Econometric Theory, **26**, 2010, 1398–1436.
doi:10.1017/S0266466609990624

ON TAIL INDEX ESTIMATION FOR DEPENDENT, HETEROGENEOUS DATA

JONATHAN B. HILL
University of North Carolina–Chapel Hill

In this paper we analyze the asymptotic properties of the popular distribution tail index estimator by Hill (1975) for dependent, heterogeneous processes. We develop new extremal dependence measures that characterize a massive array of linear, nonlinear, and conditional volatility processes with long or short memory. We prove that the Hill estimator is weakly and uniformly weakly consistent for processes with extremes that form mixingale sequences and asymptotically normal for processes with extremes that are near epoch dependent (NED) on some arbitrary mixing functional. The extremal persistence assumptions in this paper are known to hold for mixing, L_p-NED, and some non-L_p-NED processes, including ARFIMA, FIGARCH, explosive GARCH, nonlinear ARMA-GARCH, and bilinear processes, and nonlinear distributed lags like random coefficient and regime-switching autoregressions.

Finally, we deliver a simple nonparametric estimator of the asymptotic variance of the Hill estimator and prove consistency for processes with NED extremes.

1. INTRODUCTION

This paper develops an asymptotic theory for the popular distribution tail index estimator due to B.M. Hill (1975) under general conditions. Many time series in finance, macroeconomics, and meteorology exhibit extreme values that appear to cluster (Leadbetter, Lindgren, and Rootzén, 1983; Embrechts, Klüppelberg, and Mikosch, 1997). In order to deliver a Gaussian limit theory that is robust to the nature of persistence and heterogeneity in extremes, we introduce new extremal dependence measures and develop an associated weak and uniform limit theory for dependent, heterogeneous tail arrays.

Denote by $\{X_t\} = \{X_t : -\infty < t < \infty\}$ a stochastic process on some probability measure space, write $F_t(x) := P(X_t \leq x)$, and assume F_t has support on $[0, \infty)$. Assume $\bar{F}_t(x) := P(X_t > x)$ is regularly varying at ∞: for all $\lambda > 0$ and each t,

$$\lim_{x\to\infty} \frac{\bar{F}_t(\lambda x)}{\bar{F}_t(x)} = \lambda^{-\alpha}, \tag{1}$$

I would like to thank Enno Mammen for insight into the extremal dependence properties developed here, Oliver Linton for discussions concerning the strong-GARCH case, and Holger Drees for comments on an earlier version. In particular, I kindly thank three anonymous referrees, co-editor Bruce Hansen, and editor Peter C.B. Phillips for expert commentary that led to substantial improvements. All errors, of course, are mine alone. Address correspondence to Jonathan B. Hill, Dept. of Economics, University of North Carolina–Chapel Hill; e-mail: jbhill@email.unc.edu.

  0266-4666/10 $15.00

where $\alpha > 0$ denotes the index of regular variation. Equivalently,

$$\bar{F}_t(x) = x^{-\alpha} L(x), \qquad x > 0, \quad \text{where } L(x) \text{ is slowly varying.} \tag{2}$$

The distribution class (2) 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 Bingham, Goldie, and Teugels (1987), Resnick (1987), and Basrak, Davis, and Mikosch (2002).

Let $X_{(i)} > 0$ denote the ith order statistic of a sample path $\{X_t\}_{t=1}^n$ with sample size $n \geq 1$, $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$, and let $\{m_n\}$ be an intermediate order sequence: $1 \leq m_n < n$, $m_n \to \infty$ as $n \to \infty$, and $m_n = o(n)$. B.M. Hill's (1975) estimator of α^{-1} is simply the average log-exceedance

$$\hat{\alpha}_{m_n}^{-1} := \frac{1}{m_n} \sum_{t=1}^{n} \left(\ln \left(X_t / X_{(m_n+1)} \right) \right)_+ = \frac{1}{m_n} \sum_{i=1}^{m_n} \ln \left(X_{(i)} / X_{(m_n+1)} \right),$$

where $(z)_+ := \max\{z, 0\}$. 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), Quintos, Fan, and Phillips (2001), Resnick and Rootzén (2000), Chan, Deng, Peng, and Xia (2007), and Hill (2008b), to name a few. For alternative estimation techniques consult Pickands (1975), Smith (1987), Rootzén, Leadbetter, and de Haan (1990), Smith and Weissman (1994), Drees, Ferreira, and de Haan (2004), Csörgö and Viharos (1995), Beirlant, Dierckx, and Gaillou (2005), and Iglesias and Linton (2008).

We are interested in the asymptotic properties of $\hat{\alpha}_{m_n}^{-1}$ under minimal but verifiable conditions. Asymptotic normality has been established for i.i.d., strong mixing, and l-dependent approximable sequences including GARCH(1,1) processes; and consistency was shown for l-dependent approximable sequences, infinite order moving averages, bilinear, ARCH(1), and stochastic recurrence equations (e.g., GARCH). See Mason (1982), Hall (1982), Davis and Resnick (1984), Hall and Welsh (1984), Haeusler and Teugels (1985), Rootzén et al. (1990), Hsing (1991, 1993), Resnick and Stărică (1995, 1998), de Haan and Resnick (1998), and Quintos et al. (2001).

Hsing (1991) develops an asymptotic theory under remarkably general conditions and proves asymptotic normality for strong mixing processes. Sufficient conditions include restrictions on tail decay (2) and the existence of probability and distribution limits for nonlinear tail arrays based on $\{X_t\}$ (see Section 2). It is not obvious whether such limit theory holds beyond the strong mixing case and $\{m_n\}$ is intimately tied to tail decay.

Mixing properties are convenient because functions of mixing random variables are mixing, and a well-established limit theory exists (e.g., Ibragimov and Linnik, 1971). Nevertheless, it is typically difficult to verify a mixing condition, and many time series are not mixing, or are mixing only under harsh conditions. Infinite order distributed lags, for example, need not be mixing due to density smoothness requirements, including ARFIMA, nonlinear ARMA-GARCH, and some long memory processes (see Gorodetskii, 1977; Andrews, 1984; Guegan and Ladoucette, 2001; Carrasco and Chen, 2002; and Wu, 2005).

The near epoch dependence (NED) property (Ibragimov, 1962; Ibragimov and Linnik, 1971; Gallant and White, 1988), however, has substantial practical advantages because it only requires computation of a conditional expectation, it is typically easy to verify, it carries over to a large class of functions of NED random variables, and powerful central limit theory is available (Davidson, 1992; de Jong, 1997). NED characterizes any mixing process, infinite order distributed lags of a mixing process, and many nonmixing processes, since density smoothness is irrelevant (Davidson, 1994, 2004). McLeish's (1975) broader mixingale concept is advantageous for theoretical reasons: Processes that are NED on a mixing process form mixingale sequences that satisfy useful inequalities and laws of large numbers, and mixingales decompose to martingale differences for which central limit theory is available. A related conditional moment-based concept, L_p-weak dependence, and associated central limit theory are treated in Wu (2005) and Wu and Min (2005). NED and L_p-weak dependence appear to cover many of the same processes, where neither seems to dominate the other.

In a purely extreme value theoretic environment, however, the analyst may not want to commit to superfluous assumptions involving nonextremes. Leadbetter (1974) and Leadbetter et al. (1983) provide some relief with a so-called D-mixing property for serial extremes, but the property does not necessarily carry over to arbitrary functions of D-mixing random variables (see Section 2).

Further, there are no details in the literature on how to characterize the asymptotic variance of $\hat{\alpha}_{m_n}^{-1}$ in general without specifying a parametric model or exploiting independence or a mixing property (e.g., Hall, 1982; Hsing, 1991).

In Section 2 we control for memory and heterogeneity in extremes by introducing extremal versions of mixingale and NED properties. By exploiting primitive results in Hsing (1991), we prove in Section 3 that $\hat{\alpha}_{m_n}^{-1}$ and the intermediate order statistic $X_{(m_n+1)}$ are weakly and uniformly weakly consistent by assuming that extremes of $\{X_t\}$ form mixingale arrays and delivering new uniform laws for tail arrays. See Hall and Welsh (1985) for uniform consistency of the Hill estimator for i.i.d. data and Smith (1982) for uniform convergence of sample maxima of i.i.d. data. We then prove that $\hat{\alpha}_{m_n}^{-1}$ is asymptotically normal when $\{X_t\}$ has extremes that are NED on a mixing functional of some arbitrary process $\{\epsilon_t\}$.

The generality afforded by an extremal version of NED is important if we wish to analyze X_t itself, rather than a prefiltered series based on a possibly misspecified model or a filter that erodes information reflecting tail shape.[1] The property characterizes a massive array of stochastic processes, including any geometrically mixing process (e.g., nonlinear GARCH with sufficiently smoothly distributed errors), both L_p-NED (e.g., ARFIMA, stationary GARCH) and non-L_p-NED (e.g., explosive GARCH) processes where underlying errors are only required to be L_p-bounded, as well as bilinear processes, and random coefficient and regime-switching autoregressions.

Finally, in Section 4 we develop a nonparametric kernel estimator of the asymptotic variance of $\hat{\alpha}_{m_n}^{-1}$ and prove consistency for processes with NED extremes. As far as we know this is the first of its kind in the extreme value theory literature. An underlying structure that may affect the parametric form of the limiting variance need not be specified (e.g., ARFIMA, GARCH, regime switching). Nevertheless, the asymptotic variance in the i.i.d. case, α^{-2}, may hold for nonidentically distributed weakly orthogonal processes, including stochastic volatility (Hill 2008a).[2]

In related work, Quintos et al. (2001) also work with results due to Hsing (1991). They deliver a functional Gaussian limit for $\hat{\alpha}_{m_n}^{-1}$ for GARCH(1,1) processes by extending Hsing's (1991, Cor. 3.3) proof of asymptotic normality for tail mixing data. See Section 2.2 below for a definition of tail mixing. Quintos et al. (2001) use the theory to deliver a unique structural break test with respect to the tail index. Although their approach undoubtedly extends to other processes by case, their arguments closely exploit GARCH model dynamics and rely on a case-dependent semiparametric construction of the asymptotic variance (cf. Hsing, 1991). By comparison we do not require stationarity in general, and our results cover GARCH, IGARCH, explosive GARCH, nonlinear GARCH (e.g., quadratic GARCH), and much more. Similarly, we do require any information on the asymptotic variance other than existence in order to deliver the consistent kernel estimator. See, also, Hill (2009a) for functional limit theory for D-valued, dependent heterogeneous tail arrays of the same broad class of processes covered here.

Appendix A contains proofs of the main results, Appendix B contains preliminary results, and Appendix C compiles variable definitions for quick reference.

We employ the following notation conventions: $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{a.s.}$ $a.s.$ almost sure convergence, and $\Longrightarrow$ convergence in distribution; $[x]$ denotes the integer part of x; $K > 0$ denotes an arbitrary finite constant whose value may change from line to line; $\imath > 0$ is an arbitrarily tiny constant whose value may change; and $x_n \sim y_n$ implies $x_n / y_n \to 1$.

2. EXTREMAL DEPENDENCE

Assume $\bar{F}_t(x)/\bar{F}_t(x-) \to 1$ uniformly in $t \in \mathbb{Z}$ such that there exists a sequence of positive real numbers $\{b_{m_n}\}_{n \geq 1}$ satisfying (e.g., Leadbetter et al., 1983, Thm. 1.7.13)

$$\frac{n}{m_n} P\left(X_t > b_{m_n}\right) \to 1. \tag{3}$$

We implicitly assume $\{m_n, b_m\}$ satisfy (3) for all t. Intuitively, b_{m_n} estimates the intermediate order statistic $X_{(m_n+1)}$, since $P(X_t > b_{m_n}) \sim m_n/n$ and $1/n \sum_{t=1}^{n} I(X_t > X_{(m_n+1)}) \sim m_n/n$ by construction.

Hsing (1991, Thms. 2.2, 2.4) proves under a mild second order constraint on tail decay (2) that asymptotics concerning $\hat{\alpha}_{m_n}^{-1}$ are grounded on triangular tail arrays based on tail exceedances and events:

$$\left\{ \left(\ln\left(X_t/b_{m_n}\right)\right)_+, I\left(X_t > b_{m_n}e^u\right) : 1 \le t \le n \right\}_{n\ge 1}.$$

Hsing (1991, Thm. 3.3) then imposes a mixing property on $\{(\ln(X_t/b_{m_n}))_+, I(X_t > b_{m_n}e^u)\}$ to prove $\hat{\alpha}_{m_n}^{-1}$ is asymptotically normal. We impose new tail dependence properties on $\{I(X_t > b_{m_n}e^u)\}$ that cover and substantially generalize Hsing's mixing condition.

2.1. Extremal Mixingale and Extremal NED

Let $\{\Im_{n,t}\} = \{\Im_{n,t} : 1 \le t \le n\}_{n\ge 1}$ be an increasing triangular array of σ-fields induced by some arbitrary, possibly vector-valued stochastic array $\{E_{n,t}\} = \{E_{n,t} : 1 \le t \le n\}_{n\ge 1}$:

$$\Im_{n,t} := \sigma\left(E_{n,\tau} : 1 \le \tau \le t\right).$$

Since the objects of interest $\{(\ln(X_t/b_{m_n}))_+, I(X_t > b_{m_n}e^u)\}$ are tail arrays dependent on the sample size, we restrict information to sample time periods $t \in \{1, ..., n\}$. By convention, $\Im_{n,s}^t = \{\varnothing, \Xi\}$ if $t \le 0$ or $s > n$, hence $\Im_{n,s}^t = \Im_{n,1}^t = \Im_{n,-\infty}^t$ if $s \le 0$, $\Im_{n,s}^t = \Im_{n,s}^n = \Im_{n,s}^{+\infty}$ if $t \ge n$, and $\Im_{n,s} \subset \Im_{n,t}$ $\forall 1 \le s < t \le n$.

Consider two extremal dependence properties for $\{X_t\}$ that characterize how well information induced from $\{E_{n,t}\}$ can be used to predict extreme values of $\{X_t : 1 \le t \le n\}$ as $n \to \infty$. Throughout, $\{q_n\}$ denotes an arbitrary sequence of integer displacements satisfying $1 \le q_n < n$, and $q_n \to \infty$.

Property 1. L_p-E-MIXL $\{X_t, \Im_{n,t}\}$ forms an L_p-extremal mixingale array, $p > 0$, with size $\lambda > 0$ if

$$\left\| P\left(X_t > b_{m_n}e^u\right) - P\left(X_t > b_{m_n}e^u | \Im_{n,t-q_n}\right)\right\|_p \le e_{n,t}(u) \times \varphi_{q_n}$$
$$\left\| I\left(X_t > b_{m_n}e^u\right) - P\left(X_t > b_{m_n}e^u | \Im_{n,t+q_n}\right)\right\|_p \le e_{n,t}(u) \times \varphi_{q_n+1},$$

where $e_{n,t} : \mathbb{R}_+ \to \mathbb{R}_+$ is Lebesgue measurable, $\sup_{1\le t\le n}\sup_{u\ge 0} e_{n,t}(u) = O((m_n/n)^{1/p})$, and $\varphi_{q_n} = o(q_n^{-\lambda})$.

Property 2. L_p-E-NED $\{X_t\}$ is L_p-extremal NED on $\{\Im_{n,t}\}$, $p > 0$, with size $\lambda > 0$ if

$$\left\| I\left(X_t > b_{m_n}e^u\right) - P\left(X_t > b_{m_n}e^u | \Im_{n,t-q_n}^{t+q_n}\right)\right\|_p \le f_{n,t}(u) \times \psi_{q_n},$$

where $f_{n,t} : \mathbb{R}_+ \to \mathbb{R}_+$ is Lebesgue measurable, $\sup_{1\le t\le n}\sup_{u\ge 0} f_{n,t}(u) = O((m_n/n)^{1/p})$, and $\psi_{q_n} = o(q_n^{-\lambda})$.

Remark 1. In the spirit of conventional mixingale and NED definitions, the "constants" $e_{n,t}(u)$ and $f_{n,t}(u)$ permit time dependence in the L_p-norm and allow the "coefficients" φ_{q_n} and ψ_{q_n} to be scale independent. Thus, without loss of generality, assume

$$\sup_{n\geq 1}\left\{\varphi_{q_n},\psi_{q_n}\right\}\in[0,1).$$

We say $\{X_t\}$ is geometrically L_p-E-NED if $\psi_{q_n}=o(\rho^{q_n})$ for some $\rho\in(0,1)$, in which case size $\lambda>0$ is arbitrary.

Remark 2. L_p-E-NED and L_p-E-MIXL are simply NED and mixingale properties assigned to $\{I(X_t>b_{m_n}e^u)\}$, with adjustments to scale since $I(X_t>b_{m_n}e^u)$ is asymptotically degenerate. For example, after multiplying out terms and invoking the law of iterated expectations, the L_2-E-NED property implies

$$\lim_{n\to\infty}\frac{n}{m_n}q_n^{2\lambda}\sup_{1\leq t\leq n}\sup_{u\geq 0}\left(P\left(X_t>b_{m_n}e^u\right)-E\left[P\left(X_t>b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n}\right)^2\right]\right)=0,$$

and since $(n/m_n)P(X_t>b_{m_n}e^u)\to e^{-\alpha u}$ for all t under (1)–(3),

$$\lim_{n\to\infty}q_n^{2\lambda}\sup_{1\leq t\leq n}\sup_{u\geq 0}\left(e^{-\alpha u}-\frac{n}{m_n}E\left[P\left(X_t>b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n}\right)^2\right]\right)=0.$$

Literally, $\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ when also $(n/m_n)E[P(X_t>b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n})^2]\to e^{-\alpha u}$ sufficiently fast. Thus, $f_{n,t}(u)=O((m_n/n)^{1/2})$ ensures the norm does not collapse to zero simply due to degeneracy associated with the tail fractile (or "bandwidth") $m_n\to\infty$ and $m_n=o(n)$, as opposed to (near epoch) dependence.

Remark 3. We exploit a displacement sequence $\{q_n\}$ rather than fixed q due to the degenerate nature of $I(X_t>b_{m_n}e^u)$. Unless X_t is l-dependent for finite l or the base $E_{n,t}$ is independent, in general $q_n\to\infty$ must be satisfied to be able to discern degeneracy from the ability to use $\{E_{n,\tau}\}_{t-q_n}^{t+q_n}$ to predict $I(X_t>b_{m_n}e^u)$. See comments following the proof of Lemma 2 in Appendix A. Displacement sequences have been exploited by Leadbetter (1974), Leadbetter et al. (1983), Hsing (1991, 1993) and Davis and Hsing (1995) for tail mixing properties, and de Jong (1997) for mixingale arguments associated with Bernstein block arrays. See, e.g., Ibragimov and Linnik (1971), McLeish (1975), and Gallant and White (1988) for traditional usage of fixed q.

