5-27-2005

On Tail Index Estimation Using Dependent, Heterogeneous Data

Jonathan Hill
Department of Economics, Florida International University

On Tail Index Estimation Using Dependent, Heterogenous Data

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

May 27, 2005

Abstract

In this paper we analyze the asymptotic properties of the popularly used distribution tail estimator by B. Hill (1975), for heavy-tailed heterogenous, dependent processes. We prove the Hill estimator is weakly consistent for functionals of mixingales and L_1-approximable processes with regularly varying tails, covering ARMA, GARCH, and many IGARCH and FIGARCH processes. Moreover, for functionals of processes near-epoch-dependent on a mixing process, we prove a Gaussian distribution limit exists. In this case, as opposed to all existing prior results in the literature, we do not require the limiting variance of the Hill estimator to be bounded, and we develop a Newey-West kernel estimator of the variance. We expedite the theory by defining "extremal mixingale" and "extremal NED" properties to hold exclusively in the extreme distribution tails, disbanding with dependence restrictions in the non-extremal support, and prove a broad class of linear processes are extremal NED. We demonstrate that for greater degrees of serial dependence more tail information is required in order to ensure asymptotic normality, both in theory and practice.

1. Introduction Denote by $X = \{X_t\} = \{X_t : -\infty < t < \infty\}$ a stochastic process defined on the probability space $(\Omega, \Im, \mu_\Omega)$, $\Im = \cup_{t\in\mathbb{Z}}\Im_t$. The sigma algebra induced by X_t is strictly increasing: $\Im_{t-1} \subset \Im_t = \sigma(X_s : s \leq t)$. Let F denote the marginal distribution of X_t. Without loss of generality, assume F has support on $(0, \infty)$. We assume $\bar{F}(x) \equiv P(X_t > x)$ is regularly varying at ∞: there exists some $\alpha > 0$ such that for all $\lambda > 1$

$$\bar{F}(\lambda x)/\bar{F}(x) \to \lambda^{-\alpha} \tag{1}$$

*Dept. of Economics, Florida International University, Miami, FL; www.fiu.edu/~hilljona; jonathan.hill@fiu.edu.
JEL classifications: C12, C16, C52.
Keywords: Hill estimator; regular variation; infinite variance; near epoch dependence; mixingales.

as $x \to \infty$, where $\alpha > 0$ denotes the index of regular variation. Equivalently, X has common marginal tail distributions that satisfy

$$\bar{F}(x) = x^{-\alpha}L(x), \quad x > 0, \tag{1'}$$

as $|x| \to \infty$, where $L(x)$ is slowly varying[1].

We are interested in the asymptotic properties of a moment-based estimator of the tail index α under minimal restrictions on serial dependence and heterogeneity. From a theoretical point of view, the existence and type of central limit theorem hinges intimately on the property of regular variation in the distribution tails, hence the magnitude of α and the structure of the slowly varying component $L(x)$ are of paramount interest. See Ibragimov and Linnik (1971).

Interest in estimation of the tail index spans the statistics, economics, econometrics, finance and telecommunications literatures: tail thickness has been rigorously studied in the returns of financial and macroeconomic time series (e.g. Akgiray and Booth, 1988; Mittnick and Rachev, 1993; Cheng and Rachev, 1995; McCulloch 1996; Bidarkota and McCulloch, 1998, 2004), and estimators of α have been used recently as foundations of tests of covariance stationarity (Loretan, 1991; Loretan and Phillips, 1994), extremal structural change (Quintos *et al*, 2001) and extremal bivariate, GARCH, and serial dependence (Stărică, 1999; Quintos, 2004; Hill, 2005). See also Rachev (2003) and the extensive list of citations therein. Thus, it is of central importance that the asymptotic properties of estimators of α be established in the most general environment regarding serial dependence and heterogeneity.

Denote by $X_{(i)} > 0$ the i^{th} order statistic of X_t, $X_{(1)} \geq X_{(2)} \geq ...$ Consider a sequence of integers m such that $m \to \infty$ as $n \to \infty$, and $m = o(n)$, and define the function $b_n(m) > 0$ by the inverse probability

$$(n/m)P(X_t > b_n(m)) \to 1. \tag{2}$$

An intuitive approach to estimating the tail index α was conceived by B. Hill (1975). Denoting $z_+ \equiv \max\{z, 0\}$, using (1) and the definition of $b_n(m)$, for any integer $k > 0$ the k^{th} moment of the excess of $\ln X_t$ over $\ln b_n(m)$ is evaluated as (see also Hsing, 1991: eq. 1.5)

$$\begin{aligned} E\left(\ln X_t - \ln b_n(m)\right)_+^k &= \int_0^\infty P\left(\left(\ln X_t - \ln b_n(m)\right)^k > u\right) du \\ &= \bar{F}\left(b_n(m)\right) \int_0^\infty \bar{F}\left(b_n(m) e^{u^{1/k}}\right) / \bar{F}\left(b_n(m)\right) du \\ &\approx (m/n) \int_0^\infty e^{-\alpha u^{1/k}} du = (m/n) k! \alpha^{-k}, \end{aligned} \tag{3}$$

[1] Although we allow for otherwise heterogenous processes, we assume the tail index α is time-invariant. It seems that we may relax the time independence assumption on the tail of X through $L(x)$. In this case, we would assume $\bar{F}_t(\lambda x)/\bar{F}_s(x) \to \lambda^{-\alpha}$ for every $s, t \in \mathbb{Z}$. Subsequent notational changes, however, will only complicate the discourse.

as $n \to \infty$, hence $(n/m)E\left(\ln X_t - \ln b_n(m)\right)_+ \to \alpha^{-1}$ suggesting a method of moments estimator for α^{-1}:

$$\hat{\alpha}_m^{-1} \equiv \frac{1}{m}\sum_{i=1}^{m}\left(\ln X_{(i)} - \ln X_{(m+1)}\right) = \frac{1}{m}\sum_{t=1}^{n}\left(\ln X_t - \ln X_{(m+1)}\right)_+. \tag{4}$$

Known as the Hill estimator, $\hat{\alpha}_m^{-1}$ has been utilized pervasively in the applied finance, macroeconomics, physics, and telecommunications literatures.

Mason (1982) and Hall (1982) established asymptotic normality for the estimator $\hat{\alpha}_m^{-1}$ for *iid* processes; see also Davis and Resnick (1984), Hall and Welsh (1984), and de Haan and Resnick (1998). Quintos (2001,2004) allows for covariance stationary conditional heteroscedasticity with finite variance independent shocks. Stărică (1999) and Resnick and Stărică (1995, 1996, 1998) prove consistency for processes approximable by a q-dependent process, including infinite order moving averages, simple bilinear processes, and stochastic difference equations that include GARCH processes. A distribution limit, however, is not established.

For the remaining discourse, define $T_{n,t} \equiv (\ln X_t - \ln b_n(m))_+$ and $T^*_{n,t}(\epsilon,\rho) \equiv I(\ln X_t - \ln b_n(\rho m) > \epsilon)$ for any $\epsilon \in \mathbb{R}$ and ρ in a neighborhood of 1, where $I(\cdot)$ denote the indicator function. Similarly, define the processes

$$\{U_{n,t}, U^*_{n,t}\} \equiv \{T_{n,t} - E[T_{n,t}], \quad T^*_{n,t}(\epsilon,\rho) - E[T^*_{n,t}(\epsilon,\rho)]\}. \tag{5}$$

Hsing (1991), in a seminal paper, proves consistency and establishes a general distribution limit for $\hat{\alpha}_m^{-1}$ under suitable conditions: for consistency, $\{U_{n,t}, U^*_{n,t}\}$ must satisfy law of large numbers; and for a distribution limit, the slowly varying component $L(x)$ is restricted, and a central limit theorem must hold for $\{U_{n,t}, U^*_{n,t}\}$. The theory is exemplified by proving asymptotic normality for strong mixing processes provided summability conditions for moments of $\{U_{n,t}, U^*_{n,t}\}$ are satisfied[2]. The mixing assumption on X_t is useful because the functionals $\{U_{n,t}, U^*_{n,t}\}$ will necessarily have the same mixing property, and a law of large numbers and central limit theorem directly apply. Mixing properties, however, in general do not extend to many important time series processes, including infinite distributed lags, and therefore has limited appeal in the financial and macroeconomic literatures[3]. Moreover, it is not obvious upon first glance that the required law of large numbers or central limit theorem for $\{U_{n,t}, U^*_{n,t}\}$ will hold in arbitrary environments of dependence and heterogeneity, nor is it clear which processes satisfy the rather abstract restrictions on $L(x)$.

The purpose of this paper is to develop an asymptotic theory for the Hill estimator for dependent, heterogenous data with regularly varying tails (1). We prove consistency for mixingale and L_1-approximable processes $\{U_{n,t}, U^*_{n,t}\}$, covering mixing, NED, and L_0-approximable processes, which includes ARMA, GARCH, and many IGARCH and FIGARCH processes, cf. Davidson (2004):

[2] See, also, Rootzen *et al* (1990) for properties of the Hill estimator for mixing processes

[3] For example, a stationary AR(1) processes with *iid* shocks which have unbounded support are not strong mixing.

see Appendix 1 for dependence definitions. We then prove a Gaussian limit for NED processes $\{U_{n,t}, U^*_{n,t}\}$ of size $-1/2$ with a uniform or strong mixing base of respective size $-r/[2(r-1)]$, $r \geq 2$, and $-r/(r-2)$, $r > 2$, for a sub-class of (1) that includes the domain of attraction of the stable laws with $\alpha < 2$.

Specifically, we prove

$$\sqrt{m}(\hat{\alpha}_m^{-1} - \alpha^{-1})/\sigma_m \Longrightarrow N(0,1) \tag{6}$$

where $\sigma_m^2 = mE[(\hat{\alpha}_m^{-1} - \alpha^{-1})]^2 = O(n^\gamma)$, $\gamma \geq 0$, and therefore need not be finite asymptotically. A straightforward Newey-West kernel estimator suffices to generate a consistent estimate of σ_m^2, and we do not need to impose moment summability conditions on $\{U_{n,t}, U^*_{n,t}\}$. In every existing Gaussian limit proved in the literature (that we know of), by comparison, the limit $\lim \sigma_m^2$ is assumed to exist and therefore be finite (i.e. $\gamma = 0$).

Our initial dependence assumptions focus on the functionals $\{U_{n,t}, U^*_{n,t}\}$. In order to generalize the assumptions to X_t itself, we define "extremal" mixingale and NED properties to hold explicitly for X_t in the extreme right tail in terms of tail probabilities, allowing us to disband with unnecessary restrictions on the dependence structure of X_t in its non-extremal support. We prove that processes of the form $\sum_{i=0}^{\infty} \psi_i Z_{t-i}$ are extremal NED, where $\{Z_t\}$ is a process, not necessarily mean-zero, that satisfies (1) and $\sum_{i=0}^{\infty} |\psi_i|^\alpha < \infty$.

We demonstrate that for a given sample of size n more serial dependence implies more tail information is required for $\hat{\alpha}_m$ to approximately normally distributed, both in theory and practice. Dumouchel's (1983) famous rule of thumb for independent processes, in which $X_{(m+1)}$ denotes the sample upper 10th percentile of X_t, is shown to generate potentially vastly positively skewed estimates when the data are serially dependent. Similarly, popular plotting techniques like the "Hill plot" (e.g. Drees *et al*, 2000), which are used to pinpoint values $\hat{\alpha}_m$ are near, are shown to be essentially useless when the data are even mildly serially dependent.

To the best of our knowledge the results presented here are the most general available, even allowing for the fact that our distribution theory is restricted to a sub-class of (1). Although we prove a Gaussian limit for processes extremal NED on a mixing base, it would be straightforward to extend the result to mixingales (and "extremal mixingales") with few additional assumptions.

As a simple extension of the developed theory, we prove asymptotic normality for an estimator of the scale parameter c in tail forms $\bar{F}(x) = cx^{-\alpha}(1 + o(x^{-\theta}))$. An estimator of c has been used to derive covariance stationarity tests (Loretan, 1991; Loretan and Phillips, 1994) and tests of serial extremal dependence (Hill, 2005).

The remainder of the paper is organized as follows. In Section 2 we detail the major assumptions. Sections 3 and 4 contain the main results and extensions to new extremal dependence measures. Sections 5 and 6 contain an application of the theory to a scale estimator and a simulation study. Parting comments are left for Section 7. Appendix 1 contains notation conventions and dependence definitions; Appendix 2 contains tables; and Appendix 3 contains all formal proofs.

2. Assumptions The environment of this paper is detailed in the following three assumptions. Assumption A.1 defines a general regularly varying tail, while A.2 includes processes that belong to the domain of attraction of the stable laws when $\alpha < 2$ (see Feller, 1971; and Ibragimov and Linnik, 1971). Assumption B restricts the rate $m \to \infty$; and Assumption C defines mixingale and NED processes.

Assumption A

1. The distribution tails satisfy

$$\bar{F}(x) = x^{-\alpha} L(x), \quad x > 0, \tag{7}$$

for some $\alpha > 0$, as $|x| \to \infty$, where L is slowly varying.

2. The distribution tails satisfy

$$\bar{F}(x) = cx^{-\alpha}(1 + o(x^{-\theta})), \quad x > 0, \tag{8}$$

for some $c > 0$, $\alpha > 0$ and $\theta > 0$, as $|x| \to \infty$.

Assumption B

1. $m = o(n)$;
2. $m = [n^{\delta}]$, $0 < \delta < 2\theta/(2\theta + \alpha)$ where $\theta > 0$ is defined in Assumption A.2.

Assumption C Let $\mathcal{F}_t$ be an increasing sigma field on the measure space $(\Xi, \mathcal{F}, \mu_{\Xi})$ such that $\mathcal{F}_t = \sigma(\varepsilon_s < s \leq t)$ where $\{\varepsilon_t\}_{-\infty}^{\infty}$ is a stochastic process. Let $\epsilon \in \mathbb{R}$ and ρ in a neighborhood of 1 be arbitrary.

1. For each $U \in \{U_{n,t}, U_{n,t}^{*}(\epsilon, \rho)\}$[4], the sequence $\{U_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ is an L_p-mixingale, $p \geq 1$, with constants $d_t \in \{d_{n,t}, d_{n,t}^{*}\}$ and coefficients $\{\psi_q, \psi_q^{*}\}$ of size $-\lambda(1/p - 1/r)$ for some $\lambda > 0$, $r > p \geq 1$;
2. For each $U \in \{U_{n,t}, U_{n,t}^{*}(\epsilon/\sqrt{m}, 1)\}$, the sequence $\{U_t\}$ is L_r-bounded, $r \geq 1$, L_2-NED of size $-1/2$ on $\{\varepsilon_t\}_{-\infty}^{\infty}$, where ε_t is a uniform mixing process of size $-r/[2(r-1)]$, $r \geq 2$; or strong mixing of size $-r/(r-2)$, $r > 2$.

Remark 1: The L_r-boundedness component of Assumption C for each $\{U_{n,t}, U_{n,t}^{*}\}$ follows from Lemma 1, below, and therefore is a non-binding assumption.

3. Main Results We require the following lemma in order to make explicit mixingale and NED constants for $\{U_{n,t}, U_{n,t}^{*}(\epsilon, \rho)\}$.

[4] Throughout we write $\{U_{n,t}, U_{n,t}^{*}\}$ to denote the bivariate process. We write $U \in \{U_{n,t}, U_{n,t}^{*}\}$ when $U = \{U_t\}$ represents either process $\{U_{n,t}\}$ or $\{U_{n,t}^{*}\}$.

Lemma 1 *Consider a process X such that Assumption A.1 holds. Each $\{U_{n,t}\}$ and $\{U^*_{n,t}(\epsilon,\rho)\}$ is an L_r-bounded process for any $r \geq 1$. Specifically, for every $\epsilon \in \mathbb{R}$ and any ρ in a neighborhood of 1 there exists finite constant sequences $\{A_r, B_r, C_r\}_{r\geq 1} > 0$ such that for each $t \in \mathbb{N}$ as $n \to \infty$*

$$\left\|U^*_{n,t}(\epsilon,\rho)\right\|_r \leq A_r(m/n)^{1/r}; \tag{9}$$
$$\left\|U_{n,t}\right\|_r \leq B_r\,(m/n)^{1/r},\ r \in \mathbb{N};$$
$$\left\|U_{n,t}\right\|_r \leq C_r\,(m/n)^{1/[r+1]},\ r \in \mathbb{R}_+.$$

3.1. Weak Probability Convergence

The first result proves weak consistency of $\hat{\alpha}_m^{-1}$ for mixingales that have regularly varying tails (1) with $m = o(n)$.

Theorem 2 (Weak Limit for Mixingales) *Consider a process X such that Assumptions A.1, B.1 and C.1 hold for each $U \in \{U_{n,t}, U^*_{n,t}(\epsilon,\rho)\}$. Then*

$$m^{-1}\sum\nolimits_{t=1}^{n} U_t \to 0 \tag{10}$$
$$\hat{\alpha}_m^{-1} \to \alpha^{-1}.$$

Moreover,

$$\ln X_{([\rho m])} - \ln b_n(\rho m) \to 0. \tag{11}$$

for any $\epsilon \in \mathbb{R}$ and ρ in a neighborhood of 1.

Remark 1: Because strong and uniform mixing processes, and processes NED on a mixing base, are special cases of mixingales, the above theorem holds in those cases for $\{U_{n,t}, U^*_{n,t}\}$.

Remark 2: Result (11) will be useful for deriving a Gaussian limit for NED processes, below, and substantiates the use of $X_{(m+1)}$ in (4).

Remark 3: We generalize the dependence assumptions on $\{U_{n,t}, U^*_{n,t}\}$ to X_t itself in Section 3.3.

An NED process X_t is simply a process L_p-approximable by $E[X_t|\digamma_{t-q}^{t+q}]$. It is straightforward to extend Theorem 2 to other forms of approximability.

