Limit theory and robust evaluation methods for the extremal properties of GARCH\((p, q)\) processes

Fabrizio Laurini 1 · Paul Fearnhead 2 · Jonathan Tawn 2

Received: 6 May 2022 / Accepted: 1 October 2022 / Published online: 1 November 2022
© The Author(s) 2022, corrected publication 2023

Abstract

Generalized autoregressive conditionally heteroskedastic (GARCH) processes are widely used for modelling financial returns, with their extremal properties being of interest for market risk management. For GARCH\((p, q)\) processes with \(\max(p, q) = 1\) all extremal features have been fully characterised, but when \(\max(p, q) \geq 2\) much remains to be found. Previous research has identified that both marginal and dependence extremal features of strictly stationary GARCH\((p, q)\) processes are determined by a multivariate regular variation property and tail processes. Currently there are no reliable methods for evaluating these characterisations, or even assessing the stationarity, for the classes of GARCH\((p, q)\) processes that are used in practice, i.e., with unbounded and asymmetric innovations. By developing a mixture of new limit theory and particle filtering algorithms for fixed point distributions we produce novel and robust evaluation methods for all extremal features for all GARCH\((p, q)\) processes, including ARCH and IGARCH processes. We investigate our methods’ performance when evaluating the marginal tail index, the extremogram and the extremal index, the latter two being measures of temporal dependence.

Keywords Cluster of extremes · Extremal index · Fixed point distributions · GARCH process · Multivariate regular variation · Particle filtering · Stochastic recurrence equations

Mathematics Subject Classification 62M10 · 62G32 · 60G70 · 60G10

1 Introduction

Risk management in the stock markets, commonly called market risk management, requires the use of statistical tools and models which aim at reducing the potential size of losses, occurring by sudden drops or growth in the value of stock. Losses can be amplified during periods of large volatility. Risk managers routinely use strategies to handle, model and predict the volatility of daily log-returns, defined as \(X_t = \log P_t - \log P_{t-1}\), where \(P_t, t = 1, 2, \ldots\), is the price of a generic asset. The behaviour of the extreme values of daily log-returns is critically important for market risk management. Marginally these processes have power-law heavy tail decay and determining the associated (power) tail index is therefore important to quantify this. Additionally, isolated extreme values of daily log-return can often be managed, but there is major risk when there is a clustering of these extreme values, and so the study of this dependence structure during extreme events is essential.

It is standard to assume that the series \(\{X_t\}\) is a stationary series. The most widely adopted models for \(\{X_t\}\) are the generalised autoregressive conditionally heteroskedastic (GARCH) models (Bollerslev 1986) and stochastic volatility models (Taylor 1986). These models are capable of capturing many of the empirical features of daily log-returns. Both processes have heavy tailed marginal distributions but they differ in terms of their extremal dependence structure. One of the most common ways to measure this is through the lag \(\tau\) tail dependence

\[ \chi_{X}(\tau) = \lim_{x \to \infty} \Pr(X_{t+\tau} > x \mid X_t > x), \]
proposed by Ledford and Tawn (2003), with Davis and Mikosch (2009b) terming \( \chi_X(\tau) \) the extremogram. For stochastic volatility models Breidt and Davis (1998) show that there is no clustering of extreme values, so that \( \chi_X(\tau) = 0 \) for all \( \tau > 0 \). Thus extreme values from this stochastic volatility process occur in temporal isolation, although there are examples of other such processes for which this is not the case (Mikosch and Reszep 2013). In contrast, for any GARCH(\( p, q \)) process \( \chi_X(\tau) > 0 \) for at least one value of \( \tau > 0 \). The values of the extremogram, and other extremal dependence features, are only known for a very restricted subclass of GARCH(\( p, q \)) processes.

