



Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Extremal memory of stochastic volatility with an application to tail shape inference

Jonathan B. Hill

Department of Economics, University of North Carolina-Chapel Hill, USA

ARTICLE INFO

Article history:

Received 14 July 2008

Received in revised form

10 July 2010

Accepted 13 July 2010

Available online 18 July 2010

Keywords:

Stochastic volatility

Tail dependence

Convolution tail

Hill estimator

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.

© 2010 Elsevier B.V. All rights reserved.

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 $\{\Sigma_t\}_{t \in \mathbb{Z}}$,

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

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

$$x_{i,t} = \sigma_{i,t} \varepsilon_{i,t}$$

for the sake of brevity, but all that follows extends to the general case. We are interested in an exact characterization of the marginal and joint tails

$$P(x_{i,t} > z_i) \quad \text{and} \quad P(x_{i,s} > z_i, x_{j,t} > z_j) \quad \text{as } z's \rightarrow \infty$$

E-mail address: jbhill@email.unc.eduURL: <http://www.unc.edu/~jbhill>

0378-3758/\$ - see front matter © 2010 Elsevier B.V. All rights reserved.

doi:10.1016/j.jspi.2010.07.007

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 (de Haan et al., 1989; Chernick et al., 1991; Basrak et al., 2002a, b; Mikosch and Stărică, 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 Stărică (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 $\{\varepsilon_t\}$ that have independent extremes exhibits tail dependence as a function of volatility dependence. In particular, we show in Corollary 2.3 of Section 2

$$\lim_{z \rightarrow \infty} \frac{P(x_{i,t-h} > z, x_{j,t} > z)}{P(x_{i,t-h} > z)P(x_{j,t} > z)} = 1 + r_{ij}(h) \tag{2}$$

for some function $r_{ij} : \mathbb{N} \rightarrow \mathbb{R}$ that decays according to linear dependence in $\{\sigma_{i,t}^{\alpha_i}, \sigma_{j,t-h}^{\alpha_j}\}$ where α_i are indices of regular variation. In the univariate case write $r(h) = r_{ii}(h)$. In order to address (2) we characterize multivariate product convolution tails for x_t 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. $\{\varepsilon_{i,t}\sigma_{i,t-1}\}$, and not the multivariate case, e.g. $\{\varepsilon_{i,s}\sigma_{i,s-1}, \varepsilon_{j,t}\sigma_{j,t-1}\}$. 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)

$$\tilde{r}_{ij}(h) := \lim_{z \rightarrow \infty} \frac{P(x_{i,t-h} > z, x_{j,t} > z) - P(x_{i,t-h} > z)P(x_{j,t} > z)}{P(x_{i,t-h} > z)^{1/2}P(x_{j,t} > z)^{1/2}}. \tag{3}$$

Stationary ARMA and GARCH processes exhibit $\tilde{r}_{ii}(h) \rightarrow 0$ as $h \rightarrow \infty$ at a geometric rate. If $\tilde{r}_{ij}(h) \neq 0$ then clearly limit (2) does not exist. Conversely, under (2) the extremogram for SV is simply $\tilde{r}_{ij}(h) = 0$ (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 $\tilde{r}_{ij}(h)$ only measures linear dependence in tail events $I(x_{i,t} > z)$ where $I(\cdot) \in \{0,1\}$. Thus, $\tilde{r}_{ij}(h) = 0$ 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 $z \rightarrow \infty$ are predictable in a non-trivial way even though they are asymptotically uncorrelated $\tilde{r}_{ij}(h) = 0$. This implies tail arrays of $\{x_t\}$ have Gaussian limits, including log-high threshold exceedances $\max\{0, \ln(|x_t|/z)\}$ as $z \rightarrow \infty$, and therefore Gaussian asymptotics are available for tail index and tail dependence estimators. See also Hill (forthcoming, 2009, in press).

Second, $r_{ij}(h)$ in (2) is non-trivially summable which ensures x_t is D' -mixing (Leadbetter's, 1974, 1983). Third, $r_{ij}(h)$ 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, in press).¹ See, e.g., Pickands (1975), Hall (1982), Davis and Resnick (1984), Smith (1987), Rootzén et al. (1990), Smith and Weissman (1994), Beirlant et al. (1996, 2005), de Haan and Resnick (1998), Resnick and Stărică (1995), Drees et al. (2004), and Hill (2009, in press), and see Hill (in press) 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 $x_t = \sigma_t \varepsilon_t$ is a scalar, and recall a GARCH(1,1) process (Bollerslev, 1986)

$$x_t = \sigma_t \varepsilon_t, \quad \text{where } \sigma_t^2 = \omega + \beta x_{t-1}^2 + \gamma \sigma_{t-1}^2, \quad \omega > 0, \beta, \gamma \geq 0. \tag{4}$$

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

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

$$\lim_{z \rightarrow \infty} P(x_t > z | x_{t-h} > z) = 0 \tag{5}$$

¹ Hsing (1991) offers some details for mixing sequences, but under fairly complicated assumptions on the limits of sample covariances of nonlinear tail arrays. Hill (in press) 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

$$\lim_{z \rightarrow \infty} P(x_t > z | x_{t-h} > z) = (1+r(h)) \times \lim_{z \rightarrow \infty} P(x_t > z) = 0,$$

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

$$P(x_{t-h}^* > z, x_t^* > z) = z^{-1/\eta_h} L_h(z) \quad \text{as } z \rightarrow \infty. \tag{6}$$

All values $\eta_h < 1$, or $\eta_h = 1$ with $\lim_{z \rightarrow \infty} L_h(z) = 0$, imply $P(x_t^* > z | x_{t-h}^* > z) \rightarrow 0$ hence “asymptotic independence”, and if $\eta_h = 1$ and $L_h(z) \rightarrow 0$ then $P(x_t^* > z | x_{t-h}^* > z) \rightarrow 0$ hence “asymptotic dependence” (Ledford and Tawn, 1996, 1997). See also Coles and Tawn (1994), Hefferman 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 $\eta_h < \frac{1}{2}$, $\eta_h = \frac{1}{2}$ and $\eta_h \in (\frac{1}{2}, 1)$ 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

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

if and only if $\eta_h = \frac{1}{2}$ and $L_h(z) \rightarrow 1 + r(h)$. 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)

$$\tilde{r}(h) := \lim_{z \rightarrow \infty} \frac{P(x_{t-h} > z, x_t > z)}{P(x_{t-h} > z)^{1/2} P(x_t > z)^{1/2}} = O(\rho^h) \quad \text{for } |\rho| < 1.$$

