
Extreme Value Theory for GARCH Processes

Richard A. Davis¹ and Thomas Mikosch²

¹ Department of Statistics, Colorado State University, Fort Collins, CO 80523, U.S.A., www.stat.colostate.edu/~rdavis, rdavis@stat.colostate.edu

² Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, www.math.ku.dk/~mikosch, mikosch@math.ku.dk

We consider the extreme value theory for a stationary GARCH process with iid innovations. One of the basic ingredients of this theory is the fact that, under general conditions, GARCH processes have power law marginal tails and, more generally, regularly varying finite-dimensional distributions. Distributions with power law tails combined with weak dependence conditions imply that the scaled maxima of a GARCH process converge in distribution to a Fréchet distribution. The dependence structure of a GARCH process is responsible for the clustering of exceedances of a GARCH process above high and low level exceedances. The size of these clusters can be described by the extremal index. We also consider the convergence of the point processes of exceedances of a GARCH process toward a point process whose Laplace functional can be expressed explicitly in terms of the intensity measure of a Poisson distribution and a cluster distribution.

1 The model

We consider a *GARCH process* $(X_t)_{t \in \mathbb{Z}}$ of order (p, q) (GARCH(p, q)) given by the equations³

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}. \quad (1)$$

In order to avoid ambiguity as regards the order (p, q) we assume that $\alpha_p \beta_q \neq 0$ and that all coefficients α_i and β_j are non-negative. Since we are not interested in the trivial solution $X_t \equiv 0$ a.s. to (1) we also assume $\alpha_0 > 0$.

³ Following the tradition in time series analysis, we index any stationary sequence (A_t) by the integers \mathbb{Z} . For practical purposes, one would consider the sequence $(X_t)_{t \in \mathbb{N}}$ corresponding to observations at the times $t = 1, 2, \dots$. If (A_t) is strictly stationary we write A for a generic element of the sequence.

Throughout, the *noise* or *innovations* sequence $(Z_t)_{t \in \mathbb{Z}}$ is an iid sequence with mean zero and unit variance.⁴ We refer to (σ_t) as the *volatility* sequence of the GARCH process.

2 Strict stationarity and mixing properties

To develop a reasonable extension of extreme value theory for iid sequences, the assumption of strict stationarity is often required. Therefore, we will always assume that there exists a unique strictly stationary causal⁵ solution (X_t) to (1). Notice that stationarity⁶ of the GARCH process (X_t) is equivalent to stationarity of the volatility sequence (σ_t) . Necessary and sufficient conditions for the existence and uniqueness of a stationary ergodic solution to (1) are given in Nelson [35] for the GARCH(1,1) case and for the general GARCH(p, q) case in Bougerol and Picard [2]; cf. Lindner [30]. These conditions will be discussed in Section 3.

Under general conditions such as the existence of a Lebesgue density of Z in some neighborhood of the origin, a stationary GARCH process (X_t) is ϕ -mixing, i.e.,

$$\sup_{C \in \mathcal{F}_{-\infty}^0, D \in \mathcal{F}_t^\infty} |P(D | C) - P(D)| = \phi_t \rightarrow 0, \text{ as } t \rightarrow \infty,$$

where $\mathcal{F}_{a,b}^b$, $a \leq b$, is the σ -field generated by $(X_s)_{a \leq s \leq b}$ with the obvious modifications for infinite a or b . Moreover, the mixing rate ϕ_t decays to 0 geometrically. These properties (under conditions on the distribution of Z more general than mentioned above) follow from work by Mokkadem [34], cf. Doukhan [14], Boussama [5].

Substantial insight into the probabilistic structure of a GARCH process (X_t) is gained by embedding the squares X_t^2 and σ_t^2 into a stochastic recurrence equation. This procedure offers one a way to find conditions for stationarity of (X_t) , but also for its marginal tail behavior and, in turn, for its extreme value behavior and the existence of moments.

3 Embedding a GARCH process in a stochastic recurrence equation

By the definition of a GARCH, the quantities X_t and σ_t are inextricably linked. Therefore any statement about the existence of the stationary sequence

⁴ A standardization like unit variance of Z is necessary in order to avoid a trade-off in the scaling between σ_t and Z_t which would lead to non-identifiability of the parameters of the model (1).

⁵ This means that X_t has representation as a measurable function of the past and present noise values Z_s , $s \leq t$.

⁶ In what follows, we will use stationarity as a synonym for strict stationarity.

