Extremal Memory of Stochastic Volatility with an Application to Tail Shape Inference

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

July 2010

ABSTRACT We characterize joint tails and tail dependence for a class of stochastic volatility processes. We derive the exact joint tail shape of multivariate stochastic volatility with innovations that have a regularly varying distribution tail. This is used to give four new characterizations of tail dependence. In three cases tail dependence is a non-trivial function of linear volatility memory parametrically represented by tail scales, while tail power indices do not provide any relevant dependence information. Although tail dependence is associated with linear volatility memory, tail dependence itself is nonlinear. In the fourth case a linear function of tail events and exceedances is linearly independent. Tail dependence falls in a class that implies the celebrated Hill (1975) tail index estimator is asymptotically normal, while linear independence of nonlinear tail arrays ensures the asymptotic variance is the same as the iid case. We illustrate the latter finding by simulation.

1. INTRODUCTION

Stochastic volatility models have risen in popularity in finance and macroeconomics due to their flexibility for explaining and forecasting volatility. Applications include equity returns, option prices, and exchange rates for which evidence suggests in many cases heavy tails. See Ghysels et al (1995), Embrechts et al (1997), Shephard (2005) and Davis and Mikosch (2009b). It is therefore natural to characterize tail shape and tail dependence for stochastic volatility [SV].

Consider a $k$-vector stationary process $\{x_t\}_{t \in \mathbb{Z}}$ with stochastic volatility $\{\Sigma_t\}_{t \in \mathbb{Z}}$,

$$x_t = \Sigma_{t-1} \epsilon_t, \quad \epsilon_t \in \mathbb{R}^k, \quad \Sigma_{t-1} = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ \cdots & \ddots & \cdots & \cdots \\ 0 & \cdots & \sigma_{k,t} \end{bmatrix}, \quad \sigma_{i,t} \geq 0 \ a.s., \quad (1)$$

---

1We thank a referee for helpful suggestions.

2Dept. of Economics, University of North Carolina-Chapel Hill, www.unc.edu/~jbhill, jbhill@email.unc.edu.

Keywords and phrases: stochastic volatility; tail dependence; convolution tail; Hill estimator.

where \( \{ \epsilon_t \}_{t \in \mathbb{Z}} \) and \( \{ \Sigma_t \}_{t \in \mathbb{Z}} \) are independent of each other, governed respectively by non-degenerate distributions, and \( E|\epsilon_t|^p < \infty, p > 0 \). We only consider the diagonal case

\[
\lim_{z \to 0} \frac{P(x_{i,t} - h > z, x_{j,t} > z)}{P(x_{i,t} > z)} = 1 + r_{i,j}(h)
\]

under relatively general conditions on the data generating processes of \( \{ \sigma_{i,t} \} \). A natural outcome will be characterizations of serial and bivariate tail dependence.

Although the moment, memory and aggregation properties of volatility processes have received extensive treatment (e.g. Nelson 1990, Ghysels et al 1995, Giraitis et al 2000, Basrak et al 2002a, Carrasco and Chen 2002, Davidson 2004, Meddahi and Renault 2004), only recently have volatility extremes been considered (Chernick 1981, de Haan et al 1989, Basrak et al 2002a,b, Mikosch and Stårică 2000, Klüppelberg and Lindner 2008, Davis and Mikosch 2009a,b). A large class of stationary GARCH processes, for example, have marginal distribution tails that decay according to a power law, and exhibit extremal clustering. See Mikosch and Stårică (2000), Basrak et al (2002a,b), and Davis and Mikosch (2009a,c), and see details below.

In this paper we show SV with heavy tailed innovations \( \{ \epsilon_t \} \) that have independent extremes exhibits tail dependence as a function of volatility dependence. In particular, we show in Corollary 2.3 of Section 2

\[
\lim_{z \to 1} \frac{P(x_{i,t} > z_i, x_{j,t} > z_j)}{P(x_{i,t} > z_i, x_{j,t} > z_j)} = 1 + \rho_{i,j}(h)
\]

for some function \( r_{i,j} : \mathbb{N} \to \mathbb{R} \) that decays according to linear dependence in \( \{ \sigma_{i,t}, \sigma_{j,t-h} \} \) where \( \alpha_l \) are indices of regular variation. In the univariate case write \( r(h) = r_{i,t}(h) \). In order to address (2) we characterize multivariate product convolution tails for \( x_i \) in (1) and link the result in a simple way to tail memory. Extant theory for product convolution tails exists only in the univariate case, e.g. \( \{ \epsilon_{i,t}^{\alpha_i} \sigma_{i,t} \} \), and not the multivariate case, e.g. \( \{ \epsilon_{i,t}^{\alpha_i} \sigma_{i,t}^{\alpha_{j-t-1}} \} \). See especially Breiman (1965) and Cline (1986).

Notice (2) need not, and does not, hold for a wide range of processes. Processes with a well defined tail event correlation or "extremogram", for example, satisfy (Davis and Mikosch 2009c)

\[
\tilde{r}_{i,j}(h) := \lim_{z \to 1} \frac{P(x_{i,t-h} > z, x_{j,t} > z) - P(x_{i,t-h} > z) P(x_{j,t} > z)}{P(x_{i,t-h} > z)^{1/2} P(x_{j,t} > z)^{1/2}} = 0
\]

Stationary ARMA and GARCH processes exhibit \( \tilde{r}_{i,j}(h) \to 0 \) as \( h \to \infty \) at a geometric rate. If \( \tilde{r}_{i,j}(h) \neq 0 \) then clearly limit (2) does not exist. Conversely, under (2) the extremogram for SV is simply \( \tilde{r}_{i,j}(h) = 0 \) (Davis and Mikosch 2009c) hence SV is repeatedly cited as not exhibiting extremal clustering (e.g. Davis and Mikosch 2009b,c). In Section 3 we show this is an artifact of the
natural degenerateness of tail events, and the fact that $\tau_{i,j}(h)$ only measures linear dependence in tail events $I(x_{i,t} > z)$ where $I(\cdot) \in \{0,1\}$. Thus, $\tau_{i,j}(h) = 0$ need not align with independence in extremes, in particular SV satisfies nonlinear tail event dependence.

In Section 3 we characterize three other forms of tail dependence for SV processes $\{x_t\}$ for a wide range of volatility processes $\{\Sigma_t\}$ with geometric or hyperbolic memory. First, $\{x_t\}$ is Extremal-Near Epoch Dependent [E-NED]; its extreme indicators $I(|x_t| > z)$ as $z \to \infty$ are predictable in a non-trivial way even though they are asymptotically uncorrelated $\tau_{i,j}(h) = 0$. This implies tail arrays of $\{x_t\}$ have Gaussian limits, including log-high threshold exceedances $\max\{0, \ln(|x_t|/z)\}$ as $z \to \infty$, and therefore Gaussian asymptotics are available for tail index and tail dependence estimators. See also Hill (2008, 2009, 2010).

Second, $\tau_{i,j}(h)$ in (2) is non-trivially summable which ensures $x_t$ is $D^3$-mixing (Leadbetter’s 1974, 1983). Third, $\tau_{i,j}(h)$ summability ensures nonlinear tail arrays of $\{x_t\}$ exhibit a form of linear tail independence. The value here is the asymptotic variance of the Hill (1975) tail index estimator is the same as in the iid case. See Section 3.3. Although tail index estimation and inference has received a wide range of treatments, as far as we know it has not been shown that inference for any particular estimator for any class of random volatility is identical to an iid process. In particular, there has been very little study into the asymptotic variance of tail index estimators under data dependence (Hill 2010)\(^3\). See, e.g., Pickands (1975), Hall (1982), Davis and Resnick (1984), Smith (1987), Rootzén et al (1990), Smith and Weissman (1994), Beirlant, Vynckier and Teugels (1996), de Haan and Resnick (1998), Resnick and Stărică (1998), Drees et al (2004), Beirlant, Dierckx, and Guillou (2005), and Hill (2009, 2010), and see Hill (2010) for recent a review. We illustrate the finding by simulation in Section 4.

There are now a multitude of ways to define tail dependence, and several well known measures have been used to characterize GARCH and SV. The most popular depict GARCH as tail dependent and SV as independent in the tails. Assume for notational clarity $x_t = \sigma_t \epsilon_t$ is a scalar, and recall a GARCH(1,1) process (Bollerslev 1986)

$$x_t = \sigma_t \epsilon_t \quad \text{where} \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \gamma \epsilon_{t-1}^2, \quad \omega > 0, \quad \beta, \gamma \geq 0.$$  \hspace{1cm} (4)

Assume the distribution of $\epsilon_t$ is strictly positive and continuous on $\mathbb{R}$, $\beta + \gamma > 0$, $E[\max\{0, \ln|\epsilon_t|\}] < \infty$ and $E[\ln(\beta + \gamma \epsilon_t^2)] < 0$. Then $x_t$ is stationary, geometrically $\alpha$-mixing and has regularly varying distribution tails (Bougerol and Picard 1992, Basrak et al 2002: Theorem 3.1).