Remark 4. If $\Im_{n,t}$ is adapted to X_t or simply $I(X_t>b_{m_n}e^u)$, then E-NED is trivial: $I(X_t>b_{m_n}e^u)-P(X_t>b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n})=I(X_t>b_{m_n}e^u)-I(X_t>b_{m_n}e^u)=0$, hence size is arbitrary.

Remark 5. A process $\{X_t\}$ is L_p-E-NED with size λ *if and only if* it is L_s-E-NED with size $\lambda p/\max\{p,s\}$ for any $s\gtreqless p$ since $|I(X_t>b_{m_n}e^u)-P(X_t>b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n})|\leq 1$ a.s. (see Hill, 2008c). But this suggests p is irrelevant, since L_p-E-NED is equivalent to L_s-E-NED. It is nevertheless convenient to assume that $\{X_t\}$ is L_2-E-NED to ensure both exceedance and event processes $\{(\ln(X_t/b_{m_n}))_+, I(X_t > b_{m_n}e^u)\}$ have the same memory property, since the two form the stochastic basis of $\hat{\alpha}_{m_n}^{-1}$.

2.2. Functional Mixing

In the E-MIXL and E-NED definitions, the σ-fields $\{\Im_{n,t}\}$ are induced by some triangular array $\{E_{n,t}\}$. We restrict persistence in $E_{n,t}$ by imposing a mixing condition. Assume $\{E_{n,t}\}$ is a possibly vector-valued functional of some process $\{\epsilon_t\}$ with σ-field

$$G_t = \sigma(\epsilon_\tau : \tau \leq t) \quad \text{and} \quad G_s^t = \sigma(\epsilon_\tau : s \leq \tau \leq t), \quad \text{where } \Im_{n,t} \subseteq G_t.$$

Let $E_{n,t} = 0$ for $t \notin \{1,..,n\}$, and the remaining $E_{n,t}$ may, for example, be some lag or lags of ϵ_t or of the extreme event $I(\epsilon_t > a_{n,t})$, peak over threshold $(\epsilon_t - a_{n,t})_+$, or extreme value $\epsilon_t I(\epsilon_t > a_{n,t})$ each for some triangular array $\{a_{n,t}\}$ of constants, $a_{n,t} \to \infty$ as $n \to \infty$. Because nonsample $E'_{n,t}s$ are constants, the associated σ-fields are trivial: $\Im_{n,s}^t = \{\varnothing, \Xi\}$ if $t \leq 0$ or $s > n$.

The generality behind $E_{n,t}$ is not vacuous, since ϵ_t may be the innovations in a parametric model like strong-GARCH, or simply $\epsilon_t = X_t$. In the former case ϵ_t is i.i.d., so any functional $E_{n,t}$ of ϵ_t is trivially mixing. In the latter case, since under mild conditions $\hat{\alpha}_{m_n}^{-1}$ is grounded on $\{(\ln(X_t/b_{m_n}))_+, I(X_t > b_{m_n}e^u)\}$, we may assume $E_{n,t} = I(\epsilon_t > b_{m_n}e^u)$ and impose a mixing condition on $E_{n,t}$ as in Hsing (1991).

Now define mixing coefficients, where $\{q_n\}$ again denotes a sequence of integer displacements, $1 \leq q_n < n$ and $q_n \to \infty$:

$$\varepsilon_{n,q_n} := \sup_{A \in \Im_{n,-\infty}^t, B \in \Im_{n,t+q_n}^{+\infty} : t \in \mathbb{Z}} |P(A \cap B) - P(A)P(B)|$$

$$\varpi_{n,q_n} := \sup_{A \in \Im_{n,-\infty}^t, B \in \Im_{n,t+q_n}^{+\infty} : t \in \mathbb{Z}} |P(B|A) - P(B)|.$$

F-Mixing. If $(n/m_n)q_n^\lambda \varepsilon_{n,q_n} \to 0$ as $n \to \infty$ we say $\{\epsilon_t\}$ is functional-strong mixing with size $\lambda > 0$. If $(n/m_n)q_n^\lambda \varpi_{n,q_n} \to 0$ as $n \to \infty$ we say $\{\epsilon_t\}$ is functional-uniform mixing with size $\lambda > 0$.

Remark 6. F-mixing on $\{\epsilon_t\}$ is simply mixing assigned to the triangular array $\{E_{n,t}\}$. There are, therefore, many variations on this concept. If, for example, $E_{n,t} = I(\epsilon_t > b_{m_n}e^u)$ and $(n/m_n)q_n^\lambda \varepsilon_{n,q_n} \to 0$, we might say $\{\epsilon_t\}$ is *extremal-strong mixing* since tail events mix asymptotically.

Remark 7. The coefficients ε_{n,q_n} and ϖ_{n,q_n} intrinsically depend on sample size n due to the triangular array nature of $\Im_{n,t}$, similar to the E-MIXL and E-NED constants $e_{n,t}(u)$ and $f_{n,t}(u)$. Mixing conditions applied to triangular arrays have a range of applications in the dependence and limit theory literatures (e.g., Andrews, 1985), in particular for sample-size dependent extremal arrays (Leadbetter, 1974; Leadbetter et al., 1983).

Remark 8. The scale $n/m_n \to \infty$ is required in general, since we use F-mixing $\{\epsilon_t\}$ as an E-NED base, and E-NED characterizes degenerate $I(X_t > b_{m_n}e^u)$. Thus, $q_n \to \infty$ must also hold since, for example, $(n/m_n)q^{\lambda}\varepsilon_{n,q} \to \infty$ is possible unless $\varepsilon_{n,q} = 0$ uniformly in n and q (e.g., $E_{n,t}$ is independent). See especially the proof of Lemma 2. In general there is much room for interpretation, since $q_n \to \infty$ is otherwise arbitrary. By σ-field dominance $\Im_{n,t} \subseteq G_t$, for example, it is easy to show that a strong mixing process $\{\epsilon_t\}$ of size 2 satisfies $\lim_{n\to\infty} q_n^2\varepsilon_{n,q_n} \to 0$. Now put $q_n = [n/m_n]$ and note that

$$\lim_{n\to\infty} [n/m_n]^2 \varepsilon_{n,[n/m_n]} = \lim_{n\to\infty} (n/m_n) q_n \varepsilon_{n,q_n} = 0$$

implies F-strong mixing of size 1. Hill (2009b, Lem. C.1) shows that asymptotically infinite order lags of F-mixing random variables are F-mixing, and standard inequalities apply, as in Ibragimov (1962) and Serfling (1968).

Remark 9. By the construction of $\{\Im_{n,t}\}$, note identically

$$\varepsilon_{n,q_n} = \sup_{A\in\Im_{n,1}^t,\, B\in\Im_{n,t+q_n}^n : 1\le t\le n-q_n} |P(A\cap B) - P(A)P(B)|$$

are Hsing's (1991, p. 1555) mixing coefficients. Using our notation, Hsing (1991) only considers the case $\epsilon_t = X_t$, $E_{n,t} = [(\ln(X_t/b_{m_n}))_+, I(X_t > b_{m_n}e^u)]'$, $q_n = o(n)$, and $(n/q_n)\varepsilon_{n,q_n} \to 0$ to prove that $\hat{\alpha}_{m_n}^{-1}$ is asymptotically normal. Since Hsing's displacement $q_n = o(n)$ is otherwise arbitrary, suppose $q_n = m_n^a$ for some $a \in (0,1)$. Then $(n/q_n)\varepsilon_{n,q_n} = (n/m_n)q_n^{(1-a)/a}\varepsilon_{n,q_n} \to 0$ implies F-strong mixing of size $(1-a)/a \in (0,\infty)$.

Remark 10. F-strong mixing is also a generalized, uniform version of Leadbetter's (1974) D-mixing concept (cf. Leadbetter et al., 1983). For any triangular array $\{\epsilon_t : 1 \le t \le n\}_{n\ge1}$ and any sequence of integers $1 \le t_1 < \cdots < t_{p_1} < s_1 < \cdots < s_{p_2} \le n$ for which $s_1 - t_{p_1} > q_n \to \infty$, define

$$\delta_{q_n} := \left| F_{t_1,\ldots,t_{p_1};\, s_1,\ldots,s_{p_2}}(a_n) - F_{t_1,\ldots,t_p}(a_n) F_{s_1,\ldots,t_{p_2}}(a_n) \right|,$$

where $F_{t_1,\ldots,t_{p_1}}(a_n) := P(\epsilon_{t_1} \le a_{n,t_1}, \ldots, \epsilon_{t_{p_1}} \le a_{n,t_{p_1}})$, $\{a_{n,t}\}$ is some deterministic array where $a_{n,t} \to \infty$ as $n \to \infty$, and p_1 and p_2 are arbitrary positive integers. Then $\{\epsilon_t\}$ is D-mixing if $\delta_{q_n} \to 0$ as $n \to \infty$. D-mixing implies joint independence of the events $\{\epsilon_i \le a_{n,i}\}_{i=1}^t$ and $\{\epsilon_i \le a_{n,i}\}_{i=t+q_n}^n$ as $n \to \infty$, and strong mixing implies D-mixing. If ϵ_t is F-strong mixing with respect to $E_{n,t} = I(\epsilon_t \le a_{n,t})$, then ϵ_t is necessarily D-mixing, since $\delta_{q_n} \le \varepsilon_{n,q_n}$ due to the sup-operator in ε_{n,q_n}. In this case D-mixing is a weaker condition, but D-mixing does not necessarily carry over to finite measurable functions of D-mixing random variables, while asymptotically infinite order lag functions of F-mixing random variables are F-mixing. In this regard F-mixing has a superlative advantage that we exploit in the proof of asymptotic normality of $\hat{\alpha}_{m_n}^{-1}$.

The following examples of F-mixing and E-NED processes are verified in Section 5:

Example 1 (Finite dependence)
Let $\{y_t\}$ be a one-sided l-dependent process for finite $l \in \mathbb{N}$. Then $X_t := |y_t|$ is trivially F-strong mixing with arbitrary size, since $\varepsilon_{n,q_n} = 0 \; \forall q_n \geq l$. If the E-NED base is simply X_t itself, and $E_{n,t} = X_t$ for $t = 1, \ldots, n$ and 0 otherwise, then $\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ where E-NED and F-mixing sizes are arbitrary.

Example 2 (Strong mixing GARCH)
Let $y_t = h_t u_t$, where u_t is i.i.d. and h_t^2 is stationary, geometrically strong mixing, and measurable with respect to $\sigma(y_\tau : \tau \leq t-1)$. Examples include linear and nonlinear GARCH processes. See Carrasco and Chen (2002) and Meitz and Saikkonen (2008) for sufficient conditions for geometric strong mixing in GARCH processes. Define $X_t := |y_t|$ and let $E_{n,t} = X_t$ for $t \in \{1, \ldots, n\}$ and 0 otherwise. Then $\{X_t\}$ is geometrically F-strong mixing by Lemma C.1 in Hill (2009b), and since $\{\Im_{n,t}\}$ is adapted to $\{X_t\}$, the E-NED property is trivial: $\{X_t\}$ is geometrically L_2-E-NED on $\{\Im_{n,t}\}$ where E-NED and F-mixing sizes are arbitrary.

Example 3 (Hsing's mixing)
Strong mixing is far stronger than actually required. Let $E_{n,t} = I(X_t > b_{m_n} e^u)$ for $t \in \{1, \ldots, n\}$ and 0 otherwise, and assume F-strong mixing coefficients ε_{n,q_n} satisfy $(n/q_n)\varepsilon_{n,q_n} \to 0$, where $q_n = m_n^a$ for any $q \in (0, 1]$. Then $\{X_t\}$ satisfies Hsing's (1991, p. 1555) mixing condition by Remark 9. But $\Im_{n,t}$ is adapted to $I(X_t > b_{m_n} e^u)$ and $(n/m_n) q_n^{(1-a)/a} \varepsilon_{n,q_n} = (n/q_n)\varepsilon_{n,q_n} \to 0$, hence $\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ with arbitrary E-NED size and F-mixing size $(1-a)/a$.

Example 4 (Nonlinear distributed lag)
Consider $y_t = \sum_{i=0}^{\infty} \pi_{t,i} \epsilon_{t-i}$, where $|\epsilon_t|$ has tail (2) with index $\alpha > 1$ and $\lim_{\epsilon \to \infty} L(\epsilon) = K$. The innovations ϵ_t are strictly stationary, uniformly $L_{\alpha-\iota}$-bounded, and strong mixing with size $r/(r-2)$, $r > 2$. The coefficients $\{\pi_{t,i}\}$ are for each i measurable with respect to $\sigma(\epsilon_\tau : \tau \leq t-i)$, strong mixing with size $r/(r-2)$, and $\sup_{t \in \mathbb{Z}} |\pi_{t,i}| \leq |\pi_i| = O(i^{-\mu})$ with probability 1 for some $\mu > 1/\min\{1, p/2\}$. Examples include regime switching and random coefficient autoregressions, and ARFIMA processes each with GARCH innovations. Assume $X_t := |y_t|$, and $E_{n,t} = [\epsilon_{t-i}]_{i=0}^{[q_n/2]}$ for $t \in \{1, \ldots, n\}$ and 0 otherwise. The lag structure of $E_{n,t}$ ensures that $\{X_t\}$ is L_2-E-NED with size 1/2 on an F-strong mixing base by ensuring that $(n/m_n)^{1/2} q_n^{1/2} ||I(X_t > b_{m_n} e^u) - P(X_t > b_{m_n} e^u | \Im_{n,t-q_n}^{t+q_n})||_2 \to 0$ for each $1 \leq t \leq n$ as $n \to \infty$.

Example 5 (Explosive GARCH)
Let $y_t = h_t \epsilon_t$, where ϵ_t is i.i.d. and $h_t^2 = \beta + \gamma y_{t-1}^2 + \delta h_{t-1}^2$, $\beta > 0$, and $\gamma, \delta \geq 0$. Write $X_t := |y_t|$ and let $E_{n,t} = [\epsilon_{t-i}]_{i=0}^{[q_n/2]}$ for $t \in \{1, \ldots, n\}$ and 0 otherwise. By independence, $\{\epsilon_t\}$ is trivially F-strong mixing with arbitrary size. If the GARCH

process has a unit root, and in many cases an explosive root, then $\{X_t\}$ is still geometrically L_2-E-NED on $\{\Im_{n,t}\}$ with arbitrary E-NED and F-mixing base sizes (Hill, 2008c), although $\{X_t\}$ itself need not be mixing nor population L_p-NED (Carrasco and Chen, 2002; Davidson, 2004).

3. MAIN RESULTS

We require two sets of assumptions concerning tail dependence and tail decay.

Assumption A.

(1) Let $\{\Im_{n,t}\}$ be an arbitrary array of σ-fields, and let $\{X_t, \Im_{n,t}\}$ form an L_2-E-MIXL array with coefficients φ_{q_n} of size $1/2$ and constants $e_{n,t}(u)$. In particular, $e_{n,t}(u)$ is integrable with respect to Lebesgue measure on $\mathbb{R}_+$ and $\sup_{1\leq t\leq n}\int_0^\infty e_{n,t}(u)\,du = O((m_n/n)^{1/2})$.

(2) $\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ with coefficients ψ_{q_n} of size $1/2$ and constants $f_{n,t}(u)$. In particular, $f_{n,t}(u)$ is integrable with respect to Lebesgue measure on $\mathbb{R}_+$ and $\sup_{1\leq t\leq n}\int_0^\infty f_{n,t}(u)\,du = O((m_n/n)^{1/2})$. The base $\{\epsilon_t\}$ is F-uniform mixing with size $r/[2(r-1)], r \geq 2$, or F-strong mixing with size $r/(r-2), r > 2$.

Remark 11. We work with the L_2-norm and assume Lebesgue integrability of $e_{n,t}(u)$ and $f_{n,t}(u)$ to ensure $\{(\ln(X_t/b_{m_n}))_+\}$ satisfies a corresponding mixingale or NED property. See Lemma B.1 in Appendix B, and see Section 5 for examples.

Remark 12. It is easy to show that L_2-E-NED Assumption A.2 ensures the L_2-E-MIXL Assumption A.1 by an argument identical to Theorem 17.5 of Davidson (1994).

In order to prove uniform consistency and characterize the limit distribution of $\hat{\alpha}_{m_n}^{-1}$, we appeal to the concept of slow variation with remainder as in condition (SR1) of Goldie and Smith (1987). See also Smith (1982), Haeusler and Teugels (1985), and Hsing (1991).

Assumption B. There exists a positive measurable function g on $(0,\infty)$ such that for any $\lambda > 0$,

$$L(\lambda x)/L(x) - 1 = O(g(x)) \quad \text{as } x \to \infty. \tag{SR1}$$

In particular, g has a bounded increase: There exist $0 < D, z_0 < \infty$, and $\tau \leq 0$ such that $g(\lambda z)/g(z) \leq D\lambda^{\tau}$ some for $\lambda \geq 1$ and $z \geq z_0$. We require m_n, b_{m_n}, and g to satisfy

$$m_n^{1/2} g(b_{m_n}) \to 0.$$

Remark 13. Assumption B implies the rate $m_n \to \infty$ must be made explicit depending on $\bar{F}_t(x)$. For example, if $\bar{F}_t(x) = cx^{-\alpha}(1 + O(x^{-\theta}))$, $\alpha, \theta > 0$, then

$m_n^{1/2} g(b_{m_n}) \to 0$ only if $m_n = o(n^{2\theta/(2\theta+\alpha)})$. See Haeusler and Teugels (1985) for this and other examples, and see, inter alia, Hall (1982), Cline (1983), Chan and Tran (1989), Caner (1998), and Hill (2008b) for applications with this tail shape. Regularly varying tails with $L(x) = c(\ln x)^\theta$, on the other hand, do not satisfy (SR1) but property (SR2) in Goldie and Smith (1987), which leads to uncentered limit laws for $\hat{\alpha}_{m_n}^{-1}$ (e.g., Haeusler and Teugels, 1985; Hsing, 1991).

