# Stochastic volatility models with possible extremal clustering

THOMAS MIKOSCH[1] and MOHSEN REZAPOUR[2]

[1]*University of Copenhagen, Department of Mathematics, Universitetsparken 5, DK-2100 Copenhagen, Denmark. E-mail: mikosch@math.ku.dk*
[2]*Department of Statistics, Shahid Bahonar University, Kerman, Iran. E-mail: mohsenrzp@gmail.com*

In this paper we consider a heavy-tailed stochastic volatility model, $X_t = \sigma_t Z_t$, $t \in \mathbb{Z}$, where the volatility sequence $(\sigma_t)$ and the i.i.d. noise sequence $(Z_t)$ are assumed independent, $(\sigma_t)$ is regularly varying with index $\alpha > 0$, and the $Z_t$'s have moments of order larger than $\alpha$. In the literature (see *Ann. Appl. Probab.* **8** (1998) 664–675, *J. Appl. Probab.* **38A** (2001) 93–104, In *Handbook of Financial Time Series* (2009) 355–364 Springer), it is typically assumed that $(\log \sigma_t)$ is a Gaussian stationary sequence and the $Z_t$'s are regularly varying with some index $\alpha$ (i.e., $(\sigma_t)$ has lighter tails than the $Z_t$'s), or that $(Z_t)$ is i.i.d. centered Gaussian. In these cases, we see that the sequence $(X_t)$ does not exhibit extremal clustering. In contrast to this situation, under the conditions of this paper, both situations are possible; $(X_t)$ may or may not have extremal clustering, depending on the clustering behavior of the $\sigma$-sequence.

*Keywords:* EGARCH; exponential AR(1); extremal clustering; extremal index; GARCH; multivariate regular variation; point process; stationary sequence; stochastic volatility process

## 1. Introduction

The *stochastic volatility model*

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z}, \tag{1.1}$$

has attracted some attention in the financial time series literature. Here the *volatility sequence* $(\sigma_t)$ is (strictly) stationary and consists of non-negative random variables independent of the i.i.d. sequence $(Z_t)$. We refer to [1] for a recent overview of the theory of stochastic volatility models. The popular GARCH model has the same structure (1.1), but every $Z_t$ feeds into the future volatilities $\sigma_{t+k}$, $k \geq 1$, and thus $(\sigma_t)$ and $(Z_t)$ are dependent in this case; see, for example, the definition of a GARCH(1, 1) process in Example 4.1. However, neither $\sigma_t$ nor $Z_t$ is directly observable, and thus whether we prefer a stochastic volatility a GARCH, or any other model for returns depends on our modeling efforts.

Previous research on extremes (e.g., [7,12,13,25]) has focused mainly on stochastic volatility models, where $(\log \sigma_t)$ constitutes a Gaussian stationary process and $(Z_t)$ is light-tailed (e.g., centered Gaussian) or rather heavy-tailed in the sense that there exists $\alpha > 0$, a slowly varying function $L$ and constants $p, q \geq 0$ such that $p + q = 1$ and

$$P(Z > x) \sim p x^{-\alpha} L(x) \quad \text{and} \quad P(Z \leq -x) \sim q x^{-\alpha} L(x), \qquad x \to \infty. \tag{1.2}$$

Here and in what follows, for any (strictly) stationary sequence $(Y_t)$, $Y$ denotes a generic element. A random variable $Z$ satisfying (1.2) will be called *regularly varying with index $\alpha$*.

Under the foregoing conditions, the sequence $(X_t)$ exhibits extremal behavior similar to an i.i.d. sequence whatever the strength of dependence in the Gaussian log-volatility sequence. In particular, $(X_t)$ does not have extremal clusters. It is common to measure extremal clustering in a stationary sequence $(Y_t)$ by considering the *extremal index*; suppose that an i.i.d. sequence $(\widetilde{Y}_t)$ with the same marginal distribution as $Y$ satisfies the limit relation

$$\lim_{n\to\infty} P\big(c_n^{-1}\big(\max(\widetilde{Y}_1,\ldots,\widetilde{Y}_n) - d_n\big) \le x\big) = H(x), \qquad x \in \mathbb{R}$$

for suitable constants $c_n > 0$, $d_n \in \mathbb{R}$ and a nondegenerate limit distribution function $H$ (which is necessarily continuous). If the same limit relation holds with $\max(\widetilde{Y}_1,\ldots,\widetilde{Y}_n)$ replaced by $\max(Y_1,\ldots,Y_n)$ and $H$ replaced by $H^\theta$ for some $\theta \in [0,1]$, then $\theta$ is called the *extremal index* of $(Y_t)$. Clearly, that the smaller the $\theta$, the stronger the extremal clustering effect present in the sequence. Under the aforementioned conditions, the stochastic volatility model $(X_t)$ has extremal index 1; that is, this process does not exhibit extremal clustering. However, real-life financial returns typically cluster around large positive and small negative values. This effect is described by the GARCH model, which under general conditions has an extremal index $\theta < 1$ (see [3,27]).

The aim of this paper is to show that the lack of extremal clustering in stochastic volatility models is due to the conditions on the tails of distributions of the sequences $(\sigma_t)$ and $(Z_t)$. In particular, we focus on the heavy-tailed situation when the distribution of $\sigma$ has power law tails in the sense that there exist $\alpha > 0$ and a slowly varying function $L$ such that

$$P(\sigma > x) \sim x^{-\alpha} L(x), \qquad x \to \infty;$$

that is, $\sigma$ is regularly varying with index $\alpha$, and $Z$ has lighter tail in the sense that $E|Z|^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$. By a result of Breiman [8], we then have

$$P(X > x) \sim EZ_+^\alpha P(\sigma > x) \quad \text{and} \quad P(X \le -x) \sim EZ_-^\alpha P(\sigma > x), \qquad x \to \infty.$$

This means that the tail behavior of $X$ is essentially determined by the right tail of $\sigma$. This is in contrast to the situation mentioned earlier. In that case, also by Breiman's result, $P(X > x) \sim E\sigma^\alpha P(Z > x)$. The latter relation is responsible for the lack of clustering; it indicates that extreme values of the sequence $(X_t)$ are essentially determined by the extremes in the i.i.d. sequence $(Z_t)$, an extremal index $\theta = 1$ result. We mention in passing that extremal clustering also can be expected when both the tails of $Z$ and $\sigma$ are regularly varying with the same index $\alpha > 0$. In that case it is well known (see [17]) that $X$ has regularly varying tails with a slowly varying function $L$, which in general is rather difficult to determine. We will not treat this case because it is of limited interest and will lead to rather technical conditions.

The paper is organized as follows. In Section 2 we introduce the notion of a regularly varying sequence and review point process convergence for such a sequence which was developed by [10]. We then state a result (Theorem 2.6) which translates mixing and regular variation of the sequence $(\sigma_t)$ to the stochastic volatility model $(X_t)$ defined in (1.1). Our results in Sections 3–5 are concerned with three major examples. In Section 3 we study the stochastic volatility model (1.1), where $(\sigma_t)$ is an exponential AR(1) process with regularly varying marginals. We show that

this model does not exhibit extremal clustering, due to the lack of extremal clustering in $(\sigma_t)$. We also show that an EGARCH model with the same volatility dynamics has no extremal clustering either. In Section 4 we assume that a positive power of $(\sigma_t)$ satisfies a random coefficient autoregressive equation, which we call stochastic recurrence equation. In this case, the extremal clustering of $(\sigma_t)$ translates to the stochastic volatility model. In Section 5 we consider another model with genuine extremal clustering. Here we assume that $(\sigma_t)$ is some positive power of the absolute values of a regularly varying moving average process.

## 2. Preliminaries

### 2.1. Regularly varying sequences

A strictly stationary sequence $(X_t)$ is said to be regularly varying with index $\alpha > 0$ if for every $d \geq 1$, the vector $\mathbf{X}_d = (X_1, \ldots, X_d)'$ is regularly varying with index $\alpha > 0$. This means that there exists a sequence $(a_n)$ with $a_n \to \infty$ and a sequence of non-null Radon measures $(\mu_d)$ on the Borel $\sigma$-field of $\overline{\mathbb{R}}_0^d = \overline{\mathbb{R}}^d \setminus \{0\}$ such that for every $d \geq 1$,

$$n P(a_n^{-1} \mathbf{X}_d \in \cdot) \xrightarrow{v} \mu_d(\cdot),$$

where $\xrightarrow{v}$ denotes vague convergence and $\mu_d$ satisfies the scaling property $\mu_d(t\cdot) = t^{-\alpha} \mu_d(\cdot)$, $t > 0$. The latter property justifies the term "regular variation with index $\alpha > 0$." The sequence $(a_n)$ can be chosen as such that $n P(|X_1| > a_n) \to 1$. We refer to [30,31] for more reading on regular variation and vague convergence of measures. Examples of regularly varying sequences are GARCH processes with i.i.d. Student or normal noise and ARMA processes with i.i.d. regularly varying noise. [12] studied the extremes of the stochastic volatility model (1.1) under the assumptions that $E\sigma^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$ and $(Z_t)$ is i.i.d. regularly varying with index $\alpha > 0$. Then $(X_t)$ is regularly varying with index $\alpha$, and the measures $\mu_d$ are concentrated on the axes. This property is shared with an i.i.d. regularly varying sequence $(X_t)$.

In this paper, we consider the opposite situation. We assume that $(\sigma_t)$ is regularly varying with index $\alpha > 0$, normalizing constants $(a_n)$ such that $n P(\sigma > a_n) \to 1$, and limiting measures $\nu_d$, $d = 1, 2, \ldots$. This means that for $\mathbf{\Sigma}_d = (\sigma_1, \ldots, \sigma_d)'$, $d \geq 1$, the relations

$$n P(a_n^{-1} \mathbf{\Sigma}_d \in \cdot) \xrightarrow{v} \nu_d(\cdot)$$

hold. We also assume that $E|Z|^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$.

**Lemma 2.1.** *Under the foregoing conditions, $(X_t)$ is regularly varying with index $\alpha$ and limiting measures $\mu_d$, $d = 1, 2, \ldots$, given by the relation*

$$\mu_d(\cdot) = E\nu_d\{\mathbf{s} \in \mathbb{R}_+^d \colon (Z_1 s_1, \ldots, Z_d s_d) \in \cdot\}. \tag{2.1}$$

**Proof.** Assuming that all vectors are written in column form, we have $\mathbf{X}_d = \mathbf{A} \mathbf{\Sigma}_d$, where $\mathbf{A} = \text{diag}(Z_1, \ldots, Z_d)$. The matrix $\mathbf{A}$ has moment of order $\alpha + \varepsilon$ and then regular variation of $\mathbf{X}_d$

with normalizing constants $(a_n)$ given by $nP(\sigma > a_n) \to 1$ and the form of the limit measures $\mu_d$ follow from the multivariate Breiman result (see [3]). □

The limits (2.1) are generally difficult to evaluate. We consider some simple examples.

