Regular Variation of GARCH Processes

Bojan Basrak[1]
EURANDOM

Richard A. Davis[2]
Colorado State University

Thomas Mikosch[3]
University of Copenhagen

Short Title: Regular Variation of GARCH Processes

Abstract

We show that the finite-dimensional distributions of a GARCH process are regularly varying, i.e., the tails of these distributions are Pareto-like and hence heavy-tailed. Regular variation of the joint distributions provides insight into the moment properties of the process as well as the dependence structure between neighboring observations when both are large. Regular variation also plays a vital role in establishing the large sample behavior of a variety of statistics from a GARCH process including the sample mean and the sample autocovariance and autocorrelation functions. In particular, if the 4th moment of the process does not exist, the rate of convergence of the sample autocorrelations becomes extremely slow, and if the 2nd moment does not exist, the sample autocorrelations have non-degenerate limit distributions.

[1]This research supported by an NWO PhD grant.
[2]This research supported in part by NSF DMS Grant No. DMS-9972015.
[3]This research supported by the European Research Mobility Programme Network DYNSTOCH.
AMS 1991 Subject Classification: Primary: 62M10 Secondary: 62G20 60G55 62P05 60G10 60G70
Key Words and Phrases. Point process, vague convergence, multivariate regular variation, mixing condition, stationary process, heavy tail, sample autocovariance, sample autocorrelation, GARCH, finance, Markov chain.

1 Introduction

In this paper we study the tail behavior of a generalized autoregressive conditionally heteroscedastic (GARCH) process (see Section 3.1 for a definition). Such models are widely used for modeling financial returns, i.e., relative changes of prices such as stock indices, share prices of stock, foreign exchange rates, etc. We refer to the collection [16] of original articles on GARCH processes and their applications in finance. It turns out that the finite-dimensional distributions of such processes exhibit quite an interesting feature: they are in most instances multivariate regularly varying. Regular variation is a consequence of the fact that the squares of a stationary GARCH process can be embedded in a multivariate linear stochastic recurrence equation. For this type of recursion equation, an advanced theory exists that provides conditions for the existence of a unique stationary solution to the system and describes the tail behavior of the distribution of the stationary solution.

One of the aims of this paper is to prove that GARCH processes have regularly varying tails. This implies in particular that sufficiently high-order moments of these processes do not exist. This is a well known fact; see for example Bollerslev [6]. Our results, however, are more refined since we can make precise statements about the asymptotic form of the tails, not only of the univariate marginal distribution, but also about the tails of the finite-dimensional distributions.

The regular variation of the finite-dimensional distributions of GARCH processes is consistent with the "heavy-tailedness" exhibited by real-life log-return data. Indeed, there is plenty of statistical evidence that financial log-returns of foreign exchange rates, composite stock indices or share prices of stock can have infinite 5th, 4th or even 3rd moments; see for example Chapters 6 and 7 of Embrechts et al. [15] where statistical methods for measuring the thickness of tails are also provided. This in turn requires study of the behavior of standard statistical tools such as the sample autocorrelations under the assumption that the data come from a GARCH model with non-existing 2nd or 4th moments. Perhaps not surprising, the heavier the tails of the process, the slower the rate of convergence of the sample autocorrelations, or even worse, the sample autocorrelations converge weakly without normalization to a non-degenerate limit that involves ratios of infinite variance stable random variables. This implies standard theory for the sample autocorrelations does not apply for GARCH processes when certain moments are infinite. In view of the common practice to consider not only the sample autocorrelations of log-returns but also their squares, absolute values and other powers, a comprehensive limit theory of autocorrelations of functions of GARCH processes is needed.

In this paper we focus on the tail behavior of the finite-dimensional distributions of GARCH processes and its consequences for the large sample behavior of the sample autocovariances and autocorrelations. Our efforts are a continuation of the work started in Davis and Mikosch [13] for the ARCH(1) case and in Mikosch and Stărică [29] for the GARCH(1,1) case. As in the latter paper, the squares of a GARCH process will be embedded in a linear stochastic recurrence equation. For this reason we give in Section 2 some theory for linear stochastic recurrence equations including conditions on the noise distribution and model coefficients for the existence of a stationary solution that has regularly varying tail probabilities. Part of these results are known, but we include them here because they are needed in various proofs throughout the paper. Section 2 gives a survey of results about stochastic recurrence equations which are scattered over the literature and which may be useful also for other kinds of models such as multivariate GARCH processes. The conditions and results given there also show that the probabilistic properties of solutions to general stochastic recurrence equations, and of GARCH models in particular, require some advanced technology which does not always yield results in a sufficiently explicit form for practical implementation. In Section 3 we apply the stochastic recurrence equation results to the GARCH process. In particular, we show how the squares of a GARCH process can be embedded in a stochastic recurrence equation and therefore the finite-dimensional distributions of such processes are regularly varying. We also study the consequences for the asymptotic behavior of the sample autocovariances and autocorrelations.

2 Basic theory for stochastic recurrence equations

Consider a d-dimensional time series $(\mathbf{X}_t)$ given by a *stochastic recurrence equation* (SRE)

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t\,, \quad t \in \mathbb{Z}\,, \tag{2.1}$$

for some iid sequence $((\mathbf{A}_t, \mathbf{B}_t))$ of random $d \times d$ matrices $\mathbf{A}_t$ and d-dimensional vectors $\mathbf{B}_t$. By $|\cdot|$ we denote the Euclidean norm in $\mathbb{R}^d$, and by $\|\cdot\|$ the corresponding operator norm, i.e., for any $d \times d$-matrix $\mathbf{A}$,

$$\|\mathbf{A}\| = \sup_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|\,.$$

By $\mathbf{A} > 0$ we mean that all entries of $\mathbf{A}$ are positive.

2.1 Existence of a stationary solution

There exist various results about the existence of a strictly stationary solution to (2.1); see for example Kesten [23], Vervaat [36], Bougerol and Picard [7]. Below we recall a sufficient condition which remains valid for ergodic sequences $((\mathbf{A}_n, \mathbf{B}_n))$ (see Brandt [10]) and which is close to necessity (see Babillot et al. [1]). These conditions involve the notion of the *Lyapunov exponent for a sequence of random $d \times d$ matrices* $(\mathbf{A}_n)$ given by

$$\gamma = \inf \left\{ \frac{1}{n} E \ln \|\mathbf{A}_1 \cdots \mathbf{A}_n\|\,, \quad n \in \mathbb{N} \right\}\,. \tag{2.2}$$

If $E\ln^+\|\mathbf{A}_1\| < \infty$, an application of the subadditive ergodic theorem (see Kingman [24] or results in Furstenberg and Kesten [18]) yields that

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathbf{A}_1 \cdots \mathbf{A}_n\| \quad \text{a.s.} \tag{2.3}$$

In most cases of interest, γ cannot be calculated explicitly when $d > 1$. However, relation (2.3) offers a potential method for determining the value of γ, via Monte-Carlo simulations of the random matrices $\mathbf{A}_n$. Work by Goldsheid [20] even allows one to give asymptotic confidence bands through a central limit theorem.

Theorem 2.1 *Assume $E\ln^+\|\mathbf{A}_1\| < \infty$, $E\ln^+|\mathbf{B}_1| < \infty$ and $\gamma < 0$. Then the series $(\mathbf{X}_t)$ defined by*

$$\mathbf{X}_n = \mathbf{B}_n + \sum_{k=1}^{\infty} \mathbf{A}_n \cdots \mathbf{A}_{n-k+1} \mathbf{B}_{n-k} \tag{2.4}$$

converges a.s., and is the unique strictly stationary causal solution of (2.1).

Notice that $\gamma < 0$ holds if $E\ln\|\mathbf{A}_1\| < 0$. The condition on γ in Theorem 2.1 is particularly simple in the case $d = 1$ since then

$$\frac{1}{n} E\ln|A_1 \cdots A_n| = E\ln|A_1| = \gamma\,.$$

Corollary 2.2 *Assume $d = 1$, $-\infty \le E\ln|A_1| < 0$ and $E\ln^+|B_1| < \infty$. Then the unique stationary solution of* (2.1) *is given by* (2.4).

