ON TAIL INDEX ESTIMATION FOR DEPENDENT, HETEROGENEOUS DATA*

Jonathan B. Hill†
Dept. of Economics
Florida International University

Original version: May 2005
This version:

March 24, 2006

Abstract

In this paper we analyze the asymptotic properties of the popular distribution tail index estimator by B. Hill (1975) for possibly heavy-tailed, heterogenous, dependent processes. We prove the Hill estimator is weakly consistent for processes with extremes that form mixingale sequences, and asymptotically normal for processes with extremes that are near-epoch-dependent on the extremes of a mixing process. Our limit theory covers infinitely many ARFIMA and FIGARCH processes, stochastic recurrence equations, and simple bilinear processes. Moreover, we develop a simple non-parametric kernel estimator of the asymptotic variance of the Hill estimator, and prove consistency for extremal-NED processes.

1. INTRODUCTION

This paper develops an asymptotic theory for the popular distribution tail index estimator due to B. Hill (1975) under general conditions. Denote by $\{X_t\} = \{X_t : -\infty < t < \infty\}$ a stochastic process on some probability measure space $(\ ,\Im,\mu)$, $\Im = \cup_{t\in\mathbb{Z}}\Im_t$, $\Im_{t-1} \subset \Im_t \equiv \sigma(X_\tau : \tau \leq t)$. We assume X_t has for each t a common marginal distribution F, and without loss of generality assume F

*I would like to thank three anonymous referrees, Bruce Hansen, and Holger Drees for expert commentary that lead to substantial improvements. All errors, of course, are mine alone.

†Dept. of Economics, Florida International University, Miami, FL; www.fiu.edu/~hilljona; jonathan.hill@fiu.edu.

JEL classifications: C12, C16, C52.
Keywords: Hill estimator; regular variation; infinite variance; near epoch dependence; mixingale; kernel estimator; tail array sum.

has support on $(0, \infty)$. A fundamental assumption is the tail probability $\bar{F}(x) \equiv P(X_t > x)$ is regularly varying at ∞: there exists some $\alpha > 0$ such that for all $\lambda > 0$

$$(1) \quad \bar{F}(\lambda x)/\bar{F}(x) \rightarrow \lambda^{-\alpha}$$

as $x \rightarrow \infty$, where α denotes the index of regular variation. Equivalently,

$$(2) \quad \bar{F}(x) = x^{-\alpha}L(x), \quad x > 0,$$

where $L(x)$ is slowly varying. The class of distributions satisfying (1) includes the domain of attraction of the stable laws, coincides with the maximum domain of attraction of the extreme value distributions $\exp\{-x^{-\alpha}\}$, and characterizes the tails of many stochastic recurrence equations, including GARCH processes. See de Haan (1970), Feller (1971), Ibragimov and Linnik (1971), Leadbetter *et al* (1983), Bingham *et al* (1987), Resnick (1987), and Basrak *et al* (2002a,b).

Denote by $X_{(i)} > 0$ the i^{th} order statistic of the sample path $\{X_1, ..., X_n\}$: $X_{(1)} \geq X_{(2)} \geq ...$ Let m be a sequence of integers satisfying $m \rightarrow \infty$ as $n \rightarrow \infty$, and $m = o(n)$. A popular estimator for the inverted tail index α^{-1}, due to B. Hill (1975), is simply

$$\hat{\alpha}_m^{-1} \equiv \frac{1}{m}\sum\nolimits_{t=1}^{n}\left(\ln X_t/X_{(m+1)}\right)_+ = \frac{1}{m}\sum\nolimits_{i=1}^{m}\ln X_{(i)}/X_{(m+1)},$$

where $(z)_+ = \max\{z, 0\}$. See Hsing (1991: eq. 1.5) for an intuitive characterization of $\hat{\alpha}_m^{-1}$ as a method of moments estimator. See, also, Haeusler and Teugels (1985).

The so-called Hill estimator has been used pervasively in the applied finance, economics, statistics and telecommunications literatures. Consider Akgiray and Booth (1988), Cheng and Rachev (1995), Loretan and Phillips (1994), Resnick and Rootzén (2000), Rachev (2003), Chan *et al* (2005), and Hill (2005), to name a few. For alternative techniques for estimating α, consult Pickands (1975), Smith (1987), Rootzén *et al* (1990), Smith and Weissman (1994), Drees *et al* (2004), and Beirlant *et al* (1996, 2005).

We are interested in the asymptotic properties of $\hat{\alpha}_m^{-1}$ under minimal conditions. Asymptotic normality was proved for *iid* and strong mixing processes; and consistency was shown specifically for infinite order moving averages, simple bilinear and ARCH(1) processes; and in general for stochastic recurrence equations (e.g. GARCH), and processes approximable by a finite-dependent process. See Mason (1982), Hall (1982), Davis and Resnick (1984), Hall and Welsh (1984), Haeusler and Teugels (1985), Rootzén *et al* (1990), Hsing (1991), Resnick and Stărică (1995, 1998), and de Haan and Resnick (1998).

Hsing (1991), in a seminal paper, develops an asymptotic theory under remarkably general sufficient conditions, and exemplifies the theory by proving asymptotic normality for strong mixing processes. The sufficient conditions include the existence of probability and distribution limits for nonlinear tail arrays $\{U_{n,t}, U_{n,t}^*\}$ based on $\{X_t\}$ (see Section 3), and for $L(x)$ to be restricted in the manner of Smith (1982) and Goldie and Smith (1987). It is not obvious when the law of large numbers or central limit theorem for $\{U_{n,t}, U^*_{n,t}\}$ will hold outside the strong mixing case, and in general concrete cases for the slowly varying component $L(x)$ will ultimately have to be considered (see, e.g., Haeusler and Teugels, 1985).

Mixing assumptions are not practical for every financial and macroeconomics context because they either do not hold, or a mixing condition holds only under hash conditions. Examples include linear distributed lags (Andrews, 1984), long memory processes (Guegan and Ladoucette, 2001), and general nonlinear GARCH processes (e.g. Carrasco and Chen, 2002).

Moreover, there are no details in the literature on how to characterize the asymptotic variance of $\hat{\alpha}_m^{-1}$ in general, without specifying a parametric model or exploiting a mixing property (e.g. Hsing, 1991).

Our starting point is the fact that the Hill estimator only utilizes sample information from the extreme tail of the distribution. Using a generalization of the D-mixing property in Leadbetter *et al* (1983), we define an "extremal" mixing base. We then re-define the mixingale and near-epoch-dependence properties to hold exclusively in the extreme support of the distribution.

We prove the Hill estimator is consistent for processes with extremes that form mixingale sequences, and asymptotically normal for processes with extremes that are near-epoch-dependent on the extremes of a mixing base. The extremal-mixingale property implies the extremal-NED property, which characterizes the memory of infinitely many L_p-NED processes satisfying (1), $p > 0$, hence strong mixing, ARFIMA, FIGARCH, and bilinear processes, and many stochastic recurrence equations. See Section 3.4, and see Basrak *et al* (2002a,b), Davidson (2004) and Hill (2006a). The generality afforded by the extremal-NED property is important if we wish to analyze $\{X_t\}$ itself, rather than a pre-filtered series based on a possibly mis-specified model, or a filter that erodes information reflecting tail shape[1].

We then prove a non-parametric kernel estimator of the asymptotic variance of $\hat{\alpha}_m^{-1}$ is consistent for extremal-NED processes. Therefore an underlying structure (e.g. GARCH) that may affect the parametric form of the limiting variance need not be specified. See Hill (2006b) for a broad simulation study that demonstrates the exceptional merits of the kernel estimator for ARMA$(1,\infty)$, GARCH$(1,1)$, IGARCH and Davidson's (2004) long-memory Hyperbolic-GARCH processes. See, also, Smith and Weissman (1994), Poon *et al* (2002), and the citations therein for alternative data cluster techniques to improve standard error estimation.

A useful consequence of our results is a complete limit theory for the intermediate tail quantile estimator $X_{(m+1)}$, and tail array sums $m^{-1}\sum_{t=1}^{n}(\ln X_t - b_n(m))_+$ and $m^{-1}\sum_{t=1}^{n} I(X_t > b_n(m))_+$ for some increasing threshold $b_n(m) \to \infty$ as $n \to \infty$. Limit theory for such tail array sums and a general tail empirical

[1] For example, GARCH processes are known to have regularly varying tails: see Basrak *et al* (2002a,b). The scaled residuals $\{\hat{\epsilon}_t/\hat{\sigma}_t\}$ of a GARCH time series $X_t = \sigma_t \epsilon_t$, however, may have subtantially thinner tails than the original series itself, and need not have regularly varying tails (e.g. $\epsilon_t \overset{iid}{\sim} N(0,1)$).

quantile process has been developed for *iid*, mixing, and q-dependent processes in Leadbetter *et al* (1983), Hsing (1991,1993) and Drees (2002).

Criteria for selecting the sample tail fractile m is considered in Hall (1982), Resnisk (1996), Resnick and Stărică (1997), Danielsson *et al* (1998), and Drees *et al* (2000). Consult McLeish (1975), Hall and Heyde (1980), Gallant and White (1988) and Davidson (1994) for details on population dependence properties. For dependence theory associated with univariate and multivariate extremes, see, for example, Loynes (1964), de Haan and Resnick (1977), Leadbetter *et al* (1983), and Ledford and Tawn (1996).

In Section 2 we define extremal dependence properties. Section 3 contains assumptions, the main results, and examples of processes that are covered by our theory. Appendix 1 contains proofs of the main results and Appendix 2 contains preliminary lemmas.

Through $\rightarrow$ denotes convergence in probability and $\Rightarrow$ denotes convergence in distribution. Let $[x]$ denote the smallest integer larger than x..

2. EXTREMAL DEPENDENCE

Let $\{\epsilon_t\}$ be a stochastic process on some probability measure space (Ξ, G, μ_Ξ), with σ-algebra $G = \cup_t G_t$. Denote by $\{a_{n_t}\}_{t=1}^n$ a sequence of thresholds $a_{n_t} \rightarrow \infty$ as $n \rightarrow \infty$ for each t, and define the σ-sub-algebra associated with extremes of ϵ_t:

$$\digamma_{n,s}^t \equiv \sigma(\epsilon_\tau \ > \ a_{n_\tau} : 1 \leq s \leq \tau \leq t \leq n).$$

Define the following coefficients, where $1 \leq q < n$:

$$\varepsilon_{n,q} \equiv \sup_{1 \leq t \leq n} \sup_{A \in \digamma_{n,1}^t, B \in \digamma_{n,t+q}^n} |P(A \cap B) - P(A)P(B)|$$
$$\varpi_{n,q} \equiv \sup_{1 \leq t \leq n} \sup_{A \in \digamma_{n,1}^t, B \in \digamma_{n,t+q}^n} |P(A|B) - P(A)|.$$

E-Mixing *If* $\sup_n (n/m)\varepsilon_{n,q} = O(q^{-\lambda-\iota})$ *we say* $\{\epsilon_t\}$ *is Extremal-Strong Mixing with size* $-\lambda < 0$*, for some tiny* $\iota > 0$*. If* $\sup_n (n/m)\varpi_{n,q} = O(q^{-\lambda-\iota})$ *we say* $\{\epsilon_t\}$ *is Extremal-uniform mixing with size* $-\lambda$*.*

Remark: The E-strong mixing property is only a slight generalization of Leadbetter *et al*'s (1983) D-mixing property. Because $\digamma_{n,t} \in G_t$ it is easy to show a mixing process is E-mixing. Similarly, well known inequalities easily apply. Cf. Ibragimov (1962) and Serfling (1968).

