Central Limit Theory for Kernel-Self Normalized Tail-Trimmed Sums of Dependent, Heterogeneous Data with Applications*

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

January 31, 2011

Abstract

Although robust estimation methods were formalized by the late 1800's, data trimming and truncation for non-iid data has received little attention. We present Gaussian central limit theorems for tail-trimmed sums of a heavy tailed weakly dependent process in the Feller class. We assume distribution tails are stationary, and otherwise leave stationarity and heterogeneity unrestricted. The sum is self-normalized with a consistent kernel variance estimator, so the rate of convergence, tail thickness and memory persistence do not need to be specified beyond fairly minimal regularity conditions, hence robust inference is available. The theory applies to mixing and non-mixing processes, including linear and nonlinear distributed lags, and linear and nonlinear random volatility with smooth or non-smooth errors. We show how the results imply asymptotic normality for sample tail-trimmed variances and covariances, and a super-$\sqrt{n}$-convergent least squares estimator for infinite variance autoregressions.

1. INTRODUCTION Despite the rich history of robust estimation dating to Newcomb [61], few results for non-iid data exist, and none permitting non-parametric inference under tail-trimming. We present Gaussian central limit theorems for an intermediate order tail-trimmed sum of weakly dependent data y_t, we develop a consistent non-parametric estimator of the partial sum scale, and apply the theory to robust estimation problems.

Assume y_t is L_p-bounded, $p \in (0, 2)$, measurable on a probability space $(\Omega, \mathcal{G}, P)$, and $E[y_t^2] = \infty$ since the literature covers the finite variance case. We consider those distributions $F_t(y) := P(y_t \leq y)$ for infinite variance y_t in the Feller subclass (Feller [26]):

$$\mathcal{F} = \left\{ F : y_t \sim F \Longrightarrow \liminf_{c \to \infty} \frac{E\left[y_t^2 I\left(|y_t| \leq c\right)\right]}{c^2 P\left(|y_t| > c\right)} > 0 \right\}. \tag{1}$$

Class $\mathcal{F}$ contains all distributions with regularly varying tails and index $\kappa \in (0, 2]$ by Karamata's Theorem, but not slowly varying [s.v.] tails, and it includes distributions of

**AMS 2010 classifications*: primary 62E20, 62G05; secondary 62G35, 62G20.

Keywords and phrases: tail trimmed sum, central limit theorem, telescoping sum, near epoch dependence, kernel variance.

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

iid sequences $\{y_t\}$ with scaled and centered partial sums that have tight nondegenerate limit laws (Feller [26], Pruitt [66]).

The sample path is $\{y_t\}_{t=1}^n$ with sample size $n \geq 1$. Let $y_t^{(-)}$ and $y_t^{(+)}$ denote tail-specific observations

$$y_t^{(-)} := y_t I\left(y_t < 0\right) \quad \text{and} \quad y_t^{(+)} := y_t I\left(y_t \geq 0\right)$$

with order statistics $y_{(1)}^{(-)} \leq y_{(2)}^{(-)} \leq \cdots \leq y_{(n)}^{(-)}$ and $y_{(1)}^{(+)} \geq y_{(2)}^{(+)} \geq \cdots \geq y_{(n)}^{(+)}$. Define the stochastically trimmed triangular array $\{\hat{y}_{n,t}\} = \{\hat{y}_{n,t} : 1 \leq t \leq n\}_{n \geq 1}$ by

$$\hat{y}_{n,t} := y_t \times I\left(y_{(m_{1,n}+1)}^{(-)} \leq y_t \leq y_{(m_{2,n}+1)}^{(+)}\right),$$

where $I(A) = 1$ if A is true and $I(A) = 0$ otherwise, and $\{m_{1,n}, m_{2,n}\}$ are intermediate order sequences: $1 \leq m_{1,n} + m_{2,n} < n$, $m_{i,n} \rightarrow \infty$, and $m_{i,n}/n \rightarrow 0$. Thus $\hat{y}_{n,t} = y_t$ for any y_t between its lower $m_{1,n}/n \rightarrow 0$ and upper $m_{2,n}/n \rightarrow 0$ asymptotic quantiles. If the support is $[0, \infty)$ then $y_t = y_t^{(+)}$ and $\hat{y}_{n,t} := y_t \times I(y_t \leq y_{(m_n+1)}^{(+)})$ for some intermediate order sequence $\{m_n\}$, and so on.

We establish conditions that ensure a kernel self-normalized central limit theorem

$$\frac{\sum_{t=1}^n \left\{\hat{y}_{n,t} - E\left[\hat{y}_{n,t}\right]\right\}}{\sqrt{\sum_{s,t=1}^n k_{n,s,t} \times \{\hat{y}_{n,s} - \hat{y}_n\} \times \{\hat{y}_{n,t} - \hat{y}_n\}}} \xrightarrow{\mathfrak{L}} N(0,1), \tag{2}$$

where $\hat{y}_n := 1/n \sum_{t=1}^n \hat{y}_{n,t}$, $N(0,1)$ denotes a Gaussian law with zero mean and unit variance, and $k_{n,s,t}$ is a kernel function. We do not consider conditions under which $\{\hat{y}_{n,t}\}$ belongs to *any* domain of attraction (e.g. Csörgő et al [13], Hahn and Weiner [38], Pruitt [66], Whalen [74]).

Limit (2) follows from related theory for deterministically trimmed arrays $\{y_{n,t}\}$ defined by

$$y_{n,t} := y_t I\left(-l_n \leq y_t \leq u_n\right),$$

where the positive real sequences $\{l_n, u_n\}$ satisfy $l_n, u_n \rightarrow \infty$ as $n \rightarrow \infty$ and (see Theorem 1.7.13 of Leadbetter et al [55])

$$\frac{n}{m_{1,n}} P\left(y_t \leq -l_n\right) \rightarrow 1 \quad \text{and} \quad \frac{n}{m_{2,n}} P\left(y_t \geq u_n\right) \rightarrow 1. \tag{3}$$

In effect (3) enforces a type of tail stationarity and therefore restricts some forms of trend[1]. We require the restriction because limit theory for intermediate order statistics $y_{(m_{i,n}+1)}^{(\cdot)}$ for the broad class of dependent processes considered here is available only under tail stationarity (e.g. Hill [43], [44], [46], Hsing [49], Rootzén [69]).

We assume $\{y_{n,t}\} = \{y_{n,t} : 1 \leq t \leq n\}_{n \geq 1}$ is geometrically Near Epoch Dependent on an α-mixing process $\{\epsilon_t\}$. See Section 2 for a definition. In theory ϵ_t can be anything, including regression errors or unobserved shocks to a system, while $\epsilon_t = y_t$ is a trivial special case if y_t is α-mixing. We impose NED on $\{y_{n,t}\}$ because potentially non-NED sequences $\{y_t\}$ have NED trimmed arrays $\{y_{n,t}\}$ as shown in Hill [46]; NED is more general than mixing properties and usually easier to verify; it is related to mixingale (McLiesh [57]) and synonymous moment-based properties like L_p-Weak Dependence (Wu and Min [76]); and it is well known that many processes are not mixing (e.g. Andrews [1], Dedecker

[1] Consider stochastic trend $y_t := \sum_{i=1}^{t-1} \epsilon_{t-i}$ where ϵ_t is an iid stable law with tail exponent $\kappa \in (0,2)$ and unit scale. Then y_t is stable with exponent κ and scale $t^{1/\kappa}$, hence thresholds satisfying (3) must be trending (Ibragimov and Linnik [50]). Such trending processes are therefore ruled out here.

et al [24], Guegan and Ladoucette [34]) or must be driven by sufficiently smooth stochastic innovations to be mixing, like nonlinear AR, GARCH and Stochastic Volatility (Carrasco and Chen [8], Meitz and Saikkonen [59]). The majority of results in the limit theory literature concern mixing sequences. See Dedecker and Doukhan [22] and their references.

Key steps in our proofs exploit the NED property and geometric memory, so it is unknown whether our framework extends to other dependence properties or hyperbolic memory (e.g. Bickel and Bühlmann [5], Dedecker and Doukhan [22], Dedecker and Merlevéde [23], Doukhan and Louhichi [25], Hill [43], [44], [46], Peligrad and Utev [64], Wu and Min [76]). Nevertheless, a wide range of processes are covered, including linear or nonlinear autoregressions with linear or nonlinear random volatility errors like ARMA, Asymmetric GARCH, Threshold AR-Threshold GARCH, and Stochastic Volatility.

We prove (2) by first delivering a CLT for NED deterministically tail-trimmed arrays $\{y_{n,t}\}$ in Lemma 3.3:

$$\frac{1}{\sigma_n}\sum_{t=1}^{n}\{y_{n,t}-E[y_{n,t}]\}\stackrel{\mathfrak{L}}{\to}N(0,1)\ \text{ where }\ \sigma_n^2:=E\left[\left(\sum_{t=1}^{n}\{y_{n,t}-E[y_{n,t}]\}\right)^2\right],\tag{4}$$

and then by demonstrating $\hat{y}_{n,t}$ is sufficiently close to $y_{n,t}$ in Lemma 3.4:

$$\sum_{t=1}^{n}\{\hat{y}_{n,t}-y_{n,t}\}=o_p(\sigma_n).\tag{5}$$

Finally, we prove kernel variance consistency $\hat{\sigma}_n^2/\sigma_n^2 := 1/\sigma_n^2\sum_{s,t=1}^{n}k_{n,s,t}\times\{\hat{y}_{n,s}-\hat{y}_n\}\times\{\hat{y}_{n,t}-\hat{y}_n\}\stackrel{p}{\to}1$ in Theorem 4.1.

Nonparametric variance and sandwich covariance estimation focus mainly on stationary L_2-bounded data (e.g. Hall et al [39], Newey and West [62]), extended to nonstationary arrays, spatial data, and dependent tail arrays (Andrews [2], Davidson and de Jong [19], Hill [44], Sampson and Guttorp [71]), with improvements over bias and attention to estimator dispersion (Cribari-Neto et al [12], Kauermann and Carroll [53]). Evidently nonparametric inference has been ignored in the tail-trimming literature, although a tail-trimmed covariance matrix and trimming in nonparametric contexts has been treated (Horowitz [48], Schennach [72]). Since $\sigma_n^2\to\infty$ in general at an unknown rate, we must scale $\hat{\sigma}_n^2/\sigma_n^2$ to prove $\hat{\sigma}_n^2$ and σ_n^2 are similar asymptotically. The issue is irrelevant if y_t is finite dependent since (2) can be greatly simplified. See Section 6.

Our proof of (4) partially resembles the martingale approximation approach of Gordin [29], [30], McLeish [56], [57], extended in Davidson [15], [16], de Jong [21], Hill [43], [44], Peligrad and Utev [64], Wu and Woodroofe [77] and many others. There are, however, crucial differences. First, we show

$$\frac{1}{\sigma_n}\sum_{t=1}^{n}\{y_{n,t}-E[y_{n,t}]\}=\sum_{i=1}^{r_n}\mathcal{W}_{n,i}+o_p(1),\tag{6}$$

for some martingale difference array $\{\mathcal{W}_{n,i},\mathcal{F}_{n,i}\}_{i=1}^{r_n}$ based on telescoping sums of $y_{n,t}$, some array of σ-fields $\{\mathcal{F}_{n,i}\}$, and integer sequence $r_n\to\infty$ and $r_n=o(n)$. This devise is well known in the mixingale and martingale approximation literatures (Davidson [15], [16], de Jong [21], Gordin [29], Hill [43], [44], McLeish [57]). But under tail trimming $\{\mathcal{W}_{n,i}\}$ need not be uniformly L_2-bounded, and at present no restrictions on $\sigma_n\to\infty$ have been levied. Further, we require a sharper bound than McLeish's [57] for a partial sum of a mixingale array.

Second, although we exploit the convenience of geometric memory and tail-stationarity, we do not otherwise restrict heterogeneity. This runs in stark contrast with the existing

NED and mixingale literatures where abstract restrictions on heterogeneity are evidently universally exploited (e.g. Chen and White [9], Davidson [15], [16], de Jong [21], Hill [43], [44], Wooldridge and White [75]).

Finally, partial sums form the cornerstone of dependence estimation and asymptotic theory for minimum distance estimators. In Section 5 we prove asymptotic normality of a self-normalized sample tail-trimmed variance and covariance, and a tail-trimmed least squares estimator of the slope parameter in an infinite variance autoregression.

The tail trimming literature focuses primarily on techniques for proving the central limit property, establishing tightness and characterizing domains of attraction for sums of iid data. See Csörgő et al [13], Csörgő et al [14], Griffin and Pruitt [32], Griffin and Qazi [33], Hahn et al [36], [37], Hahn and Weiner [38], Kim [54], Pruitt [67] and Whalen [74], and consult their references for historical details. Elsewhere tail-trimming is used to ease limit theory arguments since trimmed arrays are bounded (e.g. de Jong [21], Hannan [40], Hill [43], [44]).

Hahn et al [36], [37] deliver a version of (2) for iid data under asymptotically bounded tail truncation: $y_{n,t} := sign\{y_t\} \times \min\{|y_t|, c_n\}/c_n$. Weiner [73] treats independent but non-identically distributed data; Hahn et al [35] work with stationary, strong mixing processes in a Hilbert space; and except for Hahn et al [37], robust variance estimation has not been explored in the tail-trimming literature. Hilbert space rules out infinite variance processes, and a large array of processes are either non-mixing or are not known to be mixing, including infinite order lags of mixing random variables, and linear and random volatility processes with a non-smooth error distribution (e.g. Asymmetric GARCH with non-Gaussian errors). See Davidson [18], Dedecker and Doukhan [22], Gorodetskii [31] and Wu and Min [76].

Berkes et al [4] propose a CUSUM test based on tail-trimming an iid process $\{y_t\}$ in the domain of attraction of a stable law with index $0 < \kappa < 2$. Although we do not tackle weak convergence here, by Feller class (1) and our NED assumption we cover such processes $\{y_t\}$ and far more.

The remainder of the paper proceeds as follows. In Section 2 we define NED and the telescoping sums. Section 3 contains assumptions and the main central limit theorems, and Section 4 establishes kernel variance consistency. We present examples and applications in Section 5 and treat a special case of limited dependence in Section 6.

Throughout K denotes a positive finite constant whose value may change from line to line; similarly $\iota > 0$ is an arbitrarily tiny constant; N is an arbitrary positive integer; ρ is a constant in $(0, 1)$; and $L(n)$ is a s.v. function with $L(n) \to \infty$ and whose value or rate may change from line to line. $(z)_+ := \max\{0, z\}$. $\Sigma_{a_n}^{b_n}(\cdot) = 0$ and $\int_{a_n}^{b_n}(\cdot) = 0$ $\forall n \geq N$ for finite $N \in \mathbb{N}$ if $\liminf_{n\to\infty} a_n/b_n > 1$. $||\cdot||_p$ denotes the usual L_p-matrix norm $||x||_p = (\sum_{i,j} E|x_{i,j}|^p)^{1/p}$, and $||\cdot|| = ||\cdot||_2$. $[z]$ denotes the integer-part of z. $\{a_n\} = \{a_n\}_{n\geq 1}$. $a_n \sim b_n$ signifies $a_n/b_n \to 1$. $x_t \overset{iid}{\sim} (0, 1)$ implies x_t is iid with mean zero and unit variance. $\overset{p}{\to}$ and $\overset{\mathfrak{L}}{\to}$ denote convergence in probability and in law.

2. NED TAIL-TRIMMED ARRAYS

This section defines the required NED property and telescoping sums.

2.1 NEAR EPOCH DEPENDENCE

We require both the trimmed level $\{y_{n,t}\}$ and tail array $\{I(|y_t| > c_n e^u)\}$ for non-stochastic $u \geq 0$ to be L_2-NED on an α-mixing process $\{\epsilon_t\}$. The former ensures central limit (4), and the latter ensures approximation (5). Assume ϵ_t is measurable on the space $(\Omega, \mathcal{G}, P)$ with non decreasing σ-fields $\Im_t := \sigma(\epsilon_\tau : \tau \leq t) \subseteq \Im_{t+1}$, $\sigma(\cup_{t\in\mathbb{Z}}\Im_t) \subseteq \mathcal{G}$, and write

$$\Im_a^b := \sigma\left(\epsilon_\tau : a \leq \tau \leq b\right).$$

Recall the definition of α-mixing coefficients with *size* $\theta > 0$ (Ibragimov and Linnik [50], Rosenblatt [70],):

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

where $\iota > 0$ is infinitessimal[2]. Geometric α-mixing implies $\alpha_d = O(\rho^d)$ for some $\rho \in (0,1)$, hence size θ is arbitrary. The property is weaker than geometric ergodicity, β-, ϕ-, ρ-, and s-mixing (Dedecker and Doukhan [22], Dedecker et al [24]) and other weak dependence properties (e.g. Doukhan and Louhichi [25])

The array $\{y_{n,t}\}$ is L_2-Near Epoch Dependent on $\{\Im_t\}$ or on $\{\epsilon_t\}$ if for some array of non-negative non-stochastic *constants* $\{e_{n,t}^*\}$, any sequence of finite positive integer *displacements* $\{d_n\}$ that are bounded $\inf_{n\to\infty} d_n = d \geq 1$, and a sequence of positive *coefficients* $\{\psi_{d_n}\}$, $\psi_{d_n} \to 0$ monotonically as $d_n \to \infty$,

$$\left\|y_{n,t} - E\left[y_{n,t}|\Im_{t-d_n}^{t+d_n}\right]\right\| \leq e_{n,t}^*\psi_{d_n}. \tag{7}$$

If $\psi_{d_n} = O(d_n^{-\lambda-\iota})$ for some $\lambda > 0$ and tiny $\iota > 0$ we call λ the *size*, and $\{y_{n,t}\}$ is *geometrically* NED when $\psi_{d_n} = O(\rho^{d_n})$ for some $\rho \in (0,1)$. If y_t is $\Im_{t-h}^{t+h}$-measurable for finite $h \in \mathbb{N}$ then $\{y_{n,t}\}$ is trivially L_2-NED with $\psi_{d_n} = o(d_n^{-\lambda})$ for any $\lambda > 0$. By construction $||y_{n,t} - E[y_{n,t}|\Im_{t-d_n}^{t+d_n}]|| \leq 2c_n$, so we can always assume bounded constants

$$e_{n,t}^* \leq Kc_n.$$

The NED property dates in some form to Ibragimov and Linnik [50] and McLeish [57], [58], and was evidently coined in Gallant and White [27]. The idea is $y_{n,t}$ is perfectly predicted in L_p-norm by the near-epoch $\{\epsilon_\tau\}_{t-d_n}^{t+d_n}$ as $d_n \to \infty$. NED is a version of mixingale (McLeish [57]), which is a special case of L_p-Weak Dependence (Wu and Min [76]), and related to s-weak dependence (Dedecker and Doukhan [22]). See Davidson [17], Dedecker et al [24], Hill [43], [44], [46] and Nze and Doukhan [63] for historical notes.

