Ann. Inst. Statist. Math.
Vol. 50, No. 2, 337–359 (1998)

INFERENCE FOR THE TAIL PARAMETERS OF A LINEAR PROCESS WITH HEAVY TAIL INNOVATIONS

SOMNATH DATTA AND WILLIAM P. MCCORMICK

Department of Statistics, Franklin College of Arts and Sciences, University of Georgia, 204 Statistics Building, Athens, GA 30602-1952, U.S.A.

(Received December 5, 1995; revised January 6, 1997)

Abstract. Consider a linear process $X_t = \sum_{i=0}^{\infty} c_i Z_{t-i}$, where the innovations Z's are i.i.d. satisfying a standard tail regularity and balance condition, viz., $P(Z > z) \sim rz^{-\alpha}L_1(z)$, $P(Z < -z) \sim sz^{-\alpha}L_1(z)$, as $z \to \infty$, where $r + s = 1$, $r, s \geq 0$, $\alpha > 0$ and L_1 is a slowly varying function. It turns out that in this setup, $P(X > x) \sim px^{-\alpha}L(x)$, $P(X < -x) \sim qx^{-\alpha}L(x)$, as $x \to \infty$, where α is the same as above, p is a convex combination of r and s, $p + q = 1$, $p, q \geq 0$ and $L = \|\underline{c}\|_\alpha^\alpha L_1$ where $\|\underline{c}\|_\alpha = (\Sigma|c_i|^\alpha)^{1/\alpha}$. The quantities α and $\beta = 2p - 1$ can be regarded as tail parameters of the marginal distribution of X_t. We estimate α and β based on a finite realization $X_1, \ldots, X_n$ of the time series. Consistency and asymptotic normality of the estimators are established. As a further application, we estimate a tail probability under the marginal distribution of the X_t. A small simulation study is included to indicate the finite sample behavior of the estimators.

Key words and phrases: Linear processes, heavy tailed distribution, tail parameters, tail probability.

1. Introduction and summary

Consider a linear process

$$X_t = \sum_{i=0}^{\infty} c_i Z_{t-i}, \tag{1.1}$$

$c_0 = 1$, where the innovations are i.i.d. satisfying the usual assumptions of heavy tailed modeling:

$$P(Z_1 > z) \sim rz^{-\alpha}L_1(z), \qquad \text{and} \qquad P(Z_1 < -z) \sim sz^{-\alpha}L_1(z), \tag{1.2}$$

as $z \to \infty$, where $\alpha > 0$, $r, s \geq 0$, $r + s = 1$, and L_1 is a slowly varying function at ∞ (i.e., $L_1(cz)/L_1(z) \to 1$ as $z \to \infty$, for all $0 < c < \infty$). Here, the notation $a(z) \sim b(z)$, as $z \to \infty$, is used to denote the fact that $a(z)/b(z) \to 1$, as $z \to \infty$.

The sequence $\{c_i\}$ of reals satisfies certain mild summability conditions to be specified later. It then turns out (see Lemma 5.2) that the (common) marginal distribution of X_t also satisfies analogues of (1.2), i.e.,

$$(1.3)\qquad P(X_1 > x) \sim px^{-\alpha}L(x) \quad \text{and} \quad P(X_1 < -x) \sim qx^{-\alpha}L(x),$$

as $x \to \infty$, where $\alpha > 0$ is the same as in (1.2), $p, q \geq 0$, $p+q = 1$ and $L = \|\underline{c}\|_\alpha^\alpha L_1$ is another slowly varying function.

The quantity α is known as the index of regular variation and p can be viewed as a tail-balance parameter. Let $\beta = 2p - 1$. Note that $\beta = 0$ iff $p = q = \frac{1}{2}$ in which case the tails $P(X_1 < -x)$ and $P(X_1 > x)$ are asymptotically the same. Therefore β can be regarded as an asymmetry parameter.

The main problem we consider in this paper is the estimation of the tail parameters α and β of the (common) marginal distribution of the X's based on a finite realization of the series $X_1, \ldots, X_n$. When the X's are i.i.d. (which corresponds to $c_i = 0$ for $i > 0$), Hill (1975) proposed an estimator of α using the order statistics. Variations and extensions of Hill's estimator for the i.i.d. case were later considered by de Haan and Resnick (1980) and Csörgo *et al.* (1985). Hahn and Weiner (1991) considered estimation of both α and β and the joint asymptotic normality of the estimators in the i.i.d. case. Asymptotic properties of the Hill's estimator of α for a stationary strongly mixing sequence was recently studied by Rootzén *et al.* (1990). Moreover, a point process approach to prove consistency of the Hill's estimator for dependent data was recently obtained by Resnick and Starica (1995). In the i.i.d. case, new estimators of α and β based on the empirical c.d.f. were recently proposed by Athreya *et al.* (1992) as simpler alternatives to earlier estimators mentioned above. In this paper we introduce an estimator of α based on the empirical c.d.f. which has a least square interpretation and generalizes the estimator of Athreya *et al.* (1992). Consistency and joint asymptotic normality of the estimators of α and β are established in the linear process setup (1.1). Estimation of their joint covariance matrix is also considered. As an application, we estimate a tail probability using the estimates of α and β. Although we feel that our results can be extended to the more general stationary strongly mixing sequence (eg., the setup of Rootzén *et al.* (1990)) at the expense of more abstract sufficient conditions, we do not pursue it here because linear processes form a rich class of stationary time series models. Yet the setup is simple enough so that relatively straightforward sufficient conditions for our results can be obtained. In particular, it is possible to formulate sufficient conditions in terms of the innovation distribution (i.e., Z_t) and the coefficients $\{c_i\}$. In addition, an explicit expression of the asymptotic variances and covariances of our estimators can be obtained in this situation.

The rest of the paper is organized as follows. The estimation of α, β and the asymptotic properties of the estimators are presented in Section 2. In Section 3 we consider the estimation of a tail probability under the (common) marginal distribution of X_t using the estimators of α and β. In Section 4, a small simulation study is reported to indicate the finite sample behavior of the estimators. The simulation results do conform with the asymptotic results of Section 2. The proofs of Sections 2 and 3 results are collectively presented in Section 5.

2. Estimation of the tail parameters

Consider a linear process $\{X_t\}$ defined by (1.1) where the innovations Z_t's are i.i.d. satisfying (1.2). Throughout this paper, assume the following mild summability condition on the coefficients $\{c_i\}$.

$$\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty, \qquad \text{for some} \quad 0 < \delta < \alpha \wedge 1. \tag{2.1)(C.1}$$

Let F denote the (common) distribution function of X_t defined by (1.1). Also throughout this paper, let $G(x) = 1 - F(x) + F(-x)$. For notational convenience we will assume that F is continuous so that $P(|X| > x) = 1 - F(x) + F((-x)^-) = G(x)$. At the expense of keeping track of left limits all our results go through without this assumption. Clearly, (1.3) implies

$$\frac{1 - F(x) - F(-x)}{G(x)} \to \beta \quad \text{and} \quad \frac{G(Tx)}{G(x)} \to T^{-\alpha}, \tag{2.2}$$

as $x \to \infty$, where $0 < T < \infty$. One can obtain a set of natural estimators of α and β by replacing F by the empirical c.d.f. F_n, x by a sequence $x_n \to \infty$, in (2.2), and then solving the resulting equations (obtained by replacing the limits with equalities). This approach leads to the estimators

$$\hat{\alpha} = \hat{\alpha}(T) = -\left(\frac{1}{\log T}\right) \log\left(\frac{G_n(Tx_n)}{G_n(x_n)}\right) \tag{2.3}$$

and

$$\hat{\beta} = \frac{1 - F_n(x_n) - F_n((-x_n)^-)}{G_n(x_n)}$$

where $x_n \to \infty$, $F_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \leq x)$, $G_n(x) = 1 - F_n(x) + F_n((-x)^-)$, and $T \in (0,1) \cup (1,\infty)$. The above estimators were recently proposed and their consistency and asymptotic normality were established for the case when X_t's are i.i.d. by Athreya *et al.* (1992).

The Athreya *et al.* (1992) estimator of α given in (2.3) can be generalized as follows. For a positive integer L, select L values of T, namely, $T_1, \ldots, T_L \in (0,1) \cup (1,\infty)$. Arguing as before we see that

$$\log G_n(T_i x_n) \approx -\alpha \log T_i + \log G_n(x_n), \qquad i = 1, \ldots, L.$$

Thus α can be estimated by the method of least squares yielding

$$\hat{\alpha}_{LS} = \frac{-\sum_{i=1}^{L} \log T_i \log(G_n(T_i x_n)/G_n(x_n))}{\sum_{i=1}^{L} (\log T_i)^2} = \frac{\sum_{i=1}^{L} \hat{\alpha}(T_i)(\log T_i)^2}{\sum_{i=1}^{L} (\log T_i)^2} \tag{2.4}$$

where $\hat{\alpha}(T_i)$ denotes the estimator $\hat{\alpha}$ in (2.3) with $T = T_i$, $1 \leq i \leq L$. The estimator $\hat{\alpha}_{LS}$ is expected to be more stable w.r.t. the choice of T's and the level x_n.

