frac{t^{-\gamma}-1}{\gamma}\right)_{0<t\leqslant 1}\right) = \gamma.$

(T3)′ There exists a signed measure $\nu_{T,\gamma}$ on (0,1] with $\int_{(0,1]} t^{-\gamma-1/2}(1+|\log t|)^{1/2}\,|\nu_{T,\gamma}|(\mathrm{d}t) < \infty$ such that

$$\varepsilon_n^{-1}\left(T\left(\left(\frac{t^{-\gamma}-1}{\gamma} + \varepsilon_n z_n(t)\right)_{0<t\leqslant 1}\right) - T\left(\left(\frac{t^{-\gamma}-1}{\gamma}\right)_{0<t\leqslant 1}\right)\right) \to \int_{(0,1]} z(t)\nu_{T,\gamma}(\mathrm{d}t)$$

for all $\varepsilon_n \downarrow 0$ and z_n satisfying

$$\sup_{0<t\leqslant 1} t^{\gamma+1/2}(1+|\log t|)^{-1/2}|z_n(t)-z(t)| \to 0$$

for some continuous function z as described in (T0).

Note that, for $\gamma > 0$, here $\nu_{T,\gamma}$ has a slightly different meaning than in (T3), since here we consider a derivative of T at $(t^{-\gamma} - 1)/\gamma$.

Next we need an estimator of the scale function a. To this end, we can employ a similar approach, that is, we estimate a by a smooth functional $S(Q_n)$. Like T, the functional S should be invariant under shifts but it must be *equivariant* under scale transformations. Moreover, S should give the value 1 when applied to the standard generalized Pareto qf. Hence we impose the following conditions:

(S0) S is a Borel measurable real-valued functional on the set of functions $z \in D(0, 1]$ satisfying $t^{\gamma+1/2}|\log t|^{-1/2}z(t) \to 0$ as $t \downarrow 0$.

(S1) $S(az + b) = aS(z)$ for all $a > 0$ and $b \in \mathbb{R}$.

(S2) $S\left(\left(\dfrac{t^{-\gamma} - 1}{\gamma}\right)_{0<t\leqslant 1}\right) = 1.$

(S3) There exists a signed measure $\mu_{S,\gamma}$ on (0,1] with $\int_{(0,1]} t^{-\gamma-1/2}(1 + |\log t|)^{1/2}\,|\mu_{S,\gamma}|(\mathrm{d}t) < \infty$ such that

$$\varepsilon_n^{-1}\left(S\left(\left(\frac{t^{-\gamma} - 1}{\gamma} + \varepsilon_n z_n(t)\right)_{0<t\leqslant 1}\right) - S\left(\left(\frac{t^{-\gamma} - 1}{\gamma}\right)_{0<t\leqslant 1}\right)\right) \to \int_{(0,1]} z(t)\mu_{S,\gamma}(\mathrm{d}t)$$

for all $\varepsilon_n \downarrow 0$ and z_n satisfying

$$\sup_{0<t\leqslant 1} t^{\gamma+1/2}(1 + |\log t|)^{-1/2}|z_n(t) - z(t)| \to 0$$

for some continuous function z as described in (S0).

Example. The estimator (8) is of this type with

$$S(z) = (z(\tfrac{1}{2}) - z(1))\frac{T(z)}{2^{T(z)} - 1}$$

if $\hat{\gamma}_n = T(Q_n)$ for some T satisfying (T0) and (T1)$'$–(T3)$'$.

Conditions (S0)–(S2) are readily verified. To check (S3) note that, with $y_\gamma(t) := (t^{-\gamma} - 1)/\gamma$, condition (T3)$'$ and a Taylor expansion of $x \mapsto x/(2^x - 1)$ at γ yield

$$\frac{T(y_\gamma + \varepsilon_n z_n)}{2^{T(y_\gamma+\varepsilon_n z_n)} - 1} = \frac{\gamma}{2^\gamma - 1} + \varepsilon_n \frac{2^\gamma - 1 - \gamma 2^\gamma \log 2}{(2^\gamma - 1)^2}\int_{(0,1]} z(t)\nu_{T,\gamma}(\mathrm{d}t) + o(\varepsilon_n),$$

which for $\gamma = 0$ is to be interpreted as the limit for $\gamma \to 0$. Hence

