# Asymptotic Behavior of Hill's Estimator for Autoregressive Data

Sidney Resnick* Cătălin Stărică†
School of ORIE, Cornell University, Ithaca, NY 14853

August 14, 1996

**Abstract**

Consider a stationary, $p$th order autoregression $\{X_n\}$ satisfying

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + Z_n, \quad n = 0, \pm 1, \pm 2, \ldots$$

whose innovation sequence $\{Z_n\}$ is iid with regularly varying tail probabilities of index $-\alpha$. From observations $X_1, \ldots, X_n$, one may estimate $\alpha^{-1}$ by applying Hill's estimator to $X_1, \ldots, X_n$. Alternatively, a second procedure is to use $X_1, \ldots, X_n$ to get estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p$ of the autoregressive coefficients and then to estimate the residuals by

$$\hat{Z}_t(n) = X_t - \sum_{i=1}^{p} \hat{\phi}_i X_{t-i}, \quad t = p+1, \ldots, n,$$

and then to apply Hill's estimator to the estimated residuals. We show that from the point of asymptotic variance, the second procedure is superior.

## 1 Introduction.

Sets of data displaying large values with high probabilities are commonly encountered in fields such as finance, hydrology, reliability and teletraffic engineering. For these fields estimating the tail probability $\mathrm{P(X > x)}$ of a random variable $X$ for large $x$ has serious practical implications. Often the estimation procedures are based on the semi-parametric assumption of regular variation for the tail of the marginal distribution. Then the index of variation has to be estimated based on a sequence $X_1, X_2, \ldots, X_n$ of observations. A well studied estimator of the reciprocal of the index, Hill's estimator, is known to be consistent and under reasonable conditions also asymptotically normal for iid samples. (See Hall (1982), Mason (1982, 1988), Mason and Turova (1994), de Haan and Resnick (1996), Geluk et al (1996), Davis and Resnick (1984), Hausler and Teugels (1985), Resnick and Stărică (1996).) However, since many real life applications provide one with dependent, stationary data rather then iid data, it is important to understand the behavior of Hill's

*Partially supported by NSF Grant DMS-9400535 at Cornell University. Some support was also received from NSA Grant MDA904-95-H-1036.

†Supported by NSF Grant DMS-9400535 at Cornell University.

estimator under more general assumptions such as stationarity of the observed sequence or, even more generally, only for a common marginal distribution. Several recent papers (Hsing (1991), Rootzen et al. (1990), Rootzen (1995)) support the belief that Hill's estimator performs well even under these weaker assumptions.

Resnick and Stărică (1995) proved the consistency of Hill's estimator for an infinite order moving average sequence whose marginal distribution is regularly varying. They also considered in detail the special case when the observations $\{X_n, n \geq 0\}$ come from a $p$-th order autoregressive process whose residuals have regularly varying tail probabilities of index -$\alpha$. Since both the stationary sequence $\{X_n\}$ and the residuals have distributions with regularly varying tails of index -$\alpha$, for estimating $\alpha^{-1}$ one could either

1. apply Hill's estimator to the observed time series $X_1, X_2, \ldots, X_n$ or
2. assuming the order of the autoregression $p$ is known, fit coefficients of the autoregression and use this to estimate residuals. Then estimate $\alpha$ by applying Hill's estimator to the estimated residuals.

(Methods of estimating autoregressive coefficients in the heavy tailed case have been suggested by Davis and Resnick (1985), Feigin and Resnick (1992, 1993); Mikosch, Gadrich, Klüppelberg and Adler (1993); Davis, Knight and Liu (1992).) Resnick and Stărică (1995) show that both methods are consistent.

The main goal of this paper is to compare efficiencies of the two methods of estimation. We prove under quite general assumptions on the innovations of the AR–process and on the asymptotic behavior of the estimators for the coefficients of the autoregression that the second method based on estimated residuals, is a more efficient procedure. The asymptotic variance of the Hill estimator is always smaller when the second method is used. An important conclusion is that the asymptotic variance of the Hill estimator applied to the estimated residuals does not depend on the coefficients of the AR process and is actually equal to the asymptotic variance of the Hill estimator for independent data.

We now give some basic notations and assumptions. Let $\{X, X_i, i = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of dependent random variables having the same marginal distribution $F_X$ satisfying $(x \to \infty)$

$$\bar{F}_X(x) := 1 - F_X(x) \sim p x^{-\alpha} L_1(x), \quad F_X(-x) \sim q x^{-\alpha} L_1(x) \tag{1.1}$$

where $\alpha > 0$, $p, q \geq 0$, $p + q = 1$ and $L_1$ is a slowly varying function. The distribution of $|X|$ will be denoted by $F_{|X|}$. Note that (1.1) implies

$$\bar{F}_{|X|}(x) := \bar{F}_X(x) + F_X(-x) \sim x^{-\alpha} L_1(x)$$

(we also write $\bar{F}_{|X|} \in RV_{-\alpha}$). We are interested in estimating $\alpha$ based on observing $X_1, X_2, \ldots, X_n$. Note that by setting $p = 1$, we could specialize (1.1) to the one tail case appropriate to positive variables.

Set

$$F^{\leftarrow}_{|X|}(y) := \inf\{x : F_{|X|}(x) \geq y\}, \qquad 0 < y < 1$$

and

$$b_{|X|}(t) := \left(\frac{1}{1 - F_{|X|}}\right)^{\leftarrow}(t) = F^{\leftarrow}_{|X|}\left(1 - \frac{1}{t}\right), \quad t > 1. \tag{1.2}$$

To simplify notation, $b$ will always stand for $b_{|X|}$.

For $1 \leq i \leq n$, write $|X|_{(i)}$ for the ith largest value of $|X_1|, |X_2|, \ldots, |X_n|$. With this notation, Hill's estimator based on the $k$ largest order statistics (Hill (1975)) is defined as

$$H_{k,n}^{|X|} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{|X|_{(i)}}{|X|_{(k+1)}}. \tag{1.3}$$

Asymptotic properties of $H_{k,n}^{|X|}$ have been studied under assumptions which include that $k = k(n)$ is a function of $n$ such that $n/k \to \infty$ as $n \to \infty$.

It is useful to define the tail empirical measure which is going to play a key role. For $x \in \mathbb{R}$ and $A \subset \mathbb{R}$ define

$$\epsilon_x(A) = \begin{cases} 1, \text{if } x \in \ A \\ 0, \text{if } x \in \ A^c. \end{cases}$$

Define the tail empirical measure associated to the sequence of random variable $\{|X_n|, n \in \mathbb{N}\}$ as

$$\nu_{|X|,n}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{|X_i|/b(n/k)}(\cdot) \tag{1.4}$$

which is considered as a random element of $M_+((0,\infty])$, the space of Radon measures on the punctured set $(0,\infty]$. The tail empirical process is defined as a stochastic process on $(0,\infty)$ by

$$\{E_{k,n}^{|X|}(y) := \nu_{|X|,n}((y,\infty]), 0 < y \leq \infty\}. \tag{1.5}$$

In Section 2 we discuss the asymptotic behavior of the Hill estimator applied to MA($\infty$) processes of the form

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \quad -\infty < n < \infty, \tag{1.6}$$

where $\{Z_k, k = 0, \pm 1, \pm 2, \ldots\}$ are iid with a common distribution $G$ satisfying the analogue of (1.1). Following the method of Rootzen, Leadbetter and de Haan (1990), we discuss when $\{X_n\}$ satisfies a strong mixing condition and apply one of their results to the infinite moving average. General sufficient conditions on the density of the iid innovations $\{Z_k\}$ guaranteeing that the sequence $\{X_n\}$ satisfies the needed strong mixing condition are known (see for example Gorodetskii (1977) or Withers (1981)). The special dependence structure of the infinite moving average process (1.6) allows for derivation of the asymptotic variance of the estimator in terms of the coefficients $\{c_j\}$ of the infinite order moving average MA($\infty$) given in (1.6). Section 2 also provides some evidence that the main result in Section 4 of Rootzen, Leadbetter and de Haan (1990) is, in fact, versatile and useful although at first glance their statement seems plagued by restrictive conditions. In particular, our result applies to a stationary, autoregressive process of the form

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + Z_n, \quad n = 0, \pm 1, \pm 2, \ldots \tag{1.7}$$

since such a process (under proper assumptions) has a causal representation of the form (1.6) (cf. Brockwell and Davis (1991)).