Under the Ledford–Tawn model (6) this implies asymptotic dependence $\eta_h = 1$ and $\lim_{z \rightarrow \infty} L_h(z) = \tilde{r}(h) = O(\rho^h)$ since

$$\tilde{r}(h) := \lim_{z \rightarrow \infty} \frac{P(x_{t-h}^* > z, x_t^* > z)}{P(x_{t-h}^* > z)^{1/2} P(x_t^* > z)^{1/2}} = \lim_{z \rightarrow \infty} z^{1-1/\eta_h} L_h(z) = O(\rho^h).$$

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 $\{x_t\}_{t=1}^n$ satisfies

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

if and only if $\{\check{b}_n\}$ satisfies $nP(|x_t| > \check{b}_n) \rightarrow 1$. See also Loynes (1965) and O'Brien (1974). The extremal index θ 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 $\tilde{r}(h) = 0$ only suggests extremes are linearly tail independent, while non-trivial predictability of extreme events based on volatility memory suggests nonlinear tail dependence. This has an important repercussion: 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 $\iota > 0$ is a tiny constant. $a_n \sim b_n$ implies $a_n/b_n \rightarrow 1$. $\|\cdot\|_p$ denotes the L_p -norm: $\|x\|_p := (E|x|^p)^{1/p}$. $\eta_t \stackrel{iid}{\sim} N(0, 1)$ states η_t is iid normally distributed with zero mean and unit variance.

2. Product convolution tails

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

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

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

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

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

Property (8) is encompassed by the 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 et al., 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):

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

If $\sigma_{i,t}$ has tail shape (8) with an index $\kappa_i \leq \alpha_i$ then $x_{i,t} = \varepsilon_{i,t} \sigma_{i,t-1}$ 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 $E[\sigma_{i,t}^{\alpha_i}] = \infty$.

2.1. Preliminary results

We assume ε_t satisfies tail independence by imposing a multivariate version of $r_{i,j}(h) = 0$ in (2). The point here is technically ε_t need not be independent over the entire support of its distribution, and non-extremes can exhibit any degree or form of dependence.

Assumption A. Each $\{\varepsilon_{i,t}\}$ is independent of every $\{\sigma_{i,t}\}$. Further

$$\frac{P(\varepsilon_{i_1,t_1} > z_1, \dots, \varepsilon_{i_l,t_l} > z_l)}{P(\varepsilon_{i_1,t_1} > z_1) \times \dots \times P(\varepsilon_{i_l,t_l} > z_l)} \rightarrow 1 \quad \text{as } z_l \rightarrow \infty \tag{10}$$

for all l -tuples $\{i_1, \dots, i_l\}$ and $\{t_1, \dots, t_l\}$ 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, $0 < E\left[\prod_{j=1}^l \sigma_{i_j,t_j-1}^{\alpha_{i_j}}\right] < \infty$.

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

Remark 2. The moment bound $E\left[\prod_{j=1}^l \sigma_{i_j,t_j-1}^{\alpha_{i_j}}\right] < \infty$ holds when $E[\sigma_{i,t}^{k\alpha_i}] < \infty$.

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 $\{i_1, \dots, i_l\}$ and $\{t_1, \dots, t_l\}$, $i_{j_1} \neq i_{j_2}$ for $j_1 \neq j_2$, $1 \leq l \leq k$, as $\{z\} \rightarrow \infty$

$$\begin{aligned} P(\varepsilon_{i_1,t_1} \sigma_{i_1,t_1-1} > z_1, \dots, \varepsilon_{i_l,t_l} \sigma_{i_l,t_l-1} > z_l) &\sim E\left[\prod_{j=1}^l \sigma_{i_j,t_j-1}^{\alpha_{i_j}}\right] \times P(\varepsilon_{i_1,t_1} > z_1, \dots, \varepsilon_{i_l,t_l} > z_l) \\ &\sim \frac{E\left[\prod_{j=1}^l \sigma_{i_j,t_j-1}^{\alpha_{i_j}}\right]}{\prod_{j=1}^l E[\sigma_{i_j,t_j-1}^{\alpha_{i_j}}]} \times \prod_{j=1}^l P(\varepsilon_{i_j,t_j} \sigma_{i_j,t_j-1} > z_j). \end{aligned}$$

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

$$P(x_{i,s} > z_i, x_t > z_j) \sim P(x_{i,s} > z_i)P(x_t > z_j) \quad \text{as } \{z_i, z_j\} \rightarrow \infty,$$

if and only if the bivariate volatility process $\{\sigma_{i,t}^{\alpha_i}, \sigma_{j,t}^{\alpha_j}\}$ is linearly independent:

$$E[\sigma_{i,s}^{\alpha_i} \sigma_{j,t}^{\alpha_j}] = E[\sigma_{i,s}^{\alpha_i}]E[\sigma_{j,t}^{\alpha_j}].$$

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

$$x_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \text{ is iid with tail (8) and index } \alpha > 0,$$

$$\ln \sigma_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}, \quad \psi_0 = 1, \quad \psi_i = o(i^{-1/2}), \quad \eta_t \stackrel{iid}{\sim} N(0,1), \tag{11}$$

hence $\sigma_t^\alpha = \prod_{i=0}^{\infty} e^{\alpha \psi_i \eta_{t-i}}$. Examples include a log-autoregression $\ln \sigma_t = \phi \ln \sigma_{t-1} + \eta_t$, $|\phi| < 1$ (Taylor, 1986; Breidt and Davis, 1998; Davis and Mikosch, 2009b) and log-ARFIMA(p,d,q) with Hurst index $d < \frac{1}{2}$. 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

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

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 σ_t exhibits geometric memory $\psi_i = O(\rho^i)$, $|\rho| < 1$, then

$$\lim_{z \rightarrow \infty} \frac{P(x_{t-h} > z, x_t > z)}{P(x_{t-h} > z)P(x_t > z)} = 1 + O(\rho^h) \quad \text{hence } r(h) = O(\rho^h).$$

Similarly, under hyperbolic memory $\psi_i = O(i^{-\mu})$, $\mu > \frac{1}{2}$ (e.g. ARFIMA(p, d, q), $d < \frac{1}{2}$), then $r(h) = O(h^{-\mu})$.