(X_t) and its distributional properties is also a statement about the corresponding properties of the stationary sequence (σ_t) , and vice versa. Therefore one important approach to the understanding of the structure of a GARCH process is to express σ_t as a measurable function of past and present values Z_s , $s \leq t$. We refer to such a representation as a *causal solution* of (1).

Similarly to the representation of a causal ARMA process as a linear process of past values of the noise (see Brockwell and Davis [4]) such a representation can be obtained by iterating the defining difference equation (1) for σ_t^2 . Indeed, writing

$$\mathbf{Y}_t = \begin{pmatrix} \sigma_{t+1}^2 \\ \vdots \\ \sigma_{t-q+2}^2 \\ X_t^2 \\ \vdots \\ X_{t-p+1}^2 \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\ Z_t^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

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

we see that $((\mathbf{A}_t, \mathbf{B}_t))$ is an iid sequence, \mathbf{Y}_{t-1} and $(\mathbf{A}_t, \mathbf{B}_t)$ are independent, where the \mathbf{A}_t 's are iid random $(p+q-1) \times (p+q-1)$ matrices and the \mathbf{B}_t 's iid $(p+q-1)$ -dimensional random vectors. Then (\mathbf{Y}_t) satisfies the following vector *stochastic recurrence equation* (SRE):

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{Z}. \quad (2)$$

Iteration of the SRE (2) yields a unique stationary solution of the form

$$\mathbf{Y}_t = \mathbf{B}_t + \sum_{i=1}^{\infty} \mathbf{A}_t \cdots \mathbf{A}_{t-i+1} \mathbf{B}_{t-i}, \quad t \in \mathbb{Z}, \quad (3)$$

The crucial condition for the a.s. convergence of the infinite series in (3), hence for the existence of a strictly stationary solution to (2), is negativity of the *top Lyapunov exponent* given by

$$\gamma = \inf_{n \geq 1} n^{-1} E \log \|\mathbf{A}_n \cdots \mathbf{A}_1\|, \quad (4)$$

where $\|\cdot\|$ is the operator norm corresponding to a given norm in \mathbb{R}^{p+q-1} , see Bougerol and Picard [2]. By virtue of (3) the process (\mathbf{Y}_t) has representation

$\mathbf{Y}_t = f((Z_s)_{s \leq t})$ for some measurable function f . Therefore standard ergodic theory yields that (\mathbf{Y}_t) is an ergodic process, see Krengel [26].

In general, the top Lyapunov coefficient γ cannot be calculated explicitly, but a well known sufficient condition for $\gamma < 0$ is given by

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1, \quad (5)$$

see p. 122 in Bougerol and Picard [2]. Interestingly, (5) is also necessary and sufficient for the finiteness of $\text{var}(X)$, hence for the second order stationarity of (X_t) .

Example 1. (The GARCH(1,1) case)

This case is a real exception in the class of all GARCH processes since it is possible to calculate the Lyapunov coefficient explicitly. This is due to the fact that (2) essentially collapses into the one-dimensional SRE

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 Z_t^2 + \beta_1) \sigma_t^2, \quad (6)$$

where $A_t = \alpha_1 Z_t^2 + \beta_1$, hence

$$\gamma = n^{-1} E \log(A_n \cdots A_1) = E \log A = E \log(\alpha_1 Z^2 + \beta_1).$$

Then, following Nelson [35], the conditions $E \log(\alpha_1 Z^2 + \beta_1) < 0$ and $\alpha_0 > 0$ are necessary and sufficient for the existence of a stationary causal non-degenerate solution to (1). \square

4 The tails of a GARCH process

The embedding of the squares X_t^2 and σ_t^2 of a stationary GARCH process (X_t) and its volatility sequence (σ_t) into the SRE (2) also allows one to use classical theory about the tails of the solution to this SRE. Such a theory was developed by Kesten [25], see also Goldie [19] for an alternative approach, cf. Embrechts et al. [16], Section 8.4.