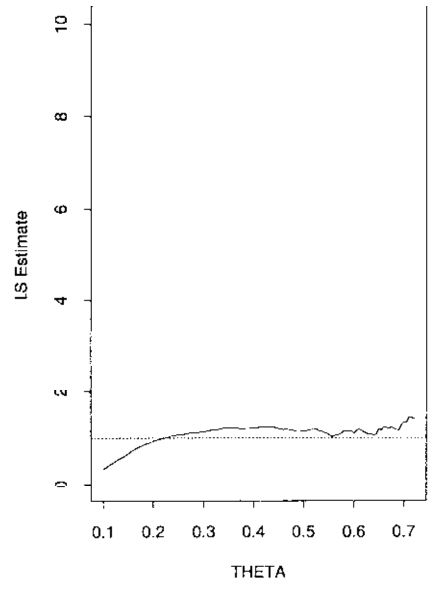

Fig. 2.1

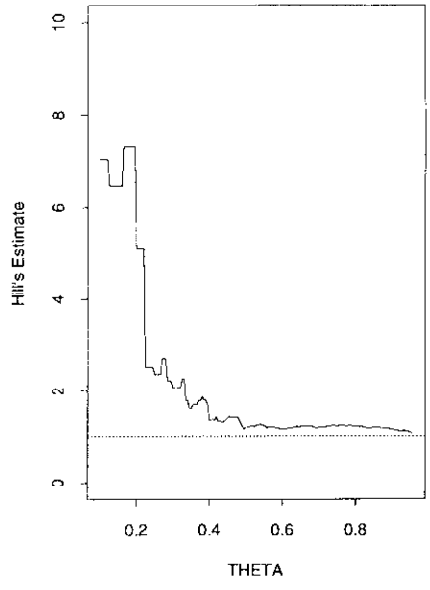

Fig. 2.2

Fig. 2.1. Plot of the LS estimate based on a sample of size 2000 from $X_t = Z_t + 0.5Z_{t-1}$, with Pareto Z and level $x_n = n^\theta$, $T_i = 0.1i + 1$, $1 \le i \le 10$, $T_i = (0.1(i-10)+1)^{-1}$, $11 \le i \le 20$. The dotted line denotes the true value of α.

Fig. 2.2. Plot of the Hill's estimate based on a sample of size 2000 from $X_t = Z_t + 0.5Z_{t-1}$, with Pareto Z and number of order statistics $c_n = [n^\theta]$. The dotted line denotes the true value of α.

It turns out, under regularity conditions, the convergence rate for the above estimators is $Op(1/\sqrt{nG(x_n)})$, which is typically comparable to Hill type estimators. However, since $\hat{\alpha}_{LS}$ is based on smoothing in the estimator space, we can expect it to be relatively stable with respect to the choice of the level x_n. On the other hand, Hill's estimator is known to be very sensitive to the choice of the number of order statistics used. To illustrate this, we calculate $\hat{\alpha}_{LS}$ with $x_n = n^\theta$ and the Hill's estimator of α with c_n, the number of order statistics used $= [n^\theta]$, respectively, for a simulated sample of size $n = 2000$ from a MA(1) process $X_t = Z_t + 0.5Z_{t-1}$ with standard Pareto innovations corresponding to $\alpha = 1$. The T used in $\hat{\alpha}_{LS}$ were $T_i = 0.1i + 1$, $1 \le i \le 10$, and $= (0.1(i-10)+1)^{-1}$, for $11 \le i \le 30$. Recall that the Hill's estimate of α is given by $\hat{\alpha} = (c_n - 1)^{-1} \sum_{i=1}^{c_n} (Y_{(i)}^n - Y_{(c_n)}^n)$, where $Y_{(n)}^n < \cdots < Y_{(1)}^n$ are the ordered $\log |X_i|$, $i = 1, \ldots, n$. The graphs of the estimate versus θ for the two estimators are given in Figs. 2.1 and 2.2, respectively. Clearly the graph of $\hat{\alpha}_{LS}$ exhibits less fluctuations in the usable range of θ values, indicating a practical advantage.

2.1 *Consistency*

We first address the issue of consistency of the estimators $\hat{\alpha}_{LS}$ and $\hat{\beta}$ in the linear process setup of this paper.

Since $\hat{\alpha}_{LS}$ is a convex combination of $\hat{\alpha}(T_i)$, it is enough to prove that each $\hat{\alpha}(T_i)$ is consistent. In other words, for the proof, we will assume $L = 1$ and $\hat{\alpha} = \hat{\alpha}(T)$, for a $T \in (0, 1) \cup (1, \infty)$.

The following mild summability condition will be needed for this purpose.

$$\sum_{i=1}^{\infty} \sum_{j=0}^{\infty} (|c_j| \wedge |c_{i+j}|)^{\delta} < \infty \tag{C.2}$$

for some $0 < \delta < \alpha$.

It is not hard to check that (C.2) holds if (C.1) is satisfied for some $0 < \delta < (\alpha \wedge 1)/2$ and for some positive integer m,

$$|c_j| \geq \bigwedge_{l=1}^{m} |c_{j+l}|, \qquad j \geq 0.$$

In particular, both (C.1) and (C.2) hold for ARMA processes.

THEOREM 2.1. *Let $x_n \to \infty$ be such that*

$$nP(|Z_1| > x_n) \to \infty. \tag{2.5}$$

Then under conditions (C.1) *and* (C.2), $\hat{\alpha} \xrightarrow{P} \alpha$ *and* $\hat{\beta} \xrightarrow{P} \beta$, *as* $n \to \infty$.

Note that in (2.5) the tail of Z_1 can be replaced by the tail of X_1, $P(|X_1| > x_n)$ (by Lemma 5.2). Although this condition involves the distribution of Z_1 (or X_1), which is unknown, such a level x_n can be chosen only on the basis of partial information about the tail. For example if we know an upper bound α^* on α (i.e., $\alpha^* > \alpha$) then one may choose $x_n = n^{1/\alpha^*}$. It is straightforward to check using (1.2) that (2.5) holds for this x_n. In particular, for infinite variance modeling one assumes that $0 < \alpha < 2$, in which case one may use $x_n = \sqrt{n}$.

2.2 *An adaptive choice*

We now consider a data based selection of the level x_n in the definition of estimators $\hat{\alpha}$ and $\hat{\beta}$. The advantage of this approach is that one does not need to have any knowledge of α. Several such choices are possible. In particular one can take $x_n = \sqrt{M_n}$, where $M_n = \max(|X_1|, \ldots, |X_n|)$. Denote the resulting estimators by $\hat{\alpha}^*$ and $\hat{\beta}^*$. In this paper, we only prove the consistency of these adaptive versions.

THEOREM 2.2. *Under the same conditions as in Theorem* 2.1, $\hat{\alpha}^* \xrightarrow{P} \alpha$ *and* $\hat{\beta}^* \xrightarrow{P} \beta$ *as* $n \to \infty$.

2.3 *Asymptotic normality*

Once again, for the simplicity of presentation we let $L = 1$, which means $\hat{\alpha}_{LS} = \hat{\alpha}(T)$ for a $T \in (0,1) \cup (1,0)$. The asymptotic normality of a linear combination can be proved exactly along the same line of reasoning. As a first step to proving asymptotic normality, we rewrite the estimators in (2.3) in a more propitious form for a central limit theorem. Define sequences of reals $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ by

$$(2.6) \qquad \tilde{\alpha}_n = \left(-\frac{1}{\log T}\right) \log\left(\frac{G(Tx_n)}{G(x_n)}\right) \quad \text{and} \quad \tilde{\beta}_n = \frac{1 - F(x_n) - F(-x_n)}{G(x_n)}, \quad n \geq 1.$$

Note $\tilde{\alpha}_n \to \alpha$ and $\tilde{\beta}_n \to \beta$ as $n \to \infty$. Furthermore, let

$$(2.7) \qquad v_n = \frac{G(Tx_n)}{G(x_n)} \quad \text{and} \quad \hat{v}_n = \frac{G_n(Tx_n)}{G_n(x_n)}, \quad n \geq 1.$$

Now set $U_n = \hat{v}_n / v_n$. Finally define variables $\xi_{nj} = I[|X_j| > Tx_n] - v_n I[|X_j| > x_n]$ and $\delta_{nj} = (1 - \tilde{\beta}_n) I[X_j > x_n] - (1 + \tilde{\beta}_n) I[X_j < -x_n]$. Next observe that

$$(2.8) \qquad \hat{\alpha} - \tilde{\alpha}_n = \left(\frac{-1}{\log T}\right) \log U_n = \left(\frac{-1}{\log T}\right) \log\left(1 + \frac{1}{n v_n G_n(x_n)} \sum_{j=1}^{n} \xi_{nj}\right)$$

and $\hat{\beta} - \tilde{\beta}_n = (\sum_{j=1}^n \delta_{nj})/(nG_n(x))$. From the last two relations, it is clear that we will need a central limit theorem for $(\sum_{j=1}^n \xi_{nj}, \sum_{j=1}^n \delta_{nj})$. To achieve such a result, we require certain additional conditions on the sequence of coefficients $\{c_i\}$ and the innovation distribution.

(C.3) The characteristic function ϕ_Z of Z_1 satisfies

$$\int_{-\infty}^{\infty} |u \phi_Z(u)| du < \infty.$$

(C.4) For some $0 < \tau < \frac{\alpha \wedge 2}{q + \alpha \wedge 2}$, $q \geq 1$,

$$\sum_{i=v}^{\infty} i |c_i|^{\tau} = O(v^{-\theta}),$$

as $v \to \infty$, for some $\theta > 2$.

Note that (C.4) implies (C.1).

THEOREM 2.3. *Assume conditions* (C.2) (C.4). *In addition assume that the levels* $x_n \to \infty$ *satisfy*

$$(2.9) \qquad n^{q(\theta-1)/(q(\theta+1)+2\theta)} P(|Z_1| > x_n) \to \infty,$$

as $n \to \infty$, *where* q *and* θ *are as in* (C.4).