Theorem 3 (Weak Limit for L_1-approximable Processes) *Let Assumptions A.1 and B.1 hold. Assume each $U \in \{U_{n,t}, U^*_{n,t}(\epsilon,\rho)\}$ is L_1-approximable by $\digamma_{t-q}^{t+q}$-measurable sequences $h^{(q)} \in \{h^{(q)}_{n,t}, h^{*(q)}_{n,t}\}$ with approximating constants $e_t \in \{e_{n,t}, e^*_{n,t}\}$. Assume $\limsup_{n\to\infty}\sum_{t=1}^{n} e_t/m \leq D < \infty$ for some constant D, and assume each $(h^{(q)}_t - E[h^{(q)}_t])/m$ satisfies a weak law of large numbers. Then each conclusion of Theorem 2 holds.*

Remark 1: In order for each $(h^{(q)}_t - E[h^{(q)}_t])/m$ to satisfy a law of large numbers, we may assume each is a uniformly integrable L_1-mixingale with constants d_t/m, $\limsup_{n\to\infty}\sum_{t=1}^{n} d_t/m < \infty$, and $\lim_{n\to\infty}\sum_{t=1}^{n}(d_t/m)^2 = 0$. In this case, a mixingale law of large numbers due to Davidson (1993a) applies and

$\sum_{t=1}^{n}(h_t^{(q)} - E[h_t^{(q)}])/m \to 0$. Although this set of sufficient assumptions seems prohibitive, each necessarily holds for mixingale U_t/m (consult the line of proof of Theorem 2).

Remark 2: L_1-approximable processes are L_0-approximable, and L_0- approximability implies L_1-approximability provided the process is L_1-bounded (see Pötscher and Prucha, 1991; see also Davidson, 1994: p. 274 and Theorem 17.21). Because $\{U_{n,t}, U_{n,t}^*\}$ are bounded in the L_r-norm for any $r \geq 1$, cf. Lemma 1, we conclude Theorem 3 holds for L_0-approximable $\{U_{n,t}, U_{n,t}^*\}$.

3.2 Distribution Convergence

The previous results deal broadly with regularly varying tails by using only Assumption A.1. In order to prove $\hat{\alpha}_m^{-1}$ converges in law to a Gaussian random variable we must refine the structure of the slowly varying component $L(x)$. Hsing (1991) does this by restricting $L(x)$ to satisfy either of equations (2.3a) or (2.3b) of that work. However, it is not obvious which class of distributions satisfy either assumption. We consider the popularly studied class of distributions that satisfies Assumptions A.2, and effectively prove this class satisfies Hsing's (2.3a). In general, the following environment (Assumptions A.2 and B.2) is quite similar to the distribution tail setting studied in Hall (1982)[5].

Let

$$\sigma_m^2 = m \times E(\hat{\alpha}_m^{-1} - \alpha^{-1})^2. \tag{12}$$

In the following, we apply a powerful central limit theorem due to de Jong (1997): see also de Jong and Davidson (2000)[6].

Theorem 4 (Gaussian Limit for NED Processes) *Let Assumptions A.2, B.2, and C.2 hold. Let $\sigma_m^2 = O(n^\gamma)$, $\gamma \geq 0$[7]. Then*

$$\sigma_m^{-1}\sqrt{m}\left(\hat{\alpha}_m^{-1} - \alpha^{-1}\right) \Rightarrow N(0,1), \tag{13}$$

if $r > 2$, $\delta \geq 1 - r\gamma/(r-2)$ and $\gamma > 0$; or $r = 2$ and $\gamma \geq 0$.

If additionally Assumption C.2 holds with $r = 2$ (i.e. the NED-mixing base size is -1) and $\gamma = 0$, then

$$\sqrt{m}\left(\hat{\alpha}_m^{-1} - \alpha^{-1}\right) \Rightarrow N(0,\sigma^2), \tag{14}$$

where $\sigma^2 = \lim_{n\to\infty}\sigma_m^2$. If X_t is iid, then $\sigma^2 = \alpha^{-2}$.

---

[5] Distributions which satisfy Assumption A.2 belong to the domain of attraction of a stable law when $\alpha < 2$ (see Ibragimov and Linnik, 1971), and have been utilized in the development of asymptotic theory for least squares estimators (Cline, 1983), t-ratios, the Durbin-Watson test (Loretan and Phillips, 1991), the Box-Pierce test (Runde, 1997), tests of covariance stationarity (Loretan, 1991; Loretan and Phillips, 1994), cointegration (Caner, 1998), unit roots (Chan and Tan, 1989; Phillips, 1990), extremal structural change (Quintos *et al*, 2001), extremal dependence (Hill, 2005), etc.

[6] Other theorems due to Wooldridge and White (1988), Davidson (1992,1993b), Chen and White (1997), and Lin and Qiu (2004) are also applicable, but do not improve the major outcome due to size restrictions, or restrictions on the implied convergence property of σ_m^2.

[7] By convention the rate $\sigma_m^2 = O(n^\gamma)$ implies $\sigma_m^2 \to \infty$ if $\gamma > 0$, and $\sigma_m^2 \to \sigma^2 \in (0,\infty)$ if $\gamma = 0$. We do not consider the degenerate case, $\gamma < 0$, such that $\sigma_m^2 \to 0$.

Remark 1: For (13), when $r = 2$ no restrictions on δ are required other than Assumption B.2. However, the NED-mixing base ϵ_t must be uniform mixing in order to allow $r = 2$. Moreover, if $\gamma = 0$ such that σ_m^2 converges, we require $r = 2$, limiting the NED-uniform mixing size to -1.

Remark 2: Limit (14) is a generalized version of Hsing's (1991) Theorem 3.3 for mixing processes, in which moments of the processes $\{U_{n,t}, U_{n,t}^*\}$ are assumed to satisfy summability conditions. The assumption $\sigma_m^2 = O(1)$ in Theorem 4 replaces Hsing's (1991) summability conditions.

Remark 3: If X_t is *iid*, then $\sigma^2 = \alpha^{-2}$, and (14) is simply the limit derived in Hall (1982) for a sub-class of distributions that satisfy Assumption A.1: $\bar{F}(x) = cx^{-\alpha}(1 + Bx^{-\theta} + o(x^{-\theta}))$. If our case, $B = 0$.

Remark 4: Using Cramér's Theorem it is straightforward to show (13) implies

$$\sqrt{m}\tilde{\sigma}_m^{-1}(\hat{\alpha}_m - \alpha) \Rightarrow N(0, \alpha^2), \tag{15}$$

where $\tilde{\sigma}_m = \alpha\sigma_m$. If the process X_t is *iid*, then $\lim_{n\to\infty} \tilde{\sigma}_m^2 = \alpha^2 \lim_{n\to\infty} \sigma_m^2 = \alpha^2\alpha^{-2} = 1$.

As r increases such that the allowable degree of dependence through the NED-mixing base size increases, the lower bound on δ, $1 - r\gamma/(r-2)$, increases, hence more tail information (large $m \approx n^\delta$) in general is required for each sample of size n. Strictly speaking, when X_t is "approximable" by a relatively "more serially dependent" process ϵ_t, more tail information is required.

Intuitively, if X_t is strongly positively serially dependent, then clusters of X_t may not reveal sufficient information concerning volatility in order to obtain a sharp estimate of the tail shape, which itself characterizes dispersion. If dependence is restricted such that the mixing base is uniform with size -1 (i.e. $r = 2$), then no restrictions on δ are required other than Assumption B.2[8].

3.3 Variance Estimator

The tail estimator $\hat{\alpha}_m^{-1}$ is simply the first sample moment of the serially dependent $(\ln X_t - \ln X_{(m+1)})_+$, which is simply an estimator of the L_2-process $(\ln X_t - \ln b_n(m))_+$, where L_2-boundedness follows from Lemma 1. Hence a Newey-West kernel estimator of the variance will suffice in practice. For background theory, see Newey and West (1987) and Gallant and White (1988). We define

$$\hat{\sigma}_m^2 = \frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w((s-t)/l_n)\hat{Z}_s\hat{Z}_t \tag{16}$$

where $\hat{Z}_t = [\left(\ln X_t/X_{(m+1)}\right)_+ - (m/n)\hat{\alpha}_m^{-1}]$, and $w((s-t)/l_n)$ denotes a standard kernel function with bandwidth l_n, an increasing sequence of positive integers $1 \leq l_n < m$, $l_n \to \infty$ as $m \to \infty$ (see, e.g., de Jong and Davidson, 2000).

[8] An important caveat is the reliance of the distribution limit (13), and the subsequent restriction (if any) on δ, on de Jong's (1997) central limit theorem. In order for de Jong's result to apply here, we must bound δ when the mixing base size is such that $r > 2$. Whether such a restriction on δ is required under *any* central limit theorem is debatable, and beyond the scope of the present paper, although de Jong's central limit theorem is evidently the most powerful and inclusive available.

Lemma 5 *Let the kernel $w_{s,t,n} \equiv w((s-t)/l_n)$ satisfy Assumption 1 of de Jong and Davidson (2000) with $l_n = O(n^{\varsigma})$, $\varsigma \in [0,1]$, and $m^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n}|w_{n,s,t}| = O(1)$. Under the conditions of Theorem 4, $\hat{\sigma}_m^2/\sigma_m^2 \to 1$.*

Remark 1: If $r = 2$ and $\gamma = 0$ such that the NED base ϵ_t is uniform mixing and $\sigma_m^2 \to \sigma^2$, then $\hat{\sigma}_m^2/\sigma^2 \to 1$.

4 Extremal Dependence It is straightforward to demonstrate that if $\ln X_t$ is L_2-NED, then $U_{n,t}$ is an L_2-NED processes. However, establishing an NED property for $U_{n,t}$ based on X_t itself, or demonstrating $U_{n,t}^*$ is NED based on either $\ln X_t$ or X_t, is less transparent.

For brevity, denote by $x_n(u)$ the term $b_n(m)e^u$ for any $u \in \mathbb{R}$. We circumvent the problem by creating dependence properties based on the minimal requirements of this paper. The problem with the process $U_{n,t}^* = T_{n,t}^* - E[T_{n,t}^*]$ lies in the fact that it is binary: in order for $U_{n,t}^*$ to be L_2-NED we require

$$\left\| U_{n,t}^* - E[U_{n,t}^* | \digamma_{t-q}^{t+q}] \right\|_2 \leq d_{n,t}^* \psi_q^*, \tag{17}$$

which, using (5), reduces to[9]

$$\begin{aligned} & \left\| I\left(X_t > x_n(u)\right) - E[I\left(X_t > x_n(u)\right) | \digamma_{t-q}^{t+q}] \right\|_2 \qquad (18) \\ = \quad & \left\| P(X_t > x_n(u) | \Im_{t-q}^{t+q}) - P\left(X_t > x_n(u) | \digamma_{t-q}^{t+q}\right) \right\|_2 \leq d_{n,t}^* \psi_q^*. \end{aligned}$$

Moreover, consistency and the distribution limit established in Sections 3.1-3.2 are, of course, limiting properties, and $b_n e^{\epsilon/\sqrt{m}} \to \infty$ as $n \to \infty$ for any $\epsilon \in \mathbb{R}$ by (2). Thus, inequality (17) need only hold in the extreme right tail of X_t: apparently the field $\digamma_{t-q}^{t+q}$ need only contain extreme tail information concerning X_t.

Assumption C.3 defines the sequence $\{X_t, \digamma_t\}$ as an L_p-Extremal-Mixingale (L_p-E-MIX), and Assumption C.4 defines the process $\{X_t\}$ as L_p-Extremal-Near-Epoch-Dependent (L_p-E-NED) on $\{\digamma_t\}$. Essentially we are defining $U_{n,t}^*$ to be, respectively, an L_1-mixingale or L_2-NED on a mixing base as $n \to \infty$, and we prove in Theorem 7, below, that the dependence property of $U_{n,t}^*$ carries over to $U_{n,t}$. Let $x_n(u) \equiv b_n e^u$, and define the field $\digamma_a^b = \sigma(\epsilon_t : a \leq t \leq b)$.

Assumption C.3 (L_p-E-MIX) Let ϵ_t satisfy the mixing properties of Assumption C.1. For each t there exists a Lebesgue measurable function $\tilde{e}_t^* : \mathbb{R} \to \mathbb{R}_+$, integrable on $\mathbb{R}_+$, and a sequence of constants $\{\tilde{v}_q^*\}_{q=0}^{\infty}$, with $\tilde{v}_q^* = O(q^{-\lambda})$, $\lambda > 0$, such that for any $u \in \mathbb{R}$ and some $p \geq 1$

$$\begin{aligned} & \limsup_{n\to\infty} \left(\frac{n}{m}\right)^{1/p} \left\| P\left(X_t > x_n(u)\right) - P\left(X_t > x_n(u) | \digamma_{t-q}\right) \right\|_p \qquad (19) \\ & \qquad \leq e_t^*(u)\tilde{v}_q^* \\ & \limsup_{n\to\infty} \left(\frac{n}{m}\right)^{1/p} \left\| P(X_t > x_n(u) | \Im_{t-q}^{t+q}) - P\left(X_t > x_n(u) | \digamma_{t+q}\right) \right\|_p \\ & \qquad \leq e_t^*(u)\tilde{v}_{q+1}^*. \end{aligned}$$

[9]Observe that $P(X_t > x_n(u) | \Im_{t-q}^{t+q}) = I(X_t > x_n(u))$ because X_t is $\Im_t^t$-measurable.

Assumption C.4 (L_p-E-NED) Let ϵ_t satisfy the mixing properties of Assumption C.2. For each t there exists a Lebesgue measurable function $\tilde{e}_t^*$: $\mathbb{R} \to \mathbb{R}_+$, integrable on $\mathbb{R}_+$, and a sequence of constants $\{\tilde{\upsilon}_q^*\}_{q=0}^{\infty}$, with $\tilde{\upsilon}_q^* = O(q^{-\lambda})$, $\lambda > 0$, such that for any $u \in \mathbb{R}$ and some $p \geq 1$

$$\limsup_{n\to\infty} \left(\frac{n}{m}\right)^{1/p} \left\| P(X_t > x_n(u)|\Im_{t-q}^{t+q}) - P\left(X_t > x_n(u)|\digamma_{t-q}^{t+q}\right)\right\|_p \leq \tilde{e}_t^*(u)\tilde{\upsilon}_q^*. \tag{20}$$

As in the standard case, an extremal NED process is also an extremal mixingale.

Lemma 6 *Let X_t be L_p-E-NED, $p \geq 1$, on some $\{\digamma_t\}$, $\digamma_t = \sigma(\varepsilon_s : s \leq t)$, with constants $\tilde{e}_t^*$ and coefficients $\tilde{\upsilon}_q^*$ of size $-\lambda_1$, where ε_t is strong or uniform mixing of size $-\lambda_0$. Then $\{X_t, \digamma_t\}$ is an L_p-E-MIX sequence with coefficients of size $-\min\{\lambda_0^{1-1/p}, \lambda_1\}$.*

Theorem 7 *Let Assumption A.1 hold.*
i. Provided Assumption C.3 holds with E-MIX size $-\lambda$, then each $\{U_t, \digamma_t\}$ is an L_1-mixingale as $n \to \infty$, with constants

$$d_{n,t} = (m/n)\int_0^{\infty} e_t^*(u)du, \quad d_{n,t}^* = (m/n)\tilde{e}_t^*(\epsilon/\sqrt{m}), \tag{21}$$

and coefficients $\{\psi_q, \psi_q^*\}$ of size $-\lambda$.
ii. Provided Assumption C.4 holds with E-NED size $-\lambda$, then each $\{U_t\}$ is L_2-NED on $\{\digamma_t\}$ as $n \to \infty$, with constants

$$d_{n,t} = (m/n)^{1/2}\int_0^{\infty} e_t^*(u)du, \quad d_{n,t}^* = (m/n)^{1/2}\tilde{e}_t^*(\epsilon/\sqrt{m}), \tag{22}$$

and coefficients $\{\psi_q, \psi_q^*\}$ of size $-\lambda$.

Remark 1: If $\{X_t, \digamma_t\}$ is an L_1-E-MIX sequence, then each $\{U_t, \digamma_t\}$ is an L_1-mixingale as $n \to \infty$, and under Assumptions A.1 and B.1, Theorem 2 holds: $\hat{\alpha}_m^{-1} \to \alpha^{-1}$. Similarly, if $\{X_t\}$ is L_2-E-NED on $\{\digamma_t\}$ with size $-1/2$, where $\digamma_t = \sigma(\varepsilon_s : s \leq t)$ for a uniform or strong mixing process $\{\varepsilon_t\}$ of appropriate size, then each $\{U_t\}$ is L_2-NED on $\{\digamma_t\}$ with size $-1/2$, and under Assumptions A.2 and B.2, Theorem 4 holds: $\sigma_m^{-1}\sqrt{m}(\hat{\alpha}_m^{-1} - \alpha^{-1}) \Rightarrow N(0,1)$.

4.1 Linear Processes

Consider the process

$$X_t = \sum\nolimits_{i=0}^{\infty} \psi_i Z_{t-i},\ \psi_0 = 1,\ \sum\nolimits_{i=0}^{\infty} |\psi_i|^{\alpha} < \infty$$

where the $Z_t^{\prime}s$ are *iid* random variables satisfying (1). Notice Z_t does not necessarily have a mean zero. By the following lemma, such processes $\{X_t\}$ are L_2-E-NED on $\{Z_t\}$.

Lemma 8 *$\{X_t\}$ is L_2-E-NED on $\{Z_t\}$ with coefficients $\tilde{v}_q^* = O([\sum_{i=q+1}^{\infty} |\psi_i|^{\alpha}]^{1/2})$ and constants $\tilde{e}_t^*(u) : \mathbb{R} \to \mathbb{R}_+$ integrable on $\mathbb{R}_+$. If $\psi_i = O(i^{-\mu})$, $\mu > 1/\alpha$, then the E-NED size is $-(\alpha\mu - 1)/2$.*

Remark 1: Notice that the E-NED size magnitude satisfies $(\alpha\mu - 1)/2 \geq 1/2$ *if and only if* $\mu \geq 2/\alpha$, in which case Theorem 4 holds. For processes with relatively thick tails (small α), μ must be relatively large such that ψ_i decays faster, effectively restricting the strength of allowed serial dependence. This result nicely compliments Theorem 4. Together, relatively more serially dependent processes, as measured by the a smaller mixing base size, require more tail information per sample for tail estimation; and relatively more serially dependent *linear* processes, as measured by a smaller decay rate μ, must have thinner distribution tails.