By convention L_2-NED is defined as $||y_{n,t} - E[y_{n,t}|\Im_{t-d}^{t+d}]|| \leq e_{n,t}^*\psi_d$ for an integer constant d (e.g. Davidson [17], Gallant and White [27]). Although our construction (7) is identical in structure, the displacement sequence $\{d_n\}$ provides a non-trivial increase in generality. Since d_n is arbitrary we can render irrelevant the notion of heterogeneity through $e_{n,t}$ since for tiny $\delta \in (0,\iota)$

$$\left\|y_{n,t} - E\left[y_{n,t}|\Im_{t-d_n}^{t+d_n}\right]\right\|_p \leq \frac{e_{n,t}^*}{\ln d_n} \times (\ln d_n) \times \psi_{d_n} \leq K \times O\left(d_n^{-\lambda-\iota+\delta}\right) \tag{8}$$

for any displacement sequence that satisfies $d_n/\exp\{\max_{1\leq t\leq n}\{e_{n,t}^*\}\} \to \infty$. Property (8) offers a key observation since $\{d_n\}$ is used solely to measure memory for the sake of proofs, and is therefore *controlled by the analyst*. We can always displace to the deep past or distant future faster than the accumulation of heterogeneous characteristics of the data generating process as measured by $e_{n,t}^*$. Thus, although we implicitly impose tail stationarity through (3), we do not otherwise need to restrict heterogeneity. See Section 3 for all assumptions. See Hill [46] for the first use of NED properties (7) and (8)

[2] There is a not a universal convention for representing the rate of memory decay (see, e.g., Davidson [17], Ibragomiv and Linnik [50], McLeish [57], Nze and Doukhan [63]). In this paper a sequence $\{a_d\}_{d\in\mathbb{N}}$ has size $\theta > 0$ when $a_d = O(d^{-\theta-\iota})$ for tiny $\iota > 0$.

to catalogue dependence properties for negligibly trimmed arrays of NED and possibly non-NED sequences $\{y_t\}$.

It is well known NED with an α-mixing base is linked to the mixingale property in general (e.g. Davidson [17]: Theorem 17.5). In this paper we focus on geometric memory.

LEMMA 2.1 (NED-MIXINGALE). *If $\{y_{n,t}\}$ is geometrically NED on geometrically α-mixing $\{\epsilon_t\}$ then $\{y_{n,t}, \Im_t\}$ forms a geometric L_2-mixingale array with constants $e_{n,t} \leq Kc_n$ and coefficients $\zeta_{d_n} = O(\rho^{d_n})$, cf. McLeish [57]:*

$$\left\| E\left[y_{n,t}\right] - E\left[y_{n,t}|\Im_{t-d_n}\right]\right\|_p \leq e_{n,t}\zeta_{d_n} \quad \text{and} \quad \left\|y_{n,t} - E\left[y_{n,t}|\Im_{t+d_n}\right]\right\|_p \leq e_{n,t}\zeta_{d_n+1}. \tag{9}$$

2.2 TELESCOPING SUMS

We require some compact representations. Write the threshold and quantile sequences

$$c_n := \max\left\{l_n, u_n\right\} \quad \text{and} \quad m_n := \min\left\{m_{1,n}, m_{2,n}\right\},$$

and define standardized trimmed arrays:

$$z_{n,t} := \frac{y_{n,t} - E\left[y_{n,t}\right]}{\sigma_n} \quad \text{and} \quad \hat{z}_{n,t} := \frac{\hat{y}_{n,t} - E\left[\hat{y}_{n,t}\right]}{\sigma_n}.$$

Define sequences of positive real numbers $\{h_n,\ j_n,\ r_n\}$ used to form telescoping or Bernstein sums of $z_{n,t}$: assume $h_n \to \infty$, $j_n \to \infty$, and

$$r_n = [n/h_n],\ 1 \leq\ j_n \leq h_n \ \text{ and } j_n = o\left(h_n\right).$$

Now define an array of σ-fields $\{\mathcal{F}_{n,i}\}$: for any sequence of positive finite numbers $\{g_n\}$, $g_n \to \infty$,

$$\mathcal{F}_{n,i} := \sigma\left(\epsilon_t : -g_n \leq t \leq ih_n\right) = \Im_{-g_n}^{ih_n}, \text{ for } i = 1, ..., r_n,$$

and define telescoping sums of $z_{n,t}$, and a martingale difference:

$$\mathcal{Z}_{n,i} := \sum_{t=(i-1)h_n+j_n+1}^{ih_n} z_{n,t} \text{ and } \mathcal{W}_{n,i} := E\left[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}\right] - E\left[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}\right]. \tag{10}$$

Thus h_n is the block size, r_n the number of blocks, and j_n the buffer between blocks.

Trivially we have the decomposition

$$\begin{aligned}\sum_{t=1}^{n} z_{n,t} \quad = \quad & \sum_{i=1}^{r_n} \mathcal{W}_{n,i} + \sum_{i=1}^{r_n} \left(\mathcal{Z}_{n,i} - E\left[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}\right]\right) + \sum_{i=1}^{r_n} E\left[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}\right] \\ & + \sum_{i=1}^{r_n} \sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_{n,t} + \sum_{t=r_n h_n+1}^{n} z_{n,t}.\end{aligned} \tag{11}$$

The bulk of effort rests on showing $\sum_{t=1}^{n} z_{n,t} = \sum_{i=1}^{r_n} \mathcal{W}_{n,i} + o_p(1)$ as in (6). The remaining martingale difference central limit theorem $\sum_{i=1}^{r_n} \mathcal{W}_{n,i} \xrightarrow{\mathfrak{L}} N(0,1)$ is relatively straightforward.

By construction $\{W_{n,i}, \mathcal{F}_{n,i} : 1 \leq i \leq r_n\}_{n\geq1}$ forms a martingale difference array based on the information $\mathcal{F}_{n,i} = \Im_{-g_n}^{ih_n}$. The standard setup in the literature exploits $\mathcal{F}_{n,i} = \Im_{-\infty}^{ih_n}$ (e.g. Davidson [16], de Jong [21], Gordin [29], Hill [43], [44]. Our use of a finite end-point $-g_n > -\infty$ implies measurable functions of $E\left[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}\right]$ and $E\left[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}\right]$ are α-mixing, reducing complications associated with the NED property when proving a martingale difference central limit theorem. The burden of proving limit theorems for mixingale-like processes has been noted elsewhere (e.g. Dedecker and Doukhan [22]).

2.3 PARTIAL SUM VARIANCE BOUND

Hill [45: Theorem 2.1] presents an improved partial sum variance bound for mixingale arrays $\{y_{n,t}\}$ that exploits $d_n \to \infty$ as $n \to \infty$ in a mixingale version of (8). We use the bound, stated below, to prove the martingale difference decomposition (6) and develop key central limit arguments. See Hill [45] for a discussion on the relative improvement over similar bounds, especially when y_t has an infinite variance.

LEMMA 2.2 (PARTIAL SUM VARIANCE BOUND). *If $\{y_{n,t}, \Im_t\}$ is a zero mean L_2-mixingale array (9) with coefficients $\zeta_{d_n} = O(d_n^{-\lambda-\iota})$ of size $\lambda = 1/2$ then $E[(\sum_{t=1}^n y_{n,t})^2] \leq K \sum_{t=1}^n E\left[y_{n,t}^2\right]$.*

3. CLT FOR STOCHASTICALLY TAIL-TRIMMED ARRAYS

We first state required assumptions, and then present the main results. The martingale difference $\{\mathcal{W}_{n,i}, \mathcal{F}_{n,i}\}$ defined in (10) is bounded by construction:

$$|\mathcal{W}_{n,i}| \leq K \frac{h_n c_n}{\sigma_n}.$$

Thus, McLeish's [56: Theorem 2.3] sufficient conditions for asymptotic normality, specifically $\max_{1 \leq i \leq r_n}\{|\mathcal{W}_{n,i}|\}$ is uniformly L_2-bounded and $o_p(1)$ and $\sum_{i=1}^{r_n} \mathcal{W}_{n,i}^2 \xrightarrow{p} 1$, are in part assured by restricting the thresholds $\{l_n, u_n\}$. This follows since if

$$c_n := \max\{l_n, u_n\} \to \infty$$

too quickly because we are trimming too few observations (i.e. $m_n \to \infty$ too slowly), then too many large values may enter into $\mathcal{W}_{n,i}$ and therefore $\sum_{i=1}^{r_n} \mathcal{W}_{n,i}^2$ may diverge.

Random variables y_t that belong to Feller subclass (1) have intermediate order thresholds $\{l_n, u_n\}$ that satisfy a sequence of fixed point bounds, implying c_n is naturally bounded and $\{y_{n,t}\}$ is relatively stable (cf. Gnedenko [28]). We state an expanded result here for use in the sequel. Consult the appendix for omitted proofs.

LEMMA 3.1 (RELATIVE STABILITY). *Under (1) and (3) $c_n = O((\sum_{t=1}^n E[y_{n,t}^2])^{1/2}/m_n^{1/2})$ hence*

$$\max_{1 \leq t \leq n}\left\{\frac{|y_{n,t}|}{\|y_{n,t}\|}\right\} = o_p(n^{1/2}). \tag{12}$$

In non-trimming contexts (12) applies to a large class of finite variance sequences $\{y_t\}$ that are stationary with weakly dependent maxima (e.g. Naveau [60]). Further, (12) aligns with a necessary and sufficient condition for the distribution limit of a sum of an iid array to be Gaussian (e.g. Kallenberg [52: Theorem 5.15]). Thus, intermediate order trimming (3) offers a powerful means to ensure domain of attraction membership (cf. Hahn and Weiner [38: p.457]).

ASSUMPTION A (non-degeneracy and fractiles) *i.* $\liminf_{n\to\infty} \min_{1\leq t\leq n}\{E[y_{n,t}^2]\} > 0$; *ii.* $\sum_{t=1}^n E[y_{n,t}^2]/\sigma_n^2 = O(1)$; *and iii.* $m_n/(\ln(n))^{2+\iota}$ *for tiny* $\iota > 0$.

Remark 1: Property (*i*) rules out degenerate cases asymptotically in lieu of trimming, while (*ii*) eases bounding σ_n^2 since we only know $\sigma_n^2 \leq K \sum_{t=1}^n E[y_{n,t}^2]$ by Lemma 2.2. Asymptotically (*ii*) effectively rules out maximum negative serial correlation in $y_{n,t}$ where $\limsup_{n\to\infty} \sigma_n^2 < \infty$. Both (*i*) and (*ii*) implicitly restrict trend non-stationarity in accordance with threshold and fractile property (3), and are examined closely in Section 5.1.

Remark 2: Property (ii) and Lemma 3.1 ensure the thresholds are bounded

$$c_n = O\left(\left(\sum_{t=1}^{n} E[y_{n,t}^2]\right)^{1/2} / m_n^{1/2}\right) = O\left(\sigma_n / m_n^{1/2}\right).$$

The role (1) and (3) and therefore Lemma 3.1 ultimately play is remarkably subtle. By restricting the block size $h_n = o(m_n^{1/2})$ in the appendix, we use $c_n = O(\sigma_n/m_n^{1/2})$ to ensure the deterministic upper bound $\max_{1\leq i\leq r_n} |\mathcal{W}_{n,i}| \leq K h_n c_n/\sigma_n = O(h_n/m_n^{1/2}) = o(1)$, a key step toward proving $\sum_{i=1}^{r_n} \mathcal{W}_{n,i} \overset{\mathfrak{L}}{\to} N(0,1)$. Notice in Lemma 3.1 we cannot prove directly $c_n = O(\sigma_n/m_n^{1/2})$ without also imposing Assumption A.ii. Thus, for self-normalized CLT (4) evidently Assumption A.ii is indispensable.

Remark 3: Property (iii) is required to ensure sufficiently many observations are trimmed. In the appendix we show when memory is non-trivially geometric (i.e. not q-dependent for finite q), the block buffer must satisfy $j_n/\ln(n) \to \infty$ to under geometric memory and heavy tails to overcome cross-block dependence, and to ensures blocks are sufficiently separated to bound dispersion in sums of $\mathcal{Z}_{n,i}$. But since $j_n/h_n \to 0$ must hold to ensure the buffers are negligible, and $h_n/m_n^{1/2} \to 0$ to ensure $\max_{1\leq i\leq r_n} |\mathcal{W}_{n,i}| \leq K h_n c_n/\sigma_n = o(1)$, we must assume $m_n \to \infty$ *faster* than $(\ln(n))^2$. Although (iii) is not the only way to enforce this balance, it is obviously convenient.

Remark 4: Under (ii) the convergence rate n/σ_n of the tail-trimmed mean $1/n \sum_{t=1}^{n}\{y_{n,t} - E[y_{n,t}]\}$ is no greater than $n^{1/2}$, and $o(n^{1/2})$ in the infinite variance case where $E[y_{n,t}^2] \to \infty$. This is obvious in the iid case since $n/\sigma_n = n/(n^{1/2}||y_{n,t}||) = o(n^{1/2})$, cf. Hahn et al [36], but it clarifies a key implication of (ii) for non-iid data, an issue we discuss in Section 5.

Both arrays $\{y_{n,t}\}$ and $\{I(|y_t| > c_n e^u)\}$ are assumed to be NED.

ASSUMPTION B (NED tail-trimmed array) $\{y_{n,t}\}$ *is geometrically* L_2*-NED on a geometrically* α*-mixing base* $\{\epsilon_t\}$.

Let z_t denote $|y_t|$, $-y_t I(y_t < 0)$ or $y_t I(y_t \geq 0)$ with associated threshold and quantile sequences $\{\tilde{c}_n, \tilde{m}_n\}$ that satisfy $1 \leq \tilde{m}_n < n$, $\tilde{m}_n \to \infty$, $\tilde{m}_n = o(n)$ and $(n/\tilde{m}_n)P(z_t > \tilde{c}_n) \to 1$.

ASSUMPTION C (Extremal-NED) *The tail event* $\{I(z_t > \tilde{c}_n e^u)\}$ *is geometrically* L_2*-NED on a geometrically* α*-mixing base* $\{\epsilon_t\}$ *with constants* $e_{n,t}^*(u)$ *that are Lebesgue integrable on* $\mathbb{R}_+$ *and* $\max_{1\leq t\leq n} \sup_{u\geq 0}\{e_{n,t}^*(u)\} = K(\tilde{m}_n/n)^{1/2}$.

Remark 1: Hill [43], [44], [46] first proposed Extremal-NED to characterize the limit distribution of tail estimators. The property leads to minimal rates of convergence $m_{i,n}^{1/2}$ for $y_{(m_{1,n}+1)}^{(-)}/l_n \overset{p}{\to} 1$ and $y_{(m_{2,n}+1)}^{(+)}/u_n \overset{p}{\to} 1$ for a massive array of dependent, heterogeneous processes, which we use to prove approximation (5).

Remark 2: We do not restrict NED heterogeneity for $y_{n,t}$ since we do not impose bounds nor summability conditions on the NED constants $e_{n,t}^*$. This is allowed since the partial sum variance bound Lemma 2.2 does not use mixingale constants $e_{n,t}$, while the standard in the literature entails summability conditions for $e_{n,t}^*$ to restrict heterogeneity (Davidson [15], [16], de Jong [21], Hill [43], [44], Jenish and Prucha [51], Qiu and Lin [68], Wooldridge and White [75]). The constants $e_{n,t}^*(u)$ for $I(z_t > \tilde{c}_n e^u)$, however, are intrinsically bounded due to $I(z_t > \tilde{c}_n e^u) \in \{0,1\}$.

The main result of this section is a central limit theorem for stochastically trimmed $\{\hat{y}_{n,t}\}$, the first step towards (2).

THEOREM 3.2 (CLT-ST). *Under Assumptions A-C* $1/\sigma_n \sum_{t=1}^{n} \{\hat{y}_{n,t} - E[\hat{y}_{n,t}]\} \stackrel{\mathfrak{L}}{\rightarrow} N(0,1)$.

Theorem 3.2 follows from a CLT for deterministically trimmed arrays $\{y_{n,t}\}$, and a weak asymptotic approximation between $\hat{y}_{n,t}$ and $y_{n,t}$.

LEMMA 3.3 (CLT-DT). *Under Assumptions A and B* $1/\sigma_n \sum_{t=1}^{n} \{y_{n,t} - E[y_{n,t}]\} \stackrel{\mathfrak{L}}{\rightarrow} N(0,1)$.

We present a general approximation result that will be used to characterize the kernel variance estimator.