The aim of our paper is to derive all the extremal features for all GARCH(\( p, q \)) models used in a wide range of financial applications and to present computationally efficient algorithms for their evaluation. We consider GARCH(\( p, q \)) models, for \( p \in \mathbb{N} \) and \( q \in \mathbb{N}_+ \), of the form

\[
X_t = \sigma_t Z_t, \quad \text{where} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,
\]

for \( t \in \mathbb{Z} \), where the parameters are \( \alpha_0 > 0, \alpha_i \geq 0, i = 1, \ldots, q - 1, \alpha_q > 0 \) and \( \beta_j \geq 0, j = 1, \ldots, p - 1, \beta_p > 0 \) for \( p \geq 1 \), and \( \{Z_t\} \) are independent and identically distributed (i.i.d.) random variables with \( E(Z_t) = 0 \) and \( \text{Var}(Z_t) = 1 \) for all \( t \in \mathbb{Z} \). The processes \( \{\sigma_t\} \) and \( \{Z_t\} \) are commonly referred to as the conditional volatility and innovations of \( \{X_t\} \). We additionally assume that \( \{Z_t\} \) are continuous random variables, which avoids technical issues that occur when \( \text{Pr}(Z_t = 0) > 0 \) (Buraczewski et al. 2016). For the process to be strictly stationary the parameters need to satisfy the constraints in Sect. 2.2. Properties of GARCH(\( p, q \)) processes are often determined by

\[
\phi := \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j.
\]

Two important special cases of GARCH(\( p, q \)) processes arise when \( p = 0 \) or \( \phi = 1 \), corresponding to ARCH(\( q \)) and IGARCH(\( p, q \)) processes respectively. The IGARCH(\( p, q \)) process is strictly stationary but not weakly stationary, due to violating the condition \( E(X_t^2) < \infty \) for all \( t \). We develop new numerically robust and efficient methods for assessing whether any GARCH(\( p, q \)) process is strictly stationary. Our results cover all of the GARCH(\( p, q \)) processes including these special cases.

Existing theoretical and computational methods for deriving the extremal properties are well established for special cases of the GARCH(\( p, q \)) process, namely: for symmetric \( Z_t \) with \( p = 0, q = 1 \), corresponding to the ARCH(1) process (de Haan et al. 1989) and for \( p = q = 1 \), corresponding to a GARCH(1,1) (Lauiri and Tawn 2012); and for asymmetric \( Z_t \) with \( p = q = 1 \) (Ehleiet al. 2015). For a class of processes which contain both the ARCH(1) and GARCH(1,1) processes, but not any GARCH(\( p, q \)) process with \( \max(p, q) \geq 2 \), Collamore et al. (2014) provide an algorithm to derive \( \text{Pr}(X_t > v) \) for any \( v \). For general GARCH(\( p, q \)) models, with arbitrary \( (p, q) \) many theoretical extremal properties have been derived by Basrak et al. (2002), Davis and Mikosch (2009a), Basrak and Segers (2009) and Collamore and Mentemeier (2018), including marginal distributions and some results for the extremal clustering properties. At first sight it seems that these results give everything that is needed for their numerical evaluation. But, as we will show, this is far from the case. In fact, until now none of these features can be evaluated for general GARCH(\( p, q \)) processes or for the processes considered by Collamore and Mentemeier (2018).

Consider the marginal tail behaviour of \(|X_t|\) and \(X_t^2\) for GARCH(\( p, q \)) processes \( X_t \). Basrak et al. (2002) give a limit expression, as \( u \to \infty \), that there exists \( \kappa > 0 \) such that for fixed \( x > 1 \)

\[
\text{Pr}(|X_t| > u x | |X_t| > u) \to x^{-2\kappa}, \quad \text{and}
\]

\[
\text{Pr}(X_t^2 > u x | X_t^2 > u) \to x^{-\kappa}.
\]

These papers only give an asymptotic limiting expression for \( \kappa \), but do not evaluate it. We find that direct computation using their expression gives very poor numerical performance. Janssen (2010) presents an alternative approach to evaluate \( \kappa \), however that approach requires \( Z_t \) to have bounded support, ruling out distributions used by practitioners, e.g., \( Z_t \) being Gaussian or \( t \)-distributed. Furthermore, the associated numerical methods are very slow and unreliable. Collamore and Mentemeier (2018) characterise the limit \( \lim_{u \to \infty} u^k \text{Pr}(X_t^2 > u) = C > 0 \), with \( k \) formulated as in Basrak et al. (2002). So, without techniques to obtain \( \kappa \) their methods cannot evaluate \( C \) or any of the other properties they characterise. We propose the first computationally efficient numerical algorithm for the evaluation of \( \kappa \), which is valid irrespective of whether \( Z_t \) are unbounded or bounded. We also provide strong empirical evidence for how \( \kappa \) influences \( \kappa \), including \( \kappa = 1 \) for all IGARCH(\( p, q \)) and how our evaluation of \( \kappa \) can be used for assessing stationarity.