Then

$$\sqrt{nG(x_n)}(\hat{\alpha} - \tilde{\alpha}_n, \hat{\beta} - \tilde{\beta}_n)^T \xrightarrow{d} N_2(\mathbf{0}, \Sigma), \qquad as \quad n \to \infty, \tag{2.10}$$

where the asymptotic dispersion matrix is given by

$$\Sigma = \begin{pmatrix} \lambda_{11}T^{2\alpha}/(\log T)^2 & -\lambda_{12}T^{\alpha}/\log T \\ -\lambda_{12}T^{\alpha}/\log T & \lambda_{22} \end{pmatrix} \tag{2.11}$$

where the λ_{ij} *are given in* (5.25)–(5.27).

Moreover if

$$\sqrt{nG(x_n)}(\tilde{\alpha}_n - \alpha) = o(1) \quad and \quad \sqrt{nG(x_n)}(\tilde{\beta}_n - \beta) = o(1) \qquad as \quad n \to \infty, \tag{2.12}$$

then

$$\sqrt{nG(x_n)}(\hat{\alpha} - \alpha, \hat{\beta} - \beta)^T \xrightarrow{d} N_2(\mathbf{0}, \Sigma), \qquad as \quad n \to \infty. \tag{2.13}$$

Remark 2.1. It is not difficult to check that $x_n = n^{-1/\alpha^*}$ satisfies (2.9) for $0 < \alpha/\alpha^* < q(\theta - 1)(q(\theta + 1) + 2\theta)^{-1}$. Note that it means that if θ and q are large then we have more choices for α^*, which sounds reasonable. In particular if the coefficients $\{c_i\}$ decrease exponentially then it holds for any $\alpha^* > \alpha$.

Remark 2.2. One can obtain simple sufficient conditions for (2.12), which says that the bias is asymptotically negligible, provided one is willing to assume slightly more than the basic tail regularity condition (1.2). These are described in Athreya *et al.* (1992). For example, suppose $G(x) = x^{-\alpha}(A + O(x^{-\gamma}))$ (see Hall (1982)), for constants A and $\gamma > 0$. Then the first statement of (2.12) holds for the choice $x_n = n^{1/\alpha^*}$ with $\alpha^* < \alpha + 2\gamma$. If on the other hand one is willing to assume a model like $G(x) = x^{-\alpha}A(\log x)^{\gamma}$, then (2.12) for $\tilde{\alpha}$ will be satisfied for levels like $x_n = n^{1/\alpha}(\log n)^{\delta}$, with $(\gamma - 2)/\alpha < \delta < 2\gamma/\alpha$, provided $\gamma > -2$. Note that by Lemma 5.2, the tail of X_1 inherits the tail behavior of Z, and it is not hard to see that (2.12) for $\tilde{\alpha}$ holds under these conditions.

Condition (2.12) for $\tilde{\beta}$ is related to higher order symmetry of the tails of X_1. For example, if one writes $1 - F(x) = px^{-\alpha}\|\underline{c}\|_{\alpha}^{\alpha}L_1(x)(1 + h_1(x))$ and $F(x) = qx^{-\alpha}\|\underline{c}\|_{\alpha}^{\alpha}L_1(x)(1 + h_2(x))$, where $\{c_i\}$ and L_1 are as in (1.1) then $h_1(x) \to 0$ and $h_2(x) \to 0$, as $x \to \infty$. If for $x \to \infty$, the difference $|h_1(x) - h_2(x)|$ between the two h functions is $O(x^{-\delta})$ for some $\delta > 0$ then $x_n = n^{1/\alpha^*}$, with $\alpha^* < \alpha + 2\delta$, is a level sequence satisfying (2.12) for $\tilde{\beta}$.

2.4 *Estimation of the dispersion matrix*

We first remark that the model dependent quantities λ_{ij} appearing in the limiting dispersion matrix may be estimated by virtue of equations (5.25)–(5.27) by estimating the linear process coefficients. Such an approach would be natural for ARMA models. More generally, as noted in the proof of Theorem 2.3, the λ appear as limiting variance—covariance values for $\frac{1}{\sqrt{nG(x_n)}}(\sum_{i=1}^k \zeta_{ni}, \sum_{i=1}^k \eta_{ni})$. Here $k = k_n = o(n)$, $k_n \uparrow \infty$ is an appropriately chosen integer sequence and $\zeta_{ni} = \sum_{j=1}^r \xi_{n,(i-1)r+j}$, $\eta_{ni} = \sum_{j=1}^r \delta_{n,(i-1)r+j}$, $1 \le i \le k$, $r = [n/k]$. Let $\hat{\xi}_{nj} = I[|X_j| > Tx_n] - \hat{v}_n I[|X_j| > x_n]$ and $\hat{\delta}_{nj} = (1-\hat{\beta}_n)I[X_j > x_n] - (1+\hat{\beta}_n)I[X_j < -x_n]$. Set $\hat{\zeta}_{ni} = \sum_{j=1}^r \hat{\xi}_{n,(i-1)r+j}$ and $\hat{\eta}_{ni} = \sum_{j=1}^r \hat{\delta}_{n(i-1)r+j}$, $i = 1, \dots, k$. Since the ζ_{ni} are approximately i.i.d., a natural estimate for λ_{11} is given by

$$\hat{\lambda}_{11} = \frac{1}{nG_n(x_n)} \sum_{i=1}^k (\hat{\zeta}_{ni} - \bar{\zeta}_n)^2 \tag{2.14}$$

where $\bar{\zeta}_n = \frac{1}{k}\sum_{i=1}^k \hat{\zeta}_{ni}$. Similarly natural estimators for λ_{12} and λ_{22} are given by

$$\hat{\lambda}_{12} = \frac{1}{nG_n(x_n)} \sum_{i=1}^k (\hat{\zeta}_{ni} - \bar{\zeta}_n)\hat{\eta}_{ni} \tag{2.15}$$

and

$$\hat{\lambda}_{22} = \frac{1}{nG_n(x_n)} \sum_{i=1}^k (\hat{\eta}_{ni} - \bar{\eta}_n)^2 \tag{2.16}$$

where $\bar{\eta}_n = \frac{1}{k}\sum_{i=1}^k \hat{\eta}_{ni}$. Using these estimators, one then obtains a plug-in estimator for Σ given by

$$\hat{\Sigma} = \begin{pmatrix} \hat{\lambda}_{11} T^{2\hat{\alpha}}/(\log T)^2 & -\hat{\lambda}_{12} T^{\hat{\alpha}}/\log T \\ -\hat{\lambda}_{12} T^{\hat{\alpha}}/\log T & \hat{\lambda}_{22} \end{pmatrix}. \tag{2.17}$$

The estimates for the asymptotic variance-covariance of $\hat{\alpha} = \hat{\alpha}_{LS}$ and $\hat{\beta}$ can be constructed in the same way where the $\hat{\xi}_{nj}$ is replaced by the appropriate convex combination of the $\hat{\xi}_{nj}$'s for various T_i, $i = 1, \dots, L$.

THEOREM 2.4. *Suppose $n^\kappa P(|Z_1| > x_n) \to \infty$, for a $0 < \kappa < q(\theta - 1)/((1+q)(1+\theta))$. Under conditions* (C.2)–(C.4), *for $k_n = n^a$ with $\frac{1}{2}(\frac{(1+q)}{q}\kappa + 1) \le a < \theta/(1+\theta)$, we obtain $\hat{\Sigma}_n \xrightarrow{P} \Sigma$ as $n \to \infty$.*

COROLLARY 2.1. *Under the assumptions of Theorem* 2.3 *and Theorem* 2.4 *and* (2.12), *we have with $k = n^a$, for $\frac{1}{2}(\frac{(1+q)}{q}\kappa + 1) \le a < \theta/(1+\theta)$,*

$$\sqrt{nG_n(x_n)}\hat{\Sigma}^{-1/2}(\hat{\alpha} - \alpha, \hat{\beta} - \beta)^T \xrightarrow{d} N(\mathbf{0}, I) \quad \text{as} \quad n \to \infty,$$

where $\hat{\Sigma}$ given by (2.17) *is non-singular and positive definite with probability tending to one.*

Note that one can construct a confidence set for (α, β) via Corollary 2.1.

3. Tail probability estimation

We now use our estimators of the tail parameters α and β to estimate the right tail probability $P(X_1 > u)$ under the (common) marginal distribution of X_t, where $u = u_n \to \infty$. The usual empirical tail probability is not very suitable because the level u is large. However, we can estimate the tail probability by the empirical value when the level is smaller, say u/T for some $T > 1$, and use the regular variation property (2.2) to estimate the required tail probability. Letting $T = u_n/x_n$, this approach leads to the estimator

$$\hat{P}(X_1 > u_n) = \hat{p}(u_n/x_n)^{-\hat{\alpha}} G_n(x_n), \tag{3.1}$$

where x_n is as in (2.5), $\hat{p} = (1+\hat{\beta})/2$. Similarly, with $\hat{q} = (1-\hat{\beta})/2$ the left and the two sided tail probabilities are estimated as

$$\hat{P}(X_1 < -u_n) = \hat{q}(u_n/x_n)^{-\hat{\alpha}} G_n(x_n) \quad \text{and} \tag{3.2}$$

$$\hat{P}(|X_1| > u_n) = (u_n/x_n)^{-\hat{\alpha}} G_n(x_n). \tag{3.3}$$

THEOREM 3.1. *Let $\hat{\alpha} = \hat{\alpha}(T_1)$. Suppose all the conditions of Theorem 2.3 hold. Let $u = u_n = Tx_n$, with $T \in (1, \infty)$, where x_n is as in the statement of Theorem 2.3. Also assume that*

$$\sqrt{nG(x_n)}\{P(X_1 > Tx_n)/P(|X_1| > x_n) - pT^{-\alpha}\} = o(1). \tag{3.4}$$

Then

$$\sqrt{\frac{n}{G(x_n)}}\{\hat{P}(X_1 > u_n) - P(X_1 > u_n)\} \xrightarrow{d} N(0, \sigma_T^2)$$

with

$$\sigma_T^2 = (p, T^{-\alpha}/2, pT^{-\alpha})A(p, T^{-\alpha}/2, pT^{-\alpha})^T,$$

where $A = ((a_{ij}))$ is the matrix given by $a_{ij} = \lambda_{ij}$, $1 \le i, j \le 2$, with $T = T_1$, $a_{13} = (T_1 \vee 1)^{-\alpha} - T_1^{-\alpha} + \|\underline{c}\|_\alpha^{-\alpha} \sum_{k,l}\{(|c_k| \wedge T_1^{-1}|c_l|)^\alpha - T_1^{-\alpha}(|c_k| \wedge |c_l|)^\alpha\}$, $a_{23} = \|\underline{c}\|_\alpha^{-\alpha} \sum_{k \ne l}((r \quad s)\operatorname{sgn}(c_k) - \beta)(|c_k| \wedge |c_l|)^\alpha$, $a_{33} = \|\underline{c}\|_\alpha^{-\alpha} \sum_{k,l}(|c_k| \wedge |c_l|)^\alpha$.