**Example 2.2.** Assume that $\nu_d$ is concentrated on the axes, that is, it has the form

$$\nu_d(\cdot) = c_d \sum_{i=1}^{d} \int_0^\infty x^{-\alpha-1} I_{\{x\mathbf{e}_i \in \cdot\}} \, dx$$

for some constants $c_d > 0$, where $\mathbf{e}_i$ denotes the $i$th unit vector in $\mathbb{R}^d$. Then (2.1) reads as

$$\mu_d(\cdot) = c_d \sum_{i=1}^{d} \int_0^\infty x^{-\alpha-1} P(x Z_1 \mathbf{e}_i \in \cdot) \, dx$$

$$= c_d \left[ E Z_+^\alpha \sum_{i=1}^{d} \int_0^\infty x^{-\alpha-1} I_{\{x\mathbf{e}_i \in \cdot\}} \, dx + E Z_-^\alpha \sum_{i=1}^{d} \int_0^\infty x^{-\alpha-1} I_{\{-x\mathbf{e}_i \in \cdot\}} \, dx \right].$$

Sometimes it is possible to characterize the limit measures $\mu_d$ by their values on all sets of the form $A_{\mathbf{c}} = \{\mathbf{x} \in \mathbf{X}^d : \mathbf{c}'\mathbf{x} > 1\}$ for any choice of $\mathbf{c}$ in the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}$ with respect to the Euclidean norm. However, in general, $\mu_d$ cannot be reconstructed from its values on the sets $A_{\mathbf{c}}$ (see [2,6,20]).

**Example 2.3.** Consider an i.i.d. sequence of symmetric $\beta$-stable random variables $(Z_t)$; that is, the characteristic function of $Z$ is given by $e^{-c|z|^\beta}$, $z \in \mathbb{R}$, for some $c > 0$. Assume that $\beta = 2$ for $\alpha \geq 2$ and $2 \geq \beta > \alpha$ for $\alpha < 2$. Then, for $\mathbf{c} \in \mathbb{S}^{d-1}$,

$$\mu_d(A_{\mathbf{c}}) = E\nu_d \left\{ \mathbf{s} \in \mathbb{R}_+^d : \sum_{i=1}^{d} c_i Z_i s_i > 1 \right\}$$

$$= E\nu_d \left\{ \mathbf{s} \in \mathbb{R}_+^d : Z \left( \sum_{i=1}^{d} |c_i|^\beta s_i^\beta \right)^{1/\beta} > 1 \right\}$$

$$= E Z_+^\alpha \nu_d \left\{ \mathbf{s} \in \mathbb{R}_+^d : \left( \sum_{i=1}^{d} |c_i|^\beta s_i^\beta \right)^{1/\beta} > 1 \right\}.$$

The measure $\mu_d$, for example, is uniquely determined by the values $\mu_d(A_{\mathbf{c}})$, $\mathbf{c} \in \mathbb{S}^{d-1}$, provided that they are positive and $\alpha$ is not an integer (see [2]) or, in view of the symmetry of the underlying distributions, if $\alpha$ is an odd integer (see [24]). By virtue of the foregoing calculations, this means

that $\mu_d$ is uniquely determined by the values of $\nu_d$ on the sets

$$\left\{ \mathbf{s} \in \mathbb{R}_+^d \colon \left( \sum_{i=1}^d |c_i|^\beta s_i^\beta \right)^{1/\beta} > 1 \right\}, \qquad \mathbf{c} \in \mathbb{S}^{d-1}, \tag{2.2}$$

provided that these values are positive. For $\beta = 2$, $Z$ is centered normal, and then (2.2) describes the complements of all ellipsoids in $\mathbb{R}^d$ with $\sum_{i=1}^d c_i^2 = 1$ intersected with $\mathbb{R}_+^d$.

## 2.2. Mixing conditions

For the reader's convenience, here we introduce mixing concepts for strictly stationary sequences $(X_t)$ used in this work. For $h \geq 1$, let

$$\alpha_h = \sup_{A \in \sigma_{(-\infty,0]}, B \in \sigma_{[h,\infty)}} |P(A \cap B) - P(A)P(B)|,$$

$$\beta_h = E \left( \sup_{B \in \sigma_{[h,\infty)}} \left| P\left(B | \sigma_{(-\infty,0]}\right) - P(B) \right| \right),$$

where $\sigma_A$ is the $\sigma$-field generated by $(X_t)_{t \in A}$ for any $A \subset \mathbb{Z}$. The sequence $(X_t)$ is *strongly mixing with rate function* $(\alpha_h)$ if $\alpha_h \to 0$ as $h \to \infty$. If $\beta_h \to 0$ as $h \to \infty$, then $(X_t)$ is $\beta$-*mixing with rate function* $(\beta_h)$. Strong mixing is known to imply $\beta$-mixing (see Doukhan [15] for examples and comparisons of different mixing concepts).

Strong mixing and $\beta$-mixing were introduced in the context of the central limit theory for partial sums of $(X_t)$. For partial maxima of $(X_t)$, other mixing concepts are more suitable (see, e.g., the conditions $D$ and $D'$ in Leadbetter *et al.* [26]). In this paper, we make use of the condition $\mathcal{A}(a_n)$ introduced by Davis and Hsing [10]: Assume that there exists a sequence $r_n \to \infty$ such that $r_n = \mathrm{o}(n)$ and

$$\Psi_f(N_n) - (\Psi_f(N_{n,r_n}))^{n/r_n} \to 0, \tag{2.3}$$

where $N_n$ is the point process of the points $(a_n^{-1} X_t)_{t=1,\ldots,n}$, $N_{n,r_n}$ is the point process of the points $(a_n^{-1} X_t)_{t=1,\ldots,r_n}$, $\Psi_f(N)$ denotes the Laplace functional of the point process $N$ evaluated at the non-negative function $f$ and $(a_n)$ satisfies $P(|X| > a_n) \sim n^{-1}$. Davis and Hsing [10] required (2.3) to hold only for non-negative measurable step functions $f$, which have bounded support in $\overline{\mathbb{R}}_0$. The mixing condition $\mathcal{A}(a_n)$ is very general. It ensures that $N_n$ has the same limit (provided that it exists) as a sum of $[n/r_n]$ i.i.d. copies of the point process $N_{n,r_n}$. Condition $\mathcal{A}(a_n)$ is implied by many known mixing conditions, particularly strong mixing (see [10]).

## 2.3. The Davis and Hsing [10] approach

Davis and Hsing presented a rather general approach to the extremes of a strictly stationary sequence $(X_t)$. We quote their Theorem 2.7 for further reference.

**Theorem 2.4.** *Assume that $(X_t)$ is regularly varying with index $\alpha > 0$ and normalization $(a_n)$ such that $P(|X| > a_n) \sim n^{-1}$, the mixing condition $\mathcal{A}(a_n)$ is satisfied, and the anticlustering condition*

$$\lim_{m \to \infty} \limsup_{n \to \infty} P\left(\max_{m \le |t| \le r_n} |X_t| > y a_n \big| |X_0| > y a_n\right) = 0, \qquad y > 0, \tag{2.4}$$

*holds. Here $(r_n)$ is an integer sequence such that $r_n \to \infty$, $r_n = o(n)$, which appears in the definition of $\mathcal{A}(a_n)$. Then the following point process convergence holds in $M_p(\overline{\mathbb{R}}_0)$, the set of point processes with state space $\overline{\mathbb{R}}_0$, equipped with the vague topology and the Borel $\sigma$-field:*

$$N_n = \sum_{i=1}^{n} \varepsilon_{X_t/a_n} \xrightarrow{d} N = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varepsilon_{P_i Q_{ij}},$$

*where $(P_i)$ are the points of a Poisson process on $(0, \infty)$ with intensity $\lambda(dx) = \theta_{|X|} \alpha x^{-\alpha-1} \, dx$ and $\sum_{i=1}^{\infty} \varepsilon_{Q_{ij}}$, $j \ge 1$, constitute an i.i.d. sequence of point processes whose points satisfy the property $|Q_{ij}| \le 1$ a.s. and $\sup_j |Q_{ij}| = 1$ a.s. Here $\theta_{|X|} \in [0, 1]$ is the extremal index of the sequence $(|X_t|)$.*

**Remark 2.5.** The anticlustering condition (2.4) ensures that clusters of extremes become separated from one another through time. (For a precise description of the distribution of the point processes $\sum_{i=1}^{\infty} \varepsilon_{Q_{ij}}$, see [10]. For more on the extremal index of a stationary sequence, see [26] and [16], Section 8.1. For an introduction to point processes and their convergence in the context of extreme value theory, see [30,31].)

An immediate consequence of Theorem 2.4 is limit theory for the maxima $M_n^{|X|} = \max_{t=1,\dots,n} |X_t|$, $n \ge 1$, of the sequence $(|X_t|)$. Indeed, we conclude with $(a_n)$ chosen such that $n P(|X| > a_n) \sim 1$,

$$\lim_{n \to \infty} P\left(a_n^{-1} M_n^{|X|} \le x\right) = \lim_{n \to \infty} P\left(N_n([-x, x]) = 0\right) = P\left(N([-x, x]) = 0\right)$$

$$= P\left(\sup_{i \ge 1} P_i \sup_{j \ge 1} |Q_{ij}| \le x\right)$$

$$= P\left(\sup_{i \ge 1} P_i \le x\right) = P(P_1 \le x)$$

$$= \Phi_\alpha^{\theta_{|X|}}(x), \qquad x > 0,$$

where $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x > 0$, denotes the Fréchet distribution function with parameter $\alpha$. Similar results can be derived for the maxima and upper-order statistics of the $X$-sequence, joint convergence of minima and maxima, and other results belonging to the folklore of extreme value theory. Theorem 2.4 is fundamental for an extreme value theory of the sequence $(X_t)$, and the results reported by [3,4,10,11,31] also show that the point process convergence can be used to derive limit results for sums, sample autocovariances and autocorrelations, and large deviation results.

## 2.4. A translation result

Our next result states that the stochastic volatility model (1.1) inherits the properties relevant for the extremal behavior of $(X_t)$ from the volatility sequence $(\sigma_t)$.

**Theorem 2.6.** *Consider the stochastic volatility model* (1.1). *Assume that* $(\sigma_t)$ *is regularly varying with index* $\alpha > 0$, *it satisfies the strong mixing property, and* $E|Z|^{\alpha+\varepsilon} < \infty$ *for some* $\varepsilon > 0$. *Then* $(X_t)$ *is regularly varying with index* $\alpha$ *and is strongly mixing with the same rate as* $(\sigma_t)$. *If* $(X_t)$ *also satisfies the anticlustering condition* (2.4), *then Theorem* 2.4 *applies.*

**Proof.** The proof of the regular variation of $(X_t)$ follows from Lemma 2.1. Strong mixing of $(\sigma_t)$ implies strong mixing of $(X_t)$ with the same rate (see page 258 in [14]). Because we assume the anticlustering condition (2.4) for $(X_t)$, the conditions of Theorem 2.4 are satisfied. $\qquad\square$

***Remark 2.7.*** If $|Z| \le M$ a.s. for some positive $M$, then the anticlustering condition for $(X_t)$ follows trivially from that for $(\sigma_t)$. If $Z$ is unbounded, then whether this conclusion remains true is not obvious. However, when dealing with concrete examples, it often is not difficult to derive the anticlustering condition for $(X_t)$; see the examples below.