Now consider the extremal dependence and the clustering features of GARCH(\( p, q \)) processes. Basrak and Segers (2009) propose algorithms for their evaluation with bounded innovations. However, Basrak and Segers (2011) identify these methods have major limitations and that they do not hold for any GARCH(\( p, q \)) process. So currently no extremal clustering features for GARCH(\( p, q \)) processes, when \( \max(p, q) \geq 2 \), can be evaluated. We propose an entirely new algorithm to evaluate a range of cluster fea-
tures regardless of \( p \) and \( q \) and for \( Z_t \) having any continuous distribution.

There are a range of features of GARCH processes that are of interest in managing financial stress periods. We focus primarily on \( \{X_t(\tau) : \tau \geq 0\} \), discussed above, and the extremal index for the process \( X_t \) denoted \( 0 < \theta_X \leq 1 \), where \( 1/\theta_X \) is the limiting mean number of extreme values per independent extreme event (Hsing et al. 1988). We also evaluate these properties for the lower tail and for the square of the GARCH process. Other features can also be derived using our methods, such as the duration of a stress period and the total losses in such a period.

All of these cluster functionals can be obtained from the tail process (Smith et al. 1997; Segers 2003; Plaminić and Soulier 2018). For a heavy tailed process \( \{X_t\} \) the tail process \( \{\hat{X}_t\}_{t \geq 0} \) is defined in the following way. If for any \( t \in \mathbb{N} \), the variables \( (X_0/u, X_1/u, \ldots, X_t/u) \mid X_0 > u \) converge weakly to \( (\hat{X}_0, \hat{X}_1, \ldots, \hat{X}_t) \), with \( \hat{X}_0 \) non-degenerate as \( u \to \infty \), then the limit process \( \{\hat{X}_t\}_{t \geq 0} \) is termed the forward tail process. For Breidt and Davis (1998) models of stochastic volatility \( \hat{X}_t = 0 \) almost surely for all \( t \geq 1 \), so large values are not followed by large values for these processes. In contrast, for any GARCH\((p, q)\) process at least one \( \hat{X}_t \), for \( t \geq 1 \), is non-degenerate, and furthermore all elements of the tail process are non-degenerate if all the coefficients of \( \sigma_t^2 \) in expression (1) are positive.

In this paper we derive the theory for obtaining the forward tail process, the extremogram, the extremal index and cluster size distribution for any GARCH\((p, q)\) process with bounded or unbounded support for the innovations. We provide a new fast, yet accurate, Monte Carlo algorithm for the numerical evaluation of these extremal features. Our approach is quite different from that of Collamore and Mentemeier (2018), who start their Markov chains in non-extreme states and use a form of importance sampling to move into the tail region. Relative to our approach, using the tail process, this is guaranteed to be inefficient for evaluation of limiting properties, as we start in the limit. However, for evaluating non-limit tail features the Collamore and Mentemeier (2018) approach has advantages over ours, but to be useful they need our evaluation of \( \kappa \).

The paper is structured in the following way. In Sect. 2 we give the required background details for the properties of stationary GARCH\((p, q)\) processes and the theory of multivariate regular variation that is required for our methodology. In Sect. 3 we give new results for robust computational testing for stationarity and in Sect. 4 a novel particle filtering algorithm for fixed point distributions that enables us to obtain the tail index \( \kappa \) and to sample from a \((p + q)\)-dimensional extreme state of the process. In Sect. 5 we derive the tail processes for the squared and original GARCH\((p, q)\) processes. Section 6 discusses the numerical evaluation of the method, including checking for stationarity, evaluating \( \kappa \), and illustrating the rapid convergence of the particle filter algorithm. Section 7 has a study of a range of extremal dependence features of the GARCH\((p, q)\) process for a variety of parameter values. In particular, the empirical evidence from the systematic study of two sub-classes of GARCH\((p, q)\) process suggests a number of other features that remain to be proved. In Sect. 8 we discuss the likely wider impact of our methods and in the Appendix give proofs. In Supplementary Material, Laurini et al. (2022), we provide some theory on the convergence properties of our methods and extensive illustrations of their reliability.