$$\begin{aligned}
&\varepsilon_n^{-1}(S(y_\gamma + \varepsilon_n z_n) - S(y_\gamma)) \\
&\quad = (z_n(\tfrac{1}{2}) - z_n(1))\frac{\gamma}{2^\gamma - 1} + (y_\gamma(\tfrac{1}{2}) - y_\gamma(1))\varepsilon_n^{-1}\left(\frac{T(y_\gamma + \varepsilon_n z_n)}{2^{T(y_\gamma+\varepsilon_n z_n)} - 1} - \frac{\gamma}{2^\gamma - 1}\right) + o(1) \\
&\quad \to (z(\tfrac{1}{2}) - z(1))\frac{\gamma}{2^\gamma - 1} + \frac{2^\gamma - 1 - \gamma 2^\gamma \log 2}{\gamma(2^\gamma - 1)}\int_{(0,1]} z(t)\nu_{T,\gamma}(\mathrm{d}t),
\end{aligned}$$

that is, (S3) with

$$\mu_{S,\gamma} = \frac{\gamma}{2^\gamma - 1}(\varepsilon_{1/2} - \varepsilon_1) + \frac{2^\gamma - 1 - \gamma 2^\gamma \log 2}{\gamma(2^\gamma - 1)} \nu_{T,\gamma}.$$

Theorem 6.2. *Suppose that the conditions of Theorem 6.1 are satisfied. If* $\hat{\gamma}_n = T(Q_n)$ *and* $\hat{a}(k_n/n) = S(Q_n)$ *with* T *and* S *satisfying* (T0), (T1)′–(T3)′ *and* (S0)–(S3), *respectively, then*

$$k_n^{1/2}(\hat{\gamma}_n - \gamma) \to \mathcal{N}(0, \sigma^2_{T,\gamma}) \tag{50}$$

and

$$k_n^{1/2}\left(\frac{\hat{a}(k_n/n)}{a(k_n/n)} - 1\right) \to \mathcal{N}(0, \sigma^2_{S,\gamma}) \tag{51}$$

weakly with

$$\sigma^2_{T,\gamma} = \int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} \nu_{T,\gamma}(\mathrm{d}s)\nu_{T,\gamma}(\mathrm{d}t),$$

$$\sigma^2_{S,\gamma} = \int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} \mu_{S,\gamma}(\mathrm{d}s)\mu_{S,\gamma}(dt).$$

Suppose, in addition, that condition (15) *holds and*

$$\lim_{n\to\infty} d_n \tilde{R}\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right) = 0 \tag{52}$$

with

$$d_n := k_n^{1/2} \frac{\gamma}{(np_n/k_n)^{-\gamma} - 1} \begin{cases} (\log(k_n/(np_n)))^{-1}, & \text{if } \gamma \geqslant 0, \\ 1, & \text{if } \gamma < 0. \end{cases}$$

Then the estimator $\tilde{x}_{p_n}$ *defined by* (7) *satisfies*

$$\frac{d_n}{a(k_n/n)}(\tilde{x}_{p_n} - x_{p_n}) \to \mathcal{N}(0, \sigma^2_{S,T,\gamma}) \tag{53}$$

where $\sigma^2_{S,T,\gamma} = \sigma^2_{T,\gamma}$ *if* $\gamma > 0$, $\sigma^2_{S,T,\gamma} = \sigma^2_{T,\gamma}/4$ *if* $\gamma = 0$, *and*

$$\begin{aligned}\sigma^2_{S,T,\gamma} = {} & \gamma^2 c(1, 1) - 2\gamma \int_{(0,1]} t^{-(\gamma+1)} c(1, t)\mu_{S,\gamma}(\mathrm{d}t) + 2\int_{(0,1]} t^{-(\gamma+1)} c(1, t)\nu_{T,\gamma}(\mathrm{d}t) \\ & + \int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} c(s, t)\mu_{S,\gamma}(\mathrm{d}s)\mu_{S,\gamma}(\mathrm{d}t) \\ & - \frac{2}{\gamma}\int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} c(s, t)\mu_{S,\gamma}(\mathrm{d}s)\nu_{T,\gamma}(\mathrm{d}t) \\ & + \frac{1}{\gamma^2}\int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} c(s, t)\nu_{T,\gamma}(\mathrm{d}s)\nu_{T,\gamma}(\mathrm{d}t)\end{aligned}$$

if $\gamma < 0$.

Remark 6.3. In view of the proof of Theorem 6.2, condition (52) is a natural generalization of (19).

Remark 6.4. Note that for $\gamma \geqslant 0$ the choice of the estimator for the scale function a does not matter asymptotically. For $\gamma < 0$, though, both the estimators of γ and of a influence the asymptotic behaviour of the quantile estimator, leading to a considerably more complicated expression for the asymptotic variance.