2.2 The multivariate regular variation property

2.2.1 Definition

The d-dimensional random vector $\mathbf{X}$ is said to be *regularly varying with index $\alpha \ge 0$* if there exists a sequence of constants (a_n) and a random vector $\mathbf{\Theta}$ with values in $\mathbb{S}^{d-1}$ a.s., where $\mathbb{S}^{d-1}$ denotes the unit sphere in $\mathbb{R}^d$ with respect to the norm $|\cdot|$, such that for all $t > 0$,

$$n\,P(|\mathbf{X}| > t\,a_n\,,\,\mathbf{X}/|\mathbf{X}| \in \cdot\,) \stackrel{v}{\rightarrow} t^{-\alpha} P(\mathbf{\Theta} \in \cdot\,)\,, \quad \text{as } n \to \infty\,.$$

This is equivalent to the condition that for all $t > 0$,

$$\frac{P(|\mathbf{X}| > t\,x\,,\,\mathbf{X}/|\mathbf{X}| \in \cdot\,)}{P(|\mathbf{X}| > x)} \stackrel{v}{\rightarrow} t^{-\alpha} P(\mathbf{\Theta} \in \cdot\,)\,, \quad \text{as } x \to \infty\,, \tag{2.5}$$

cf. de Haan and Resnick [21], Resnick [32]. The symbol $\stackrel{v}{\rightarrow}$ stands for vague convergence on $\mathbb{S}^{d-1}$; vague convergence of measures is treated in detail in Kallenberg [22]. The distribution of $\mathbf{\Theta}$ is referred to as the *spectral measure* of $\mathbf{X}$. For further information on multivariate regular variation we refer to Resnick [32] and [33], Chapter 5.

A particular consequence of the vague convergence in (2.5) is that linear combinations of the components of a regularly varying vector $\mathbf{X}$ are regularly varying with the same index α. Specifically,

$$\text{(2.6)} \qquad \text{For all } \mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}\,, \qquad \lim_{u\to\infty} \frac{P((\mathbf{x},\mathbf{X}) > u)}{L(u)u^{-\alpha}} = w(\mathbf{x}) \quad \text{exists}\,,$$

where $L(u)$ is a slowly varying function, and w is a finite-valued function, $w(\mathbf{x}) = 0$ being possible for certain choices of $\mathbf{x} \neq \mathbf{0}$. It follows directly from (2.6) that the limit function w is homogeneous and has the form

$$\text{(2.7)} \qquad w(t\mathbf{x}) = t^{-\alpha} w(\mathbf{x})\,,$$

for all $t > 0$, $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ for some $\alpha \geq 0$. That is, for all $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, the random variable $(\mathbf{x},\mathbf{X})$ is regularly varying with index α.

In Basrak et al. [2] it was shown that the two definitions (2.5) and (2.6) are essentially equivalent, The motivation for studying this equivalence was the fact that Kesten's theorem given below for solutions to stochastic recurrence equations states regular variation in the sense of (2.6), not in the more general sense of (2.5).

Theorem 2.3 *Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$. Then* (2.5) *and* (2.6) *are equivalent provided one of the following two conditions holds:*

- *α is a positive non-integer.*
- *$\mathbf{X}$ has non-negative components and α is an odd integer.*

The distinction between integer and non-integer, non-negative-valued and $\mathbb{R}^d$-valued vectors $\mathbf{X}$ is essential. Kesten [23], Remark 4, mentions in the case $\alpha = 1$ that the assumption of non-negativity of $\mathbf{X}$ is close to necessity. We conjecture that the equivalence between (2.5) and (2.6) is indeed valid for even α's and non-negative-valued $\mathbf{X}$'s, but a proof has not yet been constructed.

2.2.2 Kesten's theorem

Under general conditions, the stationary solution to the SRE (2.1) satisfies a multivariate regular variation condition. This follows from work by Kesten [23] in the general case $d \geq 1$; for an alternative proof in the case $d = 1$ see Goldie [19]. We state a modification of Kesten's fundamental result (a combination of Theorems 3 and 4 in [23]).

Theorem 2.4 *Let $((\mathbf{A}_n, \mathbf{B}_n))$ be an iid sequence of $d \times d$ matrices $\mathbf{A}_n$ with non-negative entries and d-dimensional non-negative-valued random vectors $\mathbf{B}_n \neq \mathbf{0}$ a.s. Assume that the following conditions hold:*

- *For some $\epsilon > 0$, $E\|\mathbf{A}_1\|^\epsilon < 1$.*

- $\mathbf{A}_1$ *has no zero rows a.s.*

- *The set*

$$\{\ln \|\mathbf{a}_n \cdots \mathbf{a}_1\| : \ n \geq 1, \ \mathbf{a}_n \cdots \mathbf{a}_1 > 0 \text{ and } \mathbf{a}_n, \ldots, \mathbf{a}_1 \in \text{the support of } P_{\mathbf{A}_1}\}$$

  generates a dense group in $\mathbb{R}$.

- *There exists a* $\kappa_0 > 0$ *such that*

$$E\left(\min_{i=1,\ldots,d} \sum_{j=1}^{d} A_{ij}\right)^{\kappa_0} \geq d^{\kappa_0/2} \tag{2.8}$$

  and

$$E\left(\|\mathbf{A}_1\|^{\kappa_0} \ln^+ \|\mathbf{A}_1\|\right) < \infty\,. \tag{2.9}$$

Then the following statements hold:

1. *There exists a unique solution* $\kappa_1 \in (0, \kappa_0]$ *to the equation*

$$0 = \lim_{n\to\infty} \frac{1}{n} \ln E\|\mathbf{A}_n \cdots \mathbf{A}_1\|^{\kappa_1}\,. \tag{2.10}$$

2. *There exists a unique strictly stationary causal solution* $(\mathbf{X}_n)$ *to the stochastic recurrence equation* (2.1).

3. *If* $E|\mathbf{B}_1|^{\kappa_1} < \infty$, *then* $\mathbf{X}_1$ *satisfies the following regular variation condition:*

$$\text{For all } \mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}\,, \quad \lim_{u\to\infty} u^{\kappa_1}\, P((\mathbf{x}, \mathbf{X}_1) > u) = w(\mathbf{x}) \quad \text{exists} \tag{2.11}$$

$$\text{and is positive for all non-negative-valued vectors } \mathbf{x} \neq 0.$$

Remark 2.5 In the case $d = 1$, the conditions of Kesten's theorem become particularly simple. Indeed, if A_1 is a non-negative-valued random variable with a non-lattice distribution on $[0, \infty)$, $E \ln A_1 < 0$, $1 \leq EA_1^{\kappa_0}$ and $EA_1^{\kappa_0} \ln^+ A_1 < \infty$, then the assumptions of the first part of the theorem are satisfied and (2.10) reduces to $EA_1^{\kappa_1} = 1$ which has a unique positive solution. If, in addition, $EB_1^{\kappa_1} < \infty$, then X_1 is regularly varying with index κ_1.

Clearly, (2.11) is a special case of (2.6), where the slowly varying function L is a positive constant. An appeal to Theorem 2.3 immediately gives the following result.

Corollary 2.6 *Under the assumptions of Theorem 2.4, the marginal distribution of the unique strictly stationary causal solution* $(\mathbf{X}_n)$ *of the stochastic recurrence equation* (2.1) *is regularly varying in the following sense. If the value* κ_1 *in* (2.10) *is not an even integer, then there exist a positive constant* c *and a random vector* $\mathbf{\Theta}$ *with values in the unit sphere* $\mathbb{S}^{d-1}$ *such that*

$$u^{\kappa_1} P(|\mathbf{X}_1| > t\,u\,,\ \mathbf{X}_1/|\mathbf{X}_1| \in \cdot\,) \stackrel{v}{\to} c\, t^{-\kappa_1}\, P(\mathbf{\Theta} \in \cdot\,)\,, \quad \text{as } u \to \infty\,.$$

From the latter result we conclude the following.

Corollary 2.7 *Under the conditions of Corollary 2.6 the finite-dimensional distributions of the stationary solution* $(\mathbf{X}_t)$ *of* (2.1) *are regularly varying with index* κ_1.