Now define

$$r_{i,j}(h) := \frac{E[(\sigma_{i,t}^{2\alpha_i} - E[\sigma_{i,t}^{2\alpha_i}])(\sigma_{j,t-h}^{2\alpha_j} - E[\sigma_{j,t-h}^{2\alpha_j}])]}{E[\sigma_{i,t}^{2\alpha_i}]E[\sigma_{j,t-h}^{2\alpha_j}]} \in (-\infty, \infty). \tag{12}$$

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

$$\lim_{z_1, z_2 \rightarrow \infty} \frac{P(x_{i,t-h} > z_1, x_{j,t} > z_2)}{P(x_{i,t-h} > z_1)P(x_{j,t} > z_2)} = 1 + r_{i,j}(h).$$

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 Ledford–Tawn model (6) applies with $\eta_h = \frac{1}{2}$ for all h and $L_h(z) \rightarrow 1 + r_{i,i}(h)$. 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

$$x_t = \sigma_t \varepsilon_t, \quad \{\varepsilon_t\} \text{ is independent of } \{\sigma_t\}. \tag{13}$$

Assume ε_t is iid L_p -bounded with tail

$$\bar{F}_\varepsilon(z) := P(\varepsilon_t > z) = z^{-\alpha} L_\varepsilon(z) \quad \text{as } z \rightarrow \infty,$$

where $L_\varepsilon(z)$ is slowly varying. Let $\bar{F}_\varepsilon(z)$ have support $[0, \infty)$, and assume $\{\sigma_t^2\}$ is a covariance stationary process as in Breidt and Davis (1998) and Davis and Mikosch (2009b, c). Then by (9) $P(x_t > z) = KP(\varepsilon_t > z)$, hence for slowly varying $L_x(z) = KL_\varepsilon(z)$

$$\bar{F}_x(z) := P(x_t > z) = z^{-\alpha} L_x(z) \quad \text{as } z \rightarrow \infty. \tag{14}$$

The Hill (1975) estimator of α^{-1} is the sample mean of peaks-over-thresholds

$$\hat{\alpha}_{m_n}^{-1} \equiv \frac{1}{m_n} \sum_{t=1}^n (\ln\{x_t/x_{(m_n+1)}\})_+ = \frac{1}{m_n} \sum_{i=1}^{m_n} \ln\{x_t/x_{(m_n+1)}\},$$

where n is the sample size, $x_{(i)}$ denotes the i th order statistic, $x_{(1)} \geq x_{(2)} \geq \dots$, and $\{m_n\} = \{m_n\}_{n \geq 1}$ is an intermediate order sequence as in Hall (1982): $1 \leq m_n < n$, $m_n \rightarrow \infty$ and $m_n = o(n)$. As long as $m_n \rightarrow \infty$ 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)

$$v_{m_n}^2 := E(m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}))^2 \rightarrow \alpha^{-2}.$$

See Section 3.3, below.

The estimator $\hat{\alpha}_{m_n}^{-1}$ is consistent for α^{-1} and asymptotically normal $m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/v_{m_n} \xrightarrow{d} N(0,1)$ for a vast array of dependent, heterogeneous processes $\{x_t\}$. See Hsing (1991), Resnick and Stărică (1995) and Hill (forthcoming, 2009, in press) and their citations. We first demonstrate x_t in (13) satisfies the requirements for asymptotic normality for a large class of volatility processes $\{\sigma_t\}$. Second, we show $v_{m_n}^2 \rightarrow \alpha^{-2}$ 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 θ in (7).

3.1. Tail structure

In order to demonstrate $\nu_{\check{m}_n}^2 \rightarrow \alpha^{-2}$ we use a decomposition of $\hat{\alpha}_{\check{m}_n}^{-1}$ that exploits the concept of *slow variation with remainder* proposed in Smith (1982) and Goldie and Smith (1987). See also Haeusler and Teugels (1985), Hsing (1991), de Haan and Resnick (1996), and Hill (in press).

Assume $\bar{F}_\varepsilon(z)/\bar{F}_\varepsilon(-z) \rightarrow 1$ so there exist sequences of positive numbers $\{\check{m}_n, \check{b}_n\}_{n \in \mathbb{N}}$ satisfying (Leadbetter et al., 1983, Theorem 1.7.13)

$$\frac{n}{\check{m}_n} P(x_t > \check{b}_n) \rightarrow 1. \tag{15}$$

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

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

$$L_x(\lambda z)/L_x(z) - 1 = O(g(z)) \quad \text{as } z \rightarrow \infty. \tag{16}$$

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, \check{m}_n and \check{b}_n satisfy

$$\check{m}_n^{1/2} \times g(\check{b}_n) \rightarrow 0. \tag{17}$$

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

Define

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

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

Lemma 3.1. Under (13)–(17) $(n/\check{m}_n)\tilde{r}_{\check{m}_n}(h) = r(h) + o(1/\check{m}_n^{1/2})$.

Remark 1. Although globally $\tilde{r}_{\check{m}_n}(h) \rightarrow 0$, SV exhibits $\tilde{r}_{\check{m}_n}(h) = (\check{m}_n/n)r(h) + o(\check{m}_n^{1/2}/n)$ 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/\check{m}_n) -tail event correlated”.

Remark 2. Locally correlated extreme events is similar in spirit to Ledford and Tawn's (2003) near independence ($\eta_h = \frac{1}{2}$ and $\lim_{z \rightarrow \infty} L_h(z) > 0$) and asymptotic independence with association of large values ($\eta_h \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$). 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 $m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/\nu_{m_n} \xrightarrow{d} N(0, 1)$ and $\nu_{m_n}^2 \rightarrow \alpha^{-2}$. 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 Stărică (1998), Drees et al. (2004) and Hill (in press) to name a few. In very few cases the asymptotic variance is characterizable 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.

3.2.1. 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 (in press) E-NED weak tail dependence property.

Let $\{\varepsilon_t\}$ be a possibly vector-valued stochastic process with σ -algebra

$$\mathfrak{F}_t := \sigma(\varepsilon_\tau : \tau \leq t) : \mathfrak{F}_a^b := \sigma(\varepsilon_t : a \leq t \leq b),$$

and let $\{m_n\}$ be an intermediate order sequence with associated threshold sequence $\{b_n\}$:

$$\frac{n}{m_n} \bar{F}_x(b_n) \rightarrow 1.$$

We say $\{x_t\}$ is L_p -E-NED on $\{\mathfrak{F}_t\}$ or on “base” $\{\varepsilon_t\}$, with “size” $\lambda > 0$, if for some sequence of integer displacements $\{l_n\}$, $l_n \rightarrow \infty$,

$$\|I(x_t > b_n e^u) - P(x_t > b_n e^u | \mathfrak{F}_{t-l_n}^{t+l_n})\|_p \leq d_{n,t}(u) \times \psi_{l_n},$$

