A New Moment Bound and Weak Laws for Mixingale Arrays without Memory or Heterogeneity Restrictions, with Applications to Tail Trimmed Arrays

Jonathan B. Hill*
Dept. of Economics
University of North Carolina - Chapel Hill

Dec. 2010

Abstract

We present a new weak law of large numbers for dependent heterogeneous triangular arrays $\{y_{n,t}\}$ with applications to tail trimming. The law follows from a new partial sum moment bound $E|\sum_{t=1}^{n} y_{n,t}|^p \leq K \sum_{t=1}^{n} E|y_{n,t}|^p$ for zero mean L_p-mixingale arrays $\{y_{n,t}, \Im_t\}$, $p \in [1, 2]$, where $\Im_t$ is a σ-field. We do not require uniform integrability, nor restrict mixingale dependence or heterogeneity. The weak law is applied to tail trimmed heavy tailed data, and we characterize self-scaled laws $\sum_{t=1}^{n} |y_{n,t}| / \sum_{t=1}^{n} E|y_{n,t}| \xrightarrow{p} 1$ for potentially very heavy tailed data ($E|y_{n,t}| \to \infty$ is possible). Finally, we characterize a minimal rate of convergence for the tail trimmed self-scaled law when distribution tails are regularly varying and possibly non-stationary, including trending tails.

1. Introduction We present a new partial sum moment bound and weak laws of large numbers [WLLN] for zero mean mixingale triangular arrays $\{y_{n,t}\} = \{y_{n,t} : 1 \leq t \leq n\}_{n \geq 1}$. The array $\{y_{n,t}\}$ may exhibit any rate of memory decay and any degree of heterogeneity. The primary application is to tail trimmed arrays in Sections 3 and 4.

Let $\{\Im_t\}$ be a sequence of non-decreasing σ-fields on a probability space $(\Omega, \Im, P)$, denote the L_p-norm $||x||_p := (E|x|^p)^{1/p}$, and let $K > 0$ be a finite constant that may change from place to place. We rule out degenerate cases by assuming

$$\liminf_{n \to \infty} \sum_{t=1}^{n} E|y_{n,t}|^p \geq K \text{ for some } K > 0. \tag{1}$$

We say $\{y_{n,t}, \Im_t\}$ forms a zero mean L_p-mixingale triangular array, $p > 1$, with *size* $\lambda > 0$ if for some *base field* $\Im_t$

$$\|E[y_{n,t}|\Im_{t-q_n}]\|_p \leq e_{n,t}\zeta_{q_n} \quad \text{and} \quad \|y_{n,t} - E[y_{n,t}|\Im_{t+q_n}]\|_p \leq e_{n,t}\zeta_{q_n+1}, \tag{2}$$

where $e_{n,t} > 0$ and $\zeta_{q_n} = O(q_n^{-\lambda-\iota})$ for infinitessimal $\iota > 0$. We call $\{e_{n,t}\}$ the *constants*; $\{\zeta_{q_n}\} = \{\zeta_{q_n}\}_{n \geq 1}$ are the *coefficients* where $\zeta_0 = 1$; and $\{q_n\} = \{q_n\}_{n \geq 1}$ is any sequence

*Dept. of Economics, University of North Carolina - Chapel Hill; jbhill@email.unc.edu; www.unc.edu/∼jbhill.

AMS classifications: primary 60F15; secondary 60F25.

Keywords and phrases: moment bound, law of large numbers, mixingale array, tail trimming, heavy tails.

of positive finite integer *displacements*. Notice we allow the displacement $q_n \to \infty$ as $n \to \infty$, a central tool in this paper detailed extensively below.

McLeish [53] first proposed the L_2-mixingale property for sequences $\{y_t\}$, assumed here to have a zero mean,

$$\|E[y_t|\Im_{t-q}]\|_2 \leq e_t \vartheta_q \text{ and } \|y_t - E[y_t|\Im_{t+q}]\|_2 \leq e_t \vartheta_{q+1}, \tag{$2'$}$$

where q is a fixed integer displacement, $e_t > 0$ and $\vartheta_q = O(q^{-\lambda-\iota})$, and used ($2'$) to deliver a maximal inequality and strong law. Hansen [37], [38] extended that result to L_p-mixingale sequences, $p \in (1,2)$, with generalizations to weak and strong laws for arrays in Andrews [1], Davidson [14], de Jong [19], [20], Hill [39], [40], and Yanjiao and Zhengyan [71], and further generalizations to so-called s-weak dependence in Dedecker and Doukhan [23]. In all cases a fixed displacement q is used to deliver a maximal inequality. The combined mixingale cases imply

$$E\left[\max_{1\leq j\leq n}\left|\sum_{t=1}^{j} y_{n,t}\right|^p\right] \leq K\sum_{t=1}^{n} e_{n,t}^p$$

for $p \in (1,2)$ with size $\lambda = 1$ ([37], [38], [71]) or $p = 2$ with size $\lambda = 1/2$ ([53]). Similar bounds are achieved in Davydov [18], Rio [61], [62], Dedecker and Doukhan [23], and are widely used for central limit theory ([21], [39], [40]) and HAC matrix estimation ([17], [40]).

The large discontinuity in size requirements $\lambda = 1$ if $p \in (1,2)$ and $\lambda = 1/2$ if $p = 2$ arises from the architecture of L_p-spaces. See arguments in McLeish [53] and Hansen [37], [38].

An adapted martingale difference is trivially a mixingale, and von Bahr and Esséen [68] prove L_p-bounded martingale difference sequences $\{y_t, \Im_t\}$, satisfy

$$E\left|\sum_{t=1}^{n} y_t\right|^p \leq K\sum_{t=1}^{n} E|y_t|^p \text{ for } p \in [1,2]. \tag{3}$$

Hitczenko [43], [44] generalizes Rosenthal [64] and Burkholder [9] inequalities for iid and martingale difference processes. First, for any arbitrary $\Im_t$-adapted nonnegative random variables y_t that are L_p-bounded, and some finite $K_p > 0$:

$$E\left(\sum_{t=1}^{n} y_t\right)^p \leq K_p \max\left\{\sum_{t=1}^{n} E[y_t^p],\ E\left(\sum_{t=1}^{n} E[y_t|\Im_{t-1}]\right)^p\right\} \text{ for } p > 1. \tag{4}$$

Similarly, for a martingale difference sequence $\{y_t, \Im_t\}$ of L_p-bounded y_t:

$$E\left|\sum_{t=1}^{n} y_t\right|^p \leq K_p \max\left\{\sum_{t=1}^{n} E|y_t|^p,\ E\left(\sum_{t=1}^{n} E\left[y_t^2|\Im_{t-1}\right]\right)^{p/2}\right\} \text{ for } p > 2. \tag{5}$$

See de la Peña et al [22] for a sharp characterization of the scale K_p, cf. Ibragimov and Sharakhmetov [48], [49].

Although (2) is nearly identical to ($2'$), and mixingale arrays appear as early as Andrews [1], (2) is decidedly different due to the allowed dependence of q_n on n. Notice for any sequence of finite displacements $\{q_n\}$ that satisfy $q_n/\exp\{\max_{1\leq t\leq n}\{e_{nt}\}\} \to \infty$, and any tiny $0 < \delta < \iota$

$$\|E[y_{n,t}] - E[y_{n,t}|\Im_{t-q_n}]\|_p \leq \left\{\frac{e_{n,t}}{\ln(q_n)}\right\} \times \ln(q_n)\,\zeta_{q_n} \leq Kq_n^{-\lambda-\iota+\delta} \tag{6}$$

$$\|y_{n,t} - E[y_{n,t}|\Im_{t+q_n}]\|_p \leq \left\{\frac{e_{n,t}}{\ln(q_n)}\right\} \times \ln(q_n)\,\zeta_{q_n+1} \leq Kq_n^{-\lambda-\iota+\delta}.$$

By displacing fast enough as sample size $n \to \infty$ the measure of heterogeneity $e_{n,t}$ is irrelevant: only memory captured by $q_n^{-\lambda-\iota+\delta}$ matters. Ultimately this suggests an exploitable defect in the mixingale concept itself: separating heterogeneity $e_{n,t}$ from memory ζ_{q_n} is artificial and meaningless from the perspective of (6). Hill [41] apparently first used displacement sequences to deduce a property like (6) for tail trimmed Near Epoch Dependent arrays. We exploit the idea here to prove moment bound (3) under (2), and deliver a new WLLN.

The sequence $\{q_n\}$ is arbitrary and controlled by the analyst, hence we can always displace to the deep past or distant future faster than the accumulation of heterogeneous characteristics of $y_{n,t}$ measured by $e_{n,t}$. This key quality of mixingale (2), exhibited in (6), leads to bound (3) and therefore an improvement over Burkholder-Rosenthal type bounds and their recent advances (4) for a massive array of dependent processes. See de Jong [21], Hill [39], [40], [41], Leadbetter et al [51], and Rootzén [63] to name a few for usage of displacement sequences in various settings.

In this paper we exploit $q_n \to \infty$ as $n \to \infty$ to prove (3) in Theorem 2.1 for L_p-mixingale arrays $\{y_{n,t}, \Im_t\}$, $p \in [1,2]$, with any size $\lambda > 0$ and any degree of heterogeneity $e_{n,t}$. Bound (3) then promotes our main WLLN implied by Theorem 2.2. The bound is potentially far sharper than bounds implied by McLeish [53] and Hansen [37] since $E|y_{n,t}|^p$ may be much smaller than mixingale constants $e_{n,t}$. See Sections 2 and 5 for examples.

Further, at least for asymptotic arguments (3) renders bounds like (4) irrelevant since, up to a multiplicative scale, $\sum_{t=1}^{n} E|y_{n,t}|^p$ is trivially better. Whether our scale K is sharp is not considered here since this is irrelevant for a WLLN. See [22], [48], [49] and their references.

Our proof of (3) re-vitalizes a well known decomposition of $\sum_{t=1}^{n} y_{n,t}$ into an infinite series of a martingale difference. We use $q_n \to \infty$ as $n \to \infty$ to support a *finite series* approximation that eliminates mixingale size and heterogeneity restrictions. Consult Section 2 for an intuitive explanation behind the proof of (3) and why the use of fixed displacements q fails to promote (3).

Andrews [1] and Davidson [14] deliver laws of large numbers for uniformly integrable L_1-mixingale arrays $\{y_{n,t}, \Im_t\}$ of any size, provided heterogeneity is bounded: $\sum_{t=1}^{n} e_{n,t} = O(n)$. Yanjiao and Zhengyan [71] tackle non-uniformly integrable L_p-mixingales $\{y_t, \Im_t\}$ with size 1 and bounded heterogeneity: $\sum_{t=1}^{\infty} e_t^p / b_t^p < \infty$ for monotone $\{b_t\}$, $b_t \to \infty$ as $t \to \infty$. We allow any size λ, any degree of heterogeneity since we never bound $e_{n,t}$, and non-uniform integrability, and we do not require uniform L_p-boundedness for any $p > 0$. For example, $E|y_{n,t}|^p \to \infty$ is possible for small $0 < p \leq 1$ when $y_{n,t}$ is a tail trimmed version of heavy tailed random variable y_t.

Related inequalities for mixing sequences, martingales and sums of martingale differences date to Burkholder [9], Davydov [18], Doob [25], Ibragimov [46], Marcinkiewicz and Zygmund [52], Rosenthal [64], von Bahr and Esséen [68], and Yokoyama [72], with recent contributions for associated, mixing or weakly dependent sequences by Birkel [5], Boucheron et al [6], Dedecker and Doukhan [23], Doukhan and Louhichi [27], Doukhan and Neumann [28] and Rio [61], [62]. Subsequent improvements for martingales predominately concern the constant K_p in (4): see [22], [23], and [58].