3.1. Weak Consistency for E-MIXL Arrays

Uniform consistency is delivered over a parametric class of Lipschitz continuous intermediate order sequences $\{m_n(\phi)\}$, $\phi \in \Phi$, where Φ is some compact subset of $\mathbb{R}_+$.

Assumption C. Let

$$1 \leq \inf_{\phi \in \Phi} m_n(\phi) \to \infty,\; n \geq \sup_{\phi \in \Phi} m_n(\phi) = o(n),$$

$$\liminf_{n \geq 1} \left\{ \frac{m_n(\phi)}{m_n(\phi')} \right\} \geq 1 \Longleftrightarrow \phi \geq \phi', \quad \text{and} \quad m_n(\phi)^{1/2} \times g\left(b_{m_n(\phi)}\right) \to 0. \tag{4}$$

Further, for some sequence of positive numbers $\{h_n\}$, $h_n = O(\inf_{\phi \in \Phi} m_n(\phi))$, $\forall \phi, \phi' \in \Phi$,

$$\left| m_n(\phi) - m_n(\phi') \right| \leq h_n \times \left| \phi - \phi' \right|. \tag{5}$$

Remark 14. Monotonicity $m_n(\phi)/m_n(\phi') \geq 1 \Longleftrightarrow \phi \geq \phi'$ simplifies proofs and could easily be replaced with $m_n(\phi)/m_n(\phi') \geq 1 \Longleftrightarrow \phi \leq \phi'$.

Define tail arrays of X_t: For $1 \leq t \leq n$, $n \geq 1$,

$$U_{m_n,t} := \left(\ln\left(X_t/b_{m_n}\right)\right)_+ - E\left[\left(\ln\left(X_t/b_{m_n}\right)\right)_+\right]$$

$$I_{m_n,t}(u) := I\left(X_t > b_{m_n} e^u\right) - P\left(X_t > b_{m_n} e^u\right), \quad \text{for any } u \in \mathbb{R}. \tag{6}$$

The E-MIXL property suffices for tail array strong laws.

LEMMA 1. *Under Assumption A.1, for any ρ in an arbitrary neighborhood of* 1,

$$\frac{1}{m_n} \sum_{t=1}^n U_{m_n,t} \overset{a.s.}{\to} 0, \qquad \frac{1}{\rho m_n} \sum_{t=1}^n I_{\rho m_n,t}(u) \overset{a.s.}{\to} 0 \quad \textit{and} \quad \ln\left(\frac{X_{([\rho m_n])}}{b_{\rho m_n}}\right) \overset{p}{\to} 0.$$

Lipschitz continuity and Lemma 1 imply uniform strong laws for $\{(\ln(X_t/b_{m_n}))_+, I(X_t > b_{m_n} e^u)\}$ by arguments in Andrews (1992), and therefore weak uniform consistency for $\hat{\alpha}_{m_n(\phi)}^{-1}$ by arguments in Hsing (1991).

THEOREM 1.

(i) Under Assumption A.1, $\hat{\alpha}_{m_n}^{-1} \stackrel{p}{\to} \alpha^{-1}$ *for any* $1 \leq m_n < n$, $m_n \to \infty$, *and* $m_n = o(n)$.
Now let Assumptions A.1, B, and C hold.

(ii) The following limits are uniform on Φ*:*

$$\frac{1}{m_n(\phi)} \sum_{t=1}^{n} U_{m_n(\phi),t} \stackrel{p}{\to} 0, \qquad \frac{1}{m_n(\phi)} \sum_{t=1}^{n} I_{m_n(\phi),t}(u) \stackrel{p}{\to} 0,$$
$$\ln\left(\frac{X_{(m_n(\phi)+1)}}{b_{m_n(\phi)}}\right) \stackrel{p}{\to} 0.$$

(iii) Finally, $\sup_{\phi \in \Phi} |\hat{\alpha}_{m_n(\phi)}^{-1} - \alpha^{-1}| \stackrel{p}{\to} 0.$

Remark 15. Since E-NED suffices for E-MIXL, Hill's estimator is consistent for a truly massive array of time series. See Examples 1–5 and Section 5.

There are notable limitations to Assumption C.

Example 6
If $\bar{F}_t(x) = cx^{-\alpha}(1 + O(x^{-\theta}))$, $\alpha, \theta > 0$, then $m_n(\phi) \sim n^{\xi} + n^{\phi}$ satisfies (4) and (5) for any *fixed* $\xi \in (0, 2\theta/(2\theta + \alpha))$, where $\phi \in \Phi = [0, \xi_0]$ for any $\xi_0 \in (0, \xi - 2\iota]$ and tiny $\iota > 0$. This follows since $\inf_{\phi \in \Phi} m_n(\phi) \sim n^{\xi} \to \infty$, and by the mean value theorem $|m_n(\phi) - m_n(\phi')| \leq n^{\xi - \iota} \ln(n) \times |\phi - \phi'| = O(n^{\xi}) \times |\phi - \phi'|$.

Example 7
For the same tail shape, consider $m_n(\phi) \sim \phi n^{\xi}$ for any *fixed* $\xi \in (0, 2\theta/(2\theta + \alpha))$, where $\phi \in \Phi = [\phi_0, 1]$ for any $\phi_0 \in (0, 1)$. Then $|m_n(\phi) - m_n(\phi')| \leq n^{\xi}|\phi - \phi'|$ and $\inf_{\phi \in \Phi} m_n(\phi) = \phi_0 n^{\xi}$, hence (4) and (5) hold.

Example 8
Theorem 1 does not cover $m_n(\phi) \sim n^{\phi}$, $\phi \in (0, 2\theta/(2\theta + \alpha))$, for the same tail shape because Lipschitz continuity (5) with $h_n = O\left(\inf_{\phi \in \Phi} m_n(\phi)\right)$ fails to hold. Whether $\hat{\alpha}_{m_n(\phi)}^{-1}$ is uniformly weakly consistent for such $m_n(\phi)$ is left for future consideration.

3.2. Asymptotic Normality for E-NED Processes

Hsing (1991, Thm. 2.4) proves that if the tail arrays $\{U_{m_n,t}, I_{m_n,t}(u)\}$ in (6) have a joint central limit property

$$\left(\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} U_{m_n,t},\ \alpha^{-1} \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} I_{m_n,t}(u/m_n^{1/2})\right)' \Longrightarrow (Y_1, Y_2)' \qquad \textbf{(7)}$$

in distribution to some random vector (Y_1, Y_2), $L(\lambda x)/L(x) \to 1$ as $x \to \infty$ fast enough and $\ln(X_{([\rho m])}/b_{\rho m_n}) \xrightarrow{p} 0$ for ρ in any neighborhood of 1, then

$$m_n^{1/2}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right) \Longrightarrow Y_1 - Y_2.$$

We now characterize memory in $\{U_{m_n,t}, I_{m_n,t}(u)\}$ under Assumption A.2 and deliver a key tail array central limit theory for L_2-E-NED processes $\{X_t\}$.

Construct the following tail array:

$$T_{m_n,t}(\omega, u/m_n^{1/2}) := \frac{1}{m_n^{1/2}}\left[\omega_1 U_{m_n,t} + \omega_2 \alpha^{-1} I_{m_n,t}(u/m_n^{1/2})\right], \qquad \omega = [\omega_1, \omega_2]', \tag{8}$$

and variance

$$\sigma_{m_n}^2(\omega) = \sigma_{m_n}^2(\omega_1, \omega_2) := E\left(\sum_{t=1}^n T_{m_n,t}(\omega, u/m_n^{1/2})\right)^2. \tag{9}$$

The following ensures $\sigma_{m_n}^2(\omega) > 0$ uniformly in n and $\omega \neq 0$.

Assumption D. The covariance matrix of $[1/m_n^{1/2}\sum_{t=1}^n U_{m_n,t},\ 1/m_n^{1/2}\sum_{t=1}^n I_{m_n,t}(u/m_n^{1/2})]'$ is positive definite uniformly in n.

LEMMA 2. *Let Assumption A.2 hold. For each $\omega'\omega = 1$, $\{T_{m_n,t}(\omega, u/m_n^{1/2})\}$ is L_2-NED on $\{\Im_{n,t}\}$ with constants $d_{n,t} = O(m_n^{-1/2}(m_n/n)^{1/r})$ uniformly over $1 \le t \le n$, and coefficients $\psi_{n,q_n}^* = o((m_n/n)^{1/2-1/r} q_n^{-1/2})$. Further, $\{T_{m_n,t}(\omega, u/m_n^{1/2}), \Im_{n,t}\}$ forms an L_2-mixingale array with coefficients $\psi_{q_n} = o(q_n^{-1/2})$ and constants $c_{n,t} = Kn^{-1/2}$. Neither sequence of constants $\{d_{n,t}\}$ and $\{c_{n,t}\}$ depends on ω.*

The L_2-mixingale property of $\{T_{m_n,t}(\omega)\}$ under Lemma 2 and a general central limit theorem due to de Jong (1997, Lem. 1) ensure the following central limit theorem.

LEMMA 3. *Under Assumptions A.2 and D,*

(i) $\sum_{t=1}^n T_{m_n,t}(\omega, u/m_n^{1/2})/\sigma_{m_n}(\omega) \Longrightarrow N(0,1)$ *pointwise in* $\omega'\omega = 1$ *and* $u \in \mathbb{R}$, *where* $\sup_{\omega'\omega=1}\sigma_{m_n}^2(\omega) = O(1)$;

(ii) $m_n^{1/2}\ln(X_{(m_n+1)}/b_{m_n})/\sigma_{m_n}(0,1) \Longrightarrow N(0,1)$, *where* $\sigma_{m_n}^2(0,1) = \alpha^{-2}E(1/m_n^{1/2}\sum_{t=1}^n I_{m_n,t}(u/m_n^{1/2}))^2$.

Remark 16. Invoke the Cramér-Wold theorem to deduce that $1/m_n^{1/2}\sum_{t=1}^n \{(\ln(X_t/b_{m_n}))_+ - E[(\ln(X_t/b_{m_n}))_+]\}$ and $1/m_n^{1/2}\sum_{t=1}^n\{I(X_t > b_{m_n}e^u) - P(X_t >$

$b_{m_n} e^u)\}$ have Gaussian distribution limits when $\{X_t\}$ is L_2-E-NED on an F-mixing base $\{\epsilon_t\}$. See Hsing (1991, 1993), Drees (2002), Einmahl and Lin (2006), and Rootzén (2009) for related limit theory for tail arrays of i.i.d., mixing, and l-dependent processes $\{X_t\}$, each of which is covered under E-NED (Section 5).

The Lemma 3 central limit theorem does not impose any restrictions on the slowly varying component $L(x)$ in (2). The following main result relies on slow variation with remainder (SR1).

THEOREM 2. *Under Assumptions A.2, B, and D,*

$$m_n^{1/2}\left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right)/\sigma_{m_n} \Longrightarrow N(0,1),$$

where $\sigma_{m_n}^2 = E(m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}))^2 = O(1)$ *and*

$$\left|\sigma_{m_n}^2 - E\left(\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\left\{U_{m_n,t} - \alpha^{-1} I_{m_n,t}(u/m_n^{1/2})\right\}\right)^2\right| \to 0.$$

Remark 17. If $\{X_t\}$ is i.i.d. then $\lim_{n\to\infty}\sigma_{m_n}^2 = \alpha^{-2}$ (e.g., Hall, 1982).

Remark 18. Notice the mean squared error $\sigma_{m_n}^2 = E(m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}))^2$ is not necessarily the variance, since $\hat{\alpha}_{m_n}^{-1}$ is in general biased (e.g., Hall, 1982; Segers, 2002). Nevertheless, $\sigma_{m_n}^2$ is proportional to the asymptotic variance under Assumptions A.2 and B, since by Theorem 2 $m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/\sigma_{m_n} \Longrightarrow N(0,1)$.

4. KERNEL VARIANCE ESTIMATOR

In general the parametric form of the asymptotic variance $\lim_{n\to\infty}\sigma_{m_n}^2$ may depend upon underlying memory and heterogeneity properties and therefore model parameters (e.g., ARFIMA, regime switching, GARCH). Our next goal is a nonparametric estimator that sidesteps such distributional issues, at least for L_2-E-NED data. We base our estimator on the following trivial expansion:

$$\begin{aligned}
\sigma_{m_n}^2 &= m_n \times E\left(\frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{X_{(m_n+1)}}\right)\right)_+ - \alpha^{-1}\right)^2 \\
&= m_n \times E\left(\frac{1}{m_n}\sum_{t=1}^{n}\left\{\left(\ln\left(\frac{X_t}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\alpha^{-1}\right\}\right)^2 \\
&= \frac{1}{m_n}\sum_{s,t=1}^{n} E\left[\left\{\left(\ln\left(\frac{X_s}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\alpha^{-1}\right\}\right. \\
&\qquad \left.\times\left\{\left(\ln\left(\frac{X_t}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\alpha^{-1}\right\}\right].
\end{aligned}$$

It is well known that a standard estimator of the right-hand side,

$$\frac{1}{m_n}\sum_{s,t=1}^{n}\left\{\left(\ln\left(\frac{X_s}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\hat{\alpha}_{m_n}^{-1}\right\}\left\{\left(\ln\left(\frac{X_t}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\hat{\alpha}_{m_n}^{-1}\right\},$$

is not guaranteed to be positive (Newey and West, 1987). A powerful solution is a kernel estimator

$$\hat{\sigma}_{m_n}^2 = \frac{1}{m_n}\sum_{s,t=1}^{n} w_{s,t,n}\left\{\left(\ln\left(\frac{X_s}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\hat{\alpha}_{m_n}^{-1}\right\}$$
$$\times\left\{\left(\ln\left(\frac{X_t}{X_{(m_n+1)}}\right)\right)_+ - \frac{m_n}{n}\hat{\alpha}_{m_n}^{-1}\right\},$$

where $w_{s,t,n} := w((s-t)/\gamma_n)$ denotes a kernel function with bandwidth $\gamma_n \to \infty$ as $n \to \infty$, $w(0) = 1$, and $w(z) = w(-z)$. The de Jong and Davidson (2000, Assum. 1) class of kernels ensures

$$\hat{\sigma}_{m_n}^2 > 0 \text{ a.s.},$$

and includes Bartlett, Parzen, quadratic spectral, and Tukey-Hanning kernels. See also Newey and West (1987), Gallant and White (1988), and Hansen (1992).

THEOREM 3. *Let $m_n = o(n)$ and $m_n/n^{1/2} \to \infty$, and let $w_{s,t,n}$ satisfy Assumption 1 of de Jong and Davidson (2000) with bandwidth $\gamma_n \to \infty$ and $\gamma_n = o(n)$. In particular, $\gamma_n = o(m_n/n^{1/2})$ and $1/n\sum_{s,t=1}^{n}|w_{n,s,t}| = O(\gamma_n)$. Under Assumptions A.2 and B, $|\hat{\sigma}_{m_n}^2 - \sigma_{m_n}^2| \xrightarrow{p} 0$.*

Remark 19. The number of tail observations m_n must increase sufficiently fast to ensure that the plug-ins $X_{(m_n+1)}$ and $\hat{\alpha}_{m_n}^{-1}$ that appear in every cross-product of $(\ln(X_t/X_{(m_n+1)}))_+ - (m_n/n)\hat{\alpha}_{m_n}^{-1}$ in $\hat{\sigma}_{m_n}^2$ do not affect the limit. The restriction $m_n/n^{1/2} \to \infty$ implies that some tails characterized by Assumption B are not covered here, including $\bar{F}(x) = cx^{-\alpha}(1 + O((\ln x)^{-\theta}))$, because $m_n = o((\ln n)^{2\theta})$ is required (Haeusler and Teugels, 1985).

Remark 20. As few as m_n^2 pairs $\{X_s, X_t\}$ go into the construction of $\hat{\sigma}_{m_n}^2$ due to the operator $(\cdot)_+$. Thus the bandwidth rate $\gamma_n \to \infty$, which regulates the number of included cross-products in $\hat{\sigma}_{m_n}^2$, must be restricted. The bound $\gamma_n = o(m_n/n^{1/2})$ implies that the largest bandwidth allowed is $\gamma_n \sim m_n^{1/2-\iota}$ for infinitessimal $\iota > 0$ because we then require $m_n \sim n^{1-\iota} = o(n)$.

5. APPLICATIONS: L_2-E-NED

In this section we relate mixing and L_p-NED properties to L_2-E-NED and characterize processes that have the L_2-E-NED property. In particular, we want to know when Assumption A.2 holds.

5.1. Mixing Implies L_2-E-NED

If $\Im_{n,t}$ is adapted to X_t or simply $I(X_t > b_{m_n}e^u)$, then $\{X_t\}$ is trivially L_2-E-NED on $\{\Im_{n,t}\}$ with constants $f_{n,t}(u) = 0$ and coefficients ψ_{q_n} of any size, since $||I(X_t > b_{m_n}e^u) - P(X_t > b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n})||_p = ||I(X_t > b_{m_n}e^u) - I(X_t > b_{m_n}e^u)||_p = 0$. For example, suppose X_t is geometrically strong mixing and $E_{n,t} = X_t$ for $t \in \{1,\ldots,n\}$. Then $\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ with arbitrary E-NED size and $\{\Im_{n,t}\}$ is induced by a strong mixing array $\{E_{n,t}\}$ with arbitrary size due to geometric memory, so Assumption A.2 is trivial. This covers finite dependent processes and geometrically ergodic processes like nonlinear AR-nonlinear GARCH with innovations that have a sufficiently smooth density (An and Huang, 1996; Carrasco and Chen, 2002; Leibscher, 2005; Meitz and Saikkonen, 2008). See Examples 1–3 in Section 2.

5.2. L_p-NED Implies L_2-E-NED

By definition, $\{X_t\}$ is L_p-NED on $\{\Im_{n,t}\}$ with size $\lambda > 0$ if $||X_t - E[X_t|\Im_{n,t-q}^{t+q}]||_p \leq d_{n,t}\vartheta_q$ for some constants $d_{n,t} \geq 0$, coefficients $\vartheta_q = o(q^{-\lambda})$ where $q \in \mathbb{N}$ (Gallant and White, 1988). The following composite result implies that population L_p-NED implies L_s-E-NED for any $s > 0$.