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

We call $d_{n,t}(u)$ the “constants” and ψ_{l_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 $\psi_{l_n} = o(l_n^{-\lambda})$. 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 $n \rightarrow \infty$. The constants $d_{n,t}(u)$ capture such changes unrelated to memory. See Hill (forthcoming, 2009, in press) 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 $\{\mathfrak{F}_t\}$ or on $\{\varepsilon_t\}$ with size $\lambda > 0$ if for constants $\{d_t\}$ and coefficients $\{\varphi_l\}$, $l^{\lambda} \varphi_l \rightarrow 0$ as $l \rightarrow \infty$ (Gallant and White, 1988; Davidson, 1994; Nze and Doukhan, 2004):

$$\|z_t - E[z_t | \mathfrak{F}_{t-l}^{t+l}]\|_p \leq d_t \varphi_l.$$

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_q -E-NED $\forall p \geq q$ with adjustment to size (Hill, forthcoming, Theorem 2.1). Any process is L_p -NED or L_p -E-NED on itself $x_t = \varepsilon_t$ 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 $\eta_h = \frac{1}{2}$ and $L_h(z) \rightarrow 1 + r(h)$ (Ledford and Tawn, 2003), and exhibiting “no extremal clustering” a la the extremogram $\tilde{r}(h) = 0$ and extremal index $\theta = 1$ (Breidt and Davis, 1998; Davis and Mikosch, 2009b, c), $\{x_t\}$ is non-trivially L_2 -E-NED on $\{\eta_t, \varepsilon_t\}$ with size that depends on volatility memory. Recall the definition of α -mixing coefficients for a stationary sequence (Rosenblatt, 1956; Ibragimov and Linnik, 1971):

$$\alpha_l := \sup_{t \in \mathbb{Z}} \sup_{A \subset \mathfrak{F}_{-\infty}^t, B \subset \mathfrak{F}_{t+l}^{\infty}} |P(A \cap B) - P(A)P(B)|.$$

Lemma 3.2 (E-NED). *Let (13)–(17) hold where L_p -bounded iid ε_t has tail index $0 < \alpha < p$. If $\{\sigma_t\}$ is $L_{2\alpha}$ -NED with coefficients φ_l of size $2/\min\{p, 1\}$ on an α -mixing base $\{\eta_t\}$ with coefficients $\alpha_l = o(l^{-1})$, 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 $\varphi_{l_n}^{\min\{p, 1\}/4} = o(l_n^{-1/2})$ on an α -mixing base $\{\varepsilon_t, \eta_t\}$ with size 1. In particular L_2 -E-NED is equivalent to*

$$\left| \frac{n}{m_n} P(x_t > b_n e^u) - \frac{n}{m_n} E[P(x_t > b_n e^u | \mathfrak{F}_{t-l_n}^{t+l_n})^2] \right| \leq e^{-up} \times o(l_n^{-1/2}). \tag{18}$$

Remark 1. Clearly $\sigma_t = \eta_t$ is a special case, so NED covers α -, β -, ϕ -, ρ -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, forthcoming).

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 $\mathcal{P}_n(u) := (n/m_n)P(x_t > b_n e^u) \rightarrow e^{-\alpha u}$. The scaled minimum mean-squared-error tail event predictor $(n/m_n)P(x_t > b_n e^u | \mathfrak{F}_{t-l_n}^{t+l_n})$ 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” $\mathfrak{F}_{t-l_n}^{t+l_n}$ asymptotically contains sufficient information for perfect non-trivial extreme event prediction in mean-squared-error.

3.2.2. Summability

We use E-NED to prove consistency and asymptotic normality of $\hat{\alpha}_{m_n}^{-1}$ since the property suffices for tail array limit theory under very general conditions. In order to show $v_{m_n}^2 \rightarrow \alpha^{-2}$ we work with the coefficient $r(h)$ 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 $v_{m_n}^2 \rightarrow \alpha^{-2}$.

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 $\{\check{b}_n\}$ by $nP(x_t > \check{b}_n) \rightarrow 1$, and exploit Lemmas 3.1 and 3.3 to deduce $\{x_t\}$ in (13) satisfies

$$n \sum_{i=1}^{[n/m]} P(x_1 > \check{b}_n, x_{i+1} > \check{b}_n) = \frac{1}{n} \sum_{i=1}^{[n/m]} n^2 P(x_1 > \check{b}_n, x_{i+1} > \check{b}_n) = \frac{1}{n} \sum_{i=1}^{[n/m]} (1 + r(i) + o(1)) = \frac{1}{m} \times (1 + o(1)) + \frac{1}{n} \sum_{i=1}^{[n/m]} r(i).$$

Take $\limsup_{n \rightarrow \infty}$ on both sides, and then $\lim_{m \rightarrow \infty}$ to deduce the right hand side converges to zero. Thus $\{x_t\}$ is $D'(\check{b}_n)$ -mixing (Leadbetter, 1983, Eq. (1.3)).

Lemma 3.4 (D' -mixing). Under the conditions of Lemma 3.2 $\{x_t\}$ is $D'(\check{b}_n)$ -mixing.

Remark. The above identity shows $D'(\check{b}_n)$ -mixing does not require $|\sum_{h=1}^{\infty} r(h)| < \infty$, 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) $D(\check{b}_n)$ -mixing property. If, however, σ_t is α -mixing with size 1 then trivially x_t is $D(\check{b}_n)$ -mixing and by Lemma 3.4 x_t is $D'(\check{b}_n)$ -mixing. Together they ensure x_t belongs to the maximum domain of attraction (7). See Leadbetter (1974, Theorem 3.1).

3.2.3. 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 (Hsing, 1991, Theorem 2.4, Hill, 2009, Lemma A.7)

$$m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1}) = \frac{1}{m_n^{1/2}} \sum_{t=1}^n (U_{m_n,t} - \alpha^{-1} U_{m_n,t}^*(u/m_n^{1/2})) + o_p(1).$$

Use Lemma B.1 in Appendix B to deduce

$$\frac{1}{m_n} \sum_{t=1}^n E[(U_{m_n,t} - \alpha^{-1} U_{m_n,t}^*(u/m_n^{1/2}))^2] \rightarrow \alpha^{-2}.$$

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

$$\lim_{n \rightarrow \infty} |v_{m_n}^2 - \alpha^{-2} - R_n| = 0,$$

where

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

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 $R_n \rightarrow 0$ 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

$$R_n = \frac{1}{m_n} \sum_{s \neq t} \int_0^\infty \int_0^\infty \Delta P_{n,s,t}(u,v) du dv - \alpha^{-1} \times \frac{1}{m_n} \sum_{s \neq t} \int_0^\infty \Delta P_{n,s,t}(u, v/m_n^{1/2}) du - \alpha^{-1} \times \frac{1}{m_n} \sum_{s \neq t} \int_0^\infty \Delta P_{n,s,t}(u/m_n^{1/2}, v) dv + \alpha^{-2} \times \frac{1}{m_n} \sum_{s \neq t} \Delta P_{n,s,t}(u/m_n^{1/2}, v/m_n^{1/2}). \quad (19)$$

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 $v_{m_n}^2 \rightarrow \alpha^{-2}$ as in Hall (1982). But if $\{x_t\}$ is any process, possibly tail dependent, with an asymptotically linearly independent tail array $\{U_{m_n,t} - \alpha^{-1} U_{m_n,t}^*(u/m_n^{1/2})\}$ then $R_n \rightarrow 0$. 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 $R_n \rightarrow 0$.