### 2 Known properties of GARCH\((p, q)\) processes

#### 2.1 Matrix formulation

Focusing on the squared GARCH process, \( X_t^2 \), we write the process as a stochastic recurrence equation (SRE) to enable the exploitation of a range of established results (Kesten 1973). Let the \((p + q)\) vector \( Y_t \), the \((p + q) \times (p + q)\) matrix \( A_t \) and the \((p + q)\) vector \( B_t \) be

\[
Y_t = \begin{pmatrix} X_t^2, \ldots, X_{t-q+1}^2, \alpha_t, \ldots, \alpha_{2-p+1}^2 \end{pmatrix}^T, \\
A_t = \begin{pmatrix} \alpha^{(q-1)} Z_t^2 & \alpha q Z_t^2 & \beta^{(p-1)} Z_t^2 & \beta Z_t^2 \\
I_{q-1} & 0_{q-1} & 0_{(q-1) \times (p-1)} & 0_{q-1} \\
\alpha^{(q-1)} & \alpha q & \beta^{(p-1)} & \beta p \\
0_{(p-1) \times (q-1)} & 0_{p-1} & I_{p-1} & 0_{p-1} \end{pmatrix}, \\
B_t = \begin{pmatrix} \alpha q Z_t^2, 0_{q-1}, \alpha_0, 0_{p-1} \end{pmatrix}^T
\]

where the superscript \( T \) denotes vector’s transpose, \( \alpha^{(s)} = (\alpha_1, \ldots, \alpha_s) \in \mathbb{R}^s \), \( \beta^{(s)} = (\beta_1, \ldots, \beta_s) \in \mathbb{R}^s \), \( I_s \) is the identity matrix of size \( s \), \( \theta_{(r \times s)} \) is a matrix of zeros with \( r \) rows and \( s \) columns and \( 0_s \) is a length \( s \) column vector of zeros. If \( s \leq 0 \) these terms are to be interpreted as being dimensionless, hence \( ARCH(q) \) processes arise using the top left \( q \times q \) elements of the matrix. Squared GARCH\((p, q)\) processes satisfy the SRE

\[
Y_t = A_t Y_{t-1} + B_t, \quad t \in \mathbb{Z},
\]

where \( \{A_t\} \) and \( \{B_t\} \) are each i.i.d. sequences. As \( Z_t \) is a continuous variable, \( A_t \) has no rows of zeros, this avoids problems discussed by Buraczewski et al. (2016). The expression of the GARCH\((p, q)\) in the SRE formulation (3) is due to Francq and Zakoïan (2010, Section 2.2.2, p. 29). This SRE formulation is less parsimonious than that of Bougerol and Picard (1992), but has the benefit of covering all GARCH\((p, q)\) processes, unlike that of Bougerol
and Picard (1992) which excludes the \( p = q = 1 \) case that expression (3) gives, simplifying to

\[
Y_t = \left( \frac{X_t^2}{\sigma_t^2} \right), \quad A_t = \left( \begin{array}{cc} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha_1 & \beta_1 \end{array} \right), \quad B_t = \left( \begin{array}{c} \alpha_0 Z_t^2 \\ \alpha_0 \end{array} \right),
\]

(5)

where \( \alpha^{(s)} \) and \( \beta^{(s)} \) are scalar and none of the \( I_s, 0_{r \times s} \) or \( 0_t \) terms are included.

### 2.2 Strict stationarity

Kesten (1973) presents necessary and sufficient conditions for SREs of the form (4), with general matrix \( A_t \) and vector \( B_t \), to have a unique, strictly stationary solution. The conditions that are non-trivially true for \( \text{GARCH}(p, q) \) processes are that \( E(\ln^+ \| A_t \|) < \infty \) and \( E(\ln^+ \| B_t \|) < \infty \) (here \( \ln^+ x = \ln x, \) if \( x \geq 1 \) and 0 otherwise), and that the top Lyapunov exponent \( \gamma \), given by

\[
\gamma = \lim_{t \to \infty} \frac{1}{t} E(\ln \| A_t A_{t-1} \cdots A_1 \|),
\]

(6)

has the property that \( \gamma < 0 \). This result holds for all norms. Here and subsequently we use the \( L_1 \) norm, so for a matrix \( A \) with \( a_{i,j} \) being the \( (i,j) \)th element, \( \| A \| = \sum_{i,j} |a_{i,j}| \) and for a vector \( B \) with \( b_i \) being the \( (i) \)th element, \( \| B \| = \sum_i |b_i| \).