Our method of proof evidently precludes a maximal inequality $E[\max_{1 \leq j \leq n} |\sum_{t=1}^{j} \{y_{n,t} - E[y_{n,t}]\}|^p] \leq K \sum_{t=1}^{n} E\,|y_{n,t}|^p$. We therefore do not provide a strong law of large numbers (e.g. [37], [53]). Nevertheless, we gain substantial generality for weak laws since $E|y_{n,t}|^p$ may be much smaller than $e_{n,t}^p$, while all mixingale inequalities require at least $\lambda \geq 1/2$ and/or summable powers of $e_{n,t}$, effectively restricting hyperbolic memory and/or non-stationarity.

Our WLLN covers martingale differences, α-, β-, ϕ- and ρ-mixing, and geometrically ergodic arrays, and Near Epoch Dependent and mixingale arrays, allowing stochastic trend, nonlinear distributed lags, AR-GARCH, and nonlinear GARCH. See Sections 5 and 6 for verification of the major assumptions and examples.

Laws of large numbers for dependent heterogeneous data are now well established. A small subset includes Andrews [1], Arcones [2], Chandra and Ghosal [10], Davidson [14], de Jong [19], [20], Doukhan and Wintenberger [29], Hansen [37], McLeish [53], Shao [65], and Teicher [67]. See de Jong [19] for what appears to be the lightest available restrictions on mixingale heterogeneity, to date. Marcinkiewicz-Zygmund moment bounds result in maximal inequalities for stationary uniformly integrable β-mixing sequences, from which strong laws are derived ([2], [24], [62]). Bounds are imposed that replicate mixingale coefficient summability and uniform integrability.

In Sections 3 and 4 we apply the main results to tail trimmed arrays $\{y_{n,t}\}$ of a possibly non-integrable sequences $\{y_t\}$. In the case of tail trimming $E|y_{n,t}|^s \to \infty$ as $n \to \infty$ is possible for some $s > 0$. We therefore characterize so-called *self-scaled* laws of the form

$$\frac{\sum_{t=1}^{n} |y_{n,t}|^s}{\sum_{t=1}^{n} E\,|y_{n,t}|^s} = 1 + o_p(1) \text{ for any } s > 0, \tag{7}$$

and explicitly characterize the rate of convergence in (7) when the untrimmed sequence $\{y_t\}$ has stationary or non-stationary regularly varying distribution tails. Scaling by $\sum_{t=1}^{n} E\,|y_{n,t}|^s$ implies the rate of convergence of $\sum_{t=1}^{n} |y_{n,t}|^s$ itself need not be known.

Limit theory for trimmed sums of iid data has a deep history, dating at least to Newcomb [56], cf. Stigler [66]. Very few extensions to non-iid data exist ([35], [69]), and apparently none allow general dependence, heterogeneity and heavy tails together (e.g. Csörgő et al [13], Griffin and Qazi [33], Hahn et al [36], Pruitt [59]).

Throughout K denotes a positive finite constant whose value may change from line to line. $\iota > 0$ is a tiny number that may change with the context. $\{a_n\} = \{a_n\}_{n \geq 1}$. $a_n \sim b_n$ signifies $a_n/b_n \to 1$. $\xrightarrow{p}$ denotes convergence in probability. $L(x)$ is a slowly varying [s.v.] function, the value or rate of which may change with the context[1]. $\epsilon_t \overset{iid}{\sim} (0,1)$ implies ϵ_t is iid with zero mean and unit variance. Summations $\sum_{a}^{b} = 0$ if $b < a$.

2. Weak Laws of Large Numbers

In this section we develop the moment bound and WLLN for a generic L_p-bounded mixingale array $\{y_{n,t}, \Im_t\}$.

Assumption 1. *The random variables $y_{n,t}$ are zero mean L_p-bounded $p \in [1,2]$ for each $1 \leq t \leq n$ and $n \geq 1$, such that non-degeneracy (1) holds. Further, $\{y_{n,t}, \Im_t\}$ forms an L_p-mixingale array (2) with monotonic coefficients $\zeta_{q_n} \searrow 0$ as $q_n \to \infty$ of any size $\lambda > 0$.*

Remark 1: Monotonic decay does not reduce generality and merely sharpens notation in the proof of Theorem 2.1, below.

Remark 2: The mixingale base $\Im_t$ does not need to be explicitly defined here. It suffices merely to impose (2), where L_p-boundedness for $p \in [1,2]$ suffices for a version $E[y_{n,t}|\Im_s]$ to exist for any $s \in \mathbb{Z}$ by the Radon-Nikodym theorem.

Theorem 2.1 (Partial Sum Moment Bound). *Under Assumption 1 $E|\sum_{t=1}^{n} y_{n,t}|^p \leq K \sum_{t=1}^{n} E|y_{n,t}|^p$.*

Apply Markov's inequality and Theorem 2.1 to deduce $P(|\sum_{t=1}^{n} y_{n,t}| > \varepsilon) = O(\sum_{t=1}^{n} E\,|y_{n,t}|^p)$ for any $\varepsilon > 0$. This leads to the second main result of this paper.

[1] Recall $L(\lambda x)/L(x) \to 1$ as $x \to \infty$, $\forall \lambda > 0$, where products and powers of $L(x)$ are slowly varying, cf Resnick [59].

Theorem 2.2 (Partial Sum Probability Bound). *Under Assumption 1*

$$\frac{1}{\left(\sum_{t=1}^{n} E|y_{n,t}|^{p}\right)^{1/p}} \sum_{t=1}^{n} y_{n,t} = O_{p}\left(1\right).$$

Remark 1: Although Assumption 1 presumes L_p-boundedness for each t and n, Theorem 2.2 does not require uniform integrability. In particular $E|y_{n,t}|^p \to \infty$ as $n \to \infty$ is allowed.

Remark 2: Theorem 2.2 implies a WLLN: for any $\mathcal{L}(n) \to \infty$ as $n \to \infty$

$$\frac{1}{\left(\sum_{t=1}^{n} E\left|y_{n,t}\right|^{p}\right)^{1/p} \mathcal{L}(n)} \sum_{t=1}^{n} y_{n,t} \xrightarrow{p} 0.$$

2.1 Intuition Behind Theorem 2.1

In order to understand how displacements $q_n \to \infty$ as $n \to \infty$ can alleviate size and heterogeneity restrictions, recall a standard martingale difference decomposition for L_p-mixingales $p \in (1, 2]$ (e.g. [37], [38], [53], [54]):

$$\sum_{t=1}^{n} y_{n,t} = \sum_{q=-\infty}^{\infty} \sum_{t=1}^{n} \{E[y_{n,t}|\Im_{t+q}] - E[y_{n,t}|\Im_{t+q-1}]\} =: \sum_{q=-\infty}^{\infty} \mathcal{Y}_{n,q} \ a.s. \tag{8}$$

Since we are only interested in a partial sum moment bound, for simplicity assume here $\{y_{n,t}, \Im_t\}$ is an L_2-mixingale with size $\lambda = 1$. In the spirit of McLeish [53] and Hansen [37] observe

$$\begin{aligned} \left\| \sum_{t=1}^{n} y_{n,t} \right\|_2 &\leq \sum_{q=-\infty}^{\infty} \|\mathcal{Y}_{n,q}\|_2 \\ &\leq \sum_{q=-\infty}^{\infty} \left(\sum_{t=1}^{n} E\left(E[y_{n,t}|\Im_{t+q}] - E[y_{n,t}|\Im_{t+q-1}]\right)^2 \right)^{1/2} \\ &\leq \sum_{q=-\infty}^{\infty} \left(\sum_{t=1}^{n} e_{n,t}^2 \zeta_q^2 \right)^{1/2} = \sum_{q=-\infty}^{\infty} \zeta_q \times \left(\sum_{t=1}^{n} e_{n,t}^2 \right)^{1/2}. \end{aligned} \tag{9}$$

The first inequality follows from (8) and Minkowski's inequality; the second from the martingale difference property of $E[y_{n,t}|\Im_{t+q}] - E[y_{n,t}|\Im_{t+q-1}]$; and the third from the mixingale property (2). See Hansen [37], [38] for the L_p-mixingale case for $p \in (1, 2)$, and the proof of Theorem 2.1, below, for greater detail and generality. Clearly $E(\sum_{t=1}^{n} y_{n,t})^2 \leq K \sum_{t=1}^{n} e_{n,t}^2$ since size $\lambda = 1$ ensures coefficient summability $\sum_{q=-\infty}^{\infty} \zeta_q = K < \infty$.

We necessarily arrive at $E(\sum_{t=1}^{n} y_{n,t})^2 \leq K \sum_{t=1}^{n} e_{n,t}^2$ because the mixingale property is quantified over all integer displacements q. This is precisely what occurs at the third inequality: we are literally stuck with $\sum_{t=1}^{n} e_{n,t}^2$ because $E(\{E[y_{n,t}|\Im_{t+q}] - E[y_{n,t}|\Im_{t+q-1}]\})^2 \leq e_{n,t}^2 \zeta_q^2$ is evaluated at *every* t *and* q. Since decomposition (8) involves every integer lag $q \in \mathbb{Z}$ we are faced with the infinite series $\sum_{q=-\infty}^{\infty} \zeta_q$ in (9). The right-hand-side of the equality in (9) is finite only if the coefficients ζ_q have a size 1.

Although far more sophisticated arguments exist, in all cases a conclusion like (9) is reached: by decomposing $\sum_{t=1}^{n} y_{n,t}$ with $E[y_{n,t}|\Im_{t+q}] - E[y_{n,t}|\Im_{t+q-1}]$ for every $q \in \mathbb{Z}$ the final bound involves every $\{e_{n,t}\}_{t=1}^{n}$ and $\{\zeta_q\}_{q \in \mathbb{Z}}$. This classic approach implies a very strong cost to pay: memory must be restricted and the bound is in terms of unknown $e_{n,t}$.

We solve the size restriction necessity associated with the infinite summation $\sum_{q=-\infty}^{\infty}$ and the heterogeneity restriction associated with the constants $e_{n,t}$ by spreading out the information used in $\mathcal{Y}_{n,q}$. Rather than summing over every $E[y_{n,t}|\Im_{t+q}] - E[y_{n,t}|\Im_{t+q-1}]$, we show for some sequences of finite positive integers $\{g_n, r_n\}$

$$\lim_{n\to\infty}\left\|\sum_{t=1}^{n} y_{n,t} - \sum_{t=1}^{n}\sum_{q=-r_n}^{r_n}\left\{E[y_{n,t}|\Im_{g_n(t+q)}] - E[y_{n,t}|\Im_{g_n(t+q-1)}]\right\}\right\|_2 = 0.$$

This follows since we are free to choose $g_n \to \infty$ as $n \to \infty$ so fast that the terms $E[y_{n,t}|\Im_{g_n(t+r_n)}] \to y_{n,t}$ and $E[y_{n,t}|\Im_{g_n(t-r_n-1)}] \to 0$ in L_2-norm as fast as we choose by appealing to mixingale property (6). Precisely how fast $g_n \to \infty$ will be made clear in the proof, below.

We then show $E(\sum_{t=1}^{n}\sum_{q=-r_n}^{r_n}\{E[y_{n,t}|\Im_{g_n(t+q)}] - E[y_{n,t}|\Im_{g_n(t+q-1)}]\})^2 \leq K\sum_{t=1}^{n} E[y_{n,t}^2]$ by exploiting three facts. First, property (6) implies $\lim_{n\to\infty} ||E[y_{n,t}|\Im_{g_n(t+q)}] - E[y_{n,t}|\Im_{g_n(t+q-1)}]||_2 = 0$ for all $q \notin \{-t, -t+1\}$, where convergence occurs as fast as we choose by setting $g_n \to \infty$ sufficiently fast. Second, for the remaining $q \in \{-t, -t+1\}$ the terms $||E[y_{n,t}|\Im_{g_n(t+q)}] - E[y_{n,t}|\Im_{g_n(t+q-1)}]||_2 \leq 2||y_{n,t}||_2$ by Jensen's inequality. Third, $\sum_{q=-r_n}^{r_n}$ always sums over finitely many $E[y_{n,t}|\Im_{g_n(t+q)}] - E[y_{n,t}|\Im_{g_n(t+q-1)}]$, freeing us from mixingale size restrictions.