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 $\hat{\alpha}_{m_n} \xrightarrow{p} \alpha$ and asymptotic normality $m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/\nu_{m_n} \xrightarrow{d} N(0,1)$ are satisfied, cf. Theorems 1 and 2 of Hill (in press). Together with Lemma 3.5 this proves the following claim.

Theorem 3.6. Under the conditions of Lemma 3.2 $\hat{\alpha}_{m_n} \xrightarrow{p} \alpha$ and $m_n^{1/2}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})/\alpha^{-2} \xrightarrow{d} N(0,1)$.

3.4. Examples

A variety of volatility data generating processes satisfy the conditions of Lemma 3.2. We present two here. Consult Hill (forthcoming, 2009, in press) for numerous other examples.

Example 1 (AR-GARCH). Consider an AR(1)-GARCH(1,1) volatility process

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

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

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

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

Example 2 (Log-ARFIMA). Let \mathcal{L} denote the lag operator. If $(1-\mathcal{L})^d \ln \sigma_t$ for $d < \frac{1}{2}$ is causal ARMA(p, q) with innovations $\eta_t \stackrel{iid}{\sim} N(0, 1)$ then $\ln \sigma_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}$ and $\psi_i = o(i^d)$ if $d=0$ or $\psi_i = o(i^{-1/2})$ if $d \in (0, \frac{1}{2})$. It is easy to show by Gaussianity $\{\ln \sigma_t\}$ is L_p -NED on $\{\eta_t\}$ with size $1-d > \frac{1}{2}$ for any $p > 0$. Further, η_t is L_p -bounded for all $p > 0$ hence $\{\sigma_t\}$ is L_p -NED on $\{\eta_t\}$ with size $\frac{1}{2}$ 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 $\nu_{m_n}^2$ of the Hill-estimator $\hat{\alpha}_{m_n}^{-1}$ is adequately approximated by α^{-2} for SV. We generate 100 series of

$$X_t = \sigma_{t-1}\varepsilon_t, \quad \ln \sigma_t = \phi \ln \sigma_{t-1} + \eta_t, \quad \text{and } \eta_t \stackrel{iid}{\sim} N(0, 1),$$

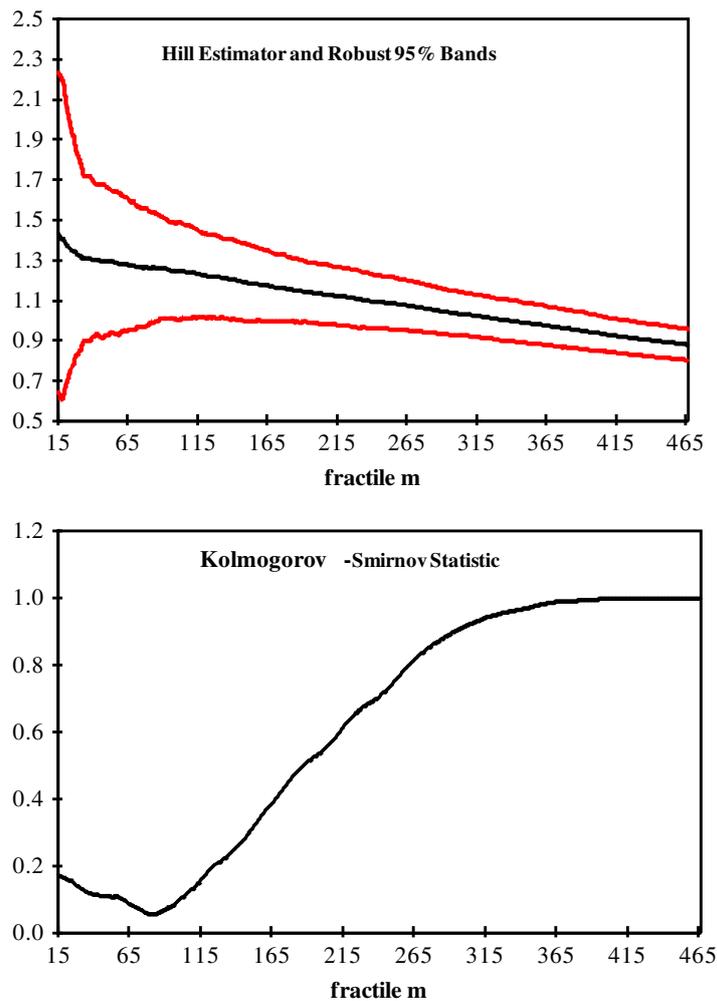


Fig. 1. $\alpha = 1.25$. Top panel: Hill estimator and robust 95% bands. Bottom panel: Kolmogorov–Smirnov statistic. Notes: (a) The KS {1%, 5%, 10%} critical values are {0.107, 0.122, 0.136}. (b) The KS statistic is for a test of standard normality on $m_n^{1/2}(\hat{\alpha}_{m_n} - 1.25)/1.25$. (c) The KS minimizing fractile m^* is 81, and $\hat{\alpha}_{m^*} = 1.259$ with band (0.98, 1.53).

where ε_t are iid, zero mean, symmetric, unit-scaled α -stable distributed with index $\alpha \in \{1.25, 1.60, 1.95\}$. The sample size is $n = 1000$. Since the index α is the same for any $\phi \in (-1, 1)$, we randomly draw ϕ from a uniform distribution on $[0.1, 0.9]$ for each series. Define $x_t := |X_t|$, and let $\hat{\alpha}_{m_n}$ 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 $m_n^{1/2}(\hat{\alpha}_{m_n} - \alpha)/\alpha \xrightarrow{d} N(0, 1)$.

Although theory reveals $v_{m_n}^2 \rightarrow \alpha^{-2}$ we do not have an expression for the rate of convergence nor how $v_{m_n}^2 \approx \alpha^{-2}$ 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 (in press) nonparametric estimator $\hat{v}_{m_n}^2$ of the mean-squared-error $v_{m_n}^2$:

$$\hat{v}_{m_n}^2 = \frac{1}{n} \sum_{s,t=1}^n k_{n,s,t} \left\{ \ln \left(\frac{x_s^{(a)}}{x_{(m_n+1)}^{(a)}} \right) - \frac{m_n}{n} \hat{\alpha}_{m_n} \right\} \times \left\{ \ln \left(\frac{x_t^{(a)}}{x_{(m_n+1)}^{(a)}} \right) - \frac{m_n}{n} \hat{\alpha}_{m_n} \right\},$$

where $k_{n,s,t} = (1 - |s-t|/\gamma_n)_+$ is a Bartlett kernel with bandwidth $\gamma_n = n^{225}$. By Lemma 3.2 the conditions for $|\hat{v}_{m_n}^2 - v_{m_n}^2| \xrightarrow{p} 0$ are satisfied by Theorem 3 of Hill (in press). Thus, by the mean-value-theorem robust 95% asymptotic bands are $\hat{\alpha}_{m_n} \pm 1.96 \hat{v}_{m_n} \hat{\alpha}_{m_n} / m_n^{1/2}$.