LEMMA 4. *Assume X_t satisfies Assumption B.*

(i) Let $\{X_t\}$ be L_p-NED on $\{\Im_{n,t}\}$, $0 < p < \alpha$, with constants $d_{n,t}$ and coefficients ϑ_q of size $\lambda > 0$. If the slowly varying component $\lim_{x\to\infty} L(x) = K > 0$, *then*

$$\left\| I\left(X_t > b_n e^u\right) - P\left(X_t > b_n e^u|\Im_{n,t-q_n}^{t+q_n}\right)\right\|_2 \leq \left\{e^{-up/2}\left(1+d_{n,t}^p\right)^{1/2}(m/n)^{p/2\alpha}\right\} \times o\left(q_n^{-\lambda\min\{p,1\}/4}\right).$$

In particular, if $p = \alpha - \iota$ for sufficiently tiny $\iota > 0$, $\sup_{n\geq 1}\sup_{1\leq t\leq n} d_{n,t} \leq K$, $\lambda \geq 1/\min\{1, p/2\}$ and $\{q_n\}$ satisfies $n/m_n = o(q_n^\delta)$ for some $\delta > 0$, then Assumption A.2 is satisfied.

(ii) Let $\{X_t\}$ be L_p-E-NED on $\{\Im_{n,t}\}$, $p > 0$, with constants $f_{n,t}(u)$ and coefficients ψ_{q_n} of size $\lambda > 0$. Then $\{X_t\}$ is L_s-E-NED on $\{\Im_{n,t}\}$ for any $s \gtreqless p$ with constants $f_{n,t}(u)^\theta$ and coefficients $\psi_{q_n}^\theta$ of size $\lambda\theta$, $\theta = p/\max\{p,s\}$.

Remark 21. Boundedness $d_{n,t} \leq K$ applies to $\{X_t\}$ with bounded forms of time dependence in the L_p-norm, like cyclical trend or stochastic breaks in variance when $p = 2$. Processes $\{X_t\}$ with tail (2) and $L(x) \to K$ include the popular class $\bar{F}_t(x) = cx^{-\alpha}(1 + o(1))$. Finally, any restriction on q_n is irrelevant, since the main results only exploit $q_n \to \infty$.

The general class of nonlinear distributed lags in Example 4 satisfies Lemma 4.

LEMMA 5. *Consider* $X_t = \sum_{i=0}^{\infty} \pi_{t,i} \epsilon_{t-i}$ *from Example 4. If* $E_{n,t} = [\epsilon_{t-i}]_{i=0}^{[q_n/2]}$ *for* $t = 1, \ldots, n$ *and* 0 *otherwise, and* $n/m_n = o(q_n^{\delta})$ *for some* $\delta > 0$, *then Assumption A.2 is satisfied.*

5.3. Non-NED and L_2-E-NED

The fact that such a large class of L_p-NED processes has the L_2-E-NED property suggests it is safe simply to impose L_p-NED on $\{X_t\}$. However, not all interesting processes are NED. Consider the following GARCH process:

$$X_t = \sigma_t \epsilon_t, \quad \epsilon_t \text{ is i.i.d. and } L_p\text{-bounded}, \quad p > 0; \tag{10}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i X_{t-i}^2 + \sum_{i=1}^{q} \gamma_i \sigma_{t-i}^2, \qquad \alpha_0 > 0, \quad \text{at least one } \beta_i, \gamma_i > 0;$$

$$\text{the roots of } 1 - \sum_{i=1}^{q} \gamma_i z^i \text{ lie outside unit circle;}$$

and the Lyapunov exponent $\gamma < 0$.[3] Class (10) has regularly varying tails of the form $P(|X_t| > x) = cx^{-\kappa}(1 + o(1))$, $c > 0$, $\alpha > 0$ (Basrak et al., 2002, Thm. 3.1). The root condition implies

$$\sigma_t^2 = \pi_0 + \sum_{i=1}^{\infty} \pi_i X_{t-i}^2, \qquad \pi_0 > 0, \qquad \pi_i \geq 0, \quad \text{at least one } \pi_i > 0.$$

Davidson (2004) shows that $\{X_t\}$ is L_1- or L_2-NED on $\{\epsilon_t\}$ if $\sum_{i=1}^{\infty} \pi_i < 1$, which neglects IGARCH and GARCH with explosive roots. The following result developed in Hill (2008c) reveals many of these latter processes are, however, E-NED. See also Example 5 in Section 2.

LEMMA 6. *Let* X_t *be generated by (10) with* $E[\epsilon_t] = 0$ *and* $E[\epsilon_t^2] = 1$. *Let* $0 \leq \pi_i \leq C\rho^i$ *for some* $\rho \in (0, 1)$ *and* $C \in (0, 1/\rho)$. *Then* $\{X_t\}$ *is geometrically* L_2*-E-NED on* $\{\Im_{n,t}\}$, *where* $\Im_{n,t}$ *is induced by* $E_{n,t} = [\epsilon_{t-i}]_{i=1}^{[q_n 2]}$ *for* $t = 1, \ldots, n$ *and* 0 *otherwise.*

Remark 22. The bound $\pi_i \leq C\rho^{-i}$ easily allows $\sum_{i=1}^{\infty} \pi_i \geq 1$ covering integrated and many explosive GARCH cases.

Remark 23. Since ϵ_t is i.i.d., all parts of Assumption A.2 hold.

5.4. L_2-E-NED: Direct Proofs

Despite knowing that E-NED covers mixing, NED, and certain non-NED processes, it is instructive to demonstrate the property from first principles. Assume that throughout $\{\epsilon_t\}$ is a symmetrically distributed process where $|\epsilon_t|$ has for each t tail (2) with index $\alpha > 0$, and $E_{n,t} = [\epsilon_{t-i}]_{i=0}^{[q_n/2]}$ for $t = 1, \ldots, n$ and 0 otherwise.

Example 9 (Linear distributed lags)
Define $X_t := \sum_{i=1}^{\infty} \pi_i \epsilon_{t-i}$, $\pi_0 = 1$, where $\pi_i \geq 0$ and $\inf_{t \in \mathbb{Z}} P(\epsilon_t \geq 0) = 1$ for brevity, and $\sum_{i=0}^{\infty} \pi_i^{\alpha} < \infty$, general cases being similar. In the following we only require $\{\epsilon_t\}$ to behave *like* an independent sequence in the tails (cf. Feller, 1971; Cline, 1983; Hill, 2008b).

LEMMA 7. *Let $\{\epsilon_t\}$ satisfy the convolution tail property $P(\sum_{i=0}^{\infty} a_i \epsilon_{t-i} > x) \sim \sum_{i=0}^{\infty} P(a_i \epsilon_{t-i} > x)$ for any deterministic sequence of real numbers $\{a_i\}$, $\sum_{i=0}^{\infty} |a_i|^{\alpha} < \infty$. Then X_t has tail* (2) *with index α. Further, $\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ with constants $f_{n,t}(u) = e^{-\alpha u/2}(m_n/n)^{1/2}$ and coefficients $\psi_{q_n} = (\sum_{i=q_n}^{\infty} \pi_i^{\alpha} / \sum_{i=0}^{\infty} \pi_i^{\alpha})^{1/2} \in (0, 1)$ for any $r \geq 2$.*

Remark 24. Given the simple parametric structure of X_t, we do not require $\lim_{x \to \infty} L(x) = K > 0$ or $n/m_n = o(q_n^{\delta})$, contrary to Lemma 4.

Remark 25. Since ϵ_t is geometrically strong mixing the F-mixing property with arbitrary size is immediate, and $\sup_{1 \leq t \leq n} f_{n,t}(u) = e^{-\alpha u/2}(m_n/n)^{1/2}$ is Lebesgue integrable on $\mathbb{R}_+$. Further, the E-NED size is 1/2 as long as π_i decays sufficiently fast. This is trivial for stationary ARMA, since $\pi_i \to 0$ geometrically as $i \to \infty$, and for ARFIMA(p, d, q) with Hurst $d < (\alpha - 1)/\alpha < 1$, since $\pi_i = O(i^{-(1-d)})$ implies both $\sum_{i=0}^{\infty} \pi_i^{\alpha} < \infty$ and $\psi_{q_n} = O(q_n^{-1})$.

Example 10 (Bilinear)
Assume $X_t = \beta X_{t-1} \epsilon_{t-1} + \epsilon_t$, ϵ_t is i.i.d., $\beta > 0$, and $\beta^{\alpha/2} E[\epsilon_t^{\alpha/2}] < 1$. Then $\{X_t\}$ has a convergent linear distributed lag representation $X_t = \sum_{j=0}^{\infty} \beta^j \epsilon_t^{(j)}$, where $\epsilon_t^{(0)} = \epsilon_t$, and $\epsilon_t^{(j)} = \epsilon_{t-j}^2 (\sqcap_{i=1}^{j-1} \epsilon_{t-i})$ has tail (2) with index $\alpha/2$. In particular, the tail behavior of X_t is dominated by $\sum_{j=1}^{\infty} \beta^j \epsilon_t^{(j)}$, which also satisfies (2) with index $\alpha/2$. See Davis and Resnick (1996, Cor. 2.4).

LEMMA 8. *$\{X_t\}$ is L_2-E-NED on $\{\Im_{n,t}\}$ with constants $f_{n,t}(u) = e^{-\alpha u/2}(m_n/n)^{1/2}$ and coefficients $\psi_{q_n} = o(q_n^{-\lambda})$ for any $\lambda > 0$.*

NOTES

1. GARCH processes, for example, are known to have regularly varying tails (Basrak et al., 2002). The scaled residuals $\{\hat{\epsilon}_t / \hat{\sigma}_t\}$ of GARCH $X_t = \sigma_t \epsilon_t$, however, may have subtantially thinner tails than the original series itself, and need not have regularly varying tails (e.g., $\epsilon_t \overset{iid}{\sim} N(0,1)$). See Iglesias and Linton (2008) for a novel, direct approach for estimating the index of GARCH processes.
2. I would like to thank Oliver Linton for pointing out this issue.
3. The exponent γ is associated with the first order difference equation form of $Z_t := [X_t^2, \ldots, X_{t-p+2}^2; \sigma_{t+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+2}^2]'$. It is easy to show $Z_t = A_t Z_{t-1} + B_t$ for some i.i.d. sequences $\{A_t, B_t\}$ of $k \times k$ matrices A_t and k-vectors B_t, $k \geq 1$. The exponent γ is defined by $\gamma = \lim_{n \to \infty} n^{-1} \ln || \sqcap_{t=1}^{n} A_t ||_o$, where $||A||_o = \sup_{x \in \mathbb{R}^k, |x|=1} |Ax|$. If ϵ_t in (10) is i.i.d. with zero mean and unit variance, then $\gamma < 0$ given the remaining properties (Basrak et al., 2002).

REFERENCES

Akgiray, V. & G.G. Booth (1988) The stable law model of stock returns. *Journal of Business and Economic Statistics* 6, 51–57.

An, H.Z. & F.C. Huang (1996) The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6, 943–956.

Andrews, D.W.K. (1984) Non-Strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.

Andrews, D.W.K. (1985) A nearly independent, but non-strong mixing, triangular array. *Journal of Applied Probability* 22, 729–731.

Andrews, D.W.K. (1992) Generic uniform convergence. *Econometric Theory* 8, 241–257.

Basrak, B., R.A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.

Beirlant, J., D. Dierckx, & A. Guillou (2005) Estimation of the extreme value index and generalized plots. *Bernoulli* 11, 949–970.

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

Caner, M. (1998) Tests for cointegration with infinite variance errors. *Journal of Econometrics* 86, 155–175.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chan, N.H., S.D. Deng, L. Peng, & Z. Xia (2007) Interval estimation for the conditional value-at-risk based on GARCH with heavy tailed innovations. *Journal of Econometrics* 137, 556–576.

Chan, N.H. & L.T. Tran (1989) On the first order autoregressive process with infinite variance. *Econometric Theory* 5, 354–362.

Cheng, B.N. & 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. Ph.D. dissertation, Colorado State University.

Csörgö, S. & L. Viharos (1995) On the asymptotic normality of Hill's estimator. *Mathematical Proceedings of the Cambridge Philosophical Society* 118, 375–382.

Davidson, J. (1992) A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric Theory* 8, 313–329.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

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. & T. Hsing (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability* 23, 879–917.

Davis, R. & S. Resnick (1984) Tail estimates motivated by extreme value theory. *Annals of Statistics* 12, 1467–1487.

Davis, R. & S. Resnick (1996) Limit theory for bilinear processes with heavy-tailed noise. *Annals of Applied Probability* 6, 1191–1210.

de Haan, L. & 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. & 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, & M. Sørensen (eds.), *Empirical Process Techniques for Dependent Data*. Birkhauser.

Drees, H., A. Ferreira, & L. de Haan (2004) On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability* 14, 1179–1201.

Einmahl, J. & T. Lin (2006) Asymptotic normality of extreme value estimators on C[0,1]. *Annals of Statistics* 34, 469–492.

Embrechts, P., C. Klüppelberg, & T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*. Springer.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, vol. 2. Wiley.

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

Goldie, C.M. & R.L. Smith (1987) Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics* 38, 45–71.

Gorodetskii, V.V. (1977) On the strong mixing property for linear sequences. *Theory of Probability and Its Applications* 22, 411–413.

Guegan D. & S. Ladoucette (2001) Non-mixing properties of long memory processes. *Comptes Rendus de l'Academie des Sciences*, Series I, Mathematics 333, 373–376.

Haeusler, E. & 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*, Series B 44, 37–42.

Hall, P. & A.H. Welsh (1985) Adaptive estimates of parameters of regular variation. *Annals of Statistics* 13, 331–341.

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. (2008b) Gaussian Tests of Extremal White Noise for Dependent, Heterogeneous, Heavy Tailed Time Series. Working paper, University of North Carolina–Chapel Hill.

Hill, J.B. (2008a) Extremal Memory of Stochastic Volatility with Applications to Tail Shape and Tail Dependence Inference. Working paper, University of North Carolina–Chapel Hill.

Hill, J.B. (2008c) Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics. Working paper, University of North Carolina–Chapel Hill.

Hill, J.B. (2009a) On functional central limit theorems for dependent, heterogeneous arrays with applications. *Journal of Statistical Planning and Inference* 139, 2091–2110.

Hill, J.B. (2009b) On Tail Index Estimation for Dependent, Heterogeneous Data. Working paper, University of North Carolina–Chapel Hill.

Hsing, T. (1991) On tail index estimation using dependent data. *Annals of Statistics* 19, 1547–1569.

Hsing, T. (1993) Extremal index estimation for a weakly dependent stationary sequence. *Annals of Statistics* 21, 2043–2071.

Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349–382.

Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff.

Iglesias, E. & O. Linton (2008) Estimation of Tail Thickness Parameters from GARCH Models. Mimeo, London School of Economics.

Leadbetter, M.R. (1974) On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 28, 289–303.

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

Leibscher, E. (2005) Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis* 26, 669–689.

Mason, D. (1982) Laws of large numbers for sums of extreme values. *Annals of Probability* 10, 754–764.
McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *Annals of Probability* 2, 620–628.
McLeish, D.L. (1975) A maximal inequality and dependent strong law. *Annals of Probability* 3, 329–339.
Meitz M. & P. Saikkonen (2008) Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis* 29, 453–475.
Newey, W.K. & 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.
Quintos, C., Z. Fan, & P.C.B. Phillips (2001) Structural change tests in tail behavior and the Asian crisis. *Review of Economic Studies* 68, 633–663.
Resnick, S. (1987) *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag.
Resnick, S. & H. Rootzén (2000) Self-Similar communication models and very heavy tails. *Annals of Applied Probability* 10, 753–778.
Resnick, S. & C. Stărică (1995) Consistency of Hill's estimator for dependent data. *Journal of Applied Probability* 32, 139–167.
Resnick, S. & C. Stărică (1998) Tail index estimation for dependent data. *Annals of Applied Probability* 8, 1156–1183.
Rootzén, H. (2009) Weak convergence of the tail empirical function for dependent sequences. *Stochastic Processes and Their Applications* 119, 468–490.
Rootzén, H., M. Leadbetter, & L. de Haan (1990) Tail and Quantile Estimation for Strongly Mixing Stationary Sequences. Technical report 292, University of North Carolina, Chapel Hill.
Segers, J. (2002) Abelian and Tauberian theorems on the bias of the Hill estimator. *Scandinavian Journal of Statistics* 29, 461–483.
Serfling, R.J. (1968) Contributions to central limit theory for dependent variables. *Annals of Mathematical Statistics* 39, 1158–1175.
Smith, R.L. (1982) 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. & I. Weissman (1994) Estimating the extremal index. *Journal of the Royal Statistical Society*, Series B 56, 515–128.
Wu, W.B. (2005) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Science USA* 102, 14150–14154.
Wu, W.B. & M. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and Their Applications* 115, 939–958.

APPENDIX A: Proofs of Main Results

The following proofs exploit Lemmas B.1–B.10 in Appendix B. Recall that $U_{m_n,t} = (\ln(X_t/b_{m_n}))_+ - E[(\ln(X_t/b_{m_n}))_+]$ and $I_{m_n,t}(u) = I(X_t > b_{m_n}e^u) - P(X_t > b_{m_n}e^u)$, $u \geq 0$.

Proof of Lemma 1. Under the maintained assumptions and Lemma B.1, $\{U_{m_n,t}, \Im_{n,t}\}$ and $\{I_{\rho m_n,t}(u), \Im_{n,t}\}$ for all ρ in an arbitrary neighborhood of 1 form L_2-mixingale arrays with size 1/2 and constants $\{e^*_{n,t}, e_{n,t}(u)\} = O((m_n/n)^{1/2})$. Now define an integer sequence $\{a_{n,t}\}$,

$$a_{n,t} := t \times I(t \neq n) + m_n \times I(t = n) \qquad t = 1, 2, \ldots,$$

and note that $a_{n,n} = m_n$ and $a_{n,t} \to \infty$ as $t \to \infty$ $\forall n \geq 1$. For some finite $K > 0$, each $\tilde{e}_{n,t} \in \{e^*_{n,t}, e_{n,t}(u)\}$ satisfies (e.g., Davidson, 1994, Thm. 2.2.3)

$$\sum_{t=1}^{\infty} (\tilde{e}_{n,t}/a_{n,t})^2 \leq K \sum_{t \neq n} t^{-2} + o(1) < \infty.$$

Thus $\sum_{t=1}^{n} U_{m_n,t}/a_{n,n} \overset{a.s.}{\to} 0$ and $\sum_{t=1}^{n} I_{\rho m_n,t}(u)/a_{n,n} \overset{a.s.}{\to} 0$ by Davidson's (1994, Cor. 20.16) generalization of McLeish's (1975) strong law for L_2-mixingales. The weak limit $\ln(X_{([\rho m])}/b_{\rho m_n}) \overset{p}{\to} 0$ then follows from arguments in Hsing (1991, p. 1551). ■

Proof of Theorem 1.