LEMMA 3.4 (APPROXIMATION). *Let* $\{w_{n,t}\}$ *be any* $\mathcal{G}$*-measurable stochastic triangular array that satisfies* $\max_{1 \leq t \leq n}\{|w_{n,t}|\} = O_p(1)$. *Under Assumptions A and C* $\sum_{t=1}^{n} w_{n,t}\{\hat{y}_{n,t} - y_{n,t}\} = O_p(c_n L(n)) = o_p(\sigma_n)$ *for s.v.* $L(n) \rightarrow \infty$.

PROOF OF THEOREM 3.2. Apply Lemmas 3.3 and 3.4 to deduce $1/\sigma_n \sum_{t=1}^{n} \{\hat{y}_{n,t} - E[y_{n,t}]\} \stackrel{\mathfrak{L}}{\rightarrow} N(0,1)$. But this implies $1/\sigma_n \sum_{t=1}^{n} \{\hat{y}_{n,t} - E[\hat{y}_{n,t}]\} \stackrel{\mathfrak{L}}{\rightarrow} N(0,1)$ and $1/\sigma_n \sum_{t=1}^{n} \{E[\hat{y}_{n,t} - y_{n,t}]\} = o(1)$ by the Helly-Bray theorem. $\mathcal{QED}$.

4. KERNEL VARIANCE AND MAIN RESULT

The last step towards (2) involves estimating σ_n^2. Define a kernel function $k : \mathbb{R} \rightarrow [-1, 1]$, and write

$$k_{n,s,t} := k\left((s-t)/\gamma_n\right)$$

for some bandwidth $\gamma_n \rightarrow \infty$ that satisfies $\gamma_n = o(n)$. Recall $\hat{y}_n := 1/n \sum_{t=1}^{n} \hat{y}_{n,t}$ and define the variance estimator

$$\hat{\sigma}_n^2 := \sum_{s,t=1}^{n} k_{n,s,t} \times \{\hat{y}_{n,s} - \hat{y}_n\} \times \{\hat{y}_{n,t} - \hat{y}_n\}.$$

Classic treatments under mixing and other weak dependence properties can be found in Andrews [2], Hansen [41], and Newey and West [62]. Davidson and de Jong's [19] utilization of NED and mixingale theory forms the basis of our argument. The kernel function $k(\cdot)$ and bandwidth γ_n are restricted as follows.

ASSUMPTION D

i. k *is a member of class* $\mathcal{K}$, *where*

$$\mathcal{K} = \{k : \mathbb{R} \rightarrow [-1,1] \mid k(0) = 1,\ k(x) = k(-x)\ \forall x \in \mathbb{R},\ k \text{ is integrable}, \int_{-\infty}^{\infty} |k(x)|dx < \infty,\ \int_{-\infty}^{\infty} |\varpi(\xi)|d\xi < \infty, k(\cdot) \text{ is continuous at } 0 \text{ and all but a finite number of points}\},$$

and $\varpi(\xi) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x)e^{i\xi x}dx < \infty$.

ii. $\max_{1 \leq s,t \leq n} |k_{n,s,t}| = O(1)$, $\sum_{s,t=1}^{n} |k_{n,s,t}| = o(n^2)$, $\sup_{1 \leq s \leq n} \sum_{t=1}^{n} |k_{n,s,t}| = o(n)$, $\gamma_n \rightarrow \infty$ *and* $\gamma_n = o(n)$.

Remark 1: Class $\mathcal{K}$ includes Bartlett, Parzen, Quadratic Spectral, Tukey-Hanning and other kernels. See Davidson and de Jong [19] and their citations.

Remark 2: The Bartlett kernel is $k(z) = \max\{0, 1 - |z|\} \geq 0$. If $\gamma_n = o(n)$ then $\sum_{s,t=1}^{n} |k_{n,s,t}| = o(n^2)$ and $\sup_{1 \leq s \leq n} \sum_{t=1}^{n} |k_{n,s,t}| = o(n)$ are immediate. Similarly, the

Parzen kernel is $k(z) = 1 - 6z^2 + 6|z|^3$ for $0 \leq |z| \leq 1/2$, $k(z) = 2(1 - |z|)^3$ for $1/2 \leq |z| \leq 1$, and $k(z) = 0$ otherwise. Since $|k(z)| \leq K \max\{0, 1 - |z|\}$, if $\gamma_n = o(n)$ then the remaining bounds under Assumption D.ii hold. Assumption D appears to provide one of the least restrictive environments available in the HAC literature (see Davidson and de Jong [19] and their references).

Theorem 4.1 (Kernel Variance). *Under Assumptions A-D* $|\hat{\sigma}_n^2/\sigma_n^2 - 1| \xrightarrow{p} 0$.

Theorems 3.2 and 4.1 together imply the main result (2).

Theorem 4.2 (Main Result). *Under Assumptions A-D* $1/\hat{\sigma}_n^2 \sum_{t=1}^n \{\hat{y}_{n,t} - E[\hat{y}_{n,t}]\} \xrightarrow{\mathfrak{L}} N(0,1)$.

5. ASSUMPTIONS AND EXAMPLES

We study in Section 5.1 data generating processes that ensure non-degeneracy Assumption A.i,ii. Throughout let $m_n/(\ln(n))^{2+\iota} \to \infty$ such that Assumption A.iii holds. NED Assumptions B and C are then verified in Section 5.2 for a variety of sequences $\{y_t\}$. Finally, Section 5.3 contains examples covering all Assumption A-C.

Assume y_t for each $t \in \mathbb{Z}$ is L_p-bounded $p > 0$, with unbounded upper support

$$0 < P(y_t \leq y) < 1 \ \forall y \in [a, \infty) \text{ and some } a \in \mathbb{R}. \tag{13}$$

Note we allow for strictly positive support $P(y_t \geq 0) = 1$. Also assume y_t is symmetrically trimmed for simplicity, with a two-tailed intermediate order sequence $\{m_n\}$ and thresholds $\{c_n\}$: for all $t \in \mathbb{Z}$

$$y_{n,t} = y_t I(|y_t| \leq c_n) \quad \text{where} \quad \frac{n}{m_n} P(|y_t| > c_n) \to 1.$$

As always $\{\Im_t\}$ denotes a sequence of σ-fields induced by an α-mixing sequence $\{\epsilon_t\}$.

5.1 ASSUMPTION A: NON-DEGENERACY

Assume y_t has a probability density $f_t(y) := (\partial/\partial y)P(y_t \leq y)$. Then (13) implies $\inf_{y \in \mathcal{S}} \min_{t \in \mathbb{Z}}\{f_t(y)\} > 0$ on some subset $\mathcal{S} \subset \mathbb{R}$ with positive Lebesgue measure, which suffices for Assumption A.i.

Assumption A.i: Use $0 < P(y_t \leq y) < 1 \ \forall y \in [a, \infty)$ and $m_n = o(n)$ to deduce $c_n \to \infty$. Therefore there exists $N \geq 1$ such that $\mathcal{S} \subset [0, \inf_{n \geq N} c_n]$, and

$$\inf_{n \geq N} \left\{ \min_{1 \leq t \leq n} \left\{ E\left[y_{n,t}^2\right] \right\} \right\} = \int_0^{\inf_{n \geq N}\{c_n\}} y^2 \min_{1 \leq t \leq n} \{f_t(dy)\} \geq \int_{y \in \mathcal{S}} |y| \min_{t \in \mathbb{Z}} \{f_t(dy)\} > 0.$$

Hence Assumption A.i holds.

If we additionally assume negative serial association in $y_{n,t}$ does not dominate in a sense defined below, then Assumption A.ii also holds.

Assumption A.ii: A simple example provides the intuition. Any two zero mean random variables $\{X, Y\}$ with finite variances $\{\sigma_x^2, \sigma_y^2\}$ and perfect negative linear dependence satisfy $E(X + Y)^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y = (\sigma_x - \sigma_y)^2$. If $\sigma_x = \sigma_y$ then $E(X + Y)^2 = 0$. Translated to $\sum_{t=1}^n y_{n,t}$ we therefore must restrict negative serial correlation in $y_{n,t}$. This is non-trivial since $E[y_{n,s}y_{n,t}]$ may not exist asymptotically and may depend on s and t.

Assume $E[y_{n,t}] = 0$ to reduce notation, and define for $n \geq N$

$$\mathcal{R}_n := \sum_{s=1}^{n-1} \sum_{t=s+1}^{n} \frac{\left(E\left[y_{n,s}^2\right]\right)^{1/2} \left(E\left[y_{n,t}^2\right]\right)^{1/2}}{\sum_{r=1}^n E\left[y_{n,r}^2\right]} p_n(s,t) \quad \text{where} \quad p_n(s,t) := \frac{E\left[y_{n,s}y_{n,t}\right]}{\left(E\left[y_{n,s}^2\right]\right)^{1/2} \left(E\left[y_{n,t}^2\right]\right)^{1/2}}.$$

Since Assumption A.i holds the ratios are well defined for $n \geq N$.

The serial correlations $p_n(s,t)$ satisfy $|p_n(s,t)| \leq 1 \; \forall n$ by the Cauchy-Schwarz inequality. Since by construction

$$0 \leq \sigma_n^2 = \sum_{t=1}^{n} E\left[y_{n,t}^2\right] \times (1 + 2\mathcal{R}_n),$$

and $\liminf_{n\to\infty} \sum_{t=1}^{n} E\left[y_{n,t}^2\right] > 0$ under Assumption A.i, trivially $\liminf_{n\to\infty} \mathcal{R}_n \geq -1/2$. Assumption A.ii $n \max_{1\leq t\leq n}\{E[y_{n,t}^2]\}/\sigma_n^2 = O(1)$ is therefore satisfied under Assumption A.i and "*sufficient positive association*" $\liminf_{n\to\infty} \mathcal{R}_n > -1/2$. Obviously $\liminf_{n\to\infty} \mathcal{R}_n > -1/2$ is trivial if y_t exhibits positive association of any magnitude at all displacements $\liminf_{n\to\infty} \inf_{s,t\geq 1} p_n(s,t) \geq 0$. The latter can be generalized to cases where sufficiently many $p_n(s,t) > 0$ and few $p_n(s,t) < 0$ as $n \to \infty$ with attention to their relative magnitudes.

EXAMPLE 1 (Martingale Difference): If $\{y_{n,t}, \Im_t\}$ is an adapted martingale difference array then $\mathcal{R}_n = 0 > -1/2$ for each n. Consider the special case $y_t = u_t x_{t-1}$ where u_t is an iid random variable symmetrically distributed at zero and independent of some random variable x_{t-1}. This covers regression contexts discussed in Example 3, below. Under symmetric trimming $y_{n,t}$ is a martingale difference relative to $\Im_t := \sigma(\{u_\tau, x_\tau\} : \tau \leq t)$ since

$$E\left[u_t I\left(|u_t| \times |x_{t-1}| \leq c_n\right)|\Im_{t-1}\right] = E\left[u_t I\left(|u_t| \times |x_{t-1}| \leq c_n\right)|x_{t-1}, x_{t-2}, ...\right] = 0,$$

hence $E[y_{n,t}|\Im_{t-1}] = x_{t-1}E[u_t I(|u_t x_{t-1}| \leq c_n)|\Im_{t-1}] = 0$.

EXAMPLE 2 (Positive Definiteness): By construction $a_{n,t} := (E[y_{n,t}^2])^{1/2}/(\sum_{t=1}^{n} E[y_{n,t}^2])^{1/2}$ is well defined under Assumption A.i, and $\sum_{t=1}^{n} a_{n,t}^2 = 1$. As long as $p_n(s,t)$ is positive definite and Assumption A.i holds then for all $n \geq N$ and some $N \in \mathbb{N}$

$$\frac{\sigma_n^2}{\sum_{t=1}^{n} E\left[y_{n,t}^2\right]} = \sum_{s,t=1}^{n} a_{n,s} a_{n,t} p_n(s,t) > 0$$

hence $\sum_{t=1}^{n} E\left[y_{n,t}^2\right]/\sigma_n^2 = O(1)$, or $\liminf_{n\to\infty} \mathcal{R}_n > -1/2$.

Linear processes with short-range memory have positive definite $p_n(s,t)$. Define $y_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$, $\psi_0 = 1$, $\psi_i = O(\rho^i)$, where iid ϵ_t has a probability density function symmetric at zero and positive on $\mathbb{R}$, with regularly varying tails:

$$P\left(|\epsilon_t| > \epsilon\right) = \epsilon^{-\kappa} L\left(\epsilon\right) \text{ where } \kappa \in (0,2] \text{ and } L\left(\epsilon\right) \text{ is s.v.}$$

Then $\{y_t\}$ has tail $P(|y_t| > y) \sim \sum_{i=0}^{\infty} |\psi_i|^\kappa y^{-\kappa} L\left(y\right)$ and therefore belongs to class (1). See Brockwell and Cline [6], cf. Ibragimov and Linnik [50].

LEMMA 5.1 (LINEAR PROCESSES). *If $y_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ with $\psi_i = O(\rho^i)$ and ϵ_t defined above then $\inf_{r'r=1} r'[p_n(s,t)]_{s,t=1}^{N} r > 0$ for all $n \geq N$ and some $N \in \mathbb{N}$, where $r \in \mathbb{R}^N$. Therefore Assumptions A.i,ii hold.*

5.2 ASSUMPTIONS B AND C: NED TAIL and NON-TAIL ARRAYS

Hill [44], [46] shows NED sequences $\{y_t\}$ have NED tail-arrays $\{I(|y_t| > c_n e^u)\}$ and tail-trimmed arrays $\{y_{n,t}\}$. Recall $\{y_t\}$ is L_p-NED on $\{\Im_t\}$ with size λ if for constants e_t^* and coefficients $\psi_d = o(d^{-\lambda})$:

$$\left\| y_t - E\left[y_t|\Im_{t-d}^{t+d}\right]\right\| \leq e_t^* \psi_d.$$

See Gallant and White [27] and Nze and Doukhan [63] for details.

THEOREM 5.2 (CLT-NED). *Let $\{y_t\}$ be geometrically L_p-NED, $p > 0$, on a geometrically α-mixing base $\{\epsilon_t\}$.*

a. Assumption B holds: $\{y_{n,t}\}$ is geometrically L_2-NED on $\{\epsilon_t\}$ with constants $e^*_{n,t}$;

b. Assumption C holds: $\{I(|y_t| > c_n e^u)\}$ is geometrically L_2-NED on $\{\epsilon_t\}$ with constants $e^*_{n,t}(u) = K(m_n/n)^{1/2}e^{-u/2}$;

c. If additionally Assumption A holds then $1/\sigma_n \sum_{t=1}^n \{\hat{y}_{n,t} - E[\hat{y}_{n,t}]\} \xrightarrow{\mathfrak{L}} N(0,1)$.

PROOF. Claim (c) follows from (a), (b) and CLT-ST Theorem 3.2. Claims (a) and (b) are verified in Theorem 2.1 and Corollary 3.5 of Hill [46]. $\mathcal{QED}$.

Remark: Tail-trimmed arrays $\{y_{n,t}\}$ and tail arrays $\{I(|y_t| > c_n e^u)\}$ may be L_2-NED even when y_t is neither L_p-NED nor mixing. Integrated and Explosive GARCH with non-smoothly distributed errors are two cases discussed in Example 7 below.

If y_t is a finite lag function of a $\mathcal{G}$-measurable geometrically α-mixing process, then Assumptions B and C are trivial. Simply define $\epsilon_t := y_t$ in Theorem 5.2.

COROLLARY 5.3 (MIXING). *Assumptions B and C hold for geometric α-mixing sequences $\{y_t\}$.*

The preceding discussions can be summarized as follows.

THEOREM 5.4 (ASSUMPTIONS A-C). *Let $m/(\ln(n))^{2+\iota} \to \infty$ and $0 < P(y_t \leq y) < 1$ $\forall y \in [a, \infty)$, $a \in \mathbb{R}$. Assumptions A-C hold under the following conditions :*

i. y_t has a density $f_t(y) = (\partial/\partial y)P(y_t \leq y)$;

ii. $y_{n,t}$ has positive definite serial correlations (e.g. $y_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$, for iid ϵ_t with regularly varying tails, $\psi_0 = 1$ and $\psi_i = O(\rho^i)$);

iii. $\{y_t\}$ is geometrically L_p-NED, $p > 0$, on a geometrically α-mixing base $\{\epsilon_t\}$.

5.3 EXAMPLES

The remainder of this section contains an application to Tail-Trimmed Least Squares, and examples of processes that satisfy Assumptions A-C. Let $Y_{h,t}$ denote an arbitrary $\sigma(y_t, ..., y_{t-h})$-measurable random variable for $h \in \mathbb{N}$. In many cases we impose symmetry to simplify trimming notation. As before assume $m_n/(\ln(n))^{2+\iota} \to \infty$.

EXAMPLE 3 (Tail-Trimmed Least Squares): The following is fully treated in Hill and Renault [47]. Consider estimating the slope parameter θ^0 in a heavy-tailed stationary AR(1)

$$x_t = \theta^0 x_{t-1} + u_t, \ \ |\theta^0| < 1, \ \ u_t \text{ is iid}, \ \ E[u_t] = 0.$$

Assume u_t is symmetrically distributed, $0 < P(u_t \leq u) < 1$ $\forall u \in \mathbb{R}$, $P(u_t \leq u)$ is absolutely continuous on $\mathbb{R}$-a.e. with a density $(\partial/\partial u)P(u_t \leq u) > 0$ $\forall u \in \mathbb{R}$, and power-law tail

$$P(|u_t| > u) = du^{-\kappa}(1 + o(1)), \ d > 0, \ \kappa \in (1, 2]. \tag{14}$$

Then x_t is geometrically α-mixing (Pham and Tran [65]) with tail (14) and the same index κ (Brockwell and Cline [6]).