The seminal approach for characterizing tail dependence dates at least to Sibuya (1960), Mardia (1964) and Loynes (1965). Assuming an infinite support, if

$$\lim_{z \to \infty} P(x_t > z \mid x_{t-h} > z) = 0 \quad \text{(5)}$$

\(^3\)Hsing (1991) offers some details for mixing sequences, but under fairly complicated assumptions on the limits of sample covariances of nonlinear tail arrays. Hill (2010) side steps the problem altogether by delivering a consistent nonparametric estimator of the asymptotic variance of Hill’s (1975) estimator for a massive array of dependent, heterogeneous processes.
then we say \( x_{t-h} \) and \( x_t \) are "asymptotically independent" (Loynes 1965). Notice if (2) holds for some \( r(h) \) then
\[
\lim_{z \to \infty} P(x_t > z \mid x_{t-h} > z) = (1 + r(h)) \times \lim_{z \to \infty} P(x_t > z) = 0,
\]

hence SV is asymptotically independent.

Ledford and Tawn (1996, 1997, 2003) propose a bivariate tail model to give more structure to the asymptotic independence case. Define a unit Fréchet transformation \( x_t^* := -1/\ln F_t(x_t) \) where \( F_t(x) := P(x_t \leq x) \), thus \( P(x_t^* \leq z) = \exp\{-1/z\} \). Ledford and Tawn (2003) extend their seminal framework to a time series setting by assuming there exists a coefficient \( \eta_h \in [0,1] \) and slowly varying \( L_h(z) \) such that
\[
P(x_{t-h}^* > z, x_t^* > z) = z^{-1/\eta_h} L_h(z) \quad \text{as} \quad z \to \infty. \tag{6}
\]

All values \( \eta_h < 1 \), or \( \eta_h = 1 \) with \( \lim_{z \to \infty} L_h(z) = 0 \), imply \( P(x_t^* > z \mid x_{t-h}^* > z) \to 0 \) hence "asymptotic independence", and if \( \eta_h = 1 \) and \( L_h(z) \to 0 \) then \( P(x_t^* > z \mid x_{t-h}^* > z) \to 0 \) hence "asymptotic dependence" (Ledford and Tawn 1996, 1997). See also Coles and Tawn (1994), Heffernan and Tawn (2004), and Schmidt and Stadtmüller (2006) to name a very few.

The usefulness of (6) lies precisely in a refined depiction of the asymptotic independence case \( \eta_h < 1 \) since \( \eta_h < 1/2 \), \( \eta_h = 1/2 \) and \( \eta_h \in (1/2,1) \) respectively imply large values are negatively associated, independent, and positively associated. Since \( x_t^* \) is a monotonic transform, \( P(x_t^* > z) = 1 - \exp\{-1/z\} = z^{-1} \times (1 + o(1)) \), and \( L_h(z) \) is slowly varying it follows model (6) implies limit (2) is valid
\[
\frac{P(x_{t-h}^* > z, x_t^* > z)}{P(x_{t-h}^* > z) P(x_t^* > z)} = z^{2-1/\eta_h} L_h(z) \times (1 + o(1)) \to 1 + r(h)
\]

if and only if \( \eta_h = 1/2 \) and \( L_h(z) \to 1 + r(h) \). Ledford and Tawn (2003) refer to this case as "near independence". By comparison, a stationary GARCH process (4) has a geometrically decaying extremogram (Davis and Mikosch 2009c)
\[
\tilde{r}(h) := \lim_{z \to \infty} \frac{P(x_{t-h} > z, x_t > z)}{P(x_{t-h} > z)^{1/2} P(x_t > z)^{1/2}} = O(\rho^h) \quad \text{for} \quad |\rho| < 1.
\]

Under the Ledford-Tawn model (6) this implies asymptotic dependence \( \eta_h = 1 \) and \( \lim_{z \to \infty} L_h(z) = \tilde{r}(h) = O(\rho^h) \) since
\[
\tilde{r}(h) := \lim_{z \to \infty} \frac{P(x_{t-h}^* > z, x_t^* > z)}{P(x_{t-h}^* > z)^{1/2} P(x_t^* > z)^{1/2}} = \lim_{z \to \infty} z^{1-1/\eta_h} L_h(z) = O(\rho^h).
\]

GARCH extremes cluster so much \( P(x_{t-h} > z, x_t > z) \) decays slower than SV by an asymptotic magnitude \( P(x_t > z) \).

Another tail exponent dependence measure is due to Leadbetter (1974, 1983), cf. Leadbetter et al (1983). The maximum \( \max_{1 \leq t \leq n} |x_t| \) of many weakly
dependent sequences $\{x_i\}_{i=1}^n$ satisfies

$$\lim_{n \to \infty} P \left( \frac{1}{b_n} \max_{1 \leq i \leq n} |x_i| \leq z \right) = e^{-\theta z^{-\kappa}}, \quad z \geq 0, \quad \theta \in [0, 1] \tag{7}$$

if and only if $\{\tilde{b}_n\}$ satisfies $nP(|x_i| > \tilde{b}_n) \to 1$. See also Loynes (1965) and O’Brien (1974). The extremal index $\theta$ is roughly the inverted mean number of high threshold exceedances, hence $\theta = 1$ implies independence, $\theta \in (0, 1)$ short-range dependence, and $\theta = 0$ long-range dependence. Breidt and Davis (1998) and Davis and Mikosch (2009b,c) show SV with iid errors satisfy $\theta = 1$, the case of "no extremal clustering" (Davis and Mikosch 2009b: p. 356). GARCH (4), by comparison, exhibits short-range extremal clustering $\theta \in (0, 1)$. See Mikosch and Stărică (2000) and Davis and Mikosch (2009a,b).

It is important to note the shortcomings of concluding SV extremes do not cluster by focusing on extant measures (3), (5), (6) and (7). First, they do cluster in the senses of (2) and E-NED. Thus, labels like "near independence" (Ledford and Tawn 2003) and "no extremal clustering" (Breidt and Davis 1998, Davis and Mikosch 2009b,c) are problematic because they are based on somewhat narrow definitions of tail dependence. In particular, that SV has an extremal index $\theta = 1$ and extremogram $\hat{r}(h) = 0$ only suggests extremes are *linearly tail independent*, while non-trivial predictability of extreme events based on volatility memory suggests *nonlinear tail dependence*. This has two important repercussions: asymptotic theory for tail arrays of SV is not trivial, and we cannot treat extremes as independent when characterizing the distribution limit of tail estimators.

Throughout $(z)_+ := \max\{z, 0\}$. $K > 0$ is an arbitrary finite constant whose value may change from line to line, and $\lambda > 0$ is a tiny constant. $a_n \sim b_n$ implies $a_n/b_n \to 1$. $\|\cdot\|_p$ denotes the $L_p$-norm: $\|x\|_p := (E|x|^p)^{1/p}$. $\eta \iid N(0, 1)$ states $\eta$ is iid normally distributed with zero mean and unit variance.

### 2. PRODUCT CONVOLUTION TAILS

Assume $\{\epsilon_{i,t}, \sigma_{i,t}\}$ are stationary, $L_p$-bounded $p > 0$, and positive with probability one, hence

$$\tilde{F}_i(z) := P(x_{i,t} > z) \text{ has support } [0, \infty).$$

This captures transforms of two-tails $|x_{i,t}| = \sigma_{i,t} |\epsilon_{i,t}|$, left-tail $-x_{i,t} \times I(\epsilon_{i,t} < 0)$ or right-tail $x_{i,t} \times I(\epsilon_{i,t} > 0)$. All subsequent measures of tail dependence are therefore tail specific: two-tailed, left-tailed, right-tailed, or cross-tailed.

We assume the marginal tails $P(\epsilon_{i,t} > z)$ are regularly varying at $\infty$: there exist $\alpha_i > 0$ and slowly varying functions $L_i(z)$ such that

$$P(\epsilon_{i,t} > z) = z^{-\alpha_i} L_i(z) \text{ as } z \to \infty. \tag{8}$$

Property (8) is encompassed by the larger class of subexponential distributions (see Embrechts et al 1997), and includes the domain of attraction of stable laws.
when \( \alpha_i < 2 \), and the maximum domain of attraction of a Type II extreme value distribution (Leadbetter et al 1983, Bingham, Goldie and Tengels 1987, Resnick 1987).

If \( E[\sigma_{i,t}^{\alpha_i}] < \infty \) then \( x_{i,t} \) also satisfies (8) with index \( \alpha_i > 0 \) since (see Breiman 1965):

\[
P(x_{i,t} > z) \sim E[\sigma_{i,t-1}^{\alpha_i}] \times P(\epsilon_{i,t} > z) = z^{-\alpha_i} \times E[\sigma_{i,t-1}^{\alpha_i}] \times L_i(z). \tag{9}
\]

If \( \sigma_{i,t} \) has tail shape (8) with an index \( \kappa_i \leq \alpha_i \) then \( x_{i,t} = \epsilon_{i,t} \sigma_{i,t-1} \) satisfies (9), but the form of the slowly varying component is not in general known when \( \kappa_i = \alpha_i \) (Cline 1986: Proposition II). We take the approach of Breidt and Davis (1998) and Davis and Mikosch (2009b,c) and do not consider the case \( E[\sigma_{i,t}^{\kappa_i}] = \infty \).