Section 3 assumes that $\{X_n\}$ is an AR($p$) process defined by the $p$-th order autoregression (1.7). Let $\hat{\boldsymbol{\phi}}^{(n)} = (\hat{\phi}_1^{(n)}, \ldots, \hat{\phi}_p^{(n)}), n \geq 1$ be a given set of consistent estimators for the coefficients of the autoregression (1.7), where $\hat{\boldsymbol{\phi}}^{(n)}$ is based on observing $X_1, \ldots, X_n$. An alternative method of estimating $\alpha$ for the AR–process (1.7) is to apply Hill's estimator to the estimated residuals $\hat{Z}_{p+1}^{(n)}, \ldots, \hat{Z}_n^{(n)}$ defined as

$$\hat{Z}_t^{(n)} := X_t - \sum_{j=1}^{p} \hat{\phi}_j^{(n)} X_{t-j}, \quad t = p+1, \ldots, n. \tag{1.8}$$

To derive the asymptotic behavior of Hill's estimator applied to the vector $\{\hat{Z}_t^{(n)}, p+1 \leq t \leq n\}$ we make the standing assumption that the estimators of the coefficients of the autoregression have an asymptotic law and that they converge at a certain rate $d(n)$; that is,

$$d(n)(\hat{\boldsymbol{\phi}}^{(n)} - \boldsymbol{\phi}) \Rightarrow \boldsymbol{S}$$

where $d(n) \to \infty$ and $\boldsymbol{S}$ is a non-degenerate, proper random vector. This assumption is satisfied by both the Yule Walker estimates (Davis and Resnick (1986)) and the linear programming estimators introduced in Feigin and Resnick (1994). See also Mikosch et al (1995) and Davis et al (1992). Section 3 uses a tail empirical measure approach: associate the tail empirical process to the sequence of estimated residuals $\hat{Z}_1^{(n)}, \hat{Z}_2^{(n)}, \ldots, \hat{Z}_n^{(n)}$, show the weak convergence of the normalized tail empirical process to a process closely related to Brownian motion and deduce from this the asymptotic behavior of the Hill estimator applied to the estimated residuals. See de Haan and Resnick (1994, 1996), Resnick and Stărică (1996), Mason (1988), Mason and Turova (1994).

For proving asymptotic normality of the Hill estimator, a second order regular variation condition and a restriction on the sequence $\{k_n\}$ is needed. Let $H$ be a distribution on $\mathbb{R}$. Then $\bar{H}$:= 1 - H is *second order* $(-\alpha, \rho)$ *regularly varying at* $\infty$ (written $\bar{F} \in 2RV(-\alpha, \rho)$) if there exists an $\alpha > 0$, $\rho \leq 0$, K$\in \mathbb{R}$ and $L$ a slowly varying function such that $\bar{H}(x) = x^{-\alpha} L(x) \in RV_{-\alpha}$ and

$$\lim_{t \to \infty} \frac{\frac{L(tx)}{L(t)} - 1}{g(t)} = K \frac{x^\rho - 1}{\rho} \tag{1.9}$$

for all $x > 0$, where $g \in RV_\rho$ and $g(t) \to 0$ as $t \to \infty$. (See, for example, de Haan and Stadtmüller (1994), Geluk and de Haan (1987)). We will say that a distribution $H$ satisfies **Condition 1** if $\bar{H} \in 2RV(-\alpha, \rho)$ and the slowly varying function $L$ associated to $\bar{H}$ satisfies (1.9).

The function $g$ appearing in (1.9) can be used to further restrict the sequence $k = k(n)$ used in the definition of the Hill estimator. In conjunction with Condition 1, we say the sequence $k(n)$ satisfies **Condition 2** if

$$\sqrt{k} g(b_H(\frac{n}{k})) \to 0 \tag{1.10}$$

for $n \to \infty$ where $b_H(n/k)$ corresponds to the cdf $H$ and is defined like in (1.2).

Condition 2 has been used by many authors. See for example Hall (1982), Häusler and Teugels (1985), Dekkers and de Haan (1993), Hsing (1991).

It will be assumed throughout that the following hold:

**Condition A:** $\{X, X_i, i = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of dependent random variables having the same marginal distribution $F_X$ satisfying

$$\bar{F}_X(x) \sim px^{-\alpha} L_1(x), \quad F_X(-x) \sim qx^{-\alpha} L_1(x) \tag{1.11}$$

where $\alpha > 0$, $p, q \geq 0$, $p + q = 1$ and $L_1$ is a slowly varying function so that (1.11) implies

$$\bar{F}_{|X|}(x) := \bar{F}_X(x) + F_X(-x) \sim x^{-\alpha} L_1(x).$$

The quantile function of $|X|$ is defined as $b(t) := (1/\bar{F}_{|X|})^{\leftarrow}(t)$, $t > 1$.

**Condition B:** $\{Z_i, i = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of iid random variables with the common distribution G where

$$\bar{G}(x) \sim rx^{-\alpha} L_2(x), \quad G(-x) \sim sx^{-\alpha} L_2(x) \tag{1.12}$$

with $\alpha > 0$, $r, s \geq 0$, $r + s = 1$ and $L_2$ a slowly varying function. Note that (1.12) implies

$$\bar{G}_{|Z|}(x) := \bar{G}(x) + G(-x) \sim x^{-\alpha} L(x).$$

The quantile function of $|Z|$ is defined as $\flat(t) := b_{|Z|}(t) = (1/\bar{G}_{|Z|})^{\leftarrow}(t)$, $t > 1$.

# 2 Asymptotics of Hill's estimator for an infinite order moving average process.

In this section we discuss the asymptotic behavior of the Hill estimator applied to the absolute value of an infinite moving average $MA(\infty)$ process of the form

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \quad -\infty < n < \infty, \tag{2.1}$$

where $Z_k$ are iid and satisfy Condition B. Our approach follows Rootzen, Leadbetter and de Haan's (1990) Theorem 4.3 which derives Hill estimator asymptotics under a strong mixing assumption on a stationary sequence $\{X_n\}$. General conditions for an infinite linear combination like (2.1) to satisfy the strong mixing conditions of their Theorem 4.3 are available (see for example Gorodetskii (1977), Withers (1981)). The special dependence structure of the infinite moving average allows us to obtain the expression of the asymptotic variance of the estimator as a function of the coefficients $c_j$ in (2.1). Our Section 2 is also related to Hsing (1991) and our Proposition 2.1 extends and clarifies Hsing's Theorem 4.5.

Denote $t \vee 0 := t^+$ and suppose that the sequence $\{c_j\} \in \mathbb{R}^\infty$ appearing in (2.1) contains at least one non-zero element and that there exists $A > 0$ and $u > 1$ such that

$$|c_j| < Au^{-j}, \quad j \in \mathbb{N}. \tag{2.2}$$

Condition (2.2) holds for example if one assumes there exists $|z_0| > 1$ such that

$$\sum_{j=0}^{\infty} c_j z^j < \infty, \quad |z| < |z_0|, \tag{2.3}$$

which is the case for causal autoregressive processes defined by (1.7). We also assume

$$0 < \sum_{j=0}^{\infty} |c_j|^\delta < \infty \tag{2.4}$$

for $0 < \delta < \alpha \wedge 1$ which implies (cf. Datta and McCormick (1995), Lemma 5.2)

$$\sum_{j=0}^{\infty} |c_j||Z_j| < \infty$$

and

$$\lim_{x\to\infty} \frac{P(|X_1| > x)}{P(|Z_1| > x)} = \sum_{j=0}^{\infty} |c_j|^\alpha \tag{2.5}$$

so that $|X_1|$ also has regularly varying tail probabilities. Next, assume that

$$\sum_{j=1}^{\infty}\sum_{k=0}^{\infty} |c_k|^\alpha \wedge |c_{j+k}|^\alpha \log\left(\frac{|c_{k+j}| \vee |c_k|}{|c_{k+j}| \wedge |c_k|}\right) < \infty, \tag{2.6}$$

a mild condition in view of (1.7), and we also suppose that Condition 1 and Condition 2 hold for $F_{|X|}$, the cdf of $|X_t|$. For a finite moving average of positive random variables with positive coefficients, it is enough to suppose Condition 1 holds for $G_Z$ since by Geluk et al (1995) Theorem 3.2, this implies Condition 1 for the distribution of the finite moving average.