Define estimating equations

$$y_t(\theta) := (x_t - \theta y_{t-1})\, y_{t-1} \ \text{ and } \ \hat{y}_{n,t}(\theta) := y_t(\theta) I\left(|y_t(\theta)| \leq y^{(a)}_{(m_n+1)}(\theta)\right)$$

where $y_t^{(a)} := |y_t|$, and $\theta \in \Theta$ a compact subset of $\mathbb{R}$ with θ^0 in the interior. The deterministically trimmed version $y_{n,t}(\theta)$ and assumed threshold property are

$$y_{n,t}(\theta) := y_t(\theta) I\left(|y_t(\theta)| \leq c_n(\theta)\right) \text{ where } P\left(|y_t(\theta)| > c_n(\theta)\right) = \frac{m_n}{n}$$

for some intermediate order sequence $\{m_n\}$, and thresholds $c_n : \Theta \rightarrow \mathbb{R}$. Write $y_t = y_t(\theta^0)$, $\hat{y}_{n,t} = \hat{y}_{n,t}(\theta^0)$ and $y_{n,t} = y_{n,t}(\theta^0)$. Continuity of the error distribution ensures such a sequence $\{c_n(\theta)\}$ exists for all θ and any $\{m_n\}$. Hill and Renault [47] propose the following estimator:

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{arginf}} \left\{ \left(\frac{1}{n} \sum_{t=1}^{n} \hat{y}_{n,t}(\theta) \right)^2 \right\}.$$

The present data generating process satisfies the conditions for asymptotic linearity:

$$\frac{n\,|J_n|}{\sigma_n} \left(\hat{\theta}_n - \theta^0\right) = \frac{1}{\sigma_n} \sum_{t=1}^{n} \hat{y}_{n,t} \times (1 + o_p(1)) + o_p(1),$$

where $J_n := -E[y_{t-1}^2 I(|u_t y_{t-1}| \leq c_n(\theta^0))]$ and $\sigma_n^2 := E(\sum_{t=1}^n y_{n,t})^2$. See Hill and Renault [47: Section 5.1].

Linearity implies if $1/\sigma_n \sum_{t=1}^n \hat{y}_{n,t}$ converges in law to $N(0,1)$ then so does $n|J_n|\sigma_n^{-1}(\hat{\theta}_n - \theta^0)$. Since u_t and x_{t-1} are independent and have the same tail index, $y_t = u_t x_{t-1}$ has tail (14) with the same index $\kappa \leq 2$ (Cline [10]). Therefore y_t belongs to class (1). Assumption A.i applies by continuity and positiveness of the error distribution. Assumption A.ii applies since $y_{n,t}$ forms an adapted martingale difference array relative to $\sigma(\{u_\tau, x_\tau\} : \tau \leq t)$ by Example 1. Therefore $\sigma_n^2 = nE[y_{n,t}^2]$. Finally, Assumptions B and C hold by Corollary 5.3 since $\{x_t, u_t\}$ are geometrically α-mixing. This proves the next result.

Theorem 5.5. *The present data generating process satisfies* $1/\sigma_n \sum_{t=1}^n \hat{y}_{n,t} \overset{\mathfrak{L}}{\rightarrow} N(0,1)$*, hence* $n\,|J_n|\,\sigma_n^{-1}(\hat{\theta}_n - \theta^0) \overset{\mathfrak{L}}{\rightarrow} N(0,1)$.

Remark: Under power-law tail decay $P(|u_t| > u) = du^{-\kappa}(1 + o(1))$ with $\kappa \in (1,2)$ it can be shown $|J_n| \sim K(n/m_n)^{2/\kappa - 1}$ and $\sigma_n^2 \sim Kn(n/m_n)^{2/\kappa-1}$ by applications of Karamata's Theorem, hence the rate of convergence is $n|J_n|\sigma_n^{-1} \sim Kn^{1/2}(n/m_n)^{1/\kappa - 1/2} > n^{1/2}$. See Lemma 3.1 of Hill and Renault [47]. Simply choose very light intermediate order trimming $m_n = [L(n)]$ to ensure $n|J_n|\sigma_n^{-1} \sim Kn^{1/\kappa}/L(n)$, a rate arbitrarily close to the highest achieved $n^{1/\kappa}$ amongst untrimmed M-estimators (e.g. Davis et al [20]).

EXAMPLE 4 (Distributed Lag): Consider a class of L_p-bounded linear processes $\{y_t\}$ with geometric memory: $y_t := \sum_{i=0}^\infty \psi_i \epsilon_{t-i}$, $||\epsilon_t||_p \leq K$, $p > 0$, $\psi_i = O(\rho^i)$, $\rho \in (0,1)$. Assume ϵ_t is stationary geometric α-mixing. If $p \geq 1$ use Minkowski's inequality and $|\psi_i| \leq K\rho^i$ to deduce, for any $d \in \mathbb{N}$: $||y_t - E[y_t|\Im_{t-d}^{t+d}]||_p \leq \sum_{i=d+1}^\infty |\psi_i| \times ||\epsilon_{t-i}||_p \leq K\rho^d$. If $p < 1$ then by Loève's inequality $||y_t - E[y_t|\Im_{t-d}^{t+d}]||_p \leq (\sum_{i=d+1}^\infty |\psi_i|^p \times E|\epsilon_{t-i}|^p)^{1/p} \leq K\rho^d$. Hence $\{y_t\}$ is geometrically L_p-NED on $\{\epsilon_t\}$, so Assumptions B and C hold by Theorem 5.2.a,b. If ϵ_t is iid with a density $f(\epsilon) > 0\ \forall \epsilon \in \mathbb{R}$ then Assumption A holds by Lemma 5.1.

EXAMPLE 5 (Strong-IGARCH): Let $y_t = h_t u_t$, $u_t \overset{iid}{\sim} N(0,1)$, where $h_t^2 = \omega + \alpha y_{t-1}^2 + (1-\alpha)h_{t-1}^2$, $\omega > 0$, $\alpha \in [0,1)$. The random variable y_t has a density positive on $\mathbb{R}$-*a.e.*, and is geometrically α-mixing with Paretian tail (14) and index $\kappa = 2$ (Basrak et al [3]: Theorem 3.1).

Suppose we want to estimate the tail-trimmed variance $E[y_t^2 I(|y_t| \leq c_n)]$ with thresholds $\{c_n\}$ defined by $(n/m_n)P(|y_t| > c_n) \rightarrow 1$ for an intermediate order $\{m_n\}$. Define $y_t^{(a)}$

$:= |y_t|$ and $y_{n,t}^2 := y_t^2 I(|y_t| \leq c_n)$. Assumptions B and C hold by Corollary 5.3 since y_t is geometrically α-mixing, and results from Section 5.1 verify Assumption A follows from distribution smoothness and $\liminf_{n\to\infty} E[(y_{n,t}^2 - E[y_{n,t}^2])(y_{n,t-h}^2 - E[y_{n,t-h}^2])] \geq 0$ for all h by construction of the GARCH process. Apply Theorem 5.4 to conclude

$$\frac{1}{\sigma_n}\sum_{t=1}^{n}\left\{y_t^2 I\left(|y_t| \leq y_{(m_n+1)}^{(a)}\right) - E\left[y_t^2 I\left(|y_t| \leq y_{(m_n+1)}^{(a)}\right)\right]\right\} \xrightarrow{\mathfrak{L}} N(0,1).$$

Similarly, $Y_{h,t} := y_t y_{t-h}$ is geometrically α-mixing and has tail (14) with index $\kappa/2 = 1$ (Basrak et al [3], Cline [10]). Since $Y_{h,t} = u_t u_{t-h} h_t h_{t-h}$ is symmetrically distributed and u_t is iid, all parts of Assumption A hold by Example 1. Now define $\{c_n\}$ by $(n/m_n)P(|Y_{h,t}| > c_n) \to 1$. The tail-trimmed covariance $E[Y_{h,t}I(|Y_{h,t}| \leq c_n)]$ is estimated by $1/n\sum_{t=1}^{n} Y_{h,t}I(|Y_{h,t}| \leq Y_{h,(m_n+1)}^{(a)})$, where

$$\frac{1}{\sigma_n}\sum_{t=1}^{n}\left\{Y_{h,t} I\left(|Y_{h,t}| \leq Y_{h,(m_n+1)}^{(a)}\right) - E\left[Y_{h,t} I\left(|Y_{h,t}| \leq Y_{h,(m_n+1)}^{(a)}\right)\right]\right\} \xrightarrow{\mathfrak{L}} N(0,1).$$

EXAMPLE 6 (AR-GARCH): A large array of nonlinear AR-GARCH processes are geometrically α-mixing provided the innovations have a sufficiently smooth density and a Lyapunov-type moment bound is satisfied. Examples include Threshold Autoregressions; Smooth Transition Autoregressions; and Asymmetric-, Multiplicative-, Exponential-, and Threshold-GARCH. See, e.g., Carrasco and Chen [8], Cline and Pu [11], and Meitz and Saikonnen [59].

Consider a stationary Self Exciting Threshold-Autoregression $y_t = \phi y_{t-1} I(y_{t-1} < 0) + \epsilon_t$, $|\phi| < 1$, with Exponential Smooth Transition GARCH(1,1) errors $\epsilon_t = h_t u_t$, $u_t \overset{iid}{\sim} (0,1)$ and

$$h_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \tilde{\alpha}\epsilon_{t-1}^2 \exp\left\{-\gamma\epsilon_{t-1}^2\right\} + \beta h_{t-1}^2$$

where $\gamma \geq 0$, $\omega > 0$, and $\{\alpha, \alpha + \tilde{\alpha}, \beta\} \geq 0$, and $E[(\beta + \max\{\alpha, \alpha + \tilde{\alpha}\}u_t^2)^r] < 1$ for some $r > 0$. Assume u_t has a density, positive on $\mathbb{R}$ Then $\{y_t\}$ is geometrically ergodic (Meitz and Saikonnen [59: Proposition 1]) hence Corollary 5.3 applies to any $Y_{h,t}$, so Assumptions B and C hold for $Y_{h,t}$. Assumption A.i holds for y_t by distribution smoothness, and Assumption A.ii holds for y_t with $\phi \in [0,1)$, or y_t^2 due to positive association. Note Example 2 is based on an infinite series of independent $\psi_i\epsilon_{t-i}$, while $y_t = \sum_{i=0}^{\infty}\psi_{t,i}\epsilon_{t-i}$ is nonlinear for $\phi \neq 0$ since $\psi_{t,i} = \phi^i \sqcap_{j=0}^{i-1} I(y_{t-j-1} < 0)$, and $\psi_{t,i}\epsilon_{t-i}$ are dependent. Thus, whether the range $\phi \in (-1,0)$ is covered is not considered here.

EXAMPLE 7 (Non-Mixing/Non-NED Stationary GARCH): Represent a lag polynomial by $\pi_r(\mathcal{L}) := \sum_{i=1}^{r}\pi_i\mathcal{L}^i$ with $\mathcal{L}$ the lag operator, and suppose $y_t = h_t\epsilon_t$ where $\epsilon_t \overset{iid}{\sim} (0,1)$ and

$$h_t^2 = \pi_0 + \pi_\infty(\mathcal{L})\, y_t^2$$
$$\pi_0 > 0,\ \pi_i \geq 0,\ \pi_i \leq C\rho^i \text{ for some } \rho \in (0,1) \text{ and } C \in (0, 1/\rho).$$

Note $S := \pi_\infty(1) \leq C\rho/(1-\rho)$ covers integrated $S = 1$ and explosive $S > 1$ cases. Hill [46: Theorem 4.3] shows the trimmed level and tail array $\{y_{n,t}, I(|y_t| > c_n e^u)\}$ are geometrically L_2-NED on iid $\{\epsilon_t\}$ with constants $\{e_{n,t}^*, (m_n/n)^{1/2}e^{-u/2}\}$, so y_t satisfies Assumptions B and C by Theorem 5.2.a,b. If ϵ_t has an absolutely continuous distribution positive on $\mathbb{R}$ then Assumption A.i,ii hold for y_t or $y_t y_{t-h}$, $h \geq 1$, by Example 1, or for y_t^2 by Example 2 given positive association. Notice y_t need not be mixing nor NED, and in many cases cannot be shown to be either. See Davidson [18], Hill [46] and their references.

6. STUDENTIZED RATIO FOR INDEPENDENT DATA In some cases a substantial simplification of (2) exists. We only treat iid Paretian data, but similar arguments extend to finite-dependent data and certain martingale differences (cf. Example 1).

Suppose y_t is stationary, iid, L_1-bounded, and symmetrically distributed with a strictly positive distribution on $\mathbb{R}$, and tail (14) with index $\kappa \in (1,2)$. We assume a first moment so we can test $E[y_t] = 0$. Define $y_{n,t} = y_t \times I(|y_t| \leq c_n)$ where $(n/m_n)P(|y_t| > c_n) \to 1$. The trimmed variance satisfies by an application of Karamata's Theorem, cf. (1),

$$\sigma_n^2 = nE\left[y_{n,t}^2\right] \sim \frac{\kappa}{2-\kappa} n c_n^2 P\left(|y_t| \geq c_n\right) \sim \frac{\kappa}{2-\kappa} c_n^2 m_n. \tag{15}$$

Assumptions A.i,ii hold by distribution positiveness and independence $\sigma_n^2 = nE[y_{n,t}^2]$, and independence implies Assumptions B and C. Our last result therefore follows from Theorem 3.2 and (15).

COROLLARY 6.1. *Define* $y_t^{(a)} := |y_t|$ *and let* $m_n/(\ln(n))^{2+\iota} \to \infty$. *Then*

$$\left(\frac{2-\kappa}{\kappa}\right)^{1/2} \frac{1}{c_n \times m_n^{1/2}} \sum_{t=1}^{n} y_t I\left(|y_t| < y_{(m_n+1)}^{(a)}\right) \xrightarrow{\mathfrak{L}} N(0,1).$$

Remark: Plug-ins for c_n and κ can be used for inference, including $y_{(\tilde{m}_n+1)}^{(a)}$ for c_n, and Hill's [42] tail index estimator $\hat{\kappa}_{\tilde{m}_n}^{-1} = 1/\tilde{m}_n \sum_{i=1}^{\tilde{m}_n} \ln(y_{(i)}/y_{(\tilde{m}_n+1)})$ to name one of many. Note $\{\tilde{m}_n\}$ is an intermediate order sequence that may differ from the sequence $\{m_n\}$ used for trimming. Under the iid assumption $y_{(\tilde{m}_n+1)}^{(a)}/c_n = 1 + O_p(1/\tilde{m}_n^{1/2})$ and $\hat{\kappa}_{\tilde{m}_n}^{-1} = \kappa^{-1} + O_p(1/\tilde{m}_n^{1/2})$ provided second order restrictions on the Paretian tail decay rate in (14) and companion restrictions on $\tilde{m}_n$ apply. A classic example is $P(|y_t| > y) = dy^{-\kappa}(1 + O(y^{-\alpha}))$ and $\tilde{m}_n = o(n^{2\alpha/(2\alpha+\kappa)})$. See Hill [44], Hsing [49] and their citations.

APPENDIX: Proofs

PROOF OF LEMMA 3.1 (Relative Stability)

Recall $m_n = \min\{m_{1,n}, m_{2,n}\}$. Use (1) and (3) to deduce for all $1 \leq t \leq n$

$$\frac{E\left[y_t^2 I\left(-l_n \leq y_t \leq 0\right)\right]}{l_n^2 P\left(y_t \leq -l_n\right)} = \frac{n}{m_{1,n}} \times \frac{E\left[y_t^2 I\left(-l_n \leq y_t \leq 0\right)\right]}{l_n^2}\left(1 + o\left(1\right)\right) \to (0,\infty).$$

Therefore for any $1 \leq t \leq n$ and some finite $K > 0$

$$l_n = K\left(\frac{n}{m_{1,n}}\right)^{1/2} \left\|y_t^2 I\left(-l_n \leq y_t \leq 0\right)\right\|\left(1 + o\left(1\right)\right) = O\left(n^{1/2}\left\|y_t^2 I\left(-l_n \leq y_t \leq 0\right)\right\|/m_n^{1/2}\right)$$

since $\{m_{1,n}\}$ is an intermediate order sequence. An identical argument shows $u_n = O(n^{1/2}||y_t^2(0 \leq y_t \leq u_n)||/m_n)$. But since

$$E\left[y_{n,t}^2\right] = E\left[y_t^2 I\left(-l_n \leq y_t \leq 0\right)\right] + E\left[y_t^2 I\left(0 \leq y_t \leq u_n\right)\right]$$

it follows instantly for any $1 \leq t \leq n$

$$c_n = O\left(\left(\frac{n}{m_n}\right)^{1/2} \max\left\{\left\|y_t^2\left(-l_n \leq y_t \leq 0\right)\right\|, \left\|y_t^2\left(0 \leq y_t \leq u_n\right)\right\|\right\}\right) = O\left(\left(\frac{n}{m_n}\right)^{1/2} \left\|y_{n,t}\right\|\right),$$

hence $c_n = O((\sum_{t=1}^{n} E[y_{n,t}^2])^{1/2}/m_n^{1/2})$. $\mathcal{QED}$.