In what follows, we consider three examples of regularly varying stochastic volatility models. In all cases, the volatility sequence $(\sigma_t)$ is stationary and dependent. We verify the regular variation, strong mixing, and anticlustering conditions for $(\sigma_t)$ and show that these properties are inherited by $(X_t)$. The exponential AR(1) model $(\sigma_t)$ of Section 3 does not cause extremal clustering of $(X_t)$ whereas a random coefficient autoregressive or linear process structure of $(\sigma_t)$ triggers extremal dependence in the stochastic volatility model; see Sections 4 and 5.

## 3. Exponential AR(1)

Our first example is an exponential autoregressive process of order 1 process [we write AR(1)] given by

$$\sigma_t = e^{Y_t}, \qquad t \in \mathbb{Z}, \tag{3.1}$$

where $(Y_t)$ is a causal stationary AR(1) process $Y_t = \varphi Y_{t-1} + \eta_t$ for some $\varphi \in (-1, 1)$ and an i.i.d. sequence $(\eta_t)$ of random variables.

***Example 3.1.*** Volatility sequences of the type (3.1) appear in the EGARCH (exponential GARCH) model introduced by [29]. In this case, $X_t = \sigma_t Z_t$, $t \in \mathbb{Z}$, $(Z_t)$ is an i.i.d. sequence and

$$\log(\sigma_t^2) = \alpha_0(1 - \varphi)^{-1} + \sum_{k=0}^{\infty} \varphi^k (\gamma_0 Z_{t-1-k} + \delta_0 |Z_{t-1-k}|), \qquad t \in \mathbb{Z} \tag{3.2}$$

for positive parameters $\alpha_0, \delta_0, \gamma_0$ and $\varphi \in (-1, 1)$. Most often, it is assumed that $(Z_t)$ is an i.i.d. standard normal sequence. In that case, $\sigma$ has all moments in contrast to the situation that

we consider in this section. This model is close to the stochastic volatility model (1.1) with an exponential AR(1) volatility sequence (3.1). However, in the EGARCH model, $Z_t$ feeds into the sequence $(\sigma_s)_{s>t}$, and thus the $\sigma$- and $Z$-sequences are dependent.

## 3.1. Mixing property

It is known that $(Y_t)$, and hence $(\sigma_t)$, are $\beta$-mixing with geometric rate if $\eta$ has a positive density in some neighborhood of $E\eta$ (cf. [15], Theorem 6, page 99).

## 3.2. Regular variation

We introduce the following conditions:

$$P(e^\eta > x) = x^{-\alpha} L(x), \qquad x > 0, \tag{3.3}$$

$$P(e^{\eta^-} > x) \leq c P(e^\eta > x), \qquad x \geq 1, \tag{3.4}$$

for some $\alpha > 0$, a slowly varying function $L$ and some constant $c > 0$, and $\eta^-$ denotes the negative part of $\eta$. Here and in what follows, $c$ denotes any positive constants that are possibly different but whose values are not of interest. Note that these conditions are satisfied if $\eta$ is gamma or Laplace distributed.

We first prove that $\sigma$ is regularly varying.

**Lemma 3.2.** *Assume* (3.3) *and also* (3.4) *if* $\varphi < 0$. *Then* $E e^{(\alpha+\varepsilon)\varphi Y} < \infty$ *for some small* $\varepsilon > 0$, *and the following relation holds*:

$$P(\sigma > x) \sim E e^{\alpha \varphi Y} P(e^\eta > x) = E e^{\alpha \varphi Y} x^{-\alpha} L(x), \qquad x \to \infty. \tag{3.5}$$

**Proof.** Because $\sigma_t = e^{\eta_t} \sigma_{t-1}^\varphi$, the random variables $e^{\eta_t}$, $\sigma_{t-1}$ are independent and, by (3.3), $e^\eta$ is regularly varying with index $\alpha > 0$, we may apply a result of Breiman [8] to conclude that (3.5) holds if we can show that there exists an $\varepsilon > 0$ such that $E e^{(\alpha+\varepsilon)\varphi Y} < \infty$. We first consider the case of positive $\varphi$. Here

$$E e^{(\alpha+\varepsilon)\varphi Y} = \prod_{i=1}^\infty E e^{(\alpha+\varepsilon)\varphi^i \eta}.$$

By (3.3), for every $\delta > 0$, there exists an $x_0 > 1$ such that $P(e^\eta > x) \leq x^{-\alpha+\delta}$ for $x \geq x_0$ (so-called *Potter bounds*; see Bingham *et al.* [5], page 25). Thus for small $\varepsilon, \delta > 0$ such that $((\alpha - \delta)/[(\alpha + \varepsilon)\varphi^1] - 1) > 0$,

$$E e^{(\alpha+\varepsilon)\varphi^i \eta} \leq x_0^{(\alpha+\varepsilon)\varphi^i} + \int_{x_0^{(\alpha+\varepsilon)\varphi^i}}^\infty P(e^{(\alpha+\varepsilon)\varphi^i \eta} > y)\, dy$$

$$\leq x_0^{(\alpha+\varepsilon)\varphi^i} + ((\alpha - \delta)/[(\alpha + \varepsilon)\varphi^i] - 1)^{-1} x_0^{-\alpha+\delta+(\alpha+\varepsilon)\varphi^i}.$$

We conclude that for small $\varepsilon, \delta > 0$, some constants $c > 0$,

$$\prod_{i=1}^{\infty} E e^{(\alpha+\varepsilon)\varphi^i \eta} \leq \exp\left\{ \sum_{i=1}^{\infty} \left[ x_0^{(\alpha+\varepsilon)\varphi^i} - 1 + \left( (\alpha - \delta)/[(\alpha + \varepsilon)\varphi^i] - 1 \right)^{-1} x_0^{-\alpha+\delta+(\alpha+\varepsilon)\varphi^i} \right] \right\}$$

$$\leq c \exp\left\{ c \sum_{i=1}^{\infty} \varphi^i \right\} < \infty.$$

We next consider the case of negative $\varphi$. We observe that

$$E e^{(\alpha+\varepsilon)\varphi Y} \leq \prod_{i=1}^{\infty} E e^{(\alpha+\varepsilon)\varphi^{2i} \eta} \prod_{i=1}^{\infty} E e^{(\alpha+\varepsilon)|\varphi|^{2i-1} \eta^{-}}.$$

Similar calculations as before, where we exploit (3.3) and (3.4), show that the right-hand side is finite for small $\varepsilon$. □

**Lemma 3.3.** *Assume the conditions of Lemma 3.2. Then the sequence $(\sigma_t)$ is regularly varying with index $\alpha$. The limit measure of the vector $\boldsymbol{\Sigma}_d = (\sigma_1, \ldots, \sigma_d)'$ is given by the following limiting relation on the Borel $\sigma$-field of $\overline{\mathbb{R}}_0^d$:*

$$\frac{P(x^{-1} \boldsymbol{\Sigma}_d \in \cdot)}{P(\sigma > x)} \xrightarrow{v} \alpha \sum_{i=1}^{d} \int_0^{\infty} y^{-\alpha-1} I_{\{y \mathbf{e}_i \in \cdot\}} \, dy, \qquad x \to \infty, \tag{3.6}$$

*where $\mathbf{e}_i$ is the $i$th unit vector in $\mathbb{R}^d$.*

**Proof.** We give the proof only for positive $\varphi$ and $\eta$. Proofs for the other cases are similar.

We observe that

$$\boldsymbol{\Sigma}_d = \text{diag}(e^{\varphi Y_0}, e^{\varphi^2 Y_0}, \ldots, e^{\varphi^d Y_0}) \begin{pmatrix} e^{\eta_1} \\ e^{\eta_2 + \varphi \eta_1} \\ \vdots \\ e^{\eta_d + \varphi \eta_{d-1} + \cdots + \varphi^{d-1} \eta_1} \end{pmatrix} = \mathbf{A}\mathbf{B}.$$

Because $E\|\mathbf{A}\|^{\alpha+\varepsilon} < \infty$ for small positive $\varepsilon$ and $\mathbf{A}$ and $\mathbf{B}$ are independent regular variation of $\boldsymbol{\Sigma}_d$ will follow from Breiman's multivariate result [3] if it can be shown that $\mathbf{B}$ is regularly varying with index $\alpha$. Indeed, we will show that $\mathbf{B}$ has the same limit measure as

$$(e^{\eta_1}, e^{\eta_2} E e^{\alpha \varphi \eta}, \ldots, e^{\eta_d} E e^{\alpha \varphi \eta} \cdots E e^{\alpha \varphi^{d-1} \eta})'.$$

This fact does not follow from the continuous mapping theorem for regularly varying vectors (see [19,21]), because the function $(r_1, \ldots, r_d) \to (r_1, r_1^{\varphi} r_2, \ldots, r_1^{\varphi^{d-1}} \cdots r_{d-1}^{\varphi} r_d)$ does not have the homogeneity property.

For simplicity, we prove the result only for $d = 2$, the general case being analogous. To ease notation, we also write $R_i = e^{\eta_i}$, $i = 1, 2$. Choose $a_n$ such that $P(e^{\eta} > a_n) \sim n^{-1}$ and take any

set $A \subset \overline{\mathbb{R}}_0^2$ that is a subset of the first orthant bounded away from 0 and continuous with respect to the limiting measure of $\Sigma_d$ in the formulation of the lemma. Write $B = \{a_n^{-1}(R_1, R_1^\varphi R_2) \in A\}$, and for any $\varepsilon, \gamma > 0$, consider the disjoint sets

$$B_1 = B \cap \{R_1 > \varepsilon a_n, R_2 > \gamma a_n\},$$

$$B_2 = B \cap \{R_1 > \varepsilon a_n, R_2 \leq \gamma a_n\},$$

$$B_3 = B \cap \{R_1 \leq \varepsilon a_n, R_2 > \gamma a_n\},$$

$$B_4 = B \cap \{R_1 \leq \varepsilon a_n, R_2 \leq \gamma a_n\}.$$

Then for any $\varepsilon, \gamma > 0$,

$$n P(B_1) \leq n P(R_1 > \varepsilon a_n) P(R_2 > \gamma a_n) \to 0.$$

Next, consider $B_3$. Choose some $M > 1$ and consider the disjoint partition of $B_3$,

$$B_{31} = B_3 \cap \{R_1 \in [1, M]\}, \qquad B_{32} = B_3 \cap \{R_1 > M\}.$$

Then

$$n P(B_{32}) \leq n P(R_2 > \gamma a_n) P(R_1 > M) \sim \gamma^{-\alpha} P(R_1 > M), \qquad n \to \infty.$$