2.2 Proof of Theorem 2.1

The case $p = 1$ is trivial, so we prove the claim for any $p \in (1, 2]$ by modifying arguments in [37], [38], and [53].

Since the displacements q_n are arbitrary put

$$q_n = q \times g_n \quad \text{where} \quad q \in \mathbb{Z},$$

for some sequence $\{g_n\}$ of positive integers $g_n \in \mathbb{N}$, $\liminf_{n\to\infty} g_n \geq 1$, $g_n < \infty$ if $n < \infty$. Write $E_m[\cdot] := E[\cdot|\Im_m]$ and define

$$\mathcal{S}_n := \sum_{t=1}^{n} y_{n,,t} \quad \text{and} \quad \mathcal{Y}_{n,q}^{*} := \sum_{t=1}^{n}\left\{E_{g_n(t+q)}[y_{n,t}] - E_{g_n(t+q-1)}[y_{n,t}]\right\}.$$

Let $\{r_n\}$ be a sequence of positive integers that satisfies $1 + 3n \leq r_n < \infty$. We first derive the decomposition

$$\mathcal{S}_n = \sum_{q=-r_n}^{r_n} \mathcal{Y}_{n,q}^{*} + \mathcal{R}_n = \sum_{q=-r_n}^{r_n} \mathcal{Y}_{n,q}^{*} + o_p\left(g_n^{-\lambda} r_n^{-\lambda}\sum_{t=1}^{n} e_{n,t}\right) \tag{10}$$

where $\{\mathcal{R}_n\}$ is a zero mean stochastic sequence, for each n bounded and measurable. We then use $r_n < \infty$ to escape any mixingale size requirements, and $g_n \to \infty$ to promote the desired bound (3) itself.

Step 1 (finite series decomposition): Define

$$\mathcal{E}_{1,n} := \sum_{t=1}^{n}\left\{y_{n,t} - E_{g_n(t+r_n)}[y_{n,t}]\right\} \quad \text{and} \quad \mathcal{E}_{2,n} := \sum_{t=1}^{n} E_{g_n(t-r_n-1)}\left[y_{n,t}\right].$$

Clearly for any n

$$\sum_{q=-r_n}^{r_n}\left\{E_{g_n(t+q)}[y_{n,t}] - E_{g_n(t+q-1)}[y_{n,t}]\right\} = E_{g_n(t+r_n)}\left[y_{n,t}\right] - E_{-g_n(r_n+1-t)}[y_{n,t}]$$

hence

$$\mathcal{S}_n = \sum_{q=-r_n}^{r_n} \mathcal{Y}^*_{n,q} + \mathcal{E}_{1,n} + \mathcal{E}_{2,n}.$$

Note the identity

$$E_{g_n(t+r_n)}[y_{n,t}] = E\left[y_{n,t}|\Im_{t+(g_n-1)t+g_n r_n}\right]$$

where $(g_n - 1)t + g_n r_n \geq g_n r_n$ holds for each $n \geq 1$, $1 \leq t \leq n$, and $g_n \geq 1$. Similarly

$$E_{-g_n(r_n+1-t)}[y_{n,t}] = E\left[y_{n,t}|\Im_{t+(g_n-1)t-g_n(r_n+1)}\right]$$

where $(g_n - 1)t - g_n(r_n + 1) < -[g_n r_n/2]$ for all $n \geq 1$ and $1 \leq t \leq n$ since we assume $r_n \geq 3n + 1$. Now apply Assumption 1 with monotonic mixingale coefficients, $E[y_{n,t}] = 0$ and Minkowski's inequality to deduce

$$\begin{aligned}
\left\|\mathcal{E}_{1,n}\right\|_p &\leq \sum_{t=1}^{n} \left\|y_{n,t} - E_{g_n(t+r_n)}[y_{n,t}]\right\|_p \\
&\leq K\sum_{t=1}^{n} \left\|y_{n,t} - E_{t+g_n r_n}[y_{n,t}]\right\|_p = o\left(g_n^{-\lambda} r_n^{-\lambda} \sum_{t=1}^{n} e_{n,t}\right) \\
\left\|\mathcal{E}_{2,n}\right\|_p &\leq \sum_{t=1}^{n} \left\|E_{-g_n(r_n+1-t)}[y_{n,t}]\right\|_p \\
&\leq K\sum_{t=1}^{n} \left\|E_{t-[g_n r_n/2]}[y_{n,t}]\right\|_p = o\left(g_n^{-\lambda} r_n^{-\lambda} \sum_{t=1}^{n} e_{n,t}\right).
\end{aligned}$$

But this implies by Minkowski's inequality

$$\left\|\mathcal{S}_n - \sum_{q=-r_n}^{r_n} \mathcal{Y}^*_{n,q}\right\|_p \leq \left\|\mathcal{E}_{1,n}\right\|_p + \left\|\mathcal{E}_{2,n}\right\|_p = o\left(g_n^{-\lambda} r_n^{-\lambda} \sum_{t=1}^{n} e_{n,t}\right).$$

An appeal to Markov's inequality leads to (10) with $\mathcal{R}_n = \mathcal{E}_{1,n} + \mathcal{E}_{2,n}$, hence

$$\left\|\mathcal{R}_n\right\|_p = o\left(g_n^{-\lambda} r_n^{-\lambda} \sum_{t=1}^{n} e_{n,t}\right). \tag{11}$$

Step 2 (moment bound): Use the finite series decomposition (10) to deduce by Minkowski's inequality

$$\left\|\mathcal{S}_n\right\|_p \leq \left\|\sum_{q=-r_n}^{r_n} \mathcal{Y}^*_{n,q}\right\|_p + \left\|\mathcal{R}_n\right\|_p. \tag{12}$$

If $\limsup_{n\to\infty} \sum_{t=1}^{n} e_{n,t} \leq K$ then (11) and $r_n > n$ imply $||\mathcal{R}_n||_p \to 0$ for any $g_n \to \infty$. Otherwise choose any sequence of positive finite integers $\{g_n\}$ that satisfies $g_n(\sum_{t=1}^{n} e_{n,t})^{-1/\lambda} \to \infty$ to force

$$\left\|\mathcal{R}_n\right\|_p \to 0 \text{ hence } \left\|\mathcal{S}_n\right\|_p \leq \left\|\sum_{q=-r_n}^{r_n} \mathcal{Y}^*_{n,q}\right\|_p + o(1). \tag{13}$$

In order to bound $|| \sum_{q=-r_n}^{r_n} \mathcal{Y}^*_{n,q}||_p$, define

$$\mathcal{X}_{n,t,q} := E_{g_n(t+q)}[y_{n,t}] - E_{g_n(t+q-1)}[y_{n,t}]$$

such that

$$\mathcal{Y}_{n,q}^{*}=\sum_{t=1}^{n}\mathcal{X}_{n,t,q}.$$

Note $\{\mathcal{X}_{n,t,q},\Im_{g_n(t+q)}:1\leq t\leq n\}_{n\geq 1}$ forms for every q a martingale difference array:

$$E\left[\mathcal{X}_{n,t,q}|\Im_{g_n(t-1+q)}\right]=E\left(E[y_{n,t}|\Im_{g_n(t+q)}]-E[y_{n,t}|\Im_{g_n(t+q-1)}]\;|\Im_{g_n(t-1+q)}\right)=0.$$

Define $s=p/(p-1)$ for $p\in(1,2]$. Any sequence of non-zero real numbers $\{a_q\}_{q=-\infty}^{\infty}$, $\sum_{q=-\infty}^{\infty}a_q^{s/p}<\infty$, satisfies by Hölder's inequality (cf. [54])

$$\begin{aligned}E\left|\sum_{q=-r_n}^{r_n}\mathcal{Y}_{n,q}^{*}\right|^{p} &= E\left|\sum_{q=-r_n}^{r_n}a_q^{1/p}a_q^{-1/p}\mathcal{Y}_{n,q}^{*}\right|^{p} \\ &\leq \left(\sum_{q=-r_n}^{r_n}a_q^{s/p}\right)^{p/s}\times\sum_{q=-r_n}^{r_n}a_q^{-1}E\left|\mathcal{Y}_{n,q}^{*}\right|^{p}\leq K\sum_{q=-r_n}^{r_n}a_q^{-1}E\left|\mathcal{Y}_{n,q}^{*}\right|^{p}.\end{aligned}\tag{14}$$

Now set $a_0=1$ and $a_{-q}=a_q=q^{-p/s-\iota}$ such that $\sum_{q=-\infty}^{\infty}a_q^{s/p}=\sum_{q=-\infty}^{\infty}q^{-1-\iota(s/p)}$ $<\infty$, and define

$$\mathcal{Z}_{n,l}:=y_{n,t}-E_l[y_{n,t}]\text{ for }l\in\mathbb{Z}.$$

We require the following bounds by Minkowski's inequality:

$$\begin{aligned}&E\left|E_{g_n(t+q)}[y_{n,t}]-E_{g_n(t+q-1)}[y_{n,t}]\right|^{p}\leq\left(\left\|E_{g_n(t+q)}[y_{n,t}]\right\|_p+\left\|E_{g_n(t+q-1)}[y_{n,t}]\right\|_p\right)^{p}\\ &E\left|E_{g_n(t+q)}[y_{n,t}]-E_{g_n(t+q-1)}[y_{n,t}]\right|^{p}\leq\left(\left\|\mathcal{Z}_{n,g_n(t+q)}\right\|_p+\left\|\mathcal{Z}_{n,g_n(t+q-1)}\right\|_p\right)^{p}.\end{aligned}\tag{15}$$

Then $\sum_{q=-r_n}^{r_n}a_q^{-1}E\left|\mathcal{Y}_{n,q}^{*}\right|^{p}$ is bounded from above:

$$\begin{aligned}\sum_{q=-r_n}^{r_n}a_q^{-1}E\left|\mathcal{Y}_{n,q}^{*}\right|^{p} &= \sum_{q=-r_n}^{r_n}a_q^{-1}E\left|\sum_{t=1}^{n}\mathcal{X}_{n,t,q}\right|^{p}\leq K\sum_{q=-r_n}^{r_n}a_q^{-1}\sum_{t=1}^{n}E\left|\mathcal{X}_{n,t,q}\right|^{p} \\ &= K\sum_{t=1}^{n}\sum_{q=-r_n}^{r_n}a_q^{-1}E\left|E_{g_n(t+q)}[y_{n,t}]-E_{g_n(t+q-1)}[y_{n,t}]\right|^{p} \\ &\leq K\sum_{t=1}^{n}\sum_{q=-r_n}^{-t-1}a_q^{-1}\left(\left\|E_{g_n(t+q)}[y_{n,t}]\right\|_p+\left\|E_{g_n(t+q-1)}[y_{n,t}]\right\|_p\right)^{p} \\ &\quad+K\sum_{t=1}^{n}\sum_{q=-t+2}^{r_n}a_q^{-1}\left(\left\|\mathcal{Z}_{n,g_n(t+q)}\right\|_p+\left\|\mathcal{Z}_{n,g_n(t+q-1)}\right\|_p\right)^{p} \\ &\quad+K\sum_{t=1}^{n}\sum_{q=-t}^{-t+1}a_q^{-1}E\left|E_{g_n(t+q)}[y_{n,t}]-E_{g_n(t+q-1)}[y_{n,t}]\right|^{p} \\ &= \mathcal{A}_{1,n}+\mathcal{A}_{2,n}+\mathcal{A}_{3n}.\end{aligned}\tag{16}$$

The first inequality follows from a generalization of the van Bahr-Esséen inequality to L_p-bounded martingale difference arrays $\{\mathcal{X}_{n,t,q},\Im_{g_n(t+q)}:1\leq t\leq n\}_{n\geq 1}$ for $p\in(1,2]$