Before stating our result we will prove the following lemma.

**Lemma 2.1** *Assume $\{Z_t\}$ satisfies Condition B and (2.6) holds and let $\{r_n\}$ be a sequence such that $r_n = o(n/k)$. Then*

*(a)*

$$\frac{n}{k}\sum_{j=1}^{r_n} P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y) \to \frac{\sum_{j=1}^{\infty}\sum_{k=0}^{\infty} |c_k|^\alpha x^{-\alpha} \wedge |c_{j+k}|^\alpha y^{-\alpha}}{\sum_{j=0}^{\infty} |c_k|^\alpha}, \tag{2.7}$$

*locally uniformly in $x$ and $y$ on $(0,\infty)$,*

*(b)*

$$\frac{n}{k}\sum_{j=1}^{r_n}\int_1^\infty\int_1^\infty P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y)\frac{dx}{x}\frac{dy}{y} \tag{2.8}$$
$$\to \frac{1}{\alpha}\frac{\sum_{j=1}^{\infty}\sum_{k=0}^{\infty} |c_k|^\alpha \wedge |c_{j+k}|^\alpha (2 + \alpha\log(|c_{k+j}| \vee |c_k|/|c_{k+j}| \wedge |c_k|))}{\sum_{j=0}^{\infty} |c_k|^\alpha},$$

*(c)*

$$\frac{n}{k}\sum_{j=1}^{r_n}\int_1^\infty P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y)\frac{dy}{y}$$
$$+\frac{n}{k}\sum_{j=1}^{r_n}\int_1^\infty P(|X_1| > b(n/k)y, |X_{j+1}| > b(n/k)x)\frac{dy}{y} \tag{2.9}$$
$$\to \frac{1}{\alpha^2}\frac{\sum_{j=1}^{\infty}\sum_{k=0}^{\infty} |c_k|^\alpha \wedge |c_{j+k}|^\alpha (2 + \alpha\log(|c_{k+j}| \vee |c_k|/|c_{k+j}| \wedge |c_k|))}{\sum_{j=0}^{\infty} |c_k|^\alpha},$$

*locally uniformly in $x$ on $(0,\infty)$.*

**Proof**: To prove (a) we fix $x > 0$, $y > 0$. Lemma 5.1 and Lemma 5.2 of Datta and McCormick (1995) yield for any $j \in \mathbb{N}$ and $x > 0$, $y > 0$

$$\frac{n}{k}P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y) \to \frac{\sum_{k=0}^{\infty} |c_k|^\alpha x^{-\alpha} \wedge |c_{j+k}|^\alpha y^{-\alpha}}{\sum_{j=0}^{\infty} |c_k|^\alpha}. \tag{2.10}$$

We intend to bound $1_{j<r_n}(n/k)P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y)$ for all $n$ bigger than some $n_0$ and for all $j > j_0(x,y)$ by a sequence $s_n(j)$ which satisfies

$$\lim_{n\to\infty} \sum_{j=1}^{\infty} s_n(j) = \sum_{j=1}^{\infty} \lim_{n\to\infty} s_n(j). \tag{2.11}$$

Due to (2.10) the result follows by a commonly used variant of Fatou's lemma sometimes called Pratt's lemma (Pratt (1960)).

The form of the bounding sequence $s_n(j)$ is obtained as follows. For any $j$

$$\begin{aligned}
P(|X_1| &> b(n/k)x, |X_{j+1}| > b(n/k)y) \\
&\le P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x, \sum_{i=0}^{\infty} |c_i||Z_{j+1-i}| > b(n/k)y) \\
&\le P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x, \sum_{i=0}^{j-1} |c_i||Z_{j+1-i}| > b(n/k)y/2) \\
&\qquad + P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x, \sum_{i=j}^{\infty} |c_i||Z_{j+1-i}| > b(n/k)y/2) \\
&= P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x) P(\sum_{i=0}^{j-1} |c_i||Z_{j+1-i}| > b(n/k)y/2) \\
&\qquad + P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x, \sum_{i=0}^{\infty} |c_{i+j}||Z_{1-i}| > b(n/k)y/2) \\
&= I_{j,n} + II_{j,n}.
\end{aligned}$$

We are interested in a bound on $(n/k)(I_{j,n} + II_{j,n})$. Let $\varepsilon > 0$ such that $\alpha - \varepsilon > 0$. Then there exists an $n_1$, independent of $j$ such that for all $n > n_1$

$$\begin{aligned}
\frac{n}{k} I_{j,n} &= \frac{n}{k} P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x) P(\sum_{i=0}^{j-1} |c_i||Z_{j+1-i}| > b(n/k)y/2) \\
&\le P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x) \frac{n}{k} P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)y/2) \qquad (2.12) \\
&\le (1+\varepsilon)(y/2)^{-\alpha+\varepsilon} P(\sum_{i=0}^{\infty} |c_i||Z_{1-i}| > b(n/k)x),
\end{aligned}$$

for a constant $K_1 > 0$ where the last inequality holds by Potter's bound (Bingham et al (1987)).

Set $Y := \sum_{i=0}^{\infty} u^{-i}|Z_{1-i}|$. For our choice of $\varepsilon$ there exist $n_2$ such that for $n > n_2$ and all $j \in \mathbb{N}$

$$\begin{aligned}
(n/k)II_{j,n} =& \frac{n}{k}P(\sum_{i=0}^{\infty}|c_i||Z_{1-i}| > b(n/k)x, \sum_{i=0}^{\infty}|c_{i+j}||Z_{1-i}| > b(n/k)y/2) \\
\leq& \frac{n}{k}P(A\sum_{i=0}^{\infty}u^{-i}|Z_{1-i}| > b(n/k)x, A\sum_{i=0}^{\infty}u^{-i-j}|Z_{1-i}| > b(n/k)y/2) \\
=& \frac{n}{k}P(A\sum_{i=0}^{\infty}u^{-i}|Z_{1-i}| > b(n/k)x, Au^{-j}\sum_{i=0}^{\infty}u^{-i}|Z_{1-i}| > b(n/k)y/2) \qquad (2.13)\\
\leq& \frac{n}{k}P(Y > A^{-1}b(n/k)x, Y > A^{-1}u^j b(n/k)y/2) = \frac{n}{k}P(Y > A^{-1}b(n/k)(x \vee (u^j y/2))) \\
\leq& M(x \vee (u^j y/2))^{-\alpha+\varepsilon}
\end{aligned}$$

where the last step results from Potter's inequality (Bingham et al (1987)). For $j$ such that $x < u^j y/2$, $n > n_1 \wedge n_2$ and $K_2 = K_1 \vee M2^{\alpha-\epsilon}$ it is then true that

$$\begin{aligned}
\frac{n}{k}P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y) &= (n/k)(I_{j,n} + II_{j,n}) \\
&\leq K_2 y^{-\alpha+\varepsilon}\left(P(\sum_{i=0}^{\infty}|c_i||Z_{1-i}| > b(n/k)x) + (u^j)^{-\alpha+\varepsilon}\right). \qquad (2.14)
\end{aligned}$$

Define

$$s_n(j) := K_2 1_{j<r_n} y^{-\alpha+\varepsilon}\left(P(\sum_{i=0}^{\infty}|c_i||Z_{1-i}| > b(n/k)x) + (u^j)^{-\alpha+\varepsilon}\right).$$

Due to the fact that $r_n = o(n/k)$, (2.11) follows since

$$\begin{aligned}
\lim_{n\to\infty}\sum_{j=1}^{\infty}s_n(j) &= \lim_{n\to\infty}K_2 y^{-\alpha+\varepsilon}\left(r_n P(\sum_{i=0}^{\infty}|c_i||Z_{1-i}| > b(n/k)x) + \sum_{j=1}^{r_n}(u^j)^{-\alpha+\varepsilon}\right) \\
&= K_2\frac{(uy/2)^{-\alpha+\varepsilon}}{1-u^{-\alpha+\varepsilon}} \\
&= \sum_{j=1}^{\infty}\lim_{n\to\infty}s_n(j).
\end{aligned}$$

The result follows by Pratt's (1960) lemma. The convergence in (a) is locally uniform in $x$ and $y$ since the left hand side is a sequence of functions monotone in $x$ and $y$ and the limit is a continuous function. This ends the proof of (a).

