Econometric Theory, **27**, 2011, 844–884.
doi:10.1017/S0266466610000514

TAIL AND NONTAIL MEMORY WITH APPLICATIONS TO EXTREME VALUE AND ROBUST STATISTICS

JONATHAN B. HILL
University of North Carolina at Chapel Hill

New notions of tail and nontail dependence are used to characterize separately extremal and nonextremal information, including tail log-exceedances and events, and tail-trimmed levels. We prove that near epoch dependence (McLeish, 1975; Gallant and White, 1988) and L_0-approximability (Pötscher and Prucha, 1991) are equivalent for tail events and tail-trimmed levels, ensuring a Gaussian central limit theory for important extreme value and robust statistics under general conditions. We apply the theory to characterize the extremal and nonextremal memory properties of possibly very heavy-tailed GARCH processes and distributed lags. This in turn is used to verify Gaussian limits for tail index, tail dependence, and tail-trimmed sums of these data, allowing for Gaussian asymptotics for a new tail-trimmed least squares estimator for heavy-tailed processes.

1. INTRODUCTION

We analyze notions of dependence separately restricted to extremal and nonextremal information. If population dependence is measured over the real line $(-\infty,\infty)$, extremal dependence is measured on $(-\infty,-b_1)\cup(b_2,\infty)$ as each $b_i \to \infty$, and nonextremal dependence on $(-b_1,b_2)$ as $b_i \to \infty$. The results permit fundamentally new Gaussian limit theory for tail shape and tail dependence estimators, and tail-trimmed sums for a large array of heavy-tailed time series, including linear and nonlinear generalized autoregressive conditional heteroskedasticity (GARCH) with unit or explosive roots, and nonstationary distributed lags with hyperbolic or geometric memory and heavy-tailed shocks. Gaussian asymptotics for tail-trimmed sums support new robust estimators, including asymptotically normal tail-trimmed versions of least squares and quasi-maximum likelihood (QML) for heavy-tailed data.

Tail shape and tail dependence are natural objects of study in extreme value theory, with applications to cost, catastrophe, damage and risk modeling in finance, meteorology, insurance, and macroeconomics (Leadbetter, Lindgren, and Rootzén 1983; Beirlant, Vynckier, and Teugels 1996; Embrechts, Klüpperberg, and Mikosch, 1997). Tail trimming, however, is used for robust inference in the

The authors kindly thanks two anonymous referees and co-editor Yuichi Kitamura for helpful suggestions. Address correspondence to Jonathan B. Hill, Dept. of Economics, University of North Carolina, Chapel Hill, NC 27599; e-mail: jbhill@email.unc.edu.

presence of heavy tails (Stigler, 1973; Csörgő, Horváhth, and Mason, 1986; Hahn, Kuelbs, and Samur, 1987; Hill, 2009; Hill and Renault, 2010). Although the motivations and uses of these statistical methods are quite disparate, the underlying theory is remarkably similar due simply to the mathematics behind choosing the threshold b and to the depiction of memory.

1.1. GARCH Dependence

Although our dependence measures are quite general, one motivating application involves GARCH processes. Denote lag polynomials $\alpha(L) = \sum_{i=1}^{p} \alpha_i L^i$ and $\beta(L) = 1 - \sum_{i=1}^{q} \beta_i L^i$ with L the lag operator, and let

$$X_t = \sigma_t \epsilon_t, \quad \epsilon_t \text{ is i.i.d. } \mathrm{E}|\epsilon_t|^r < \infty \quad \text{for some } r > 0; \tag{1}$$
$$\beta(L)\sigma_t^2 = \omega + \alpha(L) X_t^2, \quad \omega > 0, \quad \text{at least one } \alpha_i, \beta_i > 0;$$
$$\text{the roots of } \beta(z) \text{ lie outside unit circle;}$$

with Lyapunov exponent[1] $\gamma < 0$; and the density of ϵ_t is positive on $\mathbb{R}$-a.e. The processes $\{X_t, \sigma_t\}$ have regularly varying marginal distribution tails of the form

$$P(|X_t| > x) = cx^{-\kappa}(1 + o(1)), \qquad c > 0, \qquad \kappa > 0. \tag{2}$$

See Basrak et al. (2002b, Thm. 3.1); cf. Davis and Mikosch (1998) and Mikosch and Stărică (2000). Recall the tail index κ is identically the moment supremum: $\mathrm{E}|X_t|^p < \infty \; \forall p < \kappa$ and $\mathrm{E}|X_t|^{\kappa+\delta} = \infty \; \forall \delta \geq 0$ (Ibragimov and Linnik, 1971). Further, roughly speaking $1/\kappa$ measures the mean exceedance of $\ln(|X_t|)$ above a large threshold: Smaller κ are associated with larger average threshold exceedances. See, e.g., Hsing (1991, eq. (1.5)).

Power law tails (2) naturally arise in random volatility processes (de Haan, Rosnick, Rootzén, and de Vries, 1989; Davis and Mikosch, 1998) and first-price auction bids (Hill and Schneyerov, 2010); they coincide with a maximum domain of attraction and the domain of attraction of a stable law when $\kappa < 2$, and they accurately characterize the tail behavior of many time series, including financial asset returns, insurance claims, telecommunication network data, urban growth, and meteorological events. See the compendia Leadbetter et al. (1983), Resnick (1987), and Embrechts et al. (1997), and see Gabaix (2008) and Ibragomov (2009), and their citations for encyclopedic treatments.

In the GARCH case κ also arises in a Lyapunov-type moment condition. A GARCH(1,1), for example, satisfies (Mikosch and Stărică, 2000)

$$\mathrm{E}\left[\left(\alpha_1 \epsilon_t^2 + \beta_1\right)^{\kappa/2}\right] = 1. \tag{3}$$

Suppose $\mathrm{E}[\epsilon_t^2] = 1$. Then X_t has a finite variance $\kappa > 2$ if $\alpha_1 + \beta_1 < 1$, a hairline infinite variance $\kappa = 2$ in the integrated GARCH (IGARCH) case $\alpha_1 + \beta_1 = 1$, and an infinite variance in the explosive root case $\alpha_1 + \beta_1 > 1$. Thus, very roughly speaking, knowledge of the moments of ϵ_t allows the use of κ to gauge GARCH memory.

In general the analyst may want to estimate κ as a check on required moment conditions for a minimum distance estimator (e.g., Hill and Renault, 2010) or as a measure of market risk (Embrechts et al., 1997; Drees, Ferreira, and de Haan, 2004; Iglesias and Linton, 2009; Hill, 2010) or tail dependence as a measure of risk decay and risk spillover (Stărică, 1999; Ledford and Tawn, 1997, 2003; Hill, 2008b, 2009b); or estimate the lag polynomial coefficients $\{\alpha, \beta\}$ to forecast volatility in the presence of extremes (Hall and Yao, 2003; Davis and Mikosch, 2009a; Hill and Renault, 2010; Linton, Pan, and Wang, 2010).

The population memory properties of GARCH are now well known. If $\{X_t\}$ is governed by (1) then it is geometrically α-mixing (Boussama, 1998; cf. Basrak et al., 2002b), and a variety of nonlinear GARCH processes like asymmetric, multiplicative, and smooth transition GARCH are geometrically ergodic, hence β-mixing (Carrasco and Chen, 2002; Meitz and Saikkonen, 2008; cf. Doukhan, 1994). The root condition in (1) ensures an ARCH(∞) representation for σ_t^2:

$$\sigma_t^2 = \pi_0 + \sum_{i=1}^{\infty} \pi_i X_{t-i}^2, \qquad \pi_0 > 0, \qquad \pi_i \geq 0, \quad \text{at least one } \pi_i > 0,$$
$$S := \sum_{i=1}^{\infty} \pi_i. \tag{4}$$

Davidson (2004) uses (4) to analyze population memory when $\epsilon_t \overset{iid}{\sim} (0,1)$ (i.e., ϵ_t is i.i.d. with mean zero and unit variance). In the covariance stationary case $S < 1$ in general X_t is L_1- or L_2-near epoch dependent (NED), and L_0-approximable (APP) when there is a unit ($S = 1$) or explosive ($S > 1$) root. See Section 2 for dependence definitions.

Although empirical studies of extremal dependence in random volatility processes abound (e.g., Stărică, 1999; Longin and Solnik, 2001), very few results formally characterize tail memory in GARCH data. One approach exploits the fact that GARCH class (1) belongs to the maximum domain of attraction: for each $z \geq 0$ and suitable normalizing sequence $\{u_n\}$

$$\lim_{n\to\infty} P\left(\frac{1}{u_n} \max_{1\leq t\leq n} |X_t| \leq z\right) = e^{-\theta z^{-\kappa}}, \qquad z \geq 0, \qquad \theta \in [0,1].$$

See Basrak et al. (2002b); cf. Chernick (1981); de Haan et al. (1989); Mikosch and Stărică (2000); and Davis and Mikosch (2009a). The inverted extremal index $1/\theta$ roughly measures the number of high threshold exceedances, hence tail memory, while $1/\kappa$ reveals the mean distance above a high threshold, hence tail thickness (Leadbetter, 1974, 1983; Leadbetter et al., 1983). Intuitively $\theta = 1$ dictates extremal independence, $\theta \in (0,1)$ short-range dependence, and $\theta = 0$ long-range dependence (Leadbetter, 1983).

The extremal index has been characterized for autoregressions, moving averages, GARCH, and stochastic volatility (SV) (Rootzén, 1978; Chernick, Hsing, and McCormick, 1991; Mikosch and Stărică, 2000; Davis and Mikosch, 2009a, 2009b). See Section 6.1, below, for details. Nevertheless, θ does not portray memory decay per se, so knowledge of θ cannot in general ensure a central limit property for tail arrays of X_t and therefore for estimators of tail exponents like κ and θ. Indeed, inference on estimators of κ and θ invariably requires more information like population mixing or NED extremes (e.g., Chernick et al; Hsing, 1991, 1993; Mikosch and Stărică; Hill, 2010).

Another approach to modeling tail dependence exploits bivariate power-law tail decay with index η (Ledford and Tawn, 1997, 2003; Stărică, 1999; cf. Resnick, 1987; Basrak, Davis, and Mitrosch, 2002a). Ledford and Tawn (2003) estimate η in a time series framework that implicitly requires a population mixing condition. Stărică estimates the tail empirical measure for constant conditional correlation GARCH (CCC-GARCH), which embeds bivariate tail dependence information from η. The CCC-GARCH class exhibits mixing at a geometric rate, yet tail memory decay measured by η and how that relates to a central limit property for dependent data are not available.[2] See Section 6.1 for a definition of η and expanded discussion.

An arguably more robust approach is based on tail indicators $I(|X_t| > c)$ and their autocorrelation. Compactly $r(h,c) := [P(X_{t-h} > c,\ X_t > c) - P_x(c)^2]/P(X_{t-h} > c)$ is referred to as the tail event correlation in Hill (2008b, 2009b) and its limit $r(h) = \lim_{c\to\infty} r(h,c)$ the extremogram in Davis and Mikosch (2009c). The coefficient has been studied for autoregressive moving average (ARMA), GARCH, and SV, while an estimator of $r(h)$, like κ, θ, and η, requires more information on memory. Davis and Mikosch (2009c) impose population α-mixing and multivariate regular variation for the joint tail of $\{X_t, X_{t-1}, \ldots, X_{t-h}\}$, and Hill (2008b, 2009b) imposes NED on tail events without joint tail restrictions. See Section 5.2 for estimation details and Section 6 for further discussion.

Finally, apparently there are no explicit treatments on the nontail memory of random volatility processes. Thus, how such properties relate to robust estimation of such models is entirely unexplored.

1.2. Heavy-Tailed Nonlinear Distributed Lags

A second motivating example is a heavy-tailed distributed lag,

$$X_t = \sum_{i=0}^{\infty} \psi_{t,i}\epsilon_{t-i},\ \epsilon_t \sim (2) \quad \text{with } \kappa < 2 \ \text{ (i.e., } \mathrm{E}[\epsilon_t^2] = \infty), \tag{5}$$

where $\psi_{t,0} = 1$, and $\{\psi_{t,i}\}$ is for each $i > 0$ a measurable stochastic process. Class (5) covers bilinear, autoregressive fractionally integrated moving average (ARFIMA), random coefficient autoregressions, and threshold and nonlinear autoregressions, and under fairly general conditions X_t is L_0-APP or L_p-NED (see Section 4).

Studies of the tail probabilities of stationary linear processes $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$, $\epsilon_t \overset{iid}{\sim}$ (2) have a rich history (e.g., Feller, 1946, 1971; Cline, 1983), with several weak extremal dependence properties like D-mixing and the extremal index θ (Leadbetter, 1974, 1983; Rootzén, 1978; Leadbetter et al., 1983; Chernick et al., 1991; Smith, 1992; Smith and Weissman, 1994; Hsing, 1993), and the extremogram $r(h)$ (Hill, 2008b; Davis and Mikosch, 2009c). D-mixing does not necessarily carry over to functions of D-mixing random variables, and in general is known to hold for a very limited class of processes. Refer to Section 6 for details.

Only recently has a central limit theory for trimmed or truncated sums of stationary linear processes been developed (e.g., Wu, 2005; Hill, 2009a; Hill and Renault, 2010). The vast majority of this literature concerns independent data (e.g., Stigler, 1973; Csörgő et al., 1986; Hahn and Weiner, 1992) and in a rare case mixing data with finite variance (Hahn et al., 1987).

1.3. Tail and Nontail Memory

Hill (2009b, 2010), by comparison, imposes α-mixing on measurable functional arrays of some process $\{X_t\}$, covering tail events, exceedances, and trimmed levels defined below. If such a function has a mixing property, $\{X_t\}$ is said to be F-mixing, and any α-mixing $\{X_t\}$ is F-mixing covering linear and nonlinear ARMA-GARCH processes with sufficiently smooth probability densities (An and Huang, 1996; Giraitis, Kokoszka, and Leipus, 2000; Carrasco and Chen, 2002; Meitz and Saikkonen, 2008).

The shortcomings of F-mixing, however, are the same as population and D-mixing conditions: It is difficult to verify, density smoothness is required, and for extreme value and robust limit theory broader dependence properties are in demand (Iglesias and Linton, 2009; Rootzén, 2008; Hill, 2009a, 2009b, 2010). Further, little is known about the mixing characteristics of long memory processes like ARFIMA and fractionally integrated GARCH (FIGARCH) (e.g., Baillie, Bollerslev, and Mikkelsen, 1996; Guegan and Ladoucette, 2001) and hyperbolic GARCH (Davidson, 2004).

Let $\{X_t\}_{t=1}^n$ be a sample of size $n \geq 1$ and $\{k_n\}$ an intermediate order sequence: $1 \leq k_n < n$, $k_n \to \infty$, and $k_n/n \to 0$ as $n \to \infty$. Construct a sequence of asymptotic k_n/n^{th}-quantiles b_n of $|X_t|$ (Leadbetter et al., 1983; Galambos, 1987):

$$\frac{n}{k_n} P\left(|X_t| > b_n\right) \to 1. \tag{6}$$

Thus k_n is approximately the number of sample exceedances above a high threshold b_n. We implicitly assume b_n does not depend on t, although we allow X_t to be otherwise nonstationary. Now define the b_n-event

$$\bar{I}_{n,t} := I\left(|X_t| > b_n\right)$$

and b_n-exceedance, or peak-over-threshold (e.g., Smith, 1984),

$$\mathrm{E}_{n,t} := (\ln|X_t| - \ln b_n)_+, \quad \text{where } (z)_+ := \max\{0, z\}. \tag{7}$$

In the sequel we generalize to one-tailed cases. Tail arrays are exploited in a variety of contexts, including tail index and tail dependence estimation (Leadbetter et al., 1983; Davison and Smith, 1990; Hsing, 1991, 1993; Hill, 2008b, 2009b, 2010), tail trimming and robust estimation (Hahn et al., 1987; Hill, 2009a; Hill and Renault, 2010), and regression estimator performance and breakdown point analysis based on tail behavior (Jurečková, 1981; He, Jurečková, Koenker, and Portnoy, 1990).

In this paper we extend the notions of L_0-APP and L_p-NED to tail and nontail information. Our first task is to prove if X_t is L_0-APP, as in explosive GARCH, then the triangular arrays $\{\bar{I}_{n,t}, \mathrm{E}_{n,t}\}$ are also L_0-APP, and $\{\bar{I}_{n,t}, \mathrm{E}_{n,t}\}$ are L_0-APP *if and only if* they are L_2-NED. See Section 2.

Second, we prove that if $\{X_t\}$ is L_0-APP then the trimmed level

$$\hat{X}_{n,t} := X_t \times I\left(|X_t| \leq b_n\right) \tag{8}$$

is always L_2-NED, even as $n \to \infty$, no matter how heavy-tailed X_t is (Section 3). The array $\{\hat{X}_{n,t}\}$ not only serves to approximate location for heavy-tailed data, it forms the asymptotic foundation for a class of robust estimators, including self-normalized tail-trimmed sums (Pruitt, 1985; Hahn, Kuelbs, and Weiner, 1990; Hahn and Weiner, 1992; Hill, 2009a) and generalized method of tail-trimmed moments (Hill and Renault, 2010).

The primary contribution of this paper is a set of new dependence notions that permit a broad Gaussian central limit theory for tail and tail-trimmed arrays of linear and nonlinear distributed lags and random volatility processes. We never need to restrict density smoothness to ensure NED or APP for extremes or nonextremes.

Moreover, other than Hill's (2009b, 2010) F-mixing for tail or nontail arrays, we believe this to be the first study that directly explores memory in "nonextremes" as defined above, and to study rigorously dependence notions away from the tails for the sake of improved asymptotic theory for robust statistics. This is the second major contribution.

As a third contribution, in Section 4 the main dependence results are applied to GARCH and distributed lags (1), (4), and (5), and generalized to larger classes of linear and nonlinear processes including nonlinear GARCH and nonlinear autoregressions.

Since $\{\bar{I}_{n,t}, \mathrm{E}_{n,t}\}$ are directly linked to extremal statistics and $\{\bar{I}_{n,t}, \hat{X}_{n,t}\}$ to robust estimation, in Section 5 we characterize Gaussian limit theory for tail index and tail dependence estimators, and a tail-trimmed sum of L_0-APP data. Finally, we show how fusing theory for dependent tail arrays $\{\bar{I}_{n,t}\}$ and tail-trimmed sums $\{\hat{X}_{n,t}\}$ supports new robust least squares and maximum likelihood estimators. This is the fourth contribution.

We focus on L_0-APP and L_p-NED due to their relative ease of verification for a massive array of time series. Similar weak dependence measures can undoubtedly be divided into tail and nontail constructions (e.g., Doukhan and Louhichi, 1999; Nze, Buhlmann, and Doukhan, 2002; Wu and Min, 2005).

Throughout $K > 0$ denotes an arbitrary finite constant whose value may change from line to line; similarly, $\iota > 0$ is infinitessimal and may change with the context. Unless otherwise stated, $\rho \in (0,1)$ is always arbitrary and may change. Compactly write the L_p-norm $||x||_p := (\sum_{i,j} \mathrm{E}|x_{i,j}|^p)^{1/p}$.