Now let $x_n : \mathbb{R} \rightarrow \mathbb{R}_+$ be a sequence of functions satisfying $x_n(u) \rightarrow \infty$ as $n \rightarrow \infty$ for arbitrary $u \in \mathbb{R}$. Consider extremal versions of the mixingale and near-epoch-dependence concepts.

L_p**-E-MIX** *The sequence* $\{X_t, \digamma_{n,t}\}$ *is an* L_p*-Extremal-Mixingale with size* $-\lambda$ *if* $x_n(u)$ *is chosen such that*

$$\|P(X_t > x_n(u)) - P(X_t > x_n(u)|\digamma_{n,t-q})\|_p \leq \tilde{e}_{n,t}^*(u)\tilde{\varphi}_{n,q}^*$$
$$\left\|P(X_t > x_n(u)|\Im_{t-q}^{t+q}) - P(X_t > x_n(u)|\digamma_{n,t+q})\right\|_p \leq \tilde{e}_{n,t}^*(u)\tilde{\varphi}_{n,q+1}^*,$$

where $\tilde{e}^*_{n,t} : \mathbb{R} \to \mathbb{R}_+$ *is Lebesgue measurable,* $\sup_t \tilde{e}^*_{n,t}(u) = O((m/n)^{1/p})$ *for each* $u \in \mathbb{R}$, *and* $\limsup_n \tilde{\varphi}^*_{n,q} = O(q^{-\lambda-\iota})$ *for some tiny* $\iota > 0$.

L_p**-E-NED** $\{X_t\}$ *is* L_p*-Extremal-NED on* $\{F_{n,t}\}$ *with size* $-\lambda$ *if* $x_n(u)$ *is chosen such that*

$$\left\| P(X_t > x_n(u)|\Im_{t-q}^{t+q}) - P(X_t > x_n(u)|F_{n,t-q}^{t+q}) \right\|_p \leq \tilde{f}^*_{n,t}(u)\tilde{\psi}^*_{n,q},$$

where $\tilde{f}^*_{n,t} : \mathbb{R} \to \mathbb{R}_+$ *is Lebesgue measurable,* $\sup_t \tilde{f}^*_{n,t}(u) = O((m/n)^{1/r})$ *for each* $u \in \mathbb{R}$, *and* $\limsup_n (n/m)^{1/p-1/r}\tilde{\psi}^*_{n,q} = O(q^{-\lambda-\iota})$ *for some* $r \geq p$.

Remark 1: The combined $O((m/n)^{1/p})$ rate is based on a simple upper bound by observing $P(X_t > b_n(m)e^u)^{1/p} = O((m/n)^{1/p})$, where $b_n(m)$ is defined by (3). We split the E-NED rate $\tilde{f}^*_{n,t}(u) \times \tilde{\psi}^*_{n,q} = O((m/n)^{1/p})$ into the two parts in order to simplify proofs. This split is always possible given $r \geq p$ and $m/n \to 0$, hence we will not comment on it further.

Remark 2: Infinitely many L_p-NED processes $\{X_t\}$ that satisfy (1) are L_2-E-NED. See Lemma B.1 of Hill (2006a)

Remark 3: The E-NED property is remotely related to Resnick and Stărică's (1998) approximability condition (2.14). We assume the extremal-event $I(X_t > x_n(u))$ is approximable by the finite-lag process $P(X_t > x_n(u)|F_{n,t-q}^{t+q})$, where $x_n(u) \to \infty$ as $n \to \infty$. They assume $\lim_{n\to\infty}(n/m)P(|X_t - X_t^{(q)}| > \epsilon) \to 0$ as $q \to \infty$ *for all* ϵ, where $X_t^{(q)}$ is a stationary, q-dependent process. Thus, they do not restrict dependence to the extremes, *per se*. Moreover, they only prove consistency for the Hill estimator.

3. MAIN RESULTS

Assumption A details the required tail and memory properties. Consult Section 3.4 for examples of processes that satisfy the following restrictions.

Assumption A The common marginal distribution F satisfies (1) with support on $(0, \infty)$. Moreover, one of the following holds:

1. The sequence $\{X_t, F_{n,t}\}$ is L_2-E-MIX of size $-1/2$. Let $\tilde{e}^*_{n,t}(u) = O((m/n)^{1/2})$ be square integrable with respect to Lebesgue measure on $\mathbb{R}_+$: in particular $(\int_0^\infty \tilde{e}^*_{n,t}(u)^2 du)^{1/2} = O((m/n)^{1/2})$.

2. $\{X_t\}$ is L_2-E-NED on $\{\epsilon_t\}$ of size $-1/2$. Let $\tilde{f}^*_{n,t}(u) = O((m/n)^{1/r})$ be square integrable with respect to Lebesgue measure on $\mathbb{R}_+$: in particular $(\int_0^\infty \tilde{f}^*_{n,t}(u)^2 du) = O((m/n)^{1/r})$. The base $\{\epsilon_t\}$ is E-uniform mixing with size $-r/[2(r-1)]$, $r \geq 2$; or E-strong mixing with size $-r/(r-2)$, $r > 2$.

3.1 Consistency

Assume there exists a sequence of positive numbers $\{b_n(m)\}_{n\geq 1}$ satisfying

(3) $\quad (n/m)P(X_t > b_n(m)) \to 1.$

See Leadbetter *et al* (1983: Theorem 1.7.13). Define the following centered threshold processes:

$$(4) \qquad \begin{aligned} \{U_{n,t}\} &\equiv \left\{(\ln X_t - \ln b_n(m))_+ - E[(\ln X_t - \ln b_n(m))_+]\right\} \\ \left\{U^*_{n,t}(\rho, u)\right\} &\equiv \{I(X_t > b_n(\rho m)e^u) - E[I(X_t > b_n(\rho m)e^u)]\} \end{aligned}$$

for any $u \in \mathbb{R}$ and any ρ in an arbitrary neighborhood of one. Throughout we write $\{U_{n,t}, U^*_{n,t}\} = \{U_{n,t}, U^*_{n,t}(\rho, u)\}$.

Our first result presents laws of large numbers for the tail array sums $m^{-1}\sum_{t=1}^{n} U_{n,t}$ and $m^{-1}\sum_{t=1}^{n} U^*_{n,t}$ and the intermediate tail quantile estimator $X_{(m+1)}$. A proof that the Hill estimator is consistent in the extremal-mixingale case will then easily follow.

LEMMA 1 *Let $\{X_t\}$ satisfy Assumptions A.1, let ρ be in an arbitrary neighborhood of 1 and let $u \in \mathbb{R}$ be arbitrary. Then for any sequence satisfying $m \sim n^{\delta}$, $\delta \in (0,1)$,*

$$m^{-1}\sum_{t=1}^{n} U_{n,t} \to 0, \quad m^{-1}\sum_{t=1}^{n} U^*_{n,t} \to 0, \quad \left|\ln X_{([\rho m])} - \ln b_n(\rho m)\right| \to 0.$$

THEOREM 2 *Under the conditions of Lemma 1, $\hat{\alpha}_m^{-1} \to \alpha^{-1}$.*

Remark: L_p-mixingale sequences and L_2-E-NED, L_p-NED, E-strong mixing and strong mixing processes are all special cases of L_2-E-MIX sequences.

Proof of Theorem 2. The limit $\hat{\alpha}_m^{-1} \to \alpha^{-1}$ is an immediate consequence of Lemma 1 and Hsing's (1991) Theorem 2.2. ■

3.2 Asymptotic Normality

Let $m = o(n)$, consider $\{U_{n,t}\}$ from (4), and now write

$$\{U^*_{n,t}\} = \{U^*_{n,t}(u/\sqrt{m})\} \equiv \{I(X_t > b_n(m)e^{u/\sqrt{m}}) - E[I(X_t > b_n(m)^{u/\sqrt{m}})]\}.$$

Hsing (1991: Theorem 2.4) proves that if the joint process $\{U_{n,t}, U^*_{n,t}\}$ satisfies

$$m^{-1/2}\left(\sum_{t=1}^{n} U_{n,t},\ \alpha^{-1}\sum_{t=1}^{n} U^*_{n,t}\right) \Rightarrow (Y_1, Y_2)$$

in distribution for some random vector (Y_1, Y_2), $L(\lambda x)/L(x) \to 1$ as $x \to \infty$ fast enough and $|\ln X_{([\rho m])} - \ln b_n(\rho m)| \to 0$, then the Hill estimator satisfies

(5) $\quad \sqrt{m}(\hat{\alpha}_m^{-1} - \alpha^{-1}) \Rightarrow Y_1 - Y_2.$

In Lemma 4 we completely characterize the joint distribution of (Y_1, Y_2) assuming $\{X_t\}$ is E-NED on an E-mixing base. We then restrict the rate $L(\lambda x)/L(x) \to 1$ in Assumption B and derive a Gaussian limit law for $\hat{\alpha}_m^{-1}$ in Theorem 5.

The proof of Lemma 4 uses a standard big-block/little-block argument pioneered by Bernstein (1927). Define the sequences g_n, l_n and r_n as follows: $g_n \to \infty$ as $n \to \infty$ such that

$$
\begin{aligned}
&g_n/n \to 0, \;\; g_n = o(m^{1/4})), \;\; r_n = [n/g_n] \\
&1 \le l_n \le g_n - 1 \le n-1, \quad l_n/g_n \to 0.
\end{aligned}
\tag{6}
$$

The restriction $g_n = o(m^{1/4})$ is always possible and merely expedites several proofs.