Theorem 1. (Basrak et al. [1]) *Consider the process (\mathbf{Y}_t) in (3) obtained from embedding a stationary GARCH process into the SRE (2). Assume that Z has a positive density on \mathbb{R} such that $E(|Z|^h) < \infty$ for $h < h_0$ and $E(|Z|^{h_0}) = \infty$ for some $h_0 \in (0, \infty]$.⁷*

Then there exist $\kappa > 0$ and a finite-valued function w on the unit sphere

$$\mathbb{S}^{p+q-2} = \{\mathbf{x} \in \mathbb{R}^{p+q-1} : |\mathbf{x}| = 1\}$$

⁷ A condition such as $E(|Z|^{h_0}) = \infty$ is needed. It means that the distribution of Z is spread ‘‘sufficiently far out’’. Indeed, if Z is supported on too short an interval in the neighborhood of the origin, then $P(|X| > x)$ may decay to zero exponentially fast as $x \rightarrow \infty$, see Goldie and Gröbel [20].

for any fixed norm $|\cdot|$ in \mathbb{R}^{p+q-1} such that

$$\text{for all } \mathbf{x} \in \mathbb{S}^{p+q-2}, \quad \lim_{x \rightarrow \infty} x^\kappa P((\mathbf{x}, \mathbf{Y}) > x) = w(\mathbf{x}) \quad \text{exists.}$$

Moreover, if $\mathbf{x} \in \mathbb{S}^{p+q-2}$ has non-negative components then $w(\mathbf{x}) > 0$.

Furthermore, if κ is not even,⁸ then \mathbf{Y} is regularly varying with index κ_1 , i.e., there exist a constant $c > 0$ and a random vector Θ on the unit sphere \mathbb{S}^{p+q-2} such that for every $t > 0$

$$x^\kappa P(|\mathbf{Y}| > tx, \mathbf{Y}/|\mathbf{Y}| \in \cdot) \xrightarrow{w} c t^{-\kappa} P(\Theta \in \cdot), \quad \text{as } x \rightarrow \infty,$$

where \xrightarrow{w} denotes weak convergence on the Borel σ -field of \mathbb{S}^{p+q-2} .

If we specify \mathbf{x} to be a unit vector an immediate consequence is the following result.

Corollary 1. *Under the conditions on the distribution of Z in Theorem 1, the tails of the marginal distribution of a stationary GARCH process exhibit power law behavior: there exist $\kappa > 0$ and positive constants $c_{|X|}$ and c_σ such that⁹*

$$P(|X| > x) \sim c_{|X|} x^{-2\kappa} \quad \text{and} \quad P(\sigma > x) \sim c_\sigma x^{-2\kappa}.$$

Applying a result by Breiman [3], cf. Davis and Mikosch [9], Section 4, one can also derive the relation

$$P(|X| > x) = P(\sigma |Z| > x) \sim E(|Z|^{2\kappa}) P(\sigma > x).$$

Now assume that Z (hence X) is symmetric, e.g., when Z has a normal or student distribution. Then $\varepsilon = \text{sign}(Z)$ and $|Z|$ are independent, hence

$$P(X > x) = \frac{1}{2} P(|X| > x) \sim \frac{1}{2} E(|Z|^{2\kappa}) c_\sigma x^{-2\kappa}.$$

With the exception of the GARCH(1, 1) case it is unknown how to evaluate c_σ , see Goldie [19].

In the case of the general SRE model (2), the value of κ is determined as the solution to a complicated equation (see Kesten [25]) which cannot be solved explicitly. As regards the general GARCH(p, q) model, one can estimate the *tail index* κ from observations of the GARCH process by using the tools of extreme value statistics, see Drees [15] and Resnick and Stărică [37], cf. Embrechts et al. [16], Chapter 6. The GARCH(1, 1) case again offers a more tractable form for κ .

⁸ It is not clear whether this condition can be omitted, see for example Hult and Lindskog [22] for some related work.

⁹ We write $f(x) \sim g(x)$ as $x \rightarrow \infty$ for two functions f and g whenever $f(x)/g(x) \rightarrow 1$ as $x \rightarrow \infty$.

Example 2. (The GARCH(1,1) case)

In this case, the SRE (2) collapses to the one-dimensional equation (6). Then direct calculation shows that Kesten's result yields κ as the unique solution (which exists under the conditions of Theorem 1) to the equation

$$E[A^\kappa] = E[(\alpha_1 Z^2 + \beta_1)^\kappa] = 1. \quad (7)$$

This equation can be solved for κ by numerical and/or simulation methods for fixed values of α_1 and β_1 from the stationarity region of a GARCH(1,1) process and assuming a concrete density for Z . In the case $\alpha_1 = 0.1$, Table 1 reports a small study for 2κ -values (the tail index of X and σ), assuming a standard normal or a unit variance student distribution with 4 degrees of freedom.