Claim (i). Weak consistency $\hat{\alpha}_{m_n}^{-1} \overset{p}{\to} \alpha^{-1}$ under Assumption A.1 follows from Lemma 1. See Theorem 2.2 of Hsing (1991).

Claim (ii). Uniform weak consistency $\sup_{\phi \in \Phi} |1/m_n(\phi) \sum_{t=1}^{n} U_{m_n(\phi),t}| \overset{p}{\to} 0$ and $\sup_{\phi \in \Phi} |1/m_n(\phi) \sum_{t=1}^{n} I_{m_n(\phi),t}(u) \overset{p}{\to} 0$ follow instantly from Theorem 3 of Andrews (1992), cf. Davidson (1994, Thm. 21.10), given weak consistency Lemma 1 and Lemma B.3 Lipschitz properties.

The argument for $\sup_{\phi \in \Phi} |\ln(X_{(m_n(\phi)+1)}/b_{m_n(\phi)})| \overset{p}{\to} 0$ is similar to Hsing's (1991, p. 1551) consistency proof. First, note by subadditivity for any $u > 0$,

$$P\left(\sup_{\phi \in \Phi}\left|\ln\left\{\frac{X_{(m_n(\phi)+1)}}{b_{m_n(\phi)}}\right\}\right| > u\right) \leq P\left(\left|\sup_{\phi \in \Phi} \ln\left\{\frac{X_{(m_n(\phi)+1)}}{b_{m_n(\phi)}}\right\}\right| > u/2\right)$$
$$+P\left(\left|\sup_{\phi \in \Phi} -\ln\left\{\frac{X_{(m_n(\phi)+1)}}{b_{m_n(\phi)}}\right\}\right| > u/2\right).$$

We will show that the first term on the right-hand side is $o(1)$, the second term being similar. Since (1)–(3) and Assumption B imply (cf. Hsing, 1991, pp. 1553–1554; see especially Smith, 1982, eqn. 2.2; Goldie and Smith, 1987, Thm. 2.1.1, Cor. 2.2.1) that

$$\frac{n}{m_n(\phi)} E\left[I\left(X_t > b_{m_n(\phi)} e^{u/2}\right)\right] = e^{-\alpha u/2} \times \left(1 + o\left(1/m_n(\phi)^{1/2}\right)\right),$$

observe by construction

$$\ln\left\{\frac{X_{(m_n(\phi)+1)}}{b_{m_n(\phi)}}\right\} > u/2 \Longleftrightarrow \frac{1}{m_n(\phi)} \sum_{t=1}^{n} I\left(X_t > b_{m_n(\phi)} e^{u/2}\right) > 1$$
$$\Longleftrightarrow \frac{1}{m_n(\phi)} \sum_{t=1}^{n} I_{m_n(\phi),t}(u/2) > 1 - e^{-\alpha u/2} + o\left(1/m_n(\phi)^{1/2}\right).$$

Now use $\sup_{\phi \in \Phi} |1/m_n(\phi) \sum_{t=1}^{n} I_{m_n(\phi),t}(u)| \overset{p}{\to} 0$, $e^{-\alpha u/2} < 1$ and $\inf_{\phi \in \Phi} m_n(\phi) \to \infty$ under Assumption C to conclude, for some tiny $\iota > 0$,

$$\lim_{n\to\infty} P\left(\left|\sup_{\phi\in\Phi} \ln\left\{\frac{X_{(m_n(\phi)+1)}}{b_{m_n(\phi)}}\right\}\right| > u/2\right)$$

$$\leq \lim_{n\to\infty} P\left(\left|\sup_{\phi\in\Phi} \frac{1}{m_n(\phi)} \sum_{t=1}^{n} I_{m_n(\phi),t}(u/2)\right| > 1 - e^{-\alpha u/2} - |o(1)|\right)$$

$$\leq \lim_{n\to\infty} P\left(\sup_{\phi\in\Phi} \left|\frac{1}{m_n(\phi)} \sum_{t=1}^{n} I_{m_n(\phi),t}(u/2)\right| > \iota\right) = 0.$$

Claim (iii). Consider $\sup_{\phi\in\Phi} |\hat{\alpha}_{m_n(\phi)}^{-1} - \alpha^{-1}| \xrightarrow{p} 0$ and define

$$\Delta W_{m_n,t} := \ln\left(X_t/b_{m_n}\right) \times I\left(X_t > X_{(m_n+1)}\right) - \ln\left(X_t/b_{m_n}\right)_+.$$

Consistency $\hat{\alpha}_{m_n}^{-1} \xrightarrow{p} \alpha^{-1}$ under Claim (i), the Lemma B.4 identity

$$\hat{\alpha}_{m_n}^{-1} - \alpha^{-1} = \frac{1}{m_n} \sum_{t=1}^{n} \left\{U_{m_n,t} - \alpha^{-1} I_{m_n,t}\left(u/m_n^{1/2}\right)\right\}$$

$$+ \frac{1}{m_n} \sum_{t=1}^{n} \Delta W_{m_n,t} + o\left(1/m_n^{1/2}\right), \qquad \textbf{(A.1)}$$

and the Lemma 1 implication $1/m_n \sum_{t=1}^{n} \left(U_{m_n,t} - \alpha^{-1} I_{m_n,t}\left(u/m_n^{1/2}\right)\right) \xrightarrow{p} 0$ imply that $1/m_n \sum_{t=1}^{n} \Delta W_{m_n,t} \xrightarrow{p} 0$. Andrews's (1992, Thm. 3) uniform law of large numbers and Lemma B.3 Lipschitz properties therefore imply that $\sup_{\phi\in\Phi} |1/m_n(\phi) \sum_{t=1}^{n} \Delta W_{m_n(\phi),t}| \xrightarrow{p} 0$. The proof now follows from identity (A.1), the Claim (ii) uniform laws, and $\inf_{\phi\in\Phi} m_n(\phi) \to \infty$ under Assumption C:

$$\sup_{\phi\in\Phi} \left|\hat{\alpha}_{m_n(\phi)}^{-1} - \alpha^{-1}\right| \leq \sup_{\phi\in\Phi} \left|\frac{1}{m_n(\phi)} \sum_{t=1}^{n} U_{m_n(\phi),t}\right| + \sup_{\phi\in\Phi} \left|\frac{1}{m_n(\phi)} \sum_{t=1}^{n} I_{m_n(\phi),t}(u)\right|$$

$$+ \sup_{\phi\in\Phi} \left|\frac{1}{m_n(\phi)} \sum_{t=1}^{n} \Delta W_{m_n(\phi),t}\right| + o\left(\left(\inf_{\phi\in\Phi} \{m_n(\phi)\}\right)^{-1/2}\right) \xrightarrow{p} 0.$$

■

Proof of Lemma 2. Write

$$T_{m_n,t} = T_{m_n,t}\left(\omega, u/m_n^{1/2}\right) = m_n^{-1/2}\left[\omega_1 U_{m_n,t} - \omega_2 \alpha^{-1} I_{m_n,t}(u)\right], \qquad \omega'\omega = 1.$$

Step 1 (NED). Under the maintained assumptions and Lemma B.1, $\{U_{m_n,t}, I_{m_n,t}(u)\}$ are L_2-NED on $\{\Im_{n,t}\}$ with coefficients $\psi_{n,q_n}^* = (m_n/n)^{1/2-1/r} \psi_{q_n} = o\left((m_n/n)^{1/2-1/r} q_n^{-1/2}\right)$ and constants $\{f_{n,t}^*, f_{n,t}^*(u)\}$ that satisfy $\sup_{1\leq t\leq n} f_{n,t}^* = O\left((m_n/n)^{1/r}\right)$ and $\sup_{1\leq t\leq n} \sup_{u\geq 0} f_{n,t}^*(u) = O\left((m_n/n)^{1/r}\right)$. Use Minkowski's inequality and $\omega'\omega = 1$

to deduce $\{T_{m_n,t}\}$ is L_2-NED on $\{\Im_{n,t}\}$ with coefficients ψ^*_{n,q_n} and constants (Davidson, 1994, Thm. 17.8)

$$d_{n,t} = Km_n^{-1/2}\max\left\{f^*_{n,t}, \sup_{u\geq 0} f^*_{n,t}(u)\right\}$$

$$= O\left(m_n^{-1/2}(m_n/n)^{1/r}\right) \quad \text{uniformly in } 1 \leq t \leq n.$$

Step 2 (Mixingale). Assume that the base $\{\epsilon_t\}$ is F-strong mixing with coefficients $\varepsilon_{n,q_n} = o\big((m_n/n)q_n^{-r/(r-2)}\big)$. Standard inequalities for mixing random variables carry over to F-mixing, and distributed lags of F-mixing random variables are F-mixing (Hill, 2009b, Lem. C.1). Therefore Theorem 17.5 of Davidson (1994) applies. For some $r > 2$,

$$\left\|T_{m_n,t} - E[T_{m_n,t}|\Im_{n,t-q_n}]\right\|_2 \leq \max\left\{\left\|T_{m_n,t}\right\|_r, \ d_{n,t}\right\} \times \max\left\{6\varepsilon_{n,q_n}^{1/2-1/r}, \psi^*_{n,q_n}\right\}.$$

Use $\omega'\omega = 1$, Minkowski's inequality, and the Lemma B.2 moment bounds to deduce that

$$\left\|T_{m_n,t}\right\|_r \leq Km_n^{-1/2}\left(\left\|U_{m_n,t}\right\|_r + \sup_{u\geq 0}\left\|I_{m_n,t}\left(u/m_n^{1/2}\right)\right\|_r\right) = O\left(m_n^{-1/2}(m_n/n)^{1/r}\right).$$

Multiply and divide by $n^{1/2}$ and rearrange terms,

$$\left\|T_{m_n,t} - E[T_{m_n,t}|\Im_{n,t-q_n}]\right\|_2$$

$$\leq Kn^{-1/2} \times (n/m_n)^{1/2-1/r}\max\left\{6\varepsilon_{n,q_n}^{1/2-1/r}, \psi^*_{n,q_n}\right\}$$

$$= Kn^{-1/2} \times \max\left\{\left[(n/m_n)\varepsilon_{n,q_n}\right]^{1/2-1/r}, \ (n/m_n)^{1/2-1/r}\psi^*_{n,q_n}\right\} = c_{n,t} \times \psi_{q_n},$$

say, where $\psi_{q_n} = o\big(q_n^{-1/2}\big)$ under Assumption B.2 and $c_{n,t} = Kn^{-1/2}$ given F-mixing and E-NED rates.

Analogous arguments apply to the remaining mixingale inequality $||T_{m_n,t} - E[T_{m_n,t}|\Im_{n,-\infty}^{t+q_n}]||_2 \leq c_{n,t}\psi_{q_n+1}$ (e.g., Davidson, 1994; eqn. 17.19) and to the F-uniform mixing case. ■

Remark A.1. Notice that $||T_{m_n,t} - E[T_{m_n,t}|\Im_{n,t-q_n}]||_2 \leq o(n^{-1/2}q_n^{-1/2})$ requires the F-mixing coefficients to satisfy $(n/m_n)q_n^{r/(r-2)}\varepsilon_{n,q_n} \to 0$. In general, therefore, $q_n \to \infty$ *must* hold to ensure $\lim_{n\to\infty}\varepsilon_{n,q_n} = 0$, since $n/m_n \to \infty$. An obvious exception is $\varepsilon_{n,q} = 0$ uniformly in n and q (e.g., the base $E_{n,t}$ is independent).

Proof of Lemma 3. The proof exploits Lemma 2: $\{T_{m_n,t}(\omega, u/m_n^{1/2}), \Im_{n,t}\}$ forms an L_2-mixingale array with coefficients $\psi_{q_n} = o(q_n^{-1/2})$ and constants $c_{n,t} = Kn^{-1/2}$. Note by McLeish's (1975) bound for L_2-mixingales with size $1/2$,

$$\sup_{\omega'\omega=1}\sigma^2_{m_n}(\omega) = \sup_{\omega'\omega=1} E\left(\sum_{t=1}^n T_{m_n,t}(\omega, u/m_n^{1/2})\right)^2 = O\left(\sup_{\omega'\omega=1}\sum_{t=1}^n c^2_{n,t}\right) = O(1). \tag{A.2}$$

Step 1 ($\sum_{t=1}^{n} T_{m_n,t}/\sigma_{m_n}(\omega) \Longrightarrow N(0,1)$)**.** Write $T_{m_n,t} := T_{m_n,t}(\omega, u/m_n^{1/2})$. We will show conditions (a)–(f) of de Jong's (1997) Lemma 1 central limit theorem hold, replicated for reference in Lemma B.5. De Jong's argument exploits the following real-valued sequences $\{k_n, l_n, r_n\}$ and Bernstein blocks $\{Z_{n,i}, L_{n,i}\}_{i=1}^{r_n}$:

$$k_n/n \to 0,\ k_n = o(m_n^{1/4}), \qquad r_n = [n/k_n] \quad \text{where } k_n, r_n \to \infty \text{ as } n \to \infty$$

$$1 \le l_n \le k_n - 1 \le n-1 \quad \text{where } l_n/k_n \to 0 \quad \text{and} \quad l_n \to \infty \quad \text{as } n \to \infty \tag{A.3}$$

and

$$Z_{n,i} := \sum_{t=(i-1)k_n+l_n+1}^{ik_n} T_{m_n,t} \quad \text{and} \quad L_{n,i} = \sum_{t=(i-1)k_n+1}^{(i-1)k_n+l_n} T_{m_n,t}. \tag{A.4}$$

By construction, $\sum_{t=1}^{n} T_{m_n,t}$ obtains the decomposition

$$\sum_{t=1}^{n} T_{m_n,t} = \sum_{i=1}^{r_n} Z_{n,i} + \sum_{i=1}^{r_n} L_{n,i} + R_n \quad \text{for some remainder } R_n.$$

De Jong's (1997) construction $r_n = [n/k_n]$ (cf. Davidson, 1992), renders $R_n = o_p(1)$. The sequences k_n and l_n regulate the amount of information in and between the blocks $L_{n,i}$ and $Z_{n,i}$ in such a way that $\sum_{i=1}^{r_n} L_{n,i} = o_p(1)$ is also asymptotically negligible. Finally, under the stated conditions $\{Z_{n,i}\}_{i=1}^{r_n}$ is approximable by a martingale difference array that satisfies McLeish's (1974, Thm. 2.1) central limit theorem (cf. Lemma 1 of de Jong, 1997). Note that $k_n = o(m_n^{1/4})$ is always possible and merely expedites the proof.

Define a σ-subfield associated with the mixing functional $E_{n,t}$

$$\tilde{F}_{n,i} := \sigma(\{E_{n,\tau} : \tau \le ik_n\}).$$

Condition (a). Minkowski's inequality and the Lemma B.2 moment bounds imply

$$\left\|T_{m_n,t}\right\|_2 \le K m_n^{-1/2} \left(\left\|U_{m_n,t}\right\|_2 + \sup_{u \ge 0} \left\| I_{m_n,t}(u/m_n^{1/2}) \right\|_2 \right)$$

$$= O(m_n^{-1/2}(m_n/n)^{1/2}) = O(n^{-1/2}).$$

Now use Minkowski's inequality again and $r_n k_n - n \to 0$ to deduce

$$\left\| \sum_{t=r_n k_n+1}^{n} T_{m_n,t} \right\|_2 \le \sum_{t=r_n k_n+1}^{n} \left\|T_{m_n,t}\right\|_2 \le (n - r_n k_n) K n^{-1/2} = o(1).$$

Chebyshev's inequality completes the proof: $\sum_{t=r_n k_n+1}^{n} T_{m_n,t} \xrightarrow{p} 0$.

Condition (b). The mixingale property and McLeish's (1975) bound imply

$$E\left(\sum_{i=1}^{r_n} \sum_{t=(i-1)k_n+1}^{(i-1)k_n+l_n} T_{m_n,t} \right)^2 = O\left(\sum_{i=1}^{r_n} \sum_{t=(i-1)k_n+1}^{(i-1)k_n+l_n} c_{n,t}^2 \right)$$

$$= O(r_n l_n n^{-1}) = O(l_n/k_n) = o(1).$$

Condition (c). Define the index set

$$A_{n,t} = \left\{ t : t \in \bigcup_{i=1}^{r_n} [(i-1)k_n + l_n + 1, ik_n] \right\}.$$

Analogous to de Jong's (1997, A.7–A.12) argument, for $t \in A_{n,t}$ it can be shown that $\{E[T_{m_n,t}|\tilde{F}_{n,i-1}], \Im_{n,t}\}$ forms an L_2-mixingale array with constants (i.e., de Jong's "*index numbers*") $c_{n,t}\psi_{l_n}^{\iota}$ and coefficients $\psi_{l_n}^{1-\iota}$ satisfying $\psi_{l_n}^{1-\iota} = o(l_n^{-1/2})$ for sufficiently tiny $\iota > 0$. Thus, by McLeish's (1975) bound and $l_n \to \infty$ as $n \to \infty$,

$$\begin{aligned} E\left(\sum_{i=1}^{r_n} E\left[Z_{n,i}|\tilde{F}_{n,i-1}\right]\right)^2 &= E\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)k_n+l_n+1}^{ik_n} E\left[T_{m_n,t}|\tilde{F}_{n,i-1}\right]\right)^2 \\ &= O\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)k_n+l_n+1}^{ik_n} c_{n,t}^2\psi_{l_n}^{2\iota}\right) \\ &= O\left(r_nk_nn^{-1}l_n^{-\iota}\right) = O(l_n^{-\iota}) = o(1). \end{aligned}$$

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

Condition (e). Analogous to de Jong (1997, A.13–A.17) and Condition (c),

$$\begin{aligned} &\left\|\sum_{i=1}^{r_n} Z_{n,i}^2 - \sum_{i=1}^{r_n}\left(E\left[Z_{n,i}|\tilde{F}_{n,i}\right] - E\left[Z_{n,i}|\tilde{F}_{n,i-1}\right]\right)^2\right\|_1 \\ &\quad \leq 3\sum_{i=1}^{r_n}\left\|Z_{n,i} - \left(E[Z_{n,i}|\tilde{F}_{n,i}] - E[Z_{n,i}|\tilde{F}_{n,i-1}]\right)\right\|_2 \times \left\|Z_{n,i}\right\|_2 \\ &\quad = O\left(\sum_{i=1}^{r_n}\left(\sum_{t=(i-1)k_n+l_n+1}^{ik_n} c_{n,t}^2\psi_{l_n}^{2\iota}\right)^{1/2}\left(\sum_{t=(i-1)k_n+l_n+1}^{ik_n} c_{n,t}^2\right)^{1/2}\right) \\ &\quad = O\left(r_n\left(k_nn^{-1}l_n^{-\iota}\right)^{1/2}\left(k_nn^{-1}\right)^{1/2}\right) = O(l_n^{-\iota/2}) = o(1). \end{aligned}$$

Now apply Chebyshev's inequality and $\sum_{i=1}^{r_n} Z_{n,i}^2/\sigma_{m_n}^2(\omega) \stackrel{p}{\to} 1$ by Lemma B.7.