Now define the asymptotic variance: for any $\omega \in \mathbb{R}^2$, $\omega \neq 0$,

$$
\sigma_n^2(\omega) \equiv E\left(m^{-1/2} \sum\nolimits_{t=1}^{n} \left[\omega_1 U_{n,t} + \omega_2 \alpha^{-1} U_{n,t}^*\right]\right)^2 . \tag{7}
$$

Because $\omega \in \mathbb{R}^2$ we variously write $\sigma_n^2(\omega) \equiv \sigma_n^2(\omega_1, \omega_2)$. For example

$$
\sigma_n^2(1,-1) \equiv E(m^{-1/2} \sum\nolimits_{t=1}^{n} [U_{n,t} - \alpha^{-1} U_{n,t}^*]).
$$

Define

$$
\begin{aligned}
&TT_{n,t} = TT_{n,t}(\omega, u/\sqrt{m}) \equiv m^{-1/2} [\omega_1 U_{n,t} + \omega_2 \alpha^{-1} U_{n,t}^*(u/\sqrt{m})]/\sigma_n(\omega) \\
&Z_{n,i} \equiv \sum\nolimits_{t=(i-1)g_n + l_n + 1}^{i g_n} TT_{n,t} \; .
\end{aligned}
\tag{8}
$$

Lemma 4 effectively proves $\sum_{i=1}^{r_n} Z_{n,i} \Rightarrow N(0,1)$ for E-NED processes $\{X_t\}$ by exploiting a result due to de Jong (1997), and shows the remaining term is $o_p(1)$. Consult Hall and Heyde (1980), Hsing (1991), Davidson (1994), and Rootzén *et al* (1998) for similar arguments.

The subsequent assumption rules out asymptotic degeneracy.

Assumption B Let $\inf_{n \ge 1} \sigma_n^2(\omega) \ge A$ for some finite $A > 0$ for every $\omega \neq 0$.

LEMMA 3 *Let Assumptions A.2 and B hold. For each $\omega \in \mathbb{R}^2$, $\omega \neq 0$, the sequence $\{TT_{n,t}(\omega, u/\sqrt{m}), \digamma_{n,t}\}$ is an L_2-mixingale with size $-1/2$ and constants $cc_{n,t} = O(n^{-1/2})$. Moreover, $\sum_{i=1}^{r_n} Z_{n,i}^2 \to 1$.*

LEMMA 4 *If $\{X_t\}$ satisfies Assumptions A.2 and B, then for any sequence satisfying $m = o(n)$*

$$
\sum\nolimits_{t=1}^{n} TT_{n,t}(\omega, u/\sqrt{m}) \Rightarrow N(0,1)
$$

point-wise in $\omega \in \mathbb{R}^2$ and $u \in \mathbb{R}$. Moreover $\sqrt{m}(\ln X_{(m+1)} - \ln b_n(m))/\breve{\sigma}_n \Rightarrow N(0,1)$, where $\breve{\sigma}_n^2 \equiv E(\sqrt{m}(\ln X_{(m+1)} - \ln b_n(m)))^2$ and $|\breve{\sigma}_n^2 - \sigma_n^2(0,1)| \to 0$.

Remark: Lemma 3 states $TT_{n,t}(\omega, u/\sqrt{m})$ is an L_2-mixingale with size $-1/2$ and constants $cc_{n,t} = O(n^{-1/2})$. We deduce $\sigma_n^2(\omega) = O(\sum_{t=1}^n cc_{n,t}^2) = O(1)$ from a well-known bound for L_2-mixingales due to McLeish (1975). Hence $\breve{\sigma}_n^2 = O(1)$ due to $|\breve{\sigma}_n^2 - \sigma_n^2(0,1)| \to 0$.

In order to characterize the limiting distribution of $\hat{\alpha}_m^{-1}$ we must restrict the tail form of F. This is handled by appealing to conditions characterized in Goldie and Smith (1987). See Smith (1982), Haeusler and Teugels (1985) and Hsing (1991).

Assumption C For some positive measurable function g on $(0, \infty)$ such that

(SR1) $L(\lambda x)/L(x) - 1 = O(g(x))$ as $x \to \infty$.

Moreover, g has bounded increase: there exists $0 < D, z_0, \tau < \infty$ such that $g(\lambda z)/g(z) \leq D\lambda^{\tau}$ some for $\lambda \geq 1$, $z \geq z_0$. Specifically, assume $\tau \leq 0$ and $\sqrt{m}g(b_n(m)) \to 0$.

Condition (SR1) implies both $\bar{F}(b_n(m)e^u)/\bar{F}(b_n(m)) = e^{-\alpha u}(1 + O(g(x)))$ and $\sqrt{m}(\sum_{t=1}^n E(\ln X_t - \ln b_n(m))_+/m - \alpha^{-1}) = o(1)$. Both of these properties are indispensable to Hsing's (1991: Theorem 2.4) generic proof of distribution convergence (5), which we exploit below. Other means to restrict $L(\lambda x)/L(x) \to 1$ are certainly available. See, e.g., condition (SR2) in Goldie and Smith (1982), and see Haeusler and Teugels (1985) and Hsing (1991) for use of (SR2) to derive uncentered limits laws for $\hat{\alpha}_m^{-1}$. We restrict attention to (SR1) for the sake of brevity[2].

THEOREM 5 *Let $m = o(n)$. If $\{X_t\}$ satisfies Assumptions A.2, B, and C, then*

$$\sqrt{m}\left(\hat{\alpha}_m^{-1} - \alpha^{-1}\right)/\sigma_n \Rightarrow N(0,1),$$

where $\sigma_n^2 \equiv E(\sqrt{m}(\hat{\alpha}_m^{-1} - \alpha^{-1}))^2$ and $|\sigma_n^2 - \sigma_n^2(1,-1)| \to 0$.

Remark 1: From remark 1 of Lemma 4 we similarly deduce $\sigma_n^2 = O(1)$.

Remark 2: The limit $\sqrt{m}(\hat{\alpha}_m - \alpha)/\tilde{\sigma}_n \Rightarrow N(0,1)$, $\tilde{\sigma}_n^2 = \alpha^4\sigma_n^2$ follows easily. If $\{X_t\}$ is *iid* then $\lim_{n\to\infty} \sigma_n^2 = \alpha^{-2}$ and $\lim_{n\to\infty} \tilde{\sigma}_n^2 = \alpha^4 \lim_{n\to\infty} \sigma_n^2 = \alpha^4\alpha^{-2} = \alpha^2$. See Hall (1982).

Remark 3: The rate $m = o(n)$ will have to be made explicit depending on how the tail probability $\bar{F}(x)$, and therefore the slowly varying component $L(x)$, is specified to satisfy (SR1). For example, if

$$\bar{F}(x) = cx^{-\alpha}(1 + O(x^{-\theta}))$$

[2] Examples of tails satisfying (SR1) include $\bar{F}(x) = cx^{-\alpha}(1 + O(x^{-\theta}))$ and $\bar{F}(x) = cx^{-\alpha}(1 + O((\ln x)^{-\theta}))$ for some $c > 0$, $\alpha > 0$ and $\theta > 0$. See Haeusler and Teugels (1985) for these and other examples.

cf. Hall(1982), then $\sqrt{m}g(b_n(m)) \to 0$ holds only if $m = o(n^{2\theta/(2\theta+\alpha)})$. See Haeusler and Teugels (1985). The above tail structure has been widely assumed in the econometrics and statistics literatures. See Hall (1982), Cline (1983), Chan and Tran (1989), Loretan and Phillips (1994), Caner (1998) and Hill (2005) to name a very few.

Proof of Theorem 5. Recall $\sigma_n^2(\omega)$ defined in (7) and define

$$\begin{aligned}\sigma_{n,1}^2 &\equiv \sigma_n^2(1,0) = E\left(m^{-1/2}\sum\nolimits_{t=1}^n U_{n,t}\right)^2\\ \sigma_{n,2}^2 &\equiv \sigma_n^2(0,1) = E\left(m^{-1/2}\sum\nolimits_{t=1}^n \alpha^{-1}U_{n,t}^*(u/\sqrt{m}))\right)^2.\end{aligned}$$

Lemma 4 and a Crámer-Wold device suffice to prove

$$(9) \quad m^{-1/2}\left(\sum\nolimits_{t=1}^n U_{n,t}/\sigma_{n,1},\, \alpha^{-1}\sum\nolimits_{t=1}^n U_{n,t}^*(u/\sqrt{m})/\sigma_{n,2}\right) \Rightarrow (Z_1, Z_2)$$

for some random vector (Z_1, Z_2) with marginal distributions $Z_i \sim N(0,1)$. Therefore, Lemma 4, (9) and the continuous mapping theorem together imply

$$\begin{aligned}(10) \qquad & m^{-1/2}\sum\nolimits_{t=1}^n (U_{n,t} - \alpha^{-1}U_{n,t}^*(u/\sqrt{m}))/\sigma_n(1,-1)\\ &= \frac{\sigma_{n,1}}{\sigma_n(1,-1)}\frac{1}{\sqrt{m}}\sum\nolimits_{t=1}^n U_{n,t}/\sigma_{n,1}\\ &\quad - \frac{\sigma_{n,2}}{\sigma_n(1,-1)}\alpha^{-1}\frac{1}{\sqrt{m}}\sum\nolimits_{t=1}^n U_{n,t}^*(u/\sqrt{m})/\sigma_{n,2}\\ &\Rightarrow \left(\lim_{n\to\infty}\frac{\sigma_{n,1}}{\sigma_n(1,-1)}\right)Z_1 - \left(\lim_{n\to\infty}\frac{\sigma_{n,2}}{\sigma_n(1,-1)}\right)Z_2 \sim N(0,1).\end{aligned}$$

Exploiting $|\ln X_{([\rho m])} - \ln b_n(\rho m)| \to 0$ for any ρ in a neighborhood of 1, cf. Theorem 2, and using Assumption C, (9) and (10), an argument identical to Hsing's (1991: p. 1553-1554)[3] proves

$$\begin{aligned}(11) \qquad & \sqrt{m}\left(\hat{\alpha}_m^{-1} - \alpha^{-1}\right)/\sigma_n(1,-1)\\ &\Rightarrow \left(\lim_{n\to\infty}\frac{\sigma_{n,1}}{\sigma_n(1,-1)}\right)Z_1 - \left(\lim_{n\to\infty}\frac{\sigma_{n,2}}{\sigma_n(1,-1)}\right)Z_2 \sim N(0,1).\end{aligned}$$

Finally, given the definition $\sigma_n^2 \equiv E(\sqrt{m}(\hat{\alpha}_m^{-1} - \alpha^{-1}))^2$ and (11), $|\sigma_n^2(1,-1) - \sigma_n^2| \to 0$ by Crámer's Theorem. ■

3.3 Variance Estimator

In general the parametric form of the variance of $\hat{\alpha}_m^{-1}$ will depend upon an underlying parametric structure (e.g. FIGARCH). Our next task is to derive a kernel estimator that side-steps such parametric issues, at least for E-NED processes.