(see Theorem 2 of [68]) The second inequality follows from Minkowski's inequality and (15), and an obvious decomposition of the summation $\sum_{q=-r_n}^{r_n}$. Re-write the summations over q in $\mathcal{A}_{1,n}$ and $\mathcal{A}_{2,n}$ using $a_q = a_{-q}$:

$$\mathcal{A}_{1,n} = K\sum_{t=1}^{n}\sum_{q=1}^{r_n-t} a_{q+t}^{-1}\left(\left\|E_{-g_n q}[y_{n,t}]\right\|_p + \left\|E_{-g_n(q+1)}[y_{n,t}]\right\|_p\right)^p$$

$$\mathcal{A}_{2,n} = K\sum_{t=1}^{n}\sum_{q=2}^{r_n+t} a_{q-t}^{-1}\left(\left\|\mathcal{Z}_{n,g_n q}\right\|_p + \left\|\mathcal{Z}_{n,g_n(q-1)}\right\|_p\right)^p.$$

Apply the Assumption 1 mixingale property with monotonicity of coefficients, and $a_q > 0$, to obtain

$$\begin{aligned}\mathcal{A}_{1,n} &\leq K\sum_{t=1}^{n}\sum_{q=1}^{r_n} a_{q+t}^{-1}\left(\left\|E_{-g_n q}[y_{n,t}]\right\|_p + \left\|E_{-g_n(q+1)}[y_{n,t}]\right\|_p\right)^p \\ &\leq K\sum_{t=1}^{n}\sum_{q=1}^{r_n} a_{q+t}^{-1} e_{n,t}^p \zeta_{g_n q}^p \leq K g_n^{-p\lambda}\sum_{t=1}^{n} e_{n,t}^p \sum_{q=1}^{r_n} a_{q+t}^{-1} q^{-p\lambda}.\end{aligned}$$

Notice $0 < \sum_{q=1}^{r_n} a_{q+t}^{-1} q^{-p\lambda} < \infty$ by construction for each $r_n < \infty$ hence for each $n < \infty$. Since $\{g_n\}$ are arbitrary finite integers we can always choose them to satisfy (13) and

$$\mathcal{A}_{1,n} \leq K g_n^{-p\lambda}\sum_{t=1}^{n} e_{n,t}^p \sum_{q=1}^{r_n} a_{q+t}^{-1} q^{-p\lambda} \to 0. \tag{17}$$

By an identical argument and the fact that $t \leq n < r_n$, we can choose $\{g_n\}$ to satisfy (13), (17) and

$$\begin{aligned}\mathcal{A}_{2,n} &\leq K\sum_{t=1}^{n}\sum_{q=2}^{2r_n} a_{q-t}^{-1}\left(\left\|\mathcal{Z}_{n,g_n q}\right\|_p + \left\|\mathcal{Z}_{n,g_n(q-1)}\right\|_p\right)^p \\ &\leq K g_n^{-p\lambda}\sum_{t=1}^{n} e_{n,t}^p \sum_{q=2}^{2r_n} a_{q-t}^{-1} q^{-p\lambda} \to 0.\end{aligned} \tag{18}$$

Finally, apply Minkowski and Jensen inequalities, and $a_0 = 1$ and $a_{-q} = a_q = q^{-p/s-\iota}$, to deduce

$$\mathcal{A}_{3n} \leq K\sum_{t=1}^{n}\sum_{q=-t}^{-t+1} a_q^{-1} E\left|y_{n,t}\right|^p \leq K\sum_{t=1}^{n} E\left|y_{n,t}\right|^p. \tag{19}$$

Combine (12)-(14) with (16)-(19) and non-degeneracy (1) to obtain

$$\left\|\mathcal{S}_n\right\|_p \leq \left\|\sum_{q=-r_n}^{r_n}\mathcal{Y}_{n,q}^*\right\|_p + \left\|\mathcal{R}_n\right\|_p \leq K\left(\sum_{t=1}^{n} E\left|y_{n,t}\right|^p\right)^{1/p} + o(1) \leq K\left(\sum_{t=1}^{n} E\left|y_{n,t}\right|^p\right)^{1/p}.$$

This completes the proof. $\mathcal{QED}$.

3. Tail trimmed Arrays and Self-Scaled Laws An interesting application of Theorem 2.2 is to tail trimmed arrays $\{y_{n,t}\}$ of a heavy tailed stochastic process $\{y_t\}_{t\in\mathbb{Z}}$. Assume y_t is measurable on the probability space $(\Omega, \mathcal{G}, P)$, and uniformly L_p-bounded for some $p \in (0,2)$ with infinite variance

$$\inf_{t\in\mathbb{Z}} E\left[y_t^2\right] = \infty.$$

Let $\{y_t\}_{t=1}^n$ be the sample path with size $n \geq 1$, and let $\{c_{n,t}\} = \{c_{n,t} : 1 \leq t \leq n\}_{n\geq 1}$ be a triangular array of positive constants, $c_{n,t} \to \infty$ as $n \to \infty$ for all t. The tail trimmed version is

$$y_{n,t} := y_t I\left(|y_t| \leq c_{n,t}\right), \tag{20}$$

where $I(A) = 1$ if A is true, and 0 otherwise. The thresholds $\{c_{n,t}\}$ are determined by an intermediate order sequence $\{m_n\}$:

$$\frac{n}{m_n} P\left(|y_t| > c_{n,t}\right) \to 1 \text{ where } m_n \to \infty \text{ and } m_n = o(n). \tag{21}$$

Thus $c_{n,t}$ approximates the $m_n/n \to 0$ two-tailed quantile of y_t. Asymmetric trimming is a simple extension with only added notation.

Central order trimming $m_n/n \to (0,1)$ implies $y_{n,t}$ is uniformly bounded. This case is implicitly covered in the mixingale literature (e.g. [1], [14], [20]). Conversely, if y_t is non-integrable then extreme order trimming $m_n \to m$, a finite integer, results in too little trimming for a WLLN to be deduced from Theorem 2.2. See Remarks 1 and 3 of Theorem 3.3, below, and see Leadbetter et al [51] for seminal order statistic theory.

Infinite variance precludes such simple non-stationary cases as uniformly distributed y_t taking values in $\{-t^{\xi}, t^{\xi}\}$ for some $\xi > 0$. See, e.g., Davidson [14]. Examples covering (20)-(21) are detailed below.

Trivially $y_{n,t}$ is L_2-bounded for each t and n, while $\inf_{t\in\mathbb{Z}} E[y_t^2] = \infty$ ensures non-degeneracy (1) since under tail trimming

$$\liminf_{c\to\infty} \inf_{t\in\mathbb{Z}} E[y_t^2 I\left(|y_t| \leq c\right)] > 0.$$

Theorem 2.2 applies as long as $\{y_{n,t}, \Im_t\}$ is a mixingale array.

Assumption 2. *Let $\{y_{n,t}\}$ be the tail trimmed array defined by (20)-(21), and assume $\{y_{n,t}, \Im_t\}$ forms an L_2-mixingale array with monotonic coefficients $\zeta_{q_n} \searrow 0$ as $q_n \to \infty$ of any size $\lambda > 0$.*

Theorem 3.1 (Tail Trimmed Partial Sum Probability Bound). *Under Assumption 2 $\sum_{t=1}^n \{y_{n,t} - E[y_{n,t}]\} = O_p((\sum_{t=1}^n E[y_{n,t}^2])^{1/2})$.*

Since L_2-mixingales form L_p-mixingales for any $p \in (1,2]$ by Lyapunov's inequality, apply Theorems 2.1 and 3.1 to deduce for tiny $\iota > 0$

$$\sum_{t=1}^n \{y_{n,t} - E[y_{n,t}]\} = O_p\left(\left(\sum_{t=1}^n E\left|y_{n,t}\right|^{1+\iota}\right)^{1/(1+\iota)}\right). \tag{22}$$

Thus if y_t is uniformly $L_{1+\iota}$-bounded then $1/n \sum_{t=1}^n \{y_{n,t} - E[y_{n,t}]\} = O_p(n^{-\iota})$, hence a WLLN applies. Since this case is implicitly covered elsewhere, in the sequel we focus on non-integrable cases.

Corollary 3.2 (Tail Trimmed Mean WLLN). *Assume $\{y_t\}$ is uniformly $L_{1+\iota}$-bounded. Under Assumption 2 $1/n \sum_{t=1}^n \{y_{n,t} - E[y_{n,t}]\} \xrightarrow{p} 0$.*

Far greater refinement is available under a regular variation property. Assume for each t and slowly varying [s.v.] $L_t(y)$

$$P\left(|y_t| > y\right) = y^{-\kappa_t} L_t(y), \text{ where } \kappa_t \in (0,2],\ L_t(y) > 0 \text{ for each } t \in \mathbb{Z} \text{ and } y \in \mathbb{R}. \quad (23)$$

In the case of stationary tails

$$L_t(y) = L(y) \text{ and } \kappa_t = \kappa, \ \text{ hence the thresholds are } c_{n,t} = c_n. \quad (24)$$

Class (23)-(24) coincides with the maximum domain of attraction of a Type II extreme value distribution, and the domain of attraction of a stable law if $\kappa < 2$, it naturally characterizes stochastic recurrence equations like GARCH data, and has been used to model asset returns, urban growth, network data, insurance claims, and meteorological events. See Basrak et al [3], Bingham et al [4], Hill [40], [41], Kesten [50], Resnick [60], and their citations.

Define a signed power transform

$$y_{n,t}^{<s>} := \text{sign}(y_t) \times |y_{n,t}|^s \quad \text{for } s > 0.$$

A direct appeal to Karamata's Theorem under stationary regular variation (23)-(24) gives the well known trimmed moment formula (e.g. Theorem 0.6 of Resnick [60]):

$$E\left|y_{n,t}\right|^s \sim K c_n^s P\left(|y_t| > c_n\right) \sim K c_n^{s-\kappa} L\left(c_n\right) \text{ as } c_n \to \infty \ \forall s > \kappa \quad (25)$$

$$E\left|y_{n,t}\right|^\kappa \sim L\left(c_n\right) \text{ for s.v. } L\left(\cdot\right). \quad (25')$$

We can assume $L(\cdot)$ in (25) and (25′) are the same up to a multiplicative constant. Simply extend Theorem 3.1 to $\{|y_{n,t}|^s, y_{n,t}^{<s>}\}$ to deduce the following self-scaled laws. See the appendix for a proof.

Theorem 3.3 (Self-Scaled Tail Trimmed WLLN). *Let trimming properties (20) and (21) hold, assume tail property (23) under stationarity (24), choose any $s \geq \kappa$ and assume $\liminf_{n\to\infty}\{E\left|y_{n,t}\right|^s\} > 0$.*

a. If $|y_{n,t}|^s$ satisfies Assumption 2 then $1/n\sum_{t=1}^n |y_{n,t}|^s/E|y_{n,t}|^s = 1 + O_p(1/m_n^{1/2})$;

b. If $y_{n,t}^{<s>}$ satisfies Assumption 2 then $1/n\sum_{t=1}^n y_{n,t}^{<s>} = E[y_{n,t}^{<s>}] + O_p((n/m_n)^{s/\kappa-1+\iota} L\left(n\right)/n^\iota)$ for some s.v. $L(n)$ and tiny $\iota > 0$.

Remark 1: The proof only exploits the property (21) implication $c_n \to \infty$, and technically does not require $m_n \to \infty$. But extreme order trimming $m_n \to m$ implies $1/n\sum_{t=1}^n |y_{n,t}|^s/E|y_{n,t}|^s = 1 + O_p(1)$ so a WLLN cannot be deduced for higher moments $s \geq \kappa$.