Remark 3.1. If p is known to be one (e.g., when $X_t \ge 0$), the estimator in (3.1) should be adjusted by using $\hat{p} = 1$. The asymptotic normality of the resulting estimator can be proved in a similar manner.

Remark 3.2. Suppose (cf. Remark 2.2) $G(x) = x^{-\alpha}(A + O(x^{-\gamma_1}))$, for positive constants A and γ_1. Write $1 - F(x) = pG(x)(1 + r(x))$, where $r(x) \to 0$, as $x \to \infty$. Suppose $|r(x)| = O(x^{-\gamma_2})$. Then (3.4) holds with $x_n = n^{1/\alpha^*}$ for $\alpha^* < x + 2\gamma$, $\gamma = \gamma_1 \wedge \gamma_2$.

4. A simulation study

We now report the results of a simulation study which demonstrates the asymptotic normality of the tail parameter estimators. First we generate 5000 independent samples of size n each from the MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, $t \geq 1$, where the Z's are standard Pareto for which $1 - F_Z(x) = x^{-1}$ for $x \geq 1$. In this case $\alpha = 1$. For simplicity throughout we use $L = 1$, i.e., $\hat{\alpha} = \hat{\alpha}(T_1)$. We use a moderate value of T_1, namely $T_1 = 2$ which happens to be the minimizer of the asymptotic variance of $\hat{\alpha}(T_1)$ on $(1, \infty)$. Also $x_n = \sqrt{n}$ is used throughout.

Figures 4.1–2 show the normal $Q - Q$ plots of 5000 standardized $\hat{\alpha}$, i.e., $\sqrt{nG(x_n)}(\hat{\alpha} - 1)/\sqrt{\sigma_{11}}$, for $n = 500$ and 4000, respectively. Figure 4.3 shows a similar plot for the sample size $n = 8000$ where a data based studentization $\sqrt{nG_n(x_n)}(\hat{\alpha} - 1)/\sqrt{\hat{\sigma}_{11}}$ is used. Also see Table 4.1, where the bias, the variance, the skewness (based on the third moment) and the kurtosis (based on the fourth moment) of the sampling distribution of $\sqrt{nG(x_n)}\hat{\alpha}$ are empirically calculated for sample sizes $n = 500$, 4000 and 8000.

The normal approximation is somewhat unsatisfactory at $n = 500$ (specially, in the left tail). The sampling distribution has a noticeable skewness. The normal approximation is effective at $n = 4000$, and at $n = 8000$ it is very effective even with a data based normalization. The block size used in the construction of $\hat{\sigma}_{11}$ is $r_n = 80$. In order to see how the choice of the block size affects the estimation of σ_{11} and in turn the normal approximation, different block sizes were used and the 95th and the 97.5th percentiles of the sampling distributions were calculated in each case by Monte Carlo based on 5000 replications. The sample size used is $n = 8000$. More specifically we use $r_n = 40$, 50, 60, 70, 80 and 90. The two special percentiles are plotted against r in Fig. 4.4. Circles and dots are used for the 95th and the 97.5th percentiles, respectively. The corresponding standard normal percentiles are represented by the two horizontal lines. As can be seen from the picture, the normal approximation is relatively insensitive w.r.t. the choice of r in the range 40 to 80 and is reasonably accurate. For $r = 90$ in which case the blocksize is almost equal to the number of blocks ($= [8000/90] = 89$), the percentiles of the sampling distribution drop below the normal percentiles.

Note that it is expected that a large sample size will be necessary for the normal approximation to be fully effective because the estimation is based on indicators of events with small probabilities. Also the dependence in the data makes additional contribution to the skewness. A similar phenomenon was noticed for the Hill's estimator as well. See Rootzén *et al.* (1990).

Next we consider the same model as before except this time the innovations (errors) Z's are generated from a two sided Pareto with density $f_Z(x) = (2x^2)^{-1}1_{(|x|>1)}$. Note that for this example $\alpha = 1$ and $\beta = 0$. This corresponds to a non-degenerate case in β. Once again we use 5000 replications of size $n = 8000$ each and calculate the studentized $\hat{\beta}(= \sqrt{nG_n(x_n)}(\hat{\beta} - 0)/\sqrt{\hat{\sigma}_{22}})$. The normal approximation appears to be very good as illustrated by Fig. 4.5. The joint asymptotic normality of $\hat{\alpha}$ and $\hat{\beta}$ is evidenced in the 2-dimensional histogram in Fig. 4.6 for the 5000 values of $(\sqrt{nG(x_n)}(\hat{\alpha} - 1), \sqrt{nG(x_n)}(\hat{\beta} - 0))$. Here the correlation (empirically calculated) was $r_{\hat{\alpha},\hat{\beta}} = -0.014$.

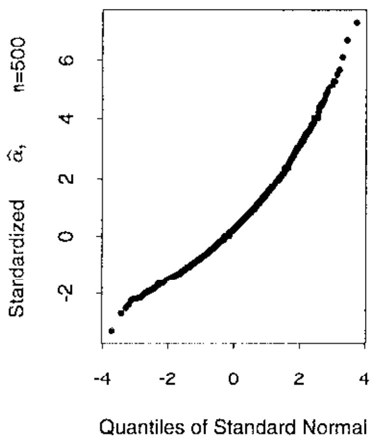

Fig. 4.1

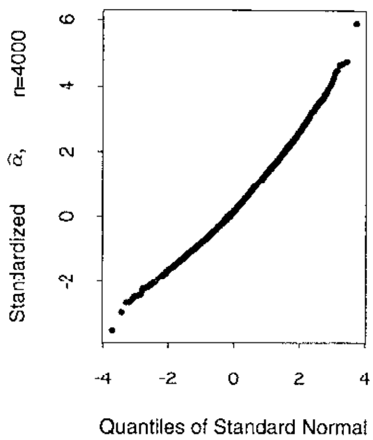

Fig. 4.2

Fig. 4.1. Normal $Q-Q$ plot of 5000 replications of standardized $\hat{\alpha}$ each based on samples of size $n = 500$ from an MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, where the Z are standard Pareto.

Fig. 4.2. Normal $Q-Q$ plot of 5000 replications of standardized $\hat{\alpha}$ each based on samples of size $n = 4000$ from an MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, where the Z are standard Pareto.

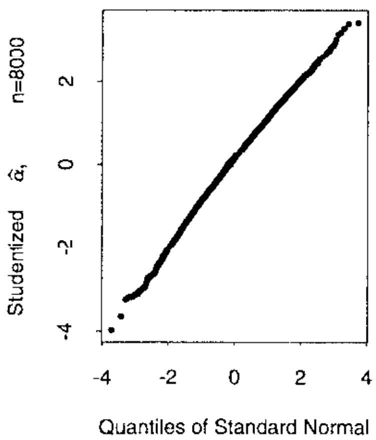

Fig. 4.3

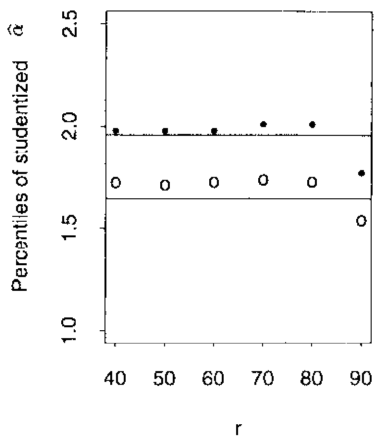

Fig. 4.4

Fig. 4.3. Normal $Q-Q$ plot of 5000 replications of studentized $\hat{\alpha}$ each based on samples of size $n = 8000$ from an MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, where the Z are standard Pareto.

Fig. 4.4 The 95-th (circles) and the 97.5-th (dots) percentiles of 5000 replications of studentized $\hat{\alpha}$ for different values of r each based on samples of size $n = 8000$ from an MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, where the Z are standard Pareto. The straight lines correspond to these percentiles for the standard normal.

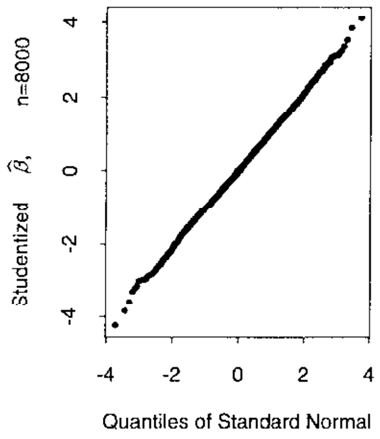

Fig. 4.5

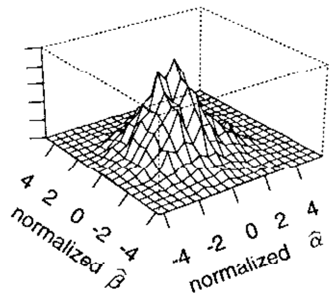

Fig. 4.6

Fig. 4.5. Normal $Q-Q$ plot of 5000 replications of studentized $\hat{\beta}$ each based on samples of size $n = 8000$ from an MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, where the Z are two sided Pareto.