[3] See, especially, equations (2.3a) and (2.4)-(2.8) of Hsing (1991). Assumption C implies Hsing's (2.3a) holds; Lemma 4 implies Hsing's (2.4) holds.

Define a kernel estimator of the asymptotic variance

(12) $\hat{\sigma}_n^2 = m^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} w((s-t)/\gamma_n)\hat{Z}_s \hat{Z}_t$

where $\hat{Z}_t \equiv [(\ln X_t / X_{(m+1)})_+ - (m/n)\hat{\alpha}_m^{-1}]$, and $w((s-t)/\gamma_n)$ denotes a standard kernel function with bandwidth $\gamma_n \to \infty$ as $n \to \infty$, $w(0) = 1$ and $w(z) = w(-z)$. The estimator satisfies $\hat{\sigma}_n^2 > 0$ almost surely for kernels defined by Assumption 1 of de Jong and Davidson (2000), including Barlett, Parzen, Quadratic Spectral and Tukey-Hanning kernels. See, also, Newey and West (1987), Gallant and White (1988), Andrews (1991) and Hansen (1992).

THEOREM 6 *Let $m = o(n)$ and $m/n^{1/2} \to \infty$. Let $w_{s,t,n} \equiv w((s-t)/\gamma_n)$ satisfy Assumption 1 of de Jong and Davidson (2000) with bandwidth $\gamma_n \to \infty$* as $n \to \infty$, $\gamma_n = O(n^{\varsigma})$, $\varsigma \in (0,1)$. *In particular $\gamma_n/m = o(n^{-1/2})$ and $\sum_{s,t=1}^{n} |w_{n,s,t}| = O(n\gamma_n)$. Under the conditions of Theorem 5, $|\hat{\sigma}_n^2 - \sigma_n^2| \to 0$.*

Remark 1: Subsequent to Remark 1 of Theorem 1, the kernel variance estimator for $\hat{\alpha}_m$ is $\hat{\hat{\sigma}}_n^2 = \hat{\alpha}_m^4 \hat{\sigma}_n^2$.

Remark 2: The number of tail observations m must increase sufficiently fast relative to both the bandwidth γ_n and sample size n to ensure consistency. For example, the largest bandwidth is $\gamma_n \sim m^{1/2-\iota}$ for some infinitessimal $\iota > 0$ because we then require $m \sim n^{1-\kappa}$ for some infinitessimal $\kappa > 0$ to ensure $\gamma_n/m = o(n^{-1/2})$. The restriction $m/n^{1/2} \to \infty$ implies some tail structures characterized by (SR1) are not covered here. For example, the tail shape $\bar{F}(x) = cx^{-\alpha}(1 + O((\ln x)^{-\theta}))$ is excluded because $m = o((\ln n)^{2\theta})$ is required: see Haeusler and Teugels (1985).

3.4 Examples

We briefly discuss processes that have regularly varying tails (1) and the L_2-E-NED properties of Assumption A.2. In this cases the Hill estimator is consistent and asymptotically normal, provided Assumptions B and C hold.

3.4.1 Linear and Power-ARCH Processes Consider a linear process $X_t = \sum_{i=0}^{\infty} \theta_i Z_{t-i}$, or power-ARCH($\infty$) process $X_t = \sigma_t Z_t$, $\sigma_t^p = \sum_{i=0}^{\infty} \theta_i |X_{t-i}|^p$, $p > 0$, $Z_t \overset{iid}{\sim} (1)$. In both cases θ_i may decay geometrically ("short" memory) or hyperbolically ("long" memory), covering ARFIMA, FIGARCH and Davidson's (2004) long-memory Hyperbolic-GARCH processes. Cline (1983) shows $\sum_{i=0}^{\infty} \theta_i Z_{t-i}$ satisfies (1) with index α. The ARCH(∞) process $X_t = \sigma_t Z_t$ satisfies (1) with index α, and under mild restrictions both distributed lag processes satisfy the L_2-E-NED Assumption A.2. See Lemmas B.4-B.6 of Hill (2006a).

3.4.1 Simple Bilinear Processes Consider the model $X_t = \beta X_{t-1} Z_{t-1} + Z_t$, $Z_t \overset{iid}{\sim} (1)$, $P(Z_t > 0) = 1$, and $\beta > 0$ satisfies $\beta^{\alpha/2} E[Z_t^{\alpha/2}] < 1$. Davis and Resnick (1996: Corollary 2.4) prove $\{X_t\}$ has regularly tails (1) with index

$\gamma = \alpha/2$, and Resnick and Stărică (1998) prove $\hat{\gamma}_m^{-1}$ is consistent for $2/\alpha$. By Lemma B.7 of Hill (2006a), $\{X_t\}$ satisfies Assumption A.2.

APPENDIX 1: PROOFS OF MAIN RESULTS

Proof of Lemma 1. Under the maintained assumptions $\{U_{n,t}, \mathcal{F}_{n,t}\}$ and $\{U^*_{n,t}(\rho,u), \mathcal{F}_{n,t}\}$ are L_2-mixingales with size $-1/2$ and constants

$$\{d_{n,t}, d^*_{n,t}\} = \left\{ \left(\int_0^\infty \tilde{e}^*_{n,t}(u)^2 du \right)^{1/2}, \ \tilde{e}^*_{n,t}(u) \right\} = O((m/n)^{1/2}).$$

See Lemma A.1. Recall $m \sim n^\delta$, $\delta \in (0,1)$, define

$$a_t \equiv t^{(1+\delta)/2}(m/n)^{1/2}$$

and notice $a_n = n^{(1+\delta)/2}(m/n)^{1/2} = n^{\delta/2}m^{1/2} \sim m$. Thus, for some finite $K > 0$ each $d_t \in \{d_{n,t}, d^*_{n,t}\}$ satisfies

$$\sum_{t=1}^\infty (d_t/a_t)^2 \leq K \sum_{t=1}^\infty t^{-(1+\delta)} < \infty.$$

Thus $\sum_{t=1}^n U_{n,t}/a_n \to 0$ $a.s.$ and $\sum_{t=1}^n U^*_{n,t}/a_n \to 0$ $a.s.$ by Corollary 20.16 of Davidson (1994) and $a_n/m \to 1$. The limit $|\ln X_{([\rho m])} - \ln b_n(\rho m)| \to 0$ now follows from an argument in Hsing (1991: p. 1551). ■

Proof of Lemma 3. Recall $\{TT_{n,t}(\omega, u/\sqrt{m})\}$ in (8):

$$\{TT_{n,t}(\omega, u/\sqrt{m})\} \equiv m^{-1/2}(\omega_1 U_{n,t} + \omega_2 \alpha^{-1} U^*_{n,t}(u/\sqrt{m}))/\sigma_n(\omega).$$

We show $\{TT_{n,t}, \mathcal{F}_{n,t}\} = \{TT_{n,t}(\omega, u/\sqrt{m}), \mathcal{F}_{n,t}\}$ is for each $\omega \in \mathbb{R}^2$ and $u \in \mathbb{R}$ an L_2-mixingale.

The limit $\sum_{i=1}^{r_n} Z^2_{n,i} \to 1$ then follows from Lemmas A.3 and A.4 in Appendix 2 by mimicking arguments in de Jong (1997: A.39-A.41).

By Assumption A.2 and Lemma A.1, $\{U_{n,t}, U^*_{n,t}\}$ are L_2-NED with coefficients $\tilde{\psi}^*_{n,q}$ and constants

$$\{d_{n,t}, d^*_{n,t}\} = \left\{ \left(\int_0^\infty \tilde{f}^*_{n,t}(u)^2 du \right)^{1/2}, \ \tilde{f}^*_{n,t}(u) \right\} = O((m/n)^{1/r}).$$

Theorems 17.5 and 17.8 of Davidson (1994) then imply $\{TT_{n,t}\}$ is an L_2-mixingale with size $-1/2$. If the base $\{\epsilon_t\}$ is E-strong mixing, Ibragimov's (1962) inequality implies for some $r > 2$

$$\begin{aligned}
&\|TT_{n,t} - E[TT_{n,t}|\mathcal{F}_{n,t-q}]\|_2 \\
&\quad \leq \max\left\{ \|TT_{n,t}\|_r, \ m^{-1/2}\sigma_n^{-1}(\omega) \times 2 \times \max\{|\omega_1| d_{n,t}, |\omega_2|\alpha^{-1} d^*_{n,t}\} \right\} \\
&\qquad \times \max\left\{ 6\varepsilon_{n,q}^{1/2-1/r}, \ \tilde{\psi}^*_{n,q} \right\}.
\end{aligned}$$

From Lemma A.2 $||TT_{n,t}||_r = O(m^{-1/2}(m/n)^{1/r})$. Hence, for some finite $K > 0$

$$\begin{aligned}&\|TT_{n,t} - E[TT_{n,t}|\digamma_{n,t-q}]\|_2 \\ &\quad \leq Kn^{-1/2} \times \max\{[(n/m)\varepsilon_{n,q}]^{1/2-1/r},\ (n/m)^{1/2-1/r}\tilde{\psi}^*_{n,q}\} = cc_{n,t}\psi_{n,q},\end{aligned}$$

where $cc_{n,t} = O(n^{-1/2})$, and $\sup_n \psi_{n,q} = O(q^{-1/2-\iota})$ for some small $\iota > 0$ by the E-mixing and E-NED coefficient properties.

Analogous arguments suffice to show $||TT_{n,t} - E[TT_{n,t}|\digamma_{n,t+q}]||_2 \leq cc_{n,t}\psi_{n,q+1}$ (e.g. Davidson, 1994: eq. 17.19), and to handle the E-uniform mixing case. ■

Proof of Lemma 4. By Lemma 3 $\{TT_{n,t}(\omega, u/\sqrt{m}), \digamma_{n,t}\}$ is an L_2-mixingale with coefficients $\psi_{n,q}$ of size $-1/2$ and constants $cc_{n,t} = O(n^{-1/2})$.

Step 1: In order to prove $\sum_{t=1}^{n} TT_{n,t}(\omega, u/\sqrt{m}) \Rightarrow N(0,1)$ we will show conditions (a)-(f) of the Lemma 1 central limit theorem of de Jong (1997) apply.