2.1 Preliminary Results

We assume \( \epsilon_t \) satisfies tail independence by imposing a multivariate version of \( \rho_{i,t} \) property (10) implies pairwise \( \lim_{z \to \infty} P(\epsilon_{i_1,t_1} > z_1, \ldots, \epsilon_{i_l,t_l} > z_l) \to 1 \) as \( z_i \to \infty \) \( \tag{10} \)

for all \( l \)-tuples \( \{i_1, \ldots, i_l\} \) and \( \{t_1, \ldots, t_l\} \) with \( 1 \leq l \leq k \), where either \( i_{j_1} \neq i_{j_2} \) for \( j_1 \neq j_2 \), or \( i_{j_1} = i_{j_2} \) and \( t_{j_1} \neq t_{j_2} \). Finally, \( 0 < E[\Pi_{j=1}^{l} \sigma_{i_{j_1},t_{j_1}}] \)

\(< \infty \).

**Remark 1:** Consider a candidate pair \( \{\epsilon_{i_1,t_1}, \epsilon_{i_2,t_2}\} \). In lieu of (2) property (10) implies pairwise \( \lim_{z \to \infty} P(\epsilon_{i_1,t_1} > z, \epsilon_{i_2,t_2} > z) / P(\epsilon_{i_1,t_1} > z)P(\epsilon_{i_2,t_2} > z) = 1 + r_{i_1,i_2}(t_2 - t_1) \) with \( r_{i_1,i_2}(t_2 - t_1) = 0 \). Thus, from (5) and (6) the \( \epsilon_{i,t}'s \) are serially and mutually asymptotically independent (Loynes 1965, Ledford and Tawn 1996, 1997, 2003).

**Remark 2:** The moment bound \( E[\Pi_{j=1}^{l} \sigma_{i_{j_1},t_{j_1}}^{\alpha_{i_{j_1}}}] < \infty \) holds when \( E[\sigma_{i,t}^{\kappa_i}] \)

\(< \infty \).

Our first result extends Breiman’s (1965) univariate product convolution tail (9) to multivariate tails.

**THEOREM 2.1 (Multivariate Product Convolution Tails)** Under Assumption A for all \( l \)-tuples \( \{i_1, \ldots, i_l\} \) and \( \{t_1, \ldots, t_l\} \), \( i_{j_1} \neq i_{j_2} \) for \( j_1 \neq j_2 \), \( 1 \leq
The remainder of this section applies Theorem 2.1 to deduce the structure of tail dependence for SV.

### 2.2 Bivariate Tails

Theorem 2.1 implies bivariate tail independence

\[
P(x_{i,s} > z_i, x_t > z_j) \sim P(x_{i,s} > z_i) P(x_t > z_j) \quad \text{as} \quad \{z_i, z_j\} \to \infty,
\]

if and only if the bivariate volatility process \(\{\sigma_{i,t}^\alpha, \sigma_{j,t}^\alpha\}\) is linearly independent:

\[
E[\sigma_{i,t}^\alpha \sigma_{j,t}^\alpha] = E[\sigma_{i,t}^\alpha] E[\sigma_{j,t}^\alpha].
\]

A simple way to demonstrate this involves a univariate Stochastic Log-Distributed Lag Volatility process

\[
x_t = \sigma_t \epsilon_t, \quad \epsilon_t \text{ iid with tail (8) and index } \alpha > 0, \quad \text{(11)}
\]

hence \(\sigma_t^\alpha = \prod_{i=0}^{\infty} e^{\alpha \psi_{i+\eta_{t-i}}}\). Examples include a log-autoregression \(\ln \sigma_t = \phi \ln \sigma_{t-1} + \eta_t, \quad |\phi| < 1\) (Taylor 1986, Breidt and Davis 1998, Davis and Mikosch 2009b) and log-ARFIMA\((p, d, q)\) with Hurst index \(d < 1/2\). Since Gaussian stationarity ensures \(E[\sigma_t^{2\alpha}] < \infty\), all parts of Assumption A are satisfied. Hence, the degree of serial tail dependence is measured by

\[
\lim_{z \to \infty} \frac{P(x_{t-h} > z, x_t > z)}{P(x_{t-h} > z) P(x_t > z)} = \frac{E[\sigma_{t-h}^\alpha \sigma_t^\alpha]}{E[\sigma_{t-h}^\alpha] E[\sigma_t^\alpha]} = \exp \left\{ 2\alpha^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h} \right\}.
\]

**COROLLARY 2.2** Process (11) satisfies (2) with \(r(h) = \exp\{2\alpha^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}\} - 1\).

Remark: If \(\sigma_t\) exhibits geometric memory \(\psi_i = O(\rho^i), \quad |\rho| < 1\), then

\[
\lim_{z \to \infty} \frac{P(x_{t-h} > z, x_t > z)}{P(x_{t-h} > z) P(x_t > z)} = 1 + O(\rho^h) \quad \text{hence} \quad r(h) = O(\rho^h).
\]
Similarly, under hyperbolic memory \( \psi_t = O(t^{-\mu}), \mu > 1/2 \) (e.g. ARFIMA\((p, d, q)\), \( d < 1/2 \)), then \( r(h) = O(h^{-\mu}) \).

Now define
\[
    r_{i,j}(h) := \frac{E \left[ (\sigma_{i,t} \alpha_i - E[\sigma_{i,t} \alpha_i]) \left( \sigma_{j,t-h} \alpha_j - E[\sigma_{j,t-h} \alpha_j] \right) \right]}{E[\sigma_{i,t} \alpha_i] E[\sigma_{j,t-h} \alpha_j]} \in (-\infty, \infty). \tag{12}
\]

Theorem 2.1 implies the following key results depicting tail dependence and bivariate regular variation.

**COROLLARY 2.3 (Bivariate Tail Dependence)** Let \( r_{i,j}(h) \) be defined by (12). Under Assumption A
\[
    \lim_{z_1, z_2 \to -\infty} \frac{P(x_{i,t-h} > z_i, x_{j,t} > z_j)}{P(x_{i,t-h} > z_i) P(x_{j,t} > z_j)} = 1 + r_{i,j}(h).
\]

Recall the unit Fréchet transform \( x_{i,t}^* \). Corollary 2.3 establishes (2), hence the discussion following (6) proves the next claim.

**COROLLARY 2.4 (Bivariate Tail Decay)** Under Assumption A the Ledfording-Tawn model (6) applies with \( \alpha_i = 1 = 2 \) for all \( i \) and \( \Theta_i(\alpha_i(z)) = 1 + r_{i,j}(h) \). In particular
\[
    P(x_{i,t-h} > z, x_{j,t}^* > z) = [1 + r_{i,j}(h)] \times z^{-2}(1 + o(1)).
\]

**3. TAIL INDEX INFERENCE**

In this section we explore three other notions of tail dependence. We then use the notions to characterize the Hill (1975) estimator for a univariate stationary SV process
\[
    x_t = \sigma_t \epsilon_t, \ \{\epsilon_t\} \text{ is independent of } \{\sigma_t\}. \tag{13}
\]
Assume \( \epsilon_t \) is iid \( L_\alpha \)-bounded with tail
\[
    \widetilde{\Phi}_\epsilon(z) := P(\epsilon_t > z) = z^{-\alpha} L_\epsilon(z) \text{ as } z \to \infty
\]
where \( L_\epsilon(z) \) is slowly varying. Let \( \widetilde{\Phi}_\epsilon(z) \) have support \([0, \infty)\), and assume \( \{\sigma_t^\alpha\} \) is a covariance stationary process as in Breidt and Davis (1998) and Davis and Mikosch (2009b,c). Then by (9) \( P(x_t > z) = K P(\epsilon_t > z) \), hence for slowly varying \( L_x(z) = KL_\epsilon(z) \)
\[
    \widetilde{\Phi}_x(z) := P(x_t > z) = z^{-\alpha} L_x(z) \text{ as } z \to \infty. \tag{14}
\]

The Hill (1975) estimator of \( \alpha^{-1} \) is the sample mean of peaks-over-thresholds
\[
    \hat{\alpha}_{mn}^{-1} = \frac{1}{m_n} \sum_{t=1}^n \left( \ln \left\{ x_t/x_{(m_n+1)} \right\} \right)_+ = \frac{1}{m_n} \sum_{t=1}^n \ln \left\{ x_t/x_{(m_n+1)} \right\}
\]
where \( n \) is the sample size, \( x_{(i)} \) denotes the \( i \)th order statistic, \( x_{(1)} \geq x_{(2)} \geq \ldots \), and \( \{m_n\} = \{m_n\}_{n \geq 1} \) is an intermediate order sequence as in Hall (1982): \( 1 \leq m_n < n, m_n \to \infty \) and \( m_n = o(n) \). As long as \( m_n \to \infty \) is restricted based on properties of slowly varying \( L_x \), cf. Haeusler and Teugels (1985), in the iid case it is known (e.g. Hall 1982)

\[
v_{m_n}^2 := E \left( m_n^{1/2} (\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}) \right)^2 \to \alpha^{-2}.
\]

See Section 3.3, below.