We mention that the inequality

$$E[(\alpha_1 Z^2 + \beta_1)^\kappa] \geq 1 \quad (8)$$

already implies that both $E(\sigma^{2\kappa})$ and $E(|X|^{2\kappa})$ are infinite. Indeed, since $\alpha_0 > 0$

$$E(\sigma_t^{2\kappa}) > E(\sigma_{t-1}^{2\kappa} (\alpha_1 Z_{t-1}^2 + \beta_1)^\kappa) = E(\sigma_{t-1}^{2\kappa}) E((\alpha_1 Z^2 + \beta_1)^\kappa).$$

By virtue of (8) this relation is impossible unless $E(\sigma^{2\kappa}) = \infty$. On the other hand, if

$$E[(\alpha_1 Z^2 + \beta_1)^\kappa] < 1 \quad (9)$$

a simple moment calculation involving Minkowski's inequality and representation (3) for $Y_t = \sigma_t^2$ imply that $E(|X|^{2\kappa}) = E(|Z|^{2\kappa}) E(\sigma^{2\kappa}) < \infty$. Similar arguments for the existence of moments of the solution to a one-dimensional SRE can be found in Ling and McAleer [28] and have been extended to GARCH and AGARCH models in Ling and McAleer [29]. \square

Example 3. (Integer valued κ and IGARCH)

Assume that we are in the GARCH(1,1) framework of Example 2. Since the solution κ of (7) is unique if it exists, it is possible to give necessary and sufficient conditions for tails of the form $P(|X| > x) \sim c_{|X|} x^{-2k}$ for $k = 1, 2, \dots$. By virtue of the discussion of the consequences of the conditions (8) and (9) we may also conclude that $E(X^{2k}) < \infty$ if and only if $E[(\alpha_1 Z^2 + \beta_1)^k] < 1$.

Relation (7) turns into

$$1 = E[(\alpha_1 Z^2 + \beta_1)^k] = \beta_1^k \sum_{l=0}^k \binom{k}{l} (\alpha_1/\beta_1)^l E(Z^{2l}). \quad (10)$$

For example, choosing $k = 1$ and recalling that $E(Z^2) = 1$, (10) turns into the equation $\alpha_1 + \beta_1 = 1$ which defines the integrated GARCH (IGARCH)

Table 1. Results for 2κ when $\alpha_1 = 0.1$. Top: Standard normal noise. Bottom: Unit variance student noise with 4 degrees of freedom.

β_1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
2κ	2.0	12.4	16.4	19.0	21.0	23.0	24.8	26.4	28.2	29.6
β_1	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89	
2κ	12.0	11.4	10.8	10.0	9.0	8.0	7.0	5.6	3.8	
β_1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
2κ	2.0	4.0	4.4	4.4	4.4	4.4	4.8	4.8	4.6	4.8
β_1	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89	
2κ	4.2	4.0	4.0	3.8	3.6	3.4	3.4	3.0	2.6	

model of Engle and Bollerslev [17]. It is a well known fact that a stationary IGARCH(p, q) process has infinite variance marginals.¹⁰ In the IGARCH(1,1) case, Kesten’s result yields the more sophisticated result $P(|X| > x) \sim c_{|X|} x^{-2}$. In applications to real-life data one often observes that the sum of the estimated parameters $\widehat{\alpha}_1 + \widehat{\beta}_1$ is close to 1 implying that moments slightly larger than two might not exist for a fitted GARCH process. An alternative explanation for this *IGARCH effect* are non-stationarities in the observed data as explained in Mikosch and Stărică [33].

For $k = 2$, (10) turns into the equation

$$1 = (\alpha_1 + \beta_1)^2 + \alpha_1^2 (E(Z^4) - 1),$$

implying that $P(|X| > x) \sim c_{|X|} x^{-4}$. Moreover, following the discussion above, the condition

$$1 > (\alpha_1 + \beta_1)^2 + \alpha_1^2 (E(Z^4) - 1)$$

is necessary and sufficient for $(EX^4) < \infty$. \square

5 Limit theory for extremes

5.1 Convergence of maxima

Generally, the asymptotic tail behavior of the marginal distribution of a stationary mixing sequence sequence (A_t) is the key factor in determining the

¹⁰ This follows immediately by taking expectations in the defining equation (1) for σ_t^2 and recalling that $\alpha_0 > 0$.