Condition (f). Define $W_{n,i} := E[Z_{n,i}|\tilde{F}_{n,i}] - E[Z_{n,i}|\tilde{F}_{n,i-1}]$. We require the Lindeberg condition $\sum_{i=1}^{r_n} E[W_{n,i}^2 I(|W_{n,i}| > \varepsilon)] \to 0$ for any $\varepsilon > 0$. By the same reasoning as Condition (a) and the conditional Jensen's inequality, $\forall r \geq 1$,

$$\left\|W_{n,i}\right\|_r \leq 2\left\|Z_{n,i}\right\|_r \leq 2\sum_{t=(i-1)k_n+l_n+1}^{ik_n}\left\|T_{m_n,t}\right\|_r = O\left(k_nm_n^{-1/2}(m_n/n)^{1/r}\right).$$

Therefore, $\forall p, s \geq 0$, $1/p + 1/s = 1$, and all $\varepsilon > 0$, under Hölder's and Markov's inequalities

$$\max_{1\leq i\leq r_n} r_n E\left[W_{n,i}^2 I(|W_{n,i}| > \varepsilon)\right] \leq K \max_{1\leq i\leq r_n}\left\{r_n \left\|W_{n,i}\right\|_{2p}^2 \times \left\|W_{n,i}\right\|_s\right\}$$
$$= O\left(r_n k_n^2 m_n^{-1}(m_n/n)^{1/p} \times k_n m_n^{-1/2}(m_n/n)^{1/s}\right)$$
$$= O\left(k_n^2 m_n^{-1/2}\right) = o(1),$$

where the last line exploits $k_n = o(m_n^{1/4})$ in (A.3).

Step 2 ($m_n^{1/2}\ln(X_{(m_n+1)}/b_{m_n})/\sigma_{m_n}(0,1) \Longrightarrow N(0,1)$)**.** Use Step 1 and a Cramér-Wold device to deduce

$$\alpha^{-1}\frac{1}{m_n^{1/2}}\sum_{t=1}^{n} I_{m_n,t}(u/m_n^{1/2})/\sigma_{m_n}(0,1) \Longrightarrow N(0,1), \tag{A.5}$$

where $\sigma_{m_n}^2(0,1) = E(\alpha^{-1}m_n^{-1/2}\sum_{t=1}^{n} I_{m_n,t}(u/m_n^{1/2}))^2 = O(1)$ by construction of $\sigma_{m_n}^2(\omega_1,\omega_2)$ in (9) and bound (A.2). It is straightforward to show that (A.5) implies $m_n^{1/2}\ln(X_{(m_n+1)}/b_{m_n})/\sigma_{m_n}(0,1) \Longrightarrow N(0,1)$ (Hsing, 1991, Thm 2.4) ■

Proof of Theorem 2. Lemma 3 and a Cramér-Wold device suffice to prove

$$\left(\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\frac{U_{m_n,t}}{\sigma_{m_n}(1,0)}, \alpha^{-1}\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\frac{I_{m_n,t}(u/m_n^{1/2})}{\sigma_{m_n}(0,1)}\right) \Longrightarrow (Z_1, Z_2) \tag{A.6}$$

for some random vector (Z_1, Z_2) with marginal distributions $Z_i \sim N(0,1)$, where $\sigma_{m_n}^2(\omega_1, \omega_2) = E(\sum_{t=1}^{n} T_{m_n,t}(\omega, u/m_n^{1/2}))^2$ and $T_{m_n,t}(\omega,u) = 1/m_n^{1/2}[\omega_1 U_{m_n,t} + \omega_2\alpha^{-1}I_{m_n,t}(u)]$. Therefore, by the continuous mapping theorem,

$$\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}(U_{m_n,t} - \alpha^{-1}I_{m_n,t}(u/m_n^{1/2}))/\sigma_{m_n}(1,-1)$$
$$= \frac{\sigma_{m_n}(1,0)}{\sigma_{m_n}(1,-1)}\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\frac{U_{m_n,t}}{\sigma_{m_n}(1,0)} - \frac{\sigma_{m_n}(0,1)}{\sigma_{m_n}(1,-1)}\alpha^{-1}\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\frac{I_{m_n,t}(u/m_n^{1/2})}{\sigma_{m_n}(0,1)}$$
$$\Longrightarrow \left(\lim_{n\to\infty}\frac{\sigma_{m_n}(1,0)}{\sigma_{m_n}(1,-1)}\right)Z_1 - \left(\lim_{n\to\infty}\frac{\sigma_{m_n}(0,1)}{\sigma_{m_n}(1,-1)}\right)Z_2 \sim N(0,1). \tag{A.7}$$

Now exploit the Theorem 1 assertion $\ln(X_{([\rho m])}/b_{\rho m_n}) \xrightarrow{p} 0$ for all ρ in a neighborhood of 1, (A.6), and (A.7), and arguments identical to Hsing's (1991, pp. 1553–1554) under tail decay Assumption B to conclude that

$$m_n^{1/2}\left(\hat{\alpha}_{m_n}^{-1}-\alpha^{-1}\right)/\sigma_{m_n}(1,-1)$$

$$\Longrightarrow\left(\lim_{n\to\infty}\frac{\sigma_{m_n}(1,0)}{\sigma_{m_n}(1,-1)}\right)Z_1-\left(\lim_{n\to\infty}\frac{\sigma_{m_n}(0,1)}{\sigma_{m_n}(1,-1)}\right)Z_2\sim N(0,1).$$

Since $\sigma_{m_n}^2 := E(m_n^{1/2}(\hat{\alpha}_{m_n}^{-1}-\alpha^{-1}))^2$, it follows instantly that $|\sigma_{m_n}^2(1,-1)-\sigma_{m_n}^2| \xrightarrow{p} 0$. ■

Proof of Theorem 3. Lemmas B.8 and B.9 together imply $|\hat{\sigma}_{m_n}^2-\sigma_{m_n}^2(1,-1)| \xrightarrow{p} 0$, and by Theorem 2, $|\sigma_{m_n}^2(1,-1)-\sigma_{m_n}^2| \xrightarrow{p} 0$. The claim $|\hat{\sigma}_{m_n}^2-\sigma_{m_n}^2| \xrightarrow{p} 0$ now follows from the triangular inequality. ■

Proof of Lemma 4.

Claim (i). Let $\{X_t\}$ be L_p-NED on $\{\Im_{n,t}\}$. For any $\eta_n > 0$ to be defined below (I would like to thank an anonymous referree for insights into the proof of Lemma 4):

$$\begin{aligned}
&E\left(I\left(X_t > b_{m_n}e^u\right)-E[I\left(X_t > b_{m_n}e^u\right)|\Im_{n,t-q_n}^{t+q_n}]\right)^2\\
&\quad\le E\left[\left(I\left(X_t > b_{m_n}e^u\right)-I(E[X_t|\Im_{n,t-q_n}^{t+q_n}] > b_{m_n}e^u)\right)^2\right.\\
&\qquad\left.\times I\left(\left|X_t-E[X_t|\Im_{n,t-q_n}^{t+q_n}]\right|\le\eta_n\right)\right]\\
&\qquad+E\left[\left(I\left(X_t > b_{m_n}e^u\right)-I(E[X_t|\Im_{n,t-q_n}^{t+q_n}] > b_{m_n}e^u)\right)^2\right.\\
&\qquad\quad\left.\times I\left(\left|X_t-E[X_t|\Im_{n,t-q_n}^{t+q_n}]\right| > \eta_n\right)\right]\\
&\quad\le E\left[I\left(b_{m_n}e^u-\eta_n < X_t < b_{m_n}e^u+\eta_n\right)\right]+P\left(\left|X_t-E[X_t|\Im_{n,t-q_n}^{t+q_n}]\right| > \eta_n\right)\\
&\quad\le\left[\bar{F}_t\left(b_{m_n}e^u-\eta_n\right)-\bar{F}_t\left(b_{m_n}e^u+\eta_n\right)\right]+\left\|X_t-E[X_t|\Im_{n,t-q_n}^{t+q_n}]\right\|_p^p/\eta_n^p\\
&\quad\le\left[\bar{F}_t\left(b_{m_n}e^u-\eta_n\right)-\bar{F}_t\left(b_{m_n}e^u+\eta_n\right)\right]+d_{n,t}^p\vartheta_{q_n}^p/\eta_n^p.
\end{aligned}\qquad\textbf{(A.8)}$$

The first inequality is due to the conditional expectations minimizing the mean squared error, and a trivial identity. The second follows from basic logic and a trivial inequality that exploits the indicator function. The third follows from Markov's inequality, and the fourth from L_p-NED, where $\vartheta_{q_n}=o(q_n^{-\lambda})$.

Define $\bar{\vartheta} := \sup_{q\ge1}\vartheta_q\in[0,1)$ and put $\eta_n=b_{m_n}e^u\vartheta_{q_n}^{1/2}$. Under Assumption B, $\bar{F}_t\left(b_{m_n}\right)=(m_n/n)\times\left(1+o\left(1/m_n^{1/2}\right)\right)$ and $\bar{F}_t\left(b_{m_n}z_{n,t}\right)/\bar{F}_t\left(b_{m_n}\right)=a_{n,t}^{-\alpha}\times\left(1+o\left(1/m_n^{1/2}\right)\right)$ for any array of nonstochastic positive real numbers $\{a_{n,t}\}$, $a_{n,t}\ge1$

(cf. Hsing, 1991, p. 1553). Therefore

$$\left[\bar{F}_t\left(b_{m_n}e^u-\eta_n\right)-\bar{F}_t\left(b_{m_n}e^u+\eta_n\right)\right]+d_{n,t}^p\vartheta_{q_n}^p/\eta_n^p$$

$$=\bar{F}_t\left(b_{m_n}\right)\frac{\bar{F}_t\left(b_{m_n}e^u\left(1-\vartheta_{q_n}^{1/2}\right)\right)}{\bar{F}_t\left(b_{m_n}\right)}-\bar{F}_t\left(b_{m_n}\right)\frac{\bar{F}_t\left(b_{m_n}e^u(1+\vartheta_{q_n}^{1/2})\right)}{\bar{F}_t\left(b_{m_n}\right)}$$

$$+\,b_{m_n}^{-p}d_{n,t}^p e^{-up}\vartheta_{q_n}^{p/2}$$

$$=(m_n/n)e^{-\alpha u}\left[(1-\vartheta_{q_n}^{1/2})^{-\alpha}-(1+\vartheta_{q_n}^{1/2})^{-\alpha}\right](1+o(1/m_n^{1/2}))$$

$$+\,b_{m_n}^{-p}d_{n,t}^p e^{-up}\vartheta_{q_n}^{p/2}$$

$$\leq K(m_n/n)e^{-\alpha u}\vartheta_{q_n}^{1/2}+b_{m_n}^{-p}d_{n,t}^p e^{-up}\vartheta_{q_n}^{p/2}$$

$$\leq K\times\max\{m_n/n,\ b_{m_n}^{-p}\}\times e^{-up}\left(1+d_{n,t}^p\right)\times\vartheta_{q_n}^{\min\{p,1\}/2}, \tag{A.9}$$

where the second inequality exploits $p<\alpha$, and the first follows from the mean value theorem:

$$(1-\vartheta_{q_n}^{1/2})^{-\alpha}-(1+\vartheta_{q_n}^{1/2})^{-\alpha}\leq\alpha 2(1-\bar{\vartheta}^{1/2})^{-\alpha-1}\vartheta_{q_n}^{1/2}\leq K\vartheta_{q_n}^{1/2}.$$

If $\lim_{x\to\infty}L(x)=K>0$, it is easy to show that $b_{m_n}^{-p}=K(m_n/n)^{p/\alpha}\geq K(m_n/n)$ from (3) and $p<\alpha$. Together (A.8), (A.9), and $\vartheta_{q_n}=o(q_n^{-\lambda})$ imply that

$$\left\|I\left(X_t>b_{m_n}e^u\right)-P\left(X_t>b_{m_n}e^u|\Im_{n,t-q_n}^{t+q_n}\right)\right\|_2$$

$$\leq\left\{e^{-up/2}\left(1+d_{n,t}^p\right)^{1/2}(m_n/n)^{p/2\alpha}\right\}\times o\left(q_n^{-\lambda\min\{p,1\}/4}\right).$$

Now suppose $p=\alpha-\iota$, $\sup_{n\geq 1}\sup_{1\leq t\leq n}d_{n,t}\leq K$, and $\lambda\geq 1/\min\{1,p/2\}$. Then the right-hand side is bounded by

$$K(m_n/n)^{1/2}e^{-up/2}\times o\left((n/m_n)^{\iota}q_n^{-\lambda\min\{p,1\}/4}\right)$$

$$=\left\{e^{-up/2}(m_n/n)^{1/2}\right\}\times o\left((n/m_n)^{\iota}q_n^{-1/2}\right)=f_{n,t}(u)\times\psi_{q_n},$$

where $\sup_{1\leq t\leq n}f_{n,t}(u)=e^{-up/2}(m_n/n)^{1/2}$ is Lebesgue integrable on $\mathbb{R}_+$. As long as $n/m_n=o(q_n^{\delta})$ for some $\delta>0$, then for sufficiently tiny $\iota>0$, $\psi_{q_n}=o((n/m_n)^{\iota}q_n^{-1/2})=o(q_n^{-1/2})$.

Claim (ii). See Hill (2008c). ■

Proof of Lemma 5. In lieu of Lemma 4, we need only prove that $\{X_t\}$ is $L_{\alpha-\iota}$-NED on $\{F_{n,t}\}$ with size $\lambda\geq 1/\min\{1,p/2\}$ and uniformly bounded constants $d_{n,t}\leq K$. Recall $E_{n,t}=[\epsilon_{t-i}]_{i=0}^{[q_n/2]}$ for $t=1,\ldots,n$ and 0 otherwise. Since $\Im_{n,-\infty}^t=\Im_{n,1}^t\subseteq G_{-\infty}^t$ and

$\Im^{+\infty}_{n,t+q_n} = \Im^{n}_{n,t+q_n} \subseteq G^{+\infty}_{t+[q_n/2]}$, it is easy to show the strong mixing property implies that ϵ_t is F-strong mixing size with $r/(r-2)$, $r > 2$.

Recall $\alpha > 1$, note that $\sup_{t\in\mathbb{Z}} ||\epsilon_i||_{\alpha-\imath} \leq K$ for tiny $\imath > 0$ by stationarity, and by construction $\Im^{t+q_n}_{n,t-q_n} = \sigma(\epsilon_\tau : \max\{1-[q_n/2],\, t-q_n-[q_n/2]\} \leq \tau \leq \min\{t+q_n, n\})$. Use $\sigma(\epsilon_\tau : \tau \leq t-i)$-measurability of $\pi_{t,i}$, $\sup_{t\in\mathbb{Z}} |\pi_{t,i}| \leq |\pi_i| = O(i^{-\mu})$ for some $\mu > 1/\min\{1, p/2\}$ by the stipulations of Example 4 and Minkowsi's and conditional Jensen's inequalities to deduce

$$\left\|x_t - E\left[x_t|\Im^{t+q_n}_{n,t-q_n}\right]\right\|_{\alpha-\imath}$$

$$\leq \sum_{i=[q_n/2]}^{\infty} \left\|\pi_{t,i}\epsilon_{t-i} - E\left[\pi_{t,i}\epsilon_{t-i}|\{\epsilon_\tau\}^{\min\{t+q_n,n\}}_{\max\{1-[q_n/2],t-q_n-[q_n/2]\}}\right]\right\|_{\alpha-\imath}$$

$$\leq K \sum_{i=[q_n/2]}^{\infty} \left\|\pi_{t,i}\epsilon_{t-i}\right\|_{\alpha-\imath} \leq K \sum_{i=[q_n/2]}^{\infty} |\pi_i| = O\left(q_n^{-\mu}\right).$$

Therefore $||x_t - E[x_t|\Im^{t+q_n}_{n,t-q_n}]||_{\alpha-\imath} \leq d_{n,t} \times o\left(q_n^{-\lambda}\right)$ for $d_{n,t} = K$ and $\lambda \geq 1/\min\{1, p/2\}$. ■

Proof of Lemma 6. See Hill (2008c). ■

Proof of Lemma 7.

Step 1 ($X_t \sim (2)$)**.** Use $\epsilon_t \sim (2)$ with index α, the convolution tail property of $\{\epsilon_t\}$ and $\sum_{i=0}^{\infty} \pi_i^{\alpha} < \infty$ to deduce, as $z \to \infty$,

$$P(X_t > z) = P\left(\sum_{i=0}^{\infty} \pi_i \epsilon_{t-i} > z\right) \sim \sum_{i=0}^{\infty} \pi_i^{\alpha} P\left(\epsilon_{t-i} > z\right) = \sum_{i=0}^{\infty} \pi_i^{\alpha} \times z^{-\alpha} L(z).$$

Therefore $X_t \sim (2)$ with index α. Further, since by construction of $\{m_n, b_{m_n}\}$

$$\lim_{n\to\infty} \frac{n}{m_n} P\left(X_t > b_{m_n}\right) = \sum_{i=0}^{\infty} \pi_i^{\alpha} \lim_{n\to\infty} \frac{n}{m_n} P\left(\epsilon_{t-i} > b_{m_n}\right) = 1,$$

identical distributedness implies $(n/m_n)P(\epsilon_t > b_{m_n}) \sim (\sum_{i=0}^{\infty} \pi_i^{\alpha})^{-1}$.