For the squared \( \text{GARCH}(p, q) \) moment conditions on \( A_t \) and \( B_t \), always hold. To illustrate this for \( A_t \), first consider the random variable \( K_t := \phi Z_t^2 + \phi + p + q \), where \( \phi > 0 \) is defined by (2). Then \( K_t > 1, 0 < \ln(K_t) < K_t \) and \( E(K_t) < \infty \) as \( E(Z_t^2) = 1 \). As \( 0 \leq \| A_t \| = \phi(Z_t^2 + 1) + p + q - 2 < K_t \), we have \( 0 \leq E(\ln^+ \| A_t \|) \leq E(\ln^+ K_t) = E(K_t) < \infty \). So for \( \text{GARCH}(p, q) \) processes \( X_t^2 \) and \( X_t \) to be strictly stationary it is necessary and sufficient that \( \gamma < 0 \). Expression (6) is not an ideal starting point for evaluating \( \gamma \). Instead we also have that

\[
\gamma_t = \frac{1}{t} \ln \| A_t A_{t-1} \cdots A_1 \| \quad \text{and} \quad \gamma_t \xrightarrow{a.s.} \gamma, \quad \text{as} \ t \to \infty,
\]

(7)

where \( a.s. \) denotes almost sure convergence, (Francq and Zakoian 2010, Theorem 2.3—p. 30). It would appear that a relatively simple simulation can be performed to obtain a reliable estimate of \( \gamma \) via expression (7). However, in Sects. 3.3 and 6.2 we show this is far from the case and a mix of careful asymptotic approximation analysis and numerical evaluation is required to evaluate \( \gamma \).

In some cases we do not need to evaluate \( \gamma \) to find if the process is strictly stationary, e.g., \( \text{GARCH}(p, q) \) processes are always strictly stationary when \( \phi \leq 1 \); this includes all IGARCH\((p, q)\) processes. Also \( \sum_{j=1}^p \beta_j < 1 \) is necessary but not sufficient for strict stationarity. So, numerical evaluation of \( \gamma \) is required whenever \( \phi > 1 \) and \( \sum_{j=1}^p \beta_j < 1 \).

### 2.3 Multivariate regular variation results

Basrak et al. (2002) and Collamore and Mentemeier (2018) show that there exists a unique stationary solution to the SRE (4) and that this solution exhibits a multivariate regular variation property, i.e., for any \( t, \) any vector norm \( \| \cdot \| \) and all \( r > 0 \),

\[
\frac{\Pr(\| Y_t \| > r, \| Y_t \| \in \cdot)}{\Pr(\| Y_t \| > x)} \xrightarrow{v} r^{-\kappa} \Pr(\hat{\Theta}_t \in \cdot),
\]

(8)

as \( x \to \infty \), where \( \xrightarrow{v} \) denotes vague convergence, \( \kappa > 0 \), and \( \hat{\Theta}_t \) is a \((p + q)\)-dimensional random vector on the unit sphere (with respect to a norm \( \| \cdot \| \)) defined by \( \mathbb{S}^{p+q} \subset \mathbb{R}^{p+q} \).

If condition (8) holds then \( Y_t \) is said to exhibit multivariate regular variation with index \( \kappa \) and the distribution of \( \hat{\Theta}_t \) is termed the spectral measure of the vector \( Y_t \) (Resnick 1987). A consequence of property (8) for \( \text{GARCH}(p, q) \) processes is that all the marginal variables of \( Y_t \) have regularly varying tails with index \( \kappa > 0 \), i.e., for \( r \geq 1 \) and all \( t \)

\[
\Pr(X_t^2 > rx | X_t^2 > x) \to r^{-\kappa},
\]

and

\[
\Pr(\sigma_t^2 > rx | \sigma_t^2 > x) \to r^{-\kappa}, \quad \text{as} \ x \to \infty.
\]