The proof of (b) follows the same path. Using Lemma 5.1 of Datta and McCormick (1995), (2.12) and (2.13) one can quickly prove that

$$\begin{aligned}
&\int_1^{\infty}\int_1^{\infty}\frac{n}{k}P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y)\frac{dx}{x}\frac{dy}{y} \\
&\to \frac{1}{\alpha^2}\frac{\sum_{k=0}^{\infty}|c_k|^{\alpha}\wedge|c_{j+k}|^{\alpha}(2+\alpha\log(|c_{k+j}|\vee|c_k|/|c_{k+j}|\wedge|c_k|))}{\sum_{j=0}^{\infty}|c_k|^{\alpha}}.
\end{aligned}$$

From (2.12) and (2.13) it follows that

$$
\begin{aligned}
1_{j<r_n} & \int_1^\infty \int_1^\infty \frac{n}{k} P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)y) \frac{dx}{x}\frac{dy}{y} \\
& \le 1_{j<r_n} \Big( \int_1^\infty P\left(\sum_{i=0}^\infty |c_i||Z_{1-i}| > b(n/k)x\right) \frac{dx}{x} \int_1^\infty \frac{n}{k} P\left(\sum_{i=0}^\infty |c_i||Z_{1-i}| > b(n/k)y/2\right) \frac{dy}{y} \\
& \qquad + M \int_1^\infty \int_1^\infty (x \vee (u^j y/2))^{-\alpha+\varepsilon} \frac{dx}{x}\frac{dy}{y} \Big) \\
& = 1_{j<r_n} \Big( \int_1^\infty P\left(\sum_{i=0}^\infty |c_i||Z_{1-i}| > b(n/k)x\right) \frac{dx}{x} \int_1^\infty \frac{n}{k} P\left(\sum_{i=0}^\infty |c_i||Z_{1-i}| > b(n/k)y/2\right) \frac{dy}{y} \\
& \qquad + ((const_1) + (const_2)j)\, u^{-j(\alpha-\varepsilon)} \Big) := s_n^{(1)}(j).
\end{aligned}
$$

Since (2.11) holds for the sequence $s_n^{(1)}(j)$ the conclusion of (b) follows. The proof for claim (c) is similar. This ends the proof of the lemma. □

**Remark 2.1:** Note that the conclusions of the lemma imply that

$$
\frac{n}{kr_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} \epsilon_{|X_i|/b(n/k)}(x,\infty]\right) \to x^{-\alpha}\left(1 + \frac{2\sum_{j=1}^\infty \sum_{k=0}^\infty |c_k|^\alpha \wedge |c_{j+k}|^\alpha}{\sum_{j=0}^\infty |c_k|^\alpha}\right) \tag{2.15}
$$

$$
\begin{aligned}
& \frac{n}{kr_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} \left(\log \frac{|X_i|}{b(n/k)}\right)^+\right) \\
& \to \frac{2}{\alpha^2}\left(1 + \frac{\sum_{j=1}^\infty \sum_{k=0}^\infty |c_k|^\alpha \wedge |c_{j+k}|^\alpha (2 + \alpha \log(|c_{k+j}| \vee |c_k| / |c_{k+j}| \wedge |c_k|))}{\sum_{j=0}^\infty |c_k|^\alpha}\right),
\end{aligned} \tag{2.16}
$$

and

$$
\begin{aligned}
& \frac{n}{kr_n} \mathrm{Cov}\left(\sum_{j=1}^{r_n} \epsilon_{|X_i|/b(n/k)}(x,\infty], \sum_{i=1}^{r_n} \left(\log \frac{|X_i|}{b(n/k)}\right)^+\right) \\
& \to \frac{x^{-\alpha}}{\alpha}\left(1 + \frac{\sum_{j=1}^\infty \sum_{k=0}^\infty |c_k|^\alpha \wedge |c_{j+k}|^\alpha (2 + \alpha \log(|c_{k+j}| \vee |c_k| / |c_{k+j}| \wedge |c_k|))}{\sum_{j=0}^\infty |c_k|^\alpha}\right).
\end{aligned} \tag{2.17}
$$

Let us quickly sketch the reasoning behind (2.15). Since

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{r_n} \epsilon_{|X_i|/b(n/k)}(x,\infty]\right) &= r_n(P(|X_1|/b(n/k) > x) - P(|X_1|/b(n/k) > x)^2) \\
+2 \sum_{j=1}^{r_n-1} (r_n - j)(P(|X_1| > b(n/k)x, |X_{j+1}| > b(n/k)x) &- P(|X_1|/b(n/k) > x)^2)
\end{aligned}
$$

it follows by Lemma 2.1, that one only has to check that

$$
\frac{1}{r_n} \sum_{j=1}^{r_n-1} j \frac{n}{k} P(|X_1| > b(n/k), |X_{j+1}| > b(n/k)x) \to 0, \tag{2.18}
$$

and

$$\frac{1}{r_n}\sum_{j=1}^{r_n-1}(r_n-j)\frac{n}{k}P(|X_1|/b(n/k)>x)^2\to 0 \tag{2.19}$$

when $n\to\infty$. By (2.14), for large $n$,

$$\begin{aligned}&\sum_{j=j_0}^{r_n-1} j\frac{n}{k}P(|X_1|>b(n/k)x,|X_{j+1}|>b(n/k)y)\\ &\qquad\le \frac{r_n(r_n-1)}{2}K_2y^{-\alpha+\varepsilon}\left(P(\sum_{i=0}^{\infty}|c_i||Z_{1-i}|>b(n/k)x)+\sum_{j=1}^{r_n-1}j(u^j)^{-\alpha+\varepsilon}\right).\end{aligned}$$

Then (2.18) leads to

$$\begin{aligned}&\frac{1}{r_n}\sum_{j=j_0}^{r_n-1} j\frac{n}{k}P(|X_1|>b(n/k),|X_{j+1}|>b(n/k)x)\\ &\qquad\le \frac{r_n-1}{2}K_2y^{-\alpha+\varepsilon}P(\sum_{i=0}^{\infty}|c_i||Z_{1-i}|>b(n/k)x)+K_2y^{-\alpha+\varepsilon}\frac{1}{r_n}\sum_{j=1}^{r_n-1}j(u^j)^{-\alpha+\varepsilon}\\ &\qquad\to 0\end{aligned}$$

since $r_n=o(n/k)$ and $\sum_{j=1}^{\infty}j(u^jy/2)^{-\alpha+\varepsilon}<\infty$. To see that (2.19) holds note that

$$\frac{1}{r_n}\sum_{j=1}^{r_n-1}(r_n-j)\frac{n}{k}P(|X_1|/b(n/k)>x)^2\le\frac{r_n-1}{2}\frac{n}{k}P(|X_1|/b(n/k)>x)^2\to 0$$

due to the fact that $r_nk/n\to 0$.

If one defines

$$\lambda_n:=\frac{n}{kr_n}\mathrm{Var}\left(\sum_{i=1}^{r_n}\left(\left(\log\frac{|X_i|}{b(n/k)}\right)^+-\frac{1}{\alpha}\epsilon_{|X_i|/b(n/k)}(1,\infty]\right)\right) \tag{2.20}$$

then (2.15), (2.16) and (2.17) imply that

$$\lambda_n\to\frac{1}{\alpha^2}\left(1+2\frac{\sum_{j=1}^{\infty}\sum_{k=0}^{\infty}|c_k|^\alpha\wedge|c_{j+k}|^\alpha}{\sum_{j=0}^{\infty}|c_k|^\alpha}\right),\quad n\to\infty. \tag{2.21}$$

Write $\mathcal{B}_{ij}$ for the $\sigma$-field $\sigma\{X_k: i\le k\le j\}$ generated by $X_i,X_{i+1},\dots,X_j$ and

$$\alpha(n)=\sup\{|P(A\cap B)-P(A)P(B)|: A\in\mathcal{B}_{1,k},B\in\mathcal{B}_{k+n,\infty},k\ge 1\}.$$

Then $\{X_j\}$ will be called $\alpha(n)$-strongly mixing if $\alpha(n)\to 0$ as $n\to\infty$.