Fig. 4.6. A 2-dimensional histogram of 5000 replications of normalized $(\hat{\alpha}, \hat{\beta})$ each based on samples of size $n = 8000$ from an MA(1) process $X_t = Z_t + 0.5Z_{t-1}$, where the Z are two sided Pareto.

Table 4.1. Bias, Variance, Skewness and Kurtosis of $\sqrt{nG(x_n)}\hat{\alpha}$.

n	Bias	Variance	Skewness	Kurtosis
500	0.53	2.74	0.81	4.37
4000	0.40	2.42	0.43	3.33
8000	0.36	2.31	0.37	3.20
Asymptotic	0	2.08	0	3

5. Proofs

We begin with a technical lemma that collects asymptotic information on the tail behavior of the d.f. of the linear process defined in (1.1).

LEMMA 5.1. *Let $\{X_t\}$ be the linear process defined in* (1.1) *with innovations satisfying* (1.2). *Also assume* (C.1) *and* (C.2). *Then for a fixed $y > 0$, as $x \to \infty$,*

(i) $P\{|X_1| > x, |X_j| > xy\}/P\{|Z_1| > x\} \sim \sum_{k=0}^{\infty}(|c_k|^\alpha \wedge y^{-\alpha}|c_{j+k-1}|^\alpha)$,

(ii) $P\{X_1 > x, |X_j| > xy\}/P\{|Z_1| > x\} \sim \sum_{k=0}^{\infty}(rI[c_k > 0] + sI[c_k < 0])(|c_k|^\alpha \wedge y^{-\alpha}|c_{j+k-1}|^\alpha)$,

(iii) $P\{|X_1| > xy, X_j < -x\}/P\{|Z_1| > x\} \sim \sum_{k=0}^{\infty}(rI[c_k < 0] + sI[c_k > 0])(|c_k|^\alpha \wedge y^{-\alpha}|c_{1-j+k}|^\alpha)$,

(iv) $P\{X_1 > x, X_j > x\}/P\{|Z_1| > x\} \sim \sum_{k=0}^{\infty}(rI[c_k \wedge c_{j+k-1} > 0] + sI[c_k \vee c_{j+k-1} < 0])(|c_k| \wedge |c_{j+k-1}|)^\alpha$,

(v) $P\{X_1 \le -x, X_j > x\}/P\{|Z_1| > x\} \sim \sum_{k=0}^{\infty}(rI[c_k < 0 < c_{j+k-1}] + sI[c_{j+k-1} < 0 < c_k])(|c_k| \wedge |c_{j+k-1}|)^\alpha$.

PROOF. (i) Fix a positive integer m and set

$$X_j^{(m)} = \sum_{i \le m} c_i Z_{j-i}. \tag{5.1}$$

Now in the heavy-tailed case, i.e. under assumption (1.2), a large value for $X_j^{(m)}$ is almost exclusively attributable to a single large innovation. A more precise statement of this observation is that for $y > 0$, as $x \to \infty$

$$P\left\{\left|\bigvee_{i\le m} c_i Z_{1-i}\right| > x, \left|\bigvee_{i\le m} c_i Z_{j-i}\right| > xy\right\} \sim P\{|X_1^{(m)}| > x, |X_j^{(m)}| > xy\}. \tag{5.2}$$

The above statement follows from the fact, which is discernible by the argument in Proposition 2.1 of Chernick *et al.* (1991), that for any $\epsilon > 0$,

$$P\left\{|X_1^{(m)}| > x, \left|\bigvee_{i\le m} c_i Z_{1-i}\right| \le x(1-\epsilon)\right\} = O(P^2\{|Z_1| > x\}) \quad \text{as} \quad x \to \infty, \tag{5.3}$$

and the fact that as $n \to \infty$, $P\{|\bigvee_{i\le m} c_i Z_{1-i}| > x, |\bigvee_{i\le m} c_i Z_{j-i}| > xy\}/P\{|Z_1| > x\}$

$$\begin{aligned} &\sim \Bigg(\sum_{i=m-j+2}^{m} P\{|c_i Z_1| > x\} \\ &\quad + \sum_{i=0}^{m-j+1} P\{|c_i Z_1| > x, |c_{i+j-1} Z_1| > xy\} \\ &\quad + \sum_{i=0}^{-m+j-2} P\{|c_i Z_1| > xy\}\Bigg)/P\{|Z_1| > x\} \\ &\sim \sum_{i=m-j+2}^{m} |c_i|^\alpha + \sum_{i=0}^{m-j+1} [|c_i| \wedge y^{-1}|c_{i+j-1}|]^\alpha + \sum_{i=0}^{-m+j-2} y^{-\alpha}|c_i|^\alpha. \end{aligned} \tag{5.4}$$

Therefore from (5.2) and (5.4) it follows that

$$\begin{aligned} &\lim_{m\to\infty}\lim_{x\to\infty} P\{|X_1^{(m)}| > x, |X_j^{(m)}| > xy\}/P\{|Z_1| > x\} \\ &\quad = \sum_{i=0}^{\infty}(|c_i| \wedge y^{-1}|c_{i+j-1}|)^\alpha. \end{aligned} \tag{5.5}$$

Straightforward albeit tedious calculations establish that (5.5) suffices to show (i). The other statements in the lemma are established by similar reasoning and hence their proofs are omitted. □

LEMMA 5.2. *Under the same conditions as Lemma* 5.1,
(i) $\lim_{x\to\infty} P\{|X_1| > x\}/P\{|Z_1| > x\} = \sum_{k=0}^{\infty} |c_k|^\alpha =: \|c\|_\alpha^\alpha$ *and*
(ii) $\lim_{x\to\infty} P\{X_1 > x\}/P\{|X_1| > x\} = (r\Sigma[c_k^+]^\alpha + s\Sigma[c_k^-]^\alpha)/\|\underline{c}\|_\alpha^\alpha =: p.$

PROOF. (i) follows from Lemma 5.1 (i) with $y = j = 1$ and (ii) follows from (i) and Lemma 5.1 taking $y = j = 1$ in (ii). □

PROOF OF THEOREM 2.1. Recall that for $\tilde{\alpha}_n$ in (2.6) we noted $\tilde{\alpha}_n \to \alpha$ as $n \to \infty$. Thus it suffices to show $\hat{\alpha}_n - \tilde{\alpha}_n \xrightarrow{P} 0$, which is easily seen to be implied by

$$Y_n = \frac{1 - F_n(x_n) + F_n((-x_n)^-)}{G(x_n)} \xrightarrow{P} 1. \tag{5.6}$$

To show (5.6), note

$$E(Y_n - 1) = 0 \tag{5.7}$$

and

$$\operatorname{Var}(Y_n) \le \frac{P\{|X_1| > x_n\} + 2\sum_{j=2}^n [P\{|X_1| > x_n, |X_j| > x_n\} - P^2\{|X_1| > x_n\}]}{nP^2\{|X_1| > x_n\}}. \tag{5.8}$$

By a lengthy computation one can show that

$$\lim_{m\to\infty} \overline{\lim_{n\to\infty}} \left\{ \frac{\sum_{j=2}^n P\{|X_1| > x_n, |X_j| > x_n\} - \sum_{j=2}^n P\{|X_1^{(m)}| > x_n, |X_j^{(m)}| > x_n\}}{nP^2\{|X_1| > x_n\}} \right\} = 0 \tag{5.9}$$

where $X_j^{(m)}$ was defined in (5.1). One may also check using (C.2) and (2.5) that

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \frac{\sum_{j=2}^n [P\{|X_1^{(m)}| > x_n, |X_j^{(m)}| > x_n\} - P^2\{|X_1| > x_n\}]}{nP^2\{|X_1| > x_n\}} \\
&\quad \le \overline{\lim_{n\to\infty}} \left[\frac{\sum_{j=2}^{2m+1} \sum_{i=-m}^{m\wedge(m-j+1)} |c_i \wedge c_{i+j-1}|^\alpha}{nP\{|X_1| > x_n\}} + \frac{\sum_{j=2m+2}^n ([\sum_{i\le m} |c_i|^\alpha]^2 - \|\underline{c}\|_\alpha^{2\alpha})}{n} \right] \\
&\quad = \left(\sum_{i\le m} |c_i|^\alpha \right)^2 - \|\underline{c}\|_\alpha^{2\alpha} = o(1), \qquad \text{as} \quad m \to \infty.
\end{aligned}$$

It thus follows from (2.5) and (5.8) that $\lim_{n\to\infty} \mathrm{Var}(Y_n) = 0$, which together with (5.7) completes the proof of (5.6). The consistency of $\hat{\beta}_n$ can be established similarly. □

PROOF OF THEOREM 2.2. By Proposition 4.28 in Resnick (1987) it follows that $a_n^{-1} M_n = a_n^{-1} \max_{1\le k\le n} |X_k| \xrightarrow{d} W$ where $a_n = \inf\{u : P\{|Z_1| > u\} \le n^{-1}\}$ and W is a nondegenerate continuous r.v. Hence for any $\epsilon > 0$ there exists $K > 1$ sufficiently large that

$$P\{K^{-1} a_n \le M_n \le K a_n\} \ge 1 - \epsilon, \qquad n \ge 1. \tag{5.10}$$