5. Application For a slightly generalized version of the tail structure in Assumption A.2, Hall (1982) considered the scale estimator

$$\hat{c}_m \equiv (m/n) X_{(m+1)}^{\hat{\alpha}_m}. \tag{23}$$

Hall (1982) analyzed $\hat{c}_m$ in an *iid* environment; and $\hat{c}_m$ has been employed for tests of covariance stationarity (Loretan, 1991; Loretan and Phillips, 1993); and tests of extremal white noise (Hill, 2005).

Theorem 9 *Consider a process X such that Assumptions A.2, B.2, and C.4 hold. Let $\sigma_m^2 = O(n^{\gamma})$, $\gamma \geq 0$. Let $X_{(m+1)}/b_n(m) = 1 + o_p(n^{-\xi})$, $\xi > (1 - \delta)\theta/\alpha$. Then*

$$\tilde{\sigma}_m^{-1} \alpha^{-1} \sqrt{m} \left[\ln(n/m)\right]^{-1} (\hat{c}_m - c) \Longrightarrow N(0, c^2), \tag{24}$$

if $r > 2$, $\delta \geq 1 - r\gamma/(r-2)$ and $\gamma > 0$; or $r = 2$ and $\gamma \geq 0$. If Assumption C.4 holds with $r = 2$ (i.e. the E-NED-mixing base size is -1) and $\gamma = 0$,

$$\sqrt{m} \left[\ln(n/m)\right]^{-1} (\hat{c}_m - c) \Longrightarrow N(0, c^2 \tilde{\sigma}^2), \tag{25}$$

where $\tilde{\sigma}^2 = \lim_{n\to\infty} \tilde{\sigma}_m^2$.

6. Simulation We now simulate linear processes X_t and analyze the performance and distribution of $\hat{\alpha}_m$. For brevity, we fix the sample size to $n = 500$. We draw random samples of *iid* mean-zero time series Z_t from a symmetric stable distribution[10], or a symmetric Pareto distribution with probability density function

$$f(z) = \alpha |z|^{-\alpha-1} \text{ if } |z| \geq \xi; \quad f(z) = \alpha \xi^{-\alpha-1} \text{ if } |z| \leq \xi, \tag{26}$$

for $\alpha = 1.7$. A tail index value near 2 will help characterize the propensity of inference on $\hat{\alpha}_m$ to lead to the false conclusion that $\alpha \geq 2$. Simulation results

[10] We use McCulloch's (1997) version of the simulation algorithm of Chambers *et al* (1976).

for other values of α are qualitatively similar. The constant $\xi > 1$ is chosen such that $f(z)$ is a proper pdf: it is straightforward to show $\xi = [2(1+\alpha)]^{1/\alpha}$ gives $\int_{-\infty}^{+\infty} f(z)dz = 1$..

We consider a family of AR(1) processes

$$X_t = \phi X_{t-1} + \epsilon_t, \quad \epsilon_t = \sum\nolimits_{i=0}^{\infty} \pi_i Z_{t-i}, \tag{27}$$

where $|\phi| < 1$, $\pi_0 = 1$ and $\pi(z) = \sum_{i=0}^{\infty} \pi_i z^i \neq 0 \ \forall z \in \mathbb{C}$, $|z| \leq 1$. In the benchmark *iid* case, $\phi = \pi_i = 0$, $i \geq 1$, and $X_t = Z_t$.

In the simple AR(1) case we use $\phi \in \{.2, .4, .6, .8, .9\}$, and set $\pi_i = 0$ for all $i \geq 1$. Using Lemma 8, it is straightforward to show $\{X_t\}$ is L_2-E-NED on $\{Z_t\}$ with coefficients $\tilde{v}_q^* = O(|\phi|^{q\alpha/2}) \leq O(q^{-\lambda})$, for some $\lambda > 0$, hence the Hill estimator is consistent and asymptotically normal, cf. Theorem 4.

We then fix $\phi = .9$, and simulate general AR(1) processes with non-*iid* shocks ϵ_t by setting $\pi_0 = 1$ and $\pi_i = i^{-\mu}$, $i \geq 1$, for $\mu > (\alpha+2)/\alpha \geq 1$. There are two ways to interpret the resulting process (27). First, it is laborious but straightforward to prove ϵ_t is strong mixing of size $-\lambda$, where

$$\lambda = [\alpha(\mu - 1) - 2]/(\alpha + 1), \tag{28}$$

under appropriate restrictions on the probability density function of Z_t, which are satisfied in the present setting: see Theorem 14.9 of Davidson (1994), cf. Chandra (1974). Clearly $\partial\lambda/\partial\mu > 0$ and $\partial\lambda/\partial\alpha > 0$. We consider $\mu = c(\alpha + 2)/\alpha$ for constants $c \in \{1.1, 1.8, 10\}$, giving mixing sizes $\lambda \in \{.137, 1.096, 12.33\}$ when $\alpha = 1.7$.

Alternatively, of course X_t obtains the linear representation

$$X_t = \sum\nolimits_{i=0}^{\infty} \psi_i Z_{t-i}, \ \psi_0 = 1, \ \psi_i = \phi^i + \sum\nolimits_{j=1}^{i} j^{-\mu}\phi^{i-j}, \ i \geq 1. \tag{29}$$

Each term ϕ^i and $j^{-\mu}\phi^{i-j}$ is $O(i^{-\mu})$[11], hence $\sum_{j=1}^{i} j^{-\mu}\phi^{i-j}$ is $O(i^{-\mu+1})$. It follows that $|\psi_i|^\alpha$ is $O(i^{-(\mu-1)\alpha})$, where $(\mu-1)\alpha > 1$ given $\mu > (\alpha+2)/\alpha > (\alpha+1)/\alpha$ by assumption, hence $\sum_{i=0}^{\infty} |\psi_i|^\alpha < \infty$. Because $\mu > (\alpha+2)/\alpha > 2/\alpha$ is satisfied, $\{X_t\}$ is L_2-E-NED on $\{Z_t\}$ with size $-1/2$, cf. Lemma 8.

We approximate the infinite distributed lag $\sum_{i=0}^{\infty} \pi_i Z_{t-i}$ by simulating the truncation $\sum_{i=0}^{1000} \pi_i Z_{t-i}$. For all AR(1) time series we simulate $3 \times n$ observations and retain the last n. We simulate 1000 series of each process, and we derive the Hill estimator for the absolute series $|X_t|$.

[11] In order to see that $j^{-\mu}\phi^{i-j}$ is $O(i^{-\mu})$, recall $0 < \phi < 1$ and $\mu > 1$, and notice $j^{-\mu}\phi^{i-j}$ is strictly convex in j:

$$(\partial/\partial j)\ln j^{-\mu}\phi^{i-j} = -\mu/j + \ln 1/\phi, \ \ (\partial/\partial j)^2 \ln j^{-\mu}\phi^{i-j} = \mu/j^2 > 0.$$

For small j, $(\partial/\partial j)\ln j^{-\mu}\phi^{i-j} < 0$ for some values of μ and ϕ, and for large j, $(\partial/\partial j)\ln j^{-\mu}\phi^{i-j} > 0$. Thus, because $1 \leq j \leq i$ we deduce for large i

$$j^{-\mu}\phi^{i-j} \leq \max\left\{\phi^{i-1}, i^{-\mu}\right\} = i^{-\mu}$$

and therefore $j^{-\mu}\phi^{i-j} = O(i^{-\mu})$.

We plot the Hill estimator $\hat{\alpha}_m$ for simple AR(1) processes driven by Paretian or stable innovations over ϕ, and for general AR(1) processes with $\phi = .9$ over mixing sizes λ: sees Figure 1 and 2. The comparative shape of the Hill estimator as the tail fractile m increases is well known for independent Paretian and stable processes: see, for example, Drees *et al* (2000). However, there does not appear to be much discussion in this literature of the impact serial dependence has on the usefulness of such plots (e.g. Resnick and Stărică, 1997). Indeed, the entire premise of the ideas presented in Resnick and Stărică (1997) and Drees *et al* (2000) is that an "alternative" Hill plot, a logarithmic plot of $\hat{\alpha}_{[n^\delta]}$ against $\delta \in [0,1]$, can help stretch the plot "giving more display space to smaller values of [the fractile index m]" (*ibid*, p. 254) and make more apparent the display space neighborhood where values of $\hat{\alpha}_m$ predominantly occur (in theory, near α).

Their premise, however, is fundamentally predicated on the assumption the process X_t is *iid*. This is well exemplified for *iid* Paretian random variables, the distribution class Drees *et al* (2000) suggest is particularly suited to an application of the Hill plot: see the bold line in Figure 1 which predominantly hovers near $\alpha = 1.7$. It is profoundly clear, however, that a plot of $\hat{\alpha}_m$ provides little, or possibly no, useful information regarding α when a Paretian-noise driven process is a simple AR(1) with $\phi \geq .8$, or general AR(1) with $\phi = .9$ and non-*iid* innovations. In either case the tail fractile m must be near $.4 \times n = 200$ for $\hat{\alpha}_m$ to be "near" α, a value substantially larger than Dumouchel's (1983) suggested 10th percentile $.1 \times n = 50$. In general, as the mixing base size decreases, implying a greater degree of serial dependence, the amount of tail information required is comparatively larger.

For stable processes, the need for tail observations is even more attenuated: as the degree of dependence increases, the general plot shape bows outward. However, for each *iid*, simple AR and general AR stable process simulated, we find that a large tail fractile m is required (near $.4 \times n$) in order for $\hat{\alpha}_m$ to be "near" α, although a slight increase in m is detectably necessary as the level of dependence increases.

As a criterion for how "near" $\hat{\alpha}_m$ is to α, for independent, simple and general stable AR(1) processes we estimate asymptotic 95% intervals using a Newey-West estimator with Bartlett kernel and bandwidth $l_n = [\sqrt{m}]$. We estimate $\hat{\alpha}_m$ for all $m = 2...n-1$, and note the minimum m at which 2 is not in the interval: this gives the minimum required number of tail observations in order to deduce correctly that the variance is infinite, at the 5%-level. Similarly, we note the largest m such that $\alpha = 1.7$ is in the 95%-bounds. See Table 1. Once again, there is a clean monotonic increase in the required number of tail observations m as the degree of serial dependence increases.

Despite the overwhelming evidence that more observations from the distribution tail are required for a sharp estimate of $\hat{\alpha}_m$, the property of consistency only requires $m = o(n)$, cf. Theorem 2. The obvious question, then, is whether asymptotic normality is influenced by the magnitude of m, as Theorem 4 predicts. We provide kernel density plots based on a Gaussian kernel (see Silverman, 1986) of a sequence of simulated $\hat{\alpha}_m' s$ for $n = 500$ symmetric stable random variables with $\alpha = 1.7$, using only $m = 5$, 10, 100, and 200. We then perform Cramér-von

Mises tests of normality on the estimated density. See Figures 3 and 4 for plots and test results. We only fail to reject normality when m is large: for $m = 5$ (not shown) or $m = 10$ we massively reject the normality hypothesis due to the extreme right skewedness and kurtosis of the sequence of $\hat{\alpha}_m$; when $m \geq 180$ (not shown) in general, and $m = 200$ in particular, the density estimates are reasonably similar to a Gaussian distribution with mean $\alpha = 1.7$, in particular when the mean, α, is estimated rather than imputed.

7. Conclusion For processes with regularly varying tails we prove weak convergence of the Hill estimator for the class of mixingale $\{U_{n,t}, U^*_{n,t}\}$, covering mixing, NED-mixing, and L_1- and L_0-approximable processes. Moreover, for a sub-class of distributions that includes the domain of attraction of the stable laws, we prove a Gaussian limit for processes $\{U_{n,t}, U^*_{n,t}\}$ near-epoch-dependent on a uniform or strong mixing process. We do not require the asymptotic variance of $\hat{\alpha}_m$ to be finite, and we establish a consistent Newey-West type kernel estimator of the variance. In theory, as the degree of permitted serial dependence increases, the minimum number of tail observations per sample used for estimation must increase.

In order to expand our dependence assumptions to X_t itself, we define "extremal mixingale" and "extremal NED" properties, disbanding with dependence assumptions on the non-extremal support. A broad class of linear processes satisfy the L_2-extremal-NED property, and therefore permit asymptotically normal estimation of the Hill estimator. A simulation study demonstrates the unavoidable need for more tail observations in order to estimate α when serial dependence is present. Even for mildly dependent processes the so-called "Hill plot" is essentially useless and data dependent methods for selection of the sample fractile m is undoubtedly required (e.g. the bootstrap method of Danielsson *et al,* 1998; or the co-relation ranking strategy of Hill, 2005).

Appendix 1: Notation and Dependence

Although m depends on n, we omit such notation. Denote by $\rightarrow$ convergence in probability, and by $\Longrightarrow$ weak convergence with respect to finite dimensional distributions. $[x]$ denotes the integer part of x, with $[|x|] \leq |x|$. $I(A)$ denotes the indicator function: $I(A) = 1$ if A is true.

Mixing[12]: Let $W_{n,t}$ denote an $\Im$-measurable function of X_t. Denote by $\Im_a^b$ the σ-field $\sigma(W_{n,t}$: $a \leq t \leq b)$ and let $\Im_t$ denote $\sigma(W_{n,s} : s \leq t)$. Define respectively uniform and strong mixing coefficients $q \geq 1$

$$\phi(\Im_{-\infty}^t, \Im_{t+q}^\infty) \equiv \sup_{A \in \Im_{-\infty}^t, B \in \Im_{t+q}^\infty} \{|P(B|A) - P(B)|$$
$$\vartheta(\Im_{-\infty}^t, \Im_{t+q}^\infty) \equiv \sup_{A \in \Im_{-\infty}^t, B \in \Im_{t+q}^\infty} \{|P(A \cap B) - P(A)P(B)|$$

We say $W_{n,t}$ is uniform mixing if $\phi_q = \sup_t \phi(\Im_{-\infty}^t, \Im_{t+q}^\infty) \rightarrow 0$ as $q \rightarrow \infty$. Similarly, X_t is strong mixing if $\vartheta_q = \sup_t \vartheta(\Im_{-\infty}^t, \Im_{t+q}^\infty) \rightarrow 0$.

Mixingale: Denote by $\digamma_t$ an arbitrary sigma-field on $(\Xi, \digamma, \mu_\Xi)$. We say $\{W_{n,t}, \digamma_{-\infty}^t\}_{-\infty}^\infty$ is an L_p-mixingale, $p \geq 1$, if for some sequence of positive constants $\{d_{n,t}\}_{-\infty}^\infty$ and $\{\psi_q\}_0^\infty$, with $\psi_q \rightarrow 0$ as $q \rightarrow \infty$,

$$\|E[W_{n,t}|\digamma_{t-q}]\|_p \leq d_{n,t}\psi_q$$
$$\|W_{n,t} - E[W_{n,t}|\digamma_{t+q}]\|_p \leq d_{n,t}\psi_{q+1}.$$

Near-Epoch-Dependence: Let $\{\varepsilon_t\}_{-\infty}^\infty$ be a processes measurable with respect to $\digamma$: specifically, $\digamma_a^b = \sigma(\varepsilon_t : a \leq t \leq b)$. We say $\{W_{n,t}\}_{-\infty}^\infty$ is L_p-near-epoch-dependent on $\{\varepsilon_t\}_{-\infty}^\infty$ (synonymously on $\{\digamma_{-\infty}^t\}_{-\infty}^\infty$) if there exist positive constants $\{d_{n,t}\}_{-\infty}^\infty$ and $\{\psi_q\}_0^\infty$, with $\psi_q \rightarrow 0$ as $q \rightarrow \infty$, such that

$$\left\|W_{n,t} - E[W_{n,t}|\digamma_{t-q}^{t+q}]\right\|_p \leq d_{n,t}\psi_q.$$

Approximability: We say $W_{n,t}$ is L_0-approximable if for each $q \in \mathbb{N}$ there exists an $\digamma_{t-q}^{t+q}$-measurable function $h_{n,t}^{(q)}$ such that for positive constants $\{e_{n,t}\}_{-\infty}^\infty$ and coefficients $\{\upsilon_q\}_{q=0}^\infty$, with $\upsilon_q \rightarrow 0$ as $q \rightarrow \infty$, and for every $\gamma > 0$

$$P\left(\left|W_{n,t} - h_{n,t}^{(q)}\right| > e_{n,t}\gamma\right) \leq \upsilon_q.$$

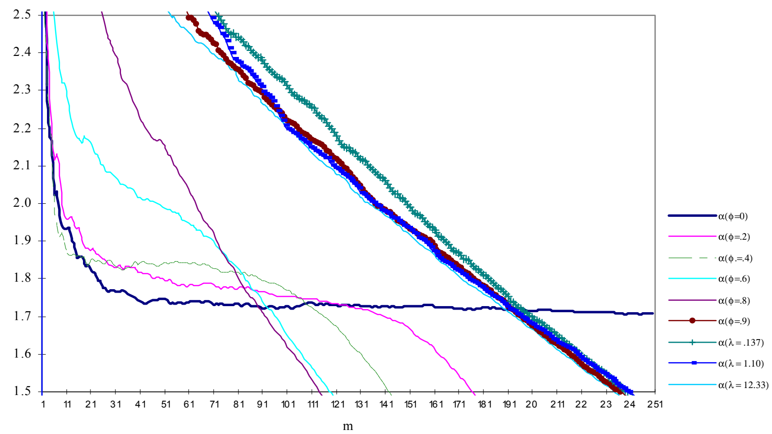
We say $W_{n,t}$ is L_p-approximable, $p > 0$, if for each $q \in \mathbb{N}$ there exists an $\digamma_{t-q}^{t+q}$-measurable function $h_{n,t}^{(q)}$ such that

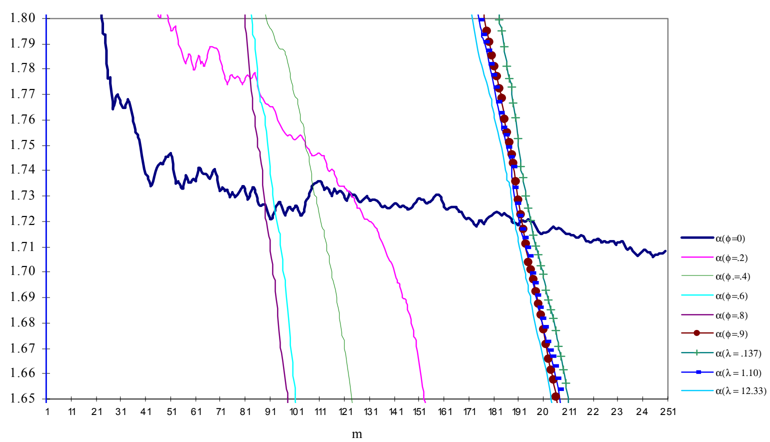
$$\left\|W_{n,t} - h_{n,t}^{(q)}\right\|_p \leq e_{n,t}\upsilon_q.$$

[12] Consult Ibragimov (1962), McLiesh (1975), Hall and Hyde (1980), Gallant and White (1988) and Davidson (1994) for theoretical and historical details on mixing, mixingale and NED usages and properties. Consult Pötscher and Prucha (1991) and Davidson (1994) for L_0- and L_p-approximability.