2. TAIL MEMORY: EVENTS AND EXCEEDANCES

For the remainder of the paper assume $X_t \geq 0$ almost surely (a.s.), capturing two-tailed $|X_t|$ or one-tailed cases $-X_t I(X_t < 0)$ or $X_t I(X_t > 0)$.

We assume X_t is L_p-bounded for some $p > 0$: $\mathrm{E}|X_t|^p < \infty$. It is easy to show by Markov's inequality the survival probability $\bar{F}_t(x) := P(X_t > x)$ is boundedly regularly varying: There exist $\kappa \in (0, p)$ and slowly varying $L_t : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{x\to\infty} \sup_{t\in\mathbb{Z}} \left\{ \frac{\bar{F}_t(x)}{x^{-\kappa} L_t(x)} \right\} \leq K < \infty. \tag{9}$$

Class (9) includes thin (exponential) and thick (regularly varying) tails, and implies

$$x^p \bar{F}_t(x) \to 0 \; \forall p < \kappa \quad \text{for each } t \in \mathbb{Z}.$$

Although uniform L_p-boundedness $\sup_{t\in\mathbb{Z}} \mathrm{E}|X_t|^p < \infty$ is not required, we assume κ is the largest index that allows (9) and does not depend on t. If $\{x_t\}$ is stationary and (9) holds exactly $\bar{F}_t(x) \sim x^{-\kappa} L(x)$ as $x \to \infty$ as in (2) then κ is the index of regular variation and therefore the moment supremum: $\kappa = \sup\{\alpha > 0 : \mathrm{E}|X_t|^\alpha < \infty\}$. See Bingham, Goldie, and Teugels (1987) and Resnick (1987).

A convenient way to characterize tail memory is to restrict NED to the tails as in Hill (2009b, 2010). The NED property dates in various forms to Ibragimov (1962) and McLeish (1975), with myriad subsequent extensions in Bierens (1987), Gallant and White (1988), and recently Nze et al. (2002) and Wu and Min (2005). See Nze and Doukhan (2004) for compendium details on NED and similar concepts, and their usage in econometrics.

Let $\{\mathcal{F}_t\}$ be a sequence of σ-fields induced by some possibly vector-valued process $\{\epsilon_t\}$,

$$\mathcal{F}_t = \sigma(\epsilon_\tau : \tau \leq t) \quad \text{and} \quad \mathcal{F}_s^t = \sigma(\epsilon_\tau : s \leq \tau \leq t).$$

If X_t is generated by (1), (4), or (5), for example, then ϵ_t is the scalar innovation.

DEFINITION 1 (L_p-near epoch dependence). *$\{X_t\}$ is L_p-NED on $\{\mathcal{F}_t\}$ or on $\{\epsilon_t\}$ with size $\lambda > 0$ if there exist deterministic sequences $\{d_t, \vartheta_l\}$, $d_t \geq 0$, $\vartheta_l \in [0, 1)$, and $\vartheta_l = o(l^{-\lambda})$ such that*

$$\left\| X_t - \mathrm{E}\left[X_t | \mathcal{F}_{t-l}^{t+l}\right] \right\|_p \leq d_t \times \vartheta_l. \tag{10}$$

Remark 1. In general we drop the addendum "on $\{\mathcal{F}_t\}$" unless the base is unclear.

Remark 2. The "constants" d_t absorb time-dependence of the norm and control for scale, where $d_t \to \infty$ as $t \to \infty$ is possible due to trend. The "coefficients" ϑ_l gauge hyperbolic memory decay according to "size" λ, and geometric memory $\vartheta_l = o(\rho^l)$ means size λ is arbitrarily large. The property characterizes linear and nonlinear distributed lags with hyperbolic or geometric memory, thin- or thick-tailed shocks, bilinear data, covariance stationary GARCH (Davidson, 1994, 2004), and SV (Hill, 2008a).

Remark 3. If $\mathcal{F}_t$ is adapted to X_t then $\{X_t\}$ is trivially L_p-NED with constants $d_t = 0$ and arbitrary size $\lambda > 0$. For example, since ϵ_t can be anything it can be mixing, and $X_t = \epsilon_t$ is always possible with $d_t = 0$ and $\vartheta_l = o(l^{-\lambda})$ for any $\lambda > 0$. Thus, a mixing process is NED on itself covering many linear and nonlinear ARMA, GARCH, and ARMA-GARCH processes. See Section 4.3 for examples.

A tail version of NED requires the b_n-event functional

$$I_{n,t}(u) := I\left(X_t \leq b_n e^u\right) \quad \text{and} \quad \bar{I}_{n,t}(u) := I\left(X_t > b_n e^u\right)$$

where

$$I_{n,t} := I_{n,t}(0) \quad \text{and} \quad \bar{I}_{n,t} := \bar{I}_{n,t}(0).$$

Throughout $\{l_n\}$ denotes a sequence of positive integers with lower bound $\liminf_{n \geq 1} l_n = l \in \mathbb{N}$. Notice $l_n \to \infty$ is allowed.

DEFINITION 2 (L_p-extremal near epoch dependence). *$\{X_t\}$ is L_p-extremal-NED on $\{\mathcal{F}_t\}$ or on $\{\epsilon_t\}$ with size $\lambda > 0$ if for some $\{l_n\}$*

$$\left\| \bar{I}_{n,t}(u) - \mathrm{E}\left[\bar{I}_{n,t}(u) | \mathcal{F}_{t-l_n}^{t+l_n}\right] \right\|_p \leq d_{n,t}(u) \times \psi_{l_n}, \quad \forall n \geq 1, \tag{11}$$

where $d_{n,t} : \mathbb{R}_+ \to \mathbb{R}_+$ is Lebesgue measurable, $\sup_{u \geq 0} \sup_{1 \leq t \leq n} d_{n,t}(u) = O((k_n/n)^{1/p})$, *$\psi_{l_n} \in [0, 1)$, and $l_n^{\lambda} \psi_{l_n} \to 0$.*

Remark 4. Extremal-NED (E-NED) is simply NED applied to $\bar{I}_{n,t}(u)$, where (11) implies

$$\frac{n}{k_n} \times l_n^{p\lambda} \times \sup_{1 \leq t \leq n} \mathrm{E}\left| \bar{I}_{n,t}(u) - \mathrm{E}\left[\bar{I}_{n,t}(u) | \mathcal{F}_{t-l_n}^{t+l_n}\right] \right|^p \to 0 \quad \text{as } n \to \infty.$$

The outer scale n/k_n controls for inherent degeneracy since $||\bar{I}_{n,t}(u)||_p^p = O(k_n/n) = o(1)$, and the inner scale $e^u \geq 1$ smoothly expands the tail threshold. The constants $d_{n,t}(u)$ are uniformly bounded in t since $|\bar{I}_{n,t}(u) - \mathrm{E}[\bar{I}_{n,t}(u)| \mathcal{F}_{t-l_n}^{t+l_n}]| \in [0, 1]$.

Remark 5. A useful distinction between NED (10) and E-NED (11) is the sample-size dependent displacement sequence l_n in (11). This permits far greater flexibility in analyzing dependence in extremes and nonextremes, and therefore deriving associated limit theory. In particular, since $l_n \to \infty$ arbitrarily fast is allowed it can be tailored to ensure E-NED and suit asymptotic arguments for nonstationary data.

Remark 6. Similar to the NED case, any mixing process is L_2-E-NED on itself with trivial constants and arbitrary size.

The case where the base σ-field $\mathcal{F}_t$ is adapted to X_t leads to trivial results. Thus all subsequent claims implicitly assume $\mathcal{F}_t$ is not adapted to X_t, unless otherwise stated.

Population L_p-NED implies L_q-E-NED for any $q \geq 2$, and L_q-E-NED is equivalent to L_r-E-NED for any $r \gtreqless q$. The latter applies because $\bar{I}_{n,t}(u)$ is bounded.

THEOREM 2.1.

(i) Let $\{X_t\}$ be L_p-NED, $p > 0$, with constants d_t and coefficients ϑ_l of size $\lambda > 0$. Then $\{X_t\}$ is L_q-E-NED for any $q \geq 2$ with constants $d_{n,t}(u) = K(k_n/n)^{1/q}e^{-up/q}$ and coefficients ψ_{l_n} of size $\theta = \lambda \times \min\{p, 1\}/(2q)$.

(ii) Let $\{X_t\}$ be L_q-E-NED, $q > 0$, with constants $d_{n,t}(u)$ and coefficients ψ_{l_n} of size $\theta > 0$. Then $\{X_t\}$ is L_r-E-NED for any $r \gtreqless q$ with constants $d_{n,t}^{q/\max\{q,r\}}(u)$ and coefficients $\psi_{l_n}^{q/\max\{q,r\}}$ of size $\theta \times (q/\max\{q, r\})$.

Remark 7. Size is irrelevant in the geometric memory case, so geometric L_p-NED implies geometric L_q-E-NED for all $q > 0$.

Remark 8. Irrespective of the degree of nonstationarity characterized by the NED constants d_t, since $l_n \to \infty$ is otherwise arbitrary the E-NED constants $d_{n,t}(u) = K(k_n/n)^{1/q}e^{-up/q}$ are uniformly bounded in $1 \leq t \leq n$, $n \geq 1$ and $u \geq 0$, and Lebesgue integrable on $\mathbb{R}_+$. Consult the proof.

Remark 9. Hill (2010, Lem. B.1) proves that if $\{X_t\}$ is L_2-E-NED with coefficients ψ_{l_n} and Lebesgue integrable constants $d_{n,t}(u)$ then the exceedance process $\{\mathrm{E}_{n,t}\} = \{(\ln(X_t/b_n)_+)\}$ is L_2-NED with the same coefficients ψ_{l_n} and constants $K\int_0^\infty d_{n,t}(u)du$. Use Theorem 2.1 to deduce that population L_p-NED implies L_2-NED exceedances $\{\mathrm{E}_{n,t}\}$.

We cannot in general apply Theorem 2.1 to GARCH processes with unit or explosive roots without further information about the error distribution. Davidson

(2004), for example, only shows that covariance stationary GARCH are L_1-NED. An alternative route is to prove that $\{X_t\}$ is mixing and use the fact that a mixing process is trivially NED on itself. This, however, invariably requires a smooth error distribution (e.g., Boussama, 1998; Basrak et al., 2002b; Carrasco and Chen, 2002).

In order to characterize the extremes of these processes without additional assumptions on the errors, we exploit probability-based L_0-approximability (L_0-APP) (Pötscher and Prucha, 1991): There exists an $\mathcal{F}_{t-l}^{t+l}$-measurable function $g_t^{(l)}$ and sequences of deterministic positive numbers $\{f_t, \upsilon_l\}$, $\inf_{t\in\mathbb{Z}} f_t \geq f > 0$, and $\upsilon_l \in [0,1)$, such that

$$P\left(\left|X_t - g_t^{(l)}\right| > f_t \times \delta\right) \leq \upsilon_l = o\left(l^{-\lambda}\right) \quad \text{for some} \lambda > 0 \text{ and any } \delta > 0. \tag{12}$$

L_0-APP is useful for heavy-tailed processes that may not satisfy certain moment conditions, as opposed to moment-based near epoch dependence, mixingale (McLeish, 1975), and so-called θ-weak dependence (Nze et al., 2002) and L_p-weak dependence (Wu and Min, 2005). Further, L_p-NED implies L_0-APP (Davidson, 1994, p. 274), and as with NED and E-NED if $\mathcal{F}_t$ is adapted to X_t then $\{X_t\}$ is trivially L_0-APP on itself.

Now repeat the logic of E-NED by replacing X_t with $\bar{I}_{n,t}(u) = I(X_t > b_n e^u)$.

DEFINITION 3 (L_0-extremal-approximability). *$\{X_t\}$ is L_0-extremal-approximable on $\{\mathcal{F}_t\}$ with size $\lambda > 0$ if there exists an $\mathcal{F}_{t-l_n}^{t+l_n}$-measurable stochastic functional $h_t^{(l_n)}(u)$ and coefficients $\varphi_{l_n} \in [0,1)$, $\varphi_{l_n} = o(l_n^{-\lambda})$, such that*

$$P\left(\left|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right| > f_{n,t} \times \delta_n\right) \leq K \times e_{n,t}(u) \times \varphi_{l_n}$$

for some $\{l_n\}$, $\liminf_{n\geq 1} l_n \in \mathbb{N}$, some deterministic triangular array $\{f_{n,t}\}$, $\liminf_{n\geq 1} \inf_{1\leq t\leq n} f_{n,t} = f > 0$, and any deterministic sequence $\{\delta_n\}$, $\inf_{n\geq 1}\delta_n = \delta > 0$. The constants $e_{n,t} : \mathbb{R}_+ \to [0,1]$ are Lebesgue measurable and $\sup_{u\geq 0} \sup_{1\leq t\leq n} e_{n,t}(u) = O(k_n/n)$. The array $\{h_t^{(l_n)}(u)\}$ is the L_0-E-APP "approximator" of X_t.

L_p-E-NED implies L_0-E-APP by Markov's inequality, just like NED implies APP, hence a proof is omitted.

LEMMA 2.2. *If $\{X_t\}$ is L_p-E-NED, $p > 0$, with size λ then it is L_0-E-APP with size $p\lambda$ and approximator $\{P(X_t > b_n e^u | \mathcal{F}_{t-l_n}^{t+l_n})\}$.*

Similarly, population L_0-APP implies L_0-E-APP, just like L_p-NED implies L_q-E-NED.

LEMMA 2.3. *Let $\{X_t\}$ be L_0-APP with size λ, constants f_t, and approximator $\{g_t^{(l)}\}$. Then $\{X_t\}$ is L_0-E-APP with coefficients φ_{l_n} of size λ, constants $e_{n,t}(u) = (k_n/n)e^{-\iota u} \in [0,1]$ for tiny $\iota > 0$, and approximator $\{I(g_t^{(l_n)} > b_n e^u)\}$.*

Remark 10. Under L_0-APP the resulting L_0-E-APP coefficients $e_{n,t}(u) = K(k_n/n)e^{-\iota u}$ are inherently Lebesgue integrable on $[0, \infty)$. Since the magnitude of $\iota > 0$ has no impact on integrability we simply write in the sequel

$$e_{n,t}(u) = K(k_n/n)e^{-u}.$$

L_0-E-APP always implies L_2-E-NED, contrary to the population case (e.g., Pötscher and Prucha, 1991; Davidson, 1994), a key link between approximability and Gaussian limit theory for tail arrays.

LEMMA 2.4. *Let $\{X_t\}$ be L_0-E-APP with size λ and constants $e_{n,t}(u) = (k_n/n)e^{-u}$. If the approximator is either $\{I(g_t^{(l_n)} > b_n e^u)\}$ for some $\digamma_{t-l_n}^{t+l_n}$-measurable array $\{g_t^{(l_n)}\}$, or $\{P(X_t > b_n e^u|\digamma_{t-l_n}^{t+l_n})\}$, then $\{X_t\}$ is L_2-E-NED with size $\lambda/2$ and constants $d_{n,t}(u) = K(k_n/n)^{1/2}e^{-u/2}$.*

The following results are easy consequences of Lemmas 2.2–2.4: Population L_0-APP is sufficient for the tail L_2-E-NED property, and NED and L_0-APP are equivalent in the tails.

THEOREM 2.5. *If $\{X_t\}$ is L_0-APP with size λ and constants f_t then it is L_2-E-NED with size $\lambda/2$ and constants $d_{n,t}(u) = K(k_n/n)^{1/2}e^{-u/2}$.*

THEOREM 2.6. *$\{X_t\}$ is L_2-E-NED with size $\lambda/2$ and constants $d_{n,t}(u) = K(k_n/n)^{1/2}e^{-u/2}$if and only if it is$L_0$-E-APP with size λ and constants $e_{n,t}(u) = (k_n/n)e^u$.*

Since L_2-E-NED with Lebesgue integrable constants $d_{n,t}(u)$ implies that the exceedance $\{\mathrm{E}_{n,t}\} = \{(\ln|X_t|/b_n)_+\}$ is L_2-NED (see Remark 9 above and see Hill, 2010), the next claim follows instantly from Theorem 2.6.

COROLLARY 2.7. *If $\{X_t\}$ is L_0-APP with size λ then $\{\bar{I}_{n,t}(u), \mathrm{E}_{n,t}\}$ are L_2-NED with size $\lambda/2$.*

These results are useful since heavy-tailed data may only be L_0-APP (e.g., GARCH with explosive roots), yet have L_2-NED extremal information $\{\bar{I}_{n,t}(u)\}$ and $\{\mathrm{E}_{n,t}\}$. The latter properties permit Gaussian limit theory for tail shape, quantile, and dependence estimators under substantially general conditions. See Section 5 for applications.

3. NONTAIL MEMORY

Since trivially $||I_{n,t}(u) - \mathrm{E}[I_{n,t}(u)|\digamma_{t-l_n}^{t+l_n}]||_p = ||\bar{I}_{n,t}(u) - \mathrm{E}[\bar{I}_{n,t}(u)|\digamma_{t-l_n}^{t+l_n}]||_p$, Theorem 2.1 instantly applies to $I_{n,t}(u)$: Near epoch dependence carries over to nontail events.

COROLLARY 3.1. *If $\{X_t\}$ is L_p-NED, $p > 0$, with constants d_t and size λ then $\{I_{n,t}(u)\}$ is L_q-NED for any $q \geq 2$ with constants $d_{n,t}(u)$ and size*

$\theta = \lambda \times \min\{p, 1\}/(2q)$. *Further, if* $\{I_{n,t}(u)\}$ *is* L_q*-NED,* $q > 0$*, with constants* $d_{n,t}(u)$ *and size* $\theta > 0$ *then it is* L_r*-E-NED for any* $r \gtreqless q$ *with constants* $d_{n,t}^{q/\max\{q,r\}}(u)$ *and size* $\theta \times (q/\max\{q, r\})$.

The next result shows that population L_0-APP carries over to power transforms of tail-trimmed levels $X_t I_{n,t}(u) = X_t I(X_t \leq b_n e^u)$ and extreme levels $X_t \bar{I}_{n,t}(u) = X_t I(X_t > b_n e^u)$.

LEMMA 3.2. *Let* $\{X_t\}$ *be* L_0*-APP with coefficients* υ_l *of size* λ *and approximator* $\{g_t^{(l)}\}$*, and choose any* $s > 0$*. Then* $\{X_t^s I_{n,t}(u)\}$ *and* $\{X_t^s \bar{I}_{n,t}(u)\}$ *are* L_0*-APP with coefficients* w_{l_n} *of size* λ *and approximators* $\{(g_t^{(l_n)})^s I(g_t^{(l_n)} \leq b_n e^u)\}$ *and* $\{(g_t^{(l_n)})^s I(g_t^{(l_n)} > b_n e^u)\}$*, respectively.*

Since central limit theory exists for NED trimmed levels $X_t I(X_t \leq b_n)$ we now link APP to NED in nonextremes. Consider a [0, 1]-bounded version $b_n^{-1} X_t I(X_t \leq b_n)$. Theorem 2.6 characterizes a two-way relationship between tail-based E-NED and E-APP. Since the relationship for $b_n^{-1} X_t I(X_t \leq b_n)$ is complicated by its product convolution structure, we only have the one-way result below.