We now verify the conditions of the main result in Section 4 of Rootzen, Leadbetter and de Haan (1990), Theorem 4.3. The numbers of the references to Rootzen and Leadbetter (1990) will carry an asterix (for example (2.5)*) and the ones referring to the current paper will be plain. We will apply Theorem 4.3 to the sequence $\{\log|X_t|\}$.

We start by verifying the Basic Assumptions of Section 2* of Rootzen et al (1990). For $(i)^*$, the strong mixing assumption, to hold, assume that $G$ has a density $G'$ and that $F$ has a density $F'$ which satisfies the Von Mises condition

$$\lim_{t\to\infty} \frac{tF'(t)}{\bar{F}(t)} = \alpha. \tag{2.22}$$

(It would be of interest to know when a Von Mises condition on $G$ implies on for $F$.) We require that the density $G$ be $L_1$-Lipschitz

$$\int_0^\infty |G'(x) - G'(x+y)|dx < Cy, \quad y > 0, \tag{2.23}$$

and there exists $d < 1$ such that

$$E(Z_1^d) < \infty. \tag{2.24}$$

By Gorodetskii (1977) these assumptions and (2.2) guarantee that $\{X_t\}$ is $\alpha(n)$-strongly mixing. Moreover, since $c_j = O(u^{-j})$ for some $u > 1$, there exists $0 < \lambda < 1$ such that $\alpha(n) = O(u^{-\lambda n})$ (Withers (1981), Corollary 4). This takes care of the strong mixing requirements of condition $(i)^*$ of the Basic Assumptions.

Suppose we are given a sequence $r_n \to \infty$ (which will be more fully specified later) satisfying also $r_n/n \to 0$. Choose $\{l_n\}$ so that $l_n/n \to 0$, $l_n/r_n \to 0$ and

$$\frac{n}{r_n}\left(u^{-\lambda l_n} + \frac{l_n}{n}\right) \to 0.$$

This assures that $(ii)^*$ is satisfied. Choose $c_n = k$ and $u_n = \log b(n/k)$. This choice guarantees that $(iii)^*$ holds.

We have to check now the conditions of Theorem 3.5 of Rootzen, Leadbetter and de Haan (1990) for $\psi_1(x) = x1_{x\geq 0}$, $\psi_2(x) = 1_{x\geq 0}$. It is easy to see that $(3.2)^*$ holds for the functions $\psi_1$, $\psi_2$. A sufficient condition for $(2.5)^*$ to hold is $(2.1)^*$ (their Lemma 2.3); that is, the sequence $r_n$ should satisfy

$$\frac{kr_n}{n} \to 0, \quad n \to \infty. \tag{2.25}$$

Condition $(3.10)^*$ asks for the existence of a sequence $w_n \to \infty$ for which

$$r_n w_n \exp(-(\alpha - \epsilon)w_n) \to 0, \quad n \to \infty \tag{2.26}$$

for some $\epsilon < \alpha \wedge 1$. Condition $(3.11)^*$ becomes

$$\frac{r_n w_n}{k^{1/2}} \to 0, \quad n \to \infty. \tag{2.27}$$

We show next that there is always a choice of sequences $r_n \to \infty$, $w_n \to \infty$ such that (2.25), (2.26) and (2.27) hold, provided $\limsup_{n\to\infty} n/k^{3/2} < \infty$ or $\liminf_{n\to\infty} n/k^{3/2} > 0$. First assume that

$$\limsup_{n\to\infty} n/k^{3/2} < \infty. \tag{2.28}$$

Then choose $\delta > 0$ and $0 < \epsilon' < 1$ such that $\delta < \epsilon'/(1-\epsilon')$ and set $r_n = (n/k)^{1-\epsilon'} \to \infty$ and $w_n = r_n^\delta \to \infty$. Then conditions (2.25) and (2.26) hold trivially and (2.27) becomes

$$\frac{r_n^{1+\delta}}{k^{1/2}} = \frac{n}{k^{3/2}} \left(\frac{n}{k}\right)^{\delta(1-\epsilon')-\epsilon'} \to 0$$

by the choice of $\epsilon'$ and $\delta$ and (2.28).

Assume now that

$$\liminf_{n\to\infty} n/k^{3/2} > 0 \tag{2.29}$$

and set

$$r_n = \frac{2(\alpha-\epsilon)k^{1/2}}{\log^2 k} \to \infty, \quad w_n = \frac{1}{2(\alpha-\epsilon)} \log k \to \infty.$$

Condition (2.25) becomes

$$\frac{kr_n}{n} = \frac{k^{3/2}}{n} \frac{2(\alpha-\epsilon)}{\log^2 k} \to 0$$

due to (2.29). Conditions (2.26) and (2.27) are easy to check. Assumptions (2.28) and (2.29) are the only new standing restrictions we impose on the sequence $k$.

The conditions on

$$\lambda_n^{(1)} = \frac{n}{kr_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} \left(\log \frac{|X_i|}{b(n/k)}\right)^+\right),$$

$$\lambda_n^{(2)} = \frac{n}{kr_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} \epsilon_{|X_i|/b(n/k)}(1,\infty]\right)$$

and

$$\lambda_n = \frac{n}{kr_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} \left(\left(\log \frac{|X_i|}{b(n/k)}\right)^+ - \frac{1}{\alpha}\epsilon_{|X_i|/b(n/k)}(1,\infty]\right)\right)$$

are satisfied due to Remark 2.1. To check condition (4.22)* denote $z_n = \log v_n$. The condition then reads

$$\frac{n}{kr_n} \mathrm{Var}\left(\sum_{i=1}^{r_n} (\epsilon_{|X_i|/b(n/k)}(1,\infty] - \epsilon_{|X_i|/b(n/k)}(v_n/b(n/k),\infty])\right) \to 0 \tag{2.30}$$

whenever $\sqrt{k}\log(v_n/b(n/k)) \to 0$. The convergence in (2.30) holds due again to Remark 2.1.

We now state our conclusion.

**Proposition 2.1** *Let $\{X_t\}$ be the infinite order moving average (2.1) and assume (2.22), (2.23) and (2.24). If (2.28) or (2.29) hold, as $n \to \infty$ and $k/n \to 0$ then*

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} \log \frac{|X|_{(i)}}{|X|_{(k+1)}} - \int_1^\infty \frac{n}{k} P\left(\frac{|X_1|}{b(n/k)} > x\right)\frac{dx}{x}\right) \Rightarrow N(0,\lambda), \tag{2.31}$$

*where*

$$\lambda := \frac{1}{\alpha^2}\left(1 + 2\frac{\sum_{j=1}^\infty \sum_{k=0}^\infty |c_k|^\alpha \wedge |c_{j+k}|^\alpha}{\sum_{j=0}^\infty |c_k|^\alpha}\right). \tag{2.32}$$

*If Condition I and Condition II also hold for $F_{|X|}$ then*

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{|X|_{(i)}}{|X|_{(k+1)}}-\frac{1}{\alpha}\right)\Rightarrow N(0,\lambda). \tag{2.33}$$

In particular, our result applies to an autoregressive process of the form

$$X_n=\sum_{i=1}^{p}\phi_i X_{n-i}+Z_n, \tag{2.34}$$

where the innovations $Z_t$ satisfy the conditions of Proposition 2.1, since such a process (under proper assumptions) has a causal representation of the form (1.6) (cf. Brockwell and Davis (1991)).

For the case when $X_t$ is a finite moving average, the $\alpha(n)$-mixing condition is trivially satisfied and there is no need to assume (2.22), (2.23) and (2.24). Moreover in this case it is enough to ask Condition I to hold for $F_{|Z|}$ (Geluk, de Haan, Resnick and Stărică (1995)). The case of the finite moving average was covered by Proposition 4.5 of Hsing (1991).