Appendix 2: Figures and Tables

Figure 1
Hill Plot, Pareto, $\alpha = 1.7$, $n = 500$

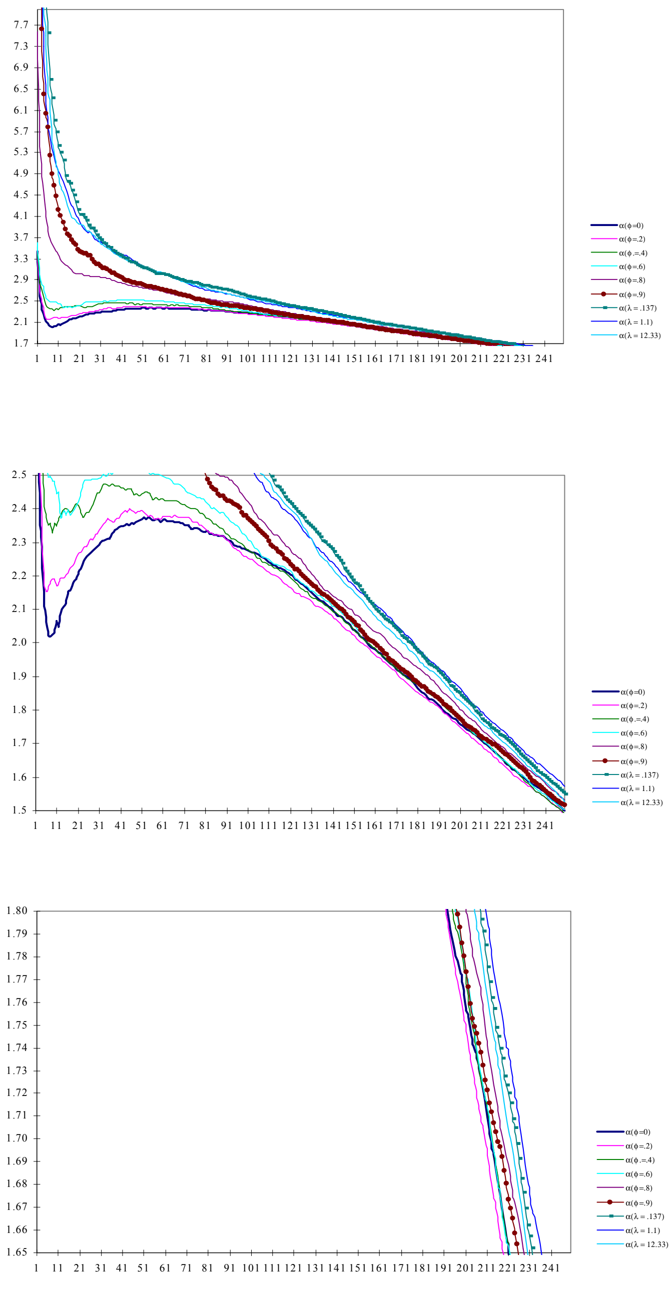




Notes: The bottom figure is a magnification of the top figure.

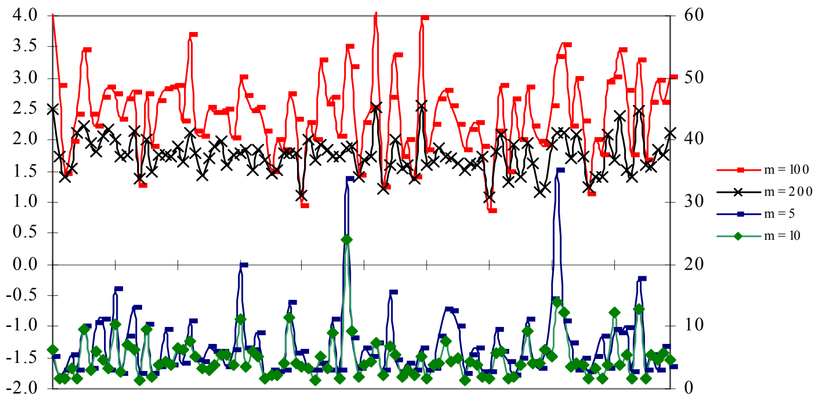
Figure 2

Hill Plot, Stable, $\alpha = 1.7$, $n = 500$



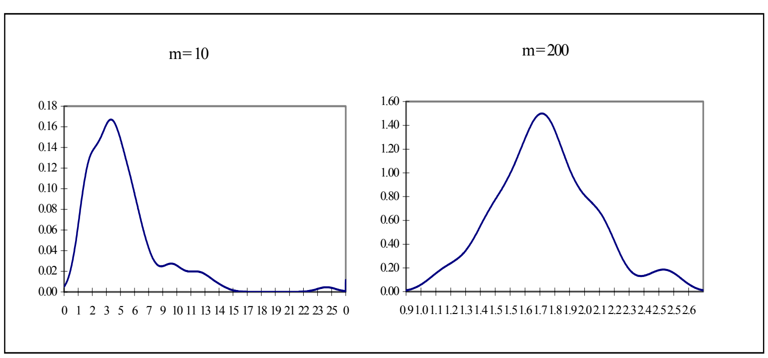
Notes: The figures, top to bottom, are progressive magnifications.

Figure 3
100 Simulated $\hat{\alpha}'_m s$ (Stable)
$\alpha = 1.7$, $m \in \{5, 10, 100, 200\}$, $n = 500$



Notes: The top two lines (m = 100 and m = 200) are plotted against the left Y-axis; the bottom two lines (m = 5 and m = 10) are plotted against the right axis. For $m = 10$, the sample mean and standard deviation are 4.92 and 3.45. For $m = 200$, the sample mean and standard derivation are 1.76 and .30.

Figure 4
Kernel Densities of $\hat{\alpha}_m$, $\alpha = 1.7$, $n = 500$ (Stable)



Notes: For $m = 10$, the Cramer-von Mises p-value is below .01 when $\alpha = 1.7$ is assumed; and the p-value is below .0001 when α is estimated. For $m = 200$, the Cramer-von Mises p-value is .146 when $\alpha = 1.7$ is assumed; and the p-value is .327 when α is estimated.

Table 1: $\hat{\alpha}_m \pm 1.96 \times \widehat{\tilde{\sigma}}_m/\sqrt{m}$ (Stable)

$\alpha = 1.7$, n = 500							
	Simple AR(1)				General AR(1)[a]		
	ϕ				λ		
	0	.2	.6	.9	12.33	1.096	.137
m	$\hat{\alpha}_m \pm k$	$\hat{\alpha}_m \pm k$	$\hat{\alpha}_m \pm k$	$\hat{\alpha}_m \pm k$	$\hat{\alpha}_m \pm k$	$\hat{\alpha}_m \pm k$	$\hat{\alpha}_m \pm k$
5	2.41±2.51	2.32±2.45	2.78±3.17	6.07±7.77	5.91±7.61	6.62±8.53	9.64±12.49
10	2.17±1.71	2.28±1.88	2.55±2.55	4.62±5.60	4.37±5.37	4.78±5.93	5.96±7.47
50	2.36±0.90	2.41±1.01	2.46±1.46	3.19±2.73	2.99±2.63	3.16±2.89	3.40±3.06
100	2.25±0.56	2.27±0.65	2.29±0.93	2.56±1.67	2.52±1.66	2.62±1.81	2.72±1.87
150	2.03±0.37	2.02±0.43	2.02±0.61	2.17±1.13	2.15±1.12	2.19±1.17	2.27±1.22
200	1.77±0.24	1.76±0.28	1.75±0.40	1.86±0.78	1.83±0.76	1.83±0.78	1.91±0.81
250	1.50±0.16	1.50±0.19	1.49±0.27	1.56±0.53	1.55±0.53	1.54±0.54	1.60±0.54
300	1.25±0.11	1.26±0.12	1.29±0.18	1.38±0.33	1.29±0.35	1.29±0.37	1.32±0.36
350	1.14±0.08	1.15±0.09	1.18±0.13	1.24±0.21	1.05±0.22	1.04±0.23	1.06±0.23
$\underline{m}_2$ $(\delta)^b$	200 (.85)	207 (.86)	220 (.87)	261 (.90)	176 (.83)	260 (.89)	265 (.90)
$\bar{m}_\alpha$ $(\delta)^b$	240 (.88)	242 (.88)	253 (.89)	304 (.92)	280 (.91)	290 (.91)	294 (.91)

Notes: a. For each general AR(1) process, ϕ= .9.

b. Minimum m at which 2 does not occur in the 95% interval ($\delta = \ln m / \ln n$).

c. Maximum m at which $\alpha = 1.7$ occur in the 95% interval ($\delta = \ln m / \ln n$).

Appendix 3: Proofs of Main Results

Proof of Lemma 1. By Minkowski's and Jensen's inequality, (1) and the definition of $b_n(\rho m)$, cf. (2), for any $\epsilon \in \mathbb{R}$ and ρ in a neighborhood of 1 (without loss of generality, let $\rho \in (0,2)$), as $n \to \infty$

$$\begin{aligned} ||U^*_{n,t+q}||_r &\leq ||T^*_{n,t+q}||_r + ||E[T^*_{n,t+q}]||_r \\ &\leq 2||T^*_{n,t+q}||_r \\ &= 2P\left(\ln X_{t+q} - \ln b_n(\rho m) > \epsilon\right)^{1/r} \\ &= 2P\left(X_{t+q} > b_n(\rho m)e^{\epsilon}\right)^{1/r} \\ &= 2\left[P\left(X_{t+q} > b_n(\rho m)\right)\frac{P\left(X_{t+q} > b_n(\rho m)e^{\epsilon}\right)}{P\left(X_{t+q} > b_n(\rho m)\right)}\right]^{1/r} \\ &\approx 2\left[(\rho m/n)e^{-\alpha\epsilon}\right]^{1/r} \leq 4(m/n)^{1/r}e^{-\alpha\epsilon/r} = A_r(m/n)^{1/r}, \end{aligned} \tag{30}$$

where $A_r = 4e^{-\alpha\epsilon/r}$.

Similarly, using (3), and Minkowski's inequality, for any positive i and q and any integer $r \geq 1$, as $n \to \infty$

$$\begin{aligned} \left\|U_{n,t+q}\right\|_r &\leq 2\left\|T_{n,t+q}\right\| \\ &= 2\left(E\left|(\ln X_{t+q} - \ln b_n(m))_+\right|^r\right)^{1/r} \\ &\approx 2\left((m/n)r!\alpha^{-r}\right)^{1/r} \\ &= 2r!^{1/r}\left(m/n\right)^{1/r}\alpha^{-1} = B_r\left(m/n\right)^{1/r}, \end{aligned} \tag{31}$$

where $B_r = 2r!^{1/r}\alpha^{-1}$. Finally, for any real-valued $r \geq 1$ and recalling $[r] \leq r$ by convention such that $[r+1] \geq r$, by Liapanov's inequality and using the above derivations,

$$\begin{aligned} \left\|U_{n,t+q}\right\|_r &\leq \left\|U_{n,t+q}\right\|_{[r+1]} \\ &\leq 2[r+1]!^{1/[r+1]}\left(m/n\right)^{1/[r+1]}\alpha^{-1} = C_r\left(m/n\right)^{1/[r+1]}. \end{aligned} \tag{32}$$

■

Proof of Theorem 2.

Step 1 ($m^{-1}\sum_{t=1}^{n} U_t \to 0$): Under Assumption C.1, $\{U_{n,t}, U^*_{n,t}\} = \{U_{n,t}, U^*_{n,t}(\rho,\epsilon)\}$ are L_p-mixingales, $p \geq 1$, with positive constants $\{d_{n,t}, d^*_{n,t}\}$ and coefficients $\{\psi_{n,i}, \psi^*_{n,i}\}$ of size $-\lambda(1/p - 1/r)$ for some $\lambda \geq 0$, $r > p \geq 1$. Fix $p = 1$. We show an L_1-mixingale weak law of large numbers due to Davidson (1993a) (see also Theorem 19.11 of Davidson, 1994) applies to each $\{U_{n,t}/m, U^*_{n,t}/m\}$.

By Lemma 1, each processes $\{U_{n,t}, U^*_{n,t}\}$ is bounded in the L_r-norm for any $r \geq 1$, hence the L_p-assumption itself is superfluous. In particular, for any $\epsilon \in \mathbb{R}$ and ρ in a neighborhood of 1

$$||U^*_{n,t}||_1 \leq A_1(m/n); \quad ||U_{n,t}||_1 \leq B_1\left(m/n\right). \tag{33}$$

Consider $U^*_{n,t}$. By construction the mixingale constants and coefficients satisfy (consult Appendix 1)

$$\begin{aligned} \left\| E\left[U^*_{n,t}|\Im_{t-q}\right]\right\|_1 &\leq (d^*_{n,t})\psi^*_q \\ \left\| U^*_{n,t} - E\left[U^*_{n,t}|\Im_{t+q}\right]\right\|_1 &\leq (d^*_{n,t})\psi^*_{q+1}. \end{aligned} \tag{34}$$

From the boundedness property, (33), we may select $d^*_{n,t} = (m/n)$, hence $U^*_{n,t}/m$ satisfies

$$\begin{aligned} \left\| E\left[U^*_{n,t}/m|\Im_{i-q}\right]\right\|_1 &\leq (d^*_{n,t}/m)\psi^*_q = (1/n)\psi^*_q = d^{*\prime}_{n,t}\psi^*_q \\ \left\| U^*_{n,t}/m - E\left[U^*_{n,t}/m|\Im_{i+q}\right]\right\|_1 &\leq (d^*_{n,t}/m)\psi^*_{q+1} = (1/n)\psi^*_{q+1} = d^{*\prime}_{n,t}\psi^*_{q+1}. \end{aligned} \tag{35}$$

From Davidson (1993a), for $\sum_{t=1}^n U^*_{n,t}/m \to 0$ we require *i.* $\{(U^*_{n,t}/m)/d^{*\prime}_{n,t}\}$ to be uniformly integrable; *ii.* $\limsup_{n\to\infty}\sum_{t=1}^n d^{*\prime}_{n,t} < \infty$; and *iii.* $\lim_{n\to\infty}\sum_{t=1}^n (d^{*\prime}_{n,t})^2$ $= 0$.

For uniform integrability (i), choose some $r, s > 0$, $1/r + 1/s = 1$. By Lemma 1, the construction $1 \leq m \leq n$, and Hölder's and Chebyshev's inequalities, for any $M > 0$

$$\begin{aligned} & \lim_{M\to\infty}\sup_{n,t} E\left[\left|(U^*_{n,t}/m)/d^{*\prime}_{n,t}\right| I\left((U^*_{n,t}/m)/d^{*\prime}_{n,t} > M\right)\right] \\ = & \lim_{M\to\infty}\sup_{n,t} E\left[\left|nU^*_{n,i}/m\right| I\left(nU^*_{n,i}/m > M\right)\right] \\ \leq & \lim_{M\to\infty}\sup_{n,t}(n/m)\left\|U^*_{n,t}\right\|_r \left\|I\left(U^*_{n,t} > (m/n)M\right)\right\|_s \\ = & \lim_{M\to\infty}\sup_{n,t}(n/m)\left\|U^*_{n,t}\right\|_r P\left(U^*_{n,t} > (m/n)M\right)^{1/s} \\ \leq & \lim_{M\to\infty}\sup_{n,t}(n/m)\left\|U^*_{n,t}\right\|_r \left[\left\|U^*_{n,t}\right\|_s^s/[(m/n)^s M^s]\right]^{1/s} \\ = & \lim_{M\to\infty}\sup_{n,t}\left\|U^*_{n,t}\right\|_r \left\|U^*_{n,t}\right\|_s/M \\ \leq & \lim_{M\to\infty}\sup_{n,t} A_r A_s(m/n)/M \\ \leq & \lim_{M\to\infty} A_r A_s/M = 0. \end{aligned} \tag{36}$$

Therefore $(U^*_{n,t}/m)/d^{*\prime}_{n,t}$ is uniformly integrable (see Billingsley, 1995: p. 216).

For (ii), using the constant sequence $\{d^{*\prime}_{n,t}\} = \{1/n\}$ we deduce

$$\limsup_{n\to\infty}\sum\nolimits_{t=1}^n d^{*\prime}_{n,t} = \limsup_{n\to\infty} n(1/n) = 1. \tag{37}$$

Similarly, for (iii) we have

$$\lim_{n\to\infty}\sum\nolimits_{t=1}^n (d^{*\prime}_{n,t})^2 = \lim_{n\to\infty} n^{-1} = 0. \tag{38}$$

Therefore $\sum_{t=1}^n U^*_{n,t}/m \to 0$.

Given the near identity of bounds in (33), an identical argument holds for $U_{n,t}$.

Step 2 ($\hat{\alpha}_m^{-1} \to \alpha^{-1}$): Having demonstrated laws of large numbers hold for $\{U_{n,t}, U^*_{n,t}\}$, Theorem 2.2 of Hsing (1991) applies such that $\hat{\alpha}_m^{-1} \to \alpha^{-1}$.