Remark 2: Under intermediate order trimming the minimum rate of convergence $m_n^{1/2} \to \infty$ is optimized by maximal trimming, hence a minimal threshold rate $c_n \to \infty$. Consider $m_n = n/L(n)$ for s.v. $L(n) \to \infty$:

$$\frac{n^{1/2}}{L\left(n\right)}\left(\frac{1}{n}\sum_{t=1}^n \frac{\left|y_{n,t}\right|^s}{E\left|y_{n,t}\right|^s} - 1\right) = O_p(1).$$

Remark 3: Self-scaling under tail stationarity (23)-(24) provides an elegant avenue for characterizing a minimal rate since by Karamata's Theorem both $E|y_{n,t}|^s$ and $(\sum_{t=1}^n E|y_{n,t}|^{2s})^{1/2}$ are proportional to the same functions of n for any $s \geq \kappa$.

The non-stationary tail case similarly follows from Theorem 3.1. Consider for brevity a Paretian tail

$$P\left(|y_t| > y\right) = d_t y^{-\kappa_t}\left(1 + o(1)\right), \ \forall t \ 0 < d_t < \infty \text{ and } \kappa_t \in (0, 2]. \tag{26}$$

Use threshold construction (21) and Karamata's Theorem to deduce for any $s \geq \max_{t\in\mathbb{N}}\{\kappa_t\}$

$$c_{n,t} = d_t^{1/\kappa_t}(n/m_n)^{1/\kappa_t} \text{ and } E\left[|y_t|^s I\left(|y_t| \leq c_{n,t}\right)\right] \sim K d_t^{s/\kappa_t}\left(n/m_n\right)^{s/\kappa_t - 1}. \tag{27}$$

Non-degeneracy $\liminf_{n\to\infty}\inf_{t\in\mathbb{Z}} E\left|y_{n,t}\right|^{2s} > 0$ implies we can always write

$$\frac{1}{\left(\sum_{t=1}^n E\left|y_{n,t}\right|^{2s}\right)^{1/2}}\sum_{t=1}^n\left\{|y_{n,t}|^s - E\left|y_{n,t}\right|^s\right\} = \frac{\sum_{t=1}^n E\left|y_{n,t}\right|^s}{\left(\sum_{t=1}^n E\left|y_{n,t}\right|^{2s}\right)^{1/2}}\left\{\frac{\sum_{t=1}^n \left|y_{n,t}\right|^s}{\sum_{t=1}^n E\left|y_{n,t}\right|^s} - 1\right\}$$

where threshold and moment formulae (27) lead to

$$\begin{aligned}
\frac{\sum_{t=1}^n E\left|y_{n,t}\right|^s}{\left(\sum_{t=1}^n E\left|y_{n,t}\right|^{2s}\right)^{1/2}} &\sim K\frac{\sum_{t=1}^n d_t^{s/\kappa_t}\left(n/m_n\right)^{s/\kappa_t-1}}{\left(\sum_{t=1}^n d_t^{2s/\kappa_t}\left(n/m_n\right)^{2s/\kappa_t-1}\right)^{1/2}} \\
&= K\frac{\sum_{t=1}^n d_t^{s/\kappa_t}\left(n/m_n\right)^{s/\kappa_t}}{\left(\sum_{t=1}^n d_t^{2s/\kappa_t}\left(n/m_n\right)^{2s/\kappa_t}\right)^{1/2}} \times \left(\frac{m_n}{n}\right)^{1/2} = \mathcal{D}_n \times \left(\frac{m_n}{n}\right)^{1/2},
\end{aligned}$$

say. It is easy to confirm under tail-stationarity $d_t = d$ and $\kappa_t = \kappa$ that $\mathcal{D}_n = Kn^{1/2}$, hence $\mathcal{D}_n(m_n/n)^{1/2} = Km_n^{1/2}$ as in Theorem 3.3. Theorem 3.1 and the above relations prove the next result.

THEOREM 3.4. *Let trimming properties (20) and (21), and power-law tail (26) hold, and let $\{|y_{n,t}|^s\}$ satisfy Assumption 2 for $s \geq \max_{t\in\mathbb{N}}\{\kappa_t\}$. Then:*

$$\mathcal{D}_n \times \left(\frac{m_n}{n}\right)^{1/2} \times \left\{\frac{\sum_{t=1}^n |y_{n,t}|^s}{\sum_{t=1}^n E\left|y_{n,t}\right|^s} - 1\right\} = O_p\left(1\right). \tag{28}$$

Notice (27) does not guarantee a WLLN since $\mathcal{D}_n(m_n/n)^{1/2} \to 0$ is evidently possible in some non-stationary cases. If the probability tails have a constant index $\kappa_t = \kappa$ and are trending in scale then a WLLN does exist.

EXAMPLE 1 (Trending Scale): Assume y_t has tail (26) with indices $\kappa_t = \kappa$ and scales $d_t = d \times t^\xi$ for some $\xi \geq 0$ and finite $d > 0$. Then $c_{n,t} = Kt^{\xi/\kappa}(n/m_n)^{1/\kappa}$. If $\xi = 0$ then $\mathcal{D}_n = n^{1/2}$, and otherwise

$$\mathcal{D}_n = \frac{\sum_{t=1}^n t^{\xi s/\kappa}}{\left(\sum_{t=1}^n t^{2\xi s/\kappa}\right)^{1/2}} \sim \frac{Kn^{s/\kappa+1}}{n^{(2s/\kappa+1)/2}} = Kn^{1/2}.$$

If $|y_{n,t}|^s$ for some $s > 0$ satisfies mixingale Assumption 2, then by Theorem 3.4

$$\frac{\sum_{t=1}^n |y_t|^s I\left(|y_t| \leq t^{1/\kappa}(n/m_n)\right)}{\sum_{t=1}^n E\left[|y_t|^s I\left(|y_t| \leq t^{1/\kappa}(n/m_n)\right)\right]} = 1 + O_p\left(1/m_n^{1/2}\right).$$

The scale properties of power-law tails (26) imply trend in tail scale has no impact on the minimum rate of convergence.

EXAMPLE 2 (Stochastic Trend): Let $\{\epsilon_t\}$ be an iid process with stationary power-law tail

$$P\left(|\epsilon_t| > \epsilon\right) = d\epsilon^{-\kappa}\left(1 + o(1)\right),\ d > 0,\ \kappa > 0, \tag{29}$$

and define a stochastic trend $y_t := \sum_{i=0}^{t} \epsilon_{t-i}$. Then y_t has tail (26) with constant index $\kappa_t = \kappa$ and trending scale $d_t = d \times t$ (see, e.g., [8]), hence $c_{n,t} = Kt^{1/\kappa}(n/m_n)^{1/\kappa}$. Note for any $s > 0$ and every $n \geq 1$ there exists a sufficiently large $q \in \mathbb{N}$ that

$$\max_{1 \leq t \leq n} E\left(E\left|y_{n,t}\right|^s - E[|y_{n,t}|^s \ \ |\Im_{t-q}]\right)^2 = 0 \text{ and } \max_{1 \leq t \leq n} E\left(|y_{n,t}|^s - E[|y_{n,t}|^s \ \ |\Im_{t+q}]\right)^2 = 0,$$

so mixingale Assumption 2 holds. Therefore $\{y_t\}$ satisfies the conditions of Example 1.

4. Weak Laws under Stochastic Trimming Let $\{y_{n,t}\}$ be the tail-trimmed array in (20), and assume tail-stationarity for brevity: $c_{n,t} = c_n$. In practice a stochastic plug-in will be used for the threshold c_n, so define $y_t^{(a)} := |y_t|$, and construct two-tailed order statistics $y_{(1)}^{(a)} \geq y_{(2)}^{(a)} \geq \cdots y_{(n)}^{(a)}$ and a stochastically trimmed array $\{\hat{y}_{n,t}\}$,

$$\hat{y}_{n,t} = y_t I\left(|y_t| \leq y_{(m_n+1)}^{(a)}\right).$$

Clearly $y_{(m_n+1)}^{(a)}$ estimates c_n. The literature on the sharpness of the approximation for iid and mixing data is substantial (cf. [45], [51]), with few results under general dependence and heterogeneity. See [39], [40], [41, [45], [63] and their references.

In order to characterize $\sum_{t=1}^{n}\{\hat{y}_{n,t} - y_{n,t}\}$ we must bound the threshold approximation rate (21), and memory in the tail-event $\bar{I}_{n,t}(u) = I(|y_t| > c_n e^u)$. The following assumptions are verified for a variety of processes in Sections 5 and 6.

Assumption 3 (Extremal-NED). *$\{\bar{I}_{n,t}(u)\}$ is L_2-Near Epoch Dependent on $\{\Im_t\}$ with size 1/2: $||\bar{I}_{n,t}(u) - E[\bar{I}_{n,t}(u)|\Im_{t-q_n}^{t+q_n}]||_2 \leq e_{n,t}^*(u)\zeta_{q_n}^*$ where $e_{n,t}^*(u)$ is Lebesgue integrable on $\mathbb{R}_+$, $\max_{1\leq t\leq n}\sup_{u\geq 0}\{e_{n,t}^*(u)\} = K(m_n/n)^{1/2}$, $\zeta_{q_n}^* = o(q_n^{-1/2})$ and $\liminf_{n\to\infty} q_n \geq q \in \mathbb{N}$.*

Assumption 4 (Tail Order). $(n/m_n)P(|y_t| > c_n) = 1 + o(1/m_n^{1/2})$.

Remark: The Extremal-NED property Assumption 3 was first explored in Hill [39], [40], [41] as a way to characterize tail memory for asymptotic theory for tail estimators. Together Assumptions 3 and 4 with a base σ-field $\Im_t$ induced by an α-mixing process $\{\epsilon_t\}$ ensure consistency and asymptotic normality of an intermediate order statistic $y_{(m_n+1)}^{(a)}$, and tail index and tail dependence estimators under general dependence conditions. We use the assumptions here to exploit $y_{(m_n+1)}^{(a)}/c_n = 1 + O_p(1/m_n^{1/2})$ for a large class of dependent sequences $\{y_t\}$.

Lemma 4.1. *Under (20), tail-stationarity and Assumptions 2-4 $\sum_{t=1}^{n}\{\hat{y}_{n,t} - y_{n,t}\} = O_p(c_n)$.*

Lemma 4.1 and partial sum probability bound Theorem 3.1 together suggest we need only bound the thresholds c_n to characterize a minimum rate of convergence.

Assumption 5 (Threshold Bound). $c_n = O(n^{1/2}||y_{n,t}||_2)$.

Remark: The property characterizes a sequence of fixed point bounds

$$c_n \leq Kn^{1/2}\left(E\left[y_t^2 I\left(|y_t| \leq c_n\right)\right]\right)^{1/2}.$$

Such a sequences $\{c_n\}$ trivially exists under non-degeneracy $\liminf_{c\to\infty} ||y_t I(|y_t| \leq c)||_2 > 0$ since any $c_n \to \infty$ with $c_n \leq Kn^{1/2}$ satisfies the bound. Further, $c_n \leq Kn^{1/2}||y_{n,t}||_2$ automatically holds when y_t has Paretian tail (29). See Section 6.

The next result follows instantly from Theorem 3.1 and Lemma 4.1.

Theorem 4.2. *Under tail-stationarity and Assumptions 2-5*

$$\frac{1}{n^{1/2}\|y_{n,t}\|_2}\sum_{t=1}^{n}\{\hat{y}_{n,t}-E[y_{n,t}]\}=O_p(1).$$

Similarly, Corollary 3.2 and Lemma 4.1 deliver a WLLN for stochastically tail-trimmed arrays $\{\hat{y}_{n,t}\}$.

Corollary 4.3. *Let* $\{y_t\}$*be uniformly* $L_{1+\iota}$*-bounded and tail-stationary. Under Assumptions 2-5* $1/n\sum_{t=1}^{n}\{\hat{y}_{n,t} - E[y_{n,t}]\} \xrightarrow{p} 0$.

5. Mixingale Arrays: Assumptions 2-3 We verify all assumptions imposed for tail trimming in Sections 3 and 4. We begin with dependence assumptions here, and tackle tail order and threshold assumptions in Section 6.