Recall the sequences g_n, l_n and r_n and block $Z_{n,i} = \sum_{t=(i-1)g_n+l_n+1}^{ig_n} TT_{n,t}(\omega, u/\sqrt{m})$ defined in (6) and (8), and define the σ-field

$$\tilde{F}_{n,i} \equiv \sigma(\{\epsilon_{n,\tau} \geq a_{n_\tau} : \tau \leq ig_n\}).$$

Condition (a): By Lemma A.2 and Minkowski's inequality

$$\left\|\sum_{t=r_n g_n+1}^{n} TT_{n,t}\right\|_2 \leq (n - r_n g_n)\|TT_{n,t}||_2 = O(m^{-1/2}(m/n)^{1/2}) = o(1).$$

Condition (b): Using a bound for L_2-mixingales with size $-1/2$ (see McLeish, 1975),

$$\begin{aligned}E\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)g_n+1}^{(i-1)g_n+l_n} TT_{n,t}\right)^2 &= O\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)g_n+1}^{(i-1)g_n+l_n} cc^2_{n,t}\right) \\ &= O(r_n l_n n^{-1}) = O(l_n/g_n) = o(1).\end{aligned}$$

Condition (c): Analogous to de Jong (1997: A.7-A.12), it can be shown that for some tiny $\eta > 0$

$$\begin{aligned}&E\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)g_n+l_n+1}^{ig_n} E\left[Z_{n,i}|\tilde{F}_{n,i-1}\right]\right)^2 \\ &= O\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)g_n+l_n+1}^{ig_n} cc^2_{n,t}\psi^{2\eta}_{n,l_n}\right) \\ &= O(r_n g_n n^{-1} l_n^{-\eta}) = O(l_n^{-\eta}) = o(1).\end{aligned}$$

Condition (d): The argument here mimics the verification of condition (c).

Condition (e): Analogous to de Jong (1997: A.13-A.17), if $\sum_{i=1}^{r_n} Z^2_{n,i} \to 1$

then

$$\left\|\sum_{i=1}^{r_n} Z_{n,i}^2 - \sum_{i=1}^{r_n}\left(E\left[Z_{n,i}|\tilde{F}_{n,i}\right] - E\left[Z_{n,i}|\tilde{F}_{n,i-1}\right]\right)^2\right\|_1$$
$$\leq 3\sum_{i=1}^{r_n}\left\|Z_{n,i} - \left(E[Z_{n,i}|\tilde{F}_{n,i}] - E[Z_{n,i}|\tilde{F}_{n,i-1}]\right)\right\|_2 \|Z_{n,i}\|_2$$
$$= O\left(\sum_{i=1}^{r_n}\left(\sum_{t=(i-1)g_n+l_n+1}^{ig_n} cc_{n,t}^2\psi_{n,l_n}^{2\eta}\right)^{1/2}\left(\sum_{t=(i-1)g_n+l_n+1}^{ig_n} cc_{n,t}^2\right)^{1/2}\right)$$
$$= O(r_n\left(g_n n^{-1} l_n^{-\eta}\right)^{1/2}\left(g_n n^{-1}\right)^{1/2}) = O(l_n^{-\eta/2}) = o(1).$$

The limit $\sum_{i=1}^{r_n} Z_{n,i}^2 \to 1$ follows from Lemma 3.

Condition (f): Define $W_{n,i} \equiv E[Z_{n,i}|\tilde{F}_{n,i}] - E[Z_{n,i}|\tilde{F}_{n,i-1}]$. We require the Lindeberg condition $\sum_{i=1}^{r_n} E[W_{n,i}^2 I(|W_{n,i}| > \varepsilon)] \to 0$ for any $\varepsilon > 0$. Choose p, $q \geq 1$ such that $1/p + 1/q = 1$. For all $\varepsilon > 0$

$$\max_i r_n E\left[W_{n,i}^2 I(|W_{n,i}| > \varepsilon)\right]$$
$$\leq \max_i r_n ||W_{n,i}||_{2p}^2 ||W_{n,i}||_q/\varepsilon$$
$$= O\left(r_n g_n^2 m^{-1}(m/n)^{1/p} g_n m^{-1/2}(m/n)^{1/q}\right)$$
$$= O\left(g_n^2 m^{-1/2}\right) = o(1).$$

The inequality follows from Hölder's and Markov's inequalities. The first equality follows from Minkowski's and the conditional Jensen's inequalities, and Lemma A.2. The last line is due to $g_n = o(m^{1/4)})$ by (6).

Step 2: From Step 1 and the definition of $\{TT_{n,t}(\omega, u/\sqrt{m})\}$ in (8), we deduce

$$\alpha^{-1}m^{-1}\sum_{t=1}^{n} U_{n,t}^*(u/\sqrt{m})/\sigma_n(0,1) \Rightarrow N(0,1).$$

The limit $\sqrt{m}(\ln X_{(m+1)} - \ln b_n(m))/\sigma_n(0,1) \Rightarrow N(0,1)$ now follows from Theorem 2.4 of Hsing (1991: eq. 2.5). From the definition $\breve{\sigma}_n^2 \equiv E(m^{1/2}(\ln X_{(m+1)} - \ln b_n(m)))^2$ we conclude $|\breve{\sigma}_n^2 - \sigma_n^2(0,1)| = 1$. ■

Proof of Theorem 6. Lemmas A.5 and A.6 together imply $|\hat{\sigma}_n^2 - \sigma_n^2(1,-1)| \to 0$, and $|\sigma_n^2(1,-1) - \sigma_n^2| \to 0$ follows from the argument following (11). Thus $|\hat{\sigma}_n^2 - \sigma_n^2| \to 0$. ■

APPENDIX 2: PRELIMINARY LEMMAS

For any ρ in a neighborhood of 1, any $u \in \mathbb{R}$, and any $\omega \in \mathbb{R}^2$ define the following processes:

$$U^*_{n,t}(\rho,u) \equiv I\left(\ln X_t - \ln b_n(\rho m) > u\right) - E[I\left(\ln X_t - \ln b_n(\rho m) > u\right)]$$
$$U_{n,t} \equiv (\ln X_t - \ln b_n(m))_+ - E(\ln X_t - \ln b_n(m))_+$$
$$TT_{n,t} = TT_{n,t}(\omega, u/\sqrt{m}) \equiv m^{-1/2}\left(\omega_1 U_{n,t} + \omega_2 \alpha^{-1} U^*_{n,t}(u/\sqrt{m})\right)/\sigma_n(\omega).$$

We write $\{U_{n,t}, U^*_{n,t}\} = \{U_{n,t}, U^*_{n,t}(\rho,u)\}$ and $U^*_{n,t}(u) = U^*_{n,t}(1,u)$.

LEMMA A.1

1. *If $\{X_t, \mathcal{F}_{n,t}\}$ is an L_p-E-MIX array with coefficients $\tilde{\varphi}^*_{n,q}$ of size $-\lambda$, then $\{U_{n,t}, \mathcal{F}_{n,t}\}$ and $\{U^*_{n,t}, \mathcal{F}_{n,t}\}$ are L_p-mixingales with coefficients $\tilde{\varphi}^*_{n,q}$ and constants $\{d_{n,t}, d^*_{n,t}\} = \{(\int_0^\infty \tilde{e}^*_{n,t}(u)^p du)^{1/p}, \tilde{e}^*_{n,t}(u)\}$ provided $\tilde{e}^*_{n,t}(u)$ is p-integrable with respect to Lebesgue measure on $\mathbb{R}_+$.*

2. *If $\{X_t\}$ is L_p-E-NED on $\{\mathcal{F}_{n,t}\}$ with coefficients $\tilde{\psi}^*_{n,q}$ of size $-\lambda$, then $\{U_{n,t}, U^*_{n,t}\}$ is L_p-NED on $\{\mathcal{F}_{n,t}\}$ with coefficients $\tilde{\psi}^*_{n,q}$ and constants $\{d_{n,t}, d^*_{n,t}\} = \{(\int_0^\infty \tilde{f}^*_{n,t}(u)^p du)^{1/p}, \tilde{f}^*_{n,t}(u)\}$ provided $\tilde{f}^*_{n,t}(u)$ is p-integrable with respect to Lebesgue measure on $\mathbb{R}_+$.*

LEMMA A.2 *The tail arrays $\{U_{n,t}\}$ and $\{U^*_{n,t}(\rho,u)\}$ are L_r-bounded for any $r \geq 1$. Specifically, for every $u \in \mathbb{R}$ and any ρ in a neighborhood of 1*

$$\lim_{n\to\infty}(n/m)^{1/r}\left\|U^*_{n,t}(\rho,u)\right\|_r \leq A_r(u) < \infty$$
$$\lim_{n\to\infty}(n/m)^{1/r}\left\|U_{n,t}\right\|_r \leq B_r < \infty,\ r \geq 1$$

where the mapping $A_r : \mathbb{R} \to \mathbb{R}_+$ is q-integrable with respect to Lebesgue measure on $\mathbb{R}_+$ for any $q > 0$..

LEMMA A.3 (de Jong, 1997: Lemma 4) *If $\{TT_{n,t}, \mathcal{F}_{n,t}\}$ is an L_2-mixingale with size $-1/2$ and constants $\sup_t cc_{n,t} = O(n^{-1/2})$, then for the sequences $\{l_n, g_n, r_n\}$ defined in (6),*

$$\lim_{n\to\infty}\left|\sum_{i=1}^{r_n}\sum_{k=i+1}^{r_n}\sum_{t=(i-1)g_n+l_n+1}^{ig_n}\sum_{s=(k-1)g_n+l_n+1}^{kg_n} E\left[TT_{n,s}TT_{n,t}\right]\right| = 0.$$

LEMMA A.4 *Let Assumptions A.2 and B hold, and recall $Z_{n,i} = \sum_{t=(i-1)g_n+l_n+1}^{ig_n} TT_{n,t}$. Then $\sum_{i=1}^{r_n}(Z^2_{n,i} - E[Z^2_{n,i}]) \to 0$.*

LEMMA A.5 *Write $U^*_{n,t} = U^*_{n,t}(u/\sqrt{m}))$, recall the kernel estimator $\hat{\sigma}^2_n$ in (12) and define*

$$\tilde{\sigma}^2_n \equiv m^{-1}\sum_{s,t=1}^{n} w_{n,s,t}(U_{n,s} - (m/n)\ln X_{(m+1)}/b_n) \times (U_{n,t} - (m/n)\ln X_{(m+1)}/b_n),$$

where $w_{s,t,n} \equiv w((s-t)/\gamma_n)$ is defined in Theorem 6. Under the assumptions of Theorem 6, $|\hat{\sigma}^2_n - \tilde{\sigma}^2_n| \to 0$.

LEMMA A.6 *Write* $U^*_{n,t} = U^*_{n,t}(u/\sqrt{m}))$ *and recall* $\sigma^2_n(1,-1) \equiv E(m^{-1/2}\sum_{t=1}^n (U_{n,t} - \alpha^{-1}U^*_{n,t}))^2$. *Under the assumptions of Theorem 6,* $|\tilde{\sigma}^2_n - \sigma^2_n(1,-1)| \to 0$.

Proof of Lemma A.1. We will prove the E-NED assertion: a proof of the E-MIX claim is similar.