Step 3: By Step 1, $m^{-1}\sum_{t=1}^{n} U_t \to 0$ for each $U \in \{U_{n,t}, U_{n,t}^*\}$, hence the result

$$\ln X_{([\rho m])} - \ln b_n(\rho m) \to 0 \tag{39}$$

follows immediately from the line of proof of Theorem 2.2 of Hsing (2.2). ■

Proof of Theorem 3. The proof follows upon application of Theorems 17.21 and 19.13 of Davidson (1994) and Theorem 2.2 of Hsing (1991). For Theorem 19.13 of Davidson (1994), see especially equation (19.42) in the line of proof: as long as the centered approximating processes satisfy a law of large numbers, the result goes through. ■

Proof of Theorem 4. Define

$$\begin{aligned} U_{n,t}^* \quad \equiv \quad & I\left(\ln X_t^+ - b_n(m) > \epsilon/\sqrt{m}\right) - E[I\left(\ln X_t^+ - b_n(m) > \epsilon/\sqrt{m}\right)] \\ & U_{n,t} \equiv (\ln X_t - \ln b_n(m))_+ - E(\ln X_t - \ln b_n(m))_+ \\ & TT_{n,t}(\omega) \equiv m^{-1/2}\sigma_m^{-1}(\omega)\left(\omega_1 U_{n,t} + \omega_2\alpha^{-1}U_{n,t}^*\right) \\ & H_n^+ \equiv m^{-1}\sum\nolimits_{t=1}^{n} (\ln X_t - \ln b_n(m))_+ \\ & \tilde{H}_n \equiv m^{-1}\sum\nolimits_{t=1}^{m} \left(\ln X_{(i)} - \ln b_n(m)\right) \\ & S_n^* \equiv m^{-1}\sum\nolimits_{t=1}^{n} U_{n,t}^* \end{aligned} \tag{40}$$

where $\omega = [\omega_1, \omega_2] \in \mathbb{R}^2$ is arbitrary, and define the variance term

$$\sigma_m^2(\omega) = m^{-1}E\left(\sum\nolimits_{t=1}^{n}\left(\omega_1 U_{n,t} + \omega_2\alpha^{-1}U_{n,t}^*\right)\right)^2. \tag{41}$$

The proof of the distribution limit (13) follows from Lemmas 10-14, below.

From Lemma 14, for some random vector (Z_1, Z_2), $Z_i \sim N(0,1)$,

$$\sqrt{m}\left(\sigma_{m,1}^{-1}\left(\tilde{H}_n - E[H_n^+]\right), \sigma_{m,2}^{-1}\left(\ln X_{(m+1)} - \ln b_n(m)\right)\right) \Longrightarrow (Z_1, Z_2), \tag{42}$$

if $r > 2$, $\delta \geq 1 - r\gamma/(r-2)$ and $\gamma > 0$, or $r = 2$ and $\gamma \geq 0$, where $\sigma_{m,1}^2$ and $\sigma_{m,2}^2$ are defined in Lemma 11. If $r = 2$, it is understood that the NED base is uniform mixing. By the continuous mapping theorem and Cramér's Theorem, we deduce for some random variable Y

$$\begin{aligned} & \frac{\sigma_{m,1}}{\sigma_m(1,-1)}\sqrt{m}\left(\frac{\tilde{H}_n - E[H_n^+]}{\sigma_{m,1}}\right) \\ & -\frac{\sigma_{m,2}}{\sigma_m(1,-1)}\sqrt{m}\left(\frac{\ln X_{(m+1)} - \ln b_n(m)}{\sigma_{m,2}}\right) \\ \Longrightarrow \quad & \lim_{n\to\infty}\left(\frac{\sigma_{m,1}}{\sigma_m(1,-1)}\right)Z_1 - \lim_{n\to\infty}\left(\frac{\sigma_{m,2}}{\sigma_m(1,-1)}\right)Z_2 \equiv Y. \end{aligned} \tag{43}$$

From Lemma 11 and the fact that each Z_i has a unit variance, we deduce the random variable Y satisfies

$$\begin{aligned} V[Y] &= \left(\lim_{n\to\infty}\frac{\sigma_{m,1}}{\sigma_m(1,-1)}\right)^2+\left(\lim_{n\to\infty}\frac{\sigma_{m,2}}{\sigma_m(1,-1)}\right)^2 \\ &\quad -2\lim_{n\to\infty}\left(\frac{\sigma_{m,1}}{\sigma_m(1,-1)}\right)\lim_{n\to\infty}\left(\frac{\sigma_{m,2}}{\sigma_m(1,-1)}\right)cov(Z_1,Z_2) \\ &= 1. \end{aligned} \tag{44}$$

Therefore, from the stability property of normal random variables and the fact that each Z_i has a zero mean, we deduce $Y \sim N(0,1)$, and from the definition of $\tilde{H}_n$, cf. (40), we obtain

$$\begin{aligned} &\frac{\sigma_{m,1}}{\sigma_m(1,-1)}\sqrt{m}\left(\frac{\tilde{H}_n-E[H_n^+]}{\sigma_{m,1}}\right) \\ &-\frac{\sigma_{m,2}}{\sigma_m(1,-1)}\sqrt{m}\left(\frac{\ln X_{(m+1)}-\ln b_n(m)}{\sigma_{m,2}}\right) \\ =\;& \frac{1}{\sigma_m(1,-1)}\sqrt{m}\left(\left(\tilde{H}_n-E[H_n^+]\right)-\left(\ln X_{(m+1)}-\ln b_n(m)\right)\right) \\ =\;& \frac{1}{\sigma_m(1,-1)}\sqrt{m}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right)-\frac{1}{\sigma_m(1,-1)}\sqrt{m}\left(E[H_n^+]-\alpha^{-1}\right) \\ \Longrightarrow\;& N(0,1), \end{aligned} \tag{45}$$

if $r > 2$, $\delta \geq 1 - r\gamma/(r-2)$ and $\gamma > 0$; or $r = 2$ and $\gamma \geq 0$.

Hsing (1991: p. 1554) shows that for the positive measurable function g defined in Lemma 13,

$$\sqrt{m}\left(E[H_n^+]-\alpha^{-1}\right)=\sqrt{m}g(b_n(m)) \tag{46}$$

where $\sqrt{m}g(b_n(m)) \to 0$, cf. Lemma 13. Because $\sigma_m^2(\omega) = O(n^{\gamma})$, $\gamma \geq 0$, for any $\omega \in \mathbb{R}^2$, from (45) and (46) we deduce upon application of Cramér's Theorem

$$\frac{1}{\sigma_m(1,-1)}\sqrt{m}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right)+o(1)\Longrightarrow N(0,1), \tag{47}$$

if $r > 2$, $\delta \geq 1 - r\gamma/(r-2)$ and $\gamma > 0$; or $r = 2$ and $\gamma \geq 0$.

The proof of (13) is complete if we show $\sigma_m^2(1,-1)/\sigma_m^2 \to 1$, where $\sigma_m^2 = E[\sqrt{m}(\hat{\alpha}_m^{-1} - \alpha^{-1})]^2$. This follows immediately from (47), Cramér's Theorem, and by the construction $E[\sqrt{m}\sigma_m^{-1}\left(\hat{\alpha}_m^{-1} - \alpha^{-1}\right)]^2 = 1$:

$$\begin{aligned} &\sqrt{m}\sigma_m^{-1}(1,-1)\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right) \\ =\;& \sqrt{m}\frac{\sigma_m}{\sigma_m(1,-1)}\sigma_m^{-1}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right) \\ =\;& \frac{\sigma_m}{\sigma_m(1,-1)}\frac{\sqrt{m}\left(\hat{\alpha}_m^{-1}-\alpha^{-1}\right)}{\sqrt{E[\sqrt{m}(\hat{\alpha}_m^{-1}-\alpha^{-1})]^2}}\Rightarrow N\left(0,1\right), \end{aligned} \tag{48}$$

if and only if $\sigma_m/\sigma_m(1,-1) \to 1$.

If X_t is *iid*, then the mean-zero $(\omega_1 U_{n,t} + \omega_2 \alpha^{-1} U^*_{n,t})$ is *iid* and

$$\begin{aligned} \sigma^2_m(1,-1) &= m^{-1} E\left(\sum\nolimits_{t=1}^n \left(U_{n,t} - \alpha^{-1} U^*_{n,t}\right)\right)^2 \\ &= m^{-1} \sum\nolimits_{t=1}^n E\left(U_{n,t} - \alpha^{-1} U^*_{n,t}\right)^2 \\ &= m^{-1} \sum\nolimits_{t=1}^n \left(E[U^2_{n,t}] - 2\alpha^{-1} E[U_{n,t} U^*_{n,t}] + \alpha^{-2} E[U^{*2}_{n,t}]\right)^2 . \end{aligned} \tag{49}$$

Using (1)-(3), it is straightforward to show as $n \to \infty$

$$\begin{aligned} E[U^2_{n,t}] &\approx \frac{m}{n} 2\alpha^{-2} - \left(\frac{m}{n}\right)^2 \alpha^{-2} \\ E[U^{*2}_{n,t}] &\approx \frac{m}{n} e^{-\alpha\epsilon/\sqrt{m}} - \left(\frac{m}{n}\right)^2 e^{-2\alpha\epsilon/\sqrt{m}} \\ E[U_{n,t} U^*_{n,t}] &\approx \frac{m}{n} \alpha^{-1} - \left(\frac{m}{n}\right)^2 \alpha^{-1} e^{-\alpha\epsilon/\sqrt{m}}, \end{aligned} \tag{50}$$

hence

$$\begin{aligned} & \lim \sigma^2_m(1,-1) \\ = & \lim \frac{n}{m} \left[\frac{m}{n} 2\alpha^{-2} - \left(\frac{m}{n}\right)^2 \alpha^{-2} - 2\alpha^{-1} \left(\frac{m}{n}\alpha^{-1} - \left(\frac{m}{n}\right)^2 \alpha^{-1} e^{-\alpha\epsilon/\sqrt{m}}\right)\right] \\ & + \lim \left[\frac{n}{m} \alpha^{-2} \left(\frac{m}{n} e^{-\alpha\epsilon/\sqrt{m}} - \left(\frac{m}{n}\right)^2 e^{-2\alpha\epsilon/\sqrt{m}}\right)\right] \\ = & 2\alpha^{-2} - 2\alpha^{-2} + \alpha^{-2} = \alpha^{-2}. \end{aligned} \tag{51}$$

■

Lemma 10 *Let Assumptions A.1, B.1 and C.2 hold. In particular, let* $m = [n^\delta]$, $0 < \delta < 1$, *and assume* $\sigma^2_m = O(n^\gamma)$, $\gamma \geq 0$. *Then for any* $\omega \in \mathbb{R}^2$

$$\sum\nolimits_{t=1}^n TT_{n,t}(\omega) \Longrightarrow N(0,1) \tag{52}$$

pointwise in $\omega \in \mathbb{R}^2$ *if* $r > 2$, $\delta \geq 1 - r\gamma/(r-2)$ *and* $\gamma > 0$; *or* $r = 2$ *and* $\gamma \geq 0$.

Lemma 11 *Under the assumptions of Lemma 10, for some Gaussian random vector* (Z_1, Z_2) *with standard normal marginal distributions,* $Z_i \sim N(0,1)$,

$$m^{-1/2}\left(\sigma^{-1}_{n,1}\left(H^+_n - E[H^+_n]\right), \ \ \sigma^{-1}_{n,2}\alpha^{-1}\left(S^*_n - E[S^*_n]\right)\right) \Longrightarrow (Z_1, Z_2) \tag{53}$$

where

$$\begin{aligned} \sigma^2_{m,1} &= m^{-1} E\left(\sum\nolimits_{t=1}^n U_{n,t}\right)^2 \\ \sigma^2_{m,2} &= m^{-1} E\left(\sum\nolimits_{t=1}^n \alpha^{-1} U^*_{n,t}\right)^2, \end{aligned} \tag{54}$$

and for any $\omega \in \mathbb{R}^2$

$$\omega_1^2 \left(\lim_{n\to\infty} \frac{\sigma_{m,1}}{\sigma_m(\omega)} \right)^2 + \omega_2^2 \left(\lim_{n\to\infty} \frac{\sigma_{m,2}}{\sigma_m(\omega)} \right)^2 \tag{55}$$
$$+2\omega_1\omega_2 \lim_{n\to\infty} \left(\frac{\sigma_{m,1}}{\sigma_m(\omega)} \right) \lim_{n\to\infty} \left(\frac{\sigma_{m,2}}{\sigma_m(\omega)} \right) cov(Z_1, Z_2) = 1.$$

Lemma 12 *If Assumption A.2 holds, then $(m/n)b_n(m)^\alpha - c = o(n^{-(1-\delta)\theta/\alpha})$. If Assumption A.2 and B.2 hold, $(m/n)b_n(m)^\alpha - c = o(1/\sqrt{m})$. Moreover, if Assumptions A.2 and B.2 hold, and $X_{(m+1)}/b_n(m) = 1 + o_p(n^{-\xi})$, $\xi > (1-\delta)\theta/\alpha$, then $(m/n)X^\alpha_{(m+1)} - c = o_p\ (1/\sqrt{m})$.*

Lemma 13 *Let Assumptions A.2 and B.2 hold. There exists a positive measurable function g on $(0, \infty)$ such that*

$$L(\lambda z)/L(z) - 1 = O(g(z)). \tag{56}$$

In particular, g has bounded increase such that $g(\lambda z)/g(z) \leq D\lambda^\tau$ for some $D > 0$, $\lambda \geq 1$ and $\tau \leq 0$[13] *and as $n \to \infty$*

$$\sqrt{m} g(b_n(m)) \to 0. \tag{57}$$

Lemma 14 *Under the conditions of Lemma 10*

$$\sqrt{m} \left(\sigma_{m,1}^{-1} \left(H_n^+ - E[H_n^+] \right),\ \ \sigma_{m,2}^{-1} \left(\ln X_{(m+1)} - \ln b_n(m) \right) \right) \Longrightarrow (Z_1, Z_2), \tag{58}$$

and

$$\sqrt{m} \left(\sigma_{m,1}^{-1} \left(\tilde{H}_n - E[H_n^+] \right),\ \ \sigma_{m,2}^{-1} \left(\ln X_{(m+1)} - \ln b_n(m) \right) \right) \Longrightarrow (Z_1, Z_2), \tag{59}$$

where $\sigma^2_{m,1}$, $\sigma^2_{m,2}$ and the random vector (Z_1, Z_2) are defined in Lemma 11.

Proof of Lemma 10. We verify the applicability of Theorem 2 of de Jong (1997) for $\sum_{t=1}^n TT_{n,t}(\omega)$ pointwise in $\mathbb{R}^2$, which follows by Assumptions 2.*a*-*d* of that work.

Assumption 2.*a* of de Jong (1997) holds by construction: $E\left(TT_{n,t}(\omega)\right) = 0$ and $V(\sum_{t=1}^n TT_{n,t}(\omega)) = 1$ for any $\omega \in \mathbb{R}^2$.

For Assumption 2.*c*, observe that by Lemma 1 the processes $\{U_{n,t}, U^*_{n,t}\}$ are L_r-bounded for arbitrary $r \geq 1$, and by assumption are L_2-NED on $\{\varepsilon_t\}$ with constants $\{d_{n,t}, d^*_{n,t}\}$ and coefficients $\{\psi_q, \psi^*_q\}$ of sizes $\{-1/2, -1/2\}$. Therefore $\{m^{-1/2}\sigma_m^{-1} U_{n,t}, m^{-1/2}\sigma_m^{-1} U^*_{n,t}\}$ are L_2-NED on $\{\varepsilon_t\}$ with coefficients $\{\psi_q, \psi^*_q\}$ and constants

$$\{m^{-1/2}\sigma_m^{-1} d_{n,t}, m^{-1/2}\sigma_m^{-1} d^*_{n,t}\}, \tag{60}$$

which implies $\{TT_{n,t}\}$ is L_2-NED on $\{\varepsilon_t\}$ with constants $\{dd_{n,t}\}$

$$dd_{n,t} = \max\{m^{-1/2}\sigma_m^{-1} d_{n,t}, m^{-1/2}\sigma_m^{-1} d^*_{n,t}\}, \tag{61}$$

[13] A function g has bounded increase if for $0 < D, z_0, \tau < \infty$, the condition $g(\lambda z)/g(z) \leq D\lambda^\tau$ holds for $\lambda \geq 1$, $z \geq z_0$.

and coefficient size $-\min\{1/2, 1/2\} = -1/2$, because it is a linear function of L_2-NED processes $\{n^{-1/2}\sigma_m^{-1}U_{n,t}, n^{-1/2}\sigma_m^{-1}U_{n,t}^*\}$ (e.g., Davidson, 1994: Theorem 17.8).This, along with the maintained mixing assumption, establishes Assumption 2.c of de Jong (1997).

For Assumption 2.b, we require[14] $TT_{n,t}(\omega)/dd_{n,t}$ to be L_r-bounded for $r \geq 2$ uniformly in t, n and ω, and additionally $TT_{n,t}(\omega)^2/dd_{n,t}^2$ to be uniformly integrable for $r = 2$. For $dd_{n,t}$, cf. (61), we require the NED constants $\{d_{n,t}, d_{n,t}^*\}$ of $\{U_{n,t}, U_{n,t}^*\}$. By Lemma 1 for any $r \geq 1$ (for notational brevity, assume r is integer-valued)

$$||U_{n,t+q}||_r \leq B_r\,(m/n)^{1/r}\,, \quad ||U_{n,t+q}^*||_r \leq A_r(m/n)^{1/r}, \tag{62}$$

hence we may arbitrarily select the constants $\{d_{n,t}, d_{n,t}^*\}$ to be

$$\{d_{n,t}, d_{n,t}^*\} = \{(m/n)^{1/r}\,, (m/n)^{1/r}\}. \tag{63}$$

Thus, for any t

$$\begin{aligned} dd_{n,t} &= \max\{m^{-1/2}\sigma_m^{-1}d_{n,t}, m^{-1/2}\sigma_m^{-1}d_{n,t}^*\} \qquad (64) \\ &= m^{-1/2}\sigma_m^{-1}\,(m/n)^{1/r} = m^{-(1/2-1/r)}n^{-1/r}\sigma_m^{-1}. \end{aligned}$$

From the boundedness properties (62), the definition of $dd_{n,t}$, and Minkowski's inequality, for $r \geq 2$ and each $\omega \in \mathbb{R}^2$ we deduce

$$\begin{aligned} & \sup_{n,t} \|TT_{n,t}(\omega)/dd_{n,t}\|_r \qquad (65) \\ = \;& \sup_{n,t} \left\| m^{-1/2}\sigma_m^{-1}\left(\omega_1 U_{n,t} + \omega_2\alpha^{-1}U_{n,t}^*\right)/(\sigma_m^{-1}m^{-(1/2-1/r)}n^{-1/r})\right\|_r \\ = \;& \sup_{n,t} m^{-1/2}m^{(1/2-1/r)}n^{1/r}\left\|\left(\omega_1 U_{n,t} + \omega_2\alpha^{-1}U_{n,t}^*\right)\right\|_r \\ = \;& \sup_{n,t}(n/m)^{1/r}\left\|\left(\omega_1 U_{n,t} + \omega_2\alpha^{-1}U_{n,t}^*\right)\right\|_r \\ \leq \;& \sup_{n,t}(n/m)^{1/r}\left(\omega_1\|U_{n,t}\|_r + \omega_2\alpha^{-1}\left\|U_{n,t}^*\right\|_r\right) \\ \leq \;& \sup_{n,t}(n/m)^{1/r}\left(\omega_1 B_r\,(m/n)^{1/r} + \omega_2\alpha^{-1}A_r(m/n)^{1/r}\right) \\ \leq \;& \omega_1 B_r + \omega_2\alpha^{-1}A_r < \infty. \end{aligned}$$

This demonstrates the L_r-boundedness of $TT_{n,t}(\omega)/dd_{n,t}$ for any $r \geq 2$. Additionally, if $r = 2$ by the Cauchy-Schwartz and Hölder's inequalities, and using

[14] de Jong (1997) uses two sets of constants for the boundedness and order conditions: the NED constants $dd_{n,t}$, and some positive constant array, say $a_{n,t}$. It is immediate that we take them to be idenitcal, and do so here: in all that follows, we simply use $dd_{n,t}$.