Step 2 (L_2-E-NED). For notational clarity assume $q_n < t$. A similar argument applies for all $1 \leq t \leq n$. By iterated expectations and the Cauchy-Schwartz inequality,

$$E\left(I(X_t > b_{m_n}e^u) - P(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n})\right)^2$$

$$= P\left(X_t > b_{m_n}e^u\right) - 2E\left[I(X_t > b_{m_n}e^u)P(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n})\right]$$

$$+ E\left[P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)^2\right]$$

$$= P\left(X_t > b_{m_n}e^u\right) - E\left[P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)^2\right]$$

$$= E\left[P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right) \times \left(1 - P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right)\right]$$

$$\leq \left\|P\left(X_t > b_{m_n}e^u)|\Im^{t+q_n}_{n,t-q_n}\right)\right\|_2 \times \left\|\left(1 - P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right)\right\|_2$$

$$\leq \left\|P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right\|_2. \quad \textbf{(A.10)}$$

Let $\epsilon^*_{t,a}$ denote a random draw from the distribution governing $\sum_{i=0}^{a}\pi_i\epsilon_{t-i}$, $a \in \mathbb{N}$, and note $\pi_i \geq 0$ and $\epsilon_t \geq 0$ a.s. $\forall t$ imply $\epsilon^*_{t,a} \geq 0$ a.s. An argument similar to Step 1, and the proof of Lemma 5, reveals as $n \to \infty$,

$$P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right) = P\left(\sum_{i=q_n+[q_n/2]+1}^{\infty}\pi_i\epsilon_{t-i} > b_{m_n}e^u - \epsilon^*_{t,q_n+[q_n/2]}\right)$$

$$\leq \left(1 - \frac{\epsilon^*_{t,q_n+[q_n/2]}}{b_{m_n}e^u}\right)^{-\alpha}\sum_{i=q_n+1}^{\infty}\pi_i^{\alpha}P\left(\epsilon_{t-i} > b_{m_n}e^u\right). \quad \textbf{(A.11)}$$

Since $\sum_{i=0}^{\infty}\pi_i^{\alpha} < \infty$, and $(n/m_n)P(\epsilon_t > b_{m_n}) \sim (\sum_{i=0}^{\infty}\pi_i^{\alpha})^{-1}$ by Step 1, for every $\varepsilon > 0$ and $a \in \mathbb{N}$,

$$P\left(\epsilon^*_{t,a}/b_{m_n} > \varepsilon\right) \leq P\left(\sum_{i=0}^{\infty}\pi_i\epsilon_{t-i} > \varepsilon \times b_{m_n}\right) \sim \varepsilon^{-\alpha}\sum_{i=0}^{\infty}\pi_i^{\alpha}P\left(\epsilon_{t-i} > b_{m_n}\right)$$

$$= O(m_n/n),$$

hence $\epsilon^*_{t,q_n+[q_n/2]}/b_{m_n} \xrightarrow{p} 0$.

Now use (A.11), Minkowski's inequality, $(n/m_n)P(\epsilon_t > b_{m_n}) \sim (\sum_{i=0}^{\infty}\pi_i^{\alpha})^{-1}$, and $||(1 - \epsilon^*_{t,q_n}/b_{m_n}e^u)^{-\alpha}||_2 \xrightarrow{p} 1$ by $\epsilon^*_{t,q_n}/b_{m_n} \xrightarrow{p} 0$ and the Helly-Bray theorem to deduce

$$\lim_{n\to\infty}\frac{n}{m_n}\left\|P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right\|_2$$

$$\leq \lim_{n\to\infty}\left\|\left(1 - \frac{\epsilon^*_{t,q_n+[q_n/2]}}{b_{m_n}e^u}\right)^{-\alpha}\right\|_2 \times \sum_{i=q_n+1}^{\infty}|\pi_i|^{\alpha}\frac{n}{m_n}P\left(\epsilon_{t-i} > b_{m_n}\right) \times e^{-\alpha u}$$

$$= \lim_{n\to\infty}\sum_{i=q_n+1}^{\infty}|\pi_i|^{\alpha}\frac{n}{m_n}P\left(\epsilon_{t-i} > b_{m_n}\right)e^{-\alpha u} = \sum_{i=q_n+1}^{\infty}|\pi_i|^{\alpha}\left(\sum_{i=0}^{\infty}|\pi_i|^{\alpha}\right)^{-1}e^{-\alpha u}. \quad \textbf{(A.12)}$$

Together, (A.10) and (A.12) imply, for any $r > 2$,

$$\left\|I\left(X_t > b_{m_n}e^u\right) - P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right\|_2$$

$$\leq \left\{e^{-\alpha u/2}\left(\frac{m_n}{n}\right)^{1/2}\right\} \times \left\{\left(\frac{\sum_{i=q_n+1}^{\infty}\pi_i^{\alpha}}{\sum_{i=0}^{\infty}\pi_i^{\alpha}}\right)^{1/2}\right\} = f_{n,t}(u) \times \psi_{q_n},$$

say, where $\sup_{1\le t\le n} f_{n,t}(u) = e^{-\alpha u/2}(m_n/n)^{1/2}$ is Lebesgue integrable on $\mathbb{R}_+$, and $\psi_{q_n} \in [0,1]$. ■

Proof of Lemma 8. The tail of $X_t = \sum_{j=0}^{\infty}\beta^j\epsilon_t^{(j)} = \epsilon_t + \sum_{j=1}^{\infty}\beta^j\epsilon_t^{(j)} = \epsilon_t + X_t^*$ is dominated by $X_t^* \sim$ (2) with index $\alpha/2$ (cf. Davis and Resnick, 1996), hence it suffices to demonstrate that X_t^* satisfies Lemma 7. Since ϵ_t is i.i.d., straightforward generalizations of Corollaries 2.3 and 2.4 of Davis and Resnick (1996) reveal $P(\beta^j\epsilon_t^{(j)} > x) \sim \beta^{j\alpha/2}\left(E|\epsilon_t|^{\alpha/2}\right)^{j-1}P(\epsilon_t^2 > x)$ *for each* $j \ge 1$ and $P(X_t^* > x) \sim \sum_{j=1}^{\infty}\beta^{j\alpha/2}\left(E|\epsilon_t|^{\alpha/2}\right)^{j-1}P(\epsilon_t^2 > x)$. But this implies that $\{\beta^j\epsilon_t^{(j)}\}_{j=1}^{\infty}$ has the same tail behavior as some stochastic sequence $\{\beta^{j\alpha/2}(E|\epsilon_t|^{\alpha/2})^{j-1}z_{t-j}\}_{j=1}^{\infty}$ where $\lim_{x\to\infty}P(z_{t-j} > x)/P(\epsilon_t^2 > x) = 1$ for all $j \in \mathbb{N}$ and $\{z_{t-j}\}_{j=1}^{\infty}$ has the convolution tail property $P(\sum_{j=1}^{\infty}a_j z_{t-j} > x) \sim \sum_{j=1}^{\infty}P(a_i\epsilon_{t-i} > x)$ for any sequence of real numbers $\{a_i\}$, $\sum_{i=0}^{\infty}|a_i|^{\alpha} < \infty$. Therefore X_t^* satisfies the conditions of Lemma 7. ■

APPENDIX B: Supporting Lemmas B1–B10

Let ρ be any number in an arbitrary neighborhood of 1, and write $T_{m_n,t} := T_{m_n,t}(\omega, u/m_n^{1/2})$. Lemmas B.1 and B.2 characterize moment and memory properties of the tail arrays $\{U_{m_n,t}, I_{m_n,t}(u)\}$, where $U_{m_n,t} := (\ln(X_t/b_{m_n}))_+ - E[[(\ln(X_t/b_{m_n}))_+]$ and $I_{m_n,t}(u) := I(X_t > b_{m_n}e^u) - P(X_t > b_{m_n}e^u)$, $u \ge 0$.

LEMMA B.1.

(i) *Under Assumption A.1,* $\{U_{m_n,t}, \Im_{n,t}\}$ *and* $\{I_{\rho m_n,t}(u), \Im_{n,t}\}$ *form* L_2*-mixingale arrays with common coefficients* φ_{q_n} *and constants* $\{e_{n,t}^*, e_{n,t}(u)\} = O\left((m_n/n)^{1/2}\right)$, *where* $e_{n,t}^* = \int_0^{\infty}e_{n,t}(u)du$, *provided* $e_{n,t}(u)$ *is Lebesgue integrable on* $\mathbb{R}_+$.

(ii) *Under Assumption A.2,* $\{U_{m_n,t}, I_{\rho m_n,t}(u)\}$ *are* L_2*-NED on* $\{\Im_{n,t}\}$ *with common coefficients* $\psi_{n,q_n}^* = (m_n/n)^{1/2-1/r}\psi_{q_n}$ *and constants* $\{f_{n,t}^*, f_{n,t}^*(u)\}$, *where* $f_{n,t}^*(u) = (n/m_n)^{1/2-1/r}f_{n,t}(u)$ *and* $f_{n,t}^* = K(n/m_n)^{1/2-1/r}\int_0^{\infty}f_{n,t}(u)du$, *provided* $f_{n,t}(u)$ *is Lebesgue integrable on* $\mathbb{R}_+$. *In particular,* $\sup_{1\le t\le n}f_{n,t}^*$ *and* $\sup_{1\le t\le n}\sup_{u\ge 0}f_{n,t}^*(u)$ *are* $O((m_n/n)^{1/r})$.

LEMMA B.2. *The tail arrays* $\{U_{m_n,t}\}$ *and* $\{I_{\rho m_n,t}(u)\}$ *are* L_r*-bounded for any* $r \ge 1$*:*

$$\lim_{n\to\infty}\left(\frac{n}{m_n}\right)^{1/r}\left\|I_{\rho m_n,t}(u)\right\|_r \le A_r(u) < \infty \quad \textit{and}$$

$$\lim_{n\to\infty}\left(\frac{n}{m_n}\right)^{1/r}\left\|U_{m_n,t}\right\|_r \le B_r < \infty,$$

where $A_r : \mathbb{R} \to \mathbb{R}_+$ *is* p*-integrable with respect to Lebesgue measure on* $\mathbb{R}_+$ *for any* $p > 0$, *and uniformly bounded on* $\mathbb{R}_+$. *In particular,* $\sup_{u\ge 0}\left\|I_{\rho m_n,t}(u)\right\|_r = O((m_n/n)^{1/r})$.

Define

$$\Delta W_{m_n,t} := \ln\left(X_t/b_{m_n}\right) \times I\left(X_t > X_{(m_n+1)}\right) - \ln\left(X_t/b_{m_n}\right)_+.$$

Lemmas B.3 and B.4 establish key Lipschitz properties and a decomposition for proving $\hat{\alpha}_{m_n}^{-1}$ is uniformly consistent for α^{-1}.

LEMMA B.3. *Define $m_* := \inf_{\phi \in \Phi} m_n(\phi)$ and let Assumptions A.1 and B hold. For each $\bar{y}_{m_n} \in \{1/m_n \sum_{t=1}^n U_{m_n,t}, 1/m_n \sum_{t=1}^n I_{m_n,t}(1, u/m_n^{1/2}), 1/m_n \sum_{t=1}^n \Delta W_{m_n,t}, \hat{\alpha}_{m_n}^{-1}\}$ there exists a stochastic array $\{B_{n,t}\}$ that is not a function of $\phi \in \Phi$ and that satisfies $1/m_* \sum_{t=1}^n E[B_{n,t}] = O(1)$, such that $|\bar{y}_{m_n(\phi)} - \bar{y}_{m_n(\phi')}| \leq 1/m_* \sum_{t=1}^n B_{n,t} \times |\phi - \phi'|$ a.s. for all $\phi, \phi' \in \Phi$.*

LEMMA B.4. *Under Assumptions A.1 and B,*

$$\hat{\alpha}_{m_n}^{-1} - \alpha^{-1} = \frac{1}{m_n} \sum_{t=1}^n \left\{ U_{m_n,t} - \alpha^{-1} I_{m_n,t}\left(u/m_n^{1/2}\right) \right\} + \frac{1}{m_n} \sum_{t=1}^n \Delta W_{m_n,t} + o\left(1/m_n^{1/2}\right)$$

where $o\left(1/m_n^{1/2}\right)$ is deterministic.

LEMMA B.5. *Let $\{X_{n,t}\}$ be a mean-zero stochastic array with $\sigma_n := ||\sum_{t=1}^n X_{n,t}||_2 > 0$ uniformly in n. Define $Z_{n,i} := \sum_{t=(i-1)k_n+l_n+1}^{ik_n} X_{n,t}$ and $\tilde{F}_{n,i} := \sigma(\{E_{n,\tau}(a_n) : \tau \leq ik_n\})$ and let the sequences $\{l_n, k_n, r_n\}$ be as in (A.3). Then $\sum_{t=1}^n X_{n,t}/\sigma_n \Longrightarrow N(0,1)$ under the following conditions:*

(a) $\sum_{t=r_n k_n+1}^n X_{n,t} \xrightarrow{p} 0$,

(b) $\sum_{i=1}^{r_n} \sum_{t=(i-1)k_n+1}^{(i-1)k_n+l_n} X_{n,t} \xrightarrow{p} 0$,

(c) $\sum_{i=1}^{r_n} E[Z_{n,i}|\tilde{F}_{n,i-1}] \xrightarrow{p} 0$,

(d) $\sum_{i=1}^{r_n} (Z_{n,i} - E[Z_{n,i}|\tilde{F}_{n,i-1}]) \xrightarrow{p} 0$,

(e) $\sum_{i=1}^{r_n} (E[Z_{n,i}|\tilde{F}_{n,i}] - E[Z_{n,i}|\tilde{F}_{n,i-1}])^2/\sigma_n \xrightarrow{p} 1$,

(f) $\sum_{i=1}^{r_n} E[W_{n,i}^2 I(|W_{n,i}| > \varepsilon)] \xrightarrow{p} 0 \; \forall \varepsilon > 0$, *where* $W_{n,i} := E[Z_{n,i}|\tilde{F}_{n,i}] - E[Z_{n,i}|\tilde{F}_{n,i-1}]$.

LEMMA B.6. *If $\{T_{m_n,t}, \Im_{n,t}\}$ forms an L_2-mixingale array with size $1/2$ and constants $c_{n,t}$, $\sup_{1 \leq t \leq n} c_{n,t} = O\left(n^{-1/2}\right)$, then for the sequences $\{l_n, k_n, r_n\}$ defined in (A.3),*

$$\lim_{n\to\infty} \left| \sum_{i=1}^{r_n} \sum_{j=i+1}^{r_n} \sum_{t=(i-1)k_n+l_n+1}^{ik_n} \sum_{s=(j-1)k_n+l_n+1}^{jk_n} E\left[T_{m_n,s} T_{m_n,t}\right] \right| = 0.$$

Recall $Z_{n,i} = \sum_{t=(i-1)k_n+l_n+1}^{ik_n} T_{m_n,t}$.

LEMMA B.7. *Under Assumptions A.2 and B, $\sum_{i=1}^{r_n} (Z_{n,i}^2 - E[Z_{n,i}^2]) \xrightarrow{p} 0$ and $\sum_{i=1}^{r_n} Z_{n,i}^2/\sigma_{m_n}^2(\omega) \xrightarrow{p} 1$.*

Compactly write the kernel function $w_{s,t,n} := w((s-t)/\gamma_n)$ from Theorem 3, and

$$\tilde{\sigma}_{m_n}^2 := \frac{1}{m_n} \sum_{s,t=1}^n w_{n,s,t} Y_{m_n,s} Y_{m_n,t}, \quad \text{where } Y_{m_n,t} := U_{m_n,t} - \frac{m_n}{n} \ln\left(\frac{X_{(m_n+1)}}{b_{m_n}}\right).$$

LEMMA B.8. *Under the conditions of Theorem 3,* $|\hat{\sigma}^2_{m_n} - \tilde{\sigma}^2_{m_n}| \xrightarrow{p} 0$.

LEMMA B.9. *Under the conditions of Theorem 3,* $|\tilde{\sigma}^2_{m_n} - \sigma^2_{m_n}(1,-1)| \xrightarrow{p} 0$.

LEMMA B.10. *Under the conditions of Theorem 3,* $\{m_n^{-1/2} Y_{m_n,t}, \Im_{n,t}\}$ *forms an* L_2-*mixingale array with* $O(n^{-1/2})$-*constants and size* $1/2$.

Proof of Lemma B.1. We will prove the E-NED assertion, the E-MIXL proof being similar. Since $\{\rho m_n\}$ forms an intermediate order sequence, under Assumption A.2 $\{I_{\rho m_n,t}(u)\}$ is by construction L_2-NED on $\{\Im_{n,t}\}$: $||I_{\rho m_n,t}(u) - E[I_{\rho m_n,t}(u)|\Im^{t+q_n}_{n,t-q_n}]||_2 \leq \{(n/m_n)^{1/2-1/r} f_{n,t}(u)\} \times \{(m_n/n)^{1/2-1/r}\psi_{q_n}\} = f^*_{n,t}(u)\psi^*_{n,q_n}$, say, where the claimed properties of $f^*_{n,t}(u)$ and ψ^*_{n,q_n} follow from Assumption A.2.