Now define processes

$$\psi_n(T) = \frac{G_n(\sqrt{a_n}T)}{G(\sqrt{a_n})}, \qquad K^{-1} \le T \le K, \quad n \ge 1, \tag{5.11}$$

where K is chosen large enough so that (5.10) holds. It follows by the method of proof of Theorem 2.1 that for each fixed $K^{-1} \le T \le K$, $\psi_n(T) \xrightarrow{P} T^{-\alpha} =: \psi(T)$ as $n \to \infty$. Let Q denote the rationals. Then by a diagonalization argument it is easy to see that for every subsequence n' there exists a further subsequence n'' such that

$$\psi_{n''}(T) \xrightarrow{\text{a.s.}} \psi(T), \qquad \text{as} \quad n'' \to \infty, \quad T \in Q \cap [K^{-1}, K]. \tag{5.12}$$

Using the uniform continuity of ψ on $[K^{-1}, K]$ and the monotonicity of the ψ_n in T, it follows from (5.12) that

$$\sup_{K^{-1}\le T\le K} |\psi_n(T) - \psi(T)| \xrightarrow{P} 0 \qquad \text{as} \quad n \to \infty. \tag{5.13}$$

Setting $Y_n(x) = G_n(x)/G(x)$, we obtain

$$Y_n(\sqrt{M_n}) = \psi_n(\sqrt{M_n/a_n}) G(\sqrt{a_n})/G(\sqrt{M_n}). \tag{5.14}$$

Using the locally uniform convergence of regularly varying functions (Bingham *et al.* (1987), p. 6) and (5.13), we obtain

$$\sup_{K^{-1}\le T\le K} \left| \psi_n(T) \frac{G(\sqrt{a_n})}{G(\sqrt{a_n}T)} - 1 \right| \xrightarrow{P} 0. \tag{5.15}$$

It then follows from (5.10), (5.14), and (5.15) that $Y_n(\sqrt{M_n}) \xrightarrow{P} 1$, as $n \to \infty$, and so the theorem holds for $\hat{\alpha}^*$. The analysis for $\hat{\beta}^*$ is similar. □

LEMMA 5.3. *Let* $Y_i = I(|X_i| > a)$, $i \ge 1$. *Then*

$$\sum_{i=2}^{\infty} |\operatorname{Cov}(Y_1, Y_i)| \le M G^{1/p}(a),$$

where M is a constant not depending on a and $p = q/(q-1)$ where q is as in (C.4).

PROOF. The proof proceeds along similar lines as that of Lemma 2.2 in Chanda (1983). Let for $i > 1$, $R_{i-1} = \sum_{k=i-1}^{\infty} c_k \tilde{Z}_{i-k}$, $X_i^* = X_i - R_{i-1}$, where $\tilde{Z} = Z$, if $\delta < 1$ and $= Z - EZ$, if $\delta \geq 1$ where δ is specified later in the proof. Denote by $L_i(y)$ the conditional expectation of Y_i given $R_{i-1} = y$. Throughout the proof, M would stand for generic constants not dependent on i or a. By condition (C.3), the density g_i of X_i^* satisfies $\|g_i\|_\infty \leq M$ and hence for any y and z,

$$\begin{aligned} |L_i(y) - L_i(z)| &< M \int |I(|u| > a) - I(|u + z - y| > a)| du \\ &\leq M|y - z|. \end{aligned}$$

Therefore, for $0 < \eta_i$, $0 \leq e_i - d_i \leq M\eta_i$ where $e_i = \max_{|y| \leq \eta_i} L_i(y)$ and $d_i = \min_{|y| \leq \eta_i} L_i(y)$. Now $\mathrm{Cov}(Y_1, Y_i) = I_i^1 + I_i^2$, say where $I_i^1 = E(Y_1 - G(a))Y_i I(|R_{i-1}| \leq \eta_i)$. Clearly $EY_1 Y_i I(|R_{i-1}| \leq \eta_i) = EY_1 L_i(R_{i-1}) I(|R_{i-1}| < \eta_i) \leq e_i EY_1 = e_i G(a)$. Also, the same is bounded below by $d_i EY_1 I(|R_{i-1}| \leq \eta_i) \geq d_i\{G(a) - G^{1/p}(a) Q_i^{1/q}\}$, by Cauchy Schwartz where $Q_i = P(|R_{i-1}| > \eta_i)$. Similary one gets

$$d_i(1 - Q_i) \leq EY_i I(|R_{i-1}| \leq \eta_i) \leq e_i.$$

Also, $|I_i^2| \leq EY_i I(|R_{i-1}| > \eta_i) \leq G^{1/p}(a) Q_i^{1/q}$. Combining the above inequalities one gets

$$|\mathrm{Cov}(Y_1, Y_i)| \leq G(a)\eta_i + G^{1/p}(a) Q_i^{1/q}.$$

Now by Theorem 2 of von Bahr and Esseen (1965), $Q_i \leq \eta_i^{-\delta} E|R_{i-1}|^\delta \leq M\eta_i^{-\delta} \sum_{k=i-1}^{\infty} |c_k|^\delta$, for $0 < \delta < 2 \wedge \alpha$. Therefore

$$\sum_{i=2}^{\infty} |\mathrm{Cov}(Y_1, Y_i)| \leq G^{1/p}(a) M,$$

provided

$$\sum_{i=2}^{\infty} \eta_i + \sum_{i=2}^{\infty} \eta_i^{-\delta/q} \left(\sum_{k=i-1}^{\infty} |c_k|^{\delta/q} \right) < \infty.$$

Let $\eta_i = (\sum_{k=i-1}^{\infty} |c_k|^{\delta/q})^{q/(q+\delta)}$. Then the above is no more than

$$2 \sum_{i=2}^{\infty} \eta_i = 2 \sum_{i=2}^{\infty} \left(\sum_{k=i-1}^{\infty} |c_k|^{\delta/q} \right)^{q/(q+\delta)} \leq 2 \sum_{k=1}^{\infty} k |c_k|^{\delta/(q+\delta)} < \infty,$$

by (C.4) with a choice of δ s.t. $\delta/(q+\delta) > \gamma$. □

To prove Theorem 2.3, we need to establish a weak limit result for $(nG(x_n))^{-1/2}(\sum_1^n \xi_{nj}, \sum_1^n \delta_{nj})$. This is achieved via the blocking technique for sums of weakly dependent random variables. The following lemma captures the basic idea.

LEMMA 5.4. *Let $k = k_n \to \infty$, $l = l_n \to \infty$ be integer sequence satisfying, $kl^{-\theta} = O(1)$, $kl/(nG^{1/q}(x_n)) \to 0$, where θ and q are as in* (C.4). *Define, with $r = [n/k]$,*

$$U_i = \frac{1}{\sqrt{nG(x_n)}} \sum_{j=1}^{r-l} \xi_{n,(i-1)r+j}, \tag{5.16}$$

$$V_i = \frac{1}{\sqrt{nG(x_n)}} \sum_{j=r-l+1}^{r} \xi_{n,(i-1)r+j}, \qquad 1 \le i \le k, \tag{5.17}$$

$$W = \frac{1}{\sqrt{nG(x_n)}} \sum_{j=1}^{n-kr} \xi_{n,kr+j}. \tag{5.18}$$

Then (i) $\sum_{i=1}^{k} V_i + W = o_p(1)$ *and* (ii) $|\phi^{(k)}(u,\dots,u) - [\phi^{(1)}(u)]^k| \to 0$, *as $n \to \infty$, for any $u \in \Re$, where for any $1 \le i \le k$, $\phi^{(i)}$ is the joint characteristic function of $(U_1,\dots,U_i)$.*

PROOF. It is easy to see that

$$|\phi^{(k)}(u,\dots,u) - [\phi^{(1)}(u)]^k| \le \sum_{j=2}^{k} |\phi^{(j)}(u,\dots,u) - \phi^{(1)}(u)\phi^{(j-1)}(u,\dots,u)|. \tag{5.19}$$

Let $N_{j-1} = \exp(iu\sum_1^{j-1} U_i) - \phi^{(j-1)}(u,\dots,u)$, $P_j = \exp(iuU_j)$ and $\mathcal{N}_j = \sigma\{Z_{(j-1)r-l+1},\dots,Z_{jr-1}\}$, $j \ge 2$. Note that N_{j-1} is independent of $\mathcal{N}_j$. Therefore with $P_j^* = P_j - E(P_j \mid \mathcal{N}_j)$, we get $|EN_{j-1}P_j| = |EN_{j-1}P_j^*| \le 2E|P_j^*|$, since $|N_{j-1}| \le 2$. Therefore RHS (5.19) is bounded by $2kE|P_1^*|$.

Note that $X_j = R_j + W_j$, where $R_j = \sum_{i=0}^{l+j-1} c_i Z_{j-i}$, $j = 1,\dots,r-l$, is $\mathcal{N}_1$ measurable, and W_j is independent of $\mathcal{N}_1$. Furthermore

$$P_1 = \exp\left\{\frac{iu}{\sqrt{nG(x_n)}} \sum_{j=1}^{r-l} h(R_j + W_j)\right\},$$

where $h(x) = I(|x| > Tx_n) - v_n I(|x| > x_n)$. Therefore, one can write

$$E(P_1 \mid \mathcal{N}_1) = E^{W_1^*,\dots,W_{r-l}^*}\left(\exp\left\{\frac{iu}{\sqrt{nG(x_n)}} \sum_{j=1}^{r-l} h(R_j + W_j^*)\right\}\right),$$

where $(W_1^*,\dots,W_{r-l}^*)$ is an independent copy of $(W_1,\dots,W_{r-l})$ and is also independent of $\mathcal{N}_1$. Hence

$$E|P_1^*| \le E\left|\exp\left\{\frac{iu}{\sqrt{nG(x_n)}} \sum_{j=1}^{r-l} h(R_j + W_j)\right\} - \exp\left\{\frac{iu}{\sqrt{nG(x_n)}} \sum_{j=1}^{r-l} h(R_j + W_j^*)\right\}\right|. \tag{5.20}$$

Using the elementary inequality $|\exp(ix) - 1| \le M|x|^\tau$, where τ is as in (C.4), for all $x \in \Re$, for a large enough constant M, we get that RHS (5.19),

$$\le 2Mk \left(\frac{u}{\sqrt{nG(x_n)}}\right)^\tau \sum_{j=1}^{r-l} E|h(R_j + W_j) - h(R_j + W_j^*)|^\tau, \tag{5.21}$$

since $0 < \tau \le 1$.