The estimator \( \hat{\alpha}_{m_n}^{-1} \) is consistent for \( \alpha^{-1} \) and asymptotically normal \( m_n^{1/2} (\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}) / v_{m_n} \sim N(0, 1) \) for a vast array of dependent, heterogeneous processes \( \{x_t\} \). See Hsing (1991), Resnick and Stàríča (1995) and Hill (2008, 2009, 2010) and their citations. We first demonstrate \( \hat{\alpha}_{m_n}^{-1} \) satisﬁes the requirements for asymptotic normality for a large class of volatility processes \( \{\sigma_t\} \). Second, we show \( v_{m_n}^2 \to \alpha^{-2} \) like iid data. The latter is quite challenging since it entails proving SV are tail independent in some sense not captured by any previous notion discussed above, including Davis and Mikosch’s (2009c) extremogram (3), Loynes’ (1965) asymptotic independence (5), Ledford and Tawn’s (2003) model (6) and the extremal index \( \theta \) in (7).

### 3.1 Tail Structure

In order to demonstrate \( v_{m_n}^2 \to \alpha^{-2} \) we use a decomposition of \( \hat{\alpha}_{m_n}^{-1} \) that exploits the concept of slow variation with remainder proposed in Smith (1982) and Goldie and Smith (1987). See also Haeusler and Tengels (1985), Hsing (1991), de Haan and Resnick (1996), and Hill (2010).

Assume \( \hat{F}_t(z)/\hat{F}_t(-z) \to 1 \) so there exist sequences of positive numbers \( \{\hat{m}_n, \hat{b}_n\}_{n \in \mathbb{N}} \) satisfying (Leadbetter et al 1983: Theorem 1.7.13)

\[
\frac{n}{\hat{m}_n} P(x_t > \hat{b}_n) \to 1. 
\]

The sequence \( \{\hat{m}_n\} \) may be an intermediate or extreme order sequence: the former \( \hat{m}_n \to \infty \) and \( \hat{m}_n = o(n) \), and the latter \( \hat{m}_n \to (0, \infty) \) (cf. Leadbetter et al 1983).

We say \( L_x \) is slowly varying with remainder when for all \( \lambda > 0 \) and some positive measurable function \( g : \mathbb{R}_+ \to \mathbb{R}_+ \)

\[
L_x(\lambda z)/L_x(z) - 1 = O(g(z)) \text{ as } z \to \infty. 
\]

Further, \( g(z) \) exhibits bounded increase: there exist \( 0 < D, z_0 < \infty \) and \( \tau \leq 0 \) such that \( g(\lambda z)/g(z) \leq D \lambda^\tau \) for some \( \lambda \geq 1 \) and \( z \geq z_0 \). Finally, \( g, \hat{m}_n \) and \( \hat{b}_n \) satisfy

\[
\hat{m}_n^{1/2} \times g(\hat{b}_n) \to 0. 
\]

Tails satisfying (14)-(17) include \( \hat{F}_t(z) = cz^{-\alpha} \) for all \( z \geq K, \hat{F}_t(z) = cz^{-\alpha}(1 + O(z^{-\xi})), \xi > 0, \) and \( \hat{F}_t(z) = cz^{-\alpha}(1 + O((\ln z)^{-\xi})) \). Property (17)
respectively requires $\bar{m}_n = o(n)$, $\bar{m}_n = o(n^{2\xi/(2\xi+\alpha)})$ and $\bar{m}_n = o((\ln n)^{2\xi})$. See Haeusler and Teugels (1985).

Define

$$\hat{r}_{\bar{m}_n}(h) = \frac{n}{\bar{m}_n} \left[ P \left( x_{t-h} > \bar{b}_n, x_t > \bar{b}_n \right) - P \left( x_t > \bar{b}_n \right)^2 \right].$$

In Section 1 we verified Davis and Mikosch’s (2009c) extremogram $\hat{r}(h) := \lim_{n \to \infty} \hat{r}_{\bar{m}_n}(h) = 0$ for $x_t$ in (13). Under slow variation the rate of decay $\hat{r}_{\bar{m}_n}(h) \to 0$ can be characterized.

**Lemma 3.1** Under (13)-(17) $(n/\bar{m}_n)\hat{r}_{\bar{m}_n}(h) = r(h) + o(1/\bar{m}_n^{1/2})$.

**Remark 1:** Although globally $\hat{r}_{\bar{m}_n}(h) \to 0$, SV exhibits $\hat{r}_{\bar{m}_n}(h) = (\bar{m}_n/n)r(h) + o(\bar{m}_n^{1/2}/n)$ hence locally correlated extreme events. This is typically referred to as Pitman drift in the inference literature on local alternatives (Pitman 1979, cf. Neyman and Pearson 1933). In this setting we can say $x_t$ is "$(n/\bar{m}_n)$-tail event correlated".

**Remark 2:** Locally correlated extreme events is similar in spirit to Ledford and Tawn’s (2003) near independence ($\eta_h = 1/2$ and $\lim_{z \to \infty} L_h(z) > 0$) and asymptotic independence with association of large values ($\eta_h \in (0, 1/2) \cup (1/2, 1)$). Each reveals tail dependence can be measured by careful consideration of scale and pre-asymptotic properties.

Tail events are naturally degenerate, so a careful re-scaling is required to deduce the nature of tail dependence for SV. We do this below with the Extremal-Near Epoch Dependence property.

### 3.2 Tail Dependence

Stochastic volatility exhibits a variety of tail dependence properties that can be used to show $m_n^{1/2}(\bar{\alpha} m_n^{-1} - \alpha^{-1})/v_{m_n} \xrightarrow{d} N(0, 1)$ and $\bar{v}_{m_n}^2 \to \alpha^{-2}$. Asymptotic normality is known for finite dependent, strong mixing, generalizations of Leadbetter’s (1983) $D'$-mixing, GARCH, Extremal-Near Epoch Dependent and Extremal-Approximable processes. Consider Davis and Resnick (1984), Rootzén et al (1990), Hsing (1991), Smith and Weissman (1994), de Haan and Resnick (1998), Resnick and Stârică (1998), Drees et al (2004) and Hill (2010) to name a few. In very few cases the asymptotic variance is characterizeable for non-finite dependent data (Hsing 1991). In general, however, we are not aware of results that lead to simple characterizations of the asymptotic variance in non-trivial cases.

#### 3.1.1 Extremal-Near Epoch Dependence

We want to prove asymptotic normality for a wide range of dependent and heterogeneous volatility processes $\{\sigma_t\}$ that covers mixing and non-mixing cases without restricting non-extremes. This is possible via Hill’s (2010) E-NED weak tail dependence property. See Hill (2008) for further theory, extensions and comparisons with other tail dependence properties.
Let \( \{ \epsilon_t \} \) be a possibly vector-valued stochastic process with \( \sigma \)-algebra
\[
\mathcal{F}_t := \sigma(\epsilon_\tau : \tau \leq t) : \mathcal{F}_a := \sigma(\epsilon_t : a \leq t \leq b),
\]
and let \( \{ m_n \} \) be an intermediate order sequence with associated threshold sequence \( \{ b_n \} \):
\[
n \frac{n}{m_n} \bar{F}_n(b_n) \to 1.
\]
We say \( \{ x_t \} \) is \( L^p \)-E-NED on \( \{ \mathcal{F}_t \} \) or on ”base” \( \{ \epsilon_t \} \), with ”size” \( \lambda > 0 \), if for some sequence of integer displacements \( \{ l_n \} \), \( l_n \to \infty \),
\[
\| I(x_t > b_n e^u) - P(x_t > b_n e^u|\mathcal{F}_{t-l_n}) \|_p \leq d_{n,t}(u) \times \psi_t n,
\]
where \( d_{n,t} : \mathbb{R} \to \mathbb{R} \) is Lebesgue measurable, \( \sup_{u \geq 0} \sup_{1 \leq t \leq n} d_{n,t}(u) = O((m_n/n)^{1/p}) \), \( \psi_t n \in [0,1] \) and \( n^\lambda \psi_t n \to 0 \).

We call \( d_{n,t}(u) \) the ”constants” and \( \psi_t n \) the ”coefficients”. E-NED characterizes predictability of the tail events \( I(x_t > b_n e^u) \) while controlling for degenerateness through the constants \( d_{n,t}(u) \), and measuring memory through the coefficients \( \psi_t n = o(l_n^{-\lambda}) \). While mixing properties typically only require coefficients to gauge memory across displacement, the E-NED norm is sensitive to time \( t \), sample size \( n \), and displacement. Even in the case of stationarity the tail event \( I(x_t > b_n e^u) \) is degenerate as \( n \to \infty \). The constants \( d_{n,t}(u) \) capture such changes unrelated to memory. See Hill (2008, 2009, 2010) for details and comparisons to other tail dependence properties, and see Doukhan (1994) and Nze and Doukhan (2004) for comparisons of mixing and conventional NED properties.