Proof. First note that we can write

$$(\mathbf{X}_1,\dots,\mathbf{X}_m) = (\mathbf{A}_1,\mathbf{A}_2\mathbf{A}_1,\dots,\mathbf{A}_m\cdots\mathbf{A}_1)\mathbf{X}_0 + \mathbf{R}_m\,,$$

where the components of $\mathbf{R}_m$ have lighter tails than the components of $\mathbf{X}_0$. The regular variation of the vector $(\mathbf{X}_1,\dots,\mathbf{X}_m)$ is assured by Proposition 5.1. □

2.3 The strong mixing condition

The Markov chain $(\mathbf{X}_n)$ satisfies a mixing condition under quite general conditions as for example provided in Meyn and Tweedie [28]. Recall that a Markov chain $(\mathbf{Y}_n)$ with state space $E\subset\mathbb{R}^d$ is said to be *μ-irreducible* for some measure μ on $(E,\mathcal{E})$ ($\mathcal{E}$ is the Borel σ-field on E), if

$$\sum_{n>0} p^n(\mathbf{y},C) > 0 \quad \text{for all } \mathbf{y}\in E, \text{ whenever } \mu(C)>0. \tag{2.12}$$

Here $p^n(\mathbf{y},C)$ denotes the *n-step transition probability* of moving from $\mathbf{y}$ to the set C in n-steps. If the function

$$E(g(\mathbf{Y}_n)\mid \mathbf{Y}_{n-1}=\mathbf{y})\,,\quad \mathbf{y}\in E\,, \tag{2.13}$$

is continuous for every bounded and continuous g on E, then the Markov chain is said to be a *Feller chain.* The Markov chain $(\mathbf{Y}_n)$ is said to be *geometrically ergodic* if there exists a $\rho\in(0,1)$ such that

$$\rho^{-n}\,\|p^n(\mathbf{y},\cdot)-\pi(\cdot)\|_{TV}\to 0\,,$$

where π denotes the *invariant measure* of the Markov chain and $\|\cdot\|_{TV}$ is the *total variation distance.* A particular consequence of geometric ergodicity is that the Markov chain is *strongly mixing with geometric rate*, i.e., if the Markov Chain is started with its stationary distribution π, then there exist constants $\widetilde{C}>0$ and $a\in(0,1)$ such that

$$\sup\nolimits_{f,g}|\operatorname{cov}(f(\mathbf{Y}_0),g(\mathbf{Y}_k))| \;=:\; \alpha_k \;\le\; \widetilde{C}\,a^k\,, \tag{2.14}$$

where the sup is taken over all measurable functions f and g with $|f|\le 1$ and $|g|\le 1$. This follows, for example, from Theorem 16.1.5 in Meyn and Tweedie [28]. The function α_k is called the *mixing rate function* of $(\mathbf{Y}_t)$ and for Markov processes, it is equal to

$$\begin{aligned}\alpha_k &= \sup_{f,g}|\operatorname{cov}(f(\dots,\mathbf{Y}_{-1},\mathbf{Y}_0),g(\mathbf{Y}_k,\mathbf{Y}_{k+1},\dots))|\\ &= \sup_{A\in\sigma(\mathbf{Y}_s\,,\,s\le 0)\,,\,B\in\sigma(\mathbf{Y}_s\,,\,s\ge k)}|P(A\cap B)-P(A)\,P(B)|\,,\end{aligned}$$

where the last equality follows from Doukhan [14], p.3.

Theorem 2.8 *For the SRE in* (2.1)*, suppose there exists an* $\epsilon > 0$ *such that* $E\|\mathbf{A}_1\|^\epsilon < 1$ *and* $E|\mathbf{B}_1|^\epsilon < \infty$*. If the Markov chain* $(\mathbf{X}_n)$ *is* μ*-irreducible, then it is geometrically ergodic and, hence, strongly mixing with geometric rate.*

Remark 2.9 The condition $E\|\mathbf{A}_1\|^\epsilon < 1$ for $\epsilon > 0$ in some neighborhood of zero is satisfied if $E\ln\|\mathbf{A}_1\| < 0$ and $E\|\mathbf{A}_1\|^\delta < \infty$ for some $\delta > 0$. Indeed, the function $h(v) = E\|\mathbf{A}_1\|^v$ then has derivative $h'(0) = E\ln\|\mathbf{A}_1\| < 0$, hence $h(v)$ decreases in a small neighborhood of zero, and since $h(0) = 1$ it follows that $h(\epsilon) < 1$ for small $\epsilon > 0$. On the other hand, $E\|\mathbf{A}_1\|^\epsilon < 1$ for some $\epsilon > 0$ implies that $E\ln\|\mathbf{A}_1\| < 0$ by an application of Jensen's inequality.

Proof. First note that by Theorem 2.1 and an application of Jensen's inequality, a unique stationary solution to the SRE exists. To show geometric ergodicity, we check the three conditions of Theorem 1 in Feigin and Tweedie [17]. The Lebesgue dominated convergence theorem ensures that (2.13) is continuous in $\mathbf{y}$ and hence the Markov chain is Feller. By assumption, the chain is μ-irreducible, so it remains to verify the drift condition, i.e., there exists a compact set K and a non-negative continuous function g such that $\mu(K) > 0$, $g(x) \geq 1$ on K, and for some $\delta > 0$, $E(g(\mathbf{X}_n) \mid \mathbf{X}_{n-1} = \mathbf{x}) \leq (1-\delta)g(\mathbf{x})$ for all $\mathbf{x} \in K^c$. For the SRE, choose

$$g(\mathbf{x}) = |\mathbf{x}|^\epsilon + 1\,, \quad \mathbf{x} \in \mathbb{R}^d\,,$$

where the ϵ is given in the assumptions. Notice that we may assume without loss of generality that $\epsilon \in (0, 1]$. Then

$$\begin{aligned} E(g(\mathbf{X}_n) \mid \mathbf{X}_{n-1} = \mathbf{x}) &\leq E|\mathbf{A}_1\mathbf{x}|^\epsilon + E|\mathbf{B}_1|^\epsilon + 1\,, \\ &\leq E\|\mathbf{A}_1\|^\epsilon|\mathbf{x}|^\epsilon + E|\mathbf{B}_1|^\epsilon + 1 \\ &=: E\|\mathbf{A}_1\|^\epsilon g(\mathbf{x}) + (E|\mathbf{B}_1|^\epsilon - E\|\mathbf{A}_1\|^\epsilon + 1)\,. \end{aligned}$$

Choose $K = [-M, M]^d$ and $M > 0$ so large that $\mu(K) > 0$ and

$$E(g(\mathbf{X}_n) \mid \mathbf{X}_{n-1} = \mathbf{x}) \leq (1-\delta)g(\mathbf{x})\,, \quad |\mathbf{x}| > M\,,$$

for some constant $1 - \delta > E\|\mathbf{A}_1\|^\epsilon$. This proves the drift condition and completes the argument. □

2.4 Point process theory

In this section we study the weak convergence of point processes generated by the stationary solution $(\mathbf{X}_t)$ to the SRE (2.1). The following result is the basis for dealing with the sample autocovariances, sample autocorrelations and extremes of this sequence. We need to consider a slightly more general sequence: for $m \geq 0$ define

$$\mathbf{X}_t(m) = \operatorname{vec}(\mathbf{X}_t, \ldots, \mathbf{X}_{t+m})\,.$$

The following results are based on work by Davis and Mikosch [13]. Since the theory developed there is quite technical we will omit details and refer to the paper.

Theorem 2.10 *Assume the conditions of Theorem 2.4 hold and that the solution to the SRE is μ-irreducible. Let (a_n) be a sequence of constants satisfying*

$$n\, P(|\mathbf{X}_1(m)| > a_n) \to 1\,. \tag{2.15}$$

Then

$$N_n = \sum_{t=1}^{n} \varepsilon_{\mathbf{X}_t(m)/a_n} \overset{d}{\to} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{P_i \mathbf{Q}_{ij}}\,,$$

where $\varepsilon_{\mathbf{x}}$ is the point measure concentrated at $\mathbf{x}$ and $\overset{d}{\to}$ denotes convergence in distribution of point measures on $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$. Here (P_i) are the points of a Poisson process on $(0,\infty)$ with intensity $\nu(dy) = \widetilde{\gamma}\kappa_1 y^{-\kappa_1 - 1}$ and $\widetilde{\gamma} > 0$ is the extremal index of the sequence $(|\mathbf{X}_t(m)|)$. The process (P_i) is independent of the iid point processes $\sum_{j=1}^{\infty} \varepsilon_{\mathbf{Q}_{ij}}$, $i \geq 1$, whose points satisfy $\sup_j |\mathbf{Q}_{ij}| = 1$ and whose distribution is described in [13].

Remark 2.11 Since Kesten's theorem implies that $P(|\mathbf{X}_1(m)| > x)| \sim c x^{-\kappa_1}$ as $x \to \infty$ for some constant $c > 0$, we have $a_n \sim (cn)^{1/\kappa_1}$.