Now consider $U_{m_n,t}$, define $P_{n,t}(u) := I(X_t > b_{m_n}e^u) - P(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n})$, invoke Assumption A.2, and let the E-NED constants $f_{n,t}(u)$ be Lebesgue integrable on $\mathbb{R}_+$. Then

$$\begin{aligned}
&\left\| U_{m_n,t} - E[U_{m_n,t}|\Im^{t+q_n}_{n,t-q_n}] \right\|_2 \\
&= \left\| (\ln(X_t/b_{m_n}))_+ - E[(\ln(X_t/b_{m_n}))_+ |\Im^{t+q_n}_{n,t-q_n}] \right\|_2 \\
&= \left[E\left(\int_0^\infty \left[I(X_t > b_{m_n}e^u) - P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right] du \right)^2 \right]^{1/2} \\
&= \left[E\int_0^\infty \int_0^\infty P_{n,t}(u_1)P_{n,t}(u_2)du_1du_2 \right]^{1/2} \\
&= \left[\int_0^\infty \int_0^\infty E\left[P_{n,t}(u_1)P_{n,t}(u_2)\right] du_1du_2 \right]^{1/2} \\
&\leq \left[\int_0^\infty \int_0^\infty \left\|P_{n,t}(u_1)\right\|_2 \left\|P_{n,t}(u_2)\right\|_2 du_1du_2 \right]^{1/2} \\
&= \int_0^\infty \left\|P_{n,t}(u)\right\|_2 du = \int_0^\infty \left\| I(X_t > b_{m_n}e^u) - P\left(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n}\right)\right\|_2 du \\
&\leq \left(\int_0^\infty f_{n,t}(u)du \right) \times \psi_{q_n} = \left(\int_0^\infty f^*_{n,t}(u)du \right) \times \psi^*_{n,q_n} = f^*_{n,t} \times \psi^*_{n,q_n},
\end{aligned}$$

say. The second equality follows from the identity $(\ln(X_t/b_m))_+ = \int_0^\infty I(X_t > b_{m_n}e^u)\,du$ and the Fubini-Tonelli theorem: $E[(\ln(X_t/b_{m_n}))_+|\Im^{t+q_n}_{n,t-q_n}] = E[\int_0^\infty I(X_t > b_{m_n}e^u)\,du\,|\Im^{t+q_n}_{n,t-q_n}] = \int_0^\infty P(X_t > b_{m_n}e^u|\Im^{t+q_n}_{n,t-q_n})$. The fourth equality follows from the Fubini-Tonelli theorem. The first inequality is Cauchy-Schwartz's. The last inequality follows from Step 1 and Lebesgue integrability of $f_{n,t}(u)$. The asserted properties of $f^*_{n,t} = \int_0^\infty f^*_{n,t}(u)\,du = (n/m_n)^{1/2-1/r}\int_0^\infty f_{n,t}(u)\,du$ follow from Assumption A.2. ∎

Proof of Lemma B.2. Use (1)–(3) to deduce for any $u \in \mathbb{R}$, any ρ in an arbitrary neighborhood of 1, and any $r \geq 1$,

$$\lim_{n\to\infty}\left(\frac{n}{m_n}\right)^{1/r}||I_{\rho m_n,t}(u)||_r \leq 2\lim_{n\to\infty}\left(\frac{n}{m_n}\right)^{1/r}P\left(X_t > b_{\rho m_n}e^u\right)^{1/r}$$

$$= 2\lim_{n\to\infty}\left[\frac{n}{m_n}P\left(X_t > b_{\rho m_n}\right)\frac{P\left(X_t > b_{\rho m_n}e^u\right)}{P\left(X_t > b_{\rho m_n}\right)}\right]^{1/r}$$

$$= 2\rho^{1/r}e^{-\alpha u/r} =: A_r(u) < \infty.$$

Trivially, $\sup_{u\geq 0} A_r(u) \leq K < \infty$, $\int_0^\infty A_r(u)^p du \leq K\int_0^\infty e^{-\alpha up/r}du < \infty$ for any $p > 0$. Similarly, for any $r \geq 1$ it is easy to show under (1)–(3) (e.g., Hsing, 1991, eqn. 1.5) that

$$\lim_{n\to\infty}\left(\frac{n}{m_n}\right)^{1/r}||U_{m_n,t}||_r \leq 2\lim_{n\to\infty}(n/m_n)^{1/r}\left\|\left(\ln\left(X_t/b_{m_n}\right)\right)_+\right\|_r$$

$$= 2\left(\int_0^\infty e^{-\alpha u^{1/r}}du\right)^{1/r} =: B_r < \infty.$$

■

Proof of Lemma B.3. See Hill (2009b, Lem. B.3). ■

Proof of Lemma B.4. Write

$$\hat{\alpha}_{m_n}^{-1} - \alpha^{-1} = \frac{1}{m_n}\sum_{j=1}^{m_n}\ln\left(\frac{X_{(j)}}{b_{m_n}}\right) - E\left[\frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+\right]$$

$$-\ln\left(\frac{X_{(m_n+1)}}{b_{m_n}}\right) + \left(E\left[\frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+\right] - \alpha^{-1}\right). \quad \textbf{(B.1)}$$

Under Assumption B, the last term satisfies (Hsing, 1991, p. 1554)

$$E\left[\frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+\right] - \alpha^{-1} = o(1/m_n^{1/2}),$$

and by construction, the first term can be written

$$\frac{1}{m_n}\sum_{j=1}^{m_n}\ln\left(\frac{X_{(j)}}{b_{m_n}}\right) = \frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+ + \frac{1}{m_n}\sum_{t=1}^{n}\Delta W_{t,m_n},$$

where $\Delta W_{t,m_n} := \ln(X_t/b_{m_n}) \times I(X_t > X_{(m_n+1)}) - (\ln(X\dot{}t/b\dot{}m\dot{}n))_+$.

Further, it is easy to show that the third term in (B.1) satisfies, for all $u \in \mathbb{R}$,

$$m_n^{1/2}\ln\left(\frac{X_{(m_n+1)}}{b_{m_n}}\right) = u \iff \alpha^{-1}\frac{1}{m_n^{1/2}}\sum_{t=1}^{n} I_{m_n,t}(u/m_n^{1/2}) = u + o(1)$$

and deterministic $o(1)$ (Hsing, 1991, p. 1553). Therefore

$$\hat{\alpha}_{m_n}^{-1} - \alpha^{-1} = \frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+ - E\left[\frac{1}{m_n}\sum_{t=1}^{n}\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+\right]$$

$$+\frac{1}{m_n}\sum_{t=1}^{n}\ln\left(\frac{X_t}{b_{m_n}}\right) \times \left[I\left(X_t > X_{(m_n+1)}\right) - I\left(X_t > b_{m_n}\right)\right]$$

$$-\alpha^{-1}\frac{1}{m_n}\sum_{t=1}^{n} I_{m_n,t}(u/m_n^{1/2}) + o(1/m_n^{1/2})$$

$$= \frac{1}{m_n}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1} I_{m_n,t}(u/m_n^{1/2})\right) + \frac{1}{m_n}\sum_{t=1}^{n}\Delta W_{t,m_n} + o(1/m_n^{1/2}). \quad \blacksquare$$

Proof of Lemma B.5. See de Jong (1997, Lem. 1). ■

Proof of Lemma B.6. See de Jong (1997, Lem. 4). ■

Proof of Lemma B.7. See Hill (2009b, Lem. B.7). ■

Proof of Lemma B.8. Recall $Y_{m_n,t} := U_{m_n,t} - (m_n/n)\ln(X_{(m_n+1)}/b_{m_n})$, write $w_{n,s,t} := w(|s-t|/\gamma_n)$, define

$$A_{n,t} := \left(\ln\left(\frac{X_t}{X_{(m_n+1)}}\right)\right)_+ - \left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+ + \frac{m_n}{n}\ln\left(\frac{X_{(m_n+1)}}{b_{m_n}}\right)$$

$$B_n := \frac{m_n}{n} \times \left\{\frac{n}{m_n}\left(E\left[\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+\right] - \alpha^{-1}\right) + \left(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}\right)\right\},$$

and decompose $\hat{\sigma}_{m_n}^2 = \tilde{\sigma}_{m_n}^2 + R_n$, where

$$\tilde{\sigma}_{m_n}^2 = \frac{1}{m_n}\sum_{s,t=1}^{n} w_{n,s,t} Y_{m_n,s} Y_{m_n,t},$$

$$R_n = \frac{1}{m_n}\sum_{s,t=1}^{n} w_{n,s,t} A_{n,s} A_{n,t} + B_n^2 \frac{1}{m_n}\sum_{s,t=1}^{n} w_{n,s,t} + 2\frac{1}{m_n}\sum_{s,t=1}^{n} w_{n,s,t} A_{n,s} Y_{m_n,t}$$

$$+ 2B_n \frac{1}{m_n}\sum_{s,t=1}^{n} w_{n,s,t} Y_{m_n,t} + 2B_n \frac{1}{m_n}\sum_{s,t=1}^{n} w_{n,s,t} A_{n,t}.$$

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

By cases it is easy to show that $|A_{n,t}| \le |\ln(X_{(m_n+1)}/b_{m_n})|$, and Assumption B implies

$$m_n^{1/2}\left\{\frac{n}{m_n}E\left[\left(\ln\left(\frac{X_t}{b_{m_n}}\right)\right)_+\right] - \alpha^{-1}\right\} = o(1). \tag{B.2}$$

Now apply Lemma 3 and Theorem 2 to deduce respectively that

$$\|A_{n,t}\|_2 \le \left\|\ln\left(\frac{X_{(m_n+1)}}{b_{m_n}}\right)\right\|_2 = O(m_n^{-1/2}) \quad \text{and} \quad \|B_n\|_2 = O(m_n^{1/2}/n). \tag{B.3}$$

Similarly, Lemma 3 and the Lemma B.2 moment bounds imply that

$$\left\|Y_{m_n,t}\right\|_2 \leq \left\|U_{m_n,t}\right\|_2 + \frac{m_n}{n}\left\|\ln\left(\frac{X_{(m_n+1)}}{b_{m_n}}\right)\right\|_2 = O((m_n/n)^{1/2}). \quad \textbf{(B.4)}$$

Finally, by supposition,

$$\frac{1}{m_n}\sum_{s,t=1}^{n}\left|\omega_{n,s,t}\right| = o\left(\gamma_n n/m_n\right) = o(n^{1/2}). \quad \textbf{(B.5)}$$

Together (B.2)–(B.5), the Minkowski and Cauchy-Schwartz inequalities, and $m_n/n^{1/2} \to \infty$ by supposition give

$$\|R_n\|_1 = o(n^{1/2}) \times \left\{O(m_n^{-1}) + O(m_n/n^2) + O(n^{-1/2}) + O(m_n/n^{3/2}) + O(n^{-1})\right\}$$

$$= o(n^{1/2}/m_n) + O(m_n/n^{3/2}) + o(1) + O(m_n/n) + O(n^{-1/2}) = o(1). \quad \blacksquare$$

Proof of Lemma B.9. Write $I_{m_n,t} := I_{m_n,t}(u/m_n^{1/2})$, recall $Y_{m_n,t} := U_{m_n,t} - (m_n/n)$ $\ln(X_{(m_n+1)}/b_{m_n})$, and

$$\sigma^2_{m_n}(1,-1) := E\left(\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\left(U_{m_n,t} - \alpha^{-1}I_{m_n,t}\right)\right)^2 \quad \text{and}$$

$$\tilde{\sigma}^2_{m_n} := \frac{1}{m_n}\sum_{s,t=1}^{n}\omega_{n,s,t}Y_{m_n,s}Y_{m_n,t}.$$

We will prove $|\tilde{\sigma}^2_{m_n} - E(1/m_n^{1/2}\sum_{t=1}^{n}Y_{m_n,t})^2| = o_p(1)$ and $|E(1/m_n^{1/2}\sum_{t=1}^{n}Y_{m_n,t})^2 - \sigma^2_{m_n}(1,-1)| = o_p(1)$.

Step 1. We will verify Assumptions 1–3 of de Jong and Davidson (2000) (JD) to show

$$\left|\tilde{\sigma}^2_{m_n} - E\left(\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}Y_{m_n,t}\right)^2\right| \to 0. \quad \textbf{(B.6)}$$

JD's Assumption 1 holds by the statement of the lemma.

By Lemma B.10, $\{m_n^{-1/2}Y_{m_n,t}, \Im_{n,t}\}$ forms an L_2-mixingale array with size 1/2 and constants $c^2_{n,t} = Kn^{-1/2}$. Thus JD's Assumption 2 is satisfied. [Equation (2.6) of de Jong and Davidson (2000) is only *sufficient* for the mixingale property to hold, but not *necessary*. By the proof of Lemma B.10 in Hill (2009b) $\left\{m_n^{-1/2}Y_{m_n,t}\right\}$ is L_2-NED on $\{\Im_{n,t}\}$ with $O((m_n/n)^{1/r})$-constants and $o\left((m_n/n)^{1/2-1/r}q_n^{-1/2}\right)$-coefficients, and $\left\{m_n^{-1/2}Y_{m_n,t}, \Im_{n,t}\right\}$ forms an L_2-mixingale sequences with constants and coefficients $c_{n,t} \times \xi_{q_n} = Kn^{-1/2} \times o(q_n^{-1/2})$. With these properties in hand, each of de Jong and Davidson's arguments that exploit their (2.6) go through.]

Finally, JD's Assumption 3 is satisfied by $\gamma_n \max_{1\leq t\leq n} c^2_{n,t} = o(1)$ given $\gamma_n = o(n)$. This proves (B.6).

Step 2. Define $U_{m_n} := m_n^{-1/2}\sum_{t=1}^{n} U_{m_n,t}$, $I_{m_n} := \alpha^{-1} m_n^{-1/2}\sum_{t=1}^{n} I_{m_n,t}$, and $B_{m_n} = m_n^{1/2}\ln(X_{(m_n+1)}/b_{m_n})$. Arguments in Hsing (1991, p. 1553) and the Helly-Bray theorem imply under Assumption B that

$$\left| E\left(\frac{1}{m_n^{1/2}}\sum_{t=1}^{n} Y_{m_n,t}\right)^2 - \sigma_{m_n}^2(1,-1)\right|$$
$$= \left| E\left(U_{m_n} - B_{m_n}\right)^2 - E\left(U_{m_n} - I_{m_n}\right)^2\right|$$
$$\leq 2\left\|U_{m_n}\right\|_2 \left\|B_{m_n} - I_{m_n}\right\|_2 + \left|E\left(B_{m_n}\right)^2 - E\left(I_{m_n}\right)^2\right| = O\left(m_n^{1/2} g\left(b_{m_n}\right)\right) = o(1). \qquad \textbf{(B.7)}$$

Together, (B.6) and (B.7) imply $|\tilde{\sigma}_{m_n}^2 - \sigma_{m_n}^2(1,-1)| = o_p(1)$ as claimed. ∎

Proof of Lemma B.10. See Hill (2009b, Lem. B.10). ∎

APPENDIX C: Symbols

The following table displays the *most frequently* used symbols and variables in order of appearance, their definitions, and the section(s) in which they first appear. If the symbol or variable first appears in a numbered equation, definition, etc., that information is also given. Consult the first appearance for a complete definition.

Symbol	Definition	Section §, (eqn.), etc.
$F_t(x), \bar{F}_t(x)$	$P(X_t \leq x), P(X_t > x)$	§1
$L(x)$	slowly varying component in $\bar{F}_t(x) = x^{-\alpha} L(x)$	§1, (2)
$X_{(i)}$	sample order statistic: $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$	§1
$\{m_n\}$	sequence of integers: $m_n \to \infty$, $m_n = o(n)$	§1
b_{m_n}	threshold sequence: $n/m_n P\left(X_t > b_{m_n}\right) \to 1$	§2, (3)
$\{E_{n,t}\}$	stochastic triangular array, mixing functional of ϵ_t	§2.1, §2.2
$\left\{\Im_{n,t}\right\}$	triangular σ-array induced by $\{E_{n,t}\}$	§2.1
$\{q_n\}$	sequence of displacements, $q_n \to \infty$, $q_n = o(n)$	§2.1
$e_{n,t}(u), \varphi_{q_n}$	E-MIXL constants and coefficients	§2.1, Defn: E-MIXL
$f_{n,t}(u), \psi_{q_n}$	E-NED constants and coefficients	§2.1, Defn: E-NED
ϵ_t	E-NED base	§2.2
G_t	σ-field induced by ϵ_t	§2.2
$\varepsilon_{n,q_n}, \varpi_{n,q_n}$	F-strong and F-uniform mixing coefficients	§2.2
g	slow variation with remainder component of $L(x)$	§3, Assumption B
$\{m_n(\phi)\}$	sequence of Lipschitz integer functions	§3.1, (4)-(5)
h_n	$O(\inf_{\phi\in\Phi} m_n(\phi))$-sequence for Lipschitz $m_n(\phi)$	§3.1, (5)
$U_{m_n,t}$	$\left(\ln\left(X_t/b_{m_n}\right)\right)_+ - E\left[\left(\ln\left(X_t/b_{m_n}\right)\right)_+\right]$	§3.1, (6)
$I_{m_n,t}(u)$	$I\left(X_t > b_{m_n} e^u\right) - E\left[I\left(X_t > b_{m_n} e^u\right)\right]$	§3.1, (6)
$T_{m_n,t}(\omega, u)$	$1/m_n^{1/2}\left[\omega_1 U_{m_n,t} - \omega_2 \alpha^{-1} I_{m_n,t}(u)\right]$	§3.2, (8)
$\sigma_{m_n}^2(\omega)$	$\sigma_{m_n}^2(\omega_1,\omega_2) := E\left(\sum_{t=1}^{n} T_{m_n,t}\left(\omega, u/m_n^{1/2}\right)\right)^2$	§3.2, (9)

Symbol	Definition	Section §, (eqn.), etc.
$d_{n,t}$	L_2-NED constants for $\left\{T_{m_n,t}\left(\omega, u/m_n^{1/2}\right)\right\}$	§3.2, Lemma 2
ψ^*_{n,q_n}	L_2-NED coefficients for $\left\{T_{m_n,t}\left(\omega, u/m_n^{1/2}\right)\right\}$	§3.2, Lemma 2
$c_{n,t}$	L_2-mixingale constants for $\left\{T_{m_n,t}\left(\omega, u/m_n^{1/2}\right)\right\}$	§3.2, Lemma 2
$\sigma^2_{m_n}$	$E\left(m_n^{1/2}\left(\hat{\alpha}^{-1}_{m_n} - \alpha^{-1}\right)\right)^2$	§3.2, Theorem 2
$w_{n,s,t}$	$w(\lvert s-t\rvert/\gamma_n)$	§4
$e^*_{n,t}, e_{n,t}(u)$	L_2-mixingale constants of $\left\{U_{m_n,t}\right\}$ and $\left\{I_{m_n,t}(u)\right\}$	Proof Lemma 1
$f^*_{n,t}, f^*_{n,t}(u)$	L_2-NED constants of $\left\{U_{m_n,t}\right\}$ and $\left\{I_{m_n,t}(u)\right\}$	Proof Lemma 2
$T_{m_n,t}$	$T_{m_n,t}\left(\omega, u/m_n^{1/2}\right)$	Proof Lemma 2
k_n, l_n, r_n	integer sequences for Bernstein blocks	Proof Lemma 3, (A.3)
$\left\{Z_{n,i}\right\}_{i=1}^{r_n}$	Bernstein blocks $\sum_{t=(i-1)k_n+l_n+1}^{ik_n} T_{m_n,t}\left(\omega, u/m_n^{1/2}\right)$	Proof Lemma 3, (A.4)
$\tilde{F}_{n,i}$	$\sigma(\{E_{n,\tau} : \tau \le ik_n\}), i = 1, ..., r_n$	Proof Lemma 3