For future reference, we say some \( \{ z_t \} \) is \( L^p \)-NED on \( \{ \mathcal{F}_t \} \) or on \( \{ \epsilon_t \} \) with size \( \lambda > 0 \) if for constants \( \{ d_t \} \) and coefficients \( \{ \varphi_t \} \), \( l^\lambda \varphi_t \to 0 \) as \( l \to \infty \) (Gallant and White 1988, Davidson 1994, Nze and Doukhan 2004):
\[
\| z_t - E[z_t|\mathcal{F}^{t+\lambda}] \|_p \leq d_t \varphi_t.
\]

Trivially \( L^p \)-NED implies \( L^q \)-NED \( \forall q \leq p \), while boundedness of the tail event \( I(x_t > b_n e^u) \) ensures \( L^p \)-E-NED is equivalent to \( L^p \)-E-NED \( \forall p \geq q \) with adjustment to size (Hill 2008: Theorem 2.1). Any process is \( L^p \)-NED or \( L^p \)-E-NED on itself \( x_t = \epsilon_t \) for any \( p > 0 \) with trivial constants \( d_t = 0 \) or \( d_{n,t}(u) = 0 \) and coefficients of any size \( \lambda > 0 \).

Despite SV being ”near independent” in the sense \( \eta_h = 1/2 \) and \( L_h(z) \to 1 + r(h) \) (Ledford and Tawn 2003), and exhibiting ”no extremal clustering” a la the extremogram \( \hat{r}(h) = 0 \) and extremal index \( \theta = 1 \) (Breidt and Davis 1998, Davis and Mikosch 2009b,c), \( \{ x_t \} \) is non-trivially \( L^2 \)-E-NED on \( \{ \eta_t, \epsilon_t \} \) with size that depends on volatility memory. Recall the definition of \( \alpha \)-mixing coefficients for a stationary sequence (Rosenblatt 1956, Ibragimov and Linnik 1971):
\[
\alpha_t := \sup_{t \in \mathbb{Z}} \sup_{A \subset \mathcal{F}_t, B \subset \mathcal{F}_{t+\lambda}} |P(A \cap B) - P(A)P(B)|.
\]
LEMMA 3.2 (E-NED) Let (13)-(17) hold where $L_\rho$-bounded iid $\epsilon_t$ has tail index $0 < \alpha < \rho$. If $\{\sigma_t\}$ is $L_{2\alpha}$-NED with coefficients $\varphi_t$ of size $2/\min\{p, 1\}$ on an $\alpha$-mixing base $\{\eta_t\}$ with coefficients $\alpha_t = o(l^{-1})$, then $\{x_t\}$ is $L_2$-E-NED with constants $d_{n,t}(u) = K(m_n/n)^{1/2}e^{-up/2}$ and coefficients $\varphi_{\min}^t(n,1)^{1/4} = o(l^{-1/2})$ on an $\alpha$-mixing base $\{\epsilon_t, \eta_t\}$ with size 1. In particular $L_2$-E-NED is equivalent to

$$
\left| \frac{n}{m_n} P(x_t > b_n e^u) - \frac{n}{m_n} E \left[ P \left( x_t > b_n e^u \cdot \eta_{t-l_n} \right)^2 \right] \right| \leq e^{-up} \times o(l_n^{-1/2}).
$$

(18)

Remark 1: Clearly $\sigma_t = \eta_t$ is a special case, so NED covers $\alpha$-, $\beta$-, $\phi$-, $\rho$-mixing, and geometrically ergodic volatility sequences $\{\sigma_t\}$.

Remark 2: In general NED does not imply mixing, or may hold when mixing is unknown. Examples include a stationary AR(1) with iid Bernoulli errors (Andrews 1984, Davidson 1994), and hyperbolic memory GARCH (Davidson 2004). Further, very heavy tailed GARCH not known to be mixing or NED may be E-NED (Hill 2008).

Remark 3: Bound (18) demonstrates how E-NED controls for degeneracy in order to reveal dependence since $E[I(x_t > b_n e^u)] = P(x_t > b_n e^u)$ and by properties of regular variation $P_n(u) := (n/m_n)P(x_t > b_n e^u) \rightarrow e^{-\alpha u}$. The scaled minimum mean-squared-error tail event predictor $(n/m_n)P(x_t > b_n e^u | \eta_{t-l_n})$ therefore approaches $(n/m_n)I(x_t > b_n e^u)$ at a rate faster than $l_n^{-1/2}$ because the "near epoch" $\eta_{t-l_n}$ asymptotically contains sufficient information for perfect non-trivial extreme event prediction in mean-squared-error.

3.1.2 Summability We use E-NED to prove consistency and asymptotic normality since the property suffices for tail limit theory under very general conditions. In order to show $\sigma_{t}^2 \rightarrow \alpha^{-2}$ we work with the coefficients $r(h)$ in (12) that depicts linear memory in $\{\sigma_t^2\}$. Volatility NED with geometric or hyperbolic memory ensures $r(h)$ is summable, which is key to showing $\sigma_{t}^2 \rightarrow \alpha^{-2}$.

LEMMA 3.3 (r(h) summability) Under the conditions of Lemma 3.2 $|\sum_{h=1}^{\infty} r(h)| < \infty$.

Summability implies Leadbetter’s (1974, 1983) $D'$-mixing property, cf. Watson (1954) and Loynes (1965). Define a sequence of extreme order thresholds $\{\hat{b}_n\}$ by $nP(x_t > \hat{b}_n) \rightarrow 1$, and exploit Lemmas 3.1 and 3.3 to deduce $\{x_t\}$ in (13) satisfies

$$
n \sum_{i=1}^{[n/m]} P(x_t > \hat{b}_n, x_{i+1} > \hat{b}_n) = \frac{1}{n} \sum_{i=1}^{[n/m]} n^2 P(x_t > \hat{b}_n, x_{i+1} > \hat{b}_n)$$

$$
= \frac{1}{n} \sum_{i=1}^{[n/m]} (1 + r(i) + o(1)) = \frac{1}{m} \times (1 + o(1)) + \frac{1}{n} \sum_{i=1}^{[n/m]} r(i).
$$
Take \( \limsup_{n \to \infty} \) on both sides, and then \( \lim_{m \to \infty} \) to deduce the right hand side converges to zero. Thus \( \{ x_t \} \) is \( D'(b_n) \)-mixing (Leadbetter 1983: eq. (1.3)).

**Lemma 3.4 (D'-mixing)** Under the conditions of Lemma 3.2 \( \{ x_t \} \) is \( D'(b_n) \)-mixing.

**Remark:** The above identity shows \( D'(b_n) \)-mixing does not imply \(| \sum_{i=1}^{\infty} r(h) | < \infty\), merely \( 1/n \sum_{i=1}^{n} r(i) = o(1) \).

In general there is not enough information to deduce \( \{ x_t \} \) satisfies Leadbetter’s (1974, 1983) \( D(\cdot \beta) \)-mixing property. If, however, \( \sigma_t \) is \( \alpha \)-mixing with size 1 then trivially \( x_t \) is \( D(b_n) \)-mixing and by Lemma 3.4 \( x_t \) is \( D_0(\cdot \beta) \)-mixing. Together they ensure \( x_t \) belongs to the maximum domain of attraction (7). See Leadbetter (1974: Theorem 3.1).

### 3.1.3 Linear Independence in Nonlinear Tail Arrays

The last property involves nonlinear tail arrays of exceedances and events

\[
U_{m,n,t} := (\ln x_t / b_n) - E[(\ln x_t / b_n)]
\]

\[
U^*_{m,n,t}(u) := I(x_t > b_n e^u) - E[I(x_t > b_n e^u)], \quad u \geq 0
\]

that form the basis of the Hill (1975) tail index estimator. Under the conditions of Lemma 3.2, \( \hat{\alpha}_{m,n}^{-1} \) obtains the decomposition (Theorem 2.4 of Hsing 1991, Lemma A.7 of Hill 2009)

\[
m_n^{1/2} (\hat{\alpha}_{m,n}^{-1} - \alpha^{-1}) = \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \left( U_{m,n,t} - \alpha^{-1} U^*_{m,n,t}(u/m_n^{1/2}) \right) + o_p(1).
\]

Use Lemma B.1 in Appendix B to deduce

\[
\frac{1}{m_n} \sum_{t=1}^{n} E \left[ \left( U_{m,n,t} - \alpha^{-1} U^*_{m,n,t}(u/m_n^{1/2}) \right)^2 \right] \to \alpha^{-2}.
\]

Then \( v_{m,n}^2 := E[m_n^{1/2}(\hat{\alpha}_{m,n}^{-1} - \alpha^{-1})]^2 \) satisfies the Helly-Bray theorem

\[
\lim_{n \to \infty} |v_{m,n}^2 - \alpha^{-2} - R_n| = 0
\]

where

\[
R_n := \frac{1}{m_n} \sum_{s \neq t}^{n} E \left[ \left( U_{m,n,s} - \alpha^{-1} U^*_{m,n,s}(u/m_n^{1/2}) \right) \left( U_{m,n,t} - \alpha^{-1} U^*_{m,n,t}(u/m_n^{1/2}) \right) \right].
\]