For any $p > 0$ let $\{X_t\}$ be L_p-E-NED on $\{\mathcal{F}_{n,t}\}$ with size $-\lambda$. By the definition of $U^*_{n,t}$ and the $\Im^{t+q}_{t-q}$-measurability of X_t,

$$(13)\qquad \left\|U^*_{n,t} - E[U^*_{n,t}|\mathcal{F}^{t+q}_{n,t-q}]\right\|_p = \left\|P(X_t > x_n(u)|\Im^{t+q}_{t-q}) - P\left(X_t > x_n(u)|\mathcal{F}^{t+q}_{n,t-q}\right)\right\|_p \le \tilde{f}^*_{n,t}(u)\tilde{\psi}^*_{n,q}.$$

We deduce from Liaponov's inequality (e.g., Corollary 9.26 of Davidson, 1994), Fubini's theorem, and (13)[4]

$$\begin{aligned}
&\left\|U_{n,t} - E[U_{n,t}|\mathcal{F}^{t+q}_{t-q}]\right\|_p \\
&= \left\|(\ln X_t - \ln b_n)_+ - E[(\ln X_t - \ln b_n)_+|\mathcal{F}^{t+q}_{n,t-q}]\right\|_p \\
&= \left\|\int_0^\infty \left[I(X_t > b_n e^u) - P\left(X_t > b_n e^u|\mathcal{F}^{t+q}_{n,t-q}\right)\right] du\right\|_p \\
&\le \left(\int_0^\infty E\left|I(X_t > b_n e^u) - P\left(X_t > b_n e^u|\mathcal{F}^{t+q}_{n,t-q}\right)\right|^p du\right)^{1/p} \\
&\le \left(\int_0^\infty \tilde{f}^*_{n,t}(u)^p du\right)^{1/p} \tilde{\psi}^*_{n,q},
\end{aligned}$$

provided $\tilde{f}^*_{n,t}(\cdot)$ is p-integrable with respect to Lebesgue measure on $\mathbb{R}_+$. ■

Proof of Lemma A.2. Exploiting (1), (2) and the construction of $b_n(\cdot)$ in (3), for any $u \in \mathbb{R}$, any ρ in an arbitrary neighborhood of one, and any $r \ge 1$

$$\begin{aligned}
&\lim_{n\to\infty}(n/m)^{1/r}||U^*_{n,t}(\rho,u)||_r \\
&\le 2\lim_{n\to\infty}(n/m)^{1/r}P\left(X_t > b_n(\rho m)e^u\right)^{1/r} \\
&= 2\lim_{n\to\infty}\left[(n/m)P\left(X_t > b_n(\rho m)\right)\frac{P\left(X_t > b_n(\rho m)e^u\right)}{P\left(X_t > b_n(\rho m)\right)}\right]^{1/r} \\
&= 2\rho^{1/r}e^{-\alpha u/r} \equiv A_r(u) < \infty.
\end{aligned}$$

Trivially $\int_0^\infty A_r(u)^q du = 2^q\rho^{q/r}\int_0^\infty e^{-\alpha uq/r}du < \infty$ for any $q > 0$. Similarly, for any $r \ge 1$ equation (1.5) of Hsing (1991) implies

$$\begin{aligned}
\lim\nolimits_{n\to\infty}(n/m)^{1/r}||U_{n,t}||_r &\le 2\lim\nolimits_{n\to\infty}(n/m)^{1/r}\|(\ln X_t - \ln b_n(m))_+\|_r \\
&= 2\left(\int_0^\infty e^{-\alpha u^{1/r}}du\right)^{1/r} \equiv B_r < \infty.
\end{aligned}$$

[4] Note $\int_0^\infty I(X_t > b_n e^u)du = \int_0^\infty I(\ln X_t/b_n > u)du = \int_0^{(\ln X_t/b_n)_+} du = (\ln X_t/b_n)_+$.

■

Proof of Lemma A.4. The following arguments borrow heavily from de Jong (1997: p. 365-366): consult that source for complete details.

For any $K \geq 0$ define the function $h_K(x)$ and process $\{\tilde{Z}_{n,i}\}$:

$$
\begin{aligned}
(14) \qquad & h_K(x) = xI(|x| \leq K) + KI(x > K) - KI(x < -K) \\
& \tilde{Z}_{n,i} \equiv h_{K/A_n}(Z_{n,i})
\end{aligned}
$$

for some sequence of constant real numbers $\{A_n\}$, $A_n \to \infty$ as $n \to \infty$, $A_n = o(m^{1/2}/g_n^2)$. For any $p, q \geq 1$, $1/p + 1/q = 1$,

$$
\begin{aligned}
\left\| \sum_{i=1}^{r_n} \left(Z_{n,i}^2 - \tilde{Z}_{n,i}^2 \right) \right\|_1 &\leq 2\sum_{i=1}^{r_n} \left\| Z_{n,i}^2 I(|Z_{n,i}| > K/A_n) \right\|_1 \\
&\leq 2A_n r_n g_n^2 \|TT_{n,t}\|_{2p}^2 g_n \|TT_{n,t}\|_q / K \\
&= O\left(A_n n g_n m^{-1} (m/n)^{1/p} g_n m^{-1/2} (m/n)^{1/q} \right) \\
&= O\left(A_n g_n^2 m^{-1/2} \right) = o(1).
\end{aligned}
$$

The first inequality follows from (14), and the second from Hölder's, Markov's and Minkowski's inequalities and stationarity. The first equality follows from Lemma A.2.

It now suffices to show

$$
\lim_{n \to \infty} \left\| \sum_{i=1}^{r_n} \left(\tilde{Z}_{n,i}^2 - E[\tilde{Z}_{n,i}^2] \right) \right\|_2 = 0.
$$

From the line of proof of Lemma 3, $\{TT_{n,t}\}$ is L_2-NED with coefficients $\tilde{\psi}_{n,q}^*$ and constants $dd_{n,t}$

$$
\tilde{\psi}_{n,q}^* = O((m/n)^{1/2-1/r}) \times O(q^{-1/2-\iota}), \quad dd_{n,t} = O(m^{-1/2}(m/n)^{1/r}).
$$

Moreover, $\{\tilde{Z}_{n,i}^2 - E[\tilde{Z}_{n,i}^2]\}$ is L_2-NED on the σ-field $\{\tilde{G}_{n,i-q}^{i+q}\}$ defined by the extremal events

$$
\tilde{G}_{n,i-q}^{i+q} = \sigma(\{\epsilon_{n,\tau} > a_{n_\tau} : (i-q-1)g_n + l_n + 1 \leq \tau \leq (i+q)g_n\}).
$$

This follows from the NED property of $\{TT_{n,t}\}$ and because $F_{n,t-ql_n}^{t+ql_n} \subseteq \tilde{G}_{n,i-q}^{i+q}$:

$$
\begin{aligned}
(15) \qquad & \left\| \tilde{Z}_{n,i}^2 - E[\tilde{Z}_{n,i}^2 | \tilde{G}_{n,i-q}^{i+q}] \right\|_2 \\
& \leq 2 \left\| h_{K/A_n}(Z_{n,i}) - h_{K/A_n}(E[Z_{n,i}|\tilde{G}_{n,i-q}^{i+q}]) \right\|_2 \times K/A_n \\
& = O\left(A_n^{-1} \sum_{t=(i-1)g_n+l_n+1}^{ig_n} \left\| TT_{n,t} - E[TT_{n,t}|F_{n,t-ql_n}^{t+ql_n}] \right\|_2 \right) \\
& = O\left(A_n^{-1} \sum_{t=(i-1)g_n+l_n+1}^{ig_n} dd_{n,t} \tilde{\psi}_{n,ql_n}^* \right) \\
& = O\left(A_n^{-1} g_n m^{-1/2} (m/n)^{1/r} (m/n)^{1/2-1/r} l_n^{-1/2-\iota} \right) \times O\left(q^{-1/2-\iota} \right) \\
& = O\left(A_n^{-1} r_n^{-1/2} \right) \times O\left(q^{-1/2-\iota} \right) = o\left(r_n^{-1/2} \right) \times O\left(q^{-1/2-\iota} \right).
\end{aligned}
$$

The last line follows by fixing $l_n = [g_n^{1/(1+2\iota)}]$, which is always possible, and by $r_n = [n/g_n]$.

Furthermore, $\{\tilde{Z}^2_{n,i}, \tilde{F}_{n,i}\}$ is an L_2-mixingale with size $-1/2$ and constants $c_{n,t} = o(r_n^{-1/2})$, where $\tilde{F}_{n,i} \equiv \sigma(\{\epsilon_{n,\tau} > a_{n_\tau} : \tau \leq ig_n\})$. If the base $\{\epsilon_{n,t}\}$ is E-strong mixing then

$$
\begin{aligned}
&\left\| E[\tilde{Z}^2_{n,i}] - E[\tilde{Z}^2_{n,i}|\tilde{F}_{n,i-2q}] \right\|_2 \\
&\quad \leq \left\| E[\tilde{Z}^2_{n,i}] - E[\tilde{Z}^2_{n,i}|\tilde{G}^{i+q}_{n,i-q}] \right\|_2 + \left\| E\left[\left(E[\tilde{Z}^2_{n,i}|\tilde{G}^{i+q}_{n,i-q}] - E[\tilde{Z}^2_{n,i}] \right) |\tilde{F}_{n,i-2q} \right] \right\|_2 \\
&\quad \leq o\left(r_n^{-1/2}\right) \times O\left(q^{-1/2-\iota}\right) + 6 \left\| \tilde{Z}^2_{n,i} \right\|_r \varepsilon^{1/2-1/r}_{n,ql_n} \\
&\quad = o\left(r_n^{-1/2}\right) \times O\left(q^{-1/2-\iota}\right) + O\left(g_n^2 m^{-1} (m/n)^{1/r} (m/n)^{1/2-1/r} l_n^{-1/2-\iota}\right) \times O\left(q^{-1/2-\iota}\right) \\
&\quad = o\left(r_n^{-1/2}\right) \times O\left(q^{-1/2-\iota}\right) + O\left(g_n m^{-1/2} r_n^{-1/2}\right) \times O\left(q^{-1/2-\iota}\right) \\
&\quad = o\left(r_n^{-1/2}\right) \times O\left(q^{-1/2-\iota}\right).
\end{aligned}
$$

The second inequality follows from (15), and Ibragimov's (1962) inequality. The first equality follows from stationarity, Lemma A.2 and the E-mixing definition:

$$
\begin{aligned}
||\tilde{Z}^2_{n,i}||_r &\leq ||Z_{n,i}||^2_{2r} \leq g_n^2 \|TT_{n,t}||^2_{2r} = O(g_n^2 m^{-1} (m/n)^{1/r}) \\
\varepsilon^{1/2-1/r}_{n,ql_n} &= O\left((m/n)^{1/2-1/r} l_n^{-1/2-\iota}\right) \times O\left(q^{-1/2-\iota}\right).
\end{aligned}
$$

The last line follows from $l_n = [g_n^{1/(1+2\iota)}]$, $r_n = [n/g_n]$, and $g_n = o(m^{1/4})$.