Remark 2.12 In the above point process result the points $(P_i, \mathbf{Q}_{ij})$ correspond to the radial and spherical parts of the limiting points $\mathbf{X}_t(m)/a_n$, respectively. In this sense, the $\mathbf{Q}_{ij}$ describe the cluster behavior in the limit point process.

Proof. The proof follows from the results in Section 2 of Davis and Mikosch [13], in particular their Theorem 2.8, for general strictly stationary sequences of random vectors. Three assumptions have to be verified. The first condition is that the finite-dimensional distributions of $(\mathbf{X}_t(m))$ are regularly varying with index κ_1. This follows from Corollary 2.7.

The second assumption is a mild mixing condition on $(\mathbf{X}_t(m))$ (the assumption $\mathcal{A}(a_n)$) which is implied by strong mixing. However, Proposition 2.8 implies that $(\mathbf{X}_t)$, and hence $(\mathbf{X}_t(m))$, are strongly mixing with a geometrically decreasing rate function.

The third condition to be verified is the following (notice that by construction of $\mathbf{X}_t(m)$ it suffices to consider the case $m = 0$):

$$\lim_{k\to\infty} \limsup_{n\to\infty} P\left(\bigvee_{k \leq |t| \leq r_n} |\mathbf{X}_t| > a_n y \,\Big|\, |\mathbf{X}_0| > a_n y \right) = 0\,, \quad y > 0\,, \tag{2.16}$$

where $r_n, m_n \to \infty$, are two integer sequences such that $n\alpha_{m_n}/r_n \to 0$, $r_n m_n/n \to 0$. Since the mixing rate function α_n (see (2.14)) decreases at a geometric rate, i.e., $\alpha_n \leq \mathit{const}\ a^n$ for some

$a \in (0,1)$, one can choose $r_n = [n^{\epsilon}]$ and $m_n = [n^{\delta}]$ for any $0 < \delta < \epsilon < 1$; see the discussion of mixing conditions in Leadbetter and Rootzén [27], Lemma 2.4.4.

Iterating (2.1), we obtain for $t > 0$,

$$\mathbf{X}_t = \prod_{j=1}^{t} \mathbf{A}_j \, \mathbf{X}_0 + \sum_{j=1}^{t} \prod_{m=j+1}^{t} \mathbf{A}_m \, \mathbf{B}_j =: \mathbf{I}_{t,1}\mathbf{X}_0 + \mathbf{I}_{t,2} \, .$$

and hence

$$\begin{aligned} P(|\mathbf{X}_t| > a_n y \,\big|\, |\mathbf{X}_0| > a_n \, y) \qquad (2.17)\\ \leq \quad P(|\mathbf{X}_0| \, \|\mathbf{I}_{t,1}\| > a_n \, y/2 \,\big|\, |\mathbf{X}_0| > a_n \, y) + P(|\mathbf{I}_{t,2}| > a_n \, y/2) \, . \end{aligned}$$

Choose $\epsilon > 0$. Then, using Markov's inequality and Karamata's theorem (see Bingham et al.[5]), the limes superior of the first term on the right of (2.17) is bounded above by

$$\limsup_{n\to\infty} E\|\mathbf{I}_{t,1}\|^{\epsilon} (2/y)^{\epsilon} \frac{E[|\mathbf{X}_0|^{\epsilon} I_{(a_n \, y,\infty)}(|\mathbf{X}_0|)]}{a_n^{\epsilon} \, P(|\mathbf{X}_0| > a_n \, y)} \leq C \, (E\|\mathbf{A}_1\|^{\epsilon})^t \, .$$

Here C is a constant independent of t. Now choose $\epsilon \in (0,1)$ such that $E\|\mathbf{A}_1\|^{\epsilon} < 1$. This is always possible in view of Remark 2.9. As for the second term in (2.17), we have

$$|\mathbf{I}_{t,2}| \stackrel{d}{=} \left| \sum_{j=1}^{t} \prod_{m=1}^{j-1} \mathbf{A}_m \, \mathbf{B}_j \right| \leq \sum_{j=1}^{t} \prod_{m=1}^{j-1} \|\mathbf{A}_m\| \, |\mathbf{B}_j| \uparrow \sum_{j=1}^{\infty} \prod_{m=1}^{j-1} \|\mathbf{A}_m\| \, |\mathbf{B}_j| = Y \quad \text{a.s.}$$

for some random variable Y. Thus we obtain by Markov's inequality and the same $\epsilon \leq 1$,

$$P(|\mathbf{I}_{t,2}| > a_n \, y/2) \leq P(Y > a_n y/2) \leq a_n^{-\epsilon} (2/y)^{\epsilon} E|\mathbf{B}_1|^{\epsilon} \sum_{j=1}^{\infty} (E\|\mathbf{A}_1\|^{\epsilon})^j \leq const \; a_n^{-\epsilon} \, .$$

According to the above remark, we can take $r_n \sim n^{\delta}$ for any small $\delta > 0$. Choosing δ so small that $r_n a_n^{-\epsilon} \to 0$ and combining the bounds for the terms in (2.17), we obtain,

$$\begin{aligned} & \lim_{k\to\infty} \limsup_{n\to\infty} P\left(\bigvee_{k \leq |t| \leq r_n} |\mathbf{X}_t| > a_n y \,\Big|\, |\mathbf{X}_0| > a_n y \right) \\ \leq & \quad \lim_{k\to\infty} \limsup_{n\to\infty} \sum_{k \leq |t| \leq r_n} P(|\mathbf{X}_t| > a_n y \,\big|\, |\mathbf{X}_0| > a_n \, y) \\ \leq & \quad \lim_{k\to\infty} (const) \, \sum_{t=k}^{\infty} (E\|\mathbf{A}_1\|^{\epsilon})^t \\ = & \quad 0 \, . \end{aligned}$$

This completes the verification of (2.16). □

2.5 Limit theory for the sample autocovariances and autocorrelations

Using the point process theory of the previous section, it is possible to derive the asymptotic behaviour of the sample cross-covariances and cross-correlations of the stationary solution $(\mathbf{X}_t)$ to the SRE (2.1) satisfying the conditions of Theorem 2.4. For ease of exposition we concentrate on the sample autocovariances of the first component process (Y_t) say of $(\mathbf{X}_t)$. Define the *sample autocovariance function*

$$\gamma_{n,Y}(h) = n^{-1} \sum_{t=1}^{n-h} Y_t Y_{t+h}\,, \quad h \geq 0\,, \tag{2.18}$$

and the corresponding *sample autocorrelation function*

$$\rho_{n,Y}(h) = \gamma_{n,Y}(h)/\gamma_{n,Y}(0)\,, \quad h \geq 1\,. \tag{2.19}$$

We also write

$$\gamma_Y(h) = EY_0Y_h \quad \text{and} \quad \rho(h) = \gamma_Y(h)/\gamma_Y(0)\,, \quad h \geq 0\,,$$

for the autocovariances and autocorrelations, respectively, of the sequence (Y_t) if these quantities exist. Mean-corrected versions of both the sample and model ACVF can also be considered—the same arguments as above show that the limit theory does not change.

The following result is an immediate consequence of Theorem 2.10 and the theory developed in Davis and Mikosch [13], in particular their Theorem 3.5. In what follows, the notion of infinite variance stable random vector is used. We refer to Samorodnitsky and Taqqu [34] for its definition and an encyclopaedic treatment of stable processes.

Theorem 2.13 *Assume that $(\mathbf{X}_t)$ is a solution to (2.1) satisfying the conditions of Theorem 2.4.*

(1) If $\kappa_1 \in (0,2)$, then

$$\left(n^{1-2/\kappa_1}\gamma_{n,Y}(h)\right)_{h=0,\dots,m} \stackrel{d}{\rightarrow} (V_h)_{h=0,\dots,m}\,,$$

$$(\rho_{n,Y}(h))_{h=1,\dots,m} \stackrel{d}{\rightarrow} (V_h/V_0)_{h=1,\dots,m}\,,$$

where the vector $(V_0,\dots,V_m)$ is jointly $\kappa_1/2$-stable in $\mathbb{R}^{m+1}$.