Note \( R_n \) captures linear dependence in the nonlinear tail array \( \{ U_{m,n,t} - \alpha^{-1} U^*_{m,n,t}(u/m_n^{1/2}) \} \). We will prove \( R_n \to 0 \) for \( x_t \) in (13). Write

\[
P_{n,s,t}(u,v) := P(x_s > b_n e^u, x_t > b_n e^v) \quad \text{and} \quad P_n(u) := P(x_t > b_n e^u)
\]

\[
\Delta P_{n,s,t}(u,v) := P_{n,s,t}(u,v) - P_n(u) P_n(v).
\]
Use $E[(\ln x_t/b_n(m_n))_+] = \int_0^\infty P_n(u) \, du$ and $E[(\ln x_s/b_n(m_n))_+ (\ln x_t/b_n(m_n))_+]$

$= \int_0^\infty \int_0^\infty P_{n,s,t}(u,v) \, du \, dv$ to deduce

$$R_n = \frac{1}{m_n} \sum_{s \neq t}^n \int_0^\infty \int_0^\infty \Delta P_{n,s,t}(u,v) \, du \, dv$$

$$\quad - \alpha^{-1} \times \frac{1}{m_n} \sum_{s \neq t}^n \int_0^\infty \Delta P_{n,s,t} \left( u, v/m_{1/2} \right) \, du$$

$$\quad - \alpha^{-1} \times \frac{1}{m_n} \sum_{s \neq t}^n \int_0^\infty \Delta P_{n,s,t} \left( u/m_{1/2}, v \right) \, dv$$

$$\quad + \alpha^{-2} \times \frac{1}{m_n} \sum_{s \neq t}^n \Delta P_{n,s,t} \left( u/m_{1/2}, v/m_{1/2} \right).$$

An analysis of the asymptotic variance of $\hat{\alpha}_{m_n}^{-1}$ is ipso facto an analysis of the joint and marginal tail probabilities, in particular tail dependence measured by $\Delta P_{n,s,t}(u,v)$. If $\{x_t\}$ is iid then $\Delta P_{n,s,t}(u,v) = 0$ so $R_n = 0$ is trivial, hence $r_{m_n}^2 \to \alpha^{-2}$ as in Hall (1982). But if $\{x_t\}$ is any process, possibly tail dependent, with an asymptotically linearly independent tail array $\{U_{m_n,t} - \alpha^{-1}U_{m_n,t}(u/m_{1/2})\}$ then $R_n \to 0$. Stochastic volatility $\{x_t\}$ has this property because $\sum_{h=1}^\infty r(h) < \infty$, as the proof of the next claim shows.

**LEMMA 3.5** Under the conditions of Lemma 3.2 $R_n \to 0$.

### 3.3 Hill Estimator Asymptotics

Lemmas 3.2 and 3.5 permit a complete characterization of the Hill (1975) estimator’s limit distribution. Under the conditions of Lemma 3.2 all conditions for consistency $\hat{\alpha}_{m_n} \overset{p}{\to} \alpha$ and asymptotic normality $m_{1/2}^{-1}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/v_{m_n} \overset{d}{\to} N(0,1)$ are satisfied, cf. Theorems 1 and 2 of Hill (2010). Together with Lemma 3.5 this proves the following claim.

**THEOREM 3.6** Under the conditions of Lemma 3.2 $\hat{\alpha}_{m_n} \overset{p}{\to} \alpha$ and $m_{1/2}^{-1}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/v_{m_n} \overset{d}{\to} N(0,1)$.

### 3.4 Examples

A variety of volatility data generating processes satisfy the conditions of Lemma 3.2. We present two here. Consult Hill (2008, 2009, 2010) for numerous other examples.
EXAMPLE 1 (AR-GARCH): Consider an AR(1)-GARCH(1,1) volatility process

\[ \sigma_t = \beta + \phi \sigma_{t-1} + \epsilon_t, \quad \epsilon_t = h_t \eta_t, \quad \beta > 0, \quad 0 \leq \phi < 1 \]

\[ \eta_t \text{ is iid, } P(\eta_t > 0) = 1, \quad E[\eta_t^2] < \infty \forall r > 0 \]

\[ h_t^2 = \omega + \gamma u_{t-1}^2 + \delta h_{t-1}^2, \quad \omega > 0, \quad \gamma, \delta \geq 0. \]

If \( \phi > 0 \) and \( \gamma + \delta < 1 \), or \( \phi = 0 \) and \( \gamma + \delta \leq 1 \), then \( \{\sigma_t\} \) is geometrically \( \alpha \)-mixing (Boussama 1998, Carrasco and Chen 2002, Meitz and Saikkonen 2008), hence \( \{\sigma_t\} \) is \( L_p\)-NED on itself as a geometrically \( \alpha \)-mixing base for any \( p > 0 \).

EXAMPLE 2 (Log-ARFIMA): Let \( L \) denote the lag operator. If \((1 - L) \ln \sigma_t \) for \( d < 1/2 \) is causal ARMA\((p, q)\) with innovations \( \eta_t \overset{iid}{\sim} N(0,1) \) then \( \ln \sigma_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i} \), and \( \psi_i = o(\rho^i) \) if \( d = 0 \) or \( \psi_i = o(i^{-1/2}) \) if \( d \in (0, 1/2) \). It is easy to show by Gaussianity \( \{\ln \sigma_t\} \) is \( L_p\)-NED on \( \{\eta_t\} \) with size \( 1 - d > 1/2 \) for any \( p > 0 \). Further, \( \eta_t \) is \( L_p\)-bounded for all \( p > 0 \) hence \( \{\sigma_t\} \) is \( L_p\)-NED on \( \{\eta_t\} \) with size \( 1/2 \) for any \( p > 0 \) by repeated application of Theorem 17.16 of Davidson (1994).

4. SIMULATION STUDY

We now perform a brief simulation study to assess the degree to which the mean-squared-error \( \hat{v}_{m_n}^2 \) of the Hill-estimator \( \hat{\alpha}_{m_n}^{-1} \) is adequately approximated by \( \alpha^{-2} \) for SV. We generate 100 series of

\[ X_t = \sigma_{t-1} \epsilon_t, \quad \ln \sigma_t = \phi \ln \sigma_{t-1} + \eta_t, \quad \eta_t \overset{iid}{\sim} N(0,1), \]

where \( \epsilon_t \) are iid, zero mean, symmetric, unit-scaled \( \alpha \)-stable distributed with index \( \alpha \in \{1.25, 1.60, 1.95\} \). The sample size is \( n = 1000 \). Since the index \( \alpha \) is the same for any \( \phi \in (-1, 1) \), we randomly draw \( \phi \) from a uniform distribution on \([-1, 1]\) for each series. Define \( x_t := |X_t| \), and let \( \hat{\sigma}_{m_n} \) be computed from \( x_t \). By Example 2 the data generating process satisfies all conditions of Lemma 3.2.

Apply Theorem 3.6 and the mean-value-theorem to deduce \( m_n^{1/2}(\hat{\alpha}_{m_n} - \alpha)/\alpha \overset{d}{\rightarrow} N(0,1) \).

Although theory reveals \( \hat{v}_{m_n}^2 \rightarrow \alpha^{-2} \) we do not have an expression for the rate of convergence nor how \( \hat{v}_{m_n}^2 \approx \alpha^{-2} \) relates to particular valid sequences \( \{m_n\} \), hence the small sample approximation need not be sharp. We therefore initially compute confidence bands using Hill’s (2010) nonparametric estimator \( \hat{v}_{m_n}^2 \) of the mean-squared-error \( \hat{v}_{m_n}^2 \) :
where \( k_{n,s,t} = (1 - |s - t|/\gamma_n)_+ \) is a Bartlett kernel with bandwidth \( \gamma_n = n^{-2/5} \).

By Lemma 3.2 the conditions for \(|\hat{\sigma}^2_n - \nu^2_{\text{m}}| \xrightarrow{p} 0\) are satisfied by Theorem 3 of Hill (2010). Thus, by the mean-value-theorem robust 95\% asymptotic bands are \( \hat{\alpha}_m \pm 1.96 \hat{\nu}_m \hat{\sigma}^2_m / m^{1/2} \).

Since \( \alpha \) is known, we assess the validity of the limit \( m_n^{1/2}(\hat{\alpha}_m - \alpha)/\alpha \xrightarrow{d} N(0,1) \) by performing Kolmogorov-Smirnov [KS] tests on

\[
z_{m} := m_n^{1/2}(\hat{\alpha}_m - \alpha)/\alpha.
\]

We generate 100 \( z'_{m,n} \),s and perform the KS test of standard normality for each fractile \( m_n \).

Consult Figure 1 for results concerning \( \alpha = 1.25 \) where we plot 95\% bands and KS statistics over the fractile window \( m_n \in \{15, ..., 465\} \). The remaining \( \alpha \in \{1.60, 1.95\} \) lead to essentially identical results. The minimum KS is below the 5\% critical value, and the KS statistic is minimized essentially at that order \( m_n \) that satisfies \( \hat{\alpha}_m \approx \alpha \). In particular \( \alpha \) is essentially in the middle of the robust 95\% band. This provides strong experimental evidence both that \( z_{m,n} \) is approximately normal and the mean-squared-error of \( \hat{\alpha}_m \) is approximately \( \alpha^2 \).