THEOREM 3.3. *Let* $\{X_t\}$ *be* L_0*-APP with size* λ *and approximator* $\{g_t^{(l)}\}$*, and choose any* $s > 0$*. Then* $\{b_n^{-s} X_t^s I_{n,t}\}$ *is* L_p*-NED for any* $p > 0$ *with size* $\lambda/\max\{p, 2\}$ *and constants* $d_{n,t} = d$.

Remark 11. The NED constants d do not depend on n and t because $b_n^{-1} X_t I(X_t \leq b_n)$ is uniformly bounded.

Remark 12. Since the proof exploits $b_n^{-1} X_t I(X_t \leq b_n) \in [0, 1]$ a similar argument for a standardized extreme level $X_t I(X_t > b_n)$ is not available.

The scale construction ensures that $b_n^{-1} X_t I(X_t \leq b_n)$ has infinitely many moments. See, for example, Hahn et al. (1990) for central limit theory for a Winsorized $b_n^{-1} \min\{X_t, b_n\}$ under independence. Asymptotic theory for the tail-trimmed sums $\sum_{t=1}^n X_t I(X_t \leq b_n)$, however, requires a scale $(\mathrm{E}(\sum_{t=1}^n X_t I(X_t \leq b_n))^2)^{1/2}$ that may deviate substantially from b_n or $\sqrt{n} b_n$ (e.g., Hill, 2009a; Hill and Renault, 2010).

THEOREM 3.4. *Let* $\{X_t\}$ *be* L_0*-APP with size* $\lambda > 0$ *and approximator* $\{g_t^{(l)}\}$*, and choose any* $s > 0$*. Then* $\{X_t^s I_{n,t}\}$ *is* L_p*-NED for any* $p > 0$ *with constants* $d_{n,t} = d$ *and coefficients* $\varpi_{l_n} = o(l_n^{-\theta})$ *of size* $\theta = \lambda/\max\{p, 2\}$.

Remark 13. The result implies that any L_0-APP process $\{X_t\}$ with arbitrarily thick tails can be negligibly trimmed into an L_2-NED process, guaranteeing a Gaussian central limit theory for $\{X_t I(X_t \leq b_n)\}$. This is immensely useful since it implies a versatile robust estimation theory. See Section 5.

Use the fact that L_p-NED with size λ implies L_0-APP with size $p\lambda$ and approximator $\{\mathrm{E}[X_t|\mathcal{F}_{t-l}^{t+l}]\}$ (Davidson, 1994, p. 274) to deduce the easy extension below of Theorems 3.3 and 3.4.

COROLLARY 3.5. *If* $\{X_t\}$ *is* L_p*-NED with size* $\lambda > 0$ *then for any* $s > 0$, $\{b_n^{-s} X_t^s I(X_t \le b_n)\}$ *and* $\{X_t^s I(X_t \le b_n)\}$ *are* L_q*-NED for any* $q > 0$ *with respective sizes* $p\lambda/\max\{q,2\}$ *and* $p\lambda/\max\{q,2\}$. *In each case the constants* $d_{n,t} = d$ *are time-invariant.*

4. TAIL AND NONTAIL MEMORY: EXAMPLES

We now characterize memory in ARCH(∞) class (4) and distributed lags (5), and discuss extensions to other nonlinear processes. In order to reduce repetitive claims, notice that Sections 2 and 3 show L_2-NED applied to tail events or exceedances $\{I(X_t > b_n e^u), (\ln(X_t/b_n)_+\}$ or tail-trimmed level power transforms $\{X_t^s I(X_t \le b_n)\}$ for any $s > 0$ results in time-independent constants, where the constants of $I(X_t > b_n e^u)$ are $O((k_n/n)^{1/2} e^{-u/2})$.

4.1. Linear Distributed Lags

Consider X_t in (5), and assume the errors ϵ_t have tail (2) with index $\kappa > 0$. Regular variation (2) and a bound on the coefficients permit a simple proof of L_0-APP. Notice that we only require the innovations to behave *like* an independent sequence and only in the tails. Recall $\mathcal{F}_t = \sigma(\epsilon_\tau : \tau \le t)$.

Assumption A.

1. X_t is the linear distributed lag (5) where ϵ_t has for each t tail (2) with index $\kappa > 0$. Further, ϵ_t is uniformly L_p-bounded, $p > 0$, and weakly tail orthogonal: $P(\sum_{i=0}^{\infty} |a_i \epsilon_{t-i}| > x) \sim \sum_{i=0}^{\infty} P(|a_i \epsilon_{t-i}| > x)$ for all real sequences $\{a_i\}_{i=0}^{\infty}$, $\sum |a_i|^\kappa < \infty$.
2. Let $\psi_{t,0} = 1$ for each t. The remaining coefficients $\psi_{t,i}$ are measurable with respect to $\mathcal{F}_{t-i-1}$, and there exists a sequence of nonstochastic real numbers $\{\psi_i\}$ satisfying $|\psi_{t,i}| \le |\psi_i|$ a.s., where $\psi_i = O(i^{-\mu})$ for some $\mu > 1/\kappa$, or $\psi_i = O(\rho^i)$.

Remark 14. Weak tail orthogonality is substantially weaker than independence while independence guarantees it (e.g., Feller, 1971; Cline, 1983). The abstraction is not empty since it, and not independence, captures the error dependence properties for a distributed lag representation of a bilinear process (Hill, 2008b, 2010). See Example 2, below.

Remark 15. The property $P(\sum_{i=0}^{\infty} |a_i \epsilon_{t-i}| > x) \sim \sum_{i=0}^{\infty} P(|a_i \epsilon_{t-i}| > x)$ is a special case of a larger set of probability and moment inequalities for an encompassing class of processes. See especially Nagev (1979, 1998) and de la Peña, Ibrahimov, and Sharakhmetov (2003) for deep results.

Use $|\psi_{t,i}| \leq |\psi_i|$ a.s., weak tail orthogonality, properties of regularly varying tails (2), and summability $\sum_{i=0}^{\infty} |\psi_i|^\kappa < \infty$ to deduce for any sequence $\{f_t\}$, $\inf_{t \in \mathbb{Z}} f_t > 0$, any $\delta > 0$, and any $l \in \mathbb{N}$ (cf. Feller, 1971; Resnick, 1987)

$$P\left(\left|X_t - \sum_{i=0}^{l} \psi_{t,i}\epsilon_{t-i}\right| > f_t\delta\right) \leq P\left(\sum_{i=l+1}^{\infty} |\psi_i| \times |\epsilon_{t-i}| > f_t\delta\right)$$
$$\sim \sum_{i=l+1}^{\infty} |\psi_i|^\kappa P\left(|\epsilon_{t-i}| > f_t\delta\right) \leq \sum_{i=l+1}^{\infty} |\psi_i|^\kappa .$$

The next claim is therefore straightforward to prove.

LEMMA 4.1. *Under Assumption A, $\{X_t\}$ is L_0-APP on $\{\epsilon_t\}$ with F_{t-l}^{t+l}-measurable approximator $\{\sum_{i=0}^{l} \psi_{t,i}\epsilon_{t-i}\}$ and arbitrary size $\lambda > 0$ if $\psi_i = O(\rho^i)$, or size $\lambda = \mu\kappa - \iota > 1$ if $\psi_i = O(i^{-\mu})$.*

Remark 16. Summability $\sum_{i=0}^{\infty} |\psi_i|^\kappa < \infty$ forces a restriction on hyperbolic memory since $\sum_{i=0}^{\infty} |\psi_i|^\kappa \leq K \sum_{i=0}^{\infty} i^{-\mu\kappa} < \infty$ requires $\mu\kappa > 1$. This reveals a standard memory-moment trade-off: As $\kappa \searrow 0$ for heavier tails we require monotonically weaker memory $\psi_i = O(i^{-\mu}) \searrow 0$ since $\mu \nearrow \infty$.

In lieu of Lemma 4.1 apply Remark 9 and Theorems 2.5 and 3.4 to deduce that the nonlinear distributed lag X_t has L_2-NED extreme events $I(X_t > b_n e^u)$, exceedances $(\ln(X_t/b_n))_+$, and nonextreme levels $X_t I(X_t \leq b_n)$.

THEOREM 4.2. *Let Assumption A hold and let $\mu\kappa > 1$.*

(*i*) *$\{X_t\}$ is L_2-E-NED on $\{\epsilon_t\}$ with arbitrary size in the geometric case, or size $\mu\kappa/2 - \iota$ if $\psi_i = O(i^{-\mu})$.*

(*ii*) *$\{X_t^s I(X_t \leq b_n)\}$ is L_2-NED on $\{\epsilon_t\}$ with time-independent constants, and arbitrary size in the geometric case, or size $\mu\kappa/2 - \iota$ if $\psi_i = O(i^{-\mu})$.*

Examples of processes that satisfy Assumption A include ARFIMA and bilinear.

Example 1 (ARFIMA). Consider an ARFIMA$(1, d, 1)$ process$\{X_t\}$, $d < 1$:

$$(1 - L)^d (1 - \phi L) X_t = (1 + \theta L)\epsilon_t, \qquad |\phi| < 1,$$
$$\epsilon_t \overset{iid}{\sim} (2) \qquad \text{with index } \kappa > 1/(1 - d).$$

Since X_t has representation $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ for i.i.d $\epsilon_t \sim (2)$ and $\psi_i = O(i^{d-1})$ (Hosking, 1981), Assumption A is satisfied: Independence implies ϵ_t is weakly tail orthogonal, and $\psi_i = O(i^{-\mu})$ for $\mu = 1 - d > 1/\kappa > 0$. Therefore $\{X_t\}$ is L_0-APP on $\{\epsilon_t\}$ with size $(1 - d)\kappa - \iota > 1$ by Lemma 4.1.

If $d = -1/2$, for example, the tail index must satisfy $\kappa > 2/3$ for Assumption A to hold. In general the memory-moment trade-off implies as $d \nearrow 1$, where $I(1)$ is the limit, the smallest allowed tail index must increase $\kappa \nearrow \infty$. The long-memory range $d \in (1/2, 1)$ requires finite variance $\kappa \in (2, \infty)$. Nevertheless, when $d \in (0,1)$ the first difference $\Delta X_t := X_t - X_{t-1}$ is ARFIMA$(1, \tilde{d}, 1)$ for some $\tilde{d} = d - 1 < 0$ so $\{\Delta X_t\}$ satisfies Assumption A for any comparatively small $\kappa > 1/(2-d)$. In this case, for example, if $\kappa \geq 1$ then all $d < 1$ are covered.

Example 2 (Bilinear). Consider a simple bilinear process

$$X_t = \phi X_{t-1} u_{t-1} + u_t, \qquad u_t \overset{iid}{\sim} (2) \text{ with index } \kappa > 0,$$
$$\phi > 0, \qquad \phi^{\kappa/2} \mathrm{E}[u_t^{\kappa/2}] < 1.$$

Then X_t has the representation $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$, $\psi_i = O(\rho^i)$, $\rho \in (0,1)$, and $\epsilon_t \sim$ (2) with index $\kappa/2$ (Davis and Resnick, 1996, Cor. 2.4). Further, ϵ_t is weakly tail orthogonal by the proof of Lemma 8 in Hill (2010).

We impose $\mathcal{F}_{t-i-1}$-measurability of $\psi_{t,i}$ in Assumption A.2 to ensure that the NED base is the weakly orthogonal $\{\epsilon_t\}$ and therefore a simple $\mathcal{F}_{t-l_n}^{t+l_n}$-measurable approximator is available. The latter are key to proving Lemma 4.1. But such measurability does not characterize many nonlinear autoregressions.

Example 3 (STAR). Consider an exponential smooth transition autoregression: $X_t = \phi X_{t-1} \exp\{-\gamma X_{t-1}^2\} + \epsilon_t$, $\gamma > 0$, $|\phi| < 1$ with $\epsilon_t \overset{iid}{\sim} (2)$. Then $X_t = \sum_{i=0}^{\infty} \psi_{t,i} \epsilon_{t-i}$ where $\psi_{t,0} = 1$ and the remaining $\psi_{t,i} = \phi^i \sqcap_{j=1}^{i} \exp\{-\gamma X_{t-j}^2\}$ are $\mathcal{F}_{t-1}$-measurable for each $i \geq 1$. Thus Assumption A.2 does not hold.

In order to allow such nonlinear feedback, further restrictions on the errors are evidently required. See Sections 4.3 and 4.4, below.

4.2. GARCH

Consider GARCH class (1).

Example 4 (GARCH). Write the GARCH(p,q) process as

$$X_t = \sigma_t \epsilon_t \quad \text{and} \quad \beta(L)\sigma_t^2 = \omega + \{\delta(L) - \beta(L)\} X_t^2$$

with lag polynomials $\delta(L) = 1 - \sum_{i=1}^{p} \delta_i L^i$ and $\beta(L) = 1 - \sum_{i=1}^{q} \beta_i L^i$. If the roots of $\beta(z)$ lie outside the unit circle then (4) holds with $\pi(z) = 1 - \delta(z)/\beta(z)$ and $\pi_i = O(\rho^i)$. Under covariance stationarity $\epsilon_t \overset{iid}{\sim} (0,1)$ and $S := \pi(1) < 1$, Davidson (2004, Thm. 3.2) proves that $\{X_t\}$ is geometrically L_1-NED on $\{\epsilon_t\}$. Apply Theorem 2.1 and Corollary 3.5 to deduce that $\{I(X_t > b_n e^u), (\ln(X_t/b_n))_+, X_t^s I(X_t \leq b_n)\}$ are geometrically L_2-NED.

Example 5 (FIGARCH). Davidson's (2004; Thms. 3.1–3.2) covariance stationarity condition $S < 1$ rules out NED for IGARCH and FIGARCH since (4) is then satisfied with

$$\pi(L) = 1 - \frac{\delta(L)}{\beta(L)}(1-L)^d, \qquad d \in (0,1]$$

for some $\delta(L)$. Both cases correspond with $S = \pi(1) = 1$, where IGARCH ($d = 1$) exhibits geometric decay $\pi_i = O(\rho^i)$, and FIGARCH ($d < 1$) hyperbolic decay since $(1 - L)^d = 1 - \sum_{i=1}^{\infty} \gamma_i L^i$ with $\gamma_i = O(i^{-1-d})$.

In order to characterize the tail and nontail memory properties of short-range memory GARCH processes with unit ($S = 1$) or explosive ($S > 1$) roots, Davidson (2004) suggests a different approach based on approximability.

Assumption B. Let $X_t = \sigma_t \epsilon_t$ satisfy (4) with $\epsilon_t \overset{iid}{\sim} (0,1)$, $0 \leq \pi_i \leq C\rho^i$ and $C \in [0, 1/\rho)$.

Under Assumption B $\{X_t\}$ is geometrically L_0-APP on $\{\epsilon_t\}$ by Theorem 3.3 of Davidson (2004). Now invoke Theorems 2.1 and 3.4 above to deduce that GARCH processes have geometrically NED extremes and nonextremes.

THEOREM 4.3. *Under Assumption B* $\{I(X_t > b_n e^u), (\ln(X_t/b_n)_+, X_t^s I(X_t \leq b_n)\}$ *are geometrically* L_2*-NED on* $\{\epsilon_t\}$.

Remark 17. Since $S = \sum_{i=0}^{\infty} \pi_i \leq C\rho/(1 - \rho)$ clearly $S = 1$ and $S > 1$ are easily feasible, covering IGARCH and many explosive GARCH cases.

Theorem 4.3 exploits L_0-APP to get around the hairline noncovariance stationary case $S = 1$ under short-range memory. Davidson (2004, eqs. (5.1) and (5.4)), however, suggests a new class of hyperbolic GARCH(p, d, γ, q) models (HYGARCH) to conquer the FIGARCH property $S = 1$. Assume $X_t = \sigma_t \epsilon_t$ is governed by (4) where

$$\pi(L) = 1 - \frac{\delta(L)}{\beta(L)} \left[1 + \gamma \left((1 - L)^d - 1\right)\right], \qquad d \in (0, 1], \qquad \gamma \geq 0 \tag{13}$$

or

$$\pi(L) = 1 - \frac{\delta(L)}{\beta(L)} \left[1 + \frac{\gamma}{\zeta(1 + d)} \sum_{i=0}^{\infty} i^{-1-d} L^i\right], \qquad d > 0, \qquad \gamma \geq 0, \tag{14}$$

and $\zeta(\cdot)$ is the Riemann zeta function. The index $d > 0$ as usual governs the degree of hyperbolic memory. The case $d > 1$ is ruled out in (13) since it permits negative coefficients. FIGARCH and GARCH correspond to (13) with $\gamma = 1$ and $\gamma = 0$, respectively; IGARCH(1, 1) can be deduced from (13) with $d = 1$ and $\delta(L) = 1$; and $\gamma \geq 1$ in (13) or (14) aligns with nonstationary cases. The tuning parameter γ is key for allowing both hyperbolic decay and covariance stationarity $S < 1$. As long as $d > 0$ then $S < 1$ in both (13) and (14) for any covariance stationary case corresponding to $\gamma < 1$.

Each condition of Davidson's (2004) Theorem 3.1 is satisfied when $\epsilon_t \overset{iid}{\sim} (0,1)$, $\gamma < 1$, and $d \in (0, 1)$ in (13) or $d > 0$ in (14): $\{X_t\}$ is L_2-bounded and L_1-NED on $\{\epsilon_t\}$ with size $\lambda = d - \iota$. Invoke Theorem 2.1 and Corollary 3.5 to deduce that such HYGARCH processes have well-defined tail and nontail memory properties.

THEOREM 4.4. *Let $X_t = \sigma_t \epsilon_t$ satisfy (4) where $\epsilon_t \overset{iid}{\sim} (0,1)$. Specifically, $\pi(L)$ satisfies HYGARCH(p,d,γ,q) (13) with $d \in (0,1]$ or (14) with $d > 0$, and in general $\gamma \in (0,1)$. Then $\{I(X_t > b_n e^u), (\ln(X_t/b_n)_+, X_t^s I(X_t \leq b_n)\}$ are L_2-NED on $\{\mathcal{F}_t\}$ with size $d/2 - \iota$.*

Remark 18. Although L_2-boundedness for X_t is a severe restriction, higher moments need not exist, and the tail-trimmed $X_t^s I(X_t \leq b_n)$ is necessarily L_2-NED. In particular, the central limit theory for a tail-trimmed sample variance and covariance exists when X_t^2 and $X_t X_{t-h}$ have an infinite variance (Hill, 2009a).

Example 6 (HYGARCH). Consider the HYGARCH$(1,2,\gamma,1)$ model

$$X_t = \sigma_t \epsilon_t \quad \text{and} \quad \sigma_t^2 = \pi_0 + \left(1 - \frac{1-\delta}{1-\beta}\left[1 + \frac{\gamma}{\zeta(2+\iota)} \sum_{i=0}^{\infty} i^{-3} L^i\right]\right) X_t^2,$$

where $\beta \in (0,1)$, $\delta > 0$ and $\gamma \in (0,1)$. By Theorem 4.4 $\{I(X_t > b_n e^u), (\ln(X_t/b_n)_+, X_t^s I(X_t \leq b_n)\}$ are L_2-NED with size $1/2$.