Proof. According to Skorohod's representation theorem there exist versions of Q_n, D_n and e such that the convergence (49) holds almost surely. Let $y_\gamma(t) := (t^{-\gamma} - 1)/\gamma$ and $z_n := k_n^{1/2}((Q_n - D_n)/a(k_n/n) - y_\gamma)$. Since the process e has almost surely continuous sample paths (see Drees 2000), (S1)–(S3) combined with (49) give

$$\begin{aligned}\frac{\hat{a}(k_n/n)}{a(k_n/n)} &= S\left(\frac{Q_n - D_n}{a(k_n/n)}\right) = S(y_\gamma + k_n^{-1/2} z_n)\\ &= 1 + k_n^{-1/2} \int_{(0,1]} t^{-(\gamma+1)} e(t) \mu_{S,\gamma}(\mathrm{d}t) + o(k_n^{-1/2}) \qquad (54)\end{aligned}$$

almost surely, from which (51) is obvious.

Likewise, one can show that

$$k_n^{1/2}(\hat{\gamma}_n - \gamma) \to \int_{(0,1]} t^{-(\gamma+1)} e(t) \nu_{T,\gamma}(\mathrm{d}t) \quad \text{a.s.} \qquad (55)$$

which implies (50) (see proof of Theorem 2.2).

To prove (53), check that

$$\begin{aligned}\frac{\tilde{x}_{p_n} - x_{p_n}}{a(k_n/n)} &= \frac{Q_n(1) - F^{-1}(1 - k_n/n)}{a(k_n/n)} - \left(\frac{(x_{p_n} - F^{-1}(1 - k_n/n)}{a(k_n/n)} - y_\gamma\left(\frac{np_n}{k_n}\right)\right)\\ &\quad + \left(\frac{\hat{a}(k_n/n)}{a(k_n/n)} - 1\right) y_{\hat{\gamma}_n}\left(\frac{np_n}{k_n}\right) + \left(y_{\hat{\gamma}_n}\left(\frac{np_n}{k_n}\right) - y_\gamma\left(\frac{np_n}{k_n}\right)\right)\\ &=: I + II + III + IV.\end{aligned}$$

Theorem 6.1, in combination with Remark 6.1, shows that

$$k_n^{1/2} I \to e(1) \qquad \text{a.s.}$$

By condition (52) we have

$$k_n^{1/2} II = \tilde{R}\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right) = o(y_\gamma(np_n/k_n)) \qquad \text{a.s.}$$

Condition (15) ensures that $y_{\hat{\gamma}_n}(np_n/k_n) = y_\gamma(np_n/k_n)(1 + o(1))$. Hence, in view of (54),

$$k_n^{1/2} III = y_\gamma\left(\frac{np_n}{k_n}\right)\int_{(0,1]} t^{-(\gamma+1)} e(t)\mu_{S,\gamma}(dt)(1+o(1)) \qquad \text{a.s.}$$

where

$$y_\gamma\left(\frac{np_n}{k_n}\right) = (1+o(1))\cdot\begin{cases}((np_n)/k_n)^{-\gamma}/\gamma, & \text{if } \gamma>0,\\ \log(k_n/(np_n)), & \text{if } \gamma=0,\\ -1/\gamma, & \text{if } \gamma<0.\end{cases}$$

Finally, similarly to the proof of Theorem 2.2 and the example given above, a Taylor expansion of $x \mapsto y_x(np_n/k_n)$ at γ in combination with (15) and (55) yields

$$\begin{aligned}k_n^{1/2} IV &= k_n^{1/2}(\hat{\gamma}_n-\gamma)\frac{1}{\gamma^2}\left(1-\left(1+\gamma\log\frac{np_n}{k_n}\right)\left(\frac{np_n}{k_n}\right)^{-\gamma}\right)(1+o(1))\\ &= \int_{(0,1]} t^{-(\gamma+1)} e(t)\nu_{T,\gamma}(\mathrm{d}t)(1+o(1))\cdot\begin{cases}\log(k_n/(np_n))(np_n/k_n)^{-\gamma}/\gamma, & \text{if } \gamma>0,\\ \log^2(k_n/(np_n))/2, & \text{if } \gamma=0,\\ 1/\gamma^2, & \text{if } \gamma<0.\end{cases}\end{aligned}$$