It is insightful to consider a slightly rearranged version of limit (8), and with the \( L_1 \) vector norm. We define radial, \( R_t \), and angular (two variants \( \Theta_t \) and \( \Theta_t^- \)) random variables by

\[
R_t = \| Y_t \| = X_t^2 + \ldots + X_{t-q+1}^2 + \sigma_t^2 + \ldots + \sigma_{t-p+1}^2, \quad \Theta_t = Y_t/\| Y_t \|
\]

\[
= (X_t^2, \ldots, X_{t-q+1}^2, \sigma_t^2, \ldots, \sigma_{t-p+1}^2)/R_t, \quad \Theta_t^- = (X_t^2, \ldots, X_{t-q+1}^2, \sigma_t^2, \ldots, \sigma_{t-p}^2)/R_t,
\]

(9)

with \( \mathbb{S}^{p+q} \) now being the \((p + q)\) dimensional unit simplex.

For \( \text{GARCH}(p, q) \) processes, with \( p \geq 1 \), we have two angular variables as the \( p + q \) dimensional variable \( \Theta_t \), has redundancy in its final dimension as \( \| \Theta_t \| = 1 \), and so for studying the distribution of angular variables it is simpler to work with the \( p + q - 1 \) dimensional variable \( \Theta_t^- \), which is related to \( \Theta_t \) by \( \Theta_t^- = (\Theta_t^-L, 1 - \| \Theta_t^- \|) \) where \( \Theta_t^-L \) is \( \Theta_t \) without its last component. We use this \( W^- \) notation to create a \((p + q - 1)\) dimensional vector from any \((p + q)\) dimensional vector \( W \) on the simplex \( \mathbb{S}^{p+q} \) throughout. Furthermore, for \( w \in \mathbb{R}^{p+q-1} \), we define

\[
H_{\Theta_t^-}(w) = \Pr(\Theta_t^- \leq w),
\]

(10)

with vector algebra, here and elsewhere, interpreted as being componentwise. We will denote the limit random variables,
that arise in limit (8), for \((R_t, \Theta_t, \Theta^-_t)\) by \((\hat{R}_t, \hat{\Theta}_t, \hat{\Theta}^-_t)\), with the distribution function of \(\hat{\Theta}^-_t\), denoted by \(H^-_{\hat{\Theta}^c_t}\), defined similarly to distribution (10). Then, for \(r \geq 1\), as \(x \to \infty\), the limit (8) becomes,

\[
\Pr(R_t > rX, \Theta^-_t \leq w \mid R_t > x) \xrightarrow{\Delta} \Pr(\hat{R}_t > r) \Pr(\hat{\Theta}^-_t \leq w) = r^{-\kappa} H^-_{\hat{\Theta}^c_t}(w).
\] (11)

From the first expression for the asymptotic form in limit (11) we see that the radial variable \(R_t\) and the angular variables \(\Theta^-_t\) become asymptotically independent, as the radial variable \(R_t\) grows due to \(x \to \infty\), i.e., the variables \(\hat{R}_t\) and \(\hat{\Theta}^-_t\) are independent. The second term in this limit shows that \(\hat{R}_t\) is a Markov random variable with tail index \(\kappa\), i.e.,

\[
\Pr(\hat{R}_t > r) = r^{-\kappa} \quad \text{for} \ r \geq 1.
\] (12)

Basrak and Segers (2009) and Janssen (2010) identify that there is additional structure imposed on \(H^-_{\hat{\Theta}^c_t}\) and \(\kappa\) respectively by the GARCH\((p, q)\) process. We discuss these restrictions below.

For the GARCH\((1, 1)\) process, Laurini and Tawn (2012) provided an expression for the spectral measure, for a different description of the angular variable to that used here. For the choice of the angular variable (9), for all \(t\), their result translates to a univariate \(\hat{\Theta}^-_t\), with distribution

\[
H^-_{\hat{\Theta}^c_t}(w) = \frac{2}{E(\|Z\|^{2\kappa})} \int_0^{\left(\frac{w}{\|w\|}\right)^{1/2}} (1 + s^2)^{-\kappa} F_Z(ds),
\] (13)

where \(F_Z\) is the distribution function of \(Z_t\). When \(\max(p, q) \geq 2\), through use of the multivariate regular variation structure, Basrak and Segers (2009, Propositions 3.3, 5.1) show that