4.3. Nonlinear AR-Nonlinear GARCH

A large array of nonlinear AR-nonlinear GARCH or SV processes with short-range memory and i.i.d. innovations ϵ_t are geometrically ergodic or β-mixing, and therefore geometrically α-mixing. In these cases $\{I(X_t > b_n e^u), (\ln(X_t/b_n))_+), X_t^s I(X_t \leq b_n)\}$ are L_2-NED with arbitrary size. Examples include threshold autoregressions, single layer feed-forward neural networks, random coefficient autoregressions, and multiplicative-, exponential-, GJR-, and threshold-GARCH, to name a few. See An and Huang (1996), Ling (1999), Carrasco and Chen (2002), Cline and Pu (2004), and Meitz and Saikkonen (2008).

Compare the following to Examples 1, 2, 4, and 5 where distribution smoothness for the errors is not required and hyperbolic memory cases are covered.

Example 7 (A-GARCH). Engle and Ng (1993) propose the following asymmetric-GARCH(1,1) process $X_t = \sigma_t \epsilon_t$ where

$$\sigma_t^2 = \omega + \alpha(\epsilon_{t-1} - c)^2 \sigma_{t-1}^2 + \beta \sigma_{t-1}^2, \qquad \omega > 0, \qquad \alpha, \beta \geq 0, \qquad c \in \mathbb{R}.$$

Assume $\epsilon_t \overset{iid}{\sim} (0,1)$ has a positive, continuous density with respect to Lebesgue measure on $\mathbb{R}$, and X_0 is initialized from the invariant distribution. The case $c = 0$ corresponds to GARCH(1,1). If $E[\beta + \alpha(\epsilon_{t-1} - c)^2] < 1$ or $\alpha + \beta < \{E[(\epsilon_{t-1} - c)^2]\}^{-1}$ then $\{X_t\}$ is geometrically β-mixing (Carrasco and Chen, 2002, Cor. 6). If the degree of asymmetry is slight ($c = .01$, for example) then $\alpha + \beta < .9999$ suffices, and if $c = 1$ then $\alpha + \beta < 1/2$ must hold, exhibiting an asymmetry-memory trade-off.

Example 8 (TAR). Consider a threshold autoregression of order one:

$$X_t = \phi_1 X_{t-1} I(X_{t-1} \leq c) + \phi_2 X_{t-1} I(X_{t-1} > c) + \epsilon_t = \phi_{t-1} X_{t-1} + \epsilon_t,$$

say, where i.i.d. ϵ_t is mean zero with positive density $\mathbb{R}$-$a.e.$, $\mathrm{E}|\epsilon_t| < \infty$, $c \in \mathbb{R}$, and $|\phi_t| \leq \max\{|\phi_1|, |\phi_2|\} \in (0,1)$. Then $\{X_t\}$ is geometrically ergodic (An and Huang, 1996, Thm. 3.1).

Example 9 (STAR). If the STAR process $\{X_t\}$ of Example 3 has innovations ϵ_t that satisfy Example 8 then $\{X_t\}$ is geometrically ergodic (An and Huang 1996, Thm. 3.1).

4.4. Distributed Lags with Random Volatility Errors

Asymptotic theory for NED, E-NED, and tail-trimmed NED processes $\{X_t\}$ permits the base $\{\epsilon_t\}$ to satisfy a mixing condition (Davidson, 1992; de Jong, 1997; Hill, 2009a, 2009b, 2010). This is helpful for modeling the memory properties of linear and nonlinear ARMA-GARCH or SV with hyperbolic memory.

By comparison the distributed lag Assumption A imposes regular variation and weak tail orthogonality on the errors and $\mathcal{F}_{t-i-1}$-measurability on $\psi_{t,i}$ to ensure the weaker L_0-APP property holds. Neither regular variation nor weak tail memory is guaranteed to hold for distributed lags with linear or nonlinear GARCH errors (Chernick et al., 1991; Mikosch and Stărică, 2000; Borkovec and Klüppelberg, 2001; Basrak et al 2002b; Cline, 2007), and coefficient measurability fails for random coefficient autoregressions like SETAR and STAR. If we require a NED array like $\{I(X_t > b_n e^u)\}$, $\{(\ln(X_t/b_n)_+\}$, or $\{X_t^s I(X_t \leq b_n)\}$ to exhibit hyperbolic memory and have a linear or nonlinear GARCH or SV base $\{\epsilon_t\}$, something more than Sections 4.1–4.3 must be developed.

Consider a generalization of (5):

$$X_t = \sum_{i=0}^{\infty} \psi_{t,i} u_{t-i}, \ \left|\psi_{t,i}\right| \leq |\psi_i| \text{ a.s.}, \qquad \sum_{i=0}^{\infty} |\psi_i| < \infty,$$
$$\sup_{t\in\mathbb{N}} \|u_t\|_p < \infty \quad \text{for } p > 0,$$

assume $\psi_{t,i}$ are measurable, and define the σ-field induced by $\{\psi_{t,i}u_{t-i}\}$:

$$G_t := \sigma\left(\psi_{\tau,i}u_{\tau-i} : t \leq \tau, i \geq 0\right).$$

If $\{\psi_{t,i}u_{t-i}\}$ is the base then $\left|\psi_{t,i}\right| \leq |\psi_i|$ and uniform L_p-boundedness imply $\{X_t\}$ is L_p-NED.

LEMMA 4.5. *$\{X_t\}$ is L_p-NED on $\{G_t\}$ with arbitrary size if $\psi_i = O(\rho^i)$ and size $\lambda = \mu - \iota$ if $\psi_i = O(i^{-\mu})$.*

Since L_p-NED with size λ implies L_0-APP with size $p\lambda$, simply replace $\mathcal{F}_t$ with G_t in Theorem 4.2 to deduce that Lemma 4.5 implies NED extremes and nonextremes.

THEOREM 4.6. *Each $\{I(X_t > b_n e^u)$, $(\ln(X_t/b_n)_+$, $X_t^s I(X_t \leq b_n)\}$ is L_2-NED on $\{G_t\}$ with arbitrary size if $\psi_i = O(\rho^i)$ or size $p\mu\kappa/2 - \iota$ if $\psi_i = O(i^{-\mu})$.*

Notice we need say nothing about the stochastic nature of $\psi_{t,i}$ nor orthogonality of ϵ_t. The key to a central limit theory for tail and nontail arrays of $\{X_t\}$ is to demonstrate that $\{\psi_{t,i}u_{t-i}\}$ satisfies a mixing property.

Example 10 (SETAR-STGARCH). Consider a stationary self exciting threshold-autoregression (SETAR) with exponential smooth transition GARCH (1,1) errors:

$$X_t = \phi X_{t-1} I(X_{t-1} < 0) + u_t, \qquad |\phi| < 1, \qquad u_t = \sigma_t \epsilon_t, \epsilon_t \overset{iid}{\sim} (0,1) \qquad \textbf{(15)}$$
$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \tilde{\alpha} u_{t-1}^2 \exp\left\{-\gamma u_{t-1}^2\right\} + \beta \sigma_{t-1}^2$$

where $\gamma \geq 0$, $\omega > 0$, and $\{\alpha, \alpha + \tilde{\alpha}, \beta\} \geq 0$. See Tong and Lim (1980) for SETAR and González-Rivera (1998) for STGARCH, inter alia. Assume $\text{E}[(\beta + \max\{\alpha, \alpha + \tilde{\alpha}\}\epsilon_t^2)^r] < 1$. Write $X_t = \phi_{t-1}X_{t-1} + u_t = \sum_{i=0}^{\infty} \psi_{t,i}u_{t-i}$, where $\phi_{t-1} = \phi I(X_{t-1} < 0)$, $\psi_{t,0} = 1\ \forall t$, and $\psi_{t,i} = \sqcap_{j=1}^{i}\phi_{t-j}$ satisfies $|\psi_{t,i}| \leq |\psi_i| = O(\rho^i)$. Both $\{u_t, x_t\}$ are geometrically ergodic (Meitz and Saikkonen, 2008, Prop. 1). Therefore $\psi_{t,i}u_{t-i} = \phi^i \sqcap_{j=1}^{i} I(X_{t-j} < 0) \times u_{t-i}$ for $i \geq 1$ is geometrically α-mixing. But this implies that $\{X_t\}$ is geometrically L_p-NED for some $p > 0$ on a geometrically α-mixing base by Lemma 4.5, so Theorem 4.6 applies.

5. EXTREMAL AND ROBUST STATISTICS

In this section we prove that tail index and tail dependence estimators, intermediate order statistics, and a tail-trimmed sum are asymptotically normal for L_0-APP data. This includes X_t in (1), (4), and (5), and the myriad nonlinear processes discussed above. We also show how the major tail and nontail results simultaneously deliver new robust estimator asymptotics.

5.1. Order Statistics and Tail Index

Consider a second-order Paretian class $P(X_t > x) = cx^{-\kappa}(1 + O(x^{-\varsigma})), c, \varsigma > 0$ (e.g., Hall, 1982; Haeusler and Teugels, 1985). The class covers linear distributed lags like ARFIMA, and bilinear as long as the errors are i.i.d. with power-law tail decay (Cline, 1983; Davis and Resnick, 1996); and it naturally characterizes GARCH (Mikosch and Stărică, 2000; Basrak et al. 2002b), linear and nonlinear AR-GARCH (Borkovec and Klüppelberg, 2001; Cline, 2007), and auction prices near the reserve price (Hill and Schneyerov, 2010).

The widely used Hill (1975) estimator of κ^{-1} is $\hat{\kappa}_n^{-1} = 1/k_n \sum_{i=1}^{k_n} \ln(X_{(i)}/X_{(k_n+1)})$. Let $k_n \to \infty$ and $k_n/n^{2\varsigma/(2\varsigma+\kappa)} \to 0$. Theorem 2.5 ensures that all conditions of Hill's (2010, Thm. 2 and Lem. 3) Gaussian limit theory for $X_{(k_n+1)}$ and $\hat{\kappa}_n^{-1}$ are satisfied. Classic treatments can be found in Galambos (1987) and Embrechts et al. (1997). See Hsing (1991) and Hill (2010) for surveys.

THEOREM 5.1. *If* $\{X_t\}$ *is* L_0*-APP with size* 1 *on* α*-mixing* $\{\epsilon_t\}$*with size* 1 *then*

$$k_n^{1/2}\ln\left(X_{(k_n+1)}/b_n\right) \stackrel{d}{\rightarrow} N(0,w^2) \quad \textit{and} \quad k_n^{1/2}\left(\hat{\kappa}_n^{-1}-\kappa^{-1}\right)/\upsilon_n \stackrel{d}{\rightarrow} N(0,1)$$

where $w^2 = \kappa^{-2}\lim_{n\to\infty} \text{E}(1/k_n^{1/2}\sum_{t=1}^n I_{n,t}(u/k_n^{1/2}))^2 < \infty$ *and* $\upsilon_n^2 := \text{E}[(k_n^{1/2} (\hat{\kappa}_n^{-1}-\kappa^{-1}))^2] = O(1)$.

Remark 19. A similar argument reveals that Hill's (2010, Thm. 3) kernel estimator $\hat{\upsilon}_n^2$ is consistent for υ_n^2 under L_0-APP.

Since Theorem 5.1 both encompasses and augments Hill's (2010) results, evidently this is the most general limit theory available for intermediate order statistics and tail index estimators under data dependence and heterogeneity. The size restrictions are irrelevant in the geometric decay case, covering IGARCH, explosive GARCH, the Example 7 asymmetric GARCH, Example 9 STAR, and Example 10 SETAR-STGARCH models.

Each hyperbolic memory process discussed above is also covered provided memory decay is carefully considered. It is easy to verify the following cases: The Example 1 ARFIMA$(1,d,1)$ with innovations $\epsilon_t \stackrel{iid}{\sim} (2)$, $\kappa > 0$, and Hurst index $d < 1 - 1/\kappa$ is L_0-APP with size 1 on i.i.d. $\{\epsilon_t\}$ covering long-memory cases when $\kappa > 2$; the Example 6 HYGARCH(p,d,γ,q) with i.i.d innovations $\{\epsilon_t\}$ is L_1-NED on i.i.d. $\{\epsilon_t\}$ with size $\lambda = d - \iota$ (Davidson, 2004) and therefore L_0-APP with size 1 for any $d > 0$.

5.2. Tail Dependence Coefficient (Extremogram)

Consider a bivariate process $\{Z_t\}$, $Z_t = [X_t, Y_t]'$, on $[0,\infty) \times [0,\infty)$ with marginal tails (2) and indices $\kappa_x, \kappa_y > 0$. Denote by $\{b_{x,n}, b_{y,n}\}$ the associated threshold sequences, e.g., $(n/k_n)P(X_t > b_{x,n}) \to 1$, where $1 \leq k_n < n$, $k_n = o(n)$ and $k_n \to \infty$, and define

$$\hat{I}_{x,n,t} := I\left(X_t > X_{(k_n+1)}\right) \quad \text{and} \quad \hat{I}_{y,n,t} := I\left(Y_t > Y_{(k_n+1)}\right).$$

The statistic

$$\hat{r}_n(h) = \frac{1}{k_n}\sum_{t=1}^n \left(\hat{I}_{x,n,t-h}\hat{I}_{y,n,t} - \left(\frac{k_n}{n}\right)^2\right)$$

nonparametrically estimates the tail event correlation

$$r_n(h) := \frac{P_{h,n} - P_{x,n}P_{y,n}}{\left(P_{x,n}P_{y,n}\right)^{1/2}} \sim \frac{n}{k_n}\left(P_{h,n} - P_{x,n}P_{y,n}\right),$$

where $P_{h,n} := P(X_{t-h} > b_{x,n}, Y_t > b_{y,n})$ and $P_{x,n} := P(X_t > b_{x,n})$. Davis and Mikosch (2009c) call the limit

$$r(h) = \lim_{n\to n} r_n(h) = \lim_{n\to n} \frac{P_{h,n} - P_{x,n}P_{y,n}}{\left(P_{x,n}P_{y,n}\right)^{1/2}}$$

the "extremogram." Trivially $P_{x,n}P_{y,n}/(P_{x,n}P_{y,n})^{1/2} = (P_{x,n}P_{y,n})^{1/2} = o(1)$, hence the extremogram is identically

$$r(h) = \lim_{n\to n} \frac{P_{h,n}}{\left(P_{x,n}P_{y,n}\right)^{1/2}}.$$

Define $\tilde{r}(h) := \lim_{n\to\infty}(n/k_n)r_n(h) = \lim_{n\to\infty}\{P_{h,n}/(P_{x,n}P_{y,n}) - 1\}$ if the limit exists.

Tail dependence at displacement h is exhibited if $\tilde{r}(h) \neq 0$ covering "local" forms $r(h) = 0$ yet $\tilde{r}(h) \neq 0$; and "distant" forms $r(h) \neq 0$, hence $\tilde{r}(h) = \infty$.[3] See Hill (2008a, 2008b) for details. Note that Davis and Mikosch (2009c) only consider distant tail dependence that neglects SV. See Section 6.1 below for details on $r_n(h)$ and its relationship with other tail dependence measures.

Write $r_{n,h} := [r_n(1), \ldots, r_n(h)]'$ and $\hat{r}_{n,h} := [\hat{r}_n(1), \ldots, \hat{r}_n(h)]'$. If $\{Z_t\}$ is L_0-APP with size 2 on some possibly vector-valued base $\{\epsilon_t\}$ then it is L_2-E-NED with size 1 on $\{\epsilon_t\}$ (Corollary 2.7), and therefore L_4-E-NED with size 1/2 on $\{\epsilon_t\}$ (Theorem 2.1). If ϵ_t is α-mixing with size 1 then $\{X_t, Y_t\}$ satisfies all relevant tail decay and tail memory properties required of Hill (2008b, Thm 3.2): As long as sufficiently many extremes are used $k_n/n^{2/3} \to \infty$ then

$$\sqrt{k_n}\left(\hat{r}_{n,h} - r_{n,h}\right) \xrightarrow{d} N(0, V),$$

where $V = \lim_{n\to\infty} k_n \mathrm{E}[(\hat{r}_{n,h} - r_{n,h})(\hat{r}_{n,h} - r_{n,h})'] \in \mathbb{R}^{h\times h}$ is positive definite. Hill (2008b, 2009b) shows that a test of tail independence based on $k_n^{1/2}\hat{r}_{n,h}$ is consistent against local (i.e., $r(h) = 0$, $\tilde{r}(h) \neq 0$) or distant (i.e., $r(h) \neq 0$, $\tilde{r}(h) = \infty$) alternatives. See Davis and Mikosch (2009a) for the same estimator under distant dependence, joint regularly variation, and α-mixing, a subset of the processes allowed here.

The hyperbolic memory case is complicated by feedback between $\hat{I}_{x,n,s-h}\hat{I}_{y,n,s}$ and $\hat{I}_{x,n,t-h}\hat{I}_{y,n,t}$.

Example 11 (Stochastic volatility). Let $\epsilon_t = [\epsilon_{x,t}, \epsilon_{y,t}]' \in \mathbb{R}^2$ be i.i.d. with symmetric marginal tails (2) and indices $\kappa = [\kappa_x, \kappa_y]' > 0$. Consider a bivariate random variable $Z_t = [X_t, Y_t]'$ with stochastic volatility

$$X_t = \sigma_{x,t}\epsilon_{x,t} \quad \text{and} \quad Y_t = \sigma_{y,t}\epsilon_{y,t}.$$

Assume $\sigma_t^{\kappa} = [\sigma_{x,t}^{\kappa_x}, \sigma_{y,t}^{\kappa_y}]' \in \mathbb{R}_+^2$ is independent of ϵ_t and governed by log-ARFIMA$(1, d, 1)$,

$$(I_2 - \Phi L)(I_2 - L)^d \ln\sigma_t^{\kappa} = \Theta + \zeta_t, \qquad \Theta \in \mathbb{R}_+^2, \qquad \Phi \in \mathbb{R}_+^{2\times 2}$$

$$\zeta_t \in \mathbb{R}^2,\ \zeta_t \overset{iid}{\sim} N(0, I_2), \qquad d < 1 - 2/\kappa,$$

where the roots of $I_2 - \Phi z$ lie outside the unit circle. By imitating Lemma 4.1 it is easy to show that $\{\ln\sigma_t^{\kappa}\}$ is L_0-APP on i.i.d. $\{\epsilon_t\}$ with size $(1 - d)\kappa - \iota \geq 2$.

Since $\zeta_t \overset{iid}{\sim} N(0, I_2)$ and each $\{\text{E}|\epsilon_{x,t}|^{\kappa_x - \iota}, \text{E}|\epsilon_{y,t}|^{\kappa_y - \iota}\} < \infty$ it similarly follows that $\{Z_t\}$ is L_0-APP on i.i.d. $\{\epsilon_t\}$ with size 2 (Davidson, 1994, Thm. 17.22). For example, if $\kappa = 1.5$ then Hurst indices $d < -1/3$ are allowed, and long-memory when $\kappa > 4$. See Hill (2008a, 2008b) for E-NED characterizations when $d = 0$.