(65),

$$
\begin{aligned}
& \lim_{M\to\infty} \sup_{n,t} E\left[\left(TT_{n,t}(\omega)^2/dd_{n,t}^2\right) I\left(TT_{n,t}^2/dd_{n,t}^2 > M\right)\right] && (66)\\
\leq\ & \lim_{M\to\infty} \sup_{n,t} \left[E\left(TT_{n,t}(\omega)^2/dd_{n,t}^2\right)^2\right]^{1/2} P\left(TT_{n,t}^2/dd_{n,t}^2 > M\right)^{1/2} \\
\leq\ & \lim_{M\to\infty} \sup_{n,t} \left[E\left(TT_{n,t}(\omega)^4/dd_{n,t}^4\right)\right]^{1/2} \left[E\left(TT_{n,t}^4/dd_{n,t}^4\right)/M^2\right]^{1/2} \\
=\ & \lim_{M\to\infty} \sup_{n,t} \left(\left\|TT_{n,t}(\omega)/dd_{n,t}\right\|_4 / M^{1/4}\right)^4 \\
\leq\ & \lim_{M\to\infty} \left(\omega_1 B_4 + \omega_2 \alpha^{-1} A_4\right)^4 / M = 0.
\end{aligned}
$$

Therefore for each point $\omega \in \mathbb{R}^2$ the sequence $TT_{n,t}(\omega)^2/dd_{n,t}^2$ is uniformly integrable, cf. Billingsley (1995: p. 216). This proves Assumption 2.b of de Jong (1997) holds.

Finally, for Assumption 2.d define the integer sequences $g_n = [n^{1-a}]$ for some $a \in (0,1]$ and $r_n = [n/g_n]$, and define $M_{n,t} = \max_{(t-1)g_n < s \leq t g_n}\{dd_{n,s}\}$ and $M_{n,r_n+1} = \max_{r_n g_n < s \leq n}\{dd_{n,s}\}$. We require for some such sequences $\{g_n\}$ and $\{r_n\}$

$$\max_{1\leq t\leq r_n+1} M_{n,t} = o(g_n^{-1/2}), \quad \sum\nolimits_{t=1}^{r_n} M_{n,t}^2 = O(g_n^{-1}). \tag{67}$$

By (64), we deduce for each t and r_n

$$\max_{1\leq t\leq r_n+1} M_{n,t} = M_{n,t} = M_{n,r_n+1} = m^{-(1/2-1/r)} n^{-1/r} \sigma_m^{-1}. \tag{68}$$

We therefore require (recalling $m \approx n^{\delta}$, $g_n \approx n^{1-a}$, $r_n \approx n^a$)

$$
\begin{aligned}
\max_{1\leq t\leq r_n+1} M_{n,t} &= \sigma_m^{-1} m^{-(1/2-1/r)} n^{-1/r} && (69)\\
&\approx \sigma_m^{-1} n^{-\delta(1/2-1/r)-1/r} = o(n^{-(1-a)/2}) \\
\sum\nolimits_{t=1}^{r_n} M_{n,t}^2 &= r_n \sigma_m^{-2} m^{-(1-2/r)} n^{-2/r} \\
&\approx \sigma_m^{-2} n^{a-\delta(1-2/r)-2/r} = O(n^{-(1-a)}).
\end{aligned}
$$

Recall $\sigma_m^2 = O(n^{\gamma})$, $\gamma \geq 0$. In turn, from (67) and (69) $\max_{1\leq t\leq r_n+1} M_{n,t} = o(g_n^{-1/2})$ *if and only if*

$$
\begin{aligned}
& \sigma_m^{-1} n^{-\delta(1/2-1/r)-1/r} = o(n^{-(1-a)/2}) && (70)\\
& \Longrightarrow \sigma_m^{-1} n^{-\delta(1/2-1/r)-1/r+(1-a)/2} = o(1) \\
& \Longrightarrow \sigma_m n^{\delta(1/2-1/r)+1/r-(1-a)/2} \to \infty,
\end{aligned}
$$

which is true if

$$
\begin{aligned}
& \gamma/2 + \delta(1/2-1/r) + 1/r - (1-a)/2 > 0 && (71)\\
& \Longrightarrow r\gamma + \delta(r-2) - (r-2) + ra > 0.
\end{aligned}
$$

If $r = 2$, we therefore require

$$\gamma + a > 0, \tag{72}$$

and if $r > 2$,

$$\delta > 1 - r(\gamma + a)/(r - 2). \tag{73}$$

Next, from (67) and (69) we deduce $\sum_{t=1}^{r_n} M_{n,t}^2 = O(g_n^{-1})$ *if and only if*

$$\begin{aligned} &\sigma_m^{-2} n^{a-\delta(1-2/r)-2/r} = O(n^{-(1-a)}) \\ &\Longrightarrow \sigma_m^{-2} n^{1-\delta(1-2/r)-2/r} = O(1) \\ &\Longrightarrow \gamma - 1 + \delta(1 - 2/r) + 2/r \geq 0. \end{aligned} \tag{74}$$

If $r = 2$, we require

$$\gamma \geq 0, \tag{75}$$

and if $r > 2$,

$$\delta \geq 1 - r\gamma/(r - 2). \tag{76}$$

Together, because $a \in (0, 1]$ is arbitrary and $\delta \in (0, 1)$ by construction, if $r = 2$, then (72) and (75) imply we must have $\gamma \geq 0$; and if $r > 2$, (73) and (76) imply we require $\delta \geq 1 - r\gamma/(r - 2)$ and $\gamma > 0$. Under these two cases each condition of Assumption 2 of de Jong (1997) holds, hence Theorem 2 of de Jong (1997) applies, which completes the proof. ■

Proof of Lemma 11. By Lemma 10 and the definitions of H_n^+, S_n^* and $TT_{n,t}$, for any $\omega \in \mathbb{R}^2$

$$\begin{aligned} &\sqrt{m}\sigma_m^{-1}(\omega) \left(\omega_1 \left(H_n^+ - E[H_n^+]\right) + \omega_2 \alpha^{-1} \left(S_n^* - E[S_n^*]\right)\right) \\ = \;& \sum_{t=1}^{n} TT_{n,t}(\omega) \Longrightarrow N(0, 1) \end{aligned} \tag{77}$$

pointwise in $\omega \in \mathbb{R}^2$, where

$$\sigma_m^2(\omega) = m^{-1} E\left(\sum_{t=1}^{n} \left(\omega_1 U_{n,t} + \omega_2 \alpha^{-1} U_{n,t}^*\right)\right)^2. \tag{78}$$

Setting $\omega = [0, 1]$ and $\omega = [1, 0]$, and invoking a Cramér-Wold device suffices to prove the joint limit

$$\sqrt{m}\left(\sigma_{m,1}^{-1}\left(H_n^+ - E[H_n^+]\right), \sigma_{m,2}^{-1}\alpha^{-1}\left(S_n^* - E[S_n^*]\right)\right) \Longrightarrow (Z_1, Z_2) \tag{79}$$

for some random vector (Z_1, Z_2), $Z_i \sim N(0, 1)$, where $\sigma_{m,1}^2 = \sigma_m^2(1, 0)$ and $\sigma_{m,2}^2 = \sigma_m^2(0, 1)$.

From (77) and (79) and Cramér's theorem, we therefore obtain

$$\begin{aligned} &\left(\omega_1 \frac{\sigma_{m,1}}{\sigma_m(\omega)} \sqrt{m} \left(\frac{H_n^+ - E[H_n^+]}{\sigma_{m,1}}\right) + \omega_2 \frac{\sigma_{m,2}}{\sigma_m(\omega)} \sqrt{m}\alpha^{-1} \left(\frac{S_n^* - E[S_n^*]}{\sigma_{m,2}}\right)\right) \\ \Longrightarrow \;\; &\omega_1 \lim_{n\to\infty} \left(\frac{\sigma_{m,1}}{\sigma_m(\omega)}\right) Z_1 + \omega_2 \lim_{n\to\infty} \left(\frac{\sigma_{m,2}}{\sigma_m(\omega)}\right) Z_2 \;\sim\; N(0, 1), \end{aligned} \tag{80}$$

pointwise in $\mathbb{R}^2$ for the same standard normal vector (Z_1, Z_2) in (79). From the standard normality of limit (80), we immediately deduce

$$\begin{aligned} & V\left[\omega_1 \lim_{n\to\infty}\left(\frac{\sigma_{n,1}}{\sigma_n(\omega)}\right) Z_1 + \omega_2 \lim_{n\to\infty}\left(\frac{\sigma_{n,2}}{\sigma_n(\omega)}\right) Z_2\right] \qquad (81) \\ = \; & \omega_1^2\left(\lim_{n\to\infty}\frac{\sigma_{n,1}}{\sigma_n(\omega)}\right)^2 + \omega_2^2\left(\lim_{n\to\infty}\frac{\sigma_{n,2}}{\sigma_n(\omega)}\right)^2 \\ & +2\omega_1\omega_2 \lim_{n\to\infty}\left(\frac{\sigma_{n,1}}{\sigma_n(\omega)}\right)\lim_{n\to\infty}\left(\frac{\sigma_{n,2}}{\sigma_n(\omega)}\right) cov(Z_1, Z_2) \\ = \; & 1. \end{aligned}$$

■

Proof of Lemma 12.

Step 1 ($(m/n)b_n^\alpha$): From (8) of Assumption A.2 and the definition of $b_n \equiv b_n(m)$, we deduce

$$(m/n) = cb_n^{-\alpha}1 + o(b_n^{-\theta})) \qquad (82)$$

hence

$$\begin{aligned} (m/n)b_n^\alpha &= c(1 + o(b_n^{-\theta})) = c + o(b_n^{-\theta})) = c + o(1) \qquad (83) \\ (m/n)b_n^\alpha - c &= o(b_n^{-\theta}). \end{aligned}$$

Define $\tilde{c} \equiv (m/n)b_n^\alpha$ and recall $m \approx n^\delta$. Hence, from (82) and (83) b_n^θ can be written as

$$\begin{aligned} (m/n)b_n^\alpha &= c(1 + o(b_n^{-\theta})) \qquad (84) \\ (m/n)^{\theta/\alpha}b_n^\theta &= c^{\theta/\alpha}(1 + o(b_n^{-\theta}))^{\theta/\alpha} \\ b_n^\theta &= (n/m)^{\theta/\alpha} c^{\theta/\alpha}(1 + o(b_n^{-\theta}))^{\theta/\alpha} \\ &= (n/m)^{\theta/\alpha} \tilde{c}^{\theta/\alpha} = \left(n/n^\delta\right)^{\theta/\alpha} \tilde{c}^{\theta/\alpha} \\ &= n^{(1-\delta)\theta/\alpha)}\tilde{c}^{\theta/\alpha} = n^{(1-\delta)\theta/\alpha)}\left(c + o(1)\right)^{\theta/\alpha} \end{aligned}$$

which gives

$$o(b_n^{-\theta}) = o(n^{-(1-\delta)\theta/\alpha)}). \qquad (85)$$

Together, (82)-(85) give the convergence rate of $(m/n)b_n^\alpha$:

$$(m/n)b_n^\alpha - c = o(n^{-(1-\delta)\theta/\alpha)}) = o(1), \qquad (86)$$

given $\delta < 1$. This implies

$$(m/n)b_n^\alpha - c = o(1/\sqrt{m}), \qquad (87)$$

if and only if

$$n^{\delta/2} \times o(n^{-(1-\delta)\theta/\alpha)}) = o(1), \qquad (88)$$

if

$$\delta/2 - (1-\delta)\theta/\alpha < 0, \qquad (89)$$

which follows from simple manipulation and Assumption B.2:

$$\begin{aligned}\delta/2-(1-\delta)\theta/\alpha &< 0 \\ \frac{\delta}{2}-\frac{\theta}{\alpha}+\frac{\delta\theta}{\alpha} &< 0 \implies \delta<\frac{2\theta}{2\theta+\alpha}.\end{aligned} \tag{90}$$

Step 2 ($(m/n)X^{\alpha}_{(m+1)}$): Under the maintained assumptions and Step 1, we can write

$$\begin{aligned}\frac{m}{n}X^{\alpha}_{(m+1)} &= \left(\frac{X_{(m+1)}}{b_n}\right)^{\alpha}\frac{m}{n}b_n^{\alpha} \\ &= (1+o_p(n^{-\xi}))^{\alpha}\frac{m}{n}b_n^{\alpha} \\ &= (1+o_p(n^{-\xi}))^{\alpha}c(1+o(1/\sqrt{m})).\end{aligned} \tag{91}$$

The term $(1+o_p(n^{-\xi}))^{\alpha}$ is bounded by $(1+o(1/\sqrt{m}))$. In order to see this, for any $\alpha>0$ let $d(\alpha)=[\alpha+1]$, the next integer great than α. Then, for any $\alpha>0$

$$\begin{aligned}\left|\left(1+o(n^{-\xi})\right)^{\alpha}\right| &\leq \left(1+|o(n^{-\xi})|\right)^{\alpha} \\ &\leq \left(1+|o(n^{-\xi})|\right)^{d(\alpha)} \\ &= \sum\nolimits_{i=0}^{d(\alpha)}1^{i}|o(n^{-\xi})|^{d(\alpha)-i}\left(\begin{array}{c}d(\alpha)\\ i\end{array}\right) \\ &= 1+o(n^{-\xi})=1+o(1/\sqrt{m}).\end{aligned} \tag{92}$$

The last line follows from the maintained assumptions, cf. $\xi>(1-\delta)\theta/\alpha$: $o(n^{-\xi})$ is $o(1/\sqrt{m})$ if $\sqrt{m}n^{-\xi}\approx n^{\delta/2-\xi}\to 0$, *if and only if* $\xi>\delta/2$. The equality $\xi>\delta/2$ holds sufficiently if $(1-\delta)\theta/\alpha>\delta/2$, which is true by Assumption B.2: see (89) and (90).

Together, (91) and 92) imply

$$\begin{aligned}(m/n)X^{\alpha}_{(m+1)} &= (1+o_p(n^{-\xi}))^{\alpha}c_2(1+o(1/\sqrt{m})) \\ &= c(1+o(1/\sqrt{m})).\end{aligned} \tag{93}$$

■

Proof of Lemma 13. By (8) of Assumption A.2 we have $L(x)=c(1+o(x^{-\theta}))$, $\theta>0$, hence it can be shown that

$$L(\lambda x)/L(x)-1=o(x^{-\theta}). \tag{94}$$

This implies we can simply choose $g(x)=x^{-\theta}$. In order to show $\sqrt{m}g(b_n(m)) = O(1)$, define $\tilde{c}\equiv(m/n)b_n^{\alpha}(m)$. From Lemma 12 we have $\tilde{c}=c+o(1)$ for any process that satisfies Assumption A.2. Moreover, using the definition of $\tilde{c}$ and Lemma 12, we can write $b_n(m)^{\theta}$ as

$$\begin{aligned}b_n(m)^{\theta} &= (n/m)^{\theta/\alpha}\tilde{c}^{\theta/\alpha}=\left(n/n^{\delta}\right)^{\theta/\alpha}\tilde{c}^{\theta/\alpha} \\ &= n^{(1-\delta)\theta/\alpha}\tilde{c}^{\theta/\alpha}=n^{(1-\delta)\theta/\alpha}\left(c+o(1)\right)^{\theta/\alpha}.\end{aligned} \tag{95}$$

Thus, $\sqrt{m}g(b_n(m)$ can be expressed as

$$\begin{aligned}\sqrt{m}g(b_n(m)) &= n^{\delta/2} \times b_n(m)^{-\theta} \qquad (96)\\ &= n^{\delta/2} \times n^{-(1-\delta)\theta/\alpha} \times (c+o(1))^{-\theta/\alpha}\\ &= n^{\delta/2-(1-\delta)\theta/\alpha} \times (c+o(1))^{-\theta/\alpha}\\ &= o(n^{\delta/2-(1-\delta)\theta/\alpha}).\end{aligned}$$

The term $o(n^{\delta/2-(1-\delta)\theta/\alpha})$ is $o(1)$ *if and only if* $\delta/2-(1-\delta)\theta/\alpha<0$, which holds by Assumption B.2, $\delta<2\theta/(2\theta+\alpha)$.