Because, for $\gamma \geqslant 0$, $I + II + III = o(IV)$, assertion (53) follows readily in this case. For $\gamma < 0$ we obtain

$$\frac{d_n}{a(k_n/n)}(\tilde{x}_{p_n} - x_{p_n}) \to -\gamma e(1) + \int_{(0,1]} t^{-(\gamma+1)} e(t)\mu_{S,\gamma}(\mathrm{d}t) - \frac{1}{\gamma}\int_{(0,1]} t^{-(\gamma+1)} e(t)\nu_{T,\gamma}(\mathrm{d}t)$$

from which (53) follows by straightforward calculations. □

Based on Theorem 6.2, one may construct confidence intervals along the lines given in Section 2. In the present situation one uses different estimators of the asymptotic variance depending on the estimated extreme value index $\hat{\gamma}_n$.

For example, if $\gamma < 0$ then one can show by arguments similar to those in the proofs of Theorems 6.2 and 2.3 that, for all $s > 0$,

$$\hat{\gamma}_n^{(i)} := T(Q_{n,i}) = \gamma + \frac{k_n^{1/2}}{i}\int_{(0,1]} t^{-(\gamma+1)} e\left(\frac{i}{k_n}t\right)\nu_{T,\gamma}(\mathrm{d}t) + o_P(k_n^{-1/2})$$

$$\frac{\hat{a}(i/n)}{a(k_n/n)} := \frac{S(Q_{n,i})}{a(k_n/n)} = \left(\frac{i}{k_n}\right)^{-\gamma}\left(1 + \frac{k_n^{1/2}}{i}\int_{(0,1]} t^{-(\gamma+1)} e\left(\frac{i}{k_n}t\right)\mu_{S,\gamma}(\mathrm{d}t) + o_P(k_n^{-1/2})\right)$$

uniformly for $sk_n \leqslant i \leqslant k_n$. From this one may conclude the existence of a sequence $s_n \downarrow 0$ such that

$$\sup_{s_n \leqslant s \leqslant 1}\left| k_n^{1/2}\frac{\tilde{x}_{p_n}^{([k_n s])} - x_{p_n}}{\hat{a}([k_n s]/n)} - Z_{S,T,\gamma}(s)\right| \to 0$$

in probability, with

$$\tilde{x}_{p_n}^{(i)} := X_{n-i:n} + \hat{a}\left(\frac{i}{n}\right)\frac{(np_n/i)^{\hat{\gamma}_n^{(i)}} - 1}{\hat{\gamma}_n^{(i)}}$$

and

$$Z_{S,T,\gamma}(s) := \frac{e(s)}{\gamma} - \frac{1}{\gamma s}\int_{(0,1]} t^{-(\gamma+1)} e(st)\mu_{S,\gamma}(\mathrm{d}t) + \frac{1}{\gamma^2 s}\int_{(0,1]} t^{-(\gamma+1)} e(st)\nu_{t,\gamma}(\mathrm{d}t).$$

In view of the proof of Theorem 2.3, this in turn implies, with $j_n := [k_n s_n] + 1$ and $\tilde{Z}_{S,T,\gamma}(u) := \mathrm{e}^{u/2} Z_{S,T,\gamma}(\mathrm{e}^u)$,

$$\begin{aligned}
\tilde{\sigma}_n^2 &:= \frac{1}{\log(k_n/j_n)}\sum_{i=j_n}^{k_n}\left(\frac{\tilde{x}_{p_n}^{(i)} - \tilde{x}_{p_n}}{\hat{a}(i/n)}\right)^2 \\
&= \frac{1}{\log(k_n/j_n)}\int_{j_n/k_n}^1 \left(Z_{S,T,\gamma}(s) - \frac{\hat{a}(k_n/n)}{\hat{a}(i/n)} Z_{S,T,\gamma}(1)\right)^2 \mathrm{d}s(1 + o_P(1)) \\
&= \frac{1}{\log(k_n/j_n)}\int_{j_n/k_n}^1 (Z_{S,T,\gamma}(s) - s^{-\gamma} Z_{S,T,\gamma}(1))^2\, \mathrm{d}s(1 + o_P(1)) \\
&= \frac{1}{\log(k_n/j_n)}\int_{\log(j_n/k_n)}^0 (\tilde{Z}_{S,T,\gamma}(u) - \mathrm{e}^{(1/2-\gamma)u}\tilde{Z}_{S,T,\gamma}(0))^2\, \mathrm{d}u(1 + o_P(1)) \\
&\to \mathrm{E}\left(\tilde{Z}_{S,T,\gamma}^2(0)\right) = \frac{\sigma_{S,T,\gamma}^2}{\gamma^2}
\end{aligned}$$

because

$$\frac{1}{\log(k_n/j_n)}\mathrm{E}\left(\int_{\log(j_n/k_n)}^0 \left(\mathrm{e}^{(1/2-\gamma)u}\tilde{Z}_{S,T,\gamma}(0)\right)\right)^2 \mathrm{d}u \to 0.$$