PROOF OF LEMMA 3.3 (CLT-DT)

Key steps of the proof rely on restricting the size h_n of the telescoping sum $\mathcal{Z}_{n,i} = \sum_{t=(i-1)h_n+j_n+1}^{ih_n} z_{n,t}$, the block buffer j_n and the number of trimming extremes $m_n = \min\{m_{1,n}, m_{2,n}\}$. As we show below, under geometric memory and heavy tails $j_n/\ln(n) \rightarrow \infty$ ensures blocks are sufficiently separated to bound dispersion in various sums of functions of $\mathcal{Z}_{n,i}$. Further, by construction we need block sizes to grow faster than the buffer $h_n/j_n \rightarrow \infty$, and bounding $\max_{1\leq i\leq r_n} |\mathcal{W}_{n,i}| \leq K h_n c_n/\sigma_n \leq K h_n/m_n^{1/2}$ under Lemma 3.1 requires $h_n/m_n^{1/2} \rightarrow 0$. Therefore, put

$$h_n \rightarrow \infty \text{ such that } h_n/\ln(n) \rightarrow \infty \text{ and } h_n = o\left((\ln(n))^{1+\delta}\right) \text{ for tiny } \delta > 0 \quad (16)$$

$$j_n/\ln(n) \rightarrow \infty \text{ and } j_n = o(h_n).$$

Recall Assumption A.iii imposes $m_n/(\ln(n))^{1+\iota} \rightarrow \infty$. For any $\delta \in (0, \iota)$ we therefore have as required $h_n = o(m_n^{1/2})$, $j_n = o(h_n)$, $j_n/\ln(n) \rightarrow \infty$.

Define

$$\mathcal{U}_{n,i} := \sigma_n \mathcal{W}_{n,i}, \text{ and } \mathcal{K}_n \sim K_0 n^{\delta} c_n^2 \text{ for some } K_0 > 0 \text{ and tiny } \delta > 0,$$

and a truncation function

$$\widetilde{\mathcal{U}}_{n,i}^2(K) := \mathcal{U}_{n,i}^2 I\left(\mathcal{U}_{n,i}^2 \leq K\right) = \sigma_n^2 \mathcal{W}_{n,i}^2 I\left(\sigma_n^2 \mathcal{W}_{n,i}^2 \leq K\right) \text{ and } \widetilde{\mathcal{U}}_{n,i}^2 = \widetilde{\mathcal{U}}_{n,i}^2(\mathcal{K}_n).$$

Notice by construction any finite lag function of $\mathcal{U}_{n,i}$ is geometrically α-mixing under Assumption B because by construction $\mathcal{U}_{n,i}$ is $\Im_{-g_n}^{ih_n}$-measurable:

$$\mathcal{U}_{n,i} = \sigma_n \left(\sum_{t=(i-1)h_n+j_n+1}^{ih_n} \left\{ E\left[z_{n,t}|\Im_{-g_n}^{ih_n}\right] - E\left[z_{n,t}|\Im_{-g_n}^{(i-1)h_n}\right]\right\}\right).$$

Define the index sets

$$\mathcal{I}_{n,i}^* := \{t : t \in [(i-1)h_n + j_n + 1, ..., ih_n]\} \text{ and } \mathcal{I}_n := \left\{t : t \in \bigcup\nolimits_{i=1}^{r_n} \mathcal{I}_{n,i}^*\right\}.$$

We first show $E[y_{n,t}|\mathcal{F}_{n,i-1}]$ and $y_{n,t} - E[y_{n,t}|\mathcal{F}_{n,i}]$ are mixingales in Lemma A.1 for any block sizes $\{h_n\}$. The mixingale property is then exploited to establish a martingale decomposition in Lemma A.2:

$$\sum_{t=1}^{n} z_{n,t} = \sum_{i=1}^{r_n} \mathcal{W}_{n,i} + o_p(1) \text{ for any } \{h_n\}.$$

Finally, in Lemma A.3 we prove a central limit theorem $\sum_{i=1}^{r_n} W_{n,i} \xrightarrow{\mathfrak{L}} N(0,1)$ for tiny block sizes $h_n = o((\ln(n))^{1+\delta})$ by (16). Lemma 3.3 therefore follows from Lemmas A.2 and A.3.

Recall the mixingale constants $e_{n,t}$ and coefficients ζ_{d_n} defined in Lemma 2.1.

Lemma A.1. *Under Assumption B* $\{y_{n,t} - E[y_{n,t}|\mathcal{F}_{n,i}], \Im_t\}_{t\in\mathcal{I}_n}$ *and* $\{E[y_{n,t}|\mathcal{F}_{n,i-1}], \Im_t\}_{t\in\mathcal{I}_n}$ *form geometric L_2-mixingale arrays with constants* $e_{n,t}\zeta_{j_n}^{\iota}$ *for tiny* $\iota > 0$.

PROOF. The claim follows essentially from de Jong's [21: eq.'s A.7-A.12] argument. By Lemma 2.1 $\{y_{n,t}, \Im_t\}$ is a geometric L_2-mixingale array with constants $e_{n,t} \leq Kc_n$ and coefficients $\zeta_{d_n} = O(\rho^{d_n})$. Further, since $\Im_t$ is increasing we have $\mathcal{F}_{n,i-1} = \Im_{-g_n}^{(i-1)h_n} \subset \Im_{-\infty}^{t-j_n}$ for each $1 \leq i \leq r_n$ and all $(i-1)h_n + j_n + 1 \leq t \leq ih_n$, and $\Im_{-\infty}^{t-j_n} \subset \Im_{-\infty}^{t-d}$ for sufficiently large n since $j_n \to \infty$. Exploit iterated expectations, Jensen's inequality and $j_n \geq d$ for sufficiently large n to deduce for arbitrarily large $\lambda > 0$

$$\begin{aligned}
\left\| E\left(E\left[y_{n,t}|\mathcal{F}_{n,i-1}\right]\right)|\Im_{-\infty}^{t-d}\right\| &\leq \left\| E\left(E\left[y_{n,t}|\Im_{-g_n}^{(i-1)h_n}\right]\right)|\Im_{-\infty}^{t-d}\right\| \\
&= \left\| E\left(E\left[E\left(y_{n,t}|\Im_{-\infty}^{t-j_n}\right)|\Im_{-g_n}^{(i-1)h_n}\right]\right)|\Im_{-\infty}^{t-d}\right\| \\
&\leq K\left\| E\left[y_{n,t}|\Im_{-\infty}^{t-j_n}\right]\right\| \leq e_{n,t} \times \zeta_{j_n} \leq e_{n,t}\zeta_{j_n}^{\iota} \times \zeta_d^{1-\iota},
\end{aligned}$$

where $\zeta_d^{1-\iota} = O(\rho^{(1-\iota)d})$. An identical argument applies to $\{y_{n,t} - E[y_{n,t}|\mathcal{F}_{n,i}], \Im_t\}_{t \in \mathcal{I}_n}$. $\mathcal{QED}$.

We now use Lemma A.1 to prove the required martingale difference decomposition.

LEMMA A.2. *Under Assumptions A and B* $\sum_{t=1}^{n} z_{n,t} = \sum_{i=1}^{r_n} \mathrm{W}_{n,i} + o_p(1)$. *In particular a.* $\sum_{t=r_nh_n+1}^{n} z_{n,t} \xrightarrow{p} 0$; *b.* $\sum_{i=1}^{r_n}\sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_{n,t} \xrightarrow{p} 0$; *c.* $\sum_{i=1}^{r_n} E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}] \xrightarrow{p} 0$; *and d.* $\sum_{i=1}^{r_n}(\mathcal{Z}_{n,i} - E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}]) \xrightarrow{p} 0$.

PROOF. In lieu of decomposition (11) claims (a)-(d) prove $\sum_{t=1}^{n} z_{n,t} = \sum_{i=1}^{r_n} \mathrm{W}_{n,i} + o_p(1)$.

Claims (a) and (b): Invoke mixingale property Lemma 2.1, variance bound Lemma 2.2, Assumption A.ii non-degeneracy $\sum_{t=1}^{n} E[y_{n,t}^2]/\sigma_n^2 = O(1)$ and $r_nh_n \sim n$ to deduce

$$E\left[\left(\sum_{t=r_nh_n+1}^{n} z_{n,t}\right)^2\right] = O\left(\frac{1}{\sigma_n^2}\sum_{t=r_nh_n+1}^{n} E[y_{n,t}^2]\right) = O\left((n - r_nh_n)/n\right) = o(1)$$

and

$$\begin{aligned}
E\left[\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_{n,t}\right)^2\right] &= O\left(\frac{1}{\sigma_n^2}\sum_{i=1}^{r_n}\sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} E[y_{n,t}^2]\right) \\
&= O\left(r_nj_n/n\right) = O(j_n/h_n) = o(1).
\end{aligned}$$

Claims (a) and (b) now follow from Chebyshev's inequality.

Claims (c) and (d): The remaining two terms $\sum_{i=1}^{r_n}\{\mathcal{Z}_{n,i} - E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}]\}$ and $\sum_{i=1}^{r_n} E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}]$ are handled by invoking the Lemma A.1 mixingale properties for $\{y_{n,t} - E[y_{n,t}|\mathcal{F}_{n,i}], \Im_t\}_{t \in \mathcal{I}_n}$ and $\{E[y_{n,t}|\mathcal{F}_{n,i-1}], \Im_t\}_{t \in \mathcal{I}_n}$. Apply McLeish's [57: Theorem 1.6] maximal inequality to deduce

$$\begin{aligned}
E\left[\left(\sum_{i=1}^{r_n}\{\mathcal{Z}_{n,i} - E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}]\}\right)^2\right] &= O\left(\sum_{i=1}^{r_n}\sum_{t=(i-1)h_n+j_n}^{ih_n}\frac{e_{n,t}^2}{\sigma_n^2}\zeta_{j_n}^{2\iota}\right) = O\left(\frac{r_nh_nc_n^2}{\sigma_n^2}\rho^{2\iota j_n}\right) \\
&= O\left(nc_n^2\sigma_n^{-2}\rho^{2\iota j_n}\right) = O\left(n\rho^{2\iota j_n}\right) = o(1).
\end{aligned}$$

The second equality follows from the geometric mixingale properties $e_{n,t} \leq Kc_n$ and $\zeta_{j_n}^{2\iota} = O(\rho^{2\iota j_n})$. The third uses $r_n h_n \sim n$, and the fourth follows from threshold bound Lemma 3.1 and non-degeneracy Assumption A.ii: $c_n^2/\sigma_n^2 \leq K\sum_{t=1}^n E[y_{n,t}^2]/\sigma_n^2 \leq K$. Since $j_n/\ln(n) \to \infty$ by assumption the equality follows since $n\rho^{2\iota j_n} \to 0$. An identical argument proves

$$E\left[\left(\sum_{i=1}^{r_n} E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}]\right)^2\right] = O\left(nc_n^2\sigma_n^{-2}\rho^{2\iota j_n}\right) = o(1).$$

Now apply Chebyshev's inequality to complete the proof. $\mathcal{QED}$.

Lastly, a martingale difference CLT.

LEMMA A.3. *Under Assumptions A-B* $\sum_{i=1}^{r_n} \mathcal{W}_{n,i} \stackrel{\mathfrak{L}}{\to} N(0,1)$.

PROOF. We will show $\max_{1\leq i\leq r_n} |\mathrm{W}_{n,i}|$ is $o_p(1)$ and uniformly L_2-bounded, and $\sum_{i=1}^{r_n} \mathrm{W}_{n,i}^2 \stackrel{p}{\to} 1$. The CLT then follows from Theorem 2.3 of McLeish [56].

By construction $\max_{1\leq i\leq r_n} |\mathcal{W}_{n,i}| \leq h_n c_n/\sigma_n$; the thresholds satisfy $c_n \leq K(\sum_{t=1}^n E[y_{n,t}^2])^{1/2}/m_n^{1/2}$ by Lemma 3.1, and by Assumption A.iii $m_n/(\ln(n))^{2+\iota} \to \infty$; and non-degeneracy Assumption A.ii states $\sum_{t=1}^n E[y_{n,t}^2]/\sigma_n^2 = O(1)$. Now use $h_n = o((\ln(n))^{1+\delta})$ to deduce for any $\delta \in (0,\iota]$

$$\max_{1\leq i\leq r_n} |\mathcal{W}_{n,i}| \leq \frac{h_n c_n}{\sigma_n} \leq Kh_n \frac{\left(\sum_{t=1}^n E\left[y_{n,t}^2\right]\right)^{1/2}/m_n^{1/2}}{\sigma_n} = o\left(\frac{h_n}{(\ln(n))^{1+\iota}}\right) = o(1).$$

But the deterministic bound $o(1)$ implies $\max_{1\leq i\leq r_n} |\mathcal{W}_{n,i}|$ is both $o_p(1)$ and uniformly L_2-bounded by Lebesgue's dominated convergence.

Next, by the triangle inequality $|\sum_{i=1}^{r_n} \mathrm{W}_{n,i}^2 - 1| \leq \sum_{i=1}^5 \mathcal{E}_{n,i}$ where

$$\begin{aligned}
\mathcal{E}_{n,1} &= \left|\frac{1}{\sigma_n^2}\sum_{i=1}^{r_n}\left(\widetilde{\mathcal{U}}_{n,i}^2 - \mathcal{U}_{n,i}^2\right)\right| \text{ and } \mathcal{E}_{n,2} = \left|\frac{1}{\sigma_n^2}\sum_{i=1}^{r_n}\left(\widetilde{\mathcal{U}}_{n,i}^2 - E\left[\widetilde{\mathcal{U}}_{n,i}^2\right]\right)\right| \\
\mathcal{E}_{n,3} &= \frac{1}{\sigma_n^2}\sum_{i=1}^{r_n}\left|E\left[\widetilde{\mathcal{U}}_{n,i}^2\right] - E\left[\mathcal{U}_{n,i}^2\right]\right| \text{ and } \mathcal{E}_{n,4} = \left|\sum_{i=1}^{r_n}\left(E\left[\mathcal{W}_{n,i}^2\right] - E\left[\mathcal{Z}_{n,i}^2\right]\right)\right| \\
\mathcal{E}_{n,5} &= \left|\sum_{i=1}^{r_n} E\left[\mathcal{Z}_{n,i}^2\right] - 1\right|.
\end{aligned}$$

Use Steps 1-3 and Markov's inequality to deduce $\mathcal{E}_{n,i} = o_p(1)$ for $i = 1-4$, and $\mathcal{E}_{n,5} = o_p(1)$ by Step 4. Therefore $\sum_{i=1}^{r_n} \mathrm{W}_{n,i}^2 \stackrel{p}{\to} 1$.

Step 1 ($\sigma_n^{-2}\sum_{i=1}^{r_n} E|\widetilde{\mathcal{U}}_{n,i}^2 - \mathcal{U}_{n,i}^2| = o(1)$): By construction $\mathcal{U}_{n,i}^2 := \sigma_n^2 \mathrm{W}_{n,i}^2 \leq Kh_n^2 c_n^2$. Use $r_n h_n \sim n$, $h_n = o((\ln(n))^{1+\delta}) = o(n^\delta)$ by supposition, $n/\sigma_n^2 = O(1)$ by Assumption A.ii, $\mathcal{K}_n \sim K_0 n^\delta c_n^2$ by construction, and stationarity to deduce for sufficiently large K_0

$$\begin{aligned}
\frac{1}{\sigma_n^2}\sum_{i=1}^{r_n} E\left|\widetilde{\mathcal{U}}_{n,i}^2 - \mathcal{U}_{n,i}^2\right| &= \frac{1}{\sigma_n^2}\sum_{i=1}^{r_n} E\left[\mathcal{U}_{n,i}^2 I\left(\mathcal{U}_{n,i}^2 > \mathcal{K}_n\right)\right] \\
&\leq K\frac{r_n}{\sigma_n^2} E\left[\mathcal{U}_{n,i}^2 I\left(Kn^\delta c_n^2 > K_0 n^\delta c_n^2\right)\right] = o(1).
\end{aligned}$$

Step 2 ($\sigma_n^{-4}E(\sum_{i=1}^{r_n}\{\widetilde{\mathcal{U}}_{n,i}^2 - E[\widetilde{\mathcal{U}}_{n,i}^2]\})^2 = o(1)$): By Lemma 2.1 $\{y_{n,t}, \Im_t\}$ forms a geometric L_2-mixingale array with constants $e_{n,t} \leq Kc_n$ and coefficients $\zeta_{d_n} = O(\rho^{d_n})$.

Similarly by Lemma A.1 $\{y_{n,t} - E[y_{n,t}|\mathcal{F}_{n,i}], \Im_t\}_{t\in\mathcal{I}_n}$ and $\{E[y_{n,t}|\mathcal{F}_{n,i-1}], \Im_t\}_{t\in\mathcal{I}_n}$ form L_2-mixingale arrays with constants $e_{n,t}\zeta_{j_n}^{\iota}$ and coefficients of any size. Since sums and products of geometric α-mixing random variables $\mathcal{U}_{n,i}$ are geometric α-mixing, and

$$\frac{\mathcal{U}_{n,i}^2}{\sigma_n^2} = \left(\sum_{t=(i-1)h_n+j_n}^{ih_n} z_{n,t} - \sum_{t=(i-1)h_n+j_n}^{ih_n} \{z_{n,t} - E[z_{n,t}|\mathcal{F}_{n,i}]\} - \sum_{t=(i-1)h_n+j_n}^{ih_n} E[z_{n,t}|\mathcal{F}_{n,i-1}]\right)^2,$$

it is straightforward to show $\{\widetilde{\mathcal{U}}_{n,i}^2/\sigma_n^2, \mathcal{F}_{n,i}\}$ forms an L_2-mixingale array with constants $\sigma_n^{-2}(\sum_{t=(i-1)h_n+j_n}^{ih_n} e_{n,t})^2\zeta_{j_n}^{2\iota}$ and coefficients of any size λ by the argument of Lemma A.1. Apply McLeish's [57: Theorem 1.6] maximal inequality to conclude

$$\begin{aligned} E\left(\frac{1}{\sigma_n^2}\sum_{i=1}^{r_n}\{\widetilde{\mathcal{U}}_{n,i}^2 - E[\widetilde{\mathcal{U}}_{n,i}^2]\}\right)^2 &= O\left(\frac{1}{\sigma_n^4}\sum_{i=1}^{r_n}\left(\sum_{t=(i-1)h_n+j_n}^{ih_n} e_{n,t}\right)^4 \zeta_{j_n}^{4\iota}\right) \\ &= O\left(\frac{r_n h_n^4 c_n^4}{\sigma_n^4}\rho^{4\iota j_n}\right) = o\left(n^{1+3\delta}\rho^{4\iota j_n}\right) = o(1). \end{aligned}$$

The second equality uses $\zeta_{j_n}^{4\iota} = O(\rho^{4\iota j_n})$ and $e_{n,t} \leq Kc_n$. The third uses $r_n h_n \sim n$, $h_n = o((\ln(n))^{1+\delta}) = o(n^{\delta})$ and $c_n^2 = O(\sigma_n^2)$ by Lemma 3.1 and Assumption A.ii. Since $j_n/\ln(n) \to \infty$ the last equality follows.