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

$$z_{m_n} := m_n^{1/2}(\hat{\alpha}_{m_n} - \alpha)/\alpha.$$

We generate 100 z_{m_n} 's and perform the KS test of standard normality for each fractile m_n .

Consult Fig. 1 for results concerning $\alpha = 1.25$ where we plot 95% bands and KS statistics over the fractile window $m_n \in \{15, \dots, 465\}$. 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 $\hat{\alpha}_{m_n} \approx \alpha$. In particular α is nearly 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 $\hat{\alpha}_{m_n}$ is approximately α^2 .

Acknowledgement

We thank a referee for helpful suggestions.

Appendix A. Proofs of main results

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

$$\mathfrak{F}_t := \sigma(\varepsilon_t, \varepsilon_{t-1}, \dots) \quad \text{and} \quad G_t := \sigma(\sigma_s : s \leq t)$$

and

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

Regular variation (8) implies

$$P(\varepsilon_{j,t} > z_j / \sigma_{j,t-1}^*) \sim (\sigma_{j,t-1}^*)^{\alpha_j} \times P(\varepsilon_{j,t} > z_j).$$

Consider any l -tuple $\{i_1, \dots, i_l\}$, $i_j \in \{1, \dots, k\}$, and $\{t_1 > \dots > t_l\}$, $1 \leq l \leq k$. If $i_j \neq i_{j_2}$ for $j_1 \neq j_2$ then $t_{j_1} \geq t_{j_2}$. Write

$$I_{j_1, \dots, j_l}(z) := I(\varepsilon_{i_1, t_1} \sigma_{i_1, t_1-1} > z_{j_1}, \dots, \varepsilon_{i_l, t_l} \sigma_{i_l, t_l-1} > z_{j_l}).$$

Since $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is independent of $\{\sigma_t\}_{t \in \mathbb{Z}}$, and ε_t is asymptotically independent (10), as $\{z\} \rightarrow \infty$

$$\begin{aligned} P(\varepsilon_{i_1, t_1} \sigma_{i_1, t_1-1} > z_1, \dots, \varepsilon_{i_l, t_l} \sigma_{i_l, t_l-1} > z_l) &\sim E[I_{2, \dots, l}(z) \times P(\varepsilon_{i_1, t_1} \sigma_{i_1, t_1-1} > z_1 | H_{t_1-1})] \\ &= E[I_{2, \dots, l}(z) \times P(\varepsilon_{i_1, t_1} > z_1 / \sigma_{i_1, t_1-1}^*)] \sim P(\varepsilon_{i_1, t_1} > z_1) \times E[I_{2, \dots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_{i_1}}]. \end{aligned}$$

Now iterate the above argument on each resulting $I_{j_1, \dots, j_l}(z)$ -term. For $E[I_{2, \dots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_{i_1}}]$ it similarly follows

$$\begin{aligned} E[I_{2, \dots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_{i_1}}] &\sim E[I_{3, \dots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_{i_1}} \times E[I(\varepsilon_{i_2, t_2} \sigma_{i_2, t_2-1} > z_2) | \sigma(H_{t_1-1} \cup G_{t_2-1})]] \\ &= E[I_{3, \dots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_{i_1}} \times P(\varepsilon_{i_2, t_2} > z_2 / \sigma_{i_2, t_2-1}^*)] \sim P(\varepsilon_{i_2, t_2} > z_2) E[I_{3, \dots, l}(z) \times \sigma_{i_1, t_1-1}^{\alpha_{i_1}} \sigma_{i_2, t_2-1}^{\alpha_{i_2}}], \end{aligned}$$

and so on. The iteration ends with

$$P(\varepsilon_{i_1, t_1} \sigma_{i_1, t_1-1} > z_1, \dots, \varepsilon_{i_l, t_l} \sigma_{i_l, t_l-1} > z_l) \sim E \left[\prod_{j=1}^l \sigma_{i_j, t_j-1}^{\alpha_{i_j}} \right] \times \prod_{j=1}^l P(\varepsilon_{i_j, t_j} > z_j). \tag{20}$$

Finally, (8)–(10) and (20) imply as $\{z\} \rightarrow \infty$

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

This completes the proof. \square

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

Proof of Lemma 3.1. Independence of the error and slow variation with remainder (13)–(17) imply (see Hsing, 1991, pp. 1552–1553)

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

and

$$\begin{aligned} P(\varepsilon_s \sigma_{s-1} > \check{b}_n, \varepsilon_t \sigma_{t-1} > \check{b}_n) &= E[I(x_s > \check{b}_n) \times P(\varepsilon_t > \check{b}_n / \sigma_{t-1}^*)] \\ &= P(\varepsilon_t > \check{b}_n) \times (1 + o(1/\check{m}_n^{1/2})) \times E\{\sigma_{t-1}^{\alpha_i} \times P(\varepsilon_s > \check{b}_n / \sigma_{s-1}^*)\} \\ &= P(\varepsilon_s > \check{b}_n) P(\varepsilon_t > \check{b}_n) \times (1 + o(1/\check{m}_n^{1/2})) \times E[\sigma_{t-1}^{\alpha_i} \sigma_{s-1}^{\alpha_j}] \\ &= (\check{m}_n/n)^2 \times (1 + o(1/\check{m}_n^{1/2})) \times E[\sigma_{t-1}^{\alpha_i} \sigma_{s-1}^{\alpha_j}]. \end{aligned}$$

Therefore

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

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

Proof of Lemma 3.2. Define σ -fields

$$G_a^b := \sigma(\{\eta_t\} : a \leq t \leq b) \quad \text{and} \quad H_a^b := \sigma(\{\varepsilon_t, \eta_t\} : a \leq t \leq b)$$

and note $E[x_t | H_{t-1}^{t+1}] = \varepsilon_t E[\sigma_{t-1} | G_{t-1}^{t+1}]$. Consider the case where ε_t is L_p -bounded, $p \leq 2$, the case $p > 2$ being similar.