5.3. Tail-Trimmed Sums

Although the use of trimmed or truncated sums for robust estimation has a substantial history,[4] the vast majority of cases cover independent data with rare exceptions including mixing data with finite variance (Hahn et al., 1987) and fixed quantile trimmed linear distributed lags (Wu, 2005). Theorem 3.3, by comparison, allows for asymptotically negligible trimming for a far larger class of dependent heterogeneous heavy-tailed processes.

Assume $X_t := |Y_t|$ for some symmetrically distributed L_p-bounded process $\{Y_t\}$, $p \in (0, 2)$, and let positive real sequences $\{k_n, b_n\}$ satisfy $(n/k_n)P(X_t > b_n) \to 1$. Recall $\hat{X}_{n,t} := X_t I(X_t \leq b_n) = X_t I_{n,t}$, define $\upsilon_n^2 := \text{E}(\sum_{t=1}^n \{\hat{X}_{n,t} - \text{E}[\hat{X}_{n,t}]\})^2$, and define the self-normalized tail-trimmed level for scalar X_t:

$$\hat{Z}_{n,t}^* := \left(\hat{X}_{n,t}^* - \text{E}\left[\hat{X}_{n,t}\right]\right)/\upsilon_n, \qquad \text{where } \hat{X}_{n,t}^* = X_t I\left(X_t \leq X_{(k_n+1)}\right).$$

Asymmetric trimming is identical in theory with only added notation. The trick is to fuse tail and nontail asymptotics simultaneously by first showing $\sum_{t=1}^n \{\hat{X}_{n,t}^* - \hat{X}_{n,t}\} = o(\upsilon_n)$, which requires the extreme value result $X_{(k_n+1)}/b_n = 1 + O_p(1/k_n^{1/2})$, and then delivering a central limit theorem for self-normalized trimmed sums $1/\upsilon_n \sum_{t=1}^n \{\hat{X}_{n,t} - \text{E}[\hat{X}_{n,t}]\}$. The theorem below exploits arguments in Hill (2009a), where only geometric memory is considered.

THEOREM 5.2. *Let* $\liminf_{n \geq \infty}\{\upsilon_n^2/n\} > 0$, $b_n/n^\iota \to \infty$ and $b_n = o(n^{1/2})$. *Further,* $\{X_t\}$ *is* L_p*-bounded geometrically* L_0*-APP on geometrically* α*-mixing* $\{\epsilon_t\}$. *Then* $\sum_{t=1}^n \hat{Z}_{n,t}^* \overset{d}{\to} N(0, 1)$.

Remark 20. The regulatory condition $\inf_{n \geq N}\{\upsilon_n^2/n\} > 0$ ensures that the trimmed sum $\sum_{t=1}^n X_t I(X_t \leq b_n)$ is not degenerate asymptotically.

Remark 21. The trimming threshold b_n must be restricted to ensure that sufficiently many tail observations are trimmed for asymptotic normality. The bound $b_n = o(n^{1/2})$ enforces $\max_{1 \leq t \leq n}\{|\hat{X}_{n,t}|\} = o_p(n^{1/2})$, the relative stability property shared by weakly dependent square integrable processes in the maximum domain of attraction of a Type II extreme value distribution (Leadbetter et al., 1983; Naveau, 2003).

The thresholds b_n are intimately related to the number of trimmed observations k_n through $P(X_t > b_n) \sim k_n/n$. Suppose X_t has tail (2) with index $\kappa \in (0, 2]$, and impose $k_n \sim n^\delta$, $\delta \in (0, 1)$, for simplicity.

Example 12 (Paretian tails and trimming). Under (2) $b_n = K(n/k_n)^{1/\kappa} \sim Kn^{(1-\delta)/\kappa}$. The bound $b_n = o(n^{1/2})$ reduces to $\delta > 1 - \kappa/2$. Heavier-tailed data generating processes $\kappa \searrow 0$ result in samples with more extremes on average, hence more observations must be trimmed $k_n \sim n^{\delta} \nearrow n$ to ensure asymptotic normality. If $\kappa = 1.5$ or $\kappa = .75$ then trimming at least $[n^{.25}]$ or $[n^{.625}]$ observations per sample $\{X_t\}_{t=1}^n$, respectively, ensures $\sum_{t=1}^n \hat{Z}_{n,t}^* \xrightarrow{d} N(0,1)$.

5.4. Tail-Trimmed Method of Moments

The tail-trimmed sum lies at the heart of a new robust minimum distance estimator. Consider estimating the autoregression parameter θ^0 of a stationary AR(1):

$$X_t = \theta^0 X_{t-1} + \epsilon_t, \qquad \left|\theta^0\right| < 1, \qquad \epsilon_t \overset{iid}{\sim} (2) \quad \text{with index } \kappa \in (1,2], \quad \text{and}$$

$$\Im_t := \sigma\left(X_\tau : \tau \leq t\right).$$

Assume for simplicity ϵ_t has an $\mathbb{R}$-*a.e.* absolutely continuous, positive distribution, symmetric at zero. The $\{X_t\}$ is geometrically α-mixing (An and Huang, 1996, Thm 3.1).

In order to robustify against heavy tails Hill and Renault (2010) propose the generalized method of tail-trimmed moments (GMTTM) estimator. Define one estimating equation,

$$m_t(\theta) = (X_t - \theta X_{t-1}) X_{t-1}.$$

The parameter of interest θ is identified by $\mathrm{E}[m_t(\theta)] = 0$ *if and only if* $\theta = \theta^0$. Denote by $m_{(j)}^{(a)}(\theta)$ the j^{th}-order statistic of $m_t^{(a)}(\theta) := |m_t(\theta)|$, $m_{(1)}^{(a)}(\theta) \geq m_{(2)}^{(a)}(\theta) \geq \cdots$, and assume there exists a sequence of deterministic functions $\{c_n(\theta)\}$ that satisfies, uniformly on compact $\Theta \subset (-1,1)$, $c_n(\theta) \to \infty$ and $(n/k_n)P(|m_t(\theta)| > c_n(\theta)) \to 1$, where $k_n \in \mathbb{N}$, $1 \leq k_n < n$, $k_n \to \infty$, $k_n/n \to 0$.

Now construct deterministically and stochastically trimmed equations $m_{n,t}(\theta) = m_t(\theta) \times I(|m_t(\theta)| \leq c_n(\theta))$ and $\hat{m}_{n,t}(\theta) = m_t(\theta) \times I(|m_t(\theta)| \leq m_{(k_n+1)}^{(a)}(\theta))$. The criterion is

$$\hat{Q}_n(\theta) = \left(\frac{1}{n}\sum_{t=1}^{n} \hat{m}_{n,t}(\theta)\right)^2$$

and the GMTTME is $\hat{\theta}_n = \arg\min_{\theta \in \Theta}\{\hat{Q}_n(\theta)\}$. Intuitively $\hat{\theta}_n$ is that θ that renders the average of nontail equations closest to zero. Since $\mathrm{E}[m_t(\theta^0)] = 0$ by integrability, the negligibly trimmed equations satisfy $\mathrm{E}[m_{n,t}(\theta^0)] \to 0$ by Lebesgue's dominated convergence, so $\hat{\theta}_n \xrightarrow{p} \theta^0$ by standard arguments (Hill and Renault, 2010, Thm. 2.1).

Define

$$v_n^2 = n\mathrm{E}\left[m_{n,t}^2(\theta^0)\right].$$

Symmetric trimming for a symmetric data generating process ensures $\mathrm{E}[m_{n,t}(\theta)] = 0$ *if and only if* $\theta = \theta^0$ for any threshold sequences $\{c_n(\theta^0)\}$. Further, absolute continuity and the linear data generating process imply $\hat{Q}_n(\theta)$ is almost surely twice differentiable at $\hat{\theta}_n$: $(\partial/\partial\theta)\hat{Q}_n(\theta)|_{\hat{\theta}_n} = 0$ a.s. (Čižek, 2008, Lem. 1). Indeed, differentiability of $\hat{Q}_n(\theta)$ along with negligibility of trimming and distribution smoothness imply asymptotic linearity

$$n^{1/2}\frac{\hat{J}_n}{\left(\mathrm{E}\left[m_{n,t}^2(\theta^0)\right]\right)^{1/2}} \times \left(\hat{\theta}_n - \theta^0\right) = \frac{1}{v_n}\sum_{t=1}^{n}\hat{m}_{n,t}(\theta^0)$$

where $\hat{J}_n := 1/n\sum_{t=1}^{n} X_{t-1}^2 I(|m_t(\theta^0)| \leq m_{(k_n+1)}^{(a)}(\theta^0))$. Under the stated assumptions $\hat{J}_n = \mathrm{E}[X_{t-1}^2 I(|m_t(\theta^0)| \leq c_n(\theta^0))] \times (1 + o_p(1))$, and $\mathrm{E}[X_{t-1}^2 I(|m_t(\theta^0)| \leq c_n(\theta^0))]/(\mathrm{E}[m_{n,t}^2(\theta^0)])^{1/2} \to \infty$ if $\kappa < 2$. In the latter infinite variance case super-$\sqrt{n}$-consistency is achieved. See Hill and Renault (2010).

Clearly if $\{\hat{m}_{n,t}(\theta^0)\}$ satisfies a central limit theorem $1/v_n\sum_{t=1}^{n}\hat{m}_{n,t}\left(\theta^0\right) \xrightarrow{d} N(0,1)$ then the GMTTME $\hat{\theta}_n$ is asymptotically normal. Theorem 5.2 contains the required limit theory. As long as $\{m_t\left(\theta^0\right)\}$ is geometrically L_0-APP on a geometrically α-mixing base, $\liminf_{n\to\infty}\{v_n^2/n\} > 0$, and sufficiently many equations $m_t(\theta)$ are trimmed $c_n(\theta^0) = o(v_n^{1/2})$, then $1/v_n\sum_{t=1}^{n}\hat{m}_{n,t}(\theta^0) \xrightarrow{d} N(0,1)$.

Example 13 (L_0-approximable equations). The product convolution $m_t(\theta^0) = \epsilon_t X_{t-1}$ of independent random variables, one i.i.d. and one geometrically α-mixing, is also geometrically α-mixing. Simply define the base as $\epsilon_t X_{t-1}$ and define the base σ-field $\mathcal{F}_t = \sigma(\epsilon_\tau X_{\tau-1} : \tau \leq t)$. The sequence $\{m_t(\theta^0)\}$ is trivially L_0-APP on $\{\mathcal{F}_t\}$ with arbitrary size on a geometrically α-mixing base.

Example 14 (Tail-trimmed QML). Since quasi-maximum likelihood (QML) can be couched in generalized method of moments (GMM), we may similarly couch a robust version of QML in GMTTM simply be redefining the estimating equations. See Hill and Renault (2010) for details and comparisons with existing trimmed M-estimators, and examples concerning GARCH and AR-GARCH estimation.

6. MEASURES OF TAIL AND NONTAIL MEMORY

We complete this paper with an expanded discussion on extant tail memory properties, and a comparison with the concepts and usages considered here. Two final

TABLE 1. Tail dependence for LASV

Property	Coeff.	Value	Interpretation
Extremal index	θ	1	no extremal clustering
Bivariate tail index	η_h	$1/2$	asymptotic independence
Extremogram	$r(h)$	0	uncorrelated tail events
Tail dep. coefficient	$\tilde{r}(h)$	$O(\rho^h)$	locally correlated tail events
Tail copula	$\Lambda_{x,x,h}$	0	uncorrelated tail events
E-NED	ψ_{l_n}	$O(\rho^{l_n})$	nonlinearly dependent tail events

TABLE 2. Tail dependence for GARCH(1,1)

Property	Coeff.	Value	Interpretation
Tail index	κ	$\mathrm{E}\left[\left(\alpha\epsilon_t^2+\beta\right)^{\kappa/2}\right]=1$	$\kappa \uparrow$ implies tail memory $\uparrow$
Extremal index	θ	$\theta(\alpha,\beta)^{\dagger}$	extremal clustering
Bivariate tail index	η_h	1	asymptotic dependence
Extremogram	$r(h)$	$O(\rho^h)$, $\rho \in (0,1)$	correlated tail events
Tail dep. coefficient	$\tilde{r}(h)$	∞	globally correlated tail events
Tail copula	$\Lambda_{x,x,h}$	$O(\rho^h)$	correlated tail events
E-NED	ψ_{l_n}	$O(\rho^{l_n})$, $l_n \to \infty$ as $n \to \infty$.	nonlinearly dependent tail events

† $\theta(\alpha,\beta) := \frac{1}{\mathrm{E}|\epsilon_1|^{\kappa}} \lim_{l\to\infty} \mathrm{E}\left(|\epsilon_1|^{\kappa} - \max_{2\le j\le l+1}\left\{\epsilon_j^2 \prod_{i=2}^{j}\left(\alpha\epsilon_i^2+\beta\right)\right\}^{\kappa/2}\right)_+ .$

unifying examples concerning SV and GARCH are given in Sections 6.2 and 6.3, with details summarized in Tables 1 and 2.

6.1. Tail Memory

Extremal Index. Leadbetter (1974, 1983) shows the maximum of many weakly dependence processes $\{X_t\}$ satisfies

$$\lim_{n\to\infty} P\left(\frac{1}{u_n}\max_{1\le t\le n}|X_t| \le z\right) = e^{-\theta z^{-\kappa}}, \qquad z \ge 0, \qquad \theta \in [0,1]$$

if and only if $nP(|X_t| > u_n) \to 1$ (cf. Loynes, 1965; O'Brien, 1974). Recall that $1/\theta$ approximates the mean number of high threshold exceedances, $\theta = 1$ implies independence, $\theta \in (0,1)$ short-range dependence, and $\theta = 0$ long-range dependence. A stationary AR(1) $X_t = \phi X_{t-1} + \epsilon_t$ with $\phi \in (0,1)$ and i.i.d. Cauchy ϵ_t satisfies $X_t \sim (2)$ with tail index $\kappa = 1$ and extremal index $\theta = 1 - \phi$. Greater geometric memory ($\phi \nearrow 1$) aligns with larger extremal clusters ($1/\theta \nearrow \infty$), irrespective of tail thickness κ. The result generalizes to linear distributed lags $X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ with $\epsilon_t \overset{iid}{\sim}$ (2). See Chernick et al. (1991).

The maximum of a GARCH sample $\{X_t\}_{t=1}^n$ has this property and exhibits power-law tail decay (2) with index $\kappa > 0$ (de Haan et al., 1989; Mikosch and Stărică, 2000; Davis and Mikosch, 2009a). Mikosch and Stărică (Thm. 4.1) characterize θ for GARCH(1,1), shown in Table 1. The formula reveals a tight moment-memory relationship: No GARCH effects ($\alpha_1 = \beta_1 = 0$) are associated with small average threshold exceedance (mean threshold exceedance $1/\kappa \searrow 0$) and no extremal clustering (mean extremal cluster size $1/\theta = 1$); and large α_1 and β_1 are associated with heavier tails ($1/\kappa \nearrow \infty$) and stronger geometric tail memory or larger extreme cluster size ($1/\theta \nearrow \infty$).

Although θ has been characterized for Markov chains, linear distributed lags, GARCH, and SV, θ alone does not contain enough information to support limit theory for tail or tail-trimmed arrays due to insufficient details on memory decay.[5] Asymptotics for estimators of θ require more information, where a mixing condition is a popular route (e.g., Smith and Weissman, 1994). See also de Haan et al. (1989), Chernick et al. (1991), Smith (1992), Hsing (1993), and Davis and Mikosch (2009b).

Tail Mixing. Mixing properties have been used to analyze extremes at least since Loynes (1965). Leadbetter (1974, 1983) and Leadbetter et al. (1983) extend the mixing concept to probability tails of weakly dependent, stationary sequences. The so-called D-mixing property has been used to deliver limit theory for sample maxima and point processes of stationary sequences (e.g., Leadbetter, 1974; de Haan, et al., 1989; Hsing, Hüsler, and Leadbetter, 1989; Hsing, 1993; Stărică, 1999; Leadbetter, Rootzén, and Chaoi, 2001).

Measurable functions of D-mixing random variables are not necessarily D-mixing, and few attempts to characterize stochastic processes as D-mixing exist. See, e.g., Chernick et al. (1991) for stationary moving averages and autoregressions.

Hsing (1991, p.1555) improves the D-mixing construction so that functions of tail-mixing random variables are tail mixing, and Hill (2010) generalizes Hsing's (1991) property to F-mixing to cover arbitrary triangular array functions of $\{X_t\}$, including extreme events $I(X_t > b_n)$ and values $X_t I(X_t > b_n)$, but also nonextreme events $I(X_t \leq b_n)$ and values $X_t I(X_t \leq b_n)$. Neither property requires stationarity; both are implied by α-mixing, hence GARCH class (1) is F-mixing; and as with other mixing properties, they are difficult to verify and have not been verified for hyperbolic memory processes (e.g., FIGARCH, HYGARCH).

Tail Event Correlation and Extremogram. The coefficient $r_n(h) := (P_{h,n} - P_{x,n} P_{y,n})/(P_{x,n} P_{y,n})^{1/2} \sim (n/k_n) \times (P_{h,n} - P_{x,n} P_{y,n})$ defined in Section 5.2 reveals a broad range of tail dependence properties including arbitrarily small deviations from tail independence, accommodates tail memory decay, and is easy to estimate. The extremogram $r(h) = \lim_{n\to\infty} r_n(h)$ characterizes large or "distant" forms of tail dependence, and has been characterized for ARMA and SV processes (Hill 2008a, 2008b; Davis and Mikosch, 2009c), and bounded for GARCH (Davis and Mikosch, 2009c). Hill (2008a, 2008b) shows that a variety of bivariate tails, including those of SV, exhibit small or "local" tail dependence $\tilde{r}(h) := \lim_{n\to\infty}(n/k_n)r_n(h) = \lim_{n\to\infty}\{P_{h,n}/(P_{x,n}P_{y,n}) - 1\} \to 0$ as $h \to \infty$. See Section 6.2.

Like extant tail dependence estimators we must say something more to deliver limit theory for the nonparametric estimator $\hat{r}_n(h) = 1/k_n \sum_{t=1}^{n}(\hat{I}_{x,n,t-h}\hat{I}_{y,n,t} - (k_n/n)^2)$. Either E-NED or E-APP without tail restrictions suffice by Section 5.2 (Hill, 2008b, 2009b) or more restrictive α-mixing with multivariate regular variation (Davis and Mikosch, 2009c).

Hsing (1993, Thm. 2.1) estimates the extremal index θ and imposes two short-range dependence properties for proving consistency of his estimator $\hat{\theta}$, including an intermediate order generalization of Leadbetter's (1983) extreme order D$'$-mixing property, $\sum_{h=1}^{\infty}|r(h)| < \infty$. Hsing (1993, pp. 2049–2050) imposes *finite* dependence to prove asymptotic normality of $\hat{\theta}$ due to substantial technical challenges.