Step 3 ($E|\sum_{i=1}^{r_n}(\mathrm{W}_{n,i}^2 - \mathcal{Z}_{n,i}^2)| = o(1)$): Invoke Lemma A.1 and an argument identical to the proof of Lemma A.2 for some tiny $\iota > 0$

$$\|E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}]\|_2 = O\left(\frac{h_n^{1/2}c_n}{\sigma_n}\rho^{\iota j_n}\right) \quad \text{and} \quad \|\{\mathcal{Z}_{n,i} - E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}]\}\|_2 = O\left(\frac{h_n^{1/2}c_n}{\sigma_n}\rho^{\iota j_n}\right).$$

Further, by construction and the conditional Jensen's inequality $||\mathrm{W}_{n,i} + \mathcal{Z}_{n,i}||_2 \leq K||\mathcal{Z}_{n,i}||_2$, while $\max_{1\leq i\leq r_n}||\mathcal{Z}_{n,i}||_2 \leq Kh_n^{1/2}c_n/\sigma_n$ follows from mixingale property Lemma A.1, variance bound Lemma 2.2 and $||y_{n,t}|| \leq c_n$. Now apply Minkowski and Cauchy-Schwarz inequalities, Lemma 3.1 and Assumption A.ii to obtain

$$\begin{aligned} \left\|\sum_{i=1}^{r_n}\{\mathcal{W}_{n,i}^2 - \mathcal{Z}_{n,i}^2\}\right\|_1 &\leq \sum_{i=1}^{r_n}\|\mathcal{W}_{n,i} - \mathcal{Z}_{n,i}\|_2 \|\mathcal{W}_{n,i} + \mathcal{Z}_{n,i}\|_2 \\ &\leq \sum_{i=1}^{r_n}\left\{\|E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}]\|_2 + \|\mathcal{Z}_{n,i} - E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}]\|_2\right\} \times \|\mathcal{Z}_{n,i}\|_2 \\ &= O\left(r_n \times \frac{h_n^{1/2}c_n}{\sigma_n}\rho^{\iota j_n} \times \frac{h_n^{1/2}c_n}{\sigma_n}\right) = O\left(\frac{nc_n^2}{\sigma_n^2}\rho^{\iota j_n}\right) = O\left(n\rho^{\iota j_n}\right). \end{aligned}$$

Since $j_n/\ln(n) \to \infty$ the last line is $o(1)$.

Step 4 ($\sum_{i=1}^{r_n} E[\mathcal{Z}_{n,i}^2] \xrightarrow{p} 1$): The proof of Lemma A.2 reveals by the definitions of

$\mathcal{Z}_{n,i}$ and $\mathrm{W}_{n,i}$

$$\begin{aligned}\sum_{t=1}^{n} z_{n,t} &= \sum_{i=1}^{r_n} \mathcal{W}_{n,i} + \sum_{i=1}^{r_n} \left(\mathcal{Z}_{n,i} - E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i}]\right) + \sum_{i=1}^{r_n} E[\mathcal{Z}_{n,i}|\mathcal{F}_{n,i-1}] \\ &\quad + \sum_{i=1}^{r_n} \sum_{t=(i-1)h_n+1}^{(i-1)h_n+j_n} z_{n,t} + \sum_{t=r_n h_n+1}^{n} z_{n,t} \\ &= \sum_{i=1}^{r_n} \mathcal{Z}_{n,i} + o_p(1) = \sum_{i=1}^{r_n} \sum_{t=(i-1)h_n+j_n+1}^{ih_n} z_{n,t} + o_p(1).\end{aligned}$$

Further $E(\sum_{t=1}^{n} z_{n,t})^2 \leq K$ by the construction of $z_{n,t}$. Therefore

$$\begin{aligned}\left|\sum_{i=1}^{r_n} E\left[\mathcal{Z}_{n,i}^2\right] - 1\right| &= \left|\sum_{i=1}^{r_n} E\left[\mathcal{Z}_{n,i}^2\right] - E\left(\sum_{t=1}^{n} z_{n,t}\right)^2\right| + o\left(1\right) \\ &\leq \left|\sum_{i=1}^{r_n} E\left[\mathcal{Z}_{n,i}^2\right] - E\left(\sum_{i=1}^{r_n} \sum_{t=(i-1)h_n+j_n+1}^{ih_n} z_{n,t}\right)^2\right| + o\left(1\right) \\ &= 2\left|\sum_{i=1}^{r_n} \sum_{j=i+1}^{r_n} \sum_{t=(i-1)h_n+j_n+1}^{ih_n} \sum_{s=(j-1)h_n+j_n+1}^{ih_n} E\left[z_{n,s} z_{n,t}\right]\right| + o\left(1\right).\end{aligned}$$

The last line is handled as follows. By Lemma 2.1 $\{y_{n,t}, \Im_t\}$ forms an L_2-mixingale array with arbitrary size and constants $e_{n,t}$. Further, variance bound Lemma 2.2 implies McLeish's [57: Theorem 1.6] maximal inequality holds with mixingale constants $e_{n,t}$ replaced with $||y_{n,t}||$, where $\sigma_n^{-2} \sum_{t=1}^{n} E[y_{n,t}^2] \leq K$ by non-degeneracy Assumption A.ii. Therefore de Jong's [21] Lemma 4 applies by replacing McLeish's [57] inequality with Lemma 2.2 in his proof, hence

$$\sum_{i=1}^{r_n} \sum_{k=i+1}^{r_n} \sum_{t=(i-1)h_n+j_n+1}^{ih_n} \sum_{s=(k-1)h_n+j_n+1}^{ih_n} E\left[z_{n,s} z_{n,t}\right] = o(1).$$

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

PROOF OF LEMMA 3.4 (Approximation)

The proof exploits the following CLT's based on tail array NED Assumption C. Recall $z_t \in \{|y_t|, -y_t I(y_t < 0), y_t I(y_t \geq 0)\}$.

LEMMA A.4. *Under Assumption C there exists a finite $w_1^2 \in \mathbb{R}_+$ such that*

$$m_n^{1/2} \ln\left(z_{(m_n+1)}/c_n\right) \stackrel{\mathfrak{L}}{\to} N\left(0, w_1^2\right),$$

and a mapping $w_2 : \mathbb{R}_+ \to \mathbb{R}_+$, $w_2^2(u) < \infty$ for each $u \geq 0$, such that

$$\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \left\{ I\left(z_t > c_n e^{u/m_n^{1/2}}\right) - P\left(z_t > c_n e^{u/m_n^{1/2}}\right) \right\} \stackrel{\mathfrak{L}}{\to} N\left(0, w_2^2(u)\right).$$

PROOF. We require a simple implication of Markov's inequality. Let $\mathcal{O} : \mathbb{R}_+ \to \mathbb{R}_+$ be a contraction mapping, $\mathcal{O}(z) \in [0, z]$, that may be different in different places. For any stationary L_p-bounded process $\{z_t\}$, $p > 0$, with $||z_t||_q = \infty$ for some finite $q > p$, and any measurable mapping $g : \mathbb{R}_+ \to \mathbb{R}_+$ that satisfies $m_n^{1/2} g(c_n) \to 0$ for positive real sequences $\{c_n, m_n\}$, $c_n \to \infty$ and $m_n \to \infty$, it follows

$$P\left(|z_t| > z\right) = \mathcal{O}\left(z^{-p}\right) \times \left(1 + \mathcal{O}\left(g\left(z\right)\right)\right) \text{ as } z \to \infty. \tag{17}$$

Simply note $P(|z_t| > z) \leq E\left|z_t\right|^{p+\iota} z^{-p-\iota} = Kz^{-p-\iota}$, say, for infinitessimal $\iota > 0$ and use $z^{-\iota} \to 0$ as $z \to \infty$ to deduce the equality.

Now define $S_n\left(u\right) := 1/m_n \sum_{t=1}^{n} I(z_t > c_n e^{u/m_n^{1/2}})$ for arbitrary $u \geq 0$. Under Assumption C and (17) $\{m_n^{-1/2} I(z_t > c_n e^u), \Im_t\}$ satisfies the conditions of Hill's [43: Theorem 2.1] central limit theorem for dependent tail arrays. See Lemma 3.1 of Hill [43] for verification[3]. Therefore, point-wise in $u \in \mathbb{R}_+$,

$$m_n^{1/2}\left\{S_n\left(u\right) - E\{S_n\left(u\right)\}\right\} \xrightarrow{\mathfrak{L}} N(0, w_2^2(u)),\; w_2^2(u) < \infty. \tag{18}$$

We will prove $m_n^{1/2} \ln(z_{(m_n+1)}/c_n) \xrightarrow{\mathfrak{L}} N(0, w_1^2)$ is a consequence of (18). By construction $m_n^{1/2} \ln(z_{(m_n+1)}/c_n) \leq u$ if $S_n(u) \leq \rho$ for $\rho \in (0, 2)$ to be chosen below, while $S_n(u)$ $\leq \rho$ if

$$\begin{aligned} m_n^{1/2}\left(S_n\left(u\right) - E\left[S_n\left(u\right)\right]\right) &\leq m_n^{1/2}\left(\rho - \frac{n}{m_n} P\left(z_t > c_n e^{u/m_n^{1/2}}\right)\right) \\ &= m_n^{1/2}\left(\rho - \frac{n}{m_n} P\left(z_t > c_n\right) \frac{P\left(z_t > c_n e^{u/m_n^{1/2}}\right)}{P\left(z_t > c_n\right)}\right). \end{aligned}$$

Apply (17) and Hsing's [49: p. 1553] argument to deduce as $n \to \infty$

$$\frac{n}{m_n} P\left(z_t > c_n\right) = \mathcal{O}(1) \times \left[1 + \mathcal{O}\left(g\left(c_n\right)\right)\right] = \mathcal{O}(1) + o\left(1/m_n^{1/2}\right)$$

and

$$\frac{P\left(z_t > c_n e^u\right)}{P\left(z_t > c_n\right)} = \mathcal{O}(e^{-u\kappa}) \times \left(1 + \mathcal{O}\left(g\left(c_n\right)\right)\right) = \mathcal{O}(e^{-u\kappa}) \times \left(1 + o\left(1/m_n^{1/2}\right)\right).$$

Now put $\rho = \mathcal{O}(1) \in [0, 1]$ and use $\mathcal{O}(z) \times \mathcal{O}(z) = \mathcal{O}(z)\ \forall z \in [0, 1]$, $\mathcal{O}(1) + o(1/m_n^{1/2})$ $= \mathcal{O}(1)$ for sufficiently large n, and $\mathcal{O}(e^{-u\kappa/m_n^{1/2}}) \leq 1$ to deduce $m_n^{1/2} \ln(z_{(m_n+1)}/c_n) \leq u$ sufficiently if

$$\begin{aligned} &\kappa^{-1} m_n^{1/2}\left(S_n\left(u\right) - E\left[S_n\left(u\right)\right]\right) \\ &\qquad \leq \kappa^{-1} m_n^{1/2}\left(\rho - \left(\mathcal{O}(1) + o\left(1/m_n^{1/2}\right)\right) \times \mathcal{O}(e^{-u\kappa/m_n^{1/2}}) \times \left(1 + o\left(1/m_n^{1/2}\right)\right)\right) \\ &\qquad = \kappa^{-1} m_n^{1/2}\left\{\rho - \mathcal{O}\left(e^{-u\kappa/m_n^{1/2}}\right) \times \left(1 + o\left(1/m_n^{1/2}\right)\right)\right\} \\ &\qquad \leq \kappa^{-1} m_n^{1/2}\left\{u\kappa/m_n^{1/2} + o\left(1/m_n^{1/2}\right)\right\} = u + o\left(1\right). \end{aligned}$$

Since $\kappa^{-1} m_n^{1/2}\{S_n(u) - E[S_n(u)]\} \xrightarrow{\mathfrak{L}} Z$ a mean-zero normal law with finite variance, it follows

$$\begin{aligned} &\lim_{n\to\infty} P\left(m_n^{1/2} \ln\left(z_{(m_n+1)}/c_n\right) \leq u\right) \\ &\qquad = \lim_{n\to\infty} P\left(\kappa^{-1} m_n^{1/2}\left(S_n\left(u\right) - E\left[S_n\left(u\right)\right]\right) \leq u + o(1)\right) = P\left(Z \leq u\right). \end{aligned}$$

[3] Hill [43] assumes $P(|z_t| > z)$ is regularly varying since the premise and major applications there concern tail index estimation. The arguments there trivially carry over to (16).

Therefore $m_n^{1/2}\ln(z_{(m_n+1)}/c_n) \overset{\mathfrak{L}}{\rightarrow} N(0,w_1^2)$ for some $w_1^2 < \infty$. $\mathcal{QED}$.

We are now ready to prove the claim.

Proof of Lemma 3.4. For notational simplicity assume $y_t \geq 0$ $a.s.$ such that $y_{n,t} := y_t I(y_t \leq c_n)$ and $\hat{y}_{n,t} := y_t I(y_t \leq y_{(m_n+1)})$ where

$$1_n := \frac{n}{m_n} P\left(y_t > c_n\right) \rightarrow 1.$$

Similarly assume $w_{n,t} \geq 0$ $a.s.$ for all t and n for simplicity. Define the sample number of exceedances $y_t > c_n$:

$$m_n^* := \sum_{t=1}^{n} I\left(y_t > c_n\right).$$

Note m_n^* is a random variable measurable with respect to $\Im_n = \sigma(y_t : t \leq n)$, and recall $\Sigma_a^b(\cdot) = 0$ if $b < a$.

Let $t_{n,i}^*$ correspond to the (stochastic) period $t \in \{1,...,n\}$ in which the i^{th} rank $y_{(i)}$ occurs, and write

$$\begin{aligned}\sum_{t=1}^{n} w_{n,t}\left\{\hat{y}_{n,t} - y_{n,t}\right\} &= \sum_{i=1}^{n} w_{n,t_{n,i}^*} y_{(i)} \left\{I\left(y_{(i)} \leq y_{(m_n+1)}\right) - I\left(y_{(i)} \leq c_n\right)\right\} \quad (19)\\ &= \sum_{i=m_n+1}^{m_n^*} w_{n,t_{n,i}^*} y_{(i)} - \sum_{i=m_n^*+1}^{m_n} w_{n,t_{n,i}^*} y_{(i)} =: \mathcal{A}_n + \mathcal{B}_n.\end{aligned}$$

By convention $\Sigma_{a_n}^{b_n}(\cdot) = 0$ if $\liminf_{n\rightarrow\infty} a_n/b_n > 1$, hence at least one $\mathcal{A}_n$ or $\mathcal{B}_n$ is identically zero, and both are zero if $m_n^* = m_n$. Which is zero depends on the sample draw $\{y_t\}$ and therefore the random variable m_n^*. We first characterize probability bounds for $\mathcal{A}_n$ and $\mathcal{B}_n$, and then use the bounds to prove the claim.