\[
E(\|A \hat{\Theta}_t\|^\kappa) = 1,
\] (14)

where \(0 \leq w \leq 1\) and \(A\) is i.i.d. to \(A_t\), and uniquely \(\Pr(\hat{\Theta}^-_t \in \cdot) = E(\|A \hat{\Theta}_t\|^\kappa; A \hat{\Theta}_t / \|A \hat{\Theta}_t\| \in \cdot),\) where the notation \(E(X; Y) := E(X \mathbf{1}(Y))\) and \(\mathbf{1}(Y)\) is the indicator of the event \(Y\). Thus

\[
H^-_{\hat{\Theta}^c_t}(w) = E(\|A \hat{\Theta}_t\|^\kappa; (A \hat{\Theta}_t / \|A \hat{\Theta}_t\|)^- \leq w).
\] (15)

Basrak and Segers (2009, p. 1075) propose an approach to simulate from \(H^-_{\hat{\Theta}^c_t}(w)\) for a SRE of the form (4), with the required distribution \(H^-_{\hat{\Theta}^c_t}\) being the invariant distribution of a Markov chain and MCMC methods used for its evaluation. However, this method cannot be used for any GARCH\((p, q)\) process for the following reasons. Firstly, they make an assumption that \(A_t\) is bounded, which excludes the possibility of \(Z_t\) being, for example, Gaussian or \(t\), distributed. Much more critically though, Basrak and Segers (2011) note that their proof was flawed and their results only hold under very specific conditions on the matrix \(A_t\) in the SRE framework (4), which excludes all GARCH\((p, q)\) processes. Our approach in Sect. 4 overcomes these restrictions.

### 2.4 Janssen's approach to determine \(\kappa\)

Next focus on how \(\kappa\) has been determined. Following methods of Kesten (1973), Basrak et al. (2002) showed that there exists a \(\kappa > 0\) which is the unique positive solution of the equation

\[
\lim_{t \to \infty} \frac{1}{t} \ln E(\|A_t A_{t-1} \cdots A_1\|^\kappa) = 0.
\] (16)

Collamore and Mentemeier (2018) found the same expression for the marginal tail index for more general stochastic recurrence relations. For the GARCH\((1, 1)\) process Mikosch and Stărică (2000) show that \(\kappa\) is simple to evaluate using expression (16). Specifically, taking \(A_t\) as in expression (5) then \(A_t A_{t-1} \cdots A_1 = A_t \prod_{k=1}^{\kappa} (\kappa 1_k Z_t^2 + \beta_1)\), from which it simply follows that expression (16) holds and \(\kappa\) satisfies \(E(\kappa \beta_1)^{1/\kappa} = 1\). Setting \(\beta_1 = 0\) for the GARCH\((1, 1)\) process gives the same result for \(\kappa\) derived by de Haan et al. (1989) for the ARCH\((1)\) process. For general GARCH\((p, q)\) processes no such simplification of expression (16) exists. So it is natural to try to evaluate \(\kappa\) by solution of the equation (16), but such an approach is impossible due to numerical instabilities.

The only existing feasible method to evaluate \(\kappa\) has been proposed by Janssen (2010, Proposition 4.3.1), which exploits Kesten (1973, Proof of Theorem 3). With \(A\) specified in (4) the conditions required for the results of Kesten (1973) apply and the equality

\[
\int_{S_p^+q} E(\|A w\|^k \left(\frac{A w}{\|A w\|}\right) H_k(ds) = \rho_k \int_{S_p^+q} g(w) H_k(dw),
\] (17)

holds for all continuous functions \(g\), where \(H_k\) is a probability measure on \(S_{p+q}\) (see step 1 in the proof of Theorem 3 in Kesten 1973), and where \(\rho_k > 0\). For all pairs \((\rho_k, H_k)\) that satisfy equality (17), for any given \(k \in (0, \infty)\) then \(\rho_k\) is determined solely by \(k\). The special case of \(g \equiv 1\) in (17) gives the simplification