A similar argument holds for $||Y_{n,i} - E[Y_{n,i}|\tilde{F}_{n,i+2q}]||_2$, and in the E-uniform mixing case. See Davidson (1994).

Finally, apply McLeish's (1975) bound for L_2-mixingales with size $-1/2$:

$$
E\left(\sum_{i=1}^{r_n} \left(\tilde{Z}^2_{n,i} - E[\tilde{Z}^2_{n,i}]\right)\right)^2 = O\left(\sum_{i=1}^{r_n} c^2_{n,t}\right) = o(1).
$$

■

Proof of Lemma A.5. Write $b_n = b_n(m)$ and define

$$
Y_{n,t} \equiv U_{n,t} - (m/n) \ln X_{(m+1)}/b_n.
$$

Decompose $\hat{\sigma}^2_n$ into

$$
\hat{\sigma}^2_n = \tilde{\sigma}^2_n + R_n,
$$

where

$$\tilde{\sigma}_n^2 = m^{-1} \sum\nolimits_{s,t=1}^{n} w_{n,s,t} Y_{n,s} Y_{n,t}$$

$$R_n = m^{-1} \sum\nolimits_{s,t=1}^{n} w_{n,s,t} A_{n,s} A_{n,t} + B_n^2 m^{-1} \sum\nolimits_{s,t=1}^{n} w_{n,s,t}$$

$$+ 2m^{-1} \sum\nolimits_{s,t=1}^{n} w_{n,s,t} A_{n,s} Y_{n,t} + 2m^{-1} B_n \sum\nolimits_{s,t=1}^{n} w_{n,s,t} Y_{n,t}$$

$$+ 2m^{-1} B_n \sum\nolimits_{s,t=1}^{n} w_{n,s,t} A_{n,t}$$

$$A_{n,t} = \left[(\ln X_t / X_{(m+1)})_+ - (\ln X_t / b_n)_+ + (m/n) \ln X_{(m+1)}/b_n\right]$$

$$B_n = (m/n) \times \left[(n/m)\left(E[(\ln X_t/b_n)_+] - \alpha^{-1}\right) + \left(\hat{\alpha}_m^{-1} - \alpha^{-1}\right)\right].$$

We need only show $\|R_n\|_1 = o(1)$.

By cases it is easy to show $|A_{n,t}| \leq |\ln X_{(m+1)}/b_n|$, and Hsing (1991: p. 1554) proves condition (SR1) of Assumption C implies

(16) $\sqrt{m}\left[(n/m)E[(\ln X_t/b_n)_+] - \alpha^{-1}\right] = o(1).$

We deduce from Lemma 4 and Theorem 5

(17) $\|A_{n,t}\|_2 \leq \left\|\ln X_{(m+1)}/b_n\right\|_2 = O(m^{-1/2})$ and $\|B_n\|_2 = O(m^{1/2}/n).$

Similarly, Lemmas 4 and A.2 imply

(18) $\|Y_{n,t}\|_2 \leq \|U_{n,t}\|_2 + (m/n)\left\|\ln X_{(m+1)}/b_n\right\|_2 = O((m/n)^{1/2}).$

Finally, the maintained assumptions imply $1/m \sum_{s,t=1}^{n} |w_{n,s,t}| = o(n^{1/2})$.

Together, the assumption $m/n^{1/2} \to \infty$, Lemma A.2, Minkowski's and the Cauchy-Schwartz inequalities, and (16)-(18) give

$$\begin{aligned}\|R_n\|_1 &= o(n^{1/2}) \times \left\{O(m^{-1}) + O(m/n^2) + O(n^{-1/2}) + O(m/n^{3/2}) + O(n^{-1})\right\} \\ &= o(n^{1/2}/m) + O(m/n^{3/2}) + o(1) + O(m/n) + O(n^{-1/2}) = o(1).\end{aligned}$$

■

Proof of Lemma A.6. Write $b_n = b_n(m)$ and define

$$\sigma_n^2(1,-1) \equiv E\left(1/\sqrt{m} \sum\nolimits_{t=1}^{n} \left(U_{n,t} - \alpha^{-1} U_{n,t}^*\right)\right)^2$$

$$\hat{\sigma}_n^2(1,-1) \equiv m^{-1} \sum\nolimits_{s,t=1}^{n} w_{n,s,t} \left(U_{n,s} - \alpha^{-1} U_{n,s}^*\right) \times \left(U_{n,t} - \alpha^{-1} U_{n,t}^*\right)$$

$$Y_{n,t} \equiv m^{-1/2}\left[U_{n,t} - (m/n) \ln X_{(m+1)}/b_n\right]$$

$$\tilde{\sigma}_n^2 \equiv \sum\nolimits_{s,t=1}^{n} w_{n,s,t} Y_{n,s} Y_{n,t}.$$

We first show $\{Y_{n,t}, \digamma_{n,t}\}$ forms an L_2-mixingale sequence with constants $cc_{n,t} = O(n^{-1/2})$. We then prove $|\tilde{\sigma}_n^2 - \sigma_n^2(1,-1)| \to 0$.

Step 1: $\{Y_{n,t}\}$ is L_2-NED on $\{\mathcal{F}_{n,t}\}$: by Minkowski's and Jensen's inequalities

$$\begin{aligned}
&\left\|Y_{n,t} - E(Y_{n,t}|\mathcal{F}_{n,t-q}^{t+q})\right\|_2 \\
&\quad \leq m^{-1/2}\left\|U_{n,t} - E(U_{n,t}|\mathcal{F}_{n,t-q}^{t+q})\right\|_2 \\
&\qquad + (m^{1/2}/n)\left\|\ln X_{(m+1)}/b_n - E\left(\ln X_{(m+1)}/b_n|\mathcal{F}_{n,t-q}^{t+q}\right)\right\|_2 \\
&\quad \leq m^{-1/2}d_{n,t}\tilde{\psi}_{n,q}^* + \alpha^{-1}(m^{1/2}/n)\left\|m^{-1}\sum\nolimits_{s=1}^{n}\left(U_{n,s}^* - E\left(U_{n,s}^*|\mathcal{F}_{n,s-q}^{s+q}\right)\right)\right\|_2 \\
&\qquad + 2n^{-1}\left(\left\|1/\sqrt{m}\sum\nolimits_{t=1}^{n}\alpha^{-1}U_{n,s}^*\right\|_2 + \left\|\sqrt{m}\ln X_{(m+1)}/b_n\right\|_2\right) \\
&\quad \leq m^{-1/2}\left(d_{n,t} + \alpha^{-1}n^{-1}\sum\nolimits_{s=1}^{n}d_{n,s}^*\right)\tilde{\psi}_{n,q}^* + o(1/n) \\
&\quad \leq Km^{-1/2}(m/n)^{1/r} \times \left[(m/n)^{1/2-1/r} \times q^{-1/2-\iota}\right] = g_{n,t}^* \times \xi_{n,q}^*,
\end{aligned}$$

say, for some finite $K > 0$. The second and third inequalities follow from the E-NED property and Lemma A.1. Lemma 4 implies the $o(1/n)$-rate.

The last inequality follows from the bound $1 \leq q < n$ and $\tilde{\psi}_{n,q}^* = O((m/n)^{1/2-1/r}) \times O(q^{-1/2-\iota})$, cf. Assumption A.2: $1/n \leq m^{-1/2}(m/n)^{1/r}(m/n)^{1/2-1/r}q^{-1/2-\iota}$ for some tiny $\iota > 0$.

Moreover, $\{Y_{n,t}, \mathcal{F}_{n,t}\}$ is an L_2-mixingale. If the base $\{\epsilon_t\}$ is E-strong mixing, then Theorem 17.5 of Davidson (1994) implies for $r > 2$

$$\|Y_{n,t} - E[Y_{n,t}|\mathcal{F}_{n,t-q}]\|_2 \leq \max\{\|Y_{n,t}\|_r, g_{n,t}^*\} \times \max\{6\varepsilon_{n,q}^{1/2-1/r}, \xi_{n,q}^*\}.$$

By Minkowski's inequality

$$\begin{aligned}
\|Y_{n,t}\|_r \quad \leq \quad & m^{-1/2}\left\|U_{n,t} - n^{-1}\sum\nolimits_{s=1}^{n}\alpha^{-1}U_{n,s}^*\right\|_r \\
& + n^{-1}\left\|\sqrt{m}\sum\nolimits_{s=1}^{n}\alpha^{-1}U_{n,s}^*/m - \sqrt{m}\ln X_{(m+1)}/b_n\right\|_r
\end{aligned}$$

The first term is $O(m^{-1/2}(m/n)^{1/r})$ by Minkowski's inequality and Lemma A.2. The second term can be shown to be $O(1/n) = O(m^{-1/2}(m/n)^{1/r})$ by exploiting Theorem 2.4 of Hsing (1991), the continuous mapping theorem and the Helly-Bray theorem. Therefore $\|Y_{n,t}\|_r = O(m^{-1/2}(m/n)^{1/r})$.

Now use $g_{n,t}^* = O(m^{-1/2}(m/n)^{1/r})$ and $\xi_{n,q}^* = O((m/n)^{1/2-1/r}) \times O(q^{-1/2-\iota})$ and repeat the remaining steps in the line of proof of Lemma 3 to get

$$\|Y_{n,t} - E[Y_{n,t}|\mathcal{F}_{n,t-q}]\|_2 \leq cc_{n,t} \times \psi_{n,q} = O(n^{-1/2}) \times O(q^{-1/2-\iota}).$$

Step 2 ($|\tilde{\sigma}_n^2 - \sigma_n^2(1,-1)| \to 0$): We will first verify Assumptions 1-3 of de Jong and Davidson (2000) to show

$$\left|\tilde{\sigma}_n^2 - E\left(\sum\nolimits_{t=1}^{n}Y_{n,t}\right)^2\right| \to 0. \tag{19}$$

Their Assumption 1 holds by the statement of the lemma.

By Step 1 $\{Y_{n,t}, \mathcal{F}_{n,t}\}$ is an L_2-mixingale with size $-1/2$ and constants $cc_{n,t}^2 = O(n^{-1/2})$. Thus Assumption 2 is satisfied[5].

Assumption 3 is satisfied by $\gamma_n \max_{1\leq t\leq n} cc_{n,t}^2 = O(n^{-(1-\varsigma)}) = o(1)$ given $\gamma_n = O(n^{\varsigma})$, $\varsigma \in [0,1)$. This proves (19).