weak limit behavior of the sequence of partial maxima

$$M_n(A) = \max_{i=1, \dots, n} A_i, \quad n \geq 1. \quad (11)$$

This property is well documented in various monographs on the topic, see e.g., Leadbetter et al. [27], Galambos [18], Resnick [36], Embrechts et al. [16]. In particular, for an iid sequence (A_t) with tail $P(A > x) \sim cx^{-\alpha}$ for some constant $c > 0$ it follows that

$$(cn)^{-1/\alpha} M_n(A) \xrightarrow{d} Y_\alpha, \quad (12)$$

where Y_α has the Fréchet distribution function,

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ e^{-x^{-\alpha}}, & x > 0. \end{cases} \quad (13)$$

Now if (A_t) is a stationary sequence that satisfies a general mixing condition and the marginal distribution has the same tail behavior as above (i.e., $P(A > x) \sim cx^{-\alpha}$), then one often can show the existence of $\theta_A \in (0, 1]$ such that

$$(cn)^{-1/\alpha} M_n(A) \xrightarrow{d} \theta^{1/\alpha} Y_\alpha, \quad (14)$$

The parameter θ_A is called the *extremal index* and measures the level of clustering for extremes. The case $\theta_A = 1$ corresponds to no clustering in which case the limiting behavior of $M_n(A)$ is the same as for the maxima of an iid sequence. The reciprocal of the extremal index $1/\theta_A$ of a stationary sequence (A_t) has the interpretation as the expected size of clusters of high level exceedances in the sequence. For iid A_t 's, $1/\theta_A = 1$, see (11), and there is no clustering of exceedances of high levels.

Returning to the GARCH setting, we assume that the conditions of Theorem 1 are satisfied. Then we know that $P(|X| > x) \sim c_{|X|} x^{-2\kappa}$ for some $\kappa, c_{|X|} > 0$, and we can even specify the value of κ in the GARCH(1, 1) case by solving the equation (7). The limit relation (12) is not directly applicable to the maxima $M_n(|X|)$ of the sequence $(|X_t|)$ because of the dependence in the sequence. However, it was mentioned in Section 2 that, under general conditions on the distribution of Z , the sequence (X_t) , hence $(|X_t|)$ is ϕ -mixing with geometric rate. Using this asymptotic independence condition it can be shown¹¹ that

$$(c_{|X|} n)^{-1/(2\kappa)} M_n(|X|) \xrightarrow{d} \theta_{|X|}^{1/(2\kappa)} Y_{2\kappa} \quad (15)$$

where $Y_{2\kappa}$ has the Fréchet distribution given in (13) and the extremal index $\theta_{|X|}$ is strictly less than one. In the aforementioned literature, it has also been

¹¹ See de Haan et al. [21] for ARCH(1); cf. Embrechts et al. [16], Section 8.4, or Davis and Mikosch [7]; Mikosch and Stărică [32] for GARCH(1,1), and Basrak et al. [1] in the general GARCH(p, q) case.

shown that results analogous to (15) hold for the maxima of the sequences (X_t) (at least when Z is symmetric) with the corresponding positive tail constants c_X, c_σ in (15) and extremal indices $\theta_X, \theta_\sigma \in (0, 1)$ in (14).

Since the extremal index $\theta_{|X|}$ is strictly less than one for a GARCH process, the expected size of clusters of high level exceedances is $1/\theta_{|X|} > 1$. This is in sharp contrast to the case of stochastic volatility processes (see [9]), which have extremal indices that are equal to one and hence possess no clustering of extremes. Formulae for calculating the value of $\theta_{|X|}$ for a general GARCH(p, q) process are unknown, but in the ARCH(1) and GARCH(1, 1) cases more explicit expressions for $\theta_{|X|}, \theta_X$ and θ_σ exist. For example, in the GARCH(1, 1) case,

$$\theta_\sigma = \int_1^\infty P \left(\sup_{t \geq 1} \prod_{j=1}^t (\alpha_1 Z_j^2 + \beta_1) \leq y^{-1} \right) \kappa y^{-\kappa-1} dy.$$

The right hand expression can be evaluated by Monte-Carlo simulations, see e.g. de Haan et al. [21] for the ARCH(1) case with standard normal noise Z , cf. Embrechts et al. [16], Section 8.1, where one can also find some advice as to how the extremal index of a stationary sequence can be estimated from data.