In general mixingale sequences $\{y_t, \Im_t\}$ satisfy tail trimmed mixingale Assumption 2, and *both* mixingale Assumption 2 and NED tail array Assumption 3 apply to L_p-Near Epoch Dependent and α-mixing sequences $\{y_t\}$. Consult Hill [39], [40], [41] for related theory for tail and tail trimmed arrays. Define the tail trimmed array $\{y_{n,t}\}$ as in (20)-(21).

5.1 Mixingale

If $\{y_t, \Im_t\}$ is an L_p-mixingale, $p > 0$ then an argument identical to Corollary 3.5 in Hill [41] shows the tail trimmed $\{y_{n,t}, \Im_t\}$ forms an L_2-mixingale array.

Lemma 5.1. *If* $\{y_t, \Im_t\}$ *forms an* L_p*-mixingale,* $p > 0$*, then Assumption 2 holds.*

A trivial example of an L_p-mixingale is an adapted martingale difference sequence $\{y_t, \Im_t\}$ (McLeish [53]).

EXAMPLE 3 (GARCH): Define a strong-GARCH(p,q) process $y_t = h_t\epsilon_t$, where the error ϵ_t is iid with zero mean and unit variance, and $h_t^2 = \omega + \sum_{i=1}^{p}\alpha_i y_{t-i}^2 + \sum_{i=1}^{q}\beta_i h_{t-i}^2$, $\omega > 0$, $\alpha_i, \beta_i \geq 0$. Each y_t and $y_t y_{t-h}$ for $h \geq 1$ satisfy Lemma 5.1 with $\Im_t := \sigma(y_\tau : \tau \leq t)$.

If $\sum_{i=1}^{p}\alpha_i + \sum_{i=1}^{q}\beta_i \leq 1$ then y_t is stationary, and if ϵ_t has a positive density on $\mathbb{R}$ and at least one α_i or β_i is strictly positive then power-law tail (29) holds with index $\kappa > 2$. See, e.g., Theorem 3.1 in Basrak et al [3]. The same methods based on results in Kesten [50] coupled with convolution tail theory in Cline [11] suffice to show $y_t y_{t-h}$ for $h \geq 1$ has tail (29) with index $\kappa/2$. See Cline [12] for similar arguments.

Tail trimmed versions of y_t and $y_t y_{t-h}$ for any $h \geq 1$ therefore satisfy a WLLN by Theorem 3.1 and Lemma 5.1. Define $Y_{t,h} := y_t y_{t-h}$, and a trimmed version $Y_{n,t,h} := Y_{t,h} I(|Y_{t,h}| \leq c_{n,h})$ where $\{c_{n,h}, m_n\}$ satisfy $(n/m_n)P(|Y_{t,h}| > c_{n,h}) \to 1$. Then

$$\frac{1}{n^{1/2}\|Y_{n,t,h}\|}\sum_{t=1}^{n}\{Y_{n,t,h}-E[Y_{n,t,h}]\}=O_p(1).$$

If y_t is uniformly $L_{2+\iota}$-bounded then the sample tail trimmed covariance is consistent by Corollary 3.2 and Lemma 5.1: $1/n\sum_{t=1}^{n}\{Y_{n,t,h} - E[Y_{n,t,h}]\} \xrightarrow{p} 0$.

5.2 Near Epoch Dependence

Recall $\{y_t\}$ is L_p-Near Epoch Dependent on $\{\Im_t\}$ or on $\{\epsilon_t\}$ with size $\lambda > 0$ if there exist positive real sequences $\{d_t\}$ and $\{\psi_q\}$ such that $d_t < \infty$ for each t, $\psi_q = O(q^{-\lambda-\iota})$ for tiny $\iota > 0$, and

$$\left\| y_t - E\left[y_t | \Im_{t-q}^{t+q} \right] \right\|_p \leq d_t \psi_q.$$

The NED property dates in some form to Ibragimov and Linnik [47] and McLeish [53], [54], and was coined in Gallant and White [31]. See Davidson [15], Nze and Doukhan [57] and Hill [41] for historical notes. Both NED and mixingale properties are relaxed versions of L_p-Weak Dependence ([70]) and related to s-weak dependence ([23]). The idea is y_t may be perfectly predicted in L_p-norm by the near-epoch $\{\epsilon_\tau\}_{t-q}^{t+q}$ as $q \to \infty$.

Similarly, $\{\epsilon_t\}$ is α-mixing or strong mixing with coefficients α_q of size $\theta > 0$ if

$$\alpha_q := \sup_{t \in \mathbb{Z}} \sup_{A \subset \Im_{-\infty}^{t-q}, B \subset \Im_t^{+\infty}} \left| P(A \cap B) - P(A)P(B) \right| = O(q^{-\theta-\iota}).$$

The property applies to β-mixing, ϕ-mixing, ρ-mixing, and geometrically ergodic sequences, as well as a variety of weak dependence properties ([23], [26], [27]. [47]).

LEMMA 5.2. *Let $\{y_t\}$ be L_r-NED, $r > 0$, on $\{\Im_t\}$ with any size $\lambda > 0$, and α-mixing base $\{\epsilon_t\}$ of any size $\theta > 0$. Then $y_{n,t}^{<s>}$ and $|y_{n,t}|^s$ for any $s > 0$ satisfy Assumption 2 with mixingale constants $e_{n,t} = c_{n,t}^s$, and Assumption 3.*

Remark 1: The mixingale constants $c_{n,t}^s$ for $y_{n,t}^{<s>}$ and $|y_{n,t}|^s$ may be much larger than $(E|y_{n,t}|^{2s})^{1/2}$. Consider power-law tail (26): if $s > \kappa_t/2$ then $E|y_{n,t}|^{2s} \sim K c_{n,t}^{2s}(m_n/n)$ *is dominated by* $c_{n,t}^s$ under tail trimming $m_n/n \to 0$. Theorem 2.1 is important for deducing a "minimal" scale $\sum_{t=1}^n E|y_{n,t}|^{2s} < \sum_{t=1}^n e_{n,t}^2$ for convergence. Clearly if it is possible to characterize constants $e_{n,t}$ that are *smaller* than $c_{n,t}^s$ this conclusion may no longer be valid. The problem is precisely that computing mixingale constants under mixing or NED assumptions typically requires higher moments and some form of bounding argument. In our setting then $e_{n,t} > (E|y_{n,t}|^{2s})^{1/2}$ is hardly surprising. See especially Theorem 17.5 of Davidson [15] and Corollary 3.5 of Hill [41].

Remark 2: Product convolutions y_t^2 or $y_t y_{t-h}$ are $L_{r/2}$-NED if y_t is L_r-NED by a straightforward generalization of Theorem 17.9 of Davidson [15][2]. Thus Lemma 5.2 extends to tail trimmed higher moments.

Since any y_t is NED on itself, the next claim follows from Lemma 5.2 by setting $\epsilon_t = y_t$.

COROLLARY 5.3. *If $\{y_t\}$ is L_p-bounded for any $p > 0$ and α-mixing with any size then Assumptions 2 and 3 hold.*

EXAMPLE 4 (Non-Stationary Distributed Lag): Let $\{\epsilon_t\}$ be a uniformly L_p-bounded process, $p > 0$, and define the distributed lag

$$y_t := \sum_{i=0}^{\infty} \psi_{t,i} \epsilon_{t-i}.$$

Assume $\psi_{0,i} = 1$, and assume $\psi_{t,i}$ are L_p-bounded $||\psi_{t,i}||_p = \varpi(t) i^{-\mu}$ for some $p > 0$, $\mu > 1/\min\{1, p\}$ and some mapping $\varpi : \mathbb{N} \to \mathbb{R}_+$, $0 \leq \varpi(t) < \infty$ for each t. Further, $\{\psi_{t,i}, \epsilon_t\}$ are α-mixing with σ-field

$$\Im_t = \sigma\left(\left\{\psi_{\tau,i}, \epsilon_\tau\right\} : 0 \leq i \leq t, \ -\infty < \tau \leq t\right).$$

[2] Davidson ([54]: p. 268) uses triangle and Cauchy-Schwartz inequalities to prove the claim for $r = 2$. Replace the triangle inequality with Loève's inequality to deduce the claim for any $r > 0$.

If $p \geq 2$ use Minkowski and Cauchy-Schwartz inequalities to deduce for any $q \in \mathbb{N}$

$$\left\| y_t - E\left[y_t | \Im_{t-q}^{t+q}\right] \right\|_{p/2} \leq \sum_{i=q+1}^{\infty} \left\| \psi_{t,i} \right\|_p \left\| \epsilon_t \right\|_p \leq \sup_{t \in \mathbb{Z}} \left\| \epsilon_t \right\|_p \sum_{i=q+1}^{\infty} \varpi(t) q^{-\mu} \leq K \varpi(t) q^{-\mu}.$$

If $p < 2$ then by $\mu > 1/p$ and Loève's and Cauchy-Schwartz inequalities

$$\left\| y_t - E[y_t | \Im_{t-q}^{t+q}] \right\|_{p/2} \leq \left(\sum_{i=q+1}^{\infty} \left(E\left|\psi_{t,i}\right|^p \times \sup_{t \in \mathbb{Z}} E\left|\epsilon_t\right|^p \right)^{1/2} \right)^{2/p} \leq K \varpi(t) q^{-\mu}.$$

Hence $\{y_t\}$ is $L_{p/2}$-NED on $\{\Im_t\}$ with size $\lambda = \mu - \iota$ and constants $d_t = K\varpi(t)$.

EXAMPLE 5 (Distributed Lag): Assume $y_t = \sum_{i=0}^{\infty} \psi_{t,i} \epsilon_{t-i}$ where ϵ_t is iid with stationary regularly varying tail (23)-(24) and index $\kappa > 0$. Assume $\psi_{t,i}$ is independent of ϵ_t and $\sum_{i=0}^{\infty} E|\psi_{t,i}|^{\kappa} < \infty$. Then y_t is a special case of Example 4, and since $E|\psi_{t,i}|^{\kappa} < \infty$ it follows $\psi_{t,i}\epsilon_t$ has tail (23)-(24) with index $\kappa > 0$, and by independence y_t has tail (23) with index κ (see [7] and [8]).

EXAMPLE 6 (ARFIMA): Let y_t be ARFIMA(p, d, q) with difference $d = 1 - \mu$ for some $\mu > 1/\min\{1, p\}$. If the innovations ϵ_t are uniformly L_p-bounded α-mixing then Example 4 applies.

EXAMPLE 7 (Threshold AR(1)-GARCH(1,1)): Assume y_t is a stationary first-order threshold autoregression

$$\begin{aligned} y_t &= \phi y_{t-1} I\left(y_{t-1} \leq a\right) + \epsilon_t, \quad |\phi| < 1 \\ \epsilon_t &= h_t u_t,\, u_t \sim (0,1)\,,\, h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2,\, \omega > 0,\, \alpha, \beta \geq 0. \end{aligned}$$

Stationarity is assured by $E[\ln(\alpha\epsilon_t^2 + \beta)] < 0$, hence $\{y_t, \epsilon_t\}$ are geometrically α-mixing. The same applies if the binary switching indicator $I(\cdot)$ is replaced with any uniformly bounded $g : \mathbb{R} \to \mathbb{R}$: $y_t = \phi y_{t-1} g(y_{t-1}) + \epsilon_t$, $|\phi| < 1$. See Meitz and Saikkonen [55]. Smooth Transition AR models, for example, are based on either $g(y_t) = \exp\{-\gamma(y_t - c)^2\}$ or $g(y_t) = [1 + \exp\{-\gamma(y_t - c)\}]^{-1}$ with threshold $c \in \mathbb{R}$ and rate of transition $\gamma \geq 0$.

EXAMPLE 8 (AR(1)-QARCH(1,1)): Let $y_t = \phi y_{t-1} + \epsilon_t$ and $\epsilon_t = h_t u_t$, where $u_t \overset{iid}{\sim} (0,1)$ and $|\phi| \in (0,1)$, with Quadratic-ARCH volatility $h_t = |\alpha + \beta y_{t-1}|$, $\alpha > 0$ and $\beta \in (0,1]$. Then ϵ_t is geometrically ergodic with stationary tail (29). See, e.g., Cline [12].