Step 1 (bounds): Define $\varepsilon_n := L(n)/m_n^{1/2}$ for an arbitrarily s.v. $L(n) \rightarrow \infty$, use the assumption $\max_{1\leq t\leq n}\{w_{n,t}\} = O_p(1)$ and the order statistic $y_{(i)}$ construction to deduce

$$\begin{aligned}\mathcal{A}_n &= O_p\left(\sum_{i=m_n+1}^{[m_n(1_n+\varepsilon_n)]} y_{(i)}\right) + O_p\left(\sum_{i=[m_n(1_n+\varepsilon_n)]+1}^{m_n^*} y_{(i)}\right) \quad (20)\\ &= O_p\left(\sum_{i=m_n+1}^{[m_n(1_n+\varepsilon_n)]} y_{(i)}\right) + O_p\left(\left(m_n^* - \left[m_n\left(1_n+\varepsilon_n\right)\right]\right)_+ \times y_{(m_n+1)}\right)\\ &= O_p\left(\sum_{i=m_n+1}^{[m_n(1_n+\varepsilon_n)]} y_{(i)}\right) + o_p(c_n)\end{aligned}$$

and

$$\begin{aligned}\mathcal{B}_n &= O_p\left(\sum_{i=m_n^*+1}^{[m_n(1_n-\varepsilon_n)]} y_{(i)}\right) + O_p\left(\sum_{i=[m_n(1_n-\varepsilon_n)]+1}^{m_n} y_{(i)}\right) \qquad (21)\\ &= O_p\left(\sum_{i=[m_n(1-\varepsilon_n)]+1}^{m_n} y_{(i)}\right) + O_p\left(\left(m_n^* - [m_n(1_n-\varepsilon_n)]\right)_+ \times y_{([m_n(1_n-\varepsilon_n)]+1)}\right)\\ &\qquad + O_p\left(\left(m_n^* - [m_n(1_n-\varepsilon_n)]\right)_+ \times \left(y_{(m_n^*+1)} - y_{([m_n(1_n-\varepsilon_n)]+1)}\right)_+\right)\\ &= O_p\left(\sum_{i=[m_n(1-\varepsilon_n)]+1}^{m_n} y_{(i)}\right) + o_p(c_n).\end{aligned}$$

The last line in (20) follows from Chebyshev's inequality, $\varepsilon_n m_n^{1/2}/L(n) = 1$ and Lemma A.4 since $y_{(m_n+1)}/c_n = 1 + O_p(1/m_n^{1/2})$ and

$$\begin{aligned}&P\left(m_n^* \geq m_n(1_n+\varepsilon_n)\right)\\ &\qquad \leq P\left(m_n^{1/2-\iota}\left|\frac{1}{m_n}\sum_{t=1}^{n} I(y_t > c_n) + 1_n\right| \geq m_n^{1/2-\iota}\varepsilon_n\right)\\ &\qquad \leq K \times E\left[\left(m_n^{1/2-\iota}\left\{\frac{1}{m_n}\sum_{t=1}^{n}\left\{I(y_t > c_n) - P(y_t > c_n)\right\}\right\}\right)^2\right] = o(1).\end{aligned}$$

The last line in (21) follows by similar reasoning because $(n/[m_n(1_n - \varepsilon_n)])P(|y_t| > c_n)$ $= 1_n/(1_n - \varepsilon_n) \to 1$ ensures Lemma A.4 applies to $y_{([m_n(1_n-\varepsilon_n)]+1)}/c_n$. Thus

$$\begin{aligned}&P\left(m_n^* \leq m_n(1-\varepsilon_n)\right)\\ &\qquad = P\left(\frac{1}{m_n}\sum_{t=1}^{n} I(y_t > c_n) - 1_n \leq -\varepsilon_n\right)\\ &\qquad \leq P\left(m_n^{1/2-\iota}\left|\frac{1}{m_n}\sum_{t=1}^{n}\left\{I(y_t > c_n) - P(y_t > c_n)\right\}\right| \geq m_n^{1/2-\iota}\varepsilon_n\right) = o(1)\end{aligned}$$

hence

$$P\left(\left(y_{(m_n^*+1)} - y_{([m_n(1_n-\varepsilon_n)]+1)}\right)_+ \geq 0\right) = P\left(m_n^* \leq m_n(1-\varepsilon_n)\right) = o(1).$$

Step 2 (approximation): We now prove the claim. Use (19)-(21), the triangle inequality and $y_{([m_n(1_n+\varepsilon_n)]+1)} \leq y_{([m_n(1_n-\varepsilon_n)]+1)}$ to obtain

$$\left|\sum_{t=1}^{n} w_{n,t}\left\{\hat{y}_{n,t}-y_{n,t}\right\}\right|$$
$$\leq \left|\sum_{i=m_n+1}^{m_n^*} w_{n,t^*_{n,i}} y_{(i)} - m_n\varepsilon_n y_{([m_n(1_n+\varepsilon_n)]+1)}\right|$$
$$+ \left|\sum_{i=m_n^*+1}^{m_n} w_{n,t^*_{n,i}} y_{(i)} - m_n\varepsilon_n y_{([m_n(1_n-\varepsilon_n)]+1)}\right| + o_p(c_n)$$
$$= O_p\left(\sum_{i=m_n+1}^{[m_n(1_n+\varepsilon_n)]} \left|y_{(i)} - y_{([m_n(1_n+\varepsilon_n)]+1)}\right|\right) + O_p\left(\sum_{i=[m_n(1_n-\varepsilon_n)]+1}^{m_n} \left|y_{(i)} - y_{([m_n(1_n-\varepsilon_n)]+1)}\right|\right)$$
$$+ O_p\left(\left|m_n\left(1_n+\varepsilon_n\right) - \left[m_n\left(1_n+\varepsilon_n\right)\right]\right| \times y_{([m_n(1_n+\varepsilon_n)]+1)}\right)$$
$$+ O_p\left(\left|m_n\left(1_n-\varepsilon_n\right) - \left[m_n\left(1_n-\varepsilon_n\right)\right]\right| \times y_{([m_n(1_n-\varepsilon_n)]+1)}\right) + o_p(c_n)$$
$$= \mathcal{C}_n + \mathcal{D}_n + \mathcal{E}_n + \mathcal{F}_n + o_p(c_n),$$

say.

We will now show $\{\mathcal{C}_n, \mathcal{D}_n, \mathcal{E}_n, \mathcal{F}_n\} = O_p\left(c_n L(n)\right)$ for s.v. $L(n) \to \infty$. The proofs for $\mathcal{C}_n$ and $\mathcal{D}_n$ are identical so consider $\mathcal{C}_n$. Since Lemma A.4 applies equally to $y_{(m_n+1)}/c_n$ and $y_{([m_n(1_n+\varepsilon_n)]+1)}/c_n$, use the triangle inequality to deduce

$$\frac{1}{c_n L(n)} \sum_{i=m_n+1}^{[m_n(1_n+\varepsilon_n)]} \left|y_{(i)} - y_{([m_n(1_n+\varepsilon_n)]+1)}\right|$$
$$= O\left(\frac{m_n}{L(n)}\varepsilon_n\right) \times \left(\left|\frac{y_{(m_n+1)}}{c_n} - 1\right| + \left|\frac{y_{([m_n(1_n+\varepsilon_n)]+1)}}{c_n} - 1\right|\right)$$
$$= O_p\left(\frac{m_n^{1/2}}{L(n)}\varepsilon_n\right) = O_p\left(1\right),$$

hence $\mathcal{C}_n = O_p\left(c_n L(n)\right)$.

Finally, for $\mathcal{E}_n$ the truncation implies $z - [z] \in [0,1]$ for positive z, hence by Lemma A.4

$$\frac{\mathcal{E}_n}{c_n L(n)} = \left|\frac{m_n\left(1_n+\varepsilon_n\right) - \left[m_n\left(1_n+\varepsilon_n\right)\right]}{L(n)}\right| \times \frac{y_{([m_n(1_n+\varepsilon_n)]+1)}}{c_n}$$
$$\leq K\frac{1}{L(n)} \times \left(1 + O_p\left(1/m_n^{1/2}\right)\right) = o_p(1).$$

An identical argument applies to $\mathcal{F}_n$. Therefore $\sum_{t=1}^{n} w_{n,t}\{\hat{y}_{n,t} - y_{n,t}\} = O_p(c_n L(n))$.

Since $L(n)$ is arbitrary and s.v., and $m_n/(\ln(n))^{1+\iota} \to \infty$ under Assumption A.iii, we can always choose $L(n) = o(m_n^{1/2})$. Now invoke Lemma 3.1 $c_n = O((\sum_{t=1}^{n} E[y_{n,t}^2])^{1/2}/m_n^{1/2})$, Assumption A.ii $\sum_{t=1}^{n} E[y_{n,t}^2] = O(\sigma_n)$, and $L(n) = o(m_n^{1/2})$ to conclude $c_n L(n) = o((\sum_{t=1}^{n} E[y_{n,t}^2])^{1/2}) = o(\sigma_n)$. $\mathcal{QED}$.

PROOF OF THEOREM 4.1 (Kernel Variance)

We require one preliminary result.

LEMMA A.5. *Under Assumptions A-D* $\mathcal{S}_{1,n} := \sum_{s,t=1}^{n} k_{n,s,t}(\hat{y}_{n,s} - y_{n,s})(\hat{y}_{n,t} - y_{n,t})$ *and* $\mathcal{S}_{2,n} := \sum_{s,t=1}^{n} k_{n,s,t}(\hat{y}_{n,s} - y_{n,s})(y_{n,t} - E[y_{n,t}])$ *are* $o_p\left(\sigma_n^2\right)$.

PROOF. We will only prove $\mathcal{S}_{1,n} = o_p(\sigma_n^2)$ since showing $\mathcal{S}_{2,n} = o_p(\sigma_n^2)$ is nearly identical. Recall by definition $k_{n,s,t}$ denotes $k((s-t)/\gamma_n)$ and define (see also Davidson and de Jong [19: p. 412-413])

$$\eta_\delta(x) := \frac{1}{\left(2\delta^2\pi\right)^{1/2}} \exp\left\{-x^2\delta^{-2}/2\right\}$$

$$\begin{aligned}\mathcal{S}_{1,n\delta} &:= \frac{1}{\sigma_n^2}\sum_{t=-n+1}^{2,n}\left(\frac{1}{\gamma_n^{1/2}}\sum_{l=1-t}^{n-t} k\left(l/\gamma_n\right)\left(\hat{y}_{n,t+l} - y_{n,t+l}\right) I\left(0 \le l \le [\gamma_n/\delta]\right)\right) \\ &\quad\times\left(\frac{1}{\gamma_n^{1/2}}\sum_{j=1-t}^{n-t} \eta_\delta\left(j/\gamma_n\right)\left(\hat{y}_{n,t+j} - y_{n,t+j}\right) I\left(0 \le j \le [\gamma_n/\delta]\right)\right) \times \left(1 + o_p(1)\right) \\ &= \frac{1}{\sigma_n^2}\sum_{t=-n+1}^{2,n} s_{1,n,t}(\delta) \times s_{2,n,t}(\delta) \times \left(1 + o_p(1)\right),\end{aligned}$$

say.

Since Lemmas 3.3 and 3.4 and the Helly-Bray theorem together imply $||\sigma_n^{-1}\sum_{t=1}^{n}(\hat{y}_{n,t} - y_{n,t})|| = o(1)$, Davidson and de Jong's [19] Lemmas A.2-A.3 suffice to prove[4]

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\left\|\mathcal{S}_{1,n} - \mathcal{S}_{1,n\delta} \times \left(1 + o_p(1)\right)\right\|_1 = 0.$$

The proof is complete if we show $\lim_{\delta\to 0}\limsup_{n\to\infty}||\mathcal{S}_{1,n\delta}||_1 = 0$. We prove each

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\max_{-n+1\le t\le 2,n}\left\|\frac{n^{1/2}}{\sigma_n}s_{i,n,t}(\delta)\right\|_2 = 0$$

since the desired limit then follows from Minkowski and Cauchy-Schwarz inequalities.

Consider $s_{1,n,t}(\delta)$, $s_{2,n,t}(\delta)$ being nearly identical, and define

$$N_n(\delta) := \min\left\{n, [\gamma_n/\delta] + 1\right\}$$

for arbitrary $\delta > 0$. Lemmas 3.3 and 3.4 extend straightforwardly to weighted versions with weights $k(t/\gamma_n)$ under kernel property Assumption D. Therefore, by the Helly-Bray theorem

$$\left\|\frac{1}{\sigma_{N_n(\delta)}}\sum_{t=1}^{N_n(\delta)} k\left(t/\gamma_n\right)\left(\hat{y}_{n,t+l} - y_{n,t+l}\right)\right\|_2 \to 0.$$

[4] Define, e.g., $X_{n,t} := \sigma_n^{-1}\{\hat{y}_{n,t} - y_{n,t}\}$. Davidson and de Jong [19: p. 414] exploit $E(\sum_{t=1}^{n} X_{n,t})^2 \le K\sum_{t=1}^{n} e_{n,t}^2 = O(1)$ for some triangular array $\{e_{n,t}\}$ by their Lemma A.1, which holds by their L_2-NED Assumption 2 and McLeish's [57: Theorem 1.6] maximal inequality. Close inspection of their arguments show $E(\sum_{t=1}^{n} X_{n,t})^2 = o(1)$ and kernel property Assumption D are the only properties required. In our case under variance bound Lemma 2.2 and non-degeneracy Assumption A.ii $E(\sigma_n^{-1}\sum_{t=1}^{n}\{\hat{y}_{n,t} - y_{n,t}\})^2 = o(1) \le \sigma_n^{-2}\sum_{t=1}^{n} E[y_{n,t}^2] \le K\sum_{t=1}^{n}(1/n^{1/2})^2 \le K$ replaces McLeish's maximal inequality with mixingale constants $1/n^{1/2}$. A similar argument extends to $y_{n,t} - E[y_{n,t}]$ in $\mathcal{S}_{2,n}$: by Lemma 2.2 and Assumption A.ii $E(\sigma_n^{-1}\sum_{t=1}^{n} y_{n,t})^2 \le K\sigma_n^{-2}\sum_{t=1}^{n} E[y_{n,t}^2] \le K\sum_{t=1}^{n}(1/n^{1/2})^2 \le K$.

Further by construction and non-degeneracy Assumption A.i,ii

$$\limsup_{n\geq N}\left\{\frac{N_n(\delta)}{\gamma_n}\right\}\leq K \quad \text{and} \quad \frac{\sigma^2_{N_n(\delta)}/N_n(\delta)}{\sigma^2_n/n}=O(1).$$

Therefore

$$\begin{aligned}
&\lim_{n\to\infty}\max_{-n+1\leq t\leq 2,n}\left\|\frac{n^{1/2}}{\sigma_n}s_{1,n,t}\right\|_2 \\
&= \lim_{n\to\infty}\max_{-n+1\leq t\leq 2,n}\left\|\frac{n^{1/2}}{\sigma_n}\frac{1}{\gamma_n^{1/2}}\sum_{l=1-t}^{n-t}k\left(l/\gamma_n\right)\left(\hat{y}_{n,t+l}-y_{n,t+l}\right)I\left(0\leq l\leq[\gamma_n/\delta]\right)\right\|_2 \\
&\leq \lim_{n\to\infty}\left\{\frac{N_n^{1/2}(\delta)}{\gamma_n^{1/2}}\frac{\sigma_{N_n(\delta)}/N_n^{1/2}(\delta)}{\sigma_n/n^{1/2}}\right\}\lim_{n\to\infty}\left\|\frac{1}{\sigma_{N_n(\delta)}}\sum_{t=1}^{N_n(\delta)}k\left(t/\gamma_n\right)\left(\hat{y}_{n,t+l}-y_{n,t+l}\right)\right\|_2 \\
&= 0.
\end{aligned}$$

But this implies $\limsup_{n\to\infty}\max_{-n+1\leq t\leq 2,n}||n^{1/2}\sigma_n^{-1}s_{1,n,t}(\delta)||_2 = 0$ for all δ, so take $\lim_{\delta\to 0}$ to deduce the required limit. $\mathcal{QED}$.

The proof of kernel estimator consistency follows.

Proof of Theorem 4.1. The triangle inequality implies

$$\left|\hat{\sigma}^2_n-\sigma^2_n\right|\leq\left|\hat{\sigma}^2_n-\tilde{\sigma}^2_n\right|+\left|\tilde{\sigma}^2_n-\sigma^2_n\right|, \tag{22}$$

where $\tilde{\sigma}^2_n$ is a kernel estimator based on deterministic trimming $y_{n,t}$:

$$\tilde{\sigma}^2_n := \sum_{s,t=1}^{n}k_{n,s,t}\times\left\{y_{n,s}-E\left[y_{n,t}\right]\right\}\times\left\{y_{n,t}-E\left[y_{n,t}\right]\right\}.$$

We show each part in (22) is $o_p(\sigma^2_n)$. The first step proceeds by a standard bounding argument based on Theorem 3.2 and Lemma 3.3 to show $\hat{\sigma}^2_n$ behaves like $\tilde{\sigma}^2_n$. The second step shows $\tilde{\sigma}^2_n/\sigma^2_n\xrightarrow{p}1$ by a generalization of arguments in Davidson and de Jong [19].

Step 1 ($\hat{\sigma}^2_n-\tilde{\sigma}^2_n$): Decompose $\hat{\sigma}^2_n-\tilde{\sigma}^2_n=\sum_{i=1}^{5}\mathcal{S}_{i,n}$ where

$$\mathcal{Y}_{1,n}=\sum_{s,t=1}^{n}k_{n,s,t}\left(\hat{y}_{n,s}-y_{n,s}\right)\left(\hat{y}_{n,t}-y_{n,t}\right), \quad \mathcal{Y}_{2,n}=2\sum_{s,t=1}^{n}k_{n,s,t}\left(\hat{y}_{n,s}-y_{n,s}\right)\left(y_{n,t}-E\left[y_{n,t}\right]\right)$$

$$\mathcal{Y}_{3,n}=-2\left(\hat{y}_n-E\left[y_{n,t}\right]\right)\sum_{s,t=1}^{n}k_{n,s,t}\left(\hat{y}_{n,s}-y_{n,s}\right), \quad \mathcal{Y}_{4,n}=-2\left(\hat{y}_n-E\left[y_{n,t}\right]\right)\sum_{s,t=1}^{n}k_{n,s,t}\left(y_{n,s}-E\left[y_{n,t}\right]\right)$$

$$\mathcal{Y}_{5,n}=\left(\hat{y}_n-E\left[y_{n,t}\right]\right)^2\sum_{s,t=1}^{n}k_{n,s,t}.$$

The first two terms $\mathcal{Y}_{1,n}$ and $\mathcal{Y}_{2,n}$ are $o_p\left(\sigma^2_n\right)$ by Lemma A.5.