The conditions of Proposition 2.1 are not perhaps as clean as desirable but the main point of the result is that for autoregressions, the variance given in (2.32) is larger than the asymptotic variance obtained when applying the Hill estimator to the estimated residuals. Thus, the procedure of applying the Hill estimator directly to an autoregressive process is inferior to the procedure of first estimating autoregressive coefficients and then estimating $\alpha$ using estimated residuals.

## 3 Asymptotics of Hill's estimator for a stationary AR($p$) process.

In this section, we assume that $\{X_n, -\infty<n<\infty\}$ is an AR($p$) process defined by the $p$-th order autoregression:

$$X_t=\sum_{i=1}^{p}\phi_i X_{t-i}+Z_t,\quad t=0,\pm1,\pm2,\dots. \tag{3.1}$$

We assume that the iid innovations $\{Z_n\}$ satisfy Condition B and suppose

$$\Phi(z)=1-\sum_{i=1}^{p}\phi_i z^i\neq 0,\quad |z|\le 1 \tag{3.2}$$

so that (Brockwell and Davis, (1991)) the autoregression (3.1) exists and has a stationary solution of the form

$$X_n=\sum_{j=0}^{\infty}c_j Z_{n-j},\qquad -\infty<n<\infty \tag{3.3}$$

where

$$C(z)=\sum_{j=0}^{\infty}c_j z^j=\frac{1}{\Phi(z)},\quad |z|\le 1. \tag{3.4}$$

We assume that we have a sequence $\hat{\boldsymbol{\phi}}^{(n)}=(\hat\phi_1^{(n)},\dots,\hat\phi_p^{(n)}), n\ge 1$ of consistent estimators for the coefficients of the autoregression such that:

$$d(n)(\hat{\boldsymbol{\phi}}^{(n)}-\boldsymbol{\phi})\Rightarrow \boldsymbol{S} \tag{3.5}$$

where $d(n) \to \infty$, $\boldsymbol{S}$ is a non-degenerate random vector and $\hat{\boldsymbol{\phi}}^{(n)}$ is based on observing $X_1, \ldots, X_n$. For this sequence of estimators, the *estimated residuals* $\hat{Z}_t^{(n)}$ are defined by (1.8) so that

$$Z_t - \hat{Z}_t^{(n)} = \sum_{i=1}^{p} (\hat{\phi}_i^{(n)} - \phi_i) X_{t-i}. \tag{3.6}$$

The purpose of this section is to show that the Hill estimator applied to $|\hat{Z}_1^{(n)}|, |\hat{Z}_2^{(n)}|, \ldots, |\hat{Z}_n^{(n)}|$ yields a consistent estimator of $\alpha^{-1}$. Following the line of proof in Resnick and Stărică (1996) and de Haan and Resnick (1996), we will show that the normalized empirical process associated with the sequence $|\hat{Z}_1^{(n)}|, |\hat{Z}_2^{(n)}|, \ldots, |\hat{Z}_n^{(n)}|$ converges to a process closely related to a Brownian motion. This is known (Resnick and Stărică (1996) and de Haan and Resnick (1996)) to imply the asymptotic normality of Hill's estimator.

We recall Proposition 2.1 of Resnick and Stărică (1996) since it is central to our proof.

**Proposition 3.1** *Assume that Condition B holds and $G_{|Z|}$ satisfies Condition 1 and $\{k\}$ satisfies Condition 2. Then, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$,*

$$W_n(y) := \sqrt{k} \left( \frac{1}{k} \sum_{i=1}^{n} \epsilon_{|Z_i|/\flat(n/k)}[y, \infty] - y^{-\alpha} \right) \Rightarrow W(y^{-\alpha}), \tag{3.7}$$

*in $D((0,\infty])$, where $\{W(t), t \geq 0\}$ is a standard Brownian motion.*

The behavior of the tail empirical process associated with $\{|\hat{Z}_t^{(n)}|\}$ is given next.

**Proposition 3.2** *Assume that the hypotheses of Proposition 3.1 hold and there exists a sequence $d(n) \to \infty$ and a non-degenerate random variable $\boldsymbol{S}$ such that the coefficient estimators $\boldsymbol{\phi}^{(n)}$ satisfy (3.5). Assume also that the sequence $k = k(n)$ is chosen to satisfy the additional requirement*

$$\frac{\sqrt{k}\flat(n/\sqrt{k})}{\flat(n/k)} = o(d(n)), \quad n \to \infty. \tag{3.8}$$

*Then*

$$\sqrt{k} \left( \frac{1}{k} \sum_{i=1}^{n} \epsilon_{|\hat{Z}_i^{(n)}|/\flat(n/k)}(y, \infty] - y^{-\alpha} \right) \Rightarrow W(y^{-\alpha}) \tag{3.9}$$

*in $D((0,\infty])$, where $\{W(t), t \geq 0\}$ is a standard Brownian motion.*

**Remark 3.1:** How restrictive is condition (3.8)? Since $\flat(t) \in RV_{1/\alpha}$, if we choose $\delta' > 0$ small then, by Potters inequality (Bingham et al (1987), for $n$ big enough

$$\frac{\sqrt{k}\flat(n/\sqrt{k})}{d(n)\flat(n/k)} < (1+\delta') \frac{k^{(1/2)(1+1/\alpha+\delta')}}{d(n)}.$$

Therefore a restriction on $k(n)$ which is sufficient for (3.8) is

$$\frac{k_n^{1+1/\alpha+\delta'}}{d(n)^2} \to 0. \tag{3.10}$$

Suppose $\hat{\phi}^{(n)}$ are the Yule–Walker estimators. If $EZ_1^2 < \infty$ (for instance if $\alpha > 2$), then $d(n) = n^{1/2}$ (Brockwell and Davis (1991), page 240) and the sufficient condition (3.10) becomes $k^{1+1/\alpha+\delta'}/n \to 0$. When $\alpha < 2$, so that $EZ_1^2 = \infty$, the Yule–Walker estimators have an inferior rate of convergence compared to the linear programming (Feigin and Resnick (1994)) or least gamma deviation estimators (Davis et al (1992)). When $Z_n \geq 0$, the linear programming estimators of Feigin and Resnick (1994) can be applied and in this case $d(n) = b(n)$ and therefore for any $\delta'' > 0$ $b(n)/n^{2/\alpha-\delta''} \to \infty$ and thus eventually

$$\frac{k^{1+1/\alpha+\delta'}}{b(n)^2} \leq \frac{k^{1+1/\alpha+\delta'}}{(1+\delta'')n^{2/\alpha-\delta''}}.$$

Then (3.10) (and hence (3.8)) hold if

$$k^{(1+\alpha)/2+\delta'''}/n \to 0$$

for any $\delta''' > 0$.

**Proof:**

We intend to show that, for a given interval $[c,d] \subset (0,\infty]$ and a given $\delta > 0$

$$P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\left|\sum_{i=1}^{n}\epsilon_{|Z_i|/b(n/k)}(x,\infty] - \sum_{i=1}^{n}\epsilon_{|\hat{Z}_i^{(n)}|/b(n/k)}(x,\infty]\right| > \delta\right) \to 0 \tag{3.11}$$

as $n \to \infty$. This together with (3.7) will imply the conclusion of the proposition. To prove (3.11), decompose the probability

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\left(\sum_{i=1}^{n}1_{[|Z_i|/b(n/k)>x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]} + \sum_{i=1}^{n}1_{[|\hat{Z}_i^{(n)}|/b(n/k)>x,|Z_i|/b(n/k)\leq x]}\right) > \delta\right)\\
&\quad\leq P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|Z_i|/b(n/k)>x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]} > \delta/2\right)\\
&\qquad+P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|\hat{Z}_i^{(n)}|/b(n/k)>x,|Z_i|/b(n/k)\leq x]} > \delta/2\right)\\
&\quad= Ia + I(b).
\end{aligned}$$

We will concentrate on proving $Ia \to 0$ as $n \to \infty$ with $\delta/2$ replaced by $\delta$ for typographical ease. The proof for $Ib$ is similar. Let $\varepsilon_n$ be a sequence of positive numbers tending to 0 at a rate to be specified later. Then

$$\begin{aligned}
Ia &= P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|Z_i|/b(n/k)>x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]} > \delta\right)\\
&\leq P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[x<|Z_i|/b(n/k)\leq(1+\varepsilon_n)x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]} > \delta/2\right)\\
&\quad+P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|Z_i|/b(n/k)>(1+\varepsilon_n)x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]} > \delta/2\right)
\end{aligned}$$

$$\leq P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[x<|Z_i|/b(n/k)\leq(1+\varepsilon_n)x]}>\delta/2\right)$$
$$+P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|Z_i|/b(n/k)>(1+\varepsilon_n)x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]}>\delta/2\right)$$
$$=Ia_1+Ia_2.$$