5.2 Convergence of point processes

Advanced insight into the limit structure of the extremes of a stationary sequence (A_t) is provided by the weak convergence of the point processes

$$N_n(\cdot) = \sum_{t=1}^n \varepsilon_{A_t/a_n}(\cdot)$$

toward a point process, whose points can be expressed in terms of products of Poisson points with independent points from a clustered distribution. Here ε_x denotes *Dirac measure* at x : for any set $B \subset \mathbb{R}$

$$\varepsilon_x(B) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B, \end{cases}$$

and (a_n) is a suitably chosen sequence of positive constants. For example, for a GARCH process (X_t) , (a_n) can be chosen such that

$$P(X > a_n) \sim n^{-1}.$$

Convergence in distribution, $N_n \xrightarrow{d} N$, of the sequence of point processes (N_n) toward the point process N is explained in standard books on point processes, see e.g. Kallenberg [24], Daley and Vere-Jones [6], Resnick [36]. Resnick's book

describes the close relationship between the convergence of (N_n) and extreme value theory. For example, choosing the set $B = (x, \infty]$, $N_n \xrightarrow{d} N$ implies for the order statistics $A_{(1)} \leq \dots \leq A_{(n)} = M_n(A)$ of the sample A_1, \dots, A_n that

$$\begin{aligned} P(N_n(x, \infty) < k) &= P(a_n^{-1} A_{(n-k+1)} \leq x) \\ &\rightarrow P(N(x, \infty) < k) \\ &= \sum_{i=0}^{k-1} P(N(x, \infty) = i), \quad x \in \mathbb{R}. \end{aligned}$$

Similar relations can be established for the joint convergence of finitely many order statistics in a sample, the joint convergence of the scaled minima $a_n^{-1} A_{(1)}$ and maxima $a_n^{-1} M_n(A)$, and various other results for extremes can be derived as well. Among others, the distribution of N determines the extremal index θ_A mentioned in Section 5.1.

The limit distribution of $(a_n^{-1} M_n(A))$ is determined by the distribution of the limiting point process N . For ARCH(1), GARCH(1, 1) and the general GARCH(p, q) processes (X_t) and their absolute values $(|X_t|)$ the form of the limit point process N was determined in Davis and Mikosch [7], Mikosch and Střičá [32] and Basrak et al. [1], respectively. The Laplace functional of the limit point process N can be expressed explicitly in terms of the intensity measure of a Poisson distribution and the distribution of clusters. However, the general form of this representation has little practical value for probability computations.

5.3 The behavior of the sample autocovariance function

Basic measures of the dependence in a stationary time series (A_t) are the *autocovariance* and *autocorrelation functions* (ACVF and ACF) given respectively by

$$\gamma_A(h) = \text{cov}(A_0, A_h) \quad \text{and} \quad \rho_A(h) = \text{corr}(A_0, A_h) = \frac{\gamma_A(h)}{\gamma_A(0)}, \quad h \geq 0.$$

Here we have assumed that $\gamma_A(0) = \text{var}(A) \in (0, \infty)$. Log-return series (X_t) usually have zero autocorrelations and therefore it is also common to consider the ACVFs and ACFs, $\gamma_{|X|^i}$ and $\rho_{|X|^i}$, $i = 1, 2$, of the absolute values $|X_t|$ and the squared returns X_t^2 .

Since the ACVF and ACF of a stationary sequence (A_t) are in general unknown functions it is common to estimate them by their sample analogs (*sample ACVF* and *sample ACF*) given at lag $h \geq 0$ by

$$\hat{\gamma}_A(h) = \frac{1}{n} \sum_{t=1}^{n-h} (A_t - \bar{A}_n)(A_{t+h} - \bar{A}_n)$$

and

$$\hat{\rho}_A(h) = \frac{\hat{\gamma}_A(h)}{\hat{\gamma}_A(0)} = \frac{\sum_{t=1}^{n-h} (A_t - \bar{A}_n)(A_{t+h} - \bar{A}_n)}{\sum_{t=1}^n (A_t - \bar{A}_n)^2},$$

where $\bar{A}_n = n^{-1} \sum_{t=1}^n A_t$ denotes the *sample mean*.