Finally, Theorem 2.4 of Hsing (1991) and the continuous mapping theorem can be used to show

(20)

$$\left| E\left(\frac{1}{\sqrt{m}}\sum\nolimits_{t=1}^{n} U_{n,t} - \sqrt{m}\ln\frac{X_{(m+1)}}{b_n}\right)^2 - E\left(\frac{1}{\sqrt{m}}\sum\nolimits_{t=1}^{n} U_{n,t} - \frac{\alpha^{-1}}{\sqrt{m}}\sum\nolimits_{t=1}^{n} U_{n,t}^*\right)^2 \right|$$
$$= \left| E\left(\sum\nolimits_{t=1}^{n} Y_{n,t}\right)^2 - \sigma_n^2(1,-1)\right| \rightarrow 0.$$

Together, (19) and (20) imply $|\tilde{\sigma}_n^2 - \sigma_n^2(1,-1)| \rightarrow 0$, as claimed. ■

References

Akgiray, Vedat and G. Geoffrey Booth (1988). The Stable Law Model of Stock Returns, Journal of Business and Economic Statistics 6, 51-57.

Andrews, D.W.K. (1984). Non-Strong Mixing Autoregressive Processes, Journal of Applied Probability 21, 930-934.

Andrews, D.W.K. (1988). Laws of Large Numbers for Dependent, Non-Identically Distributed Random Variables, Econometric Theory 4, 458-467.

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, Econometrica 59, 817-858.

Basrak, B., R.A. Davis, and T, Mikosch (2002a). Regular Variation of GARCH Processes, Stoch. Proc. and their Appl. 99, 95-115.

Basrak, B., R.A. Davis, and T, Mikosch (2002b). A Characterization of Multivariate Regular Variation, The Annals of Applied Probability 12, 908-920.

Beirlant, J., D. Dierckx, and A. Guillou (2005). Estimation of the Extreme Value Index and Generalized Plots, Bernoulli 11, 949-970.

Berstein, S. (1927). Sur L'Extension de Théorème du Calcul des Probabilités aus Sommes de Quantités Dependants, Mathematische Annalen 97, 1-59.

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

Caner, M. (1998). Tests for Cointegration with Infinite Variance Errors, Journal of Econometrics 86,155-175.

[5] Notice (2.6) of de Jong and Davidson (2000) is only sufficient for the mixingale property to hold, but not necessary. From Step 1 we know $\{Y_{n,t}\}$ is L_2-NED on $\{\mathcal{F}_{n,t}\}$ with constants and coefficients satisfying $g_{n,t}^* \times \xi_{n,q}^* = O(m^{-1/2}(m/n)^{1/r}) \times O((m/n)^{1/2-1/r}) \times O(q^{-1/2-\iota}) = O(n^{-1/2}) \times O(q^{-1/2-\iota})$, and the sequence $\{Y_{n,t}, \mathcal{F}_{n,t}\}$ is an L_2-mixingale with constants and coefficients also satisfying $cc_{n,t} \times \psi_{n,q} = O(n^{-1/2}) \times O(q^{-1/2-\iota})$. With these properties in hand, each of de Jong and Davidson's arguments that exploit their (2.6) go through.

Carrasco, M. and X. Chen (2002). Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, Econometric Theory 18, 17-39.

Chan, N.H., and L.T. Tran (1989). On the First Order Autoregressive Process with Infinite Variance, Econometric Theory 5, 354-362.

Chan, N.H., S.D. Deng, L. Peng, and Z. Xia (2005). Interval Estimation for the Conditional Value-at-Risk Based on GARCH with Heavy Tailed Innovations, Journal of Econometrics (forthcoming).

Cheng, B.N. and S.T. Rachev (1995). Multivariate Stable Futures Prices, Mathematical Finance 5, 133-153.

Cline, D. B. H. (1983). Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data, Unpublished Ph.D. Dissertation, Department of Statistics, Colorado State University.

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (1998) Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation, mimeo, Eramus University Rotterdam.

Davidson, J. (1994). Stochastic Limit Theory. Oxford Univ. Press: Oxford.

Davidson, J. (2004). Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model, Journal of Business and Economics Statistics 22, 16-29.

Davis, R., and S. Resnick (1984). Tail Estimates Motivated by Extreme Value Theory, Annals of Statistics 12, 1467-1487.

Davis, R., and S. Resnick (1996). Limit Theory for Bilinear Processes with Heavy-Tailed Noise, Annals of Applied Probability 6, 1191–1210.

de Haan, L. (1970). On Regular Variation and Its Applications to the Weak Convergence of Sample Extremes, MC Tract 32, Mathematisch Centrum, Amsterdam.

de Haan, L., and S. Resnick (1998). On Asymptotic Normality of the Hill Estimator, Stochastic Models 14, 849–867.

de Jong, R.M. (1997). Central Limit Theorems for Dependent Heterogeneous Random Variables, Econometric Theory 13, 353-367.

de Jong, R.M., and J. Davidson (2000). Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices, Econometrica 68, 407-423.

Drees, H. (2002). Tail Empirical Processes under Mixing Conditions, in H. Dehling, T. Mikosch, and M. Sørensen's (eds.) Empirical Process Techniques for Dependent Data. Birkhauser: Berlin.

Drees, H., Ferreira A, and L. de Haan (2004). On Maximum Likelihood Estimation of the Extreme Value Index, Annals of Applied Probability 14, 1179-1201.

Feller, William (1971). An Introduction to Probability Theory and its Applications, 2 ed., vol. 2. Wiley: New York.

Gallant, A. R. and H. White (1988). A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell: Oxford.

Goldie, C.M., and R.L. Smith (1987). Slow Variation with Remainder: Theory and Applications, Quarterly Journal of Mathematics 38, 45-71.

Guegan D., and S. Ladoucette (2001). Non-mixing Properties of Long Memory Processes, Comptes Rendus de l'Academie des Sciences Series I Mathematics 333, 373-376.

Haan, L. de, and S. Resnick (1977). Limit Theory for Multivariate Sample Extremes, Z. Wahr. verw. Geb. 40, 317-37.

Haeusler, E., and J. L. Teugels (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation, Annals of Statistics 13, 743-756.

Hall, P. (1982). On Some Estimates of an Exponent of Regular Variation, Journal of the Royal Statistical Society 44, 37-42.

Hall, P., and A.H. Welsh (1985). Adaptive Estimates of Parameters of Regular Variation, Annals of Statistics 13, 331-341.

Hall, P., and C. C. Heyde (1980). Martingale Limit Theory and Its Applications. Academic Press: New York.

Hansen, B. (1992). Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes, Econometrica 60, 976-972.

Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, Annals of Mathematical Statistics 3, 1163-1174.

Hill, J.B. (2005). Gaussian Tests of "Extremal White Noise" for Dependent, Heterogeneous, Heavy Tailed Time Series, Dept. of Economics, Florida International University; available at http://econwpa. wustl.edu:80/eps/em/ papers/0411/0411014.pdf.

Hill, J.B. (2006a). Technical Appendix for "On Tail Index Estimation for Dependent, Heterogeneous Data," Dept. of Economics, Florida International University; available at www.fiu.edu/∼hilljona/tech_append_tail_estim.pdf.

Hill, J.B. (2006b). Robust Tail Inference for Dependent, Heterogeneous Time Series, Dept. of Economics, Florida International University; available at www.fiu.edu /∼hilljona /robust_tail_inference.pdf.

Hsing, T. (1991). On Tail Index Estimation Using Dependent Data, Annals of Statistics 19, 1547-1569.

Ibragimov, I. A. (1962). Some Limit Theorems for Stationary Processes, Theory of Probability and its Applications 7, 349-382.

Ibragimov, I. A. and Y. V. Linnik (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhof.

Kearns, P. and A. Pagan (1997). Estimating the Density Tail Index for Financial Time Series, The Review of Economics and Statistics 79, 171-175.

Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag: New York.

Leadbetter, M.R., and H. Rootzén (1993). On Central Limit Theory for Families of Strongly Mixing Additive Set Functions. In Festschrift in Honour of Gopinath Kallianpur, eds. S. Cambanis and P. K. Sen, p. 211-223. Springer: New York.

Ledford, A. W. and J. A. Tawn (1996). Statistics for Near Independence in Multivariate Extreme Values, Biometrika 83, 169-187.

Loretan, M., and P.C.B. Phillips (1994). Testing the Covariance Stationarity of Heavy-

Tailed Economic Time Series: An Overview of the Theory with Applications to Financial Data Sets, Journal of Empirical Finance 1, 211-48.

Loynes, R.M. (1964). Extreme Values in Uniformly Mixing Stationary Stochastic Processes, Annals of Mathematical Statistics 36, 993–999.

Mason, D. (1982). Laws of Large Numbers for Sums of Extreme Values, Annals of Probability 10, 754-764.

McLeish, D.L. (1975). A Maximal Inequality and Dependent Strong Law, Annals of Probability 3, 329-339.

Newey, W.K., and K.D. West (1987). A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703-708.

Pickands, J. (1975). Statistical-Inference using Extreme Order Statistics, Annals of Statistics 3, 119-131.

Poon, S., M. Rockinger and J.A. Tawn (2002). Modelling Extreme-Value Dependence in International Stock Markets, working paper, University of Strathclyde.

Rachev, S. T. (2003). Handbook of Heavy Tailed Distributions in Finance. Elsevier Science: New York.

Resnick, S. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag: New York.

Resnick, S., and H. Rootzén (2000). Self-Similar Communication Models and Very Heavy Tails, Annals of Applied Probability 10, 753–778.

Resnick, S., and C. Stărică (1995). Consistency of Hill's Estimator for Dependent Data, Journal of Applied Probability 32, 139-167.

Resnick, S. and C. Stărică (1997). Smoothing the Hill Estimator, Advances in Applied Probability 29, 271-293.

Resnick, S. and C. Stărică (1998). Tail Index Estimation for Dependent Data, The Annals of Applied Probability 8, 1156-1183.

Rootzén, H., Leadbetter, M., and L. de Haan (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.

Rootzén, H., Leadbetter, M., and L. de Haan (1998). On the Distribution of Tail Arrays Sums for Strongly Mixing Stationary Sequences, Annals of Probability 8, 868-885.

Serfling, R.J. (1968). Contributions to Central Limit Theory for Dependent Variables, Annals of Mathematical Statistics 39, 1158-1175.

Smith, R.L. (1987). Uniform Rates of Convergence in Extreme-Value Theory, Advances in Applied Probability 14, 600-622.

Smith, R.L. (1987). Estimating Tails of Probability-Distributions, Annals of Statistics 15, 1175-1207.

Smith, R.L. and I. Weissman (1994). Estimating the Extremal Index, Journal of the Royal Statistical Society, Series B 56, 515-528.