Assume that $\sqrt{k}\alpha c^{-\alpha}\epsilon_n\to\delta/4$. We first consider $Ia_1$ and have

$$\begin{aligned}
Ia_1 &\leq P(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\left|\sum_{i=1}^{n}1_{x<|Z_i|/b(n/k)\leq(1+\varepsilon_n)x}-\sqrt{k}(1-(1+\varepsilon_n)^{-\alpha})x^{-\alpha}\right| \\
&\qquad >\delta/2-\sqrt{k}\sup_{x\in[c,d]}x^{-\alpha}(1-(1+\varepsilon_n)^{-\alpha})) \\
&\leq P(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\left|\sum_{i=1}^{n}1_{x<|Z_i|/b(n/k)\leq(1+\varepsilon_n)x}-\sqrt{k}(1-(1+\varepsilon_n)^{-\alpha})x^{-\alpha}\right| \\
&\qquad >\delta/2-\sqrt{k}(1-(1+\varepsilon_n)^{-\alpha})c^{-\alpha}) \\
&\leq P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\left|\sum_{i=1}^{n}\epsilon_{|Z_i|/b(n/k)}(x,(1+\varepsilon_n)x]-\sqrt{k}(1-(1+\varepsilon_n)^{-\alpha})x^{-\alpha}\right|>\delta/4\right) \\
&= P\left(\sup_{x\in[c,d]}|W_n(x)-W_n((1-\varepsilon)x)|>\delta/4\right) \\
&\to 0
\end{aligned}$$

where the last inequality holds for $n$ greater then some $n_0$ due to the assumption on the rate of convergence of $\epsilon_n$. The convergence to 0 of the last expression is a direct consequence of the convergence in (3.7).

Let us turn now our attention to $Ia_2$. Let $M$ be a large constant. For typographical ease, replace $\delta/2$ by $\delta$ and then

$$\begin{aligned}
Ia_2 &= P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|Z_i|/b(n/k)>(1+\varepsilon_n)x,|\hat{Z}_i^{(n)}|/b(n/k)\leq x]}>\delta\right) \\
&\leq P\left(\frac{1}{\sqrt{k}}\sup_{x\in[c,d]}\sum_{i=1}^{n}1_{[|Z_i-\hat{Z}_i^{(n)}|/b(n/k)>\epsilon_n x]}>\delta\right) \\
&\leq P\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}1_{[|Z_i-\hat{Z}_i^{(n)}|/b(n/k)>\epsilon_n c]}>\delta\right) \\
&\leq P\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}1_{[\sum_{j=1}^{p}|(\hat{\phi}_j^{(n)}-\phi_j)X_{i-j}|/b(n/k)>\epsilon_n c]}>\delta\right) \\
&\leq P\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}1_{[d(n)\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|\sum_{j=1}^{p}|X_{i-j}|/b(n/k)>d(n)\epsilon_n c]}>\delta\right) \\
&\leq P\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}1_{[d(n)\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|\sum_{j=1}^{p}|X_{i-j}|/b(n/k)>cd(n)\epsilon_n]}>\delta, d(n)\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|<M\right)
\end{aligned}$$

$$+P(d(n)\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|\geq M)$$

$$\leq P\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n}1_{[M\sum_{j=1}^{p}|X_{i-j}|/\flat(n/k)>d(n)\epsilon_n c]}>\delta\right)+P(d(n)\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|\geq M)$$

$$\leq\ \frac{n}{\delta\sqrt{k}}P\left(\sum_{j=1}^{p}|X_{p+1-j}|>\frac{c}{M}d(n)\flat(n/k)\epsilon_n\right)+P(d(n)\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|\geq M),$$

where the last step was taken by applying Chebyshev's inequality. Let first $n\to\infty$ and then $M\to\infty$. Due to the fact that $k=k(n)$ has been chosen such that

$$\frac{d(n)\flat(n/k)}{\sqrt{k}\flat(n/\sqrt{k})}\to\infty$$

and $\epsilon_n$ satisfies $\epsilon_n\sim(const)k^{-1/2}$ and since (3.5) holds, it follows that $Ia_2\to 0$ as $n\to\infty$. This ends the proof of our result. □

The asymptotic normality of the estimator follows after an argument as in Resnick and Stărică (1996). For the sake of completeness we state the result.

**Proposition 3.3** *Assume the hypotheses of Proposition 3.2 hold. Then the Hill estimator applied to the estimated residuals satisfies*

$$\sqrt{k}\left(H_{k,n}^{|\hat{Z}|}-\frac{1}{\alpha}\right)\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{|\hat{Z}^{(n)}|_{(i)}}{|\hat{Z}^{(n)}|_{(k+1)}}-\frac{1}{\alpha}\right)\Rightarrow N(0,\frac{1}{\alpha^2}). \tag{3.12}$$

# 4 Concluding remarks.

As remarked, for autoregressive processes the variance given in (2.32) is larger than the asymptotic variance obtained when applying the Hill estimator to the estimated residuals, the latter being the variance obtained when applying the Hill estimator to iid sequences. Thus, the procedure of applying the Hill estimator directly to an autoregressive process is less efficient than the procedure of first estimating autoregressive coefficients and then estimating $\alpha$ using estimated residuals.

In practice, when applying the second method of estimation based on estimated residuals, one tries to choose the set of estimators that have the fastest rate of convergence. If $\alpha>2$, the Yule–Walker estimators converge at rate $\sqrt{n}$ and in this case $\delta'$ can be chosen such that $1+1/\alpha+\delta'<3/2$. So the conclusion of Proposition 3.3 holds if $k=k(n)$ satisfies the constraint $k^{3/2}/n\to 0$. If the left end point of the distribution of $Z$ is 0 and $\alpha<2$ the linear programming estimates converge faster and it is easy to see that the same condition, i.e. $k^{3/2}/n\to 0$, turns out to be sufficient since then $(1+\alpha)/2<3/2$. The conclusion is that with a judicious choice of the estimates for the autoregressive coefficients the asymptotic normality of the Hill estimator applied to the estimated residuals follows whenever the choice of the sequence $k=k(n)$ satisfies the additional constraint $k^{3/2}/n\to 0$.

A comparison between the asymptotic variances in (2.32) and (3.12) clearly shows that for the case of an AR($p$) process the second method based on estimated residuals of estimation of the index of regular variation is a more satisfactory procedure in terms of the asymptotic variance. This conclusion is confirmed by simulation. We simulated the AR(2) process

$$X_t = 1.3X_{t-1} - 0.7X_{t-2} + Z_t, \quad t = 1, \ldots$$

where $Z_t$ are iid so that

$$P(Z_t > x) = \frac{1}{2}x^{-0.7}, \quad P(Z_t < -x) = \frac{1}{2}x^{-0.7}, \quad x \geq 1.$$

The AR(2) process is causal and therefore has an MA($\infty$) representation so that the results of Sections 2 and 3 are applicable. The coefficients $\phi_1$ and $\phi_2$ were estimated by the Yule-Walker method. The results of estimation for two simulation runs are presented, the first one consisting of 700 observations, the second one of 2000 observations.