Thus, for any $\varepsilon, \gamma > 0$,

$$\lim_{M \to \infty} \limsup_{n \to \infty} n P(B_{32}) = 0.$$

Observe that $n P(R_1 I_{\{R_1 \in [1, M]\}} > c a_n) \to 0$ for every $c > 0$ and, by Breiman's result [8], $R_2 R_1^\varphi I_{\{R_1 \in [1, M]\}}$ is regularly varying. By Lemma 3.12 of [22],

$$\left(R_1 I_{\{R_1 \in [1, M]\}}, R_2 R_1^\varphi I_{\{R_1 \in [1, M]\}}\right) = \left(R_1 I_{\{R_1 \in [1, M]\}}, 0\right) + \left(0, R_2 R_1^\varphi I_{\{R_1 \in [1, M]\}}\right)$$

is regularly varying with the same index and limiting measure as $(0, R_2 R_1^\varphi I_{\{R_1 \in [1, M]\}})$. Therefore,

$$n P(B_{31}) \sim n P\left(a_n^{-1}\left(0, R_2 R_1^\varphi I_{\{R_1 \in [1, M]\}}\right) \in A, R_2 > \gamma a_n\right)$$

$$= n P\left(a_n^{-1} R_2 R_1^\varphi I_{\{R_1 \in [1, M]\}} \in \text{proj}_2 A, R_2 > \gamma a_n\right) I_{\{\text{proj}_1 A = \{0\}\}},$$

where $\text{proj}_i A$, $i = 1, 2$, are the projections of $A$ on the $x$- and $y$-axes, respectively. Regular variation of $R_2$ with limit measure $\mu(t, \infty) = t^{-\alpha}$, $t > 0$, ensures that

$$\lim_{M \to \infty} \lim_{\gamma \to 0} \lim_{n \to \infty} n P(B_{31}) = \lim_{M \to \infty} \lim_{\gamma \to 0} E \mu\{t > \gamma: R_1^\varphi I_{\{R_1 \in [1, M]\}} t \in \text{proj}_2 A\} I_{\{\text{proj}_1 A = \{0\}\}}$$

$$= \lim_{M \to \infty} E \mu\{t > 0: R_1^\varphi I_{\{R_1 \in [1, M]\}} t \in \text{proj}_2 A\} I_{\{\text{proj}_1 A = \{0\}\}}$$

$$= \lim_{M \to \infty} E R_1^{\alpha \varphi} I_{\{R_1 \in [1, M]\}} \mu(\text{proj}_2 A) I_{\{\text{proj}_1 A = \{0\}\}}$$

$$= E R_1^{\alpha \varphi} \mu(\text{proj}_2 A) I_{\{\text{proj}_1 A = \{0\}\}}.$$

We have $A \subset \{\mathbf{x}: |x_1| + |x_2| > \delta\}$ for small $\delta > 0$. Then $B_4$ is contained in the union of the following sets for $M > 1$:

$$B_{41} = B_4 \cap \{R_1 > 0.5\delta a_n\},$$
$$B_{42} = B_4 \cap \{R_1^\varphi R_2 > 0.5\delta a_n, R_1 > M\},$$
$$B_{43} = B_4 \cap \{R_1^\varphi R_2 > 0.5\delta a_n, R_1 \in [1, M]\}.$$

Choosing $\varepsilon$ sufficiently small, $B_{41}$ is empty. Moreover, by Breiman's result,

$$n P(B_{42}) \leq n P(R_1^\varphi R_2 > 0.5\delta a_n, R_1 > M) \sim c E\big[R_1^{\alpha\varphi} I_{\{R_1 > M\}}\big], \qquad n \to \infty.$$

Choosing $\gamma$ sufficiently small, the set $B_{43}$ is empty. Therefore, and because $E[R_1^{(\alpha+\varepsilon)\varphi}] < \infty$,

$$\lim_{M \to \infty} \limsup_{n \to \infty} n P(B_{4i}) = 0, \qquad i = 1, 2, 3.$$

It remains to consider the set $B_2$. Consider the disjoint partition of $B_2$ for $M > 1$,

$$B_{21} = B_2 \cap \{R_2 \leq M\} \quad \text{and} \quad B_{22} = B_2 \cap \{R_2 > M\}.$$

Because $P(B_{22}) \leq P(R_1 > \varepsilon a_n) P(R_2 > M)$, we have

$$\lim_{M \to \infty} \limsup_{n \to \infty} n P(B_{22}) = 0.$$

Moreover,

$$n P(B_{21}) \sim n P\big(R_1 > \varepsilon a_n, a_n^{-1}(R_1, 0) \in A\big).$$

Thus, for every $M > 0$,

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} n P(B_{21}) = \lim_{\varepsilon \to 0} \mu\{t > \varepsilon: t \in \text{proj}_1 A\} I_{\{\text{proj}_2 A = \{0\}\}}$$
$$= \mu\{\text{proj}_1 A\} I_{\{\text{proj}_2 A = \{0\}\}}.$$

Summarizing the foregoing arguments, we have proven that

$$n P\big(a_n^{-1}(R_1, R_1^\varphi R_2) \in A\big) \to \alpha \int_0^\infty x^{-\alpha-1}\big[I_{\{x\mathbf{e}_1 \in A\}} + E R_1^{\alpha\varphi} I_{\{x\mathbf{e}_2 \in A\}}\big] dx.$$

Modifying the proof above for $d \geq 2$, we obtain

$$n P(a_n^{-1}\mathbf{B} \in A)$$
$$\to \alpha \int_0^\infty x^{-\alpha-1}\big[I_{\{x\mathbf{e}_1 \in A\}} + E e^{\alpha\varphi\eta} I_{\{x\mathbf{e}_2 \in A\}} + \cdots + E e^{\alpha\varphi\eta} \cdots E e^{\alpha\varphi^{d-1}\eta} I_{\{x\mathbf{e}_d \in A\}}\big] dx.$$

We now apply the multivariate Breiman result [3] to obtain

$$
\begin{aligned}
n P(a_n^{-1}\mathbf{AB} \in A) \\
\to \alpha \int_0^\infty x^{-\alpha-1} E\big[I_{\{e^{\varphi Y}x\mathbf{e}_1 \in A\}} + E e^{\alpha\varphi\eta} I_{\{e^{\varphi^2 Y}x\mathbf{e}_2 \in A\}} + \cdots \\
+ E e^{\alpha\varphi\eta}\cdots E e^{\alpha\varphi^{d-1}\eta} I_{\{e^{\varphi^d Y}x\mathbf{e}_d \in A\}}\big]\mathrm{d}x \\
= \alpha \int_0^\infty x^{-\alpha-1}\big[E e^{\alpha\varphi Y} I_{\{x\mathbf{e}_1 \in A\}} + E e^{\alpha\varphi\eta} E e^{\alpha\varphi^2 Y} I_{\{x\mathbf{e}_2 \in A\}} + \cdots \\
+ E e^{\alpha\varphi\eta}\cdots E e^{\alpha\varphi^{d-1}\eta} E e^{\alpha\varphi^d Y} I_{\{x\mathbf{e}_d \in A\}}\big]\mathrm{d}x \\
= \alpha E e^{\alpha\varphi Y} \int_0^\infty x^{-\alpha-1}\big[I_{\{x\mathbf{e}_1 \in A\}} + I_{\{x\mathbf{e}_2 \in A\}} + \cdots + I_{\{x\mathbf{e}_d \in A\}}\big]\mathrm{d}x.
\end{aligned}
$$

This relation and Lemma 3.2 conclude the proof for $\varphi \in (0, 1)$ and $\eta > 0$ a.s. $\qquad\square$

## 3.3. Anticlustering condition

**Lemma 3.4.** *Assume* (3.3) *and also*

$$
P\big(e^{|\eta|} > x\big) \le c P(e^\eta > x), \qquad x \ge 1 \tag{3.7}
$$

*for some $c > 0$, $\varphi \in (-1, 1)$. Then the anticlustering condition* (2.4) *holds for the sequence $(\sigma_t)$ and any sequence $(r_n)$ satisfying $r_n = O(n^\gamma)$ for some $\gamma \in (0, 1)$. If $|Z|$ has finite moments of any order, then* (2.4) *is also satisfied for the stochastic volatility sequence $(X_t)$ with the same sequence $(r_n)$ as for $(\sigma_t)$. If $E|Z|^{\alpha+\xi} < \infty$ for some $\xi > 0$, then* (2.4) *holds for the sequence $(X_t)$ with $(r_n)$ such that $r_n = O(n^\gamma)$ for every $\gamma \in (0, 1)$.*

**Proof.** Throughout, we assume that $\varphi \ne 0$. If $\varphi = 0$, then both $(\sigma_t)$ and $(X_t)$ are i.i.d. regularly varying sequences, and (2.4) is trivially satisfied.

We first prove the result for $(\sigma_t)$. We begin under the assumptions $\varphi \in (0, 1)$ and $\eta > 0$, and verify that

$$
\lim_{m\to\infty} \limsup_{n\to\infty} P\Big(\max_{m\le t\le r_n} Y_t > \log(ya_n)\,\big|\,Y_0 > \log(ya_n)\Big) = 0, \qquad y > 0. \tag{3.8}
$$

Fix $y > 0$ and write $B = \{\max_{m\le t\le r_n} Y_t > \log(ya_n)\}$ and observe that

$$
Y_t = \varphi^{t-m} Y_m + \sum_{i=m+1}^t \varphi^{t-i}\eta_i, \qquad m \le t. \tag{3.9}
$$

Then $B \subset B_1 \cup B_2$, where for $\delta \in (0, 1)$,

$$
B_1 = \{Y_m > \delta \log(ya_n)\} \quad \text{and} \quad B_2 = \Bigg\{\max_{m\le t\le r_n} \sum_{i=m+1}^t \varphi^{t-i}\eta_i > (1 - \delta)\log(ya_n)\Bigg\}.
$$

Because $Y_0$ is independent of $(\eta_t)_{t \geq 1}$, $P(B_2) = P(B_2 | Y_0 > \log(ya_n))$. Therefore, and by Markov's inequality,

$$P\big(B_2 | Y_0 > \log(ya_n)\big) \leq \sum_{t=m}^{r_n} P\left(\sum_{i=m+1}^{t} \varphi^{t-i} \eta_i > (1-\delta)\log(ya_n)\right)$$

$$\leq \sum_{t=m}^{r_n} P\big(Y_t > (1-\delta)\log(ya_n)\big)$$

$$\leq r_n P\big(Y > (1-\delta)\log(ya_n)\big)$$

$$\leq r_n E e^{(\alpha-\varepsilon)Y}(ya_n)^{-(1-\delta)(\alpha-\varepsilon)}$$

for $0 < \varepsilon < \alpha$ and large $n$. Because $E e^{(\alpha-\varepsilon)Y} < \infty$ and $a_n = n^{1/\alpha}\ell(n)$ for some slowly varying function $\ell$, choosing $\delta, \varepsilon > 0$ sufficiently small, the right-hand side converges to 0 if $r_n = O(n^\gamma)$ for some $\gamma < 1$. Moreover, $B_1 \subset B_{11} \cup B_{12}$, where

$$B_{11} = \{\varphi^m Y_0 > 0.5\delta \log(ya_n)\} \quad \text{and} \quad B_{12} = \left\{\sum_{i=1}^{m} \varphi^{m-i} \eta_i > 0.5\delta \log(ya_n)\right\}.$$

For any $m$, small $\varepsilon > 0$, large $n$, we have

$$n P\big(B_{11} \cap \{Y_0 > \log(ya_n)\}\big) = n P\big(Y_0 > 0.5\delta \log(ya_n)\varphi^{-m}\big) \leq n(ya_n)^{-0.5\delta\varphi^{-m}(\alpha-\varepsilon)}.$$