In order to see that g has bounded increase, observe that for any $z>0$ and any $\lambda\geq 1$:

$$g(\lambda z)/g(z)=(z\lambda)^{-\theta}/z^{-\theta}=\lambda^{-\theta}. \qquad (97)$$

Because the right-hand-side holds for any $z>0$ and $\theta>0$, the result is proved. ■

Proof of Lemma 14. The proof follows from Lemmas 11 and 13 and Theorem 2, and essentially mimics the line of proof Theorem 2.4 of Hsing (1991). From Lemma 13, Hsing's (1991) equation (2.3a) holds, and from Theorem 2 and the maintained assumptions we have

$$\ln X_{([\rho m])}-\ln b_n([\rho m])\to 0$$

for any ρ in a neighborhood of 1. Therefore imitating the line of proof of Theorem 2.4 of Hsing (1991), it is straightforward to show that Lemma 11 implies

$$\sqrt{m}\left(\sigma_{m,1}^{-1}\left(H_n^{+}-E[H_n^{+}]\right),\sigma_{m,2}^{-1}\left(\ln X_{(m+1)}-\ln b_n(m)\right)\right)\Longrightarrow(Z_1,Z_2) \qquad (98)$$

and

$$\sqrt{m}\left(\sigma_{m,1}^{-1}\left(\tilde{H}_n-E[H_n^{+}]\right),\sigma_{m,2}^{-1}\left(\ln X_{(m+1)}-\ln b_n(m)\right)\right)\Longrightarrow(Z_1,Z_2) \qquad (99)$$

for the same random vector (Z_1,Z_2) in Lemma 10. ■

Proof of Lemma 5. Define

$$\hat{\sigma}_m^2(\omega) \equiv \frac{1}{m}\sum\nolimits_{t=1}^{n}\sum\nolimits_{s=1}^{n}w_{n,s,t}\left(\omega_1U_{n,t}+\omega_2\alpha^{-1}U_{n,t}^{*}\right)\left(\omega_1U_{n,s}+\omega_2\alpha^{-1}U_{n,s}^{*}\right), \qquad (100)$$

and recall

$$\sigma_m^2(\omega)\equiv\frac{1}{m}E\left(\sum\nolimits_{t=1}^{n}\left(\omega_1U_{n,t}+\omega_2\alpha^{-1}U_{n,t}^{*}\right)\right)^2. \qquad (101)$$

We may write

$$\frac{\hat{\sigma}_m^2(\omega)}{\sigma_m^2(\omega)}=\sum\nolimits_{t=1}^{n}\sum\nolimits_{s=1}^{n}w_{n,s,t}TT_{n,t}(\omega)TT_{n,s}(\omega), \qquad (102)$$

where $TT_{n,t}(\omega) = m^{-1/2}\sigma_m^{-1}(\omega)(\omega_1 U_{n,t} + \omega_2\alpha^{-1}U_{n,t}^*)$. We will prove in order: $\hat{\sigma}_m^2(\omega)/\sigma_m^2(\omega) \to 1$ pointwise in $\omega \in \mathbb{R}^2$, $\hat{\sigma}_m^2/\hat{\sigma}_m^2(1,-1) \to 1$, and $\sigma_m^2(1,-1)/\sigma_m^2 \to 1$, from which follows $\hat{\sigma}_m^2/\sigma_m^2 \to 1$.

Step 1 ($\hat{\sigma}_m^2(\omega)/\sigma_m^2(\omega) \to 1$): Under the maintained assumptions, from the line of proof of Lemma 10, $TT_{n,t}(\omega)$ is L_2-NED of size $-1/2$ with constants

$$dd_{n,t} = m^{-(1/2-1/r)}n^{-1/r}\sigma_m^{-1}(\omega). \tag{103}$$

Theorem 2.1 of de Jong and Davidson (2000) will be used to prove $\hat{\sigma}_m^2(\omega)/\sigma_m^2(\omega) \to 1$. We must demonstrate $TT_{n,t}(\omega)$ satisfies for each point $\omega \in \mathbb{R}^2$ Assumptions 2-3, and the kernel $w_{n,s,t}$ satisfies Assumption 1 of that work. We require for some constant array $c_{n,t}$ and $r \geq 2$

$$\begin{aligned}
&\sup_{n\geq 1}\sup_{1\leq t\leq n}\left(\|TT_{n,t}(\omega)\|_r + dd_{n,t}\right)/c_{n,t} < \infty \qquad (104)\\
&\sup_{n\geq 1}\sum\nolimits_{t=1}^{n} c_{n,t}^2 < \infty\\
&\lim_{n\to\infty}\left(l_n^{-1} + l_n \max_{1\leq t\leq n} c_{n,t}^2\right) = 0,
\end{aligned}$$

pointwise in $\omega \in \mathbb{R}^2$. If $r = 2$ we also require $TT_{n,t}(\omega)^2/c_{n,t}^2$ to be uniformly integrable. Because $c_{n,t}$ is arbitrary, we may set $c_{n,t} = dd_{n,t}$.

From the line of proof of Lemma 10, cf. (65), we have

$$\sup_{n\geq 1}\sup_{1\leq t\leq n}\left(\|TT_{n,t}(\omega)\|_r/dd_{n,t}\right) < \infty. \tag{105}$$

pointwise in $\omega \in \mathbb{R}^2$. Similarly, for each point $\omega \in \mathbb{R}^2$ (66) shows $TT_{n,t}(\omega)^2/dd_{n,t}^2$ is uniformly integrable if $r = 2$. Moreover, from (103) and $m \approx n^{\delta}$

$$\begin{aligned}
\sup_{n\geq 1}\sum\nolimits_{t=1}^{n} dd_{n,t}^2 &= \sup_{n\geq 1} n \times m^{-2(1/2-1/r)}n^{-2/r}\sigma_m^{-2}(\omega) \qquad (106)\\
&\approx \sup_{n\geq 1} n^{1-2/r-\delta(1-2/r)}\sigma_m^{-2}(\omega) < \infty
\end{aligned}$$

if $1 - 2/r - \delta(1 - 2/r) - \gamma \leq 0$ given $\sigma_m^2(\omega) = O(n^{\gamma})$, $\gamma \geq 0$. If $r = 2$, then we require $\gamma \geq 0$ which holds by assumption. If $r > 2$, then we require $d \geq 1 - r\gamma/(r-2)$.

Finally, because (103) holds for all t,

$$\begin{aligned}
\max_{1\leq t\leq n} dd_{n,t}^2 &= m^{-2(1/2-1/r)}n^{-2/r}\sigma_m^{-2}(\omega) \qquad (107)\\
&\approx n^{-\delta(1-2/r)-2/r}\sigma_m^{-2}(\omega) = o(n^{-\delta(1-2/r)-2/r-\gamma}).
\end{aligned}$$

Given $\sigma_m^2(\omega) = O(n^{\gamma})$, $\gamma \geq 0$, and $l_n = O(n^{\varsigma})$, $\varsigma \in [0,1]$, in order for $l_n \max_{1\leq t\leq n} dd_{n,t}^2 = 0$, we require

$$\varsigma \leq \delta(1 - 2/r) + 2/r + \gamma. \tag{108}$$

If $r = 2$, we require $\varsigma \leq 1 + \gamma$ which holds by construction. If $r > 2$, we require

$$\frac{r\varsigma - 2}{r - 2} - \frac{\gamma r}{r - 2} \leq \delta. \tag{109}$$

In summary, if $r = 2$ we we must have $\gamma \geq \max\{0, \varsigma\text{-}1\} = 0$; and if $r > 2$ we we must have

$$\delta \geq \max\left\{1 - \frac{\gamma r}{r - 2}, \frac{r\varsigma - 2}{r - 2} - \frac{\gamma r}{r - 2}\right\} = 1 - \frac{\gamma r}{r - 2}. \tag{110}$$

Step 2 ($\hat{\sigma}_m^2/\hat{\sigma}_m^2(1, -1) \to 1$): We can write $\hat{\sigma}_m^2$ as

$$\begin{aligned}
\hat{\sigma}_m^2 &= \frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\hat{Z}_s\hat{Z}_t && (111)\\
&= \frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left(U_{n,s} - \alpha^{-1}U_{n,s}^*\right)\left(U_{n,t} - \alpha^{-1}U_{n,t}^*\right) \\
&\quad + \sum\nolimits_{i=1}^{11} A_{n,i}
\end{aligned}$$

hence

$$\hat{\sigma}_m^2 - \hat{\sigma}_m^2(1, -1) = \sum\nolimits_{i=1}^{11} A_{n,i}, \tag{112}$$

where each $A_{n,i}$ is defined as

$$\begin{aligned}
A_{n,1} &= (m/n)^2\left[\hat{\alpha}_m^{-1} - \alpha^{-1}\right]^2 \frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t} && (113)\\
A_{n,2} &= 2(m/n)\left[\hat{\alpha}_m^{-1} - \alpha^{-1}\right]\left[E\left(\ln X_1/b_n\right)_+ - (m/n)\alpha^{-1}\right] \times \\
&\quad \frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t} \\
A_{n,3} &= -2\left\{(m/n)\left[\hat{\alpha}_m^{-1} - \alpha^{-1}\right] - \left[E\left(\ln X_t/b_n\right)_+ - (m/n)\alpha^{-1}\right]\right\} \times \\
&\quad \frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left(U_{n,s} - \alpha^{-1}U_{n,s}^*\right) \\
A_{n,4} &= -2(m/n)\left[\hat{\alpha}_m^{-1} - \alpha^{-1}\right]\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t} \times \\
&\quad \left[\left(\ln X_t/b_n\right)_+ - \left(\ln X_t/X_{(m+1)}\right)_+ - \alpha^{-1}U_{n,t}^*\right] \\
A_{n,5} &= -2\left[E\left(\ln X_1/b_n\right)_+ - (m/n)\alpha^{-1}\right]\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t} \times \\
&\quad \left\{\left(\ln X_t/b_n\right)_+ - \left(\ln X_t/X_{(m+1)}\right)_+ - \alpha^{-1}U_{n,t}^*\right\} \\
A_{n,6} &= -2\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left(\left\{U_{n,s} - \alpha^{-1}U_{n,s}^*\right\}\right) \times \\
&\quad \left\{\left(\ln X_t/b_n\right)_+ - \left(\ln X_t/X_{(m+1)}\right)_+ - \alpha^{-1}U_{n,t}^*\right\} \\
A_{n,7} &= -2\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left(\left\{U_{n,s} - \alpha^{-1}U_{n,s}^*\right\}\right) \times \left[\ln X_{(m+1)}/b_n - \alpha^{-1}U_{n,t}^*\right] \\
A_{n,8} &= -2\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left(\left\{U_{n,s} - \alpha^{-1}U_{n,s}^*\right\}\right) \times \\
&\quad \left\{\left(\ln X_t/b_n\right)_+ - \left(\ln X_t/X_{(m+1)}\right)_+ - \ln X_{(m+1)}/b_n\right\}
\end{aligned}$$

$$
\begin{aligned}
A_{n,9} &= 2\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t} \times \\
&\quad \left\{\left[(\ln X_s/b_n)_+ - \left(\ln X_s/X_{(m+1)}\right)_+\right] - \left(\ln X_{(m+1)}/b_n\right)\right\} \times \\
&\quad \left[\ln X_{(m+1)}/b_n - \alpha^{-1}U^*_{n,t}\right] \\
A_{n,10} &= 2\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left[(\ln X_s/b_n)_+ - \left(\ln X_s/X_{(m+1)}\right)_+ - \left(\ln X_{(m+1)}/b_n\right)\right] \times \\
&\quad \left\{\left[(\ln X_t/b_n)_+ - \left(\ln X_t/X_{(m+1)}\right)_+\right] - \left(\ln X_{(m+1)}/b_n\right)\right\} \\
A_{n,11} &= 2\frac{1}{m}\sum\nolimits_{s=1}^{n}\sum\nolimits_{t=1}^{n} w_{n,s,t}\left[\ln X_{(m+1)}/b_n - \alpha^{-1}U^*_{n,s}\right] \times \left[\ln X_{(m+1)}/b_n - \alpha^{-1}U^*_{n,t}\right]
\end{aligned}
$$

We assume $m^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n}|w_{n,s,t}| = O(1)$, and Theorem 2 implies $\hat{\alpha}_m^{-1} - \alpha^{-1} = o_p(1)$. Hence, $A_{n,1} = o_p(1)$. Moreover, Arguments in Hsing (1991: p. 1554), Lemma 13 and (3) can be used to show $(n/m)E(\ln X_t/b_n)_+ - \alpha^{-1} = o_p(1)$ for any t, hence $A_{n,2} = o_p(1)$.

Furthermore, by Lemma 1, $\|U_{n,t}\|_r$ and $\|U^*_{n,t}\|_r$ are $O((m/n)^{1/r})$; by Theorem 2, $m^{-1}\sum_{t=1}^{n}U^*_{n,t} = o_p(1)$; Lemma 1 implies $||U_{n,t} - \alpha^{-1}U^*_{n,t}||_1 = O(m/n)$; and Lemma 14 implies $\ln X_{(m+1)}/b_n = o_p(1/\sqrt{m})$ and $(\ln X_t/b_n)_+ - \left(\ln X_t/X_{(m+1)}\right)_+ - (\ln X_{(m+1)}/b_n) = o_p(1/\sqrt{m})$ for any t. Hence, $A_{n,i} = o_p(1)$, $i = 3...5$.

Similarly, $||(\ln X_s/b_n)_+ - \left(\ln X_s/X_{(m+1)}\right)_+||_2$ and $\left\|\ln X_{(m+1)}/b_n\right\|_2$ are $O(1/\sqrt{m})$ by Lemma 14 and functional invariance of probability limits. Invoking Markov's inequality and recalling $||U_{n,t} - \alpha^{-1}U^*_{n,t}||_1$ is $O(m/n)$ gives $A_{n,i} = o_p(1)$, $i = 6...11$.

Step 3 ($\sigma_m^2(1,-1)/\sigma_m^2 \to 1$): From the line of proof of Theorem 4, cf. (48), $\sigma_m^2(1,-1)/\sigma_m^2 \to 1$ follows immediately. ■

Proof of Lemma 6. The proof is nearly identical to a conventional proof that an NED process is a mixingale (e.g. Davidson, 1994: Theorem 17.5). For brevity, therefore, we compress the proof. Let Assumption C.4 hold. Define $k = [q/2]$, and write $E_b^a[\cdot] = E[\cdot|\mathcal{F}_b^a]$ with $\mathcal{F}_a^b = \sigma(\epsilon_a, .., \epsilon_b)$. Notice by the Minkowski inequality

$$\left\|E_{-\infty}^{t-q}\left[U^*_{n,t}\right]\right\|_p \leq \left\|E_{-\infty}^{t-q}\left(U^*_{n,t} - E_{t-k}^{t+k}[U^*_{n,t}]\right)\right\|_p + \left\|E_{-\infty}^{t-q}(E_{t-k}^{t+k}[U^*_{n,t}])\right\|_p. \tag{114}$$

For the first term on the right-hand-side of (114), we have by Jensen's inequality and the law of iterated expectations

$$
\begin{aligned}
&\left\|E_{-\infty}^{t-q}\left(U^*_{n,t} - E_{t-k}^{t+k}[U^*_{n,t}]\right)\right\|_p \leq \left\|U^*_{n,t} - E_{t-k}^{t+k}[U^*_{n,t}]\right\|_p \qquad (115)\\
= &\left\|I\left(X_t > b_n e^u\right) - P\left(X_t > b_n e^u|\mathcal{F}_{t-k}^{t+k}\right)\right\|_p,
\end{aligned}
$$

hence, cf. Assumption C.4,

$$\limsup_{n\to\infty}\left(\frac{n}{m}\right)^{1/p}\left\|E_{-\infty}^{t-q}\left(U^*_{n,t} - E_{t-k}^{t+k}[U^*_{n,t}]\right)\right\|_p \leq e_t^*(u)\upsilon_k^*. \tag{116}$$

For the second term in (114), $E_{t-k}^{t+k}[U^*_{n,t}]$ is uniform mixing of size $-\lambda_0$, hence by Serfling (1968: Theorem 2.2), Jensen's inequality and the law of iterated expectations,

$$\left\|E_{-\infty}^{t-q}(E_{t-k}^{t+k}[U_{n,t}^*])\right\|_p \leq 2\phi_k^{1-1/p}\left\|E_{t-k}^{t+k}[U_{n,t}^*]\right\|_p \leq 2\phi_k^{1-1/p}\left\|U_{n,t}^*\right\|_p \tag{117}$$

where $(n/m)^{1/p}\left\|U_{n,t}^*\right\|_p < A_p < \infty$ for some finite A_p, cf. Lemma 1, and $\phi_k^{1-1/p}$ $= O(q^{-\lambda_0(1-1/p)})$ by the uniform mixing property. Together, (116)-(117) imply

$$\begin{aligned}
&\limsup_{n\to\infty}\left(\frac{n}{m}\right)^{1/p}\left\|E_{-\infty}^{t-q}\left[U_{n,t}^*\right]\right\|_p \leq e_t^*(u)\upsilon_k^* + 2A_p\phi_k^{1-1/p} \qquad (118)\\
\leq\quad & 2\times\max\{e_t^*(u), 2A_p\}\max\{\upsilon_k^*, \phi_k^{1-1/p}\},
\end{aligned}$$

where $\max\{\upsilon_k^*, \phi_k^{1-1/p}\}$ has size $-\min\{\lambda_1, \lambda_0^{1-1/p}\}$. ■

Proof of Theorem 7. We only prove the claim for E-NED processes X. The proof in the E-MIX case is nearly identical

The proof for $U_{n,t}^*$ follows by definition and the boundedness property established in Lemma 1. In particular, under Assumption C.4, as $n \to \infty$

$$\begin{aligned}
&\left\|P(X_t > x_n(u)|\Im_{t-q}^{t+q}) - P\left(X_t > x_n(u)|\digamma_{t-q}^{t+q}\right)\right\|_2 \qquad (119)\\
&\leq \tilde{e}_t^*(u)(m/n)^{1/2}\tilde{\upsilon}_q^*,
\end{aligned}$$

and from (17), (18) and the definition of $U_{n,t}^*$,

$$\begin{aligned}
&\left\|U_{n,t}^* - E[U_{n,t}^*|\digamma_{t-q}^{t+q}]\right\|_2 \qquad (120)\\
=\quad & \left\|I\left(X_t > x_n(u)\right) - E[I\left(X_t > x_n(u)\right)|\digamma_{t-q}^{t+q}]\right\|_2\\
=\quad & \left\|P(X_t > x_n(u)|\Im_{t-q}^{t+q}) - P\left(X_t > x_n(u)|\digamma_{t-q}^{t+q}\right)\right\|_2. \leq d_{n,t}^*\psi_q^*.
\end{aligned}$$

Therefore $U_{n,t}^*$ is L_2-NED with coefficients $\psi_q^* = \tilde{\upsilon}_q^*$ and constants $d_{n,t}^* =$ $\tilde{e}_t^*(u)(m/n)^{1/2}$.