Consider the remaining three terms. First, apply kernel property Assumption D and CLT-ST Theorem 3.2 to deduce

$$\left(\hat{y}_n-E\left[y_{n,t}\right]\right)^2\sum_{s,t=1}^{n}\left|k_{n,s,t}\right|=O_p\left(\sigma^2_n/n^2\right)\times o\left(n^2\right)=o_p\left(\sigma^2_n\right).$$

Second, CLT-DT Lemma 3.3 extends in a straightforward way to the weighted triangular array $\{k_{n,s,t}(y_{n,t} - E[y_{n,t}]) : 1 \leq t \leq n\}_{n \geq 1}$ for each s since $\max_{1 \leq s,t \leq n} |k_{n,s,t}| = O(1)$ by Assumption D.ii. Similarly, approximation Lemma 3.4 instantly applies to $\{k_{n,s,t}(\hat{y}_{n,t} - y_{n,t}) : 1 \leq t \leq n\}_{n \geq 1}$ for each s. Therefore

$$\max_{1 \leq s \leq n} \left| \sum_{s,t=1}^{n} k_{n,s,t} \left(y_{n,t} - E\left[y_{n,t} \right] \right) \right| = o_p\left(n\sigma_n \right), \quad \max_{1 \leq s \leq n} \left| \sum_{s,t=1}^{n} k_{n,s,t} \left(\hat{y}_{n,t} - y_{n,t} \right) \right| = o_p\left(n\sigma_n \right).$$

But this implies

$$\mathcal{Y}_{3,n} = -2\left(\hat{y}_n - E\left[y_{n,t} \right] \right) \sum_{s,t=1}^{n} k_{n,s,t} \left(\hat{y}_{n,s} - y_{n,s} \right) = O_p\left(\sigma_n / n \right) \times o_p\left(n\sigma_n \right) = o_p\left(\sigma_n^2 \right)$$

$$\mathcal{Y}_{4,n} = -2\left(\hat{y}_n - E\left[y_{n,t} \right] \right) \sum_{s,t=1}^{n} k_{n,s,t} \left(y_{n,s} - E\left[y_{n,t} \right] \right) = O_p\left(\sigma_n / n \right) \times o_p\left(n\sigma_n \right) = o_p\left(\sigma_n^2 \right)$$

$$\mathcal{Y}_{5,n} = \left(\hat{y}_n - E\left[y_{n,t} \right] \right)^2 \sum_{s,t=1}^{n} k_{n,s,t} = O_p\left(\sigma_n^2 / n^2 \right) \times o\left(n^2 \right) = o_p\left(\sigma_n^2 \right).$$

Step 2 ($\tilde{\sigma}_n^2/\sigma_n^2 - 1$): We need only verify Assumptions 1-3 of Davidson and de Jong [19], denoted DJ, to show $|\tilde{\sigma}_n^2/\sigma_n^2 - 1| \xrightarrow{p} 0$ by their Theorem 2.1. DJ's Assumption 1 holds by Assumption D.

First we translate DJ's environment into ours for clarity. They use their NED Assumption 2 to ensure an L_2-mixingale property for a standardized uniformly L_2-bounded process $z_{n,t} = \{y_{n,t} - E[y_{n,t}]\}/\sigma_n$ with size 1/2 and square-summable constants $\check{e}_{n,t} := e_{n,t}/\sigma_n$: $\sum_{t=1}^{n} \check{e}_{n,t}^2 = \sum_{t=1}^{n} e_{n,t}^2 \sigma_n \leq K$. Their sole use of this assumption is to invoke McLeish's [57] Theorem 1.6 maximal inequality to reach a bounded partial sum variance:

$$E\left(\sum_{t=1}^{n} z_{n,t} \right)^2 \leq K \sum_{t=1}^{n} \check{e}_{n,t}^2 \leq K.$$

But under our Assumptions A and B $\{y_{n,t}, \Im_t\}$ forms an L_2-mixingale array with size 1/2 by Lemma 2.1, and has by Lemma 2.2 a partial sum variance bound

$$E\left(\frac{1}{\sigma_n} \sum_{t=1}^{n} \left\{ y_{n,t} - E\left[y_{n,t} \right] \right\} \right)^2 \leq K \frac{1}{nE\left[y_{n,t} \right]^2} \sum_{t=1}^{n} E\left[y_{n,t} \right]^2 = K \sum_{t=1}^{n} \left(1/n^{1/2} \right)^2 \leq K. \quad (23)$$

A careful inspection of DJ's proof of their Theorem 2.1 reveals a L_2-mixingale property with size 1/2, constants $n^{-1/2}$ and therefore bound (22) suffice in place of their Assumption 2.

Finally, again translating DJ's environment again into ours, their Assumption 3 $\gamma_n \max_{1 \leq t \leq n}\{\check{e}_{n,t}^2\} = o(1)$ is invoked solely to ensure a partial variance bound for an L_2-mixingale array function of $\check{w}_{n,t}$ based on McLeish's inequality. See the proofs of their Lemmas A.3 and A.4. By our Lemma 2.2 we can always side-step the use of coefficients in the required partial variance bounds: we can always replace $\check{e}_{n,t}$ with $||y_{n,t}||/(n^{1/2}||y_{n,t}||) = n^{-1/2}$ as in (23). Since $\gamma_n \times n^{-1} = o(n/n) = o(1)$ under Assumption D, DJ's Assumption 3 is satisfied. $\mathcal{QED}$.

PROOF OF LEMMA 5.1 (Positive Definiteness)

Distribution smoothness ensures Assumption A.i holds by Section 5.1. We need only prove positive definiteness since Assumption A.ii follows instantly.

Define $I_{n,t}^{(y)} := I(|y_t| \leq c_n)$ and $I_{n,N} := \prod_{i=1}^{N} I_{n,i}$ for any $1 \leq N \leq n$, and note $y_{n,t} = y_t I_{n,t}^{(y)}$. Stationarity implies $p_n(s,t) := E[y_{n,s}y_{n,t}]/E[y_{n,t}^2]$. Define $\mathcal{Y}_N(r) := \sum_{t=1}^{N} r_t y_t$ for any $r \in \mathbb{R}^N$, $r'r = 1$. By monotone convergence and $I_{h,n,t} \to 1$ *a.s.* it follows for any $i, j = 0, ..., N-1$

$$E\left[y_{t-i}y_{t-j}I_{n,N}\right] = E\left[y_{t-i}I_{n,t-i}^{(y)}y_{t-j}I_{n,t-j}^{(y)}\right] \times (1 + o(1)).$$

Therefore

$$\sum_{s,t=1}^{N} r_s r_t p_n(s,t) = \frac{1}{E\left[y_{n,t}^2\right]} \sum_{s,t=1}^{N} r_s r_t E\left[y_s I_{n,s}^{(y)} y_t I_{n,t}^{(y)}\right] = \frac{E\left[\mathcal{Y}_N^2(r) I_{n,N}\right]}{E\left[y_{n,t}^2\right]} \times (1 + o(1)).$$

The proof is complete if we show $\inf_{r'r=1} E[\mathcal{Y}_N^2(r)I_{n,N}]/E[y_{n,t}^2] > 0 \; \forall n \geq N$ and some $N \in \mathbb{N}$.

By stationarity $\mathcal{Y}_N(r) = \sum_{i=0}^{N-1} \epsilon_{N-i} \sum_{j=0}^{i} r_{N-j}\psi_{i-j} + \sum_{i=0}^{\infty} \epsilon_{-i} \sum_{j=0}^{N-1} r_{N-j}\psi_{N+i-j} = \sum_{j=0}^{\infty} \check{\psi}_{N,j}(r)\epsilon_{N-j}$ for some sequence of constant real numbers $\{\check{\psi}_{N,j}(r)\}_{j=0}^{\infty}$, $\inf_{j \geq N} \sup_{r'r=1} |\check{\psi}_{N,j}(r)| = O(\rho^j)$, and trivially $y_t = \mathcal{Y}_N(r)$ for $r = [1, 0, ..., 0]'$. Therefore $\mathcal{Y}_N(r)$ satisfies

$$\sup_{r'r=1} \left| \frac{P\left(|\mathcal{Y}_N(r)| > y\right)}{\sum_{j=0}^{\infty} \left|\check{\psi}_{N,j}(r)\right|^{\kappa}} - y^{-\kappa} L(y) \right| \to 0$$

where $\kappa \in (0, 2]$ is the same tail index and $L(\cdot)$ is the same s.v. component of the error distribution (e.g. Brockwell and Cline [6]). Further, $\liminf_{n\to\infty} \inf_{r'r=1} E[\mathcal{Y}_t^2(r)I_{n,N}] > 0$ by trimming negligibility and the fact that $E[\epsilon_t^2] = \infty$.

Since y_t and $\mathcal{Y}_N(r)$ each have the representation $\sum_{j=0}^{\infty} \check{\psi}_{N,j}(r)\epsilon_{t-j}$ with the same tail decay rate up to a constant scale $\sum_{i=0}^{\infty} \left|\check{\psi}_{N,j}(r)\right|^{\kappa} < \infty$, and are each trimmed by $I_{n,t}$, it must be the case that the tail-trimmed moments $E[\mathcal{Y}_t^2(r)I_{n,N}]$ and $E[y_t^2 I_{n,N}]$ are proportional by Karamata's Theorem: $\inf_{r'r=1} E[\mathcal{Y}_N^2(r)I_{n,N}]/E[y_t^2 I_{n,N}] > 0 \; \forall n \geq N$ and some $N \in \mathbb{N}$ which completes the proof. $\mathcal{QED}$.

REFERENCES

[1] Andrews, D.W.K. (1984). Non-Strong Mixing Autoregressive Processes, *J. Appl. Probab.* 21, 930-934.

[2] Andrews, D.W.K (1991). Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, 817-858..

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

[4] Berkes, I., L. Horváth and J. Schauer (2011). Asymptotics of Trimmed CUSUM Statistics, *Bernoulli*: in press.

[5] Bickel, P. and P. Bühlmann (1999). A New Mixing Notion and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series, *Bernoulli* 5-3, 413-446.

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

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

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

[9] Chen, X. and H. White (1998). Central Limit and Functional Central Limit Theorems for

Hilbert-Valued Dependent Heterogeneous Arrays with Applications, *Economeric Theory* 14, 260-284.

[10] Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, *Probab. Theory Related Fields* 72, 529–557.

[11] Cline, D.B.H. and H-m. H. Pu (2004). Stability and the Lyapunov Exponent of Threshold AR-ARCH Models, *Ann. Appl. Probab.* 14, 1920-1949.

[12] Cribari-Neto, F., S. Ferrari, and G.M. Cordeiro (2000). Improved Heteroscedasticity-Consistent Covariance Matrix Estimators, *Biometrika* 87, 907-918.

[13] Csörgő, S., L. Horváhth and D.M. Mason (1986). What Portion of the Sample Makes a Partial Sum Asymptotically Stable or Normal, *Probab. Theory Related Fields* 72, 1-16.

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

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

[16] Davidson, J. (1993). The Central Limit Theorem for Globally Nonstationary 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, *J. Bus. Econ. Statist.* 22, 16-29.

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

[20] Davis, R.A., K. Knight and J. Liu (1992). M-Estimation for Autoregressions with Infinite Variance, *Stoch. Proc. Appl.* 40, 145-180.

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

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

[23] Dedecker, J. and F. Merlevéde (2002). Necessary and Sufficient Conditions for the Conditional Central Limit Theorem, *Ann. Probab.* 30, 1044-1081.

[24] Dedecker, J., P. Doukhan, G. Lang, J.R. Leon and S. Louhichi (2007). Weak Dependence with Examples and Applications, *Lecture Notes in Statitics* Vol. 190: Springer.

[25] Doukhan, P. and S. Louhichi (1999). A New Weak Dependence Condition and Application to Moment Inequalities, *Stoch. Proc. Appl.* 84, 313-343.

[26] Feller, W. (1967). On Regular Variation and Local Limit Theorems, *Proc. Fifth Berkeley Symposium Math. Statist. Probab.* 2, 373-388, Univ. California Press, Berkeley.

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

[28] Gnedenko, B.V. (1943). Sur la Distribution Limite du Terme Maximum d'une Série Aleatoire, *Ann. Math.* 44, 423-453.

[29] Gordin, M. I. (1969). The Central Limit Theorem for Stationary Processes, *Soviet Math. Dokl.* 10, 1174-1176.

[30] Gordin, M. I. (1973). *Abstracts of Communications*, T.1: A-K, International Conference on Probability, Vilnius.

[31] Gorodetskii, V. V. (1977). On the Strong Mixing Property for Linear Sequences, *Theory Probab. Appl.* 22, 411-413.

[32] Griffin, P.S. and W.E. Pruitt (1987). The Central Limit Problem for Trimmed Sums, *Math. Proc. Cambridge. Philos. Soc.* 102, 329-349.

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

[34] Guegan D., and S. Ladoucette (2001). Non-mixing Properties of Long Memory Processes,

Comptes Rendus de l'Academie des Sciences Series I Mathematics 333, 373-376.
[35] Hahn, M.G., J. Kuelbs and J. Samur (1987). Asymptotic Normality of Trimmed Sums of Mixing Random Variables, *Ann. Probab.* 15, 1395-1418.
[36] Hahn, M.G., J. Kuelbs and D.C. Weiner (1990a). The Asymptotic Distribution of Magnitude-Winsorized Sums via Self-Normalization, *J. Theor. Probab.* 3, 137-168.
[37] Hahn, M.G., J. Kuelbs and D.C. Weiner (1990b). The Asymptotic Joint Distribution of Self-Normalized Censored Sums and Sums of Squares, *Ann. Probab.* 18, 1284-1341.
[38] Hahn, M.G. and D.C. Weiner (1992). Asymptotic Behavior of Self-Normalized Trimmed Sums: Nonnormal Limits, *Ann. Probab.* 20, 455-482.
[39] Hall, P. F.I. Nicholas, and B. Hoffmann (1994). On the Nonparametric Estimation of Covariance Functions, *Ann. Stat.* 22, 2115-2134.
[40] Hannan, E.J. (1973). Central Limit Theorems for Time Series Regression, *Z. Wahrsch. Verw. Gebiete* 26, 157–170.
[41] Hansen, B. E. (1992). Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes, *Econometrica* 60, 967-972.
[42] Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, *Ann. Math. Statist.* 3, 1163-1174.
[43] Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, *J. Statist. Plann. Infer.* 139, 2091-2110.
[44] Hill, J.B. (2010a). On Tail Index Estimation for Dependent, Heterogeneous Data, *Econometric Theory* 26, 1398-1436.
[45] Hill, J.B. (2010b). A New Moment Bound and Weak Laws for Mixingale Arrays without Memory or Heterogeneity Restrictions, with Applications to Tail Trimmed Arrays, Dept. of Economics, University of North Carolina - Chapel Hill.
[46] Hill, J.B. (2011). Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics, *Econometric Theory*: forthcoming.
[47] Hill, J.B. and E. Renault (2010). Generalized Method Moments with Tail Trimming, Dept. of Economics, University of North Carolina-Chapel Hill.
[48] Horowitz, J. L. (1998). Bootstrap Methods for Covariance Structures, *J. Hum. Res.* 33, 39-61.
[49] Hsing, T. (1991). On Tail Index Estimation Using Dependent Data, *Ann. Statist.* 19, 1547-1569.
[50] Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhof: Berlin.
[51] Jenish, N. and I.R. Prucha (2010). On Spatial Processes and Asymptotic Inference under Near-Epoch Dependence, mimeo, Dept. of Economics, New York University.
[52] Kallenberg, O. (2002). *Foundations of Modern Probability.* Springer: New York.
[53] Kauermann, G. and R.T. Carroll (2001). A Note on Efficiency of Sandwich Covariance Matrix Estimation, *J. Amer. Stat. Assoc.* 96, 1387-1396.
[54] Kim, S-J (1992). The Metrically Trimmed Mean as a Robust Estimator of Location, *Ann. Statist.* 20, 1534-1547.
[55] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag: New York.
[56] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles, *Ann. Probab.* 2, 620-628.
[57] McLeish, D.L. (1975). Maximal Inequality and Dependent Strong Laws, *Ann. Probab.* 3, 829-839.
[58] McLeish, D.L. (1977). On the Invariance Principle for Nonstationary Mixingales, *Ann. Probab.* 5,616-621.
[59] Meitz M. and P. Saikkonen (2008). Stability of Nonlinear AR-GARCH Models, *J. Time Ser.*

Anal. 29, 453-475.
[60] Naveau, P. (2003). Almost Sure Relative Stability of the Maximum of a Stationary Sequence, *Adv. Appl. Probab.* 35, 721-736.
[61] Newcomb, S. (1886). A Generalized Theory of the Combination of Observations so as to Obtain the Best Result, *Amer. J. Math.* 8, 343-366.
[62] Newey, W.K. and K.D. West (1987). A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.
[63] Nze, P.A., and P. Doukhan (2004). Weak Dependence: Models and Applications to Econometrics, *Econometric Theory* 20, 995-1045.
[64] Peligrad, M. and S. Utev (2006). Central Limit Theory for Stationary Linear Processes, *Ann. Probab.* 34, 1608-1622.
[65] Pham, T.D. and L.T. Tran (1985). Some Mixing Properties of Time Series Models, *Stoch. Proc. Appl.* 19, 297-303.
[66] Pruitt, W.E. (1983). The Class of Limit Laws for Stochastically Compact Normed Sums, *Ann. Probab.* 11, 962-969.
[67] Pruitt W.E. (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.
[68] Qiu, J. and Z. Lin (2010). The Functional Central Limit Theorem for Linear Processes with Strong Near-Epoch Dependent Innovations, *J. Math. Anal. Appl.* 376, 373-382.
[69] Rootzén, H. (2009). Weak Convergence of the Tail Empirical Function for Dependent Sequences, *Stoch. Proc. Appl.* 119, 468-490.
[70] Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition, *Proc. Nat. Acad. Sci. USA*. 42, 43-47.
[71] Sampson, P.D. and P. Guttorp (1992). Nonparametric Representation of Nonstationary Spatial Covariance Structure, *J. Amer. Stat. Assoc.* 87, 108-119.
[72] Schennach, S.M. (2008). Quantile Regression with Mismeasured Covariates, *Econometric Theory* 24, 1010-1043.
[73] 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 (eds.), *Sums, Trimmed Sums and Extremes*. Birkhäuser: Berlin.
[74] Whalen, E. (1992). The Asymptotic Distribution of Magnitude Trimmed Sums for Distributions in the Feller Class, *J. Theor. Probab.* 5, 447-464.
[75] Wooldridge, J. and H. White (1988). Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes, *Econometric Theory* 4, 210-230.
[76] Wu, W.B. and W. Min (2005). On Linear Processes with Dependent Innovations, *Stoch. Proc. Appl.* 115, 939-958.
[77] Wu, W.B. and M. Woodroofe (2004). Martingale Approximations for Sums of Stationary Processes, *Ann. Probab.* 32, 1674-1690.