Bivariate Tail Index. Ledford and Tawn (1997, 2003) propose a power-law bivariate tail decay model for tail dependence to improve upon methods dating at least to Loynes (1965). In a simple setting there exists $\eta_h \in [0,1]$ for a joint process $\{X_t, Y_t\}$ with unit Fréchet marginals such that

$$P(X_t > z, Y_{t-h} > z) = z^{-\eta_h} L_h(z) \quad \text{as } z \to \infty,$$

for some slowly varying L_h. All values $\eta_h < 1$ imply $P(X_t > z|Y_{t-h} > z) \to 0$ as $z \to \infty$, hence "asymptotic independence" (cf. Loynes), and if $\eta_h = 1$ then $P(X_t > z|Y_{t-h} > z) \nrightarrow 0$, hence "asymptotic dependence." The above model is argued to be useful for detailing pre-asymptotic dependence since $\eta_h < 1/2$, $\eta_h = 1/2$, and $\eta_h \in (1/2, 1)$ respectively, imply that large values are negatively associated, independent, and positively associated (Ledford and Tawn, 1997). It is, however, easy to prove that values of η_h need not logically coordinate with tail dependence. For example, $\eta_h = 1/2$ can align with tail dependence even though it means doubly "asymptotically independent" and "independence of large values" (Ledford and Tawn, 1997), and negative tail dependence can align with $\eta_h > 1/2$. See Hill (2008a, 2008b) for examples, and see Section 6.2, below. Further, bivariate tail dependence decay as $h \to \infty$ for time series data is rarely modeled by bivariate power-law decay (Ledford and Tawn, 2003; Ramos and Ledford, 2009) or theory is ignored (Ledford and Tawn, 2003), or decay is ignored (Stărică, 1999).

Apparently a characterization of η_h and its decay does not exist for parametric classes of processes like ARFIMA and GARCH. But, tautologically, if Ledford and Tawn's (1997, 2003) model is valid for GARCH then $\eta_h = 1$ with $\lim_{s\to\infty} L_h(z)$ decaying at a geometric rate in h. This aligns with "asymptotic dependence" à la Loynes (1965) and Ledford and Tawn (1997). See Section 6.3.

Further, knowledge of η_h has never been shown to suffice for tail arrays of dependent heterogeneous $\{X_t, Y_t\}$ to belong to any domain of attraction. Ledford and Tawn (2003, p. 534) estimate η_h for stationary sequences but do not characterize a limit distribution, hence a nonparametric block bootstrap is applied for inference. Although they never state memory assumptions nor prove the bootstrap works in the above environment, at least strong mixing is required given the literature cited.

Tail Copula. Copula functions have risen dramatically as a way to characterize dependence in stationary sequences (Joe, 1997). Let $F_x(x)$ and $F_y(y)$ denote the marginal distribution functions of $\{X_t, Y_t\}$, and write the survival probability $\bar{F}_z(z) := 1 - F_z(z)$ with generalized inverse $\bar{F}_z^{-1}(u) := \inf\{|z| \in \mathbb{R} | \bar{F}_z(z) \geq u\}$. Then there exists a unique mapping, the copula, $C : [0,1]^2 \to [0,1]$ satisfying $C[F_x(x), F_y(y)] = P(X_t \leq x, Y_t \leq y)$, cf. Sklar (1959). The survival copula is $\bar{C}[u, v] = P(X_t \geq \bar{F}_x^{-1}(u), Y_t \geq \bar{F}_y^{-1}(v))$, the joint probability X_t and Y_t exceed their u^{th} and v^{th} marginal quantiles.

In its simplest form the right-tail copula $\Lambda_{x,y} : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by

$$\Lambda_{x,y} := \lim_{u \to 0} \frac{1}{u} \bar{C}[u, u].$$

Since independent random variables satisfy $\bar{C}[u,u]/u = u^2/u = u \to 0$ as $u \to 0$, $\Lambda_{x,y}$ near zero implies smaller degrees of tail dependence, and $\Lambda_{x,y} = 0$ is interpreted as tail-independence.

A generalized version of $\Lambda_{x,y}$ permits displacement and arbitrary thresholds. Let $\{b_{x,n}, b_{y,n}\}$ satisfy (6) respectively for $\{X_t, Y_t\}$ with common fractile sequence $\{k_n\}$. Then the right-tailed tail copula $\Lambda_{x,y,h} : \mathbb{R}_+^2 \to \mathbb{R}_+$ over displacement h is (e.g., Schmidt and Stadtmüller, 2006; Klüppleberg, Kuhn, and Peng, 2008)

$$\Lambda_{x,y,h} := \lim_{n \to \infty} \frac{n}{k_n} \times P\left(X_{t-h} > b_{x,n}, Y_t > b_{y,n}\right)$$

and since $(n/k_n)P(X_{t-h} > b_{x,n})P(Y_t > b_{y,n}) \to 0$ trivially

$$\Lambda_{x,y,h} = \lim_{n \to \infty} r_n(h) = r(h).$$

Tail independence is assumed to be captured by $\Lambda_{x,y,h} = 0$, displacement is ignored in this literature ($h = 0$), estimators are only offered for i.i.d. marginals X_t and Y_t, and only stationary joint distributions are considered. See Hill (2008b, 2009b).

Since $\Lambda_{x,y,h} = r(h)$ the tail copula identically depicts Davis and Mikosch's (2009c) extremogram, and Hsing's (1993) short-range dependence property reduces to tail copula summability $\sum_{h=1}^{\infty} |r(h)| = \sum_{h=1}^{\infty} |\Lambda_{x,y,h}| < \infty$. Otherwise explicit usage of $\Lambda_{x,y,h}$ appears to be restricted to applied contexts.

The tail copula $\Lambda_{x,y,h}$ is nearly universally used to measure contemporaneous tail dependence $h = 0$ and is easily estimable nonparametrically (e.g., Schmidt and

Stadtmüller, 2006; Hill 2008b). In a rare instance the decay of $\Lambda_{x,y,h}$ is used to prove consistency of a short-range tail dependence estimator for stationary sequences (Hsing, 1993). E-NED and E-APP, however, are abstractions used to characterize minimal and verifiable tail dependence and heterogeneity properties required to support limit theory for a wide array of tail and nontail estimators like order statistics, tail indices, tail dependence, and tail-trimmed objects.

The differences in usage can be explained by a direct comparison. Assume E-NED size is $1/2$ with displacement $l_n \to \infty$ as $n \to \infty$ and argument $u = 0$, and use Remark 4 and iterated expectations to deduce

$$\text{E-NED} : \lim_{n\to\infty} l_n \sup_{1\le t\le n} \frac{n}{k_n}\left\{P\left(X_t > b_n\right) - \mathrm{E}\left[P\left(X_t > b_n | F_{t-l_n}^{t+l_n}\right)^2\right]\right\} = 0,$$

$$\text{Tail Copula} : \Lambda_{x,x,h} = \lim_{n\to\infty} \frac{n}{k_n}\left\{P\left(X_{t-h} > b_n, X_t > b_n\right) - P\left(X_{t-h} > b_n\right)^2\right\}.$$

E-NED is a prediction error statement of *conditional* tail memory decay: As information amasses $l_n \to \infty$ the minimum mean-squared-error predictor $P(X_t > b_n e^u | F_{t-l_n}^{t+l_n})$ converges to $I(X_t > b_n e^u)$ faster than the rate at which $I(X_t > b_n e^u)$ degenerates in L_2-norm (i.e., k_n/n). The tail copula, however, measures the asymptotic discrepancy between *joint* and *marginal* tail probabilities at a fixed displacement h.

We now give two examples demonstrating the information content of various measures of tail dependence for two random volatility models.

6.2. Tail Dependence in Log-Autoregressive Stochastic Volatility

Consider a univariate log-autoregressive stochastic volatility (LASV) model

$$y_t = \sigma_t \epsilon_t \qquad \text{where } \ln \sigma_t^\kappa = \theta + \phi \ln \sigma_{t-1}^\kappa + \zeta_t, \quad \text{and} \quad |\phi| < 1,$$

where $\epsilon_t \overset{iid}{\sim} (2)$ with index κ and $\zeta_t \overset{iid}{\sim} N(0,1)$, as in Example 11. Let X_t be $|y_t|$, $-y_t I(y_t < 0)$, or $y_t I(y_t > 0)$. It is easy to show $X_t \sim (2)$ with index κ (Hill, 2008a, 2008b; Davis and Mikosch, 2009b, 2009c).

Hill (2008a) shows the bivariate tail index $\eta_h = 1/2$, which Ledford and Tawn (1997) interpret as independence asymptotically and for large values. Similarly the extremogram $r(h) = 0$ (Davis and Mikosch, 2009b, 2009c) and copula $\Lambda_{x,x,h} = 0$ (Hill, 2008a, 2008b) for all $h \ge 1$, and extremal index $\theta = 1$ (Davis and Mikosch, 2009b) each reflect a lack of tail dependence.

Yet X_t exhibits local tail dependence at a geometric rate (Hill, 2008a):

$$\tilde{r}(h) = \lim_{n\to\infty}\left(\frac{n}{k_n}\right) r_n(h) = \lim_{n\to\infty} \frac{P\left(X_{t-h} > b_{x,n}, X_t > b_{x,n}\right)}{P\left(X_t > b_{x,n}\right) P\left(X_{t-h} > b_{x,n}\right)} - 1 = K\phi^h.$$

Further, $\{X_t\}$ is geometrically β-mixing (Carrasco and Chen, 2002) and L_2-E-NED on $\{\epsilon_t, \zeta_t\}$ with coefficients that depend on ϕ^h (Hill, 2008a). Describing tail dependence by $\tilde{r}(h)$ and E-NED better captures the tail memory characteristics of LASV data. The use of η_h, $r(h)$, $\Lambda_{x,x,h}$ or θ may be misleading, and typically does not provide enough information to push through limit theory for tail or nontail estimators. That $\theta = 1$, $\eta_h = 1/2$, and $\Lambda_{x,x,h} = r(h) = 0$ need not tell us anything about whether estimators of κ, θ, η_h, $\Lambda_{x,x,h}$, or $r_n(h)$ are consistent or asymptotically normal, while the E-NED property supports asymptotics at least for Hill's (1975) $\hat{\kappa}_n$ and the nonparametric $\hat{r}_n(h)$ (Hill 2008b, 2009b, 2010). Since estimators of θ are popularly based on intermediate order statistics as in Hsing (1993) and Smith and Weissman (1994), it is likely that they too are asymptotically normal under E-NED, cf. Theorem 5.1. See Table 1 for a summary of known LASV tail dependence properties.

6.3. Tail Dependence in GARCH(1,1)

Consider a strong-GARCH(1,1):

$$y_t = \sigma_t \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} (0,1), \qquad \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \qquad \omega > 0, \alpha, \beta \in [0,1).$$

Then $\sigma_t^2 = \pi_0 + \sum_{i=1}^{\infty} \alpha \beta^{i-1} y_{t-i}^2$ satisfies $0 \leq \alpha\beta^{i-1} \leq C\rho^i$ where $\rho \in (0,1)$ and $C \in [0, 1/\rho)$, covering covariance stationary, integrated, and explosive cases. Let X_t be $|y_t|$, $-y_t I(y_t < 0)$, or $y_t I(y_t > 0)\}$. Tail dependence properties are detailed in Table 2.

In summary, κ and θ provide ample tail dependence details but alone do not dictate if tail or nontail estimators converge in any sense. The extremogram $r(h)$ and tail copula $\Lambda_{x,x,h}$ reflect distant geometric tail dependence, hence Hsing's (1993) estimator is consistent $\hat{\theta} \overset{p}{\to} \theta$ under additional regulatory conditions, but apparently no other asymptotic theory has been deduced from summability of $r(h)$ or $\Lambda_{x,x,h}$.

The extremogram can be used to deduce a bound on the bivariate tail index, if it exists, á la Ledford and Tawn (1997, 2003). Davis and Mikosch (2009c) show the above GARCH process satisfies for $\rho \in (0,1)$

$$r(h) = \lim_{n \to n} \frac{P\left(x_{t-h} > b_{x,n}, y_t > b_{y,n}\right) - P\left(x_t > b_{x,n}\right) P\left(y_t > b_{y,n}\right)}{\left[P\left(x_t > b_{x,n}\right) P\left(y_t > b_{y,n}\right)\right]^{1/2}} = O(\rho^h),$$

hence $\lim_{n\to n} P(x_{t-h} > b_{x,n}, y_t > b_{y,n})/[P(x_t > b_{x,n})P(y_t > b_{y,n})]^{1/2} = O(\rho^h)$. Let $\{x_t^*, y_t^*\}$ be unit Fréchet transforms of $\{x_t, y_t\}$. Assuming Ledford and Tawn's (1997, 2003) model is valid, use the fact that x_t^* is a monotonic transform and $P(x_t^* > z) = 1 - \exp\{-1/z\} = z^{-1} \times (1 + o(1))$ to deduce

$$r(h) = \lim_{n \to n} \frac{P\left(x_{t-h}^* > z, x_t^* > z\right)}{P\left(x_{t-h}^* > z\right) P\left(x_t^* > z\right)} = \lim_{n \to n} z^{2-1/\eta_h} L_h(z) \times (1 + o(1))$$

if and only if $\eta_h = 1$ and $\lim_{z \to \infty} L_h(z) = r(h) = O(\rho^h)$ since $L_h(z)$ is slowly varying (Resnick, 1987). Therefore GARCH is asymptotically dependent (Loynes, 1965; Ledford and Tawn, 1997).

Finally, geometric α-mixing or E-NED ensures that Hill's (1975) $\hat{\kappa}_n$ and the nonparametric $\hat{r}_n(h)$ are consistent and asymptotically normal (Hill 2008b, 2010; Davis and Mikosch, 2009c). Use the identity $\Lambda_{x,x,h} = \lim_{n\to\infty} r_n(h)$ to deduce that a nonparametric copula estimator $\hat{\Lambda}_{x,x,h} = \hat{r}_n(h)$ is also covered (e.g., Schmidt and Stadtmüller, 2006; Hill, 2008b, 2009b).

7. SUMMARY

Tail memory properties are predominantly aimed at applied researchers: The goal here is to *characterize* tail dependence (e.g., θ, η_h, $\Lambda_{x,y,h}$, $r(h)$, and $\tilde{r}(h)$). Very few properties are consequently applicable for asymptotic theory associated with extreme value estimators since they have not been, or cannot be, shown to reveal enough information for general central and weak limit theory. Further, in general there do not exist nontail memory notions for either empirical characterizations or asymptotic theory.

By comparison, near epoch dependence and L_0-approximability can be straightforwardly applied solely to tails or nontails. The extensions are not vacuous since very heavy-tailed random volatility data may have NED extremes and nonextremes even if population NED is unknown. These tail and nontail memory properties cover nonstationary and hyperbolic memory data; they link X_t to some "base" ϵ_t within a prediction premise, so multivariate dependence is allowed; and the notion of "size" permits easy characterizations of tail and nontail memory decay. Further, since memory decay can be concisely depicted, our notions of dependence suffice for extreme value and robust central limit theory.

Extremal-NED and APP provide substantial generality beyond conventional mixing conditions since little is known about mixing in processes with hyperbolic memory, and ensuring mixing in short memory data typically requires restrictions on error distribution smoothness. The results of this paper permit a series of useful limit theory extensions to extremal and nonextremal processes based on parametric classes (e.g., nonlinear ARFIMA and explosive GARCH), and a new path for delivering robust asymptotic theory with applications to new robust estimators.

NOTES

1. The Lyapunov exponent γ is associated with the first-order difference equation form of $Z_t := [X_t^2, \ldots, X_{t-p+2}^2; \sigma_{t+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+2}^2]'$. It is easy to show Z_t is a firstorder Markov $Z_t = A_t Z_{t-1} + B_t$ for some independent and identically distributed (i.i.d.) sequences $\{A_t, B_t\}$ of $k \times k$ matrices A_t and k-vectors B_t, $k \geq 1$. Identically $\gamma = \lim_{n\to\infty} n^{-1} \ln || \sqcap_{t=1}^n A_t ||_o$, where $||A||_o = \sup_{x\in\mathbb{R}^k, |x|=1} |Ax|$. If ϵ_t in (1) is i.i.d. with zero mean and unit variance, then $\gamma < 0$ given the remaining properties. See Basrak, Davis, and Mikgsch (2002a); cf. Bougerol and Picard (1992).

2. By comparison, the tail index κ for GARCH(1,1) satisfies (3) and can therefore be computed by Monte-Carlo (Mikosch and Stărică, 2000; Basrak et al., 2002b).

3. Clearly this abuses notation since therefore the limit $\tilde{r}(h) := \lim_{n\to\infty}(n/k_n) r_n(h)$ does not exist.

4. See Stigler (1973) for historical details, and Hill (2009a) for a recent survey.

5. It is important to distinguish between domains of attraction (Ibragimov and Linnik, 1971; Leadbetter et al., 1983; Resnick, 1987). The index θ itself is a limiting distribution characteristic for weakly dependent, stationary sequences $\{X_t\}$ in the *maximum* domain of attraction: $\lim_{n\to\infty} P(u_n^{-1} \max_{1\leq t\leq n} |X_t| \leq z) = e^{-\theta z^{-\kappa}}$ (Leadbetter, 1983). But knowing θ does not provide sufficient information to conclude, e.g., the tail array $\{I(X_t > b_{k_n} e^u)\}$ or nontail array $\{X_t I(X_t \leq b_{k_n} e^u)\}$ belongs to the domain of attraction of a normal law. Further, the long-range dependence case $\theta = 0$ clearly provides insufficient knowledge since it says nothing about long memory decay.

REFERENCES

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

Baillie, R.T., T. Bollerslev, & H.O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* 74, 3–30.

Basrak., R.A. Davis, & T. Mikosch (2002a) A characterization of multivariate regular variation, *Annals of Applied Probability* 12, 908–920.

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

Beirlant, J., P. Vynckier, & J.L. Teugels (1996) *Practical Analysis of Extreme Values*. Leuven University Press.

Bierens, H.J. (1987) ARMAX model specification testing, with an application to unemployment in the Netherlands. *Journal of Econometrics* 35, 161–190.

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

Borkovec, M. & C. Klüppelberg (2001) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors, *Annals of Applied Probability* 11, 1220–1241.

Bougerol, P. & N. Picard (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

Boussama, F. (1998) Ergodicité, Mélange et Estimation dans le Modelès GARCH. Ph.D. thesis, Université 7 Paris.

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

Chernick, M.R. (1981) A limit theorem for the maximum of autoregressive processes with uniform marginal distribution. *Annals of Probability* 9, 145–149.

Chernick, M.R., T. Hsing, & W.P. McCormick (1991) Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability* 23, 835–850.

Čižek, P. (2008) General trimmed estimation: Robust approach to nonlinear and limited dependent variable models, *Econometric Theory* 24, 1500–1529.

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

Cline, D.B.H. (2007) Regular variation of order 1 nonlinear AR-ARCH models, *Stochastic Processes and Their Applications* 117, 840–861.

Cline, D.B.H. & H-m. H. Pu (2004) Stability and the Lyapunov exponent of threshold AR-ARCH models. *Annals of Applied Probability* 14, 1920–1949.