Therefore, choosing $m$ sufficiently large, the right-hand side converges to 0. Because $Y_0$ and $B_{12}$ are independent,

$$P\big(B_{12} | Y_0 > \log(ya_n)\big) = P(B_{12}).$$

The right-hand side is bounded by $P(Y > 0.5\delta \log(ya_n)) = o(1)$. Thus, we have proven

$$\lim_{n\to\infty} P\big(B_1 | Y_0 > \log(ya_n)\big) = 0$$

and that (3.8) holds. Next, we prove

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\max_{-r_n \leq t \leq -m} Y_t > \log(ya_n) \big| Y_0 > \log(ya_n)\right) = 0. \tag{3.10}$$

Write

$$C = \left\{\max_{-r_n \leq t \leq -m} Y_t > \log(ya_n), Y_0 > \log(ya_n)\right\}.$$

Again using (3.9), we see that $C \subset C_1 \cup C_2$, where, for $\delta \in (0, 1)$,

$$C_1 = \{Y_{-r_n} > \delta \log a_n, Y_0 > \log(ya_n)\},$$

$$C_2 = \left\{\max_{-r_n \leq t \leq -m} \sum_{i=-r_n+1}^{t} \varphi^{t-i} \eta_i > (1-\delta)\log(ya_n), Y_0 > \log(ya_n)\right\}.$$

Another application of (3.9) and stationarity yields

$$nP(C_1) \leq nP\big(Y_0 > \delta \log(ya_n), Y_0 > (1-\delta)\varphi^{-r_n} \log(ya_n)\big)$$

$$+ nP\bigg(Y_0 > \delta \log(ya_n), \sum_{i=1}^{r_n} \varphi^{r_n-i} \eta_i > \delta \log(ya_n)\bigg) = I_1 + I_2.$$

By regular variation, for small $0 < \varepsilon < \alpha$ and large $n$,

$$I_1 \leq n(ya_n)^{-(\alpha-\varepsilon)(1-\delta)\varphi^{-r_n}}.$$

Because $r_n \to \infty$, we have $I_1 = o(1)$ as $n \to \infty$. Moreover, it follows that

$$\limsup_{n\to\infty} I_2 \leq c \limsup_{n\to\infty} P\bigg(\sum_{i=1}^{r_n} \varphi^{r_n-i} \eta_i > \delta \log(ya_n)\bigg) = 0.$$

Thus, we have proven that $\limsup_{n\to\infty} nP(C_1) = 0$. For $C_2$, we have $C_2 \subset C_{21} \cup C_{22}$, where

$$C_{21} = \bigg\{\max_{-r_n \leq t \leq -m} \sum_{i=-r_n+1}^{t} \varphi^{t-i} \eta_i > (1-\delta)\log(ya_n), \varphi^m Y_{-m} > \delta \log(ya_n)\bigg\},$$

$$C_{22} = \bigg\{\max_{-r_n \leq t \leq -m} \sum_{i=-r_n+1}^{t} \varphi^{t-i} \eta_i > (1-\delta)\log(ya_n), \sum_{i=-m+1}^{0} \varphi^{-i} \eta_i > (1-\delta)\log(ya_n)\bigg\}.$$

Thus, for small $\varepsilon$, large $m$,

$$nP(C_{21}) \leq nP\big(Y_0 > \varphi^{-m}\delta \log(ya_n)\big) \leq n(ya_n)^{-(\alpha-\varepsilon)\delta\varphi^{-m}} \to 0, \qquad n \to \infty,$$

and for small $\varepsilon, \delta$,

$$nP(C_{22}) = nP\bigg(\max_{-r_n \leq t \leq -m} \sum_{i=-r_n+1}^{t} \varphi^{t-i} \eta_i > (1-\delta)\log(ya_n)\bigg)$$

$$\times P\bigg(\sum_{i=-m+1}^{0} \varphi^{-i} \eta_i > (1-\delta)\log(ya_n)\bigg)$$

$$\leq nr_n\big[P\big(Y > (1-\delta)\log(ya_n)\big)\big]^2 \leq nr_n(ya_n)^{-2(\alpha-\varepsilon)(1-\delta)} = o(1),$$

provided that $r_n = O(n^\gamma)$ for some $\gamma < 1$. For general $\eta$ and $|\varphi| < 1$, we see that $|Y_t| \leq \sum_{i=j} |\varphi|^j |\eta_{t-j}|$. We can apply the same reasoning as above, using (3.7).

We now turn to the proof of the anticlustering condition for $(X_t)$. An inspection of the foregoing proof shows that we have to add the terms $R_t = \log |Z_t|$ to $|Y_t|$. We restrict ourselves to the cases $\varphi \in (0,1)$, $\eta > 0$ a.s., and only show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\Big(\max_{m \leq t \leq r_n} (Y_t + R_t) > \log(ya_n) \big| Y_0 + R_0 > \log(ya_n)\Big) = 0, \qquad y > 0. \quad (3.11)$$

We use the same notation for the modified events. We start by observing that

$$B = \left\{ \max_{m \leq t \leq r_n} (Y_t + R_t) > \log(ya_n) \right\} \subset B_1 \cup B_2,$$

where $B_2$ is the same as above and

$$B_1 = \left\{ Y_m + \max_{m \leq t \leq r_n} R_t > \delta \log(ya_n) \right\}$$

$$\subset \{Y_m > 0.5\delta \log(ya_n)\} \cup \left\{ \max_{m \leq t \leq r_n} R_t > 0.5\delta \log(ya_n) \right\} = D_1 \cup D_2.$$

Now $P(D_1)$ can be treated in the same way as $P(B_1)$ in the foregoing proof. If $|Z|$ has moments of any order $h > 0$, then an application of Markov's inequality for sufficiently large $h$ yields, for any choice of $r_n = o(n)$,

$$P(D_2) = P\left( \max_{m \leq t \leq r_n} |Z_t| > (ya_n)^{0.5\delta} \right) \leq r_n P\left( |Z| > (ya_n)^{0.5\delta} \right) \leq cr_n(ya_n)^{-0.5h\delta} = o(1).$$

On the other hand, if $r_n = O(n^\gamma)$ for every small $\gamma$, then Markov's inequality of order $h = \alpha + \xi$ yields the same result by choosing $\gamma$ close to 0. This completes the proof of (3.11). $\square$

## 3.4. Main result for the exponential AR(1) process

Here we give sufficient conditions for the validity of Theorem 2.4 when $(X_t)$ is a stochastic volatility process and the volatility process $(\sigma_t)$ is an exponential AR(1) process. The result is a consequence of the translation result Theorem 2.6 and the foregoing calculations.

**Theorem 3.5.** *Consider the stochastic volatility model* (1.1), *where the volatility sequence* $(\sigma_t)$ *is an exponential* AR(1) *process* (3.1) *for some* $\varphi \in (-1, 1)$. *Assume the following conditions*:

- *The regular variation conditions* (3.3) *and* (3.7) *hold for some index* $\alpha > 0$.
- *The random variable* $\eta$ *has positive density in some neighborhood of* $E\eta$.

*Then the following properties hold for* $(\sigma_t)$.

(1) *Regular variation with index* $\alpha$ *and limiting measures given in* (3.6).
(2) $\beta$-*mixing with geometric rate and* $\mathcal{A}(a_n)$ *are satisfied for any sequence* $(r_n)$ *satisfying* $r_n \geq c \log n$ *for some* $c > 0$ *and* $r_n = o(n)$.
(3) *The anticlustering condition for* $r_n = O(n^\gamma)$ *for any* $\gamma \in (0, 1)$.

*The following properties hold for the stochastic volatility process* $(X_t)$:

(4) *The strong mixing property with geometric rate and* $\mathcal{A}(a_n)$ *are satisfied for any sequence* $(r_n)$ *satisfying* $r_n \geq c \log n$ *for some* $c > 0$ *and* $r_n = o(n)$.

*Also assume that*

- $E|Z|^{\alpha+\delta} < \infty$ *for some* $\delta > 0$.

*Then*

(5) $(X_t)$ is regularly varying with index $\alpha$ and limiting measures given in Example 2.2.
(6) $(X_t)$ satisfies the anticlustering condition (2.4) for $(r_n)$ such that $r_n = \mathrm{O}(n^\gamma)$ for every $\gamma < 1$.

*Moreover, if*

- *Z has all moments,*

*then*

(7) $(X_t)$ satisfies the anticlustering condition (2.4) for any sequence $(r_n)$ such that $r_n = \mathrm{O}(n^\gamma)$ for some $\gamma < 1$.

*In particular, Theorem 2.4 applies to the sequences $(\sigma_t)$ and $(X_t)$.*

**Proof.** We first give the proof for the volatility sequence $(\sigma_t)$. Regular variation of $(\sigma_t)$ follows from Lemma 3.3, and $\beta$-mixing with geometric rate follows from Section 3.1. It follows from [10] and references therein that condition $\mathcal{A}(a_n)$ is satisfied with $r_n \geq c \log n$ for some $c > 0$. Condition (2.4) for $(\sigma_t)$ follows from Lemma 3.4 under the assumption that $r_n = \mathrm{O}(n^\gamma)$ for some $\gamma \in (0, 1)$.

Because $\beta$-mixing with geometric rate implies strong mixing with geometric rate and, using the argument on page 258 of [14], it follows that $(X_t)$ is strongly mixing with geometric rate. It follows from [10] and references therein that condition $\mathcal{A}(a_n)$ is satisfied for any $r_n \geq c \log n$ for some $c > 0$. Regular variation of $(X_t)$ follows from Theorem 2.6, and the limiting measures are derived in Example 2.2. Finally, condition (2.4) was verified in Lemma 3.4. $\square$

Using the machinery in [3,10–12], we can now derive various limit results for the sequence $(X_t)$. These include infinite variance limits for the normalized partial sums $\sum_{t=1}^{n} X_t$ and sample covariances $\sum_{t=1}^{n-h} X_t X_{t+h}$ in the case where $\alpha < 2$. For general $\alpha > 0$, the fact that the limit measures of the regular variation of $(X_t)$ are concentrated on the axes implies that the normalized partial maxima of $(X_t)$ converge to a Fréchet distribution,

$$\lim_{n \to \infty} P\left(a_n^{-1} \max_{t=1,\dots,n} X_t \leq x\right) = \Phi_\alpha(x) = \mathrm{e}^{-px^{-\alpha}}, \qquad x > 0, \tag{3.12}$$

where $(a_n)$ satisfies $n P(|X| > a_n) \to 1$ and

$$\lim_{x \to \infty} \frac{P(X > x)}{P(|X| > x)} = \frac{E Z_+^\alpha}{E |Z|^\alpha} = p \in [0, 1].$$