By (C.3), the density of R_j is bounded by M, say, uniformly in j. Therefore

$$\begin{aligned} E^{R_j}&|I(|R_j + W_j| > Tx_n) - I(|R_j + W_j^*| > Tx_n)|^\tau \\ &\le (E^{R_j}|I(|R_j + W_j| > Tx_n) - I(|R_j + W_j^*| > Tx_n)|)^\tau \\ &\le \left(M \int |I(|y| > Tx_n) - I(|y + W_j^* - W_j| > Tx_n)|dy\right)^\tau, \\ &\le M'|W_j^* - W_j|^\tau, \end{aligned}$$

for some M' not depending on j.

Since $v_n = O(1)$, by the above calculation,

$$\begin{aligned} \text{RHS (5.21)} &\le M'' \left(\frac{u}{\sqrt{nG(x_n)}}\right)^\tau k \sum_{j=1}^{r-l} E|W_j^* - W_j|^\tau, \\ &\qquad\qquad \text{for some} \quad 0 < M'' < \infty, \\ &\le M''' \left(\frac{u}{\sqrt{nG(x_n)}}\right)^\tau k \sum_{j=1}^{r-l} \sum_{i=l+j}^{\infty} |c_i|^\tau, \\ &\qquad\qquad \text{for some} \quad 0 < M''' < \infty, \\ &\le M''' \left(\frac{u}{\sqrt{nG(x_n)}}\right)^\tau k \sum_{i=l+1}^{\infty} i|c_i|^\tau, \end{aligned} \tag{5.22}$$

by interchanging the order of the summation. Now RHS (5.22) converges to zero, as $n \to \infty$, by (C.4) and the conditions on l and k, completing the proof of (ii).

To prove (i), write $V_i = V_{i,1} - V_{i,2}$, $W = W_1 - W_2$, where

$$V_{i,1} = \frac{1}{\sqrt{nG(x_n)}} \sum_{j=r-l+1}^{r} (I[|X_{(i-1)r+j}| > Tx_n] - G(Tx_n)),$$

and

$$W_1 = \frac{1}{\sqrt{nG(x_n)}} \sum_{j=1}^{n-kr} (I(|X_{kr+j}| > Tx_n) - G(Tx_n)).$$

Then

$$E\left(\sum_{i=1}^{k} V_i + W\right)^2 \le 2\left\{\text{Var}\left(\sum_{i=1}^{k} V_{i,1} + W_1\right) + \text{Var}\left(\sum_{i=1}^{k} V_{i,2} + W_2\right)\right\}$$

$$\leq M\frac{(kl+n-rk)}{nG(x_n)}\left\{EY_1^2+2\sum_{i=1}^{\infty}|\operatorname{Cov}(Y_1,Y_{i+1})| + v_n^2\left(E\tilde{Y}_1^2+2\sum_{i=1}^{\infty}|\operatorname{Cov}(\tilde{Y}_1,\tilde{Y}_{1+i})|\right)\right\}$$

where $Y_i = I(|X_i| > Tx_n)$, $\tilde{Y}_i = I(|X_i| > x_n)$,

$$= O\left(\frac{kl}{nG^{1/q}(x_n)}\right),$$

by Lemma 5.3, which converges to zero as $n \to \infty$, by the conditions on l and k. □

PROOF OF THEOREM 2.3. Take $k = n^a$, $l = n^{a/\theta}$, where $a = (q+1)\theta/(q(\theta+1)+2\theta)$. Then the conditions of Lemma 5.4 are satisfied. Also note for later use

$$\frac{n}{k^2P(|Z_1| > x_n)} = o(1) \tag{5.23}$$

as $n \to \infty$.

Therefore by Lemma 5.4 in order to prove

$$\frac{1}{\sqrt{nG(x_n)}}\sum_1^n \xi_{nj} \xrightarrow{d} N(0,\lambda_{11}),$$

it is enough to show

$$\sum_1^k U_i^* \to N(0,\lambda_{11}),$$

where $U_i^* \stackrel{d}{=} U_i$ given in (5.16) and U^* are i.i.d. In the same way by arguing with a linear combination of ξ and δ, to prove

$$\frac{1}{\sqrt{nG(x_n)}}\left(\sum_1^n \xi_{nj}, \sum_1^n \delta_{nj}\right)^T \xrightarrow{d} N_2(0,\Lambda)$$

it suffices to show that

$$\left(\sum_1^k U_i^*, \sum_1^k \tilde{U}_i^*\right)^T \xrightarrow{d} N(0,\Lambda) \tag{5.24}$$

where $(U_i^*, \tilde{U}_i^*)$ are i.i.d. and $(U_i^*, \tilde{U}_i^*) \stackrel{d}{=} (U_i, \tilde{U}_i)$ with

$$\tilde{U}_i = \frac{1}{\sqrt{nG(x_n)}}\sum_{j=1}^{r-l}\delta_{n,(i-1)r+j}.$$

Applying Lemmas 5.1 and 5.2, we obtain

$$\begin{aligned}(5.25)\quad \lambda_{11} &= \lim k \operatorname{Var}(U_1^*)\\ &= |T^{-\alpha} - T^{-2\alpha}| + 2\|\underline{c}\|_\alpha^{-\alpha} \sum_{j=2}^{\infty} \Bigg[(T^{-\alpha} + T^{-2\alpha}) \sum_{k=-\infty}^{\infty} (|c_k| \wedge |c_{j+k-1}|)^\alpha \\ &\qquad - T^{-\alpha} \sum_{k=0}^{\infty} \{ (T^{-1}|c_k| \wedge |c_{j+k-1}|)^\alpha \\ &\qquad + (|c_k| \wedge T^{-1}|c_{j+k-1}|)^\alpha \} \Bigg],\end{aligned}$$

$$\begin{aligned}(5.26)\quad \lambda_{22} &= \lim k \operatorname{Var}(\tilde{U}_1^*)\\ &= -2\beta(p-q) + \frac{1}{\|\underline{c}\|_\alpha^\alpha} \sum_{k,l} (1+\beta^2) \operatorname{sgn}(c_k c_l)(|c_k| \wedge |c_l|)^\alpha \\ &\quad + \frac{2\beta}{\|\underline{c}\|_\alpha^\alpha}(r-s) \sum_{k,l} (I[c_k \vee c_l < 0] - I[c_k \wedge c_l > 0])(|c_k| \wedge |c_l|)^\alpha,\end{aligned}$$

where $\operatorname{sgn}(x) = I[x>0] - I[x<0]$, and

$$\begin{aligned}(5.27)\quad \lambda_{12} &= \lim k \operatorname{Cov}(U_1^*, \tilde{U}_1^*)\\ &= \|\underline{c}\|_\alpha^{-\alpha} \Bigg[(1-\beta) \sum_{k,l} (rI[c_k > 0] + sI[c_k < 0])(|c_k| \wedge T^{-1}|c_l|)^\alpha \\ &\quad - (1+\beta) \sum_{k,l} (rI[c_k < 0] + sI[c_k > 0])(|c_k| \wedge T^{-1}|c_l|)^\alpha \\ &\quad - (1-\beta)T^{-\alpha} \sum_{k,l} (rI[c_k > 0] + sI[c_k < 0])(|c_k| \wedge |c_l|)^\alpha \\ &\quad + (1+\beta)T^{-\alpha} \sum_{k,l} (rI[c_k < 0] + sI[c_k > 0])(|c_k| \wedge |c_l|)^\alpha \Bigg].\end{aligned}$$

Thus from (5.25)–(5.27) we see that for $s_1, s_2 \subset \Re$,

$$(5.28)\qquad \lim_{n\to\infty} \operatorname{Var}\left(s_1 \sum_1^k U_i^* + s_2 \sum_1^k \tilde{U}_i^* \right) = s_1^2 \lambda_{11} + 2 s_1 s_2 \lambda_{12} + s_2^2 \lambda_{22}.$$

Next note that for all large n, $|\sum_1^{r-l} \zeta_{n,(i-1)r+j}| \le 2(1+T^{-\alpha})r$ and $|\sum_1^{r-l} \delta_{n,(i-1)r+j}| \le 2r$ so that the Lindeberg-Feller conditions hold for $s_1 \sum_1^k U_i^* + s_2 \sum_1^k \tilde{U}_i^*$ by virtue of (5.23). Hence by the Lindeberg-Feller CLT and (5.28),

$$s_1 \sum_1^k U_i^* + s_2 \sum_1^k \tilde{U}_i^* \xrightarrow{d} N(0, s_1^2\lambda_{11} + 2s_1 s_2 \lambda_{12} + s_2^2 \lambda_{22})$$

and this yields (5.24), since s_1 and s_2 are arbitrary.