EXAMPLE 9 (Hyperbolic-GARCH): Davidson [16] proposed a hyperbolic memory GARCH(p, q) model that escapes discontinuities and memory perversions of the FI-GARCH class. Let $y_t = h_t\epsilon_t$, $\epsilon_t \overset{iid}{\sim} (0,1)$, and $\beta(L)h_t^2 = \omega + \alpha(L)y_t^2$, $\omega > 0$, where the lag polynomials are $\alpha(L) = \sum_{i=1}^{p} \alpha_i L^i$ and $\beta(L) = \sum_{i=1}^{q} \beta_i L^i$ with $\alpha_i, \beta_i \geq 0$ Define $\delta(L) = \alpha(L) + \beta(L)$ and assume the roots of $\beta(L)$ lie outside he unit circle. Since h_t^2 has the form $\omega + \pi(L)y_t^2$, Davidson [16] suggests the Hyperbolic-GARCH model

$$\pi(L) = 1 - \frac{\delta(L)}{\beta(L)} \left[1 - \frac{\gamma}{\zeta\left(1 + d\right)} \sum_{i=0}^{\infty} i^{-1-d} L^i \right],\ d > 0, \gamma \geq 0,$$

where $\zeta(\cdot)$ is the Riemann zeta function. The index $d > 0$ governs the degree of hyperbolic memory. As long as $\gamma \in [0, 1)$ then y_t is L_2-bounded L_1-NED on $\{\epsilon_t\}$ with size $\lambda = d -$

ι (Theorem 3.1 of Davidson [16]).

EXAMPLE 10 (Stochastic Volatility): Consider a univariate Log-Autoregressive Stochastic Volatility process $y_t = h_t \epsilon_t$ where $\ln h_t^\kappa = \omega + \phi \ln h_{t-1}^\kappa + u_t$, and $|\phi| < 1$. Assume ϵ_t is α-mixing with regularly varying tail (15)-(16), and $u_t \overset{iid}{\sim} N(0,1)$. Then $\{y_t\}$ satisfies (23)-(24) and is geometrically L_p-NED on $\{\epsilon_t\}$. See Hill [42].

6. Tail Order and Threshold Bounds: Assumptions 4-5 Finally, we verify tail order Assumption 4 and threshold bound Assumption 5. Assume tail stationarity throughout.

Assumption 4 (tail order): Assumption 4 holds for smooth distributions, and under second order regular variation properties like slow variation with remainder, including Pareto-type tails (21). See Goldie and Smith [32] and Haeusler and Teugels [34] for numerous examples.

Absolute Continuity: If y_t has a distribution $F(y) := P(y_t \leq y)$ absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$-*a.e.* then there exists of a probability density $f(y) := (\partial/\partial y)F(y)$ on $\mathbb{R}$-*a.e.* Continuity ensures arbitrarily many threshold sequences $\{c_n\}$ can be found to satisfy $\int_{-c_n}^{c_n} f(y)dy = 1 - m_n/n$ for any chosen intermediate order $\{m_n\}$, hence Assumption 4 holds: $(n/m_n)P(|y_t| > c_n) = 1$.

Second Order Power-Law: If $P(|y_t| > y) = dy^{-\kappa}(1 + O(y^{-\alpha}))$, $\alpha, \kappa > 0$, then it is easy to show $(n/m_n)P(|y_t| > c_n) = 1 + o(1/m_n^{1/2})$ holds for $c_n = d^{1/\kappa}(n/m_n)^{1/\kappa}$ and any intermediate order sequence $\{m_n\}$ with $m_n = o(n^{2\alpha/(2\alpha+\kappa)})$.

Assumption 5 (threshold bound): We demonstrate a restricted version of Assumption 5 in general, and exactly Assumption 5 for power-law tails.

Sharp Bound: If $\liminf_{c\to\infty} ||y_t I(|y_t| \leq c)||_2 > 0$ then $c_n \leq Kn^{1/2}||y_{n,t}||_2$ automatically holds for any $c_n \to \infty$ and $c_n \leq Kn^{1/2}$. There is a potential cost to pay for such a tight bound. Consider if y_t has stationary tail (29) then $c_n = K(n/m_n)^{1/\kappa} \leq Kn^{1/2}$ always holds if $\kappa \geq 2$, and otherwise only for fractiles $\liminf_{n\to\infty}\{m_n/n^{1-\kappa/2}\} \geq K$. This threshold bound is achieved only if sufficiently many observations are trimmed.

Power-Law: In order to exploit $||y_{n,t}||_2 \to \infty$ for non-square integrable y_t and any threshold $c_n \to \infty$ it helps to assume y_t has tail (29). If variance is finite $\kappa > 2$ then $(n/m_n)^{1/\kappa} \leq (n/m_n)^{1/2} \leq n^{1/2}$, hence $c_n \leq Kn^{1/2} = O(n^{1/2}||y_{n,t}||_2)$ so Assumption 5 holds.

Conversely, if variance is infinite $\kappa \in (0,2)$ then by Karamata's Theorem, cf. (25),

$$E\left[y_t^2 I\left(|y_t| \leq c_n\right)\right] \sim Kc_n^2(m_n/n) = K(n/m_n)^{2/\kappa - 1}\left(1 + o(1)\right),$$

and if $\kappa = 2$ then $E[y_t^2 I(|y_t| \leq c_n)] \sim L(c_n)$ for s.v. $L(c_n)$. Now since $n/m_n \leq n$ under tail trimming it follows $c_n \leq K(n/m_n)^{1/\kappa} \leq Kn^{1/2}(n/m_n)^{1/\kappa - 1/2}$ for any $\kappa \in (0,2)$, and $c_n \leq K(n/m_n)^{1/2} \leq n^{1/2}L(c_n)$ if $\kappa = 2$ since by tail trimming $m_n = o(n)$. Therefore, $c_n = O(n^{1/2}||y_{n,t}||_2)$ for any intermediate order sequence $\{m_n\}$, so again Assumption 5 holds.

Appendix A: Proofs

Proof of Theorem 3.3.

Claim (a): By trimmed moment formula (25) it follows for any $s > \kappa$ and s.v. $L(u)$, $L(u) > 0$ $\forall u \in \mathbb{R}$,

$$\frac{nE\left|y_{n,t}\right|^{s}}{n^{1/2}\left(E\left|y_{n,t}\right|^{2s}\right)^{1/2}} \sim \frac{nc_n^{s-\kappa}L\left(c_n\right)}{n^{1/2}c_n^{s-\kappa/2}L\left(c_n\right)^{1/2}} = \frac{n^{1/2}L\left(c_n\right)^{1/2}}{c_n^{\kappa/2}}. \tag{30}$$

Further, regular variation and tail stationarity (23)-(24) imply by construction c_n satisfies the fixed point identity $c_n = (n/m_n)^{1/\kappa}L(c_n)^{1/\kappa}$ where $L(c_n)$ is the same s.v. function in (30). Hence

$$\frac{n^{1/2}L\left(c_n\right)^{1/2}}{c_n^{\kappa/2}} = \frac{n^{1/2}L\left(c_n\right)^{1/2}}{(n/m_n)^{1/2}L(c_n)^{1/2}} = m_n^{1/2}. \tag{31}$$

Combine (30) and (31) and invoke tail-stationarity and Theorem 3.1 to deduce as claimed

$$\begin{aligned}
m_n^{1/2}\left(\frac{\hat{\mu}_n^{(s)}}{\mu_n^{(s)}}-1\right) &\sim \frac{n\mu_n^{(s)}}{n^{1/2}\left(E\left|y_{n,t}\right|^{2s}\right)^{1/2}}\left(\frac{\hat{\mu}_n^{(s)}}{\mu_n^{(s)}}-1\right) \\
&= \frac{n}{\left(\sum_{t=1}^{n}E\left|y_{n,t}\right|^{2s}\right)^{1/2}}\left(\hat{\mu}_n^{(s)}-\mu_n^{(s)}\right) \\
&= \frac{1}{\left(\sum_{t=1}^{n}E\left|y_{n,t}\right|^{2s}\right)^{1/2}}\sum_{t=1}^{n}\left\{\left|y_{n,t}\right|^{s}-E\left|y_{n,t}\right|^{s}\right\} = O_p\left(1\right).
\end{aligned}$$

The case $s = \kappa$ is identical since $E|y_{n,t}|^{\kappa} \sim L\left(c_n\right)$ and $E|y_{n,t}|^{2\kappa} \sim c_n^{\kappa/2}L\left(c_n\right)^{1/2}$, where $L(\cdot)$ are the same s.v. functions up to a multiplicative constant. This implies (30) with $Kn^{1/2}L\left(c_n\right)^{1/2}/c_n^{\kappa/2}$ for some s.v. $L(\cdot)$.

Claim (b): We require $L(c_n) \sim L(n)$, where s.v. $L(\cdot)$ may be different in different places. This follows from the construction of c_n and properties of regular variation (15)-(16). Observe for any $a > 0$

$$1 \leftarrow \frac{an}{m_n}P\left(|y_t| > c_{an}\right) \sim \frac{n}{m_n}P\left(|y_t| > a^{-1/\kappa}c_{an}\right),$$

hence we can assume $c_{an} = a^{1/\kappa}c_n$. Since this implies by slow variation $L(c_{an})/L(c_n) = L(a^{1/\kappa}c_n)/L(c_n) \rightarrow 1$, we may write $L(c_n) \sim L(n)$ for some s.v. $L(n)$.

By the same argument as claim (a), and $L(c_n) \sim L(n)$, for any $s > \kappa$ (case $s = \kappa$ is similar) and tiny $\delta > 0$

$$\left(\sum_{t=1}^{n}E\left|y_{n,t}\right|^{s(1+\delta)}\right)^{1/(1+\delta)} \sim n^{1/(1+\delta)}c_n^{s-\kappa/(1+\delta)}L\left(c_n\right)^{1/(1+\delta)} \sim Kn^{1-\iota}(n/m_n)^{s/\kappa-1+\iota}L\left(n\right)^{1-\iota},$$

where $\iota := \delta/(1+\delta) > 0$ is tiny. Now invoke consequence (22) of Theorem 3.1 to deduce the claim. $\mathcal{QED}$.

Proof of Lemma 4.1. Define $I_{n,t} := I(|y_t| \leq c_n)$, $\bar{I}_{n,t} := 1 - I_{n,t}$ and $\hat{I}_{n,t} := I(|y_t|$

$\leq y^{(a)}_{(m_n+1)})$. By construction and the triangle inequality

$$\begin{aligned}\left|\sum_{t=1}^{n}\left\{\hat{y}_{n,t}-y_{n,t}\right\}\right| &\leq \max_{1\leq t\leq n}\left\{\left|y_t\left\{\hat{I}_{n,t}-I_{n,t}\right\}\right|\right\}\times\sum_{t=1}^{n}\left|\hat{I}_{n,t}-I_{n,t}\right| \\ &\leq K\left|y^{(a)}_{(m_n+1)}-c_n\right|\times\left(\left|\sum_{t=1}^{n}\left\{\bar{I}_{n,t}-E\left[\bar{I}_{n,t}\right]\right\}\right|+m_n\left|\frac{n}{m_n}E\left[\bar{I}_{n,t}\right]-1\right|\right) \\ &= \mathcal{A}_n\times\left(\mathcal{B}_n+\mathcal{C}_n\right).\end{aligned}$$

Notice $\mathcal{C}_n = O(m_n^{1/2})$ is implied by threshold Assumption 4.

Use Lemma 3.3.1 of Hill [39], cf. Lemma 3 of Hill [40], to deduce $\mathcal{A}_n \leq 2c_n|y^{(a)}_{(m_n+1)}/c_n - 1| = O_p(c_n/m_n^{1/2})$. Further, Extremal-NED Assumption 3 and variance bound Theorem 2.1 imply $E[\mathcal{B}_n/m_n^{1/2}] \leq 1/m_n \sum_{t=1}^{n} E\left[\bar{I}_{n,t}\right] \leq K$ hence $\mathcal{B}_n = O_p(m_n^{1/2})$ by Chebyshev's inequality. Therefore $\mathcal{A}_n \times (\mathcal{B}_n + O(m_n^{1/2})) = O_p(c_n m_n^{-1/2} m_n^{1/2}) = O_p(c_n)$ as required. $\mathcal{QED}$.