Relation (3.12) means that the extremal index of the sequence $(X_t)$ is 1; that is, we get the same result as for an i.i.d. sequence $(\widetilde{X}_t)$ with $\widetilde{X} \stackrel{d}{=} X$. In other words, the stochastic volatility model does not exhibit extremal clustering. This is analogous to stochastic volatility models in which $E\sigma^{\alpha+\delta} < \infty$ and $Z$ is regularly varying with index $\alpha$ (see [12,13]), although the reasons are very different in the two cases. Figure 1 presents graphs of regularly varying stochastic volatility models with light-tailed and heavy-tailed multiplicative noise. In the present case, the structure of
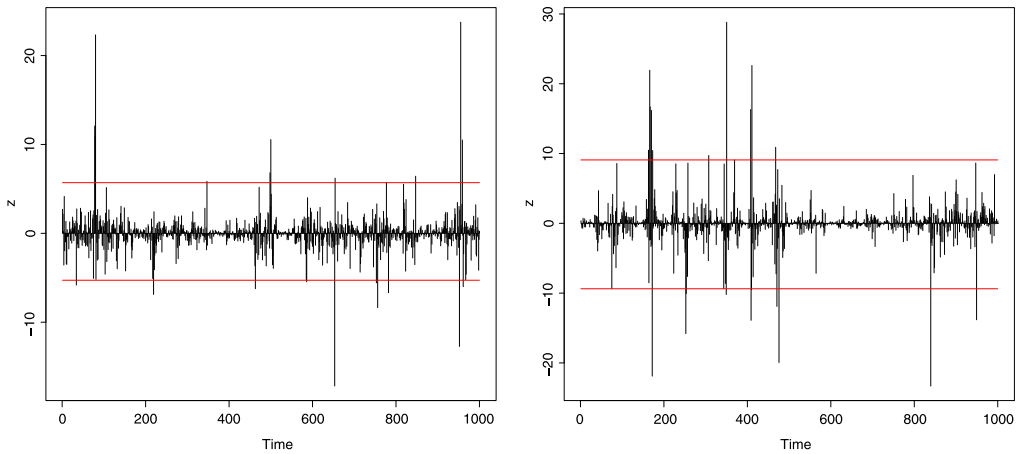
**Figure 1.** 1000 realizations of a stochastic volatility model, where $(\log \sigma_t)$ is an AR(1) process with $\varphi = 0.9$. The parallel lines indicate the 0.01 and 0.99 quantiles of the distribution of $X$. *Left:* The random variable $\eta$ is Laplace distributed: $P(X > x) = P(X \le -x) = 0.5 e^{-4x}$, $x > 0$, and $Z$ standard normal. *Right:* The random variable $\eta$ is $N(0, 0.25)$-distributed and $Z$ is $t$-distributed with 4 degrees of freedom standardized to unit variance. In both graphs, $(X_t)$ is regularly varying with index 4, and there is no extremal clustering in the sense that high and low exceedances of the lines occur separated through time.

the limiting measures for the regularly varying finite-dimensional distributions of the $\sigma$-sequence is responsible for the limiting measures of the $X$-sequence.

In passing, we mention that a condition of type (3.3) limits the choice of the distributions of the noise variable $\eta$ in the exponential AR(1) process. If $\eta$ has a slightly heavier right tail than suggested by (3.3) the random variable $Y$ will not have any moments. This occurs, for example, when $\eta$ has a lognormal or Student distribution. Thus regular variation of $(\sigma_t)$ and $(X_t)$ is possible only for a relatively thin class of noise variables $\eta$.

Before we consider other stochastic volatility models with genuine extremal clustering, we show that the EGARCH model from Example 3.1 is regularly varying and does not have extremal clusters.

***Example 3.6.*** Recall the definition of the EGARCH model from Example 3.1, particularly the dynamics of $(\sigma_t^2)$ given by (3.2). Writing $\eta_t = 0.5(\alpha_0(1 - \varphi)^{-1} + \gamma_0 Z_t + \delta_0|Z_t|)$ and assuming the conditions of Lemma 3.2, we conclude that $(\sigma_t)$ is regularly varying with index $\alpha$, and the limiting measures are concentrated on the axes. Using the modified Breiman lemma from [22], an inspection of the proof of Lemma 3.3 shows that $\boldsymbol{\Sigma}_d = (\sigma_1, \ldots, \sigma_d)'$ and $(e^{\eta_0}, \ldots, e^{\eta_{d-1}})' E e^{\alpha \varphi \eta}$ have the same limit measures of regular variation. Therefore, regular variation of $\mathbf{X}_d = (X_1, \ldots, X_d)'$ will follow if we can show that $\mathbf{R}_d = (Z_1 e^{\eta_0}, \ldots, Z_d e^{\eta_{d-1}})'$ is regularly varying with limit measures concentrated on the axes. By Breiman's result, $Z_1 e^{\eta_0}$ is regularly varying with index $\alpha$. Let $(a_n)$ be such that $nP(e^{\eta} > a_n) \to 1$. By construction, $Z$ has all moments, and thus we can choose a sequence $c_n \to \infty$ such that $nP(|Z| > c_n) \to 0$ and

$a_n/c_n \to \infty$. Then, for $d \geq 2$, $\delta > 0$,

$$
\begin{aligned}
nP(|Z_i \mathrm{e}^{\eta_{i-1}}| > \delta a_n, i = 1, \ldots, d) &\leq nP(|Z_1 \mathrm{e}^{\eta_0}| > \delta a_n, |Z_2 \mathrm{e}^{\eta_1}| > \delta a_n) \\
&\leq nP(|Z_1 \mathrm{e}^{\eta_0}| > \delta a_n) P(\mathrm{e}^{\eta_1} > \delta a_n/c_n) + nP(|Z| > c_n) \\
&= \mathrm{o}(1).
\end{aligned}
$$

Thus, if $nP(a_n^{-1} \mathbf{R}_d \in \cdot)$ has a non-vanishing vague limit, then it must be concentrated on the axes. To show this, we focus on the case where $d = 2$. Here, for $x, \delta > 0$, by Breiman's result and the previous calculations,

$$
\begin{aligned}
nP(a_n^{-1}|Z_1|\mathrm{e}^{\eta_0} \leq \delta, a_n^{-1} Z_2 \mathrm{e}^{\eta_1} > x) &= nP(a_n^{-1} Z_2 \mathrm{e}^{\eta_1} > x) \\
&\quad - nP(a_n^{-1}|Z_1|\mathrm{e}^{\eta_0} > \delta, a_n^{-1} Z_2 \mathrm{e}^{\eta_1} > x) \\
&\sim x^{-\alpha} E Z_+^{\alpha},
\end{aligned}
$$

$$
\begin{aligned}
nP(a_n^{-1}|Z_1|\mathrm{e}^{\eta_0} \leq \delta, a_n^{-1} Z_2 \mathrm{e}^{\eta_1} \leq -x) &= nP(a_n^{-1} Z_2 \mathrm{e}^{\eta_1} \leq -x) \\
&\quad - nP(a_n^{-1}|Z_1|\mathrm{e}^{\eta_0} > \delta, a_n^{-1} Z_2 \mathrm{e}^{\eta_1} \leq -x) \\
&\sim x^{-\alpha} E Z_-^{\alpha}.
\end{aligned}
$$

Therefore, $(X_t)$ is regularly varying with index $\alpha$, and the limiting measures are concentrated on the axes. Furthermore, if $\eta$ has a positive density in some neighborhood of $E\eta$, then $(\log \sigma_t)$, hence $(X_t)$, is strongly mixing with geometric rate, and then $\mathcal{A}(a_n)$ holds for any sequence $(r_n)$ satisfying $r_n = \mathrm{o}(n)$ and $r_n \geq c \log n$ for some $c > 0$. For the proof of the anticlustering condition of $(X_t)$, we can follow along the lines of the proof of Lemma 3.7, observing that $Z$ has all moments. Thus, the conditions of Theorem 2.4 are satisfied, in particular because the limiting measures of the regularly varying finite-dimensional distributions of $(X_t)$ are concentrated on the axes the extremal index $\theta_{|X|} = 1$, that is, there is no extremal clustering in this sequence.

# 4. Stochastic recurrence equations

We assume that the stationary sequence $(\sigma_t)$ satisfies the relation

$$
\sigma_t^p = A_t \sigma_{t-1}^p + B_t, \qquad t \in \mathbb{Z} \tag{4.1}
$$

for an i.i.d. sequence $((A_t, B_t))_{t \in \mathbb{Z}}$ of non-negative random variables and some positive $p$. Throughout we assume the conditions of Kesten [23], which ensure that (4.1) has a strictly stationary solution, namely $E \log A < 0$ and $E \log^+ B < \infty$.

***Example 4.1.*** For $p = 2$, a model of the type (4.1) has attracted major attention in the financial time series literature [1]: the GARCH process of order $(1, 1)$ (we write GARCH$(1, 1)$) given by $\widetilde{X}_t = \sigma_t \eta_t$, $t \in \mathbb{Z}$, $(\eta_t)$ is an i.i.d. centered sequence with unit variance and $\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2(\alpha_1 \eta_{t-1}^2 + \beta_1)$ for positive parameters $\alpha_i, \beta_1$. The main difference from the stochastic volatility model (1.1) with the same sequence $(\sigma_t)$ is that $\eta_t$ feeds into $(\sigma_{t+k})_{k \geq 1}$, and thus the noise $(\eta_t)$ and $(\sigma_t)$ are dependent.

## 4.1. Mixing property

It follows from [28] that $(\sigma_t^p)$ is strongly mixing with geometric rate if $A$, $B$ satisfy some regularity condition. In particular, if $A_t$ and $B_t$ are polynomials of an i.i.d. sequence $(\eta_t)$ and $\eta$ has a positive density in some neighborhood of $E\eta$, then $(\sigma_t)$ is $\beta$-mixing with geometric rate. Thus the GARCH(1, 1) model satisfies this condition for $p = 2$ if $\eta$ has a positive density in some neighborhood of the origin.

## 4.2. Regular variation

Regular variation of the marginal distribution of the solution to the stochastic recurrence equation (4.1) was proven by Kesten [23] and Goldie [18]. In particular, they showed that

$$P(\sigma^p > x) \sim cx^{-\alpha}, \qquad x \to \infty \tag{4.2}$$

for some constant $c > 0$. The index $\alpha$ is then obtained as the unique positive solution to the equation $EA^\kappa = 1$. Relation (4.2) holds under general conditions on $(A, B)$, which we do not give here. Regular variation of $(\sigma_t)$ is inherited by the solution to (4.1).