(2) If $\kappa_1 \in (2,4)$ and for $h = 0,\dots,m$,

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty} \operatorname{var}\left(n^{-2/\kappa_1}\sum_{t=1}^{n-h} Y_tY_{t+h} I_{\{|Y_tY_{t+h}|\leq a_n^2\epsilon\}}\right) = 0\,, \tag{2.20}$$

then

$$\left(n^{1-2/\kappa_1}(\gamma_{n,Y}(h) - \gamma_Y(h))\right)_{h=0,\dots,m} \stackrel{d}{\rightarrow} (V_h)_{h=0,\dots,m}\,, \tag{2.21}$$

$$\left(n^{1-2/\kappa_1}(\rho_{n,X}(h) - \rho_X(h))\right)_{h=1,\dots,m} \stackrel{d}{\rightarrow} \gamma_X^{-1}(0)\,(V_h - \rho_X(h)\,V_0)_{h=1,\dots,m}\,, \tag{2.22}$$

where $(V_0,\dots,V_m)$ is jointly $\kappa_1/2$-stable in $\mathbb{R}^{m+1}$.

(3) If $\kappa_1 > 4$ then (2.21) and (2.22) hold with normalization $n^{1/2}$, where $(V_1, \ldots, V_m)$ is multivariate normal with mean zero and covariance matrix $[\sum_{k=-\infty}^{\infty} \operatorname{cov}(Y_0 Y_i, Y_k Y_{k+j})]_{i,j=1,\ldots,m}$ and $V_0 = E(Y_0^2)$.

Remark 2.14 The limit random vectors in parts (1) and (2) of the theorem can be expressed in terms of the P_i's and $\mathbf{Q}_{ij}$'s defined in Theorem 2.10. For more details, see Davis and Mikosch [13], Theorem 3.5, where the proof of (1) and (2) is provided. Part (3) follows from a standard central limit theorem for strongly mixing sequences; see for example Doukhan [14].

Remark 2.15 The conclusions of Theorem 2.13 are also valid for other functions of $\mathbf{X}_t$ including linear combinations of powers of the components. Indeed, the constructed process inherits strong mixing as well as joint regular variation from the $(\mathbf{X}_t)$ process and point process convergence follows from the continuous mapping theorem.

3 Application to GARCH processes

3.1 Definition of GARCH process

One of the major applications of SRE's is to the class of GARCH processes. A *generalized autoregressive conditionally heteroscedastic process* (X_t) of order (p, q) with $p, q \geq 0$ (GARCH(p, q)) is given by the equations

$$
\begin{aligned}
X_t &= \sigma_t \, Z_t \,, \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \,,
\end{aligned}
$$

where (Z_t) is an iid sequence of random variables, and the α_i's and β_j's are non-negative constants with the convention that $\alpha_p > 0$ if $p \geq 1$ and $\beta_q > 0$ if $q \geq 1$. This class of processes was introduced by Bollerslev [6] and Taylor [35] and has since found a multitude of applications for modeling financial time series. For $q = 0$ the process is called an ARCH(p) process.

3.2 Embedding in a stochastic recurrence equation

The squared processes (X_t^2) and (σ_t^2) satisfy the following SRE:

(3.1) $$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t \,,$$

where

$$\mathbf{X}_t = \left(\sigma_{t+1}^2, \ldots, \sigma_{t-q+2}^2, X_t^2, \ldots, X_{t-p+2}^2\right)' \,,$$

$$
\mathbf{A}_t = \begin{pmatrix}
\alpha_1 Z_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_p \\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\
Z_t^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix}, \tag{3.2}
$$

$$
\mathbf{B}_t = (\alpha_0, 0, \ldots, 0)'.
$$

3.3 Basic properties of a GARCH process

In the following proposition we collect some of the basic properties of the process $(\mathbf{X}_t)$. Some of them are well known, in particular parts (A) and (C), see Remarks 3.2 and 3.3.

Theorem 3.1 *Consider the SRE* (3.1). *Assume that $\alpha_0 > 0$ and the Lyapunov exponent γ of this stochastic recurrence equation is negative.*

(*A*) (Existence of stationary solution)

Assume that the following condition holds:

$$
E \ln^+ |Z_1| < \infty.
$$

Then there exists a unique strictly stationary causal solution of the SRE (3.1).

(B) (Regular variation of the finite-dimensional distributions)

Assume the following conditions:

1. *Z_1 has a positive density on $\mathbb{R}$ such that $E|Z_1|^h < \infty$ for all $h < h_0$ and $E|Z_1|^{h_0} = \infty$ for some $h_0 \in (0, \infty]$.*

2. *Not all of the parameters α_i and β_i vanish.*

Then there exists a $\kappa_1 > 0$ and a finite-valued function $w(\mathbf{x})$ such that

$$
\text{For all } \mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}, \qquad \lim_{u\to\infty} u^{\kappa_1} P((\mathbf{x}, \mathbf{X}_1) > u) = w(\mathbf{x}) \quad \text{exists},
$$

i.e., $(\mathbf{x}, \mathbf{X}_1)$ is regularly varying with index κ_1. Moreover, if $\mathbf{x} \in [0,\infty)^d \setminus \{\mathbf{0}\}$ with $d = p+q$, then $w(\mathbf{x}) > 0$. Furthermore, if κ_1 is not even, then $\mathbf{X}_1$ is regularly varying with index κ_1, i.e., there *exists a $\mathbb{S}^{d-1}$-valued random vector $\boldsymbol{\Theta}$ such that*

$$\frac{P(|\mathbf{X}_1| > t\,x\,,\ \mathbf{X}_1/|\mathbf{X}_1| \in \cdot\,)}{P(|\mathbf{X}_1| > x)} \stackrel{v}{\to} t^{-\kappa_1}\, P(\boldsymbol{\Theta} \in \cdot\,)\,, \quad \text{as } x \to \infty\,,$$

(C) If Z_1 has a density positive in an interval containing zero, then $(\mathbf{X}_t)$ is strongly mixing with geometric rate.

Remark 3.2 Necessary and sufficient conditions for the Lyapunov exponent $\gamma < 0$ in terms of the parameters α_i and β_i and the distribution of Z_1 are known only in a few cases. This includes the ARCH(1) (see Goldie [19]; cf. Embrechts et al. [15], Section 8.4) and the GARCH(1,1) cases (see Nelson [30]). The latter case can be reduced to a one-dimensional SRE for (σ_t^2); see for example Mikosch and Stărică [29]. The general case can be found in Bougerol and Picard [8], where to the best of our knowledge the most general sufficient conditions are given. Some of their results are formulated below subject to the assumptions $\alpha_0 > 0$, $EZ_1 = 0$ and $EZ_1^2 = 1$.

- $\gamma < 0$ is necessary and sufficient for the existence of a unique strictly stationary causal solution to (3.1).
- $\sum_{i=1}^q \beta_i < 1$ is necessary for $\gamma < 0$.
- $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ implies $\gamma < 0$.
- If Z_1 has infinite support and no atom at zero, $\alpha_i > 0$ and $\beta_j > 0$ for all i and j then $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1$ implies $\gamma < 0$.

Since it is in general not possible to calculate γ explicitly, a potential method to verify whether or not $\gamma < 0$ is via Monte-Carlo simulation using relation (2.3).

Remark 3.3 Part (C) of the theorem is due to Boussama [9], Chapter 3.

Remark 3.4 As for checking whether $\gamma < 0$, it is in general difficult to determine the index κ_1 of regular variation by direct calculation. Again, the ARCH(1) and GARCH(1,1) processes are the two exceptions where κ_1 can be calculated by the method described in Remark 2.5; κ_1 is the unique solution to $E(\alpha_1 Z_1^2)^{\kappa_1} = 1$ in the first case and to $E(\alpha_1 Z_1^2 + \beta_1)^{\kappa_1} = 1$ in the second case where we again assume that $EZ_1^2 = 1$ and $EZ_1 = 0$. In either case, κ_1 can be solved by Monte-Carlo simulation if the distribution of Z_1 is known. A table of values κ_1 as a function of α_1 for the ARCH(1) case with standard normal Z_1 can be found in Embrechts et al. [15], Section 8.4. The theory for the GARCH(1,1) case is dealt with in Mikosch and Stărică [29]. Note that in the IGARCH(1,1) case, i.e., $\alpha_1 + \beta_1 = 1$, $\kappa_1 = 1$ is the unique solution to the above equation. Hence $P(X_t^2 > x) \sim c_1 x^{-1}$ and $P(\sigma_t^2 > x) \sim c_2 x^{-1}$ as $x \to \infty$ for some positive constants c_1 and c_2.

Proof. Part (A). This is an immediate consequence of Theorem 2.1.