Csörgő, S., L. Horváhth, & D.M. Mason (1986) What portion of the sample makes a partial sum asymptotically stable or normal. *Probability Theory and Related Fields* 72, 1–16.

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

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

Davidson, J. (2004) Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economics Statistics* 22, 16–29.

Davis, R.A. & T. Mikosch (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* 26, 2049–2080.

Davis, R.A. & T. Mikosch (2009a) Extreme value theory for GARCH processes. In T.G. Andersen, R.A. Davis, J.-P. Kreiss, & T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 187–200. Springer.

Davis, R.A. & T. Mikosch (2009b) Extremes of stochastic volatility models. In T.G. Andersen, R.A. Davis, J.-P. Kreiss, & T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 355–364. Springer.

Davis, R.A. & T. Mikosch (2009c) The extremogram: A correlogram for extreme events. *Bernoulli* 15, 977–1009.

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

Davison, A.C. & R.L. Smith (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series* B 52, 393–442.

de Haan, L., S.I. Resnick, H. Rootzén, & C.G. de Vries (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and Their Applications* 32, 213–224.

de Jong, R.M. (1997) Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353–367.

de la Peña, V.H., R. Ibragimov, & S. Sharakhmetov (2003) On extremal distributions and sharp Lp-bounds for sums of multilinear forms. *Annals of Probability* 31, 630–675.

Doukhan, P. (1994) Mixing: Properties and Examples. Lecture Notes in Statistics 85. Springer.

Doukhan, P. & S. Louhichi (1999) A new weak dependence condittion and applications to moment inequalities. *Stochastic Processes and Their Applications* 84, 313–342.

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

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

Engle, R. & V. Ng (1993). Measuring and testing the impact of news on volablity. *Journal of Finance* 48, 1749–1778.

Feller, W. (1946) A limit theorem for random variables with infinite moments. *American Journal of Mathematics* 68, 257–262.

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

Gabaix, X. (2008) Power laws. In S.N. Durlauf & L.E. Blume (eds.), *The New Palgrave Dictionary of Economics,* 2nd ed. Palgrave Macmillan.

Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics.* Kreiger.

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

Giraitis, L., P. Kokoszka, & R. Leipus (2000) Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.

González-Rivera, G. (1998) Smooth-Transition GARCH models, *Studies in Non-Linear Dynamics and Econometrics* 3, 61–78.

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

Haeusler, E. & J.L. Teugels (1985) On asymptotic normality of Hill's estimator for the exponent of regular variation. *Annals of Statistics* 13, 743–756.

Hahn, M.G., J. Kuelbs, & J.D. Samur (1987) Asymptotic normality of trimmed sums of mixing random variables. *Annals of Probability* 15, 1395–1418.

Hahn, M.G., J. Kuelbs, & D.C. Weiner (1990) The asymptotic joint distribution of self-normalized censored sums and sums of squares. *Annals of Probability* 18, 1284–1341.

Hahn, M.G. & D.C. Weiner (1992) Asymptotic behavior of self-normalized trimmed sums: Nonnormal limits. *Annals of Probability* 20, 455–482.

Hall, P. (1982) On some estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B* 44, 37–42.

Hall, P. & Q, Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *econometrica* 71, 285–317.

He, X., J. Jurecková, R. Koenker, & S. Portnoy (1990) Tail behavior of regression estimators and their breakdown points. *Econometrica* 58, 1195–1214.

Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Mathematical Statistics* 3, 1163–1174.

Hill, J.B. (2008) Robust Estimation and Inference for Extremal Dependence in Time Series. University of North Carolina at Chapel Hill.

Hill, J.B. (2009a). Central Limit Theory for Kernel-Self Normalized Tail-Trimmed Sums of Dependent Data with Applications. Working paper, University of North Carolina at Chapel Hill.

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

Hill, J.B. (2010) On tail index estimation for dependent, heterogeneous data. *Econometric Theory* 26, 1398–1436.

Hill, J.B. (2011) Extremal memory of stochastic volatility with an application to tail shape inference. *Journal of Statistical Planning and Inference* 141, 663–676.

Hill, J.B. & E. Renault (2010) Generalized Method of Moments with Tail Trimming. Working paper, University of North Carolina at Chapel Hill.

Hill, J.B. & A. Shneyerov (2010) Are There Common Values in First-Price Auctions? A Tail-Index Nonparametric Test. Working paper, University of North Carolina at Chapel Hill.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

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

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

Hsing, T., J. Hüsler, & M.R. Leadbetter (1989) On the exceedance point process for a stationary Sequence. *Probability Theory and Related Fields* 78, 97–112.

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

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

Ibragimov, R. (2009) Heavy-Tailed densities. In S.N. Durlauf & L.E. Blume (eds.), *The New Palgrave Dictionary of Economics Online.* Palgrave Macmillan.

Iglesias, E.M. & O. Linton (2009) Estimation of Tail Thickness Parameters from GJR-GARCH Models. Mimeo, London School of Economics.

Joe, H. (1997) *Multivariate Models and Dependence Concepts. Monograms in Statistics and Applied Probability* 73, Chapman and Hall.

Jurečková, J. (1981) Tail behavior of location estimators. *Annals of Statistics* 9, 578–585.

Klüppelberg, C., G. Kuhn, & L. Peng (2008) Semi-Parametric models for the multivariate tail dependence function—the asymptotically dependent case. *Scandinavian Journal of Statistics* 35, 701–718.

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

Leadbetter, M.R. (1983) Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete* 65, 291–306.

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

Leadbetter, M.R., H. Rootzén, & H. Choi (2001) On central limit theory for random additive functions under weak dependence restrictions. *Lecture Notes-Monograph Series* 36, 464–476.

Ledford, A.W. & J.A. Tawn (1997) Modeling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B* 59, 475–499.

Ledford, A.W. & J.A. Tawn (2003) Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society, Series B* 65, 521–543.

Ling, S. (1999) On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied Probability* 36, 688–705.

Linton, O., J. Pan, & H. Wang (2010) Estimation for a nonstationary semi-strong GARCH(1,1) model with heavy-tailed errors. *Econometric Theory* 26, 1–28.

Longin, F. & B. Solnik (2001) Extreme correlation of international equity markets. *Journal of Finance* 56, 649–676.

Loynes, R.M. (1965) Extreme values in uniformly mixing stationary stochastic processes. *Annals of Mathematical Statistics* 36, 993–999.

McLeish, D.L. (1975) A maximal inequality and dependent strong law. *Annals of Probability* 3, 329–339.

Meitz M. & P. Saikkonen (2008) Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis* 29, 453–475.

Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28, 1427–1451.

Nagev, S.V. (1979) Large deviations of sums of independent random variables. *Annals of Probability* 7, 745–789.

Nagev, S.V. (1998) Some refinements of probabilistic and moment inequalities. *Theory of Probability and Their Applications* 42, 707–713.

Naveau, P. (2003). Almost sure relative stability of the maximum of a stationary sequence. *Advances in Applied Probability* 35, 721–736.

Nze, P.A., P. Buhlmann, & P. Doukhan (2002) Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics* 30, 397–430.

Nze, P.A. & P. Doukhan (2004) Weak dependence: Models and applications to econometrics. *Econometric Theory* 20, 995–1045.

O'Brien, G.L. (1974) The maximum term of uniformly mixing stationary sequences. *Zeitschrift für Wahrscheinlichkeit Theorie und Verwandte Gebeite* 30, 57–63.

Pötscher, B.M. & I.R. Prucha (1991) Basic structure of the asymptotic theory in dynamic nonlinear econometrics models, Part I: Consistency and approximation concepts. *Econometric Reviews* 10, 125–216.

Pruitt W. (1985) Sums of independent random variables with the extreme terms excluded. In J.N. Srivastava (ed.), *Probability and Statistics: Essays in Honor of Franklin A. Graybill.* Elsevier.

Ramos, A. & A. Ledford (2009) A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society, Series B* 71, 219–241.

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

Rootzén, H. (1978) Extremes of moving averages of stable processes. *Annals of Probability* 6, 847–869.

Rootzén, H. (2008) Weak convergence to the tail empirical function for dependent sequences. *Stochastic Processes and their Applications* 119, 468–490.

Schmidt R. & U. Stadtmüller (2006) Non-Parametric estimation of tail dependence. *Scandinavian Journal of Statistics* 33, 67–335.

Sklar, A. (1959) Fonctions de répartitions á n dimensions et leurs marges. *Publications de L'Institut de Statistique de L'Université de Paris* 8, 229–231.

Smith, R. (1984) Threshold methods for sample extremes. In J. Tiago de Oliveira (ed.), *Statistical Extremes and Applications,* pp. 621–638. Reidel.

Smith, R.L. (1992) The extremal index for a markov chain. *Journal of Applied Probability* 29, 37–45.

Smith, R.L. & I. Weissman (1994) Estimating the extremal index. *Journal of the Royal Statistical Society, Series B* 56, 515–528.

Stărică, C. (1999) Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance* 6, 515–553.
Stigler, S.M. (1973) The asymptotic distribution of the trimmed mean. *Annals of Statistics* 1, 472–477.
Tong, H. & K. S. Lim (1980) Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B* 42, 245–292.
Wu, W.B. (2005) On the Badahur representation of sample quantiles for dependent sequences. *Annals of Statistics* 15, 20–36.
Wu, W.B. & M. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and Their Applications* 115, 939–958.

APPENDIX: Proofs

We list for future reference several primitive results concerning moment and probability bounds. We only prove nontrivial assertions. Here $\{U_t, X_t, Y_t, Z_t\}$ are arbitrary random variables on some measure space; $\{A_{n,t}, B_{n,t}, C_{n,t}\}$ are positive, deterministic triangular arrays; and R is an arbitrary subset of $\mathbb{R}$ with positive Lebesgue measure. Define

$$I_{x,t}(A_{n,t}) := I\left(X_t \leq A_{n,t}\right).$$

LEMMA A.1. *For any X_t in $L_2(\Omega, \mathcal{F}, P)$, any σ-field $\Im \subseteq \mathcal{F}$, and all $\Im$-measurable random variables Y_t,*

$$\mathrm{E}\left(X_t - \mathrm{E}[X_t|\Im]\right)^2 \leq \mathrm{E}\left(X_t - Y_t\right)^2. \tag{A.1}$$

LEMMA A.2. *(i.) If U_t is uniformly almost surely bounded then, for any $q > 0$,*

$$\mathrm{E}|U_t|^q \leq \mathrm{E}\left[|U_t|^q I\left(Z_t \in R\right)\right] + K \times P\left(Z_t \notin R\right). \tag{A.2}$$

(ii) If U_t is $L_{r\times q}$-bounded for $r > 1$ and $q > 0$ then

$$\mathrm{E}|U_t|^q \leq \mathrm{E}\left[|U_t|^q I\left(Z_t \in R\right)\right] + \|U_t\|^q_{r\times q} \times P\left(Z_t \notin R\right)^{(r-1)/r}. \tag{A.2$'$}$$

LEMMA A.3. *For any $q > 0$*

$$\begin{aligned}\mathrm{E}\Big[\ \left|I_{x,t}(A_{n,t}) - I_{y,t}(A_{n,t})\right|^q \times I\left(|X_t - Y_t| \leq B_{n,t}\right)\ \Big]&\\ \leq 2P\left(A_{n,t} - B_{n,t} \leq X_t \leq A_{n,t} + B_{n,t}\right).&\end{aligned} \tag{A.3}$$

LEMMA A.4. *Let $\{X_t\}$ be L_p-bounded, $p > 0$, and define $\bar{F}_t(x) := P(X_t > x)$. For all arrays $\{A_{n,t}, C_{n,t}\}$ satisfying* $\inf_{t\in\mathbb{Z}}\{A_{n,t}\} \to \infty$ *and* $\sup_{n\geq 1}\sup_{1\leq t\leq n}\{C_{n,t}\} \in (0,1)$, *there exists a triangular array $\{r_{n,t}\}$ satisfying* $\liminf_{n\geq 1}\inf_{1\leq t\leq n}\{r_{n,t}\} \geq 0$ *and $r_{n,t} = o(1)$ for each $1 \leq t \leq n$, such that*

$$\bar{F}_t(A_{n,t}) = A_{n,t}^{-p} \times r_{n,t} \quad \text{and} \quad \bar{F}_t\left(A_{n,t}(1 \pm C_{n,t})\right) = K \times A_{n,t}^{-p} \times C_{n,t} \times r_{n,t}. \tag{A.4}$$

LEMMA A.5. *For any $\{B_{n,t}\}$ almost surely*

$$I\left(\left|I_{x,t}(A_{n,t}) - I_{y,t}(A_{n,t})\right| > B_{n,t}\right) \leq \left|I_{x,t}(A_{n,t}) - I_{y,t}(A_{n,t})\right|. \tag{A.5}$$

LEMMA A.6. *If* $\limsup_{n\geq 1}\sup_{1\leq t\leq n}\{B_{n,t}\}\in(0,1)$ *then*

$$\mathrm{E}\left[\left|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})\right|^{q} I\left(\left|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})\right|\leq B_{n,t}\right)\right]=0. \tag{A.6}$$

Proof of Lemma A.2. Since $\mathrm{E}|U_t|^q=\mathrm{E}[|U_t|^q I(Z_t\in R)]+\mathrm{E}[|U_t|^q I(Z_t\notin R)]$ claim (i) follows from boundedness $\mathrm{E}[|U_t|^q I(Z_t\notin R)]\leq K\times\mathrm{E}[I(Z_t\notin R)]=K\times P(Z_t\notin R)$, and (ii) from Hölder's inequality. ■

Proof of Lemma A.3. Note $|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})|^q\times I(|X_t-Y_t|\leq B_{n,t})=1$ if $-B_{n,t}\leq X_t-Y_t\leq B_{n,t}$, and $X_t\leq A_{n,t}<Y_t$ or $Y_t\leq A_{n,t}<X_t$; and $|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})|^q\times I(|X_t-Y_t|\leq B_{n,t})=0$ otherwise. Now use $A_{n,t},B_{n,t}>0$ to deduce

$$\begin{aligned}
&\mathrm{E}\left[\left|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})\right|^{q}\times I\left(|X_t-Y_t|\leq B_{n,t}\right)\right]\\
&\quad=P\left(-B_{n,t}\leq X_t-Y_t\leq B_{n,t}\cap X_t\leq A_{n,t}\cap A_{n,t}<Y_t\right)\\
&\qquad+P\left(-B_{n,t}\leq X_t-Y_t\leq B_{n,t}\cap Y_t\leq A_{n,t}\cap A_{n,t}<X_t\right)\\
&\quad\leq P\left(A_{n,t}<Y_t\cap Y_t-B_{n,t}\leq X_t\leq A_{n,t}\right)\\
&\qquad+P\left(A_{n,t}<X_t\leq B_{n,t}+Y_t\cap Y_t\leq A_{n,t}\right)\\
&\quad\leq P\left(A_{n,t}-B_{n,t}\leq X_t\leq A_{n,t}+B_{n,t}\right)\\
&\qquad+P\left(A_{n,t}-B_{n,t}<X_t\leq A_{n,t}+B_{n,t}\right).
\end{aligned}$$

■

Proof of Lemma A.4. Apply the L_p-boundedness implications (9), $\lim_{x\to\infty}x^p\bar{F}_t(x)=0\;\forall p<\kappa$, and $\inf_{t\in\mathbb{Z}}\{A_{n,t}\}\to\infty$, to deduce $\bar{F}_t(A_{n,t})=A_{n,t}^{-p}\times r_{n,t}$ where $r_{n,t}=o(1)$ for each $1\leq t\leq n$. Since $C_{n,t}>0$ and $\sup_{n\geq 1}\sup_{1\leq t\leq n}\{C_{n,t}\}<1$ clearly $A_{n,t}(1\pm C_{n,t})\to\infty$. Now apply the first claim and the mean-value theorem to obtain $\bar{F}_t(A_{n,t}(1\pm C_{n,t})=A_{n,t}^{-p}(1\pm C_{n,t})^{-p}r_{n,t}=KA_{n,t}^{-p}C_{n,t}r_{n,t}$. ■

Proof of Lemma A.6. Since uniformly $B_{n,t}\in(0,1)$ it follows $|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})|\leq B_{n,t}$ only if $I(X_t\leq A_{n,t})=I(Y_t\leq A_{n,t})=0$ *or* 1, in which cases $|I_{x,t}(A_{n,t})-I_{y,t}(A_{n,t})|=0$ a.s. ■

Proof of Theorem 2.1.

Claim (i): Recall $\bar{I}_{n,t}(u):=I(X_t>b_n e^u)$ and define $\eta_n:=b_n e^u\vartheta_{l_n}^{1/2}$. For any $q\geq 2$ and some uniformly positive triangular array $\{r_{n,t}\}$

$$\begin{aligned}
&\mathrm{E}\left|\bar{I}_{n,t}(u)-\mathrm{E}\left[\bar{I}_{n,t}(u)|F_{t-l_n}^{t+l_n}\right]\right|^{q}\leq\mathrm{E}\left(\bar{I}_{n,t}(u)-\mathrm{E}\left[\bar{I}_{n,t}(u)|F_{t-l_n}^{t+l_n}\right]\right)^{2}\\
&\quad\leq 2\times P\left(b_n e^u-\eta_n\leq X_t\leq b_n e^u+\eta_n\right)+K\times P\left(\left|X_t-\mathrm{E}\left[X_t|F_{t-l_n}^{t+l_n}\right]\right|>\eta_n\right)\\
&\quad\leq Kr_{n,t}b_n^{-p}e^{-pu}\vartheta_{l_n}^{1/2}+K\times\mathrm{E}\left(\left|X_t-\mathrm{E}\left[X_t|F_{t-l_n}^{t+l_n}\right]\right|^{p}\right)\times\eta_n^{-p}\\
&\quad\leq Kr_{n,t}b_n^{-p}e^{-pu}\vartheta_{l_n}^{1/2}+Kd_t^{p}\vartheta_{l_n}^{p}\times\left\{e^{-pu}b_n^{-p}\vartheta_{l_n}^{-p/2}\right\}\\
&\quad\leq K\left(1+r_{n,t}\right)b_n^{-p}e^{-pu}\max_{1\leq t\leq n}\left\{d_t^{p}\right\}\vartheta_{l_n}^{\min\{p,1\}/2}.
\end{aligned}$$

The first inequality follows from $\bar{I}_{n,t}(u) - \mathrm{E}[\bar{I}_{n,t}(u)|F_{t-l_n}^{t+l_n}] \in [-1,1]$ and $q \geq 2$; the second from (A.1), (A.2) and (A.3); the third from (A.4) given $\sup_{l\geq 1}\vartheta_l \in [0,1)$, and Markov's inequality, where $r_{n,t} = o(1)$ for each $1 \leq t \leq n$; and the fourth is L_p-NED. Therefore $||\bar{I}_{n,t}(u) - \mathrm{E}[\bar{I}_{n,t}(u)|F_{t-l_n}^{t+l_n}]||_q$ is bounded by

$$\left\{\left(\frac{k_n}{n}\right)^{1/q} e^{-up/q}\right\} \times \left[K\left(1+r_{n,t}\right)\left(\frac{n}{k_n}\right)^{1/q} l_n^{-\iota} b_n^{-p/q} \max_{1\leq t\leq n}\left\{d_t^{p/q}\right\}\right] l_n^{\iota}\vartheta_{l_n}^{\min\{p,1\}/(2q)}$$

$$= d_{n,t}(u) \times \psi_{l_n},$$

say, for arbitrarily tiny $\iota > 0$. Clearly $\sup_{1\leq t\leq n}\{d_{n,t}(u)\} = K(k_n/n)^{1/q}e^{-up/q}$ is Lebesgue integrable on $\mathbb{R}_+$ and $\sup_{1\leq t\leq n}\sup_{u\geq 0}\{d_{n,t}(u)\} = O((k_n/n)^{1/q})$. Finally, we can always choose $l_n \to \infty$ sufficiently fast such that $(K + r_{n,t})(n/k_n)^{1/q} l_n^{-\iota} b_n^{-p/q} \max_{1\leq t\leq n}\{d_t^{p/q}\} = O(1)$, and for $\iota > 0$ sufficiently small $l_n^{\iota}\vartheta_{l_n}^{\min\{p,1\}/(2q)} = o(l_n^{-\lambda\min\{p,1\}/(2q)})$ by continuity and $\vartheta_{l_n} = o(l_n^{-\lambda})$.