Therefore, one may use the asymptotic $(1-\alpha)$ confidence interval

$$[\tilde{x}_{p_n} - k_n^{-1/2}\tilde{\sigma}_n z_{\alpha/2}, \tilde{x}_{p_n} + k_n^{-1/2}\tilde{\sigma}_n z_{\alpha/2}]$$

if $\hat{\gamma}_n < 0$, since $d_n \sim -\gamma k_n^{1/2}$ if $\gamma < 0$. In order to keep the paper to manageable proportions, we do not discuss the case $\gamma \geqslant 0$ and all the ramifications considered in Section 2.

Acknowledgement

The author was partly supported by a DFG Heisenberg grant. He thanks an anonymous referee for valuable remarks that helped to improve the presentation of the results.

References

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge: Cambridge University Press.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes.* Berlin: Springer-Verlag.

Caserta, S., Danielsson, J. and de Vries, C.G. (1998) Abnormal returns, risk and options in large data sets. *Statist. Neerlandica*, **52**, 324–335.

Cramér, H. and Leadbetter, M.R. (1967) *Stationary and Related Processes.* New York: Wiley.

Datta, S. and McCormick, W.P. (1998) Inference for the tail parameters of a linear process with heavy tail innovations. *Ann. Inst. Statist. Math.*, **50**, 337–359.

de Haan, L. and Rootzén, H. (1993) On the estimation of high quantiles. *J. Statist. Plann. Inference*, **35**, 1–13.

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.

Doukhan, P. (1994) *Mixing. Properties and Examples.* New York: Springer-Verlag.

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. (2000) Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.*, **10**, 1274–1301.

Drees, H. (2002) Tail empirical processes under mixing conditions. In H.G. Dehling, T. Mikosch and M. Sørensen (eds), *Empirical Process Techniques for Dependent Data*. Boston: Birkhäuser.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events.* Berlin: Springer-Verlag.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

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

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

Jansen, D.W. and de Vries, C.G. (1991) On the frequency of large stock returns: putting booms and busts into perspective. *Rev. Econom. Statist.*, **73**, 18–24.

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*. Berlin: Springer-Verlag.

Longin, F. (1996) The asymptotic distribution of extreme stock market returns. *J. Business*, **63**, 383–408.

Mikosch, T. and Samorodnitsky, G. (2000) The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Appl. Probab.*, **10**, 1025–1064.

Mikosch, T. and Stărică, C. (2003) Long range dependence effects and ARCH modeling. In G. Oppenheim, M. Taqqu and P. Doukhan (eds), *Theory and Applications of Long-Range Dependence.* Boston: Birkhäuser.

Müller, U.A., Dacorogna, M.M. and Pictet, O.V. (1998) Heavy tails in high-frequency financial data. In R.J. Adler, R.E. Feldman and M.S. Taqqu (eds), *A Practical Guide to Heavy Tails*, pp. 55–77. Boston: Birkhäuser.

Novak, S.Y. (2002) Inference on heavy tails from dependent data. *Siberian Adv. Math.*, **12**(2), 73–96.

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

Pratt, J.W. (1960) On interchanging limits and integrals. *Ann. Math. Statist.*, **31**, 74–77.
Reiss, R.-D. and Thomas, M. (2001) *Statistical Analysis of Extreme Values* (2nd edn). Basel: Birkhäuser.
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., Leadbetter, M.R. and de Haan, L. (1992) Tail and quantile estimators for strongly mixing stationary processes. Report, Department of Statistics, University of North Carolina.
Shao, Q.-M. (1995) Maximal inequalities for partial sums of ρ-mixing sequences. *Ann. Probab.*, **23**, 948–965.
Smith, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.
Stărică, C. (1998) On the tail empirical process of solutions of stochastic difference equations. Preprint, Chalmers University Gothenburg. Available at www.math.chalmers.se/~starica/publi1.html.
Stărică, C. and Granger, C. (2001) Non-stationarities in stock returns. Preprint, Chalmers University Gothenburg.
Vervaat, W. (1979) On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Probab.*, **11**, 750–783.

Received September 2001 and revised August 2002