Use independence between $\{\varepsilon_t\}$ and $\{\sigma_t\}$, the supposition that $\{\sigma_t\}$ is L_q -NED on $\{\eta_t\}$ with size $2/\min\{p, 1\}$, $2\alpha \geq p$ and Lyapunov's inequality to deduce

$$\|x_t - E[x_t | H_{t-1}^{t+1}]\|_p \leq \|\varepsilon_t\|_p \times \|\sigma_{t-1} - E[\sigma_{t-1} | G_{t-1}^{t+1}]\|_{2\alpha} \leq \check{d}_t \varphi_t,$$

where $\check{d}_t := \|\varepsilon_t\|_p \times d_t$, $\varphi_t = o(t^{-\lambda})$ and $\lambda = 2/\min\{p, 1\}$. By definition, therefore, $\{x_t\}$ is L_p -NED on $\{H_t\}$ with size λ . Hence $\{x_t\}$ is L_2 -E-NED on $\{H_t\}$ with constants $d_{n,t}(u) = K(m_n/n)^{1/2} e^{-up/2}$ and coefficients $\varphi_{m_n}^{\min\{p, 1\}/4}$ of size $\lambda \min\{p, 1\}/4 = \frac{1}{2}$ by Theorem 2.1 of Hill (forthcoming), a generalization of Lemma 4 of Hill (in press). The remaining claim follows by multiplying out the L_2 -E-NED quadratic criterion. \square

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

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

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

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

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

Proof of Lemma 3.5. By Lemma 3.1 we may write

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

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

$$\begin{aligned} \frac{1}{m_n} \sum_{s \neq t}^n \int_0^\infty \int_0^\infty \Delta P_{n,s,t}(u,v) du dv &= \left(\frac{m_n}{n}\right)^2 \frac{1}{m_n} \sum_{s \neq t}^n r(|s-t|) \left(\int_0^\infty \int_0^\infty e^{-\alpha(u+v)} du dv \right) (1 + o(1/m_n^{1/2})) \\ &= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) (1 + o(1/m_n^{1/2})) \end{aligned}$$

and

$$\begin{aligned} \alpha^{-1} \frac{1}{m_n} \sum_{s \neq t}^n \int_0^\infty \Delta P_{n,s,t}(u, v/m_n^{1/2}) du &= \alpha^{-1} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) \left(\int_0^\infty e^{-\alpha(u+v/m_n^{1/2})} du \right) (1 + o(1/m_n^{1/2})) \\ &= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) e^{-\alpha v/m_n^{1/2}} (1 + o(1/m_n^{1/2})) \end{aligned}$$

and

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

Therefore,

$$\begin{aligned} R_n &= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) (1 + o(1/m_n^{1/2})) - \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) e^{-\alpha v/m_n^{1/2}} (1 + o(1/m_n^{1/2})) \\ &\quad - \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) e^{-\alpha u/m_n^{1/2}} (1 + o(1/m_n^{1/2})) + \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) e^{-\alpha(u/m_n^{1/2} + v/m_n^{1/2})} (1 + o(1/m_n^{1/2})) \\ &= \alpha^{-2} \frac{m_n}{n^2} \sum_{s \neq t}^n r(|s-t|) (1 + o(1/m_n^{1/2})) \times (1 - e^{-\alpha u/m_n^{1/2}} - e^{-\alpha v/m_n^{1/2}} + e^{-\alpha(u/m_n^{1/2} + v/m_n^{1/2})}) \\ &= \alpha^{-2} \frac{m_n^{1/2}}{n^2} \sum_{s \neq t}^n r(|s-t|) (1 + o(1/m_n^{1/2})) \times O(1) = o(1), \end{aligned}$$

where the last line follows from Lemma 3.3. \square

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) and (17) imply (see Hsing, 1991, pp. 1548, 1553–1554)

$$\begin{aligned} P(x_t > b_n) &= (m_n/n) \times (1 + o(1/m_n^{1/2})) \\ P(x_t > b_n \lambda) &= \lambda^{-\alpha} (m_n/n) \times (1 + o(1/m_n^{1/2})) \\ E(\ln x_t b_n)_+^k &= \alpha^{-k} k! (m_n/n) \times (1 + o(1/m_n^{1/2})), \end{aligned}$$

where the $o(\cdot)$ terms do not depend on t .

By construction, therefore,

$$\begin{aligned} (n/m_n)E(U_{m_n,t}^*(u/m_n^{1/2})^2) &= (n/m_n)(P(x_t > b_n e^{u/m_n^{1/2}}) - P(x_t > b_n e^{u/m_n^{1/2}})^2) \\ &= e^{-\alpha u/m_n^{1/2}} \times (n/m_n)P(x_t > b_n) \times (1 + o(1/m_n^{1/2})) = \alpha^{-1} + O(1/m_n^{1/2}). \end{aligned}$$

Similarly

$$\begin{aligned} (n/m_n)E(U_{m_n,t}^2) &= (n/m_n)E(\ln x_t b_n)_+^2 - (n/m_n)E(\ln x_t b_n)_+^2 = (2\alpha^{-2} - (m_n/n)\alpha^{-2}) \times (1 + o(1/m_n^{1/2})) \\ &= 2\alpha^{-2} + o(1/m_n^{1/2}) + O(m_n/n), \end{aligned}$$

and

$$\begin{aligned} (n/m_n)E(U_{m_n,t} U_{m_n,t}^*(u/m_n^{1/2})) &= (n/m_n)E(I(x_t > b_n e^{u/m_n^{1/2}})(\ln x_t b_n)_+^2) - (n/m_n)P(x_t > b_n e^{u/m_n^{1/2}})E(\ln x_t b_n)_+ \\ &= (n/m_n) \int_0^\infty P(I(x_t > b_n e^{u/m_n^{1/2}})(\ln x_t b_n)_+^2 > v) dv = (n/m_n) \int_0^\infty P(x_t > b_n \max\{e^{u/m_n^{1/2}}, e^v\}) dv \end{aligned}$$

$$= (1 + o(1/m_n^{1/2})) \times \int_0^\infty \min\{e^{-\alpha u/m_n^{1/2}}, e^{-\alpha v}\} dv = \alpha^{-1} + O(1/m_n^{1/2}).$$

Therefore

$$(n/m_n)E(U_{m,t} - \alpha^{-1}U_{m,t}^*(u/m_n^{1/2}))^2 = 2\alpha^{-2} - 2\alpha^{-2} + \alpha^{-2} + o(1) = \alpha^{-2} + o(1),$$