For non-linear processes (A_t) the limit theory for the sample ACVF and sample ACF is strongly influenced by heavy tails in the marginal distribution of A_t . This has been reported early on in Davis and Resnick [13] for bilinear processes. In contrast, the sample ACF of a linear process (such as ARMA and FARIMA) consistently estimates the ACF of a Gaussian linear process with the same coefficients as the original linear process even when the distribution of A has such heavy tails that the variance of the process is infinite, in which case, the ACVF and the ACF are not defined. Interestingly, however, the rates of convergence in this heavy tailed case compare favorably to the usual \sqrt{n} -rates in the finite variance case, see Davis and Resnick [10, 11, 12], cf. Brockwell and Davis [4], Section 13.3.

The limit theory for the sample ACVF and ACF of a GARCH process (X_t) , its absolute values and squares was studied in Davis and Mikosch [7] in the ARCH(1) case, for GARCH(1,1) in Mikosch and Stărică [32] and in Basrak et al. [1] for the general GARCH(p, q) case as well as for solutions to SREs. The proofs of these results are based on the point process convergence results mentioned in Section 5.2, hence they are closely related to extreme value theory for GARCH processes. The limit distribution and the rates of convergence of the sample ACFs $\hat{\rho}_X(h)$, $\hat{\rho}_{|X|}(h)$ and $\hat{\rho}_{X^2}(h)$ critically depend on the tail index 2κ of the marginal distribution of a GARCH process, see Corollary 1. In particular, the following results hold.

1. If $2\kappa \in (0, 2)$ then $\hat{\rho}_X(h)$ and $\hat{\rho}_{|X|}(h)$ have non-degenerate limit distributions. The same statement holds for $\hat{\rho}_{X^2}(h)$ when $2\kappa \in (0, 4)$.
2. If $2\kappa \in (2, 4)$ then both $\hat{\rho}_X(h)$, $\hat{\rho}_{|X|}(h)$ converge in probability to their deterministic counterparts $\rho_X(h)$, $\rho_{|X|}(h)$, respectively, at the rate $n^{1-2/(2\kappa)}$ and the limit distribution, depending on an infinite variance stable distribution, is complex and difficult to describe.
3. If $2\kappa \in (4, 8)$, then

$$n^{1-4/(4\kappa)}(\hat{\rho}_{X^2}(h) - \rho_{X^2}(h)) \xrightarrow{d} S_\kappa(h),$$

where the random variable $S_\kappa(h)$ is a function of infinite variance stable random variables.

4. If $2\kappa > 4$ then the good mixing properties of the GARCH process (see Section 2) and standard central limit theory for stationary sequences (see e.g., Ibragimov and Linnik [23] or Doukhan [14]) imply that $(\hat{\rho}_X(h))$ and $(\hat{\rho}_{|X|}(h))$ have Gaussian limits at \sqrt{n} -rates. The corresponding result holds for (X_t^2) when $2\kappa > 8$.

These results show that the limit theory for the sample ACF of a GARCH process is rather complicated when low moments of the process do not exist.

There is empirical evidence based on extreme value statistics indicating that log-return series might not have finite 4th or 5th moment,¹² and then the limit results above would show that the usual confidence bands for the sample ACF based on the central limit theorem and the corresponding \sqrt{n} -rates are far too optimistic in this case.

It is worth noting that for stochastic volatility models, an alternative class of models to the GARCH for modeling log-returns, the situation is markedly different. If the noise in the stochastic volatility process is chosen so that the marginal distribution matches the power law tail of the GARCH with index 2κ , then

$$(n/\ln n)^{1/(2\kappa)}\widehat{\rho}_X(h) \quad \text{and} \quad (n/\ln n)^{1/(4\kappa)}\widehat{\rho}_{X^2}(h)$$

converge in distribution for $2\kappa \in (0, 2)$ and $2\kappa \in (0, 4)$, respectively. This illustrates the good large sample behavior of the sample ACF for stochastic volatility models even if ρ_X and ρ_{X^2} are not defined (see Davis and Mikosch [8, 9]) and illustrates another key difference between stochastic volatility and GARCH processes.