**Lemma 4.2.** *Assume the conditions of Kesten* [23] *for the stochastic recurrence equation* (4.1) *and the moment conditions $EA^{\alpha+\varepsilon} < \infty$ and $EB^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then $(\sigma_t)$ is regularly varying with index $\alpha p$ and for $\boldsymbol{\Sigma}_d = (\sigma_1, \ldots, \sigma_d)'$,*

$$\frac{P(x^{-1}\boldsymbol{\Sigma}_d \in \cdot)}{P(\sigma > x)}$$
$$\xrightarrow{v} \alpha p \int_0^\infty t^{-\alpha p - 1} P\big(t(1, A_1^{1/p}, \ldots, (A_{d-1} \cdots A_1)^{1/p})' \in \cdot\big)\,dt, \qquad x \to \infty. \tag{4.3}$$

*Moreover, if $E|Z|^{\alpha p + \delta} < \infty$ for some $\delta > 0$, then the stochastic volatility model $(X_t)$ is regularly varying with index $\alpha p$, and the limiting measure of $\mathbf{X}_d = (X_1, \ldots, X_d)'$ is given by*

$$\frac{P(x^{-1}\mathbf{X}_d \in \cdot)}{P(|X| > x)}$$
$$\xrightarrow{v} \frac{\alpha p}{E|Z|^{\alpha p}} \int_0^\infty t^{-\alpha p - 1} P\big(t(Z_1, Z_2 A_1^{1/p}, \ldots, Z_d(A_{d-1} \cdots A_1)^{1/p})' \in \cdot\big)\,dt. \tag{4.4}$$

*If $Z$ is symmetric and $P(Z = 0) = 0$, then the limit in* (4.4) *turns into*

$$\alpha p \int_0^\infty t^{-\alpha p - 1} P\big(t(\text{sign}(Z_1), (Z_2/|Z_1|)A_1^{1/p}, \ldots, (Z_d/|Z_1|)(A_{d-1} \cdots A_1)^{1/p})' \in \cdot\big)\,dt.$$

**Proof.** We take the approach in the proof of Corollary 2.7 in [3]. For every $t$, we have

$$\sigma_t^p = A_t \cdots A_1 \sigma_0^p + \sum_{i=1}^t A_t \cdots A_{i+1} B_i, \tag{4.5}$$

and thus, applying the power operation component-wise,

$$\boldsymbol{\Sigma}_d^p = \sigma_0^p (A_1, A_2 A_1, \ldots, A_d \cdots A_1)' + \mathbf{R}_d,$$

where, by virtue of the moment conditions on $(A, B)$, $E|\mathbf{R}_d|^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$. By Kesten's theorem (cf. (4.2)), as $x \to \infty$,

$$\frac{P(\sigma > xt)}{P(\sigma > x)} \to t^{-\alpha p} = \mu(t, \infty), \qquad t > 0.$$

Therefore, and in view of a version of the multivariate Breiman result (see [22]), $P(x^{-1}\boldsymbol{\Sigma}_d \in \cdot)/P(\sigma > x)$ has the same limit measure as

$$\frac{P(x^{-1}\sigma_0(A_1, A_2 A_1, \ldots, A_d \cdots A_1)'^{1/p} \in \cdot)}{P(\sigma > x)}$$

$$\xrightarrow{v} E\mu\{t > 0: t(A_1, A_2 A_1, \ldots, A_d \cdots A_1)'^{1/p} \in \cdot\}$$

$$= \alpha p E A^{\alpha} \int_0^{\infty} t^{-\alpha p - 1} P\big(t(1, A_2, \ldots, A_d \cdots A_2)'^{1/p} \in \cdot\big) \, dt$$

$$= \alpha p \int_0^{\infty} t^{-\alpha p - 1} P\big(t(1, A_1, \ldots, A_{d-1} \cdots A_1)'^{1/p} \in \cdot\big) \, dt.$$

Relation (4.4) follows by an application of the multivariate Breiman result; compare Lemma 2.1. $\qquad\square$

## 4.3. Anticlustering condition

**Lemma 4.3.** *Assume that the conditions of Lemma 4.2 are satisfied, ensuring that $(\sigma_t)$ is regularly varying with index $\alpha p$. Then the anticlustering condition (2.4) is satisfied for $(\sigma_t)$ for a sequence $(r_n)$ satisfying $r_n = O(n^{\gamma})$ for any small $\gamma > 0$. Moreover, if $E|Z|^{\alpha p + \delta} < \infty$ for some $\delta > 0$, then (2.4) also holds for $(X_t)$ with the same sequence $(r_n)$.*

**Proof.** Condition (2.4) for $(\sigma_t)$ follows from the proof of Theorem 2.10 in [3]. Indeed, [3] used (4.5) to show that

$$\lim_{m \to \infty} \limsup_{n \to \infty} \sum_{m \le |t| \le r_n} P(\sigma_t > a_n y | \sigma_0 > a_n y) = 0, \qquad y > 0. \tag{4.6}$$

The corresponding result for $(X_t)$ follows along the lines of the proof of (4.6), exploiting (4.5) and the independence of $(\sigma_t)$ and $(Z_t)$. $\qquad\square$

## 4.4. Main result for solution to stochastic recurrence equation

We formulate an analog of Theorem 3.5, summarizing the foregoing results in the case of a solution to a stochastic recurrence equation.

**Theorem 4.4.** *Assume that the volatility sequence $(\sigma_t)$ is given via the solution $(\sigma_t^p)$ of the stochastic recurrence equation* (4.1) *for some $p > 0$. Assume the following conditions*:

- $(\sigma_t^p)$ *satisfies Kesten's* [23] *conditions.*
- $(\sigma_t)$ *is strongly mixing with geometric rate.*

*Then*

(1) $(\sigma_t)$ *is regularly varying with index $\alpha p$ and limiting measures given in* (4.3).
(2) *Condition $\mathcal{A}(a_n)$ is satisfied for any $(r_n)$ satisfying $r_n = \mathrm{o}(n)$ and $r_n \geq c \log n$ for some $c > 0$.*
(3) *The anticlustering condition* (2.4) *holds for $(\sigma_t)$ with a sequence $(r_n)$ satisfying $r_n = \mathrm{O}(n^\gamma)$ for any small $\gamma > 0$.*

*Moreover, if $E|Z|^{\alpha p + \delta} < \infty$ for some $\delta > 0$, then the following hold*:

(4) $(X_t)$ *is regularly varying with index $\alpha p$ and limiting measures given in* (4.4).
(5) $(X_t)$ *is strongly mixing with geometric rate, and condition $\mathcal{A}(a_n)$ is satisfied for any $(r_n)$ satisfying $r_n = \mathrm{o}(n)$ and $r_n \geq c \log n$ for some $c > 0$.*
(6) *The anticlustering condition* (2.4) *holds for $(X_t)$ and sequences $(r_n)$ satisfying $r_n = \mathrm{O}(n^\gamma)$ for any small $\gamma > 0$.*

*In particular, Theorem* 2.4 *is applicable to the sequences $(\sigma_t)$ and $(X_t)$.*

Now we can again use the machinery of [3,10–12] to derive various limit results for functionals of the sequence $(X_t)$. We only derive the extremal index of $(X_t)$ in a special situation, to show the crucial difference between the exponential AR(1) process considered in Section 3 and the present situation.

***Example 4.5.*** Recall the definition of a GARCH(1, 1) process from Example 4.1. We assume that $(\sigma_t^2)$ is the squared volatility process of a GARCH(1, 1) process that is regularly varying with index $\alpha > 0$ and $E|Z|^{2\alpha + \delta} < \infty$ for some $\delta > 0$. Such a GARCH(1, 1) process and the corresponding stochastic volatility model are shown in Figure 2. Assume that both $(\sigma_t)$ and the corresponding stochastic volatility model $(X_t)$ satisfy the conditions of Theorem 2.4; sufficient conditions are given in Theorem 4.4. It is well known (e.g., [27], Theorem 4.1) that the extremal indices of $(\sigma_t)$ and $(\sigma_t^2)$ coincide and are given by

$$\theta_\sigma = \alpha \int_1^\infty P\left(\sup_{t \geq 1} \prod_{j=1}^t A_j \leq y^{-1}\right) y^{-1-\alpha}\, dy,$$

where $A_j = \alpha_1 \eta_j^2 + \beta_1$, $j \geq 1$. For the extremal index $\theta_{|X|}$ of the sequences $(|X_t|)$ and $(X_t^2)$, we use the expression in [10] given by

$$\theta_{|X|} = \lim_{m \to \infty} \frac{(|\theta_0^{(m)}|^\alpha - \max_{j=1,\dots,m} |\theta_j^{(m)}|^\alpha)_+}{E|\theta_0^{(m)}|^\alpha}, \tag{4.7}$$
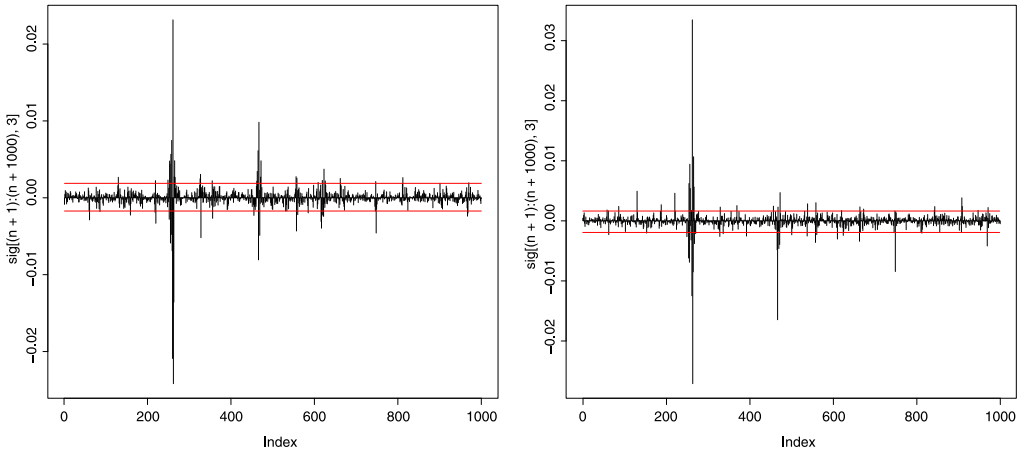
**Figure 2.** *Left:* 1000 realizations of a GARCH(1, 1) process with parameters $\alpha_0 = 10^{-7}$, $\alpha_1 = 0.1$, $\beta_1 = 0.89$ and i.i.d. standard normal noise. *Right:* Realizations of a stochastic volatility model, where $(\sigma_t)$ is taken from the GARCH(1, 1) process in the left graph and $Z$ is standard normal. In both graphs, $(\sigma_t)$ and $(X_t)$ are regularly varying with index 4 causing extremal clustering in both sequences. The parallel lines indicate the 0.99 and 0.01 quantiles of the distribution of $X$.

where $\Theta^{(m)} = (\theta_j^{(m)})_{|j|\leq m}$ is a vector with values in the unit sphere $\mathbb{S}^{2m}$ of $\mathbb{R}^{2m+1}$, which has the spectral distribution of the random vector $\widehat{\mathbf{X}}^{(m)} = (X_t^2)_{|t|\leq m}$, that is,

$$\frac{P(|\widehat{\mathbf{X}}^{(m)}| > x, \widehat{\mathbf{X}}^{(m)}/|\widehat{\mathbf{X}}^{(m)}| \in \cdot)}{P(|\widehat{\mathbf{X}}^{(m)}| > x)} \xrightarrow{w} P(\Theta^{(m)} \in \cdot), \qquad x \to \infty.$$

For any Borel set $S \subset \mathbb{S}^{2m}$ that is a continuity set with respect to $P(\Theta^{(m)} \in \cdot)$, we conclude from (4.4) with $\mathbf{R}^{(m)} = (Z_1^2, Z_2^2 A_1, \ldots, Z_{2m+1}^2 A_{2m} \cdots A_1)'$

$$\frac{P(|\widehat{\mathbf{X}}^{(m)}| > x, \widehat{\mathbf{X}}^{(m)}/|\widehat{\mathbf{X}}^{(m)}| \in S)}{P(|\widehat{\mathbf{X}}^{(m)}| > x)} \to \frac{\alpha \int_0^\infty t^{-\alpha-1} P(t|\mathbf{R}^{(m)}|I_{\{\mathbf{R}^{(m)}/|\mathbf{R}^{(m)}|\in S\}} > 1)\,\mathrm{d}t}{E|\mathbf{R}^{(m)}|^\alpha}$$

$$= \frac{E|\mathbf{R}^{(m)}|^\alpha I_{\{\mathbf{R}^{(m)}/|\mathbf{R}^{(m)}|\in S\}}}{E|\mathbf{R}^{(m)}|^\alpha} = P(\Theta^{(m)} \in S).$$

The latter relation, (4.7), and the fact that $E A^\alpha = 1$ yield

$$\theta_{|X|} = \lim_{m\to\infty} \frac{E(|Z_1|^{2\alpha} - \max_{j=2,\ldots,m}(Z_j^2 \prod_{i=2}^j A_i)^\alpha)_+}{E|Z|^{2\alpha}}.$$

A comparison with Theorem 4.1 in [27] shows that a similar expression can be derived for the extremal index $\theta_{|X|}$ of the GARCH(1, 1) process; the $Z$'s must be replaced by the corresponding $\eta$'s. (For details on the foregoing calculations, see [27].) A direct comparison of the magnitude

of the extremal indices of a GARCH$(1, 1)$ process and the corresponding stochastic volatility model seems difficult.