where $o(1)$ does not depend on t . \square

References

- Andrews, D.W.K., 1984. Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- Basrak, B., Davis, R.A., Mikosch, T., 2002a. Regular variation of GARCH processes. *Stochastic Processes and their Applications* 99, 95–115.
- Basrak, B., Davis, R.A., Mikosch, T., 2002b. A characterization of multivariate regular variation. *Annals of Applied Probability* 12, 908–920.
- Beirlant, J., Dierckx, D., Guillou, A., 2005. Estimation of the extreme value index and generalized plots. *Bernoulli* 11, 949–970.
- Beirlant, J., Vynckier, P., Teugels, J.L., 1996. Tail index estimation, Pareto quantile plots and regression diagnostics. *Journal of the American Statistical Association* 91, 1659–1667.
- Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. *Regular Variation*. Cambridge University Press, Great Britain.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Economics* 31, 307–327.
- Bougerol, P., Picard, N., 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 Modèles GARCH, Unpublished Ph.D. Thesis, Université 7 Paris.
- Breidt, F.J., Davis, R.A., 1998. Extremes of stochastic volatility models. *Annals of Applied Probability* 8, 664–675.
- Breiman, L., 1965. On some limit theorems similar to the arc-sine law. *Theory of Probability and its Applications* 10, 351–360.
- Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.
- Chernick, M.R., Hsing, T., McCormick, W.P., 1991. Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability* 23, 835–850.
- Cline, D.B.H., 1986. Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields* 72, 529–557.
- Coles, S., Tawn, J., 1994. Statistical methods for multivariate extremes: an application to structural design. *Journal of Applied Statistics* 43, 1–48.
- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.
- Davidson, J., 2004. Moment and memory properties of linear conditional heteroscedasticity models and a new model. *Journal of Business & Economic Statistics* 22, 16–29.
- Davis, R.A., Mikosch, T., 2009a. Extreme value theory for GARCH processes. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Springer, New York, pp. 187–200.
- Davis, R.A., Mikosch, T., 2009b. Extremes of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Springer, New York, pp. 355–364.
- Davis, R.A., Mikosch, T., 2009c. The extremogram for extreme events. *Bernoulli* 15, 977–1009.
- Davis, R., Resnick, S.I., 1984. Tail estimates motivated by extreme value theory. *Annals of Statistics* 12, 1467–1487.
- de Haan, L., Resnick, S.I., 1996. Second-order regular variation and rates of convergence in extreme-value theory. *Annals of Probability* 24, 97–124.
- de Haan, L., Resnick, S.I., 1998. On asymptotic normality of the Hill estimator. *Stochastic Models* 14, 849–867.
- de Haan, L., Resnick, S.I., Rootzén, H., de Vries, C.G., 1989. Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications* 32, 213–224.
- Doukhan, P., 1994. *Mixing: Properties and Examples*. Lecture Notes in Statistics. Springer, New York.
- Drees, H., Ferreira, A., de Haan, L., 2004. On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability* 14, 1179–1201.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Gallant, A.R., White, H., 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, Oxford.
- Ghysels, E., Harvey, A., Renault, E., 1995. Stochastic volatility. In: Maddala, G.S., Rao, C.R. (Eds.), *Handbook of Statistics, Statistical Methods in Finance*, vol. 15. North Holland, Amsterdam.
- Giraitis, L., Kokoszka, P., Leipus, R., 2000. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.
- Goldie, C.M., Smith, R.L., 1987. Slow variation with remainder: theory and applications. *Quarterly Journal of Mathematics* 38, 45–71.
- Hausler, E., Teugels, J.L., 1985. On asymptotic normality of Hill's estimator for the exponent of regular variation. *Annals of Statistics* 13, 743–756.
- Hall, P., 1982. On some estimates of an exponent of regular variation. *Journal of the Royal Statistical Society* 44, 37–42.
- Hefferman, J., Tawn, J.A., 2004. A conditional approach to multivariate extreme values. *Journal of the Royal Statistical Society Series B* 66, 497–546.
- 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., forthcoming. Tail and non-tail memory with applications to extreme value and robust statistics, *Econometric Theory*.
- Hill, J.B., 2009. On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* 139, 2091–2110.
- Hill, J.B., in press. On tail index estimation for dependent, heterogeneous data. *Econometric Theory*, doi:10.1017/S0266466609990624.
- Hsing, T., 1991. On tail index estimation using dependent data. *Annals of Statistics* 19, 1547–1569.
- Ibragimov, I.A., Linnik, Y.V., 1971. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Berlin.
- Klüppelberg, C., Lindner, A., 2008. Extremes of random volatility models. Mimeo, Technische Universität München.
- Leadbetter, M.R., 1974. On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 28, 289–303.
- Leadbetter, M.R., 1983. Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 65, 291–306.
- Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.
- Ledford, A.W., Tawn, J.A., 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169–187.
- Ledford, A.W., Tawn, J.A., 1997. Modeling dependence within joint tail regions. *Journal of the Royal Statistical Society Series B* 59, 475–499.
- Ledford, A.W., Tawn, J.A., 2003. Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society Series B* 65, 521–543.
- Loynes, R.M., 1965. Extreme values in uniformly mixing stationary stochastic processes. *Annals of Mathematical Statistics* 36, 993–999.
- Mardia, K.V., 1964. Asymptotic independence of bivariate extremes. *Calcutta Statistical Association Bulletin* 13, 172–178.
- McLeish, D.L., 1975. A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829–839.
- Meddahi, N., Renault, E., 2004. Temporal aggregation of volatility models. *Journal of Econometrics* 119, 355–379.
- Meitz, M., Saikkonen, P., 2008. Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis* 29, 453–475.
- Mikosch, T., Stărică, C., 2000. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28, 1427–1451.
- Nelson, D.B., 1990. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.
- Neyman, J., Pearson, E.S., 1933. On the problem of the most efficient tests of statistical hypotheses. *Philosophic Transactions of the Royal Society of London, Series A* 231, 289–337.
- Nze, P.A., Doukhan, P., 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 Wahrscheinlichkeitstheorie und Verwandte Gebiete* 30, 57–63.
- Pickands, J., 1975. Statistical-inference using extreme order statistics. *Annals of Statistics* 3, 119–131.
- Pitman, E.J.G., 1979. *Some Basic Theory for Statistical Inference*. Wiley, New York.
- Resnick, S.I., 1987. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.
- Resnick, S.I., Stărică, C., 1995. Consistency of Hill's estimator for dependent data. *Journal of Applied Probability* 32, 139–167.
- Rootzén, H., Leadbetter, M., de Haan, L., 1990. Tail and quantile estimation for strongly mixing stationary sequences. Technical Report, University of North Carolina.
- Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences USA* 42, 43–47.
- Schmidt, R., Stadtmüller, U., 2006. Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics* 33, 307–335.
- Shephard, N., 2005. *Stochastic Volatility: Selected Readings*. Oxford University Press, Oxford.
- Sibuya, M., 1960. Bivariate extremal statistics. *Annals of the Institute of Statistical Mathematics* XI, 195–210.
- Smith, R.L., 1982. Uniform rates of convergence in extreme-value theory. *Advances in Applied Probability* 14, 600–622.
- Smith, R.L., 1987. Estimating tails of probability-distributions. *Annals of Statistics* 15, 1175–1207.
- Smith, R.L., Weissman, I., 1994. Estimating the extremal index. *Journal of the Royal Statistical Society Series B* 56, 128–151.
- Taylor, S., 1986. *Modelling Financial Time Series*. Wiley, New York.
- Watson, G.S., 1954. Extreme values in samples from m -dependent stationary stochastic processes. *Annals of Mathematical Statistics* 25, 798–800.