References

1. BASRAK, B., DAVIS, R.A. AND MIKOSCH, T. (2002) Regular variation of GARCH processes. *Stoch. Proc. Appl.* **99**, 95–116.
2. BOUGEROL, P. AND PICARD, N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.
3. BREIMAN, L. (1965) On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10**, 323–331.
4. BROCKWELL, P.J. AND DAVIS, R.A. (1991) *Time Series: Theory and Methods*, 2nd edition. Springer, Berlin, Heidelberg, New York.
5. BOUSSAMA, F. (1998) *Ergodicité, mélange et estimation dans le modèles GARCH*. PhD Thesis, Université 7 Paris.
6. DALEY, D. AND VERE-JONES, D. (1988) *An Introduction to the Theory of Point Processes*. Springer, Berlin.
7. DAVIS, R.A. AND MIKOSCH, T. (1998) Limit theory for the sample ACF of stationary process with heavy tails with applications to ARCH. *Ann. Statist.* **26** (1998), 2049–2080.
8. DAVIS, R.A. AND MIKOSCH, T. (2001) The sample autocorrelations of financial time series models. In: W.J. Fitzgerald, R.L. Smith, A.T. Walden and P.C. Young (Eds.) *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press, Cambridge (2001), 247–274.
9. DAVIS, R.A. AND MIKOSCH, T. (2007) Probabilistic properties of stochastic volatility models. This volume, pp. ???
10. DAVIS, R.A. AND RESNICK, S.I. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179–195.

¹² See e.g. Embrechts et al. [16], Chapter 6, and Mikosch [31].

11. DAVIS, R.A. AND RESNICK, S.I. (1985) More limit theory for the sample correlation function of moving averages. *Stoch. Proc. Appl.* **20**, 257–279.
12. DAVIS, R.A. AND RESNICK, S.I. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533–558.
13. DAVIS, R.A. AND RESNICK, S.I. (1996) Limit theory for bilinear processes with heavy tailed noise. *Ann. Appl. Probab.* **6**, 1191–1210.
14. DOUKHAN, P. (1994) *Mixing. Properties and Examples*. Lecture Notes in Statistics **85**. Springer Verlag, New York.
15. DREES, H. (2000) Weighted approximations of tail processes for β -mixing random variables. *Ann. Appl. Probab.* **10**, 1274–1301.
16. EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
17. ENGLE, R.F. AND BOLLERSLEV, T. (1986) Modelling the persistence of conditional variances. With comments and a reply by the authors. *Econometric Rev.* **5**, 1–87.
18. GALAMBOS, J. (1987) *Asymptotic Theory of Extreme Order Statistics*, 2nd edition. Krieger, Malabar, Florida.
19. GOLDIE, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.
20. GOLDIE, C.M. AND GRÜBEL, R. (1996) Perpetuities with thin tails. *Adv. Appl. Probab.* **28**, 463–480.
21. HAAN, L. DE, RESNICK, S.I., ROOTZÉN, H. AND VRIES, C.G. DE (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Proc. Appl.* **32**, 213–224.
22. HULT, H. AND LINDSKOG, F. (2006) On Kesten’s counterexample to the Cramér-Wold device for regular variation. *Bernoulli* **12**, 133–142.
23. IBRAGIMOV, I.A. AND LINNIK, YU.V. (1971) *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
24. KÄLLENBERG, O. (1983) *Random Measures*, 3rd edition. Akademie-Verlag, Berlin.
25. KESTEN, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.
26. KRENGEL, U. (1985) *Ergodic Theorems*. De Gruyter, Berlin.
27. LEADBETTER, M.R., LINDGREN, G. AND ROOTZÉN, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.
28. LING, S. AND MCALEER, M. (2002) Stationarity and the existence of moments of a family of GARCH processes. *J. Econometrics* **106**, 109–117.
29. LING, S. AND MCALEER, M. (2002) Necessary and sufficient conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory* **18**, 722–729.
30. LINDNER, A. (2007) Fundamental properties of GARCH processes. This volume, pp.???
31. MIKOSCH, T. (2003) Modelling dependence and tails of financial time series. In: B. FINKENSTÄDT AND H. ROOTZÉN (EDS.) *Extreme Values in Finance, Telecommunications and the Environment*. Chapman and Hall, pp. 185–286.
32. MIKOSCH, T. AND STĂRICĂ, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Stat.* **28**, 1427–1451.
33. MIKOSCH, T. AND STĂRICĂ, C. (2004) Non-stationarities in financial time series, the long-range dependence and the IGARCH effects. *Review of Economics and Statistics* **86**, 378–390.

34. MOKKADEM, A. (1990) Propriétés de mélange des processus autoregressifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.* **26**, 219–260.
35. NELSON, D.B. (1990) Stationarity and persistence in the GARCH(1, 1) model. *Econom. Theory* **6**, 318–334.
36. RESNICK, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
37. RESNICK, S.I. AND STĂRICĂ, C. (1998) Tail index estimation for dependent data. *Ann. Appl. Probab.* **8**, 1156–1183.