## 5. Moving average processes

In this section we assume that the volatility process $(\sigma_t)$ is given in the form $\sigma_t^p = |Y_t|$ for some $p > 0$ and $Y_t = \sum_{j=0}^q \psi_j \eta_{t-j}$, $t \in \mathbb{Z}$, for some $q \geq 1$ and an i.i.d. sequence $(\eta_t)$ such that $\eta$ is regularly varying in the sense of (1.2) with tail balance coefficients $\widetilde{p}, \widetilde{q} \geq 0$, $\widetilde{p} + \widetilde{q} = 1$ and index $\alpha > 0$. Because $\mathbf{Y}_d = (Y_1, \ldots, Y_d)'$ has representation as a linear transformation of a finite vector of the $Z$'s, an application of the continuous mapping theorem implies that the vector $\mathbf{Y}_d$ is regularly varying with index $\alpha$. Writing $\psi_d = 0$ for $d \notin \{0, \ldots, q\}$, we conclude from [9], Theorem 2.4, that

$$
\frac{P(x^{-1}\mathbf{Y}_d \in \cdot)}{P(|\eta| > x)} \overset{v}{\to} \alpha \sum_{j=0}^{q+d-1} \int_{\mathbb{R}_0} |x|^{-\alpha-1} \big[\widetilde{p} I_{(0,\infty)}(x) + \widetilde{q} I_{(-\infty,0)}(x)\big]
$$
$$
\times I_{\{x(\psi_{j-d+1}, \ldots, \psi_j) \in \cdot\}} \, \mathrm{d}x.
\tag{5.1}
$$

The mixing condition $\mathcal{A}(a_n)$ and the anticlustering condition (2.4) are automatically satisfied for $(\sigma_t)$ and $(X_t)$. Thus Theorem 2.6 holds. We conclude from (5.1) and Breiman's result that

$$
\frac{P(x^{-1}(Z_1\sigma_1, \ldots, Z_d\sigma_d) \in \cdot)}{P(|\eta| > x)}
$$
$$
\overset{v}{\to} \alpha p \sum_{j=0}^{q+d-1} \int_0^\infty |x|^{-\alpha p-1} P\big(x(Z_1|\psi_{j-d+1}|^{1/p}, \ldots, Z_d|\psi_j|^{1/p}) \in \cdot\big) \, \mathrm{d}x.
$$

An application of (4.7) yields

$$
\theta_{|X|} = \frac{E \max_{j=0,\ldots,q} |Z_j|^{\alpha p}|\psi_j|^\alpha}{E|Z|^{\alpha p} \sum_{j=0}^q |\psi_j|^\alpha}.
\tag{5.2}
$$

In the degenerate case when $Z = 1$, we get the well-known form of the extremal index of the absolute values of a moving average process (see [9]; cf. [16], page 415). Again, a direct comparison of the value (5.2) with the corresponding one for $Z = 1$ seems difficult.

**Remark 5.1.** The foregoing techniques can be applied in the case where $(Y_t)$ constitutes an infinite moving average process as well. However, in this case mixing conditions are generally difficult to check; instead, [9] used approximations of an infinite moving average by finite moving averages. This technique does not completely fit into the framework of [10]; see Theorem 2.4 above. However, if $(Y_t)$ is an ARMA process with i.i.d. noise $(\eta_t)$ that is regularly varying with index $\alpha > 1$ and has a positive density in some neighborhood of $E\eta$, then $(Y_t)$ is strongly mixing with geometric rate. Then $\mathcal{A}(a_n)$ holds for every sequence $(r_n)$ with $r \geq c \log n$ for some $c > 0$,

and $\mathcal{A}(a_n)$ also holds for $(X_t)$ and the same sequence $(r_n)$. The anticlustering condition for $(X_t)$ can be checked in this case as well, but the calculations are lengthy. We omit further details.

# 6. Concluding remarks

The aim of this paper was to show that the stochastic volatility model $(X_t)$ given by (1.1) may exhibit extremal clustering provided that $(\sigma_t)$ is a regularly varying sequence with index $\alpha > 0$ and the i.i.d. noise sequence $(Z_t)$ has $(\alpha + \varepsilon)$th moment for some $\varepsilon > 0$. Extremal clustering is inherited from the volatility sequence $(\sigma_t)$. If $(\sigma_t)$ does not have extremal clusters, then neither does the sequence $(X_t)$. An example of this lack of clustering is given by an exponential AR(1) process $\sigma_t = \mathrm{e}^{Y_t}$, $Y_t = \varphi Y_{t-1} + \eta_t$ for $\varphi \in (-1, 1)$ and an i.i.d. regularly varying sequence $(\mathrm{e}^{\eta_t})$. The results of Section 3 show that the sequence $(X_t)$ above high levels essentially behaves like the i.i.d. sequence $(\mathrm{e}^{\eta_t})$, resulting in an extremal index $\theta_{|X|} = 1$. This is surprising, given that the autocorrelation function of $(|X_t|)$ is not negligible. This example includes $(\sigma_t)$ given by the dynamics of an EGARCH process. The EGARCH process itself does then not exhibit extremal clustering either.

In contrast to an exponential AR(1), the stochastic volatility model (1.1) exhibits extremal clustering if the dynamics of $(\sigma_t)$ or some positive power of it are given by a moving average or the solution to a stochastic recurrence equation. The latter case captures the example of the volatility sequence of a GARCH(1, 1) process.

We have chosen to describe extremal clustering in terms of the extremal index of the sequence $(X_t)$. If $\theta_{|X|} < 1$, then evaluating this quantity is difficult in the examples considered. We would depend on numerical or Monte Carlo methods if we were interested in numerical values of $\theta_{|X|}$. These methods also would depend on the model.

The literature on the extremes of the stochastic volatility model focuses on the case where $(\sigma_t)$ is lognormal and $(Z_t)$ is i.i.d. normal or regularly varying (cf. [13]). In these cases, $(X_t)$ does not have extremal clusters. The latter property can be considered a disadvantage for modeling return series that are believed to have the clustering property, often referred to as volatility clusters. From a modeling standpoint, neither the stochastic volatility model with or without extremal clusters nor any standard model such as GARCH or EGARCH can be discarded as long as no efficient methods for distinguishing between these models exist. For example, the volatility dynamics of an EGARCH model and a stochastic volatility model with exponential AR(1) volatility are rather similar and so are the volatility dynamics of a GARCH(1, 1) and a stochastic volatility model with GARCH(1, 1) volatility.

# Acknowledgements

# References

[1] Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, T. (2009). *Handbook of Financial Time Series*. Berlin: Springer.

[2] Basrak, B., Davis, R.A. and Mikosch, T. (2002). A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12** 908–920. MR1925445

[3] Basrak, B., Davis, R.A. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99** 95–115. MR1894253

[4] Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Process. Appl.* **119** 1055–1080. MR2508565

[5] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR0898871

[6] Boman, J. and Lindskog, F. (2009). Support theorems for the Radon transform and Cramér-Wold theorems. *J. Theoret. Probab.* **22** 683–710. MR2530109

[7] Breidt, F.J. and Davis, R.A. (1998). Extremes of stochastic volatility models. *Ann. Appl. Probab.* **8** 664–675. MR1627756

[8] Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10** 323–331.

[9] Davis, R.A. and Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13** 179–195. MR0770636

[10] Davis, R.A. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23** 879–917. MR1334176

[11] Davis, R.A. and Mikosch, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26** 2049–2080. MR1673289

[12] Davis, R.A. and Mikosch, T. (2001). Point process convergence of stochastic volatility processes with application to sample autocorrelation. *J. Appl. Probab.* **38A** 93–104. MR1915537

[13] Davis, R.A. and Mikosch, T. (2009). Extremes of stochastic volatility models. In *Handbook of Financial Time Series* (T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch, eds.) 355–364. Berlin: Springer.

[14] Davis, R.A. and Mikosch, T. (2009). Fundamental properties of stochastic volatility models. In *Handbook of Financial Time Series* (T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch, eds.) 255–267. Berlin: Springer.

[15] Doukhan, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**. New York: Springer. MR1312160

[16] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics* (*New York*) **33**. Berlin: Springer. MR1458613

[17] Embrechts, P. and Veraverbeke, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.* **1** 55–72. MR0652832

[18] Goldie, C.M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1** 126–166. MR1097468

[19] Hult, H. and Lindskog, F. (2005). Extremal behavior of regularly varying stochastic processes. *Stochastic Process. Appl.* **115** 249–274. MR2111194

[20] Hult, H. and Lindskog, F. (2006). On Kesten's counterexample to the Cramér–Wold device for regular variation. *Bernoulli* **12** 133–142. MR2202325

[21] Hult, H. and Lindskog, F. (2006). Regular variation for measures on metric spaces. *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **80(94)** 121–140. MR2281910

[22] Jessen, A.H. and Mikosch, T. (2006). Regularly varying functions. *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **80(94)** 171–192. MR2281913

[23] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248. MR0440724

[24] Klüppelberg, C. and Pergamenchtchikov, S. (2007). Extremal behaviour of models with multivariate random recurrence representation. *Stochastic Process. Appl.* **117** 432–456. MR2305380

[25] Kulik, R. and Soulier, P. (2011). The tail empirical process for long memory stochastic volatility sequences. *Stochastic Process. Appl.* **121** 109–134. MR2739008

[26] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. *Springer Series in Statistics*. New York: Springer. MR0691492

[27] Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process. *Ann. Statist.* **28** 1427–1451. MR1805791

[28] Mokkadem, A. (1990). Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. Henri Poincaré Probab. Stat.* **26** 219–260. MR1063750

[29] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59** 347–370. MR1097532

[30] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. New York: Springer. MR0900810

[31] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2271424