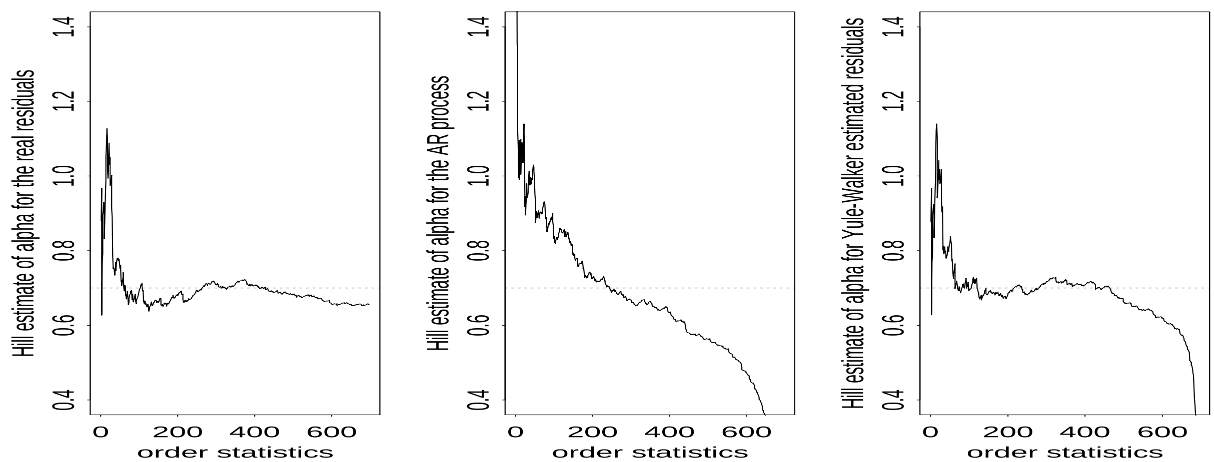


Figure 1

For the first simulation, the estimation based on the AR process $\{X_t\}$ is not very informative. Estimation based on the estimated residuals comes quite close to the correct answer. Figure 1 gives Hill estimator plots as a function of the number of order statistics. In each graph, the dotted line represents the true value of $\alpha$. The left graph applies the Hill estimator to the absolute values of the actual residuals. The middle graph applies it to the absolute values of the time series $\{|X_t|\}$ and the right graph gives the Hill plot for the absolute values of the estimated residuals $\{|\hat{Z}_t|\}$. The estimated coefficients are in this case $\hat{\phi}_1 = 1.2896$ and $\hat{\phi}_2 =$ -0.6906.

The results of the estimation for the longer second run are presented in Figure 2. The estimated coefficients were $\hat{\phi}_1 = 1.2998$ and $\hat{\phi}_2 =$ -0.6996. The second example suggests that by the time the sample size is large enough for the plot based on the time series to level off, the estimation of the coefficients tends to be so accurate that the graph displaying the result of the estimation based on the estimated residuals and the one based on the actual residuals differ very little. Of course, in practice the graph based on the actual residuals is not available.

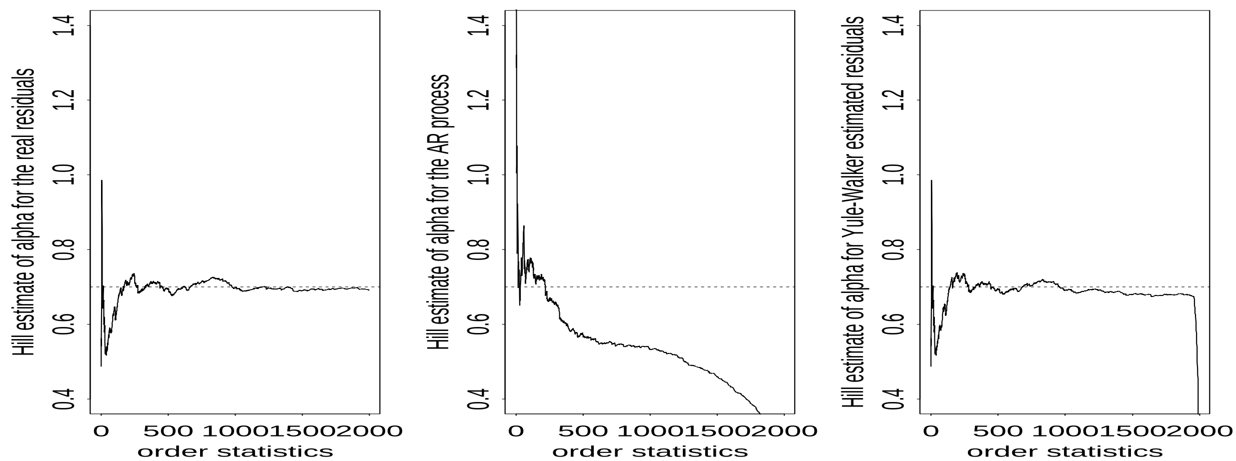


Figure 2

# References


[1] Bingham, N., Goldie, C. and Teugels, J. (1987) *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge University Press Cambridge, UK.

[2] Brockwell, P. and Davis, R. (1991). Time Series: Theory and Methods, 2nd edition, Springer-Verlag, New York.

[3] Datta, S. and McCormick, W. P. (1995). Inference for the tail parameters of a linear process with heavy tailed innovations. Preprint.

[4] Davis, R., Knight, K. and Liu, J.(1992). M estimation for autoregressions with infinite variance. Stoch. Proc and their Appl., 40, 145–180.

[5] Davis, R. and Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. 13, 179–195.

[6] Davis, R. and Resnick, S. (1986). Limit theory for sample covariance and correlation functions. Ann.Statist., 14, 533–558.

[7] Dekkers, A. L. M. and Haan, L. de (1993). Optimal choice of sample fraction in extreme value estimation. J. Mult. Analysis, 47, 173–195.

[8] Feigin, P. D. and Resnick, S. (1992). Estimation for autoregressive processes with positive innovations. Stochastic Models 8, 479–498.

[9] Feigin, P. D. and Resnick, S. (1994). Limit distributions for linear programming time series estimators. Stochastic Processes and their Applications, 51, 135–166.

[10] Geluk, J and Haan, L. de (1987). Regular Variation, Extensions and Tauberian Theorems. CWI tract 40, Centre for Mathematics and Computer Science, Amsterdam, Netherlands.

[11] Geluk, J and Haan, L. de, Resnick, S. and Stărică, C. (1995). Second order regular variation, convolution and the central limit theorem. Preprint. Available as TR1133 at http://www.orie.cornell.edu/trlist/trlist.html

[12] Goldie, C. and Smith, R. (1987). Slow variation with remainder: Theory and applications. Quarterly Journal of Mathematics 38, 45–71.

[13] Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. Theory Probability Appl. 22, 411–413.

[14] Haan, L. de and Resnick, S. I. (1994). Estimating the limit distribution of multivariate extremes. Stochastic Models 9, 275–309.

[15] Haan, L. de and Resnick, S. I. (1996). On asymptotic normality of the Hill estimator. Preprint available as TR1155.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html.

[16] Haan, L. de and Stadtmüller, U. (1994). Generalized regular variation of second order. To appear: J. Austr. Math. Soc.

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

[18] Häusler, E. and Teugels, J. (1985). On the asymptotic normality of Hill's estimator for the exponent of regular variation. Ann. Statist. 13, 743–756

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

[20] Hsing, T. (1991). On tail estimation using dependent data. Ann. Statist. 19, 1547–1569.

[21] Mason, D. (1982). Laws of large numbers for sums of extreme values. Ann. Probability, 10:754–764, 1982.

[22] Mason, D. (1988). A strong invariance theorem for the tail empirical process. Ann. Inst. Henri Poincaré 24, 491-506.

[23] Mason, D. and Turova, T. (1994). Weak convergence of the hill estimator process. In J. Galambos, J. Lechner, and Simiu, editors, *Extreme Value Theory and Applications*. Kluwer Academic Publishers, Dordrecht, Holland.

[24] Mikosch,T., Gadrich, T., Klüppelberg, C., and Adler, R. (1993) Estimation for infinite variance ARMA models. Ann. Statist., 23, 305–321.

[25] Pratt, J. (1960) On interchanging limits and integrals. Ann. Math. Statist. 31, 74–77.

[26] Resnick, S. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag, Berlin.

[27] Resnick, S. and Stărică, C. (1995) Consistency of Hill's estimator for dependent data. Journal of Applied Probability 32, 139-167.

[28] Resnick, S. and Stărică, C. (1996) Smoothing the Hill estimator. To appear: Journal of Applied Probability.

[29] Rootzen, H., Leadbetter, M. and de Haan, L. (1990) Tail and quantile estimation for strongly mixing stationary sequences. Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260

[30] Withers, C.S. (1981). Conditions for linear processes to be strong mixing. Z. Wahrscheinlichkeitstheorie verw. Geb. 57, 477–480