Part (B). It is easy to see that (B) implies (A) and hence a unique strictly stationary solution to (3.1) exists. To establish (B) we consider a subsequence $\widetilde{\mathbf{X}}_t = \mathbf{X}_{tm}$ of $(\mathbf{X}_t)$ for some integer m. Along this subsequence the underlying SRE can be written as

$$\mathbf{X}_{tm} = \mathbf{A}_{tm} \cdots \mathbf{A}_{t(m-1)+1}\mathbf{X}_{t(m-1)} + \mathbf{B}_t + \sum_{k=1}^{m-1} \mathbf{A}_{tm} \cdots \mathbf{A}_{tm-k+1}\mathbf{B}_{tm-k} = \widetilde{\mathbf{A}}_t\mathbf{X}_{t(m-1)} + \widetilde{\mathbf{B}}_t \,,$$

where $((\widetilde{\mathbf{A}}_t, \widetilde{\mathbf{B}}_t))$ is an iid sequence. Hence $(\widetilde{\mathbf{X}}_t)$ satisfies the SRE

$$\widetilde{\mathbf{X}}_t = \widetilde{\mathbf{A}}_t\widetilde{\mathbf{X}}_{t-1} + \widetilde{\mathbf{B}}_t \,, \quad t \in \mathbb{Z} \,. \tag{3.3}$$

We will apply Kesten's Theorem 2.4 to this SRE for m sufficiently large. By stationarity the regular variation property then follows for the distribution of $\mathbf{X}_t$. From the moment condition on Z_1, (2.2), (2.3), and Remark 2.9, there exists an $\epsilon > 0$ small such that $E\|\widetilde{\mathbf{A}}_1\|^\epsilon < 1$ and $E|\widetilde{\mathbf{B}}_1|^\epsilon < \infty$ by choosing m sufficiently large. Observe that the entries of $\widetilde{\mathbf{A}}_1$ are multilinear forms of the Z_t^2's. Moreover, $E|Z_1|^h$ becomes arbitrarily large for $h < h_0$ chosen sufficiently large. Hence (2.8) is satisfied for $\widetilde{\mathbf{A}}_1$ when κ_0 is sufficiently large, and so is (2.9) since $h < h_0$. We next show that the set of real numbers $\ln\|\widetilde{\mathbf{a}}_1 \cdots \widetilde{\mathbf{a}}_n\|$, where the $\widetilde{\mathbf{a}}_i$'s are from the support of $\widetilde{\mathbf{A}}_1$, generates a dense group in $\mathbb{R}$. To see this we first observe that $\widetilde{\mathbf{A}}_1$ has positive entries for m chosen sufficiently large. This follows from the fact that, for large m, those entries are multilinear forms of the Z_t^2's which have a density on $(0, \infty)$. Since multilinear forms are continuous functions of the Z_t^2's, the support of $\widetilde{\mathbf{A}}_1$ is a connected set, and so is the support of $\|\widetilde{\mathbf{A}}_1\|$, as a continuous function of the matrix $\widetilde{\mathbf{A}}_1$. Hence the support of $\ln\|\widetilde{\mathbf{A}}_1\|$ contains an interval for m sufficiently large, which yields the desired property of the numbers $\ln\|\widetilde{\mathbf{a}}_1 \cdots \widetilde{\mathbf{a}}_n\|$.

An application of Kesten's theorem finally yields the regular variation of $(\mathbf{x}, \mathbf{X}_1)$ with index $\kappa_1 > 0$, and provided κ_1 is not an even integer, Corollary 2.6 gives the regular variation of $\mathbf{X}_1$.

Part (C). The strong mixing property with geometric rate was proved by Boussama [9].

This concludes the proof of the theorem. □

The properties of the sequence (X_t^2, σ_t^2) (such as stationarity and regular variation) immediately translate into the corresponding properties for GARCH processes. This is the content of the following result.

Corollary 3.5 *Consider the SRE* (3.1). *Assume that* $\alpha_0 > 0$, *the Lyapunov exponent* $\gamma < 0$ *and the conditions of parts (B) and (C) of Theorem 3.1 hold. Then the following statements hold.*

(A) A stationary version of the process $(\mathbf{U}_t) = ((X_t, \sigma_t))$ *exists.*

(B) There exists $\kappa > 0$ *such that the limits*

$$\lim_{u\to\infty} u^\kappa P(X_1 > u) \quad \textit{and} \quad \lim_{u\to\infty} u^\kappa P(\sigma_1 > u)$$

exist and are positive. Moreover, if $\kappa/2$ is not an even integer, then the finite-dimensional distributions of the process $(\mathbf{U}_t)$ are regularly varying with index κ.
(C) If (Z_t) is iid symmetric, then the sequence $(\mathbf{U}_t)$ is strongly mixing with geometric rate.

Proof. (A) From Theorem 3.1 a strictly stationary version of the process $(\mathbf{X}_t)$ exists, hence of $\mathbf{V}_t = (|X_t|, \sigma_t)$ and $\mathbf{U}_t = (X_t, \sigma_t)$.
(B) From part (B) of Theorem 3.1 we know that $(\mathbf{x}, \mathbf{X}_1)$ is regularly varying with index $\kappa_1 > 0$. Hence σ_1 is regularly varying with index $\kappa = 2\kappa_1$, and so is $X_1 = \sigma_1 Z_1$ by an application of (5.2) in the Appendix.

If κ_1 is not an even integer then we also know that $\mathbf{X}_1$ is regularly varying with index κ_1, and so are the finite-dimensional distributions of $(\mathbf{X}_t)$, by Corollary 2.7. It is an easy exercise to conclude that the finite-dimensional distributions of $(\mathbf{V}_t)$ are regularly varying with index $\kappa = 2\kappa_1$. It suffices to show that for all $k \geq 1$, the random vector $\mathbf{Y}_k = (\sigma_1, X_1, \ldots, \sigma_k, X_k)'$ is regularly varying with index κ. This is proved by induction on k. For $k = 1$, $(\sigma_1, X_1)' = \sigma_1(Z_1, 1)'$ and since σ_1 is regularly varying, so is the vector by Corollary 5.2. Now suppose $\mathbf{Y}_k$ is regularly varying with $k \geq \max(p, q)$. Using the representation $\sigma_{k+1}^2 = \alpha_0 + \alpha_1 X_k^2 + \cdots + \alpha_p X_{k+1-p}^2 + \beta_1 \sigma_k^2 + \cdots + \beta_q \sigma_{k+1-q}^2$, it follows that $(\mathbf{Y}_k', \sigma_{k+1})'$ is regularly varying with exponent κ. Writing

$$\mathbf{Y}_{k+1} = \begin{pmatrix} I_{2k+1} & 0 \\ 0 & Z_{k+1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_k \\ \sigma_{k+1} \end{pmatrix},$$

we conclude once again from Corollary 5.2 that $\mathbf{Y}_{k+1}$ is regularly varying with exponent κ which completes the induction argument.
(C) Theorem 3.1 (C) tells us that $(\mathbf{X}_t)$ is strongly mixing with geometric rate. The mixing of $(\mathbf{X}_t)$ implies that the process $\mathbf{V}_t = (|X_t|, \sigma_t)$ is also strongly mixing with rate function (α_k), say. Fix two Borel sets in $\mathcal{B}(\mathbb{R}^\infty)$. Using the independence of the $(\mathbf{U}_t) = ((X_t, \sigma_t))$ process conditional on $(\mathbf{V}_t)$, we have

$$|P((\ldots, \mathbf{U}_{-1}, \mathbf{U}_0) \in A, (\mathbf{U}_k, \mathbf{U}_{k+1}, \ldots) \in B) - P((\ldots, \mathbf{U}_{-1}, \mathbf{U}_0) \in A) P((\mathbf{U}_k, \mathbf{U}_{k+1}, \ldots) \in B)|$$

$$= |E[f(\ldots, \mathbf{V}_{-1}, \mathbf{V}_0) g(\mathbf{V}_k, \mathbf{V}_{k+1}, \ldots)] - E[f(\ldots, \mathbf{V}_{-1}, \mathbf{V}_0)] E[g(\mathbf{V}_k, \mathbf{V}_{k+1}, \ldots)]|,$$

(3.4)

where

$$\begin{aligned} f(\ldots, \mathbf{V}_{-1}, \mathbf{V}_0) &= P\left((\ldots, \mathbf{U}_{-1}, \mathbf{U}_0) \in A \mid \mathbf{V}_s,\, s \leq 0\right), \\ g(\mathbf{V}_k, \mathbf{V}_{k+1}, \ldots) &= P\left((\mathbf{U}_k, \mathbf{U}_{k+1}, \ldots) \in A \mid \mathbf{V}_s,\, s \geq k\right). \end{aligned}$$

Now applying a standard result on functions of mixing sequences (see for example Doukhan [14]), one can show that (3.4) is bounded by $4\alpha_k$. □

3.4 The sample autocovariances and sample autocorrelations of GARCH processes

In what follows the theory of Section 2.5 is applied to derive the distributional limits of the sample autocovariances and autocorrelations of GARCH(p,q) processes. The case of an ARCH(1) process, its absolute values and squares has been treated in Davis and Mikosch [13]. The case of a GARCH$(1,1)$ was dealt with in Mikosch and Stărică [29]. Below we derive the limit theory for the sample autocovariance function (ACVF) and sample autocorrelation function (ACF) of a general GARCH(p,q) process.