Claim (ii): The assertion follows from boundedness $|\bar{I}_{n,t}(u) - \mathrm{E}[\bar{I}_{n,t}(u)|F_{t-l_n}^{t+l_n}]| \leq 1$ and Lyapunov's inequality. ■

Proof of Lemma 2.3. Let $\{\eta_n\}$ be any positive sequence that satisfies $\eta_n \geq l_n^{\lambda+\iota}$, where $\lambda > 0$ is the L_0-APP size. For some sequence $\{f_t\}$, $\inf_{t\in\mathbb{Z}} f_t = f > 0$, and constant $\delta > 0$ let $\{f_{n,t}\}$ be any triangular array and $\{\delta_n\}$ any sequence that satisfies

$$f_{n,t}\delta_n = b_n e^u \eta_n l_n^{-\lambda-\iota} f_t\delta \geq b_n e^u f_t\delta \geq e^u f_t\delta \geq f_t\delta > 0.$$

Define $h_t^{(l_n)}(u) := I(g_t^{(l_n)} > b_n e^u)$. It follows for tiny $\iota > 0$

$$P\left(\left|I\left(X_t > b_n e^u\right) - h_t^{(l_n)}(u)\right| > f_{n,t}\delta_n\right)$$

$$\leq \mathrm{E}\left[I\left(\left|I\left(X_t > b_n e^u\right) - h_t^{(l_n)}(u)\right| > b_n e^u l_n^{-\lambda-\iota} f\delta\right) \times I\left(\left|X_t - g_t^{(l_n)}\right| \leq b_n e^u l_n^{-\lambda-\iota} f\delta\right)\right]$$

$$+ \left\|I\left(\left|I\left(X_t > b_n e^u\right) - h_t^{(l_n)}(u)\right| > e^u f\delta\right)\right\|_{1/\iota} \times P\left(\left|X_t - g_t^{(l_n)}\right| > f_t\delta\right)^{1-\iota}$$

$$\leq \mathrm{E}\left[I\left(\left|I\left(X_t > b_n e^u\right) - h_t^{(l_n)}(u)\right| > b_n e^u l_n^{-\lambda-\iota} f\delta\right) \times I\left(\left|X_t - g_t^{(l_n)}\right| \leq b_n e^u l_n^{-\lambda-\iota} f\delta\right)\right] + Ke^{-\iota u} \times P\left(\left|X_t - g_t^{(l_n)}\right| > f_t\delta\right)^{1-\iota}$$

$$\leq K \times P\left(b_n e^u - b_n e^u l_n^{-\lambda-\iota} f\delta < X_t < b_n e^u + b_n e^u l_n^{-\lambda-\iota} f\delta\right) + Ke^{-\iota u} \times v_{l_n}^{1-\iota}$$

$$\leq K b_n^{-p} e^{-pu} l_n^{-\lambda-\iota} + Ke^{-\iota u} v_{l_n}^{1-\iota}$$

$$= K\left\{\frac{k_n}{n}e^{-\iota u} \times \left(b_n^{-p}+1\right) \times \frac{n}{k_n} l_n^{-\iota}\right\} \times \left\{l_n^{\iota} \times o\left(l_n^{-\lambda}\right)\right\}$$

$$\leq K \times e_{n,t}(u) \times \varphi_{l_n}.$$

The first inequality follows from $f_{n,t}\delta_n \geq b_n e^u l_n^{-\lambda-\iota} f\delta$ and $f_{n,t}\delta_n \geq e^u f_t\delta \geq e^u f\delta$, and property (A.2′). The second exploits $\{f,\delta\} > 0$ and Markov's inequality. The third follows from (A.5), then (A.3), and L_0-APP. The fourth follows from (A.4). Simply choose $l_n \to \infty$ sufficiently fast to ensure $e_{n,t}(u) = (k_n/n)e^{-\iota u} \in [0,1]$ and $\iota > 0$ sufficiently small such that $\varphi_{l_n} = o\left(l_n^{-\lambda}\right)$ by a continuity argument. ■

Proof of Lemma 2.4. Define $\bar{I}_{n,t}(u) := I(X_t > b_n e^u)$.

Claim 1: Let $h_t^{(l_n)}(u) := I(g_t^{(l_n)} > b_n e^u)$ for some $F_{t-l_n}^{t+l_n}$-measurable random variable $g_t^{(l_n)}$. Use (A.1), (A.2), (A.6) and the L_0-E-APP property to deduce for any $\eta \in (0,1)$

$$\begin{aligned} \mathrm{E}\left(\bar{I}_{n,t}(u) - \mathrm{E}\left[\bar{I}_{n,t}(u)|F_{t-l_n}^{t+l_n}\right]\right)^2 &\leq \mathrm{E}\left(\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right)^2 I\left(\left|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right| \leq \eta\right) \\ &\quad + K \times P\left(\left|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right| > \eta\right) \\ &\leq 0 + K \times e_{n,t}(u) \times \varphi_{l_n}. \end{aligned}$$

The E-NED claim follows from the stated L_0-E-APP properties of $e_{n,t}(u)$ and φ_{l_n}.

Claim 2: Let $h_t^{(l_n)}(u) := P(X_t > b_n e^u | F_{t-l_n}^{t+l_n})$ and invoke the L_0-E-APP property with constants $e_{n,,t}(u) = (k_n/n)e^{-u}$ for each $n \geq 1$. For any positive sequences $\{\eta_n, \delta_n\}$, $\eta_n \in (0,1)$ and $\delta_n < \eta_n$, and $K > 0$

$$\begin{aligned} &\mathrm{E}\left(\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right)^2 I\left(\left|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right| \leq \eta_n\right) \qquad \textbf{(A.7)} \\ &\quad \leq \delta_n + \int_{\delta_n}^{\eta_n} P\left(\left|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right| > x^{1/2}\right) dx \\ &\quad \leq K\delta_n + K e_{n,,t}(u)\varphi_{l_n}. \end{aligned}$$

Similarly, $\mathrm{E}[(\bar{I}_{n,t}(u) - h_t^{(l_n)}(u))^2 I(|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)| > \eta_n)]$ is bounded by

$$\int_{\eta_n}^{1} P\left(\left|\bar{I}_{n,t}(u) - h_t^{(l_n)}(u)\right| > x^{1/2}\right) dx \leq K e_{n,t}(u)\varphi_{l_n}. \qquad \textbf{(A.8)}$$

Together, (A.7) with $\delta_n = e_{n,,t}(u)\varphi_{l_n}$ and (A.8) imply L_2-E-NED. ■

Proof of Lemma 3.2. Under the stated suppositions $\{X_t\}$ is L_0-APP with coefficients $\upsilon_{l_n} = o(l_n^{-\lambda})$ and approximator $\{g_t^{(l_n)}\}$, and L_0-E-APP with coefficients $\varphi_{l_n} = o(l_n^{-\lambda})$ and approximator $\{I(g_t^{(l_n)} > b_n e^u)\}$ by Lemma 2.3. By construction this implies $I_{n,t}(u)$ is L_0-APP with coefficients φ_{l_n} and approximator $\{I(g_t^{(l_n)} \leq b_n e^u)\}$.

We need only demonstrate that the conditions of Davidson (1994, Thm 17.22) are satisfied to prove that $X_t^s I_{n,t}(u)$ and $X_t^s \bar{I}_{n,t}(u)$ are also L_0-APP with coefficients $\upsilon_{l_n} + \varphi_{l_n}$ of size $\min\{\lambda,\lambda\} = \lambda$ and approximators $\{(g_t^{(l_n)})^s I(g_t^{(l_n)} \leq b_n e^u)\}$ and $\{(g_t^{(l_n)})^s I(g_t^{(l_n)} > b_n e^u)\}$, respectively.

Fix $s = 1$ and consider $X_t I_{n,t}(u)$; the proofs for arbitrary $s > 0$ and $X_t^s \bar{I}_{n,t}(u)$ are identical. Define vectors $a = [a_1, a_2]' \in \mathbb{R}^2$ and $b = [b_1, b_2]' \in \mathbb{R}^2$, and define a mapping $B : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by $B(a,b) = |a_1| + |b_2|$. Thus $B(a,b)$ sums the absolute values of the first element of a and the second element of b. Let $\{X_t^{(1)} I_{n,t}^{(1)}(u),\ X_t^{(2)} I_{n,t}^{(2)}(u)\}$ be two copies of $X_t I_{n,t}(u)$. Then

$$\left|X_t^{(1)} I_{n,t}^{(1)}(u) - X_t^{(2)} I_{n,t}^{(2)}(u)\right|$$
$$\leq \left\{\left|X_t^{(1)}\right| + \left|I_{n,t}^{(2)}(u)\right|\right\} \times \left\{\left|X_t^{(1)} - X_t^{(2)}\right| + \left|I_{n,t}^{(1)}(u) - I_{n,t}^{(2)}(u)\right|\right\}$$
$$= B\left(\left\{X_t^{(1)}, I_{n,t}^{(1)}(u)\right\}, \left\{X_t^{(2)}, I_{n,t}^{(2)}(u)\right\}\right) \times \left\{\left|X_t^{(1)} - X_t^{(2)}\right| + \left|I_{n,t}^{(1)}(u) - I_{n,t}^{(2)}(u)\right|\right\}.$$

But since X_t and $I_{n,t}(u)$ are L_p-bounded, clearly

$$\left\|B\left(\left\{X_t, I_{n,t}(u)\right\}, \left\{g_t^{(l_n)}, I\left(g_t^{(l_n)} \leq b_n e^u\right)\right\}\right)\right\|_p = \left\||X_t| + \left|I\left(g_t^{(l_n)} \leq b_n e^u\right)\right|\right\|_p$$
$$\leq \|X_t\|_p + 1 \leq K \qquad \text{if } p \geq 1$$
$$\leq K\left(\mathrm{E}|X_t|^p + 1\right)^{1/p}$$
$$\leq K \qquad \text{if } p \in (0,1),$$

by Minkowski's and Loève's inequalities, respectively. This verifies the conditions of Davidson's (1994, Thm. 17.22) ■

Proof of Theorem 3.3. We prove the claim for $b_n^{-1} X_t I(|X_t| \leq b_n)$ since the argument for $b_n^{-s} X_t^s I(|X_t| \leq b_n)$ and arbitrary $s > 0$ is identical.

Write $\hat{Z}_{n,t}^* := b_n^{-1}\hat{X}_{n,t} = b_n^{-1} X_t I(|X_t| \leq b_n)$ and $\hat{g}_{n,t}^{(l_n)} := g_t^{(l_n)} I(g_t^{(l_n)} \leq b_n)$. Since $|\hat{Z}_{n,t}^*| \leq 1$ a.s. uniformly in n and t, clearly

$$\limsup_{n\geq 1} \sup_{1\leq t\leq n} \left\{\left|\hat{Z}_{n,t}^* - \mathrm{E}\left[\hat{Z}_{n,t}^* | \mathcal{F}_{t-l_n}^{t+l_n}\right]\right|\right\} \leq K < \infty.$$

Further, $\{\hat{X}_{n,t}\}$ is L_0-APP on $\{\mathcal{F}_t\}$ with coefficients $w_{l_n} \in [0,1)$ of size λ, and approximator $\{\hat{g}_{n,t}^{(l_n)}\}$ by Lemma 3.2. Now use (A.1), boundedness $b_n^{-1}|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}| \leq 1$ a.s. with (A.2), and the L_0-APP property to deduce for any sequence of uniformly positive numbers $\{\eta_n\}$ and any $p \geq 2$

$$\mathrm{E}\left|\hat{Z}_{n,t}^* - \mathrm{E}\left[\hat{Z}_{n,t}^* | \mathcal{F}_{t-l_n}^{t+l_n}\right]\right|^p \leq \mathrm{E}\left[\left(\hat{Z}_{n,t}^* - \mathrm{E}\left[\hat{Z}_{n,t}^* | \mathcal{F}_{t-l_n}^{t+l_n}\right]\right)^2\right]$$
$$\leq b_n^{-1}\mathrm{E}\left[\left(\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right)^2 \times I\left(\left|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right| \leq \eta_n\right)\right]$$
$$+ K \times P\left(\left|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right| > \eta_n\right)$$
$$\leq b_n^{-1}\mathrm{E}\left[\left(\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right)^2 \times I\left(\left|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right| \leq \eta_n\right)\right]$$
$$+ K \times w_{l_n}. \tag{A.9}$$

Similar to (A.7), use L_0-APP to deduce for any sequence $\{\delta_n\}$, $0 < \delta_n < \eta_n$,

$$\mathrm{E}\left[\left(\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right)^2 \times I\left(\left|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right| \leq \eta_n\right)\right]$$

$$= \int_0^{\delta_n} P\left(\left|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right| > \upsilon^{1/2}\right) d\upsilon + \int_{\delta_n}^{\eta_n} P\left(\left|\hat{X}_{n,t} - \hat{g}_{n,t}^{(l_n)}\right| > \upsilon^{1/2}\right) d\upsilon$$

$$\leq \delta_n + (\eta_n - \delta_n) \times w_{l_n}. \quad \textbf{(A.10)}$$

Together, (A.9) and (A.10) imply

$$\mathrm{E}\left|\hat{Z}_{n,t}^* - \mathrm{E}\left[\hat{Z}_{n,t}^* | F_{t-l_n}^{t+l_n}\right]\right|^p \leq b_n^{-1}\left[\delta_n + (\eta_n - \delta_n) \times w_{l_n}\right] + w_{l_n}.$$

Choose $\delta_n = w_{l_n} \in [0, 1)$ and $\eta_n \in (1, 2]$ to get

$$\mathrm{E}\left|\hat{Z}_{n,t}^* - \mathrm{E}\left[\hat{Z}_{n,t}^* | F_{t-l_n}^{t+l_n}\right]\right|^p \leq K(1 + b_n) w_{l_n} \leq K w_{l_n}.$$

Hence $||\hat{Z}_{n,t}^* - \mathrm{E}[\hat{Z}_{n,t}^* | F_{t-l_n}^{t+l_n}]||_p \leq K w_{l_n}^{1/p} = o(l_n^{-\lambda/p})$.

Now apply Lyapunov's inequality to deduce for any $p \in (0, 2)$, $||\hat{Z}_{n,t}^* - \mathrm{E}[\hat{Z}_{n,t}^* | F_{t-l_n}^{t+l_n}]||_p \leq ||\hat{Z}_{n,t}^* - \mathrm{E}[\hat{Z}_{n,t}^* | F_{t-l_n}^{t+l_n}]||_2 \leq o(l_n^{-\lambda/2})$. This completes the proof. ■

Proof of Theorem 3.4. Apply Theorem 3.3 to deduce for any $s > 0$ and $p > 0$, $||\hat{X}_{n,t}^s - \mathrm{E}[\hat{X}_{n,t}^s | F_{t-l_n}^{t+l_n}]||_p \leq b_n^s l_n^{-\iota} o(l_n^{-\lambda/\max\{p,2\}})$. Choose $\iota > 0$ sufficiently tiny and $l_n \to \infty$ sufficiently fast to complete the proof. ■

Proof of Lemma 4.5. If $p \geq 1$ the proof follows from Minkowski's inequality (e.g., Davidson, 1994, p. 263). If $p \in (0, 1)$ apply Loève's inequality, $\sup_{t \in \mathbb{N}} ||u_t||_p$, and $|\psi_{t,i}| \leq |\psi_i|$ to deduce $\mathrm{E}|X_t - \mathrm{E}[X_t | G_{t-l}^{t+l}]|^p \leq K(\sum_{i=l+1}^{\infty} |\psi_i|)^{1/p}$. Exploit $\psi_i = O(\rho^i)$ or $\psi_i = O(i^{-\mu})$ to finish the proof. ■

Proof of Theorem 5.1. Under the stated conditions and Theorem 2.5 $\{X_t\}$ is L_2-E-NED with size 1/2, constants $d_{n,t}(u) = K(k_n/n)^{1/2} e^{-u/2}$, and α-mixing base with size 1. Coupled with the tail property $P(X_t > x) = cx^{-\kappa}(1 + O(x^{-\varsigma}))$ and fractile bound $k_n = O(n^{2\varsigma/(2\varsigma+\kappa)})$ all conditions of Lemma 3 and Theorem 2 of Hill (2010) are satisfied, delivering both limits. ■

Proof of Theorem 5.2. By supposition $\{X_t\}$ is L_p-bounded, $p > 0$ and $\mathrm{E}[\hat{X}_{n,t}] = 0$ for each $n \geq 1$ and $1 \leq t \leq n$. We need only verify Assumptions A–C of Hill (2009b) to invoke his Theorem 3.1 central limit theorem: $\sum_{t-1}^{n} (\hat{X}_{n,t}^* - \mathrm{E}[\hat{X}_{n,t}])/\upsilon_n \xrightarrow{d} N(0, 1)$. The assumptions are listed here for reference using our notation.

(A) $\liminf_{n\to\infty}\{\upsilon_n^2/n\} > 0$; $b_n = O(n^{1/2-\iota})$; and $k_n/n^{\iota} \to \infty$ for tiny $\iota > 0$.

(B) $\{\hat{X}_{n,t}\}$ is geometrically L_2-NED on geometrically α-mixing $\{\epsilon_t\}$.

(C) $\{I(|X_t| > c_n e^u)\}$ is geometrically L_2-NED on geometrically α-mixing $\{\epsilon_t\}$ with constants $e_{n,t}(u)$ Lebesgue integrable on $\mathbb{R}_+$ and $\sup_{1 \leq t \leq n} \sup_{u \geq 0}\{e_{n,t}(u)\} \leq K(n/k_n)^{1/2}$, and coefficients with size 1/2.

Assumption A holds by supposition. Since α-mixing implies NED and NED implies L_0-APP, Assumptions B and C hold by Theorems 3.3 and 2.5 respectively. ■