**APPENDIX A: PROOFS OF MAIN RESULTS**

**Proof of Theorem 2.1.** Let \( \sigma_t \) be the diagonal entries of \( \Sigma_t \), \( \sigma_t = [\sigma_{i,t}]_{i=1}^k \), let \( \sigma^*_t \) denote a random draw from the distribution governing \( \sigma_t \), define the sigma fields

\[
\mathfrak{F}_t := \sigma(\epsilon_t, \epsilon_{t-1}, \ldots) \quad \text{and} \quad G_t := \sigma(\sigma_s : s \leq t)
\]

and

\[
H_t := \sigma(\mathfrak{F}_t \cup G_t).
\]

Regular variation (8) implies

\[
P(\epsilon_{j,t} > z_j / \sigma^*_{j,t-1}) \sim (\sigma^*_{j,t-1})^{\alpha_j} \times P(\epsilon_{j,t} > z_j).
\]

Consider any \( t \)-tuple \( \{i_1, \ldots, i_l\} \), \( i_j \in \{1, \ldots, k\} \), and \( \{t_1 > \cdots > t_l\} \), \( 1 \leq l \leq k \). If \( i_{j_1} \neq i_{j_2} \) for \( j_1 \neq j_2 \) then \( t_{j_1} < t_{j_2} \). Write

\[
I_{j_1, \ldots, j_l}(z) := I(\epsilon_{i_{j_1}, t_{j_1}} \sigma_{i_{j_1}, t_{j_1}-1} > z_{j_1}, \cdots, \epsilon_{i_{j_l}, t_{j_l}} \sigma_{i_{j_l}, t_{j_l}-1} > z_{j_l}).
\]

Since \( \{\epsilon_t\}_{t \in \mathbb{Z}} \) is independent of \( \{\sigma_t\}_{t \in \mathbb{Z}} \), and \( \epsilon_t \) is asymptotically independent (10), as \( \{z\} \to \infty \)

\[
P(\epsilon_{i_1, t_1} \sigma_{i_1, t_1-1} > z_1, \cdots, \epsilon_{i_l, t_l} \sigma_{i_l, t_l-1} > z_l) \sim E[I_{2, \ldots, l}(z) \times P(\epsilon_{i_1, t_1} \sigma_{i_1, t_1-1} > z_1|H_{t_1-1})]
\]

\[
= E[I_{2, \ldots, l}(z) \times P(\epsilon_{i_1, t_1} > z_1 / \sigma^*_{i_1, t_1-1})]
\]

\[
\sim P(\epsilon_{i_1, t_1} > z_1) \times E[I_{2, \ldots, l}(z) \times \sigma^*_{i_1, t_1-1}].
\]
Now iterate the above argument on each resulting $I_{j_1, \ldots, j_l}(z)$-term. For $E[I_{2, \ldots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_i}]$ it similarly follows

$$E[I_{2, \ldots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_i}]$$

$$\sim E\{I_{3, \ldots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_i} \times E[I(\epsilon_{i_2, t_2} \sigma_{i_2, t_2-1} > z_2) \mid \sigma(H_{t_1-1} \cup G_{t_2-1})]\}$$

$$= E[I_{3, \ldots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_i} \times P(\epsilon_{i_2, t_2} > z_2/\sigma_{i_2, t_2-1}^{\alpha_i})]$$

$$\sim P(\epsilon_{i_2, t_2} > z_2) E[I_{3, \ldots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_i} \sigma_{i_2, t_2-1}^{\alpha_i}],$$

and so on. The iteration ends with

$$P(\epsilon_{i_1, t_1}, \sigma_{i_1, t_1-1} > z_1, \ldots, \epsilon_{i_l, t_l} \sigma_{i_l, t_l-1} > z_l) \sim E\left[\prod_{j=1}^{l} \sigma_{i_j, t_j-1}^{\alpha_i}\right] \times \prod_{j=1}^{l} P(\epsilon_{i_j, t_j} > z_j).$$

(20)

Finally, (8)-(10) and (20) imply as \{z\} \to \infty

$$E\left[\prod_{j=1}^{l} \sigma_{i_j, t_j-1}^{\alpha_i}\right] \prod_{j=1}^{l} P(\epsilon_{i_j, t_j} > z_j) = E\left[\prod_{j=1}^{l} \sigma_{i_j, t_j-1}^{\alpha_i}\right] \times \prod_{j=1}^{l} E\left[\sigma_{i_j, t_j-1}^{\alpha_i}\right] \times P(\epsilon_{i_j, t_j} > z_j)$$

$$\sim E\left[\prod_{j=1}^{l} \sigma_{i_j, t_j-1}^{\alpha_i}\right] \times \prod_{j=1}^{l} P(\epsilon_{i_j, t_j} \sigma_{i_j, t_j-1} > z_j).$$

This completes the proof. QED.

**Proof of Corollary 2.3.** The claim follows instantly from Theorem 2.1 and the construction of $r_{i, j}(h)$. QED.

**Proof of Lemma 3.1.** Independence of the error and slow variation with remainder (13)-(17) imply (see Hsing 1991: p. 1552-1553)

$$P(\epsilon_t > \tilde{b}_n/\sigma_{t-1}^*) = (\sigma_{t-1}^*)^{\alpha_i} \times P(\epsilon_t > \tilde{b}_n) \times (1 + o(1/\tilde{m}_n^{1/2}))$$

$$= (\sigma_{t-1}^*)^{\alpha_i} \times (\tilde{m}_n/n) \times (1 + o(1/\tilde{m}_n^{1/2})).$$

17
and

\[ P(\epsilon_s \sigma_{s-1} > \tilde{b}_n, \epsilon_{t-1} \sigma_{t-1} > \hat{b}_n) \]

\[ = E\left[I(x_s > \hat{b}_n) \times P(\epsilon_t > \tilde{b}_n/\sigma_{t-1}^*)\right] \]

\[ = P(\epsilon_t > \tilde{b}_n) \times \left(1 + o(1/\tilde{m}_n^{1/2})\right) \times E\left\{\alpha_{t-1} \times P(\epsilon_s > \tilde{b}_n/\sigma_{s-1}^*)\right\} \]

\[ = P(\epsilon_s > \tilde{b}_n) P(\epsilon_t > \tilde{b}_n) \times \left(1 + o(1/\tilde{m}_n^{1/2})\right) \times E\left[\sigma_{t-1}^\alpha \sigma_{s-1}^\alpha\right] \]

\[ = (\tilde{m}_n/n)^2 \times \left(1 + o(1/\tilde{m}_n^{1/2})\right) \times E\left[\sigma_{t-1}^\alpha \sigma_{s-1}^\alpha\right] \]

Therefore

\[ \frac{P(x_{t-h} > \tilde{b}_n, x_t > \tilde{b}_n)}{P(x_{t-h} > \tilde{b}_n) P(x_t > \tilde{b}_n)} = \frac{E[\sigma_{t-1}^\alpha \sigma_{t-h-1}^\alpha]}{E[\sigma_{t-1}^\alpha] E[\sigma_{t-h-1}^\alpha]} + o(1/\tilde{m}_n^{1/2}), \]

and hence by construction and stationarity \((n/\tilde{m}_n)\tilde{r}_n(h) = r(h) + o(1/\tilde{m}_n^{1/2})\).

QED.

**Proof of Lemma 3.2.** Define \(\sigma\)-fields

\[ G_a^b := \sigma\{\eta_t : a \leq t \leq b\} \quad \text{and} \quad H_a^b := \sigma\{\epsilon_t, \eta_t : a \leq t \leq b\} \]

and note \(E[x_t|H_{t-1}^{t+h}] = \epsilon_t E[\sigma_{t-1}|G_{t-1}^{t+h}]\). Consider the case where \(\epsilon_t\) is \(L^p\)-bounded, \(p \leq 2\), the case \(p > 2\) being similar.

Use independence between \(\{\epsilon_t\}\) and \(\{\sigma_t\}\), the supposition that \(\{\sigma_t\}\) is \(L^\alpha\)-NED on \(\eta_t\) with size \(2/\min\{p,1\}\), \(2\alpha \geq p\) and Lyapunov’s inequality to deduce

\[ \|x_t - E[x_t|H_{t-1}^{t+h}]\|_p \leq \|\epsilon_t\|_p \times \|\sigma_{t-1} - E[\sigma_{t-1}|G_{t-1}^{t+h}]\|_2 \leq \bar{d}_t \varphi_t, \]

where \(\bar{d}_t := \|\epsilon_t\|_p \times d_t, \varphi_t = o(t^{-\lambda})\) and \(\lambda = 2/\min\{p,1\}\). By definition, therefore, \(\{x_t\}\) is \(L^p\)-NED on \(\{H_t\}\) with size \(\lambda\). Hence \(\{x_t\}\) is \(L^2\)-NED on \(\{H_t\}\) with constants \(d_{n,t}(u) = K(m_n/n)^{1/2}e^{-u^{2p}/2}\) and coefficients \(\varphi_t^{\min\{p,1\}/4}\) of size \(\lambda\min\{p,1\}/4 = 1/2\) by Theorem 2.1 of Hill (2008), a generalization of Lemma 4 of Hill (2010). The remaining claim follows by multiplying out the \(L^2\)-NED quadratic criterion. QED.