Theorem 3.6 *Under the conditions of Corollary 3.5 the sample ACF and the sample ACVF of the GARCH(p,q) process (X_t) with iid symmetric noise (Z_t) have the limit distributions as described in Theorem 2.13.*

Proof. We apply Theorem 3.5 of Davis and Mikosch [13]. By Corrolary 3.5, the process (X_t) is strictly stationary with regularly varying finite-dimensional distributions and is strongly mixing with geometric rate. Moreover, using the same argument given for the proof of Theorem 2.10, condition (2.16) is easily checked and hence convergence of the associated sequence of point processes in Theorem 3.5 follows.

The case $\kappa \in (0,2)$. This is a direct application of Theorem 3.5 of [13].

The case $\kappa \in (2,4)$. For $h \geq 1$, condition (3.4) of Theorem 3.5 of [13] is easy to verify since, by symmetry of the Z_t's, the random variables $X_t X_{t+h} I_{\{|X_t X_{t+h}| \leq a_n^2 \epsilon\}}$ are uncorrelated. For the case $h=0$, this condition is more difficult to verify directly and so we adopt a different approach. We have

$$\begin{aligned} & a_n^{-2} \sum_{t=1}^n (X_t^2 - EX_1^2) \\ = \; & a_n^{-2} \sum_{t=1}^n \sigma_t^2 (Z_t^2 - 1) + a_n^{-2} \sum_{t=1}^n (\sigma_t^2 - E\sigma_1^2) \\ = \; & a_n^{-2} \sum_{t=1}^n \sigma_t^2 (Z_t^2 - 1) I_{\{\sigma_t \leq \epsilon a_n\}} + a_n^{-2} \sum_{t=1}^n \sigma_t^2 (Z_t^2 - 1) I_{\{\sigma_t > \epsilon a_n\}} + n a_n^{-2} (\gamma_{n,\sigma}(0) - E\sigma_1^2) \\ = \; & I + II + III\,. \end{aligned}$$

Using Karamata's theorem on regular variation, it follows that

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \operatorname{var}(I) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} n a_n^{-2} \operatorname{var}(Z_1^2) E\sigma_t^2 I_{\{\sigma_t \leq \epsilon a_n\}} = 0\,. \tag{3.5}$$

As for the third term, we have

$$III \;=\; a_n^{-2} \sum_{t=1}^n \left(\sum_{j=1}^p \alpha_j (X_{t-j}^2 - E\sigma_1^2) + \sum_{j=1}^q \beta_j (\sigma_{t-j}^2 - E\sigma_1^2) \right)$$

$$= \; n\, a_n^{-2} \, \left[(\alpha_1+\cdots+\alpha_p)(\gamma_{n,X}(0)-EX_1^2)+(\beta_1+\cdots+\beta_q)\,(\gamma_{n,\sigma}(0)-E\sigma_1^2)\right]+o_P(1)$$

$$= \; (\alpha_1+\cdots+\alpha_p)\, n\, a_n^{-2}\,(\gamma_{n,X}(0)-EX_1^2)+(\beta_1+\cdots+\beta_q)\, III + o_P(1)\,.$$

Hence

$$III \;=\; \frac{\alpha_1+\cdots+\alpha_p}{1-(\beta_1+\cdots+\beta_q)}\, n\, a_n^{-2}\,(\gamma_{n,X}(0)-EX_1^2)\;+o_P(1)\,,$$

where we use that fact that $1-(\beta_1+\cdots+\beta_q)>0$ is a necessary condition for stationarity; see Remark 3.2. So we may conclude that

$$n\, a_n^{-2}\,(\gamma_{n,X}(0)-EX_1^2) \;=\; \frac{1-(\beta_1+\cdots+\beta_q)}{1-(\alpha_1+\cdots+\alpha_p)-(\beta_1+\cdots+\beta_q)}\,(I+II)+o_P(1)\,.$$

Here we use the fact that $1-(\alpha_1+\cdots+\alpha_p)-(\beta_1+\cdots+\beta_q)>0$ is a necessary condition for the existence of the second moment of X_t; see Bollerslev [6]. Using the latter relation and (3.5), the point process convergence methods of Davis and Mikosch [13], Section 4, and a continuous mapping argument, it can be shown that $n\, a_n^{-2}\,(\gamma_{n,X}(0)-EX_1^2)$ converges in distribution to a $\kappa/2$-stable random variable. Moreover, since the convergence for the sample ACVF at lags $h\ge 1$ is based on the same point process result, one has joint convergence to a $\kappa/2$-stable limit for any finite vector of sample autocovariances. This fact together with the continuous mapping theorem implies that the conclusion of part 2 of Theorem 2.13 holds for both the sample ACF and ACVF of (X_t).

The case $\kappa\in(4,\infty)$. This follows from a standard central limit theorem for strongly mixing as can be found in [14]. □

In the analysis of financial returns it is common practice to study the autocorrelations of the absolute values and their powers in order to detect the non-linearity in the dependence structure. The sample ACVF and sample ACF of the absolute values and any powers of the process can be treated in a similar way by applying the same kind of argument; see for example Davis and Mikosch [13] for the ARCH(1) case and Mikosch and Stărică [29] for the GARCH(1, 1) case.

4 Some final remarks

The results of Sections 2 and 3 show the power of the theory for solutions to stochastic recurrence equations when applied to a GARCH(p,q) process. Results on the existence of a stationary version of a GARCH process and properties of the distributional tails follow from this general theory for SREs. The verification of the required conditions, however, is in general quite involved for GARCH processes. Explicit formulae, in terms of the parameters and noise distribution of the GARCH model, for determination of the Lyapunov exponent γ and the index of regular variation κ_1 are obscure. Monte-Carlo techniques can be used to determine γ and statistical methods for tail estimation may be implemented to estimate κ_1.

The results on the regular variation of the finite-dimensional distributions of GARCH processes are quite surprising, especially in the case when the noise sequence (Z_t) has light tails. For example, if the Z_t's have a normal distribution, then the resulting GARCH process has power law tails. Many real-life log-returns such as long daily log-return series of foreign exchange rates can often be well modeled by a GARCH(1,1) or IGARCH(1,1) model in which the sum of the estimated ARCH and GARCH parameters is close to or equal to 1. In such cases the index of regular variation is close to or equal to 2 and hence inference procedures based on the sample autocorrelations in the time domain and the periodogram in the frequency domain have to be treated with enormous care. Analysis for log-returns with infinite 5th, 4th, 3rd, etc. moments require a large sample theory which is determined by the very large values in the sample and leads to 95% confidence bands much wider than the classical $\pm 1.96/\sqrt{n}$ bands for the sample ACF and to unusual limit distributions. Our results for the sample ACF document that wide confidence bands and slow rates of convergence are typical for data which are modeled by GARCH processes with infinite 4th moment. These results have to be understood as qualitative ones. The limiting distributions for the sample ACF are defined via point processes and functions of multivariate stable random vectors; to date, little is known about the properties of these distributions and one must resort to simulation for exploring the sampling behavior of these statistics.

5 Appendix

5.1 Regularly varying vectors under random affine mappings

In what follows we consider a regularly varying random vector $\mathbf{X}$ with index $\alpha \geq 0$ and spectral measure $P_{\boldsymbol{\Theta}}$. For convenience we will work here with the following characterization of a regularly varying vector $\mathbf{X}$ which is equivalent to (2.5): There exist a measure μ on $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ and a sequence (a_n) of non-negative numbers such that

$$n\,P(a_n^{-1}\mathbf{X} \in \cdot) \stackrel{v}{\to} \mu(\cdot) \tag{5.1}$$

Our first result extends a well-known one-dimensional lemma of Breiman [11] to $d > 1$. It says that, for any independent non-negative random variables ξ and η such that η is regularly varying with index α and $E\xi^\gamma < \infty$ for some $\gamma > \alpha$,

$$P(\xi\,\eta > x) \sim E\xi^\alpha P(\eta > x)\,. \tag{5.2}$$

The multivariate version of Breiman's lemma reads as follows.