Proof of Lemma 5.2.

Assumption 2: Under the L_r-NED supposition $\{|y_{n,t}|^s/c^s_{n,t}\}$ is, for any $s > 0$, L_2-NED on $\{\Im_t\}$ by Corollary 3.5 of Hill [41] . Since the base is α-mixing and $|y_{n,t}|^s/c^s_{n,t} \leq 1$, apply Theorem 17.5 of Davidson [15] to deduce $\{|y_{n,t}|^s/c^s_{n,t}, \Im_t\}$ is an L_2-mixingale array with bounded constants K, hence $\{|y_{n,t}|^s, \Im_t\}$ is an L_2-mixingale array with constants $Kc^s_{n,t}$. An identical argument applies to $y^{<s>}_{n,t}$.

Assumption 3: The L_r-NED supposition implies $\{I(|y_t| > c_n e^u)\}$ is L_2-NED on $\{\Im_t\}$ with constants $e^*_{n,t}(u) = K(m_n/n)^{1/2}e^{-u/2}$ and coefficients $\vartheta_{q_n} = o(q_n^{-1/2})$ by Theorem 2.1 of Hill [41]. $\mathcal{QED}$.

ACKNOWLEDGEMENTS

We thank Paul Doukhan for comments on an earlier version of the manuscript.

REFERENCES

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

[2] Arcones, M.A. (1998). The Law of Large Numbers for U-Statistics under Absolute Regularity, Elect. Comm. Prob. 3, 13-19.

[3] Basrak, B., R.A. Davis, and T. Mikosch (2002). Regular Variation of GARCH Processes, Stoch. Proc. Appl. 12, 908-920.

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

[5] Birkel, T. (1988). Moment Bounds for Associated Sequences, Ann. Prob. 16, 1184-1193.

[6] Boucheron, S., O. Bousquet, G. Lugosi, and P. Massart (2005). Moment Inequalities for Functions of Independent Random Variables, Ann. Prob. 33, 514-460.

[7] Breiman, L. (1965). On Some Limit Theorems Similar to the Arc-Sine Law, Theory Prob. Appl. 10, 351–360.

[8] Brockwell, P.J. and D.B.H. Cline (1985). Linear Prediction of ARMA Processes with Infinite Variance, Stoch. Proc. Appl. 19, 281-296.

[9] Burkholder, D.L. (1973). Distribution Function Inequalities for Martingales, Ann. Prob. 1, 19-42.

[10] Chandra, T.K., and S. Ghosal (1996). The Strong Law of Large Numbers for Weighted Averages Under Dependence Assumptions, J. Theor. Prob. 9, 797-809.

[11] Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, Prob. Theory Rel. Fields 72, 529–557.

[12] Cline, D.B.H. (2007). Regular Variation of Order 1 Nonlinear AR-ARCH models, Stoch. Proc. Appl. 117, 840-861.

[13] Csörgő, S., E. Haeusler, and D.M. Mason (1988). The Asymptotic Distribution of Trimmed Sums, Ann. Prob. 16, 672-699.

[14] Davidson, J. (1993). An L1-Convergence Theorem for Heterogeneous Mixingale Arrays with Trending Moments, Stat. Prob. Let. 16, 301-304.

[15] Davidson, J. (1994). Stochastic Limit Theory. Oxford University Press: Oxford.

[16] Davidson, J. (2004). Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model, J. Bus. Econ. Statistics 22, 16-29.

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

[18] Davydov, Y.A. (1968). Convergence of Distributions Generated by Stationary Stochastic Processes, Theory Prob. App. 13, 691-696.

[19] de Jong, R.M. (1995). Laws of Large Numbers for Dependent Heterogeneous Processes, Econometric Theory 11, 347-358.

[20] de Jong, R.M. (1996). A Strong Law of Large Numbers for Triangular Mixingale Arrays, Stat. Prob. Let. 27, 1-9.

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

[22] de la Peña, V.H., R. Ibragimov, and S. Sharakhmetov (2003). On Extremal Distributions and Sharp Lp-Bounds for Sums of Multilinear Forms, Ann. Prob. 31, 630-675.

[23] Dedecker, J. and P. Doukhan (2003). A New Covariance Inequality and Applications, Stoch. Proc. Appl. 106, 63-80.

[24] Dehling, H.G. and O.Sh. Sharipov (2009). Marcinkiewicz–Zygmund Strong Laws for U-Statistics of Weakly Dependent Observations, Stat. Prob. Let. 79, 2028-2036.

[25] Doob, J.L. (1953). Stochastic Processes, John Wiley: New York.

[26] Doukhan, P. (1994). Mixing: Properties and Examples, Lecture Notes in Statistics 85. Springer: New York.

[27] Doukhan, P. and S. Louhichi (1999). A New Weak Dependence Condition and Applications to Moment Inequalities, Stoch. Proc. Appl. 84, 313-342.

[28] Doukhan, P. and M. Neumann (2007). Probability and Moment Inequalities for Sums of Weakly Dependent Random Variables, with Applications, Stoch. Proc. Appl. 117, 878-903.

[29] Doukhan, P. and O. Wintenberger (2008). Weakly Dependent Chains with Infinite Memory, Stoch. Proc. Appl. 118, 1997-2013.

[30] Feller, W. (1971). An Introduction to Probability Theory and its Applications, 2nd ed., Vol. 2. Wiley: New York.

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

[32] Goldie, C.M., and R.L. Smith (1987). Slow Variation with Remainder: Theory and Applications, Q. J. Math. 38, 45-71.

[33] Griffin, P.S. and F.S. Qazi (2002). Limit Laws of Modulus Trimmed Sums, Ann. Prob. 30, 1466-1485.

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

[35] Hahn, M.G., J. Kuelbs and J. Samur (1987). Asymptotic Normality of Trimmed Sums of Mixing Random Variables, Ann. Prob. 15, 1395-1418.

[36] Hahn, M.G., J. Kuelbs and D.C. Weiner (1990). The Asymptotic Joint Distribution of Self-Normalized Censored Sums and Sums of Squares, Ann. Prob. 18, 1284-1341.

[37] Hansen, B.E. (1991). Strong Laws for Dependent Heterogeneous Processes, Econometric Theory 7, 213-221.

[38] Hansen, B.E. (1992). Erratum: Strong Laws for Dependent Heterogeneous Processes, Econometric Theory 8, 421-422.

[39] Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, Journal of Statistical Planning and Inference: 139, 2091-2110.

[40] Hill, J.B. (2010). On Tail Index Estimation for Dependent, Heterogeneous Data, Econometric Theory 26, 1398-1436.

[41] Hill, J.B. (2011a). Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics, Econometric Theory: forthcoming.

[42] Hill, J.B. (2011b). Extremal Memory of Stochastic Volatility with an Application to Tail Shape Inference, J. Stat. Plan. Infer. 141, 663-676

[43] Hitczenko, P. (1990). Best Constants in Martingale Version of Rosenthal's Inequality, Ann. Prob. 18, 1656-1668.

[44] Hitczenko, P. (1994). On a Domination of Sums of Random Variables by Sums of Conditionally Independent Ones, Ann. Prob. 22, 453-468.

[45] Hsing, T. (1991). On Tail Index Estimation Using Dependent Data, Ann. Statist. 19, 1547-1569.

[46] Ibragimov, I.A. (1962). Some Limit Theorems for Stationary Processes, Theory Prob. Appl. 7, 349-382.

[47] Ibragimov, I.A. and Yu. V. Linnik (1971). Independent and Stationary Sequences of Random Variables, Walters-Noordhoff: Groningen.

[48] Ibragimov, R. and Sh. Sharakhmetov (1997). On An Exact Constant for the Rosenthal Inequality, Theory Prob. Appl.42 294-302.

[49] Ibragimov, R. and Sh. Sharakhmetov (1998). Exact Bounds on the Moments of Symmetric Statistics. In Seventh Vilnius Conference on Probability Theory and Mathematical Statistics, 22nd European Meeting of Statisticians, Abstracts of Communications 243-244.

[50] Kesten, H. (1973) Random Difference Equations and Renewal Theory for Products of Random Matrices, Acta Mathematica 131, 207–248.

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

[52] Marcinkiewicz, J. and A. Zygmund (1937). Sur Les Foncions Independantes, Fundamenta Mathematicae 28, 60–90.

[53] McLeish, D.L. (1975a). Maximal Inequality and Dependent Strong Laws, Ann. Prob. 3, 829-839.

[54] McLeish, D.L. (1975b). Invariance Principles for Dependent Variables, Prob. Theory Rel. Fields 32, 165-178.

[55] Meitz M. and P. Saikkonen (2008). Stability of Nonlinear AR-GARCH Models, J. Time Ser. Anal. 29, 453-475.

[56] Newcomb, S. (1886). A Generalized Theory of the Combination of Observations so as to Obtain the Best Result, Amer. J. Math. 8, 343-366.

[57] Nze, P.A., and P. Doukhan (2004). Weak Dependence: Models and Applications to Econometrics, Econometric Theory 20, 995-1045.

[58] Osękowski, A. (2011). Sharp Maximal Inequality for the Moments of Martingales and Non-negative Submartingales, Bernoulli: in press.

[59] Pruitt W. (1985). Sums of Independent Random Variables with the Extreme Terms Excluded, in J.N. Srivastava (eds.), Probability and Statistics: Essays in Honor of Franklin A. Graybill. Elsevier: North Holland.

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

[61] Rio, E. (1993). Covariance Inequalities for Strong Mixing Processes, Ann. de l'Inst. Henri Poincaré B 29, 587-597.

[62] Rio, E. (1995). A Maximal Inequality and Dependent Marcinkiewicz-Zygmund Strong Laws, Ann. Prob. 23, 918-937.

[63] Rootzén, H. (2009). Weak Convergence of the Tail Empirical Function for Dependent Sequences, Stoch. Proc. Appl. 119, 468-490.

[64] Rosenthal, H.P. (1970). On the Subspace Lp (p > 2) Spanned by Sequences of Independent Random Variables, Isreal J. Math. 8, 273-303.

[65] Shao, Q-M. (1995). Maximal Inequalities for Partial Sums of p-mixing Sequences, Ann. Stat. 23, 948-965.

[66] Stigler, S.M. (1973). The Asymptotic Distribution of the Trimmed Mean, Ann. Stat. 1, 472-477.

[67] Teicher, H. (1985). Almost Certain Convergence in Double Arrays, Prob. Theory Rel. Fields 69, 331-345.

[68] von Bahr, B. and C-G. Esséen (1965). Inequalities for the rth Absolute Moment of a Sum of Random Variables, 1≤r≤2, Ann. Math.l Stat. 36, 299-303.

[69] Weiner, D.C. (1991). Center, Scale and Asymptotic Normality for Censored Sums of Independent, Nonidentically Distributed Random Variables, in M.G. Hahn, J. Kuelbs and D.C. Weiner (ed.'s), Sums, Trimmed Sums and Extremes. Birkhäuser: Berlin.

[70] Wu, W.B., W. Min (2005). On linear processes with dependent innovations. Stoch. Proc.. Appl. 115, 939–958.

[71] Yanjiao, M. and L. Zhengyan (2009). Maximal Inequalities and Laws of Large Numbers for Lq-Mixingale Arrays, Stat. Prob. Let. 79, 1539 1547.

[72] Yokoyama, R. (1980). Moment Bounds for Stationary Mixing Sequences, Prob. Theory Rel. Fields 52, 45-57.

Jonathan B. Hill
Dept. of Economics
University of North Carolina - Chapel Hill

jbhill@email.unc.edu
www.unc.edu/∼jbhill