For $U_{n,t}$, we have

$$\begin{aligned}
&\left\|U_{n,t} - E[U_{n,t}|\digamma_{t-q}^{t+q}]\right\|_2 \qquad (121)\\
=\quad & \left\|(\ln X_t - \ln b_n)_+ - E[(\ln X_t - \ln b_n)_+|\digamma_{t-q}^{t+q}]\right\|_2\\
=\quad & \left\|(\ln X_t - \ln b_n)_+ - \int_0^\infty P\left((\ln X_t - \ln b_n) > u|\digamma_{t-q}^{t+q}\right)du\right\|_2\\
=\quad & \left\|\int_0^\infty \left(I(X_t > b_n e^u) - P\left(X_t > b_n e^u|\digamma_{t-q}^{t+q}\right)\right)du\right\|_2.
\end{aligned}$$

In order to evaluate the second moment of the integral, define

$$g_{n,i}(u_i) \equiv I(X_t > b_n e^{u_i}) - P\left(X_t > b_n e^{u_i}|\digamma_{t-q}^{t+q}\right),\quad i = 1, 2, \tag{122}$$

and observe that

$$
\begin{aligned}
& E\left(\int_0^\infty \left(I(X_t > b_n e^u) - P\left(X_t > b_n e^u | \mathcal{F}_{t-q}^{t+q}\right)\right) du\right)^2 \qquad (123) \\
= \; & E\left(\int_0^\infty g_{n,1}(u_1) du_1 \int_0^\infty g_{n,2}(u_2) du_2\right) \\
= \; & \int_0^\infty \int_0^\infty E\left[g_{n,1}(u_1) g_{n,2}(u_2)\right] du_1 du_2,
\end{aligned}
$$

where the last line follows from the Fubini theorem. Using the Cauchy-Schwartz inequality and properties of the Lebesgue integral, we deduce

$$
\begin{aligned}
& E\left(\int_0^\infty \left(I(X_t > b_n e^u) - P\left(X_t > b_n e^u | \mathcal{F}_{t-q}^{t+q}\right)\right) du\right)^2 \qquad (124) \\
= \; & \int_0^\infty \int_0^\infty E\left[g_{n,1}(u_1) g_{n,2}(u_2)\right] du_1 du_2 \\
\leq \; & \int_0^\infty \int_0^\infty \left\|g_{n,1}(u_1)\right\|_2 \left\|g_{n,2}(u_2)\right\|_2 du_1 du_2 \\
= \; & \left(\int_0^\infty \left\|g_{n,1}(u_1)\right\|_2 du_1\right)^2 \\
= \; & \left(\int_0^\infty \left\|I(X_t > b_n e^u) - P\left(X_t > b_n e^u | \mathcal{F}_{t-q}^{t+q}\right)\right\|_2 du\right)^2.
\end{aligned}
$$

Therefore, by Fatou's Lemma and Assumption C.4,

$$
\begin{aligned}
& \limsup_{n\to\infty} \sqrt{\frac{n}{m}} \left\|U_{n,t} - E[U_{n,t} | \mathcal{F}_{t-q}^{t+q}]\right\|_2 \qquad (125) \\
\leq \; & \limsup_{n\to\infty} \sqrt{\frac{n}{m}} \int_0^\infty \left\|I(X_t > b_n e^u) - P\left(X_t > b_n e^u | \mathcal{F}_{t-q}^{t+q}\right)\right\|_2 du \\
\leq \; & \int_0^\infty \limsup_{n\to\infty} \sqrt{\frac{n}{m}} \left\|I(X_t > b_n e^u) - P\left(X_t > b_n e^u | \mathcal{F}_{t-q}^{t+q}\right)\right\|_2 du \\
\leq \; & \left(\int_0^\infty e_t^*(u) du\right) \nu_q^*.
\end{aligned}
$$

Therefore, as $n \to \infty$, $U_{n,t}$ is L_2-NED on $\{\mathcal{F}_t\}$, with constants $d_{n,t} = (m/n)^{1/2} \int_0^\infty e_t^*(u) du$ and coefficients $\psi_q = \nu_q^*$. ■

Proof of Lemma 8. By iterated expectations and the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
& \frac{n}{m} E\left(I(X_t > x_n(u)) - P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\right)^2 \qquad (126) \\
= & \frac{n}{m} E\left(E\left[\left(I(X_t > x_n(u)) - P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\right)^2 | \mathcal{F}_{t-q}^{t+q}\right]\right) \\
= & \frac{n}{m} E\left(P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right) - P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)^2\right) \\
= & \frac{n}{m} E\left[P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\left(1 - P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\right)\right] \\
\leq & \frac{n}{m} \left\|P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\right\|_2 \left\|\left(1 - P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\right)\right\|_2 \\
\leq & \frac{n}{m} \left\|P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right)\right\|_2 .
\end{aligned}$$

Now, because Z_t is *iid* and $\mathcal{F}_{t-q}^{t+q} = \sigma(Z_s : t - q \leq s \leq t + q)$, $P(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q})$ can be written as

$$\begin{aligned}
P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right) &= P\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i} > b_n(m) e^u | \mathcal{F}_{t-q}^{t+q}\right) \qquad (127) \\
&= P\left(\sum_{i=q+1}^{\infty} \psi_i Z_{t-i} > b_n(m) e^u - a_{t,q}\right)
\end{aligned}$$

for some number $a_{t,q}$ satisfying $P(\sum_{i=}^{q} \psi_i Z_{t-i} = a_{t,q} | \mathcal{F}_{t-q}^{t+q}) = 1$. Using (1) and (2), and the fact that Z_t is *iid*, as $n \to \infty$

$$\begin{aligned}
& P\left(X_t > x_n(u) | \mathcal{F}_{t-q}^{t+q}\right) \qquad (128) \\
= & P\left(\sum_{i=q+1}^{\infty} \psi_i Z_{t-i} > b_n(m) e^u - a_{t,q}\right) \\
= & \sum_{i=q+1}^{\infty} |\psi_i|^{\alpha} P\left(Z_{t-i} > b_n(m) e^u - a_{t,q}\right) \\
= & \sum_{i=q+1}^{\infty} |\psi_i|^{\alpha} P\left(Z_{t-i} > b_n(m) e^u (1 - a_{t,q}/[b_n(m) e^u])\right) \\
= & \sum_{i=q+1}^{\infty} |\psi_i|^{\alpha} P\left(Z_{t-i} > b_n(m)\right) \frac{P\left(Z_{t-i} > b_n(m) e^u (1 - a_{t,q}/[b_n(m) e^u])\right)}{P\left(Z_{t-i} > b_n(m)\right)} \\
\approx & \frac{m}{n} \frac{\sum_{i=q+1}^{\infty} |\psi_i|^{\alpha}}{\sum_{i=0}^{\infty} |\psi_i|^{\alpha}} e^{-\alpha u} \left(1 - a_{t,q}/[b_n(m) e^u]\right)^{-\alpha},
\end{aligned}$$

where

$$P\left(Z_{t-i} > b_n(m)\right) \approx \frac{m/n}{\sum_{i=0}^{\infty} |\psi_i|^{\alpha}} \qquad (129)$$

as $n \to \infty$ follows from

$$\begin{aligned}
m/n &\approx P\left(X_t > b_n(m)\right) = P\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i} > b_n(m)\right) \qquad (130) \\
&= \sum_{i=0}^{\infty} |\psi_i|^{\alpha} P\left(Z_{t-i} > b_n(m)\right).
\end{aligned}$$

Hence, as $n \to \infty$

$$\begin{aligned}
& \frac{n}{m} E\left(I(X_t > x_n(u)) - P\left(X_t > x_n(u) | F_{t-q}^{t+q}\right)\right)^2 && (131) \\
\leq \quad & \frac{n}{m} \left\| P\left(X_t > x_n(u) | F_{t-q}^{t+q}\right)\right\|_2 \\
\approx \quad & \frac{\sum_{i=q+1}^{\infty} |\psi_i|^{\alpha}}{\sum_{i=0}^{\infty} |\psi_i|^{\alpha}} e^{-\alpha u} \left\| (1 - a_{t,q}/b_n(m))^{-\alpha} \right\|_2 \\
\to \quad & \frac{\sum_{i=q+1}^{\infty} |\psi_i|^{\alpha}}{\sum_{i=0}^{\infty} |\psi_i|^{\alpha}} e^{-\alpha u} = \tilde{e}_t^*(u)^2 \tilde{v}_q^{*2},
\end{aligned}$$

where

$$\tilde{e}_t^*(u) = \frac{e^{-\alpha u/2}}{\sqrt{\sum_{i=0}^{\infty} |\psi_i|^{\alpha}}}, \quad \tilde{v}_q^{*2} = \sqrt{\sum_{i=q+1}^{\infty} |\psi_i|^{\alpha}}, \tag{132}$$

and $\tilde{e}_t^* : \mathbb{R} \to \mathbb{R}_+$ is integrable on $\mathbb{R}_+$: $\int_0^{\infty} e^{-\alpha u/2} du = 2/\alpha$.

Finally, if $\psi_i = O(i^{-\mu})$, $\mu > 1/\alpha$, then $\sum_{i=q+1}^{\infty} |\psi_i|^{\alpha} = O(q^{1-\alpha\mu})$ and $\tilde{v}_q = O(q^{(1-\alpha\mu)/2})$. $\blacksquare$

Proof of Theorem 9. For brevity, we prove only (25). Recall $\hat{c}_m = (m/n) X_{(m+1)}^{\hat{\alpha}_m}$, define $\check{c} \equiv (m/n) X_{(m+1)}^{\alpha}$, and write

$$\begin{aligned}
\ln \hat{c}_m &= \ln(m/n) + \hat{\alpha}_m \ln X_{(m+1)} && (133) \\
&= \ln(m/n) + (\hat{\alpha}_m - \alpha) \ln X_{(m+1)} + \alpha \ln X_{(m+1)} \\
&= \ln \check{c} + (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \ln X_{(m+1)}^{\alpha} \\
&= \ln \check{c} + (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \ln \check{c} + \ln(n/m)\, (\hat{\alpha}_m - \alpha)\, \alpha^{-1}.
\end{aligned}$$

From Lemma 12, under Assumptions A.2 and B.2, and provided $X_{(m+1)}/b_n(m) = 1 + o_p(n^{-\xi})$, $\xi > (1-\delta)\theta/\alpha$, we deduce $\ln \check{c} = \ln c + o_p(1/\sqrt{m})$. From Theorem 4, under Assumptions A.2, B.2 and C.2, $\sqrt{m}(\hat{\alpha}_m - \alpha) \Longrightarrow N(0, \alpha^2[1 + \chi + \lambda - 2\zeta])$, implying

$$\sqrt{m} [\ln(n/m)]^{-1} (\hat{\alpha}_m - \alpha) = o_p(1). \tag{134}$$

Hence, by Theorem 4 and Cramér's Theorem

$$\begin{aligned}
\ln \hat{c}_m &= \ln \check{c} + (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \ln \check{c} && (135) \\
&\quad + \ln(n/m)\, (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \\
\frac{\sqrt{m}}{\ln(n/m)} \ln \hat{c}_m / c &= \frac{\sqrt{m}}{\ln(n/m)} \ln \check{c}/c + \frac{\sqrt{m}}{\ln(n/m)} (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \ln \check{c} \\
&\quad + \sqrt{m}\, (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \\
\frac{\sqrt{m}}{\ln(n/m)} \ln \hat{c}_m / c &= o_p(1/\ln(n/m)) + o_p(1) + \sqrt{m}\, (\hat{\alpha}_m - \alpha)\, \alpha^{-1} \\
&\Longrightarrow N(0, \tilde{\sigma}^2),
\end{aligned}$$

where $\tilde{\sigma}^2 = \lim_{n\to\infty} \tilde{\sigma}_m^2$. Therefore, by the mean-value-theorem, for some $c_* \in (\hat{c}_m, c)$

$$\frac{\sqrt{m}}{\ln(n/m)} \ln \hat{c}_m/c = \frac{\sqrt{m}}{\ln(n/m)} \frac{1}{c_*} (\hat{c}_m - c) \Longrightarrow N(0, \tilde{\sigma}^2). \tag{136}$$

Because $c_* \in (\hat{c}_m, c) \to (c, c)$, the proof is complete upon application of Cramér's theorem. ■

References

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

[2] Bidarkota, P. V., and J. H. McCulloch, 1998, Optimal Univariate Inflation Forecasting with Symmetric Stable Shocks, Journal of Applied Econometrics **13**, 659-670.

[3] Bidarkota, P. V., and J. H. McCulloch (2004) Testing for Persistence in Stock Returns with GARCH-Stable Shocks, Quantitative Finance **4**, 256-265.

[4] Billingsley, P. (1995) Probability and Measure. John Wiley & Sons: New York.

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

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

[7] Chambers, J. M., C. L. Mallows and B.W. Stuck (1976) A Method for Simulating Stable Random Variable, Journal of the American Statistical Association **71**, 340-344.

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

[9] Chandra, K.C. (1974) Strong Mixing Properties of Linear Stochastic Processes, Journal of Applied Probability **11**, 401-408.

[10] Chen, X., and H. White (1997) Central Limit and Functional Central Limit Theorems for Hilbert-Valued Dependent Heterogeneous Arrays with Applications, University of California-San Diego, Dept. of Economics Discussion Paper 92-35R.

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

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

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

[14] Davidson, J. (1992) A Central Limit Theorem for Globally Nonstationary Near-Epoch-Dependent Functions of Mixing Processes, Econometric Theory **8**, 313-329.

[15] Davidson, J. (1993a) An L_1-Convergence Theorem for Heterogeneous Mixing Arrays with Trending Moments, Statistics and Probability Letters **16**, 301-304.

[16] Davidson, J. (1993b) The Central Limit Theorem for Globally Non-Stationary Near-Epoch-Dependent Functions of Mixing Processes: The Asymptotically Degenerate Case, Econometric Theory **9**, 402-412.

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

[18] 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.

[19] de Haan, L., Drees, H. and S. Resnick, How to Make a Hill Plot (2000) Annals of Statistics **28**, 254-274.

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

[21] de Jong, R. M., and J. Davidson (2000) Central Limit Theorems for Dependent Heterogenous Random Variables, Econometric Theory **13**, 353-367.

[22] Drees, H., L. de Haan and S. Resnick (2000) How to Make a Hill Plot, Annals of Statistics **28**, 254-274.

[23] Dumouchel, W. H. (1983) Estimating the Stable Index α in order to Measure Tail Thickness, Annals of Statistics **11** 1019-1036.

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

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

[26] Hall, Peter, and C. C. Hyde (1980) Martingale Limit Theory and Its Applications. New York: Academic Press.

[27] Hall, Peter (1982) On Some Estimates of an Exponent of Regular Variation, Journal of the Royal Statistical Society **44**, 37-42.

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

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

[30] Hsing, Tailen (1991) On Tail Index Estimation Using Dependent Data, Annals of Statistics **19**, 1547-1569.

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

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

[33] Lin, Z. and J. Qui (2004) A Central Limit Theorem for Strong Near-Epoch-Dependent Random Variables, Chinese Annals of Mathematics **25B**, 263-274.

[34] Loretan, M. (1991) Testing Covariance Stationarity of Heavy-Tailed Economic Time Series, Unpublished Ph.D. Dissertation, Dept. of Economics, Yale University.

[35] Loretan, M., and P.C.B. Phillips (1991) The Durbin-Watson Ratio under Infinite-Variance Errors, Journal of Econometrics **47**, 85-114 .

[36] Loretan, M., and P.C.B. Phillips (1994) Testing the Covariance Stationarity of Heavy-Tailed Economic Time Series: An Overview of the Theory with Applications to Financial Data Sets, Journal of Empirical Finance **1**, 211-48.

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

[38] McCulloch, J. Houston (1996) Financial Applications of Stable Distributions. In: G.S. Maddala and C.R. Rao, eds., Statistical Methods in Finance, Handbook of Statistics, Volume 14, 393-425. Elsevier: New York.

[39] McCulloch, J. Houston (1997) Measuring Tail Thickness to Estimate the Stable Index α: A Critique, Journal of Business and Economic Statistics **15**, 74-81.

[40] Mittnick, S and S.T. Rachev (1993) Modeling Assets Returns with Alternative Stable Distributions, Economic Reviews **12**, 261-330.

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

[42] Phillips, P.C.B. (1990) Time Series Regression with a Unit Root and Infinite Variance Errors, Econometric Theory **6**, 44-62.

[43] Pötscher, B. M., and I. R. Prucha (1991) Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts, Econometric Reviews **10**, 125-216.

[44] Quintos, C., Z. Fan, and P.C.B. Phillips (2001) Structural Change Tests in Tail Behavior and the Asian Crisis, Review of Economics Studies **68**, 633-663.

[45] Quintos, C. (2004) Extremal Correlation for GARCH Data, manuscript, Dept. of Applied Statistics, University of Rochester.

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

[47] Resnick, S. (1996) Why Non-Linearities Can Ruin the Heavy Tailed Modeler's Day, mimeo, Dept. Statistics, Cornel University.

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

[49] Resnick, S., and C. Stărică (1996) Behavior of Hill's Estimator for Autoregressive Data, mimeo, School of ORIE, Cornell University.

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

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

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

[53] Rootzen, H., Leadbetter, M., and L. de Haan (1990) Tail and Quantile Estimation for Strongly Mixing Stationary Sequences, Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill.

[54] Samorodnitsky, G. and M.S. Taqqu (1994) Stable Non-Gaussian Random Processes. Chapman and Hall: New York.

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

[56] Silverman, B. W. (1986) Density Estimation for Statistics and Data Analysis. Chapman and Hall: London.

[57] Stărică, C. (1999) Multivariate Extremes for Models with Constant Conditional Correlations, Journal of Empirical Finance **6**, 515-553.

[58] Wooldridge, J. and H. White (1988) Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes, Econometric Theory **4**, 210-230.