**Proof of Lemma 3.3.** Let \(\gamma(h)\) denote the autocovariance function of \(\{\sigma_t^\alpha\}\) so that by construction

\[ \gamma(h) = r(h) \times (E[\sigma_t^\alpha])^2. \]

Under the conditions of Lemma 3.2 it is easy to show \(\{\sigma_t^\alpha\}\) is \(L^2\)-NED with uniformly bounded constants and size \(2/\min\{p,1\} > 1/2\) on \(\alpha\)-mixing \(\{\eta_t\}\).
with coefficients $\alpha_t = o(t^{-1})$ (e.g. Davidson 1994: Theorem 17.16). Therefore \{\sigma_t^u, \sigma_t^v\} forms an $L_2$-mixingale sequence with size $1/2$ and uniformly bounded constants (Davidson 1994: Theorem 17.5). McLeish’s (1975: Theorem 1.6) maximal inequality and $|r(0)| \leq E[\sigma_0^2]/(E[\sigma_0^2])^2 + 1 < K$ therefore imply for all $n$

$$E \left( \sum_{t=1}^{n} \sigma_t^u \right)^2 = K \sum_{s,t=1}^{n} \gamma(s-t) = K \sum_{s,t=1}^{n} r(s-t) = Knr(0) + 2K \sum_{i=1}^{n-1} (n-i) \times r(i) \leq Kn.$$ 

Therefore $|\sum_{i=1}^{n} r(i)| = O(1)$ hence $|\sum_{i=1}^{\infty} r(i)| < \infty$ as claimed. QED.

**Proof of Lemma 3.5.** By Lemma 3.1 we may write

$$\Delta P_{n,s,t}(u,v) = (m_n/n)^2 r(|s-t|) \times e^{-\alpha(u+v)} \left(1 + o \left(1/m_n^{1/2} \right) \right)$$

where the $o(\cdot)$-term is not a function of $\{s,t\}$. Each component of $R_n$ in (19) reduces to

$$\frac{1}{m_n} \sum_{s \neq t}^{n} \int_{0}^{\infty} \int_{0}^{\infty} \Delta P_{n,s,t}(u,v) \, du \, dv$$

$$= \left(\frac{m_n}{n} \right)^2 \frac{1}{m_n} \sum_{s \neq t}^{n} r(|s-t|) \left(\int_{0}^{\infty} \int_{0}^{\infty} e^{-\alpha(u+v)} \, du \, dv \right) \left(1 + o \left(1/m_n^{1/2} \right) \right)$$

$$= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s-t|) \left(1 + o \left(1/m_n^{1/2} \right) \right),$$

and

$$\alpha^{-1} \frac{1}{m_n} \sum_{s \neq t}^{n} \int_{0}^{\infty} \Delta P_{n,s,t}(u,v/m_n^{1/2}) \, du$$

$$= \alpha^{-1} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s-t|) \left(\int_{0}^{\infty} e^{-\alpha(u+v/m_n^{1/2})} \, du \right) \left(1 + o \left(1/m_n^{1/2} \right) \right)$$

$$= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s-t|) e^{-\alpha v/m_n^{1/2}} \left(1 + o \left(1/m_n^{1/2} \right) \right),$$

and

$$\alpha^{-2} \frac{1}{m_n} \sum_{s \neq t}^{n} \Delta P_{n,s,t}(u/m_n^{1/2}, v/m_n^{1/2})$$

$$= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s-t|) e^{-\alpha(u/m_n^{1/2}+v/m_n^{1/2})} \left(1 + o \left(1/m_n^{1/2} \right) \right).$$
Therefore,

\[
R_n = \frac{\alpha^{-2}m_n}{n^2} \sum_{s \neq t}^{n} r(|s - t|) \left(1 + o\left(1/m_n^{1/2}\right)\right)
\]

\[
- \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s - t|) e^{-\alpha u/m_n^{1/2}} \left(1 + o\left(1/m_n^{1/2}\right)\right)
\]

\[
- \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s - t|) e^{-\alpha u/m_n^{1/2}} \left(1 + o\left(1/m_n^{1/2}\right)\right)
\]

\[
+ \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^{n} r(|s - t|) e^{-\alpha (u/m_n^{1/2} + v/m_n^{1/2})} \left(1 + o\left(1/m_n^{1/2}\right)\right)
\]

\[
= \frac{\alpha^{-2}m_n}{n^2} \sum_{s \neq t}^{n} r(|s - t|) \left(1 + o\left(1/m_n^{1/2}\right)\right)
\]

\[
\times \left(1 - e^{-\alpha u/m_n^{1/2}} - e^{-\alpha v/m_n^{1/2}} + e^{-\alpha (u/m_n^{1/2} + v/m_n^{1/2})}\right)
\]

\[
= \frac{\alpha^{-2}m_n^{1/2}}{n^2} \sum_{s \neq t}^{n} r(|s - t|) \left(1 + o\left(1/m_n^{1/2}\right)\right) \times O(1) = o(1),
\]

where the last line follows from Lemma 3.3. QED.

**Appendix B**

**LEMMA B.1** Under (16) and (17) for any intermediate order \{m_n\}, \((n/m_n)E(U_{m_n,t}) - \alpha^{-1}U_{m_n,t}^* (u/m_n^{1/2})^2 = \alpha^2 + o(1)\) where \(o(1)\) does not depend on \(t\).

**Proof.** Slow variation with remainder and bounded increase (16)-(17) imply (see Hsing 1991: p. 1548, 1553-1554)

\[
P(x_t > b_n) = (m_n/n) \times \left(1 + o(1/m_n^{1/2})\right)
\]

\[
P(x_t > b_n) = \lambda^{-\alpha} (m_n/n) \times \left(1 + o(1/m_n^{1/2})\right)
\]

\[
E \left(\ln x_t b_n\right)^k = \alpha^{-k} k! (m_n/n) \times \left(1 + o(1/m_n^{1/2})\right),
\]

where the \(o(\cdot)\) terms do not depend on \(t\).
By construction, therefore,
\[
(n/m_n) E \left( U_{m,t}^* (u/m_n^{1/2})^2 \right) = (n/m_n) \left( P \left( x_t > b_n e^{u/m_n^{1/2}} \right) - P \left( x_t > b_n e^{u/m_n^{1/2}} \right)^2 \right) = e^{-\alpha u/m_n^{1/2}} \times (n/m_n) P \left( x_t > b_n \right) \times \left( 1 + o(1/m_n^{1/2}) \right) = \alpha^{-1} + O(1/m_n^{1/2}).
\]
Similarly
\[
(n/m_n) E \left( U_{m,t}^{2} \right) = (n/m_n) E \left( \ln x_t b_n \right)^2 - (n/m_n) \left( E \left( \ln x_t b_n \right) \right)^2 = (2\alpha^{-2} - (m_n/n) \alpha^{-2}) \times \left( 1 + o(1/m_n^{1/2}) \right) = 2\alpha^{-2} + o(1/m_n^{1/2}) + O(m_n/n),
\]
and
\[
(n/m_n) E \left( U_{m,t}^* U_{m,t}^* (u/m_n^{1/2}) \right) = (n/m_n) E \left( I \left( x_t > b_n e^{u/m_n^{1/2}} \right) \left( \ln x_t b_n \right)^2 \right) - (n/m_n) P \left( x_t > b_n e^{u/m_n^{1/2}} \right) E \left( \ln x_t b_n \right) = (n/m_n) \int_0^\infty P \left( x_t > b_n \max \left\{ e^{u/m_n^{1/2}}, e^{\alpha} \right\} \right) dv = (n/m_n) \int_0^\infty P \left( x_t > b_n \max \left\{ e^{u/m_n^{1/2}}, e^{\alpha} \right\} \right) dv = (1 + o(1/m_n^{1/2})) \times \int_0^\infty \min \left\{ e^{-\alpha u/m_n^{1/2}}, e^{-\alpha \alpha} \right\} dv = \alpha^{-1} + O(1/m_n^{1/2}).
\]
Therefore
\[
(n/m_n) E \left( U_{m,t} - \alpha^{-1} U_{m,t}^* (u/m_n^{1/2}) \right)^2 = 2\alpha^{-2} - 2\alpha^{-2} + \alpha^{-2} + o(1) = \alpha^{-2} + o(1).
\]
where o(1) does not depend on t. \textbf{QED.}

**REFERENCES**


**Figure 1:** $\alpha = 1.25$

Insert Figure 1 - Top Panel

Insert Figure 1 - Bottom Panel

Notes:  

a. The KS {1%, 5%, 10%} critical values are {.107, .122, .136}.

b. The KS statistic is for a test of standard normality on $m_n^{1/2}(\hat{x}_{m_n} - 1.25)/1.25$.

c. The KS minimizing fractile $m^*$ is 81, and $\hat{\alpha}_{m^*} = 1.259$ with band (.98, 1.53).
Figure 1: Top Panel

Hill Estimator and Robust 95% Bands

fractile m