The proof of (2.10) concludes upon noting that

$$\sqrt{nG(x_n)}\begin{pmatrix}\hat{\alpha}-\tilde{\alpha}\\ \hat{\beta}-\tilde{\beta}\end{pmatrix}=\frac{1}{\sqrt{nG(x_n)}}\begin{pmatrix}[\frac{-T^{-\alpha}}{\log T}]\sum_1^n \xi_{nj}\\ \sum_1^n \delta_{nj}\end{pmatrix}+o_p(1).$$

Of course, (2.13) is obvious from (2.10) given the assumptions in (2.12). □

PROOF OF THEOREM 2.4. We first establish that the $\hat{\lambda}_{ij}$ in (2.14)–(2.16) are consistent for the λ_{ij} in (5.25)–(5.27). To that end consider

$$A_n=\frac{1}{nG(x_n)}\sum_{i=1}^k \hat{\zeta}_{ni}^2=\frac{1}{nG(x_n)}\sum_{i=1}^k[\zeta_{ni}+(v_n-\hat{v}_n)B_{ni}]^2, \tag{5.29}$$

where $B_{ni}=\sum_{j=1}^r I[|X_{(i-1)r+j}|>x_n]$.

Let $l=n^b$, where $b=a/\theta$. Consider

$$\frac{1}{nG(x_n)}\sum_1^k \zeta_{ni}^2=\sum_1^k(U_i+V_i)^2=\sum_1^k U_i^2+\sum_1^k V_i^2+2\sum_1^k U_iV_i,$$

where U_i and V_i are as in Lemma 5.4. Now

$$E\sum_1^k V_i^2\le 2\sum_{i=1}^k\{\mathrm{Var}(V_{i1})+\mathrm{Var}(V_{i2})\}=O\left(\frac{kl}{n}\right)\to 0,$$

since $a<\theta/(1+\theta)$.

Also by (5.25) $E(\Sigma U_i^2)\to\lambda_{11}$. Therefore

$$E|\Sigma U_iV_i|\le\sqrt{E(\Sigma U_i^2)}\sqrt{E(\Sigma V_i^2)}\to 0.$$

Therefore in order to show $\sum_{i=1}^k \zeta_{ni}^2/(nG(x_n))\xrightarrow{P}\lambda_{11}$, it is enough to show that

$$\sum_1^k U_i^2\xrightarrow{P}\lambda_{11}. \tag{5.30}$$

Going through similar arguments as in the proof of Lemma 5.4, and the fact that $|U_i|\le Mr/\sqrt{nG(x_n)}$, where M is a constant, one can obtain

$$|\psi^{(k)}(x,\ldots,x)-[\psi^{(1)}(x)]^k|\le \text{constant}\left(\frac{xr}{nG(x_n)}\right)^r kl^{-\theta}, \tag{5.31}$$

where $x\in\Re$, $\psi^{(j)}$ is the characteristic function of $(U_1^2,\ldots,U_j^2)$, $1\le j\le k$. It is easy to check that RHS (5.31) $\to 0$ by choice of k and l. Therefore (5.30) will follow if we can show $\sum_1^k U_i^{*2}\xrightarrow{P}\lambda_{11}$, where U_i^* are i.i.d. and $U_i^*\stackrel{d}{=}U_i$. Define

$$W_i^*=\frac{U_i^{*2}}{\lambda_{11}^*(n)}-\frac{1}{k},\qquad 1\le i\le k.$$

where $\lambda_{11}^*(n) = kEU_1^{*2} \to \lambda_{11}$, by (5.25). Then $EW_i^* = 0$ and

$$E\left(\sum_{i=1}^{k} W_i^*\right)^2 = kEW_1^{*2} = \frac{kEU_1^{*4}}{[\lambda_{11}^*(n)]^2} - \frac{1}{k}$$
$$= O\left(\frac{n}{k^2 G^{1+1/q}(x_n)}\right) = O((n^{(2a-1)q/(1+q)}G(x_n))^{(q+1)/q})$$

by a similar calculation as in the proof of Lemma 5.3 which goes to zero by assumption. This shows $\sum_1^k U_i^{*2} \xrightarrow{P} \lambda_{11}$.

By (C.2) and Lemma 5.1 (i) it follows that

$$\sup_{n\geq 1} \frac{1}{nG(x_n)} E \sum_1^k B_{ni}^2 < \infty. \tag{5.32}$$

Thus we get that $A_n \xrightarrow{P} \lambda_{11}$, which in turn shows $\hat{\lambda}_{11} \xrightarrow{P} \lambda_{11}$ since $\frac{k}{nG(x_n)}(\bar{\zeta}_n)^2 = \frac{1}{k}\{\sum_{i=1}^k (U_i + V_i)\}^2 \xrightarrow{P} 0$. One similarly establishes the consistency of $\hat{\lambda}_{12}(n)$ and $\hat{\lambda}_{22}(n)$. The theorem then follows from this and Theorem 2.1. □

The following result will be needed in the proof of Theorem 3.1. Since its proof follows the same lines as that of Theorem 2.3 it will be omitted. Define variables

$$\chi_{nj} = I[|X_j| > x_n] - G(x_n), \qquad j \geq 1, \ n \geq 1. \tag{5.33}$$

Let $A = (a_{ij})$ be the matrix given in the statement of Theorem 3.1.

LEMMA 5.5. *Assume that the hypothesis of Theorem 2.3 holds. Then with ξ_{nj}, δ_{nj} as in Section 2.3 and χ_{nj} as defined in* (5.33),

$$\frac{1}{\sqrt{nG(x_n)}} \sum_{j=1}^{n} (\xi_{nj}, \delta_{nj}, \chi_{nj})' \xrightarrow{d} N(0, A). \tag{5.34}$$

PROOF OF THEOREM 3.1. By virtue of (3.4), it is enough to show

$$\sqrt{\frac{n}{G(x_n)}}\{\hat{P}(X_1 > u_n) - p(u_n/x_n)^{-\alpha}G(x_n)\} \xrightarrow{d} N(0, \sigma_T^2)$$

where σ_T^2 is defined in the statement of Theorem 3.1. Now note

$$\sqrt{\frac{n}{G(x_n)}}\{\hat{P}(X_1 > u_n) - p(u_n/x_n)^{-\alpha}G(x_n)\}$$
$$= \sqrt{nG(x_n)}(\hat{p} - p)(u_n/x_n)^{-\hat{\alpha}} \frac{G_n(x_n)}{G(x_n)}$$

$$
\begin{aligned}
&+ p\sqrt{nG(x_n)}\{(u_n/x_n)^{-\hat{\alpha}} - (u_n/x_n)^{-\alpha}\}\frac{G_n(x_n)}{G(x_n)} \\
&+ p(u_n/x_n)^{-\alpha}\sqrt{\frac{n}{G(x_n)}}(G_n(x_n) - G(x_n)) \\
&= \frac{1}{2}T^{-\alpha}\sqrt{nG(x_n)}(\hat{\beta} - \beta) - pT^{-\alpha}\log T\sqrt{nG(x_n)}(\hat{\alpha} - \alpha) \\
&+ pT^{-\alpha}\frac{1}{\sqrt{nG(x_n)}}\sum_{j-1}^{n}\chi_{nj} + o_p(1), \\
&= \frac{1}{\sqrt{nG(x_n)}}\left(\frac{1}{2}T^{-\alpha}\sum_{j=1}^{n}\delta_{nj} + p\sum_{j=1}^{n}\xi_{nj} + pT^{-\alpha}\sum_{j=1}^{n}\chi_{nj}\right) + o_p(1) \\
&\xrightarrow{d} N(0, \sigma_T^2). \qquad \square
\end{aligned}
$$

Acknowledgements

We wish to thank an anonymous reviewer for some useful comments which led to a better paper. Research of the first author was partially supported by a grant from the United States National Security Agency.

REFERENCES

Athreya, K. B., Lahiri, S. N. and Wu, W. (1992). Inference for heavy tailed distributions, *J. Statist. Plann. Inference* (to appear).

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation, Encyclopedia of Mathematics and Its Applications*, **27**, Cambridge University Press, Cambridge.

Chanda, K. C. (1983). Density estimation for linear processes, *Ann. Inst. Statist. Math.*, **35**, 439–446.

Chernick, M. R., Hsing, T. and McCormick, W. P. (1991). Calculating the extremal index for a class of stochastic processes, *Adv. in Appl. Probab.*, **23**, 835–850.

Csörgö, S., Deheuvels, P. and Mason, D. M. (1985). Kernel estimates of the tail index of a distribution, *Ann. Statist.*, **13**, 1050–1077.

de Haan, L. and Resnick, S. I. (1980). A simple asymptotic estimate for the index of a stable distribution, *J. Roy. Statist. Soc. Ser. B*, **42**, 83–88.

Hahn, M. G. and Weiner, D. C. (1991). On joint estimation of an exponent of regular variation and an asymmetry parameter for tail distributions, *Sums, Trimmed Sums and Extremes* (eds. M. G. Hahn, D. M. Mason and D. C. Weiner), Progress in Probability, **23**, Birkhäuser, Boston.

Hall, P. (1982). On some simple estimates of an exponent of regular variation, *J. Roy. Statist. Soc. Ser. B*, **44**, 37–42.

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

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*, Springer, New York.

Resnick, S. I. and Starica, C. (1995). Consistency of Hill's estimator for dependent data, *J. Appl. Probab.*, **32**, 139–167.

Rootzén, H., Leadbetter, M. R. and de Haan, L. (1990). Tail and quantile estimation for strongly mixing stationary sequences, Tech. Report, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill.

von Bahr, B. and Esseen, C. G. (1965). Inequalities for the r-th absolute moment of random variables $1 \le r \le 2$, *Annals of Mathematical Statistics*, **36**, 299–303.