Proposition 5.1 *Assume the random vector* $\mathbf{X}$ *is regularly varying in the sense of* (5.1) *and* $\mathbf{A}$ *is a random* $q \times d$ *matrix, independent of* $\mathbf{X}$. *If* $0 < E\|\mathbf{A}\|^\gamma < \infty$ *for some* $\gamma > \alpha$, *then*

$$n\,P\left(a_n^{-1}\mathbf{A}\mathbf{X} \in \cdot\right) \stackrel{v}{\to} \widetilde{\mu}(\cdot) := E[\mu \circ \mathbf{A}^{-1}(\cdot)]\,,$$

where $\stackrel{v}{\to}$ *denotes vague convergence on* $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$.

Proof. For a fixed bounded $\widetilde{\mu}$-continuity set B define

$$A_n(B) = \left\{a_n^{-1}\mathbf{A}\mathbf{X} \in B\right\}\,.$$

Then for every $0 < \varepsilon < M < \infty$,

$$\begin{aligned} & P(A_n(B)) \\ = \quad & P(A_n(B) \cap \{\|\mathbf{A}\| \le \varepsilon\}) + P(A_n(B) \cap \{\varepsilon < \|\mathbf{A}\| \le M\}) + P(A_n(B) \cap \{\|\mathbf{A}\| > M\}) \\ =: \quad & p_1 + p_2 + p_3\,. \end{aligned}$$

Note that by (5.2) for some $t > 0$ (one can choose t to be the distance of the set B from 0),

$$\limsup_{n\to\infty} \frac{p_3}{P(|\mathbf{X}| > a_n)} \le \lim_{n\to\infty} \frac{P\left(|\mathbf{X}|\,\|\mathbf{A}\|I_{(M,\infty)}(\|\mathbf{A}\|) > a_n t\right)}{P\left(|\mathbf{X}| > a_n\right)} = t^{-\alpha} E[\|\mathbf{A}\|^\alpha I_{(M,\infty)}(\|\mathbf{A}\|)]\,.$$

Since $E\|\mathbf{A}\|^\alpha < \infty$ we conclude by Lebesgue dominated convergence that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{p_3}{P(|\mathbf{X}| > a_n)} = 0\,. \tag{5.3}$$

Now consider p_2:

$$\begin{aligned} & \lim_{n\to\infty} \frac{p_2}{P(|\mathbf{X}| > a_n)} \\ = \quad & \lim_{n\to\infty} \int_{\varepsilon < \|\mathbf{A}\| \le M} \frac{P\left(A_n(B) \mid \mathbf{A}\right)}{P\left(|\mathbf{X}| > a_n\right)} P(d\mathbf{A}) = E\left[I_{(\varepsilon,M]}(\|\mathbf{A}\|)\mu(\mathbf{A}^{-1}B)\right]\,. \end{aligned} \tag{5.4}$$

In the limit relation we made use of a Pratt's lemma; cf. Pratt [31]. The right-hand side of (5.4) converges to the desired $E\mu(\mathbf{A}^{-1}B)$ if we first let $M \to \infty$ and then $\varepsilon \to 0$. The so obtained limit is finite since $E\|\mathbf{A}\|^\alpha < \infty$ and

$$\begin{aligned} E\left[I_{(\varepsilon,M]}(\|\mathbf{A}\|)\mu(\mathbf{A}^{-1}B)\right] &= E\left[I_{(\varepsilon,M]}(\|\mathbf{A}\|)\mu\{\mathbf{x} : \mathbf{A}\mathbf{x} \in B\}\right] \\ &\le E\left[\|\mathbf{A}\|^\alpha\right]\, \mu\{\mathbf{x} : |\mathbf{x}| > t\} < \infty\,. \end{aligned}$$

Finally, we consider p_1. Then

$$\limsup_{n\to\infty} \frac{p_1}{P(|\mathbf{X}| > a_n)} \le \lim_{n\to\infty} \frac{P\left(\varepsilon|\mathbf{X}| > a_n\, t\right)}{P(|\mathbf{X}| > a_n)} = (t^{-1}\varepsilon)^\alpha\,.$$

We conclude that

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{p_1}{P(|\mathbf{X}| > a_n)} = 0\,. \tag{5.5}$$

Combining the limit results for p_1, p_2, p_3, we obtain the desired conclusion. □

Corollary 5.2 *Let* $\mathbf{X}$ *be regularly varying with index* α*, independent of the vector* $(Y_1, \ldots, Y_d)$ *which has independent components. Assume that* $E|Y_i|^{\alpha+\epsilon} < \infty$ *for some* $\epsilon > 0$*,* $i = 1, \ldots, d$*. Then* $(Y_1X_1, \ldots, Y_dX_d)$ *is regularly varying with index* α*.*

Acknowledgement. This research has been conducted during mutual visits of each of the authors at the other's home institution. The authors would like to thank Colorado State University and the University of Groningen for their hospitality. We would like to thank the referees and the Associate Editor for insightful comments which led to an improvement of the presentation of the paper.

References

[1] Babillot, M., Bougerol, P. and Elie, L. (1997) The random difference equation $X_n = A_nX_{n-1} + B_n$ in the critical case. *Ann. Probab.* **25** 478–493.

[2] Basrak, B., Davis, R.A. and Mikosch, T. (2000) A characterization of multivariate regular variation. Preprint. Available as mvreg.ps under www.math.ku.dk/∼mikosch/

[3] Billingsley, P. (1968) *Convergence of Probability Measures* Wiley, New York.

[4] Billingsley, P. (1986) *Probability and Measure* 2nd edition. Wiley, New York.

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

[6] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

[7] Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

[8] Bougerol, P. and Picard, N. (1992b) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.

[9] Boussama, F. (1998) *Ergodicité, mélange et estimation dans le modelès GARCH.* PhD Thesis, Université 7 Paris.

[10] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_nY_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.* **18**, 211–220.

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

[12] 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.

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

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

[15] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[16] Engle, R.F. (1995) *ARCH Selected Readings.* Oxford University Press, Oxford (UK).

[17] Feigin, P.D. and Tweedie, R.L. (1985) Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6**, 1–14.

[18] Furstenberg, H. and Kesten, H. (1960) Products of random matrices. *Ann. Math. Statist.* **31**, 457–469.

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

[20] Goldsheid, I. Ya. (1991) Lyapunov exponents and asymptotic behaviour of the product of random matrices. In: *Lecture Notes in Math.* **1486**, pp. 23–37. Springer, Berlin.

[21] Haan, L. de and Resnick, S.I. (1981) On the observation closest to the origin. *Stoch. Proc. Appl.* **11**, 301–308.

[22] Kallenberg, O. (1983) *Random Measures,* 3rd edition. Akademie–Verlag, Berlin.

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

[24] Kingman, J.F.C. (1973) Subadditive ergodic theory. *Ann. Probab.* **1**, 883–909.

[25] Lancaster, P. (1969) *Theory of Matrices.* Academic Press, New York.

[26] Lang, S. (1961) *Analysis I.* Addison Wesley, Reading, Massachusetts.

[27] Leadbetter, M.R. and Rootzén, H. (1988) Extremal theory for stochastic processes. *Ann. Probab.* **16**, 431–478.

[28] Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability.* Springer, London.

[29] 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.

[30] Nelson, D.B. (1990) Stationarity and persistence in the GARCH$(1,1)$ model. *Econometric Theory* **6**, 318–334.

[31] Pratt, J. (1960) On interchanging limits and integrals. *Ann. Math. Statist.*, **31**, 74–77.

[32] Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66–138.

[33] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[34] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London.

[35] Taylor, S.J. (1986) *Modelling Financial Time Series.* Wiley, Chichester.

[36] Vervaat, W. (1979) On a stochastic differential equation and a representation of a non–negative infinitely divisible random variables. *Adv. Appl. Prob.* **11**, 750–783.

Bojan Basrak
EURANDOM
P.O. Box 513
NL-5600 MB Eindhoven
THE NETHERLANDS

Richard A. Davis
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523–1877
U.S.A.

Thomas Mikosch
Laboratory of Actuarial Mathematics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen
DENMARK