\[
\int_{S_p^+q} E(\|A w\|^k) H_k(dw) = \rho_k \int_{S_p^+q} H_k(dw) = \rho_k.
\] (18)
From condition (14) we know that the $k$-th moment of $\|A \hat{\Theta}_t\|$ is equal to 1, where $\hat{\Theta}_t \sim H_{\Theta_t}$ on the space $\mathbb{S}^{p+d}$. It follows from property (18) that $p_k = 1$ when $k = \kappa$. Kesten (1973) and Janssen (2010) show that there is only one solution to equation (14), so $\kappa$ is the unique solution of
\[
\int_{\mathbb{S}^{p+d}} E \left( \|A w\|^k \right) H_k (dw) = 1, \tag{19}
\]
and that the unit measure $H_k$ must be $H_{\Theta_t}$. Thus if we can find, or simulate from, $H_{\Theta_t}$ we can find $\kappa$. To evaluate $\kappa$ all that is required is to define a class of unit measures $H_k$, over $k \in (0, \infty)$, which contains within it as an interior point $H_{\Theta_t}$, and then vary $k$ until property (19) is found.

By adapting the algorithm of Basrak and Segers (2009), Janssen (2010) proposes a valid algorithm to simulate from a class of functions $H_k$, which enables the evaluation of $\kappa$ and $H_{\Theta_t}$. Critically this only applies when the innovations $Z_t$ have bounded support. In Sect. 4 we describe an algorithm that applies even if $Z_t$ is unbounded and which has substantial computational efficiency gains for calculating $\kappa$, compared to the algorithm of Janssen (2010), even when $Z_t$ is bounded.

### 3 Alternative formulations for assessing strict stationarity of GARCH processes

#### 3.1 Preliminaries

Let $\phi = (\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p) = (\phi_1, \ldots, \phi_{p+q})$, with $\phi_i \geq 0$ for all $i = 1, \ldots, p + q$. From (2) it follows that $\phi = \sum_{i=1}^{p+q} \phi_i$. Also let $\phi^{(M)} = \max\{\phi_i; i = 1, \ldots, p + q\}$ so $0 < \phi^{(M)} \leq \phi$ for all GARCH($p, q$) processes as they require that $\alpha_q > 0$, even for ARCH($q$) processes. As we use the $L_1$ norm of the matrix, $\|A_t\| = \phi(Z_t^2 + 1) + p + q - 2$, so it is linear in $Z_t^2$.

Results (7) and (16), for $\gamma$ and $\kappa$ respectively, involve the behaviour of the $\|A_t A_{t-1} \cdots A_1\|$ as $t \to \infty$. As this term tends to zero we need to capture its dominant behaviour to stabilise it. The approach we will take uses products of simple scalar summaries of the individual matrices. One possibility is the trace of $A_t$, with trace($A_t$) $= \alpha_1 Z_t^2 + \beta_1$ for all $t, p$ and $q$, as the trace of a square matrix is the sum of all its eigenvalues. As we can have $\alpha_1 = \beta_1 = 0$ when $\min(p, q) \geq 2$, we instead focus on the related quantities
\[
\lambda_t^* = \alpha_q Z_t^2 + \beta_p \quad \text{and} \quad \lambda^* = \alpha_q Z^2 + \beta_p, \tag{20}
\]
where $Z$ is i.i.d. from $\{Z_t\}$. Here $\alpha_q > 0$ for all GARCH($p, q$) processes so $\lambda_t^* > \beta_p > 0$, but for ARCH($q$) processes $\beta_p = 0$ so $\lambda_t^*$ can be arbitrarily close to 0. Another possible summary of $A_t$ is $\lambda_t$, the largest magnitude eigenvalue of all of its eigenvalues, and similarly we use the term $\lambda$ for this when $Z_t^2$ is replaced by $Z^2$. For $A_t$, given by representation (3), it is guaranteed that $\lambda_t$ is simple and exceeds all other eigenvalues in absolute value; see Seneta (1981, Theorem 1.1). In general $\lambda_t$ does not have a simple closed form expression.

At first sight the two proposals for stabilising the matrix product seems quite different. However, their properties are highly related. For example in the GARCH(1,1) case $\lambda_t = \lambda_t^* = \alpha_1 Z_t^2 + \beta_1$, which is an important term as the property $E[(\alpha_1 Z_t^2 + \beta_1)^k] = 1$ determines $\kappa > 0$ for this process. So both proposals form natural extensions to the case with $p \neq 1$ and/or $q \neq 1$. In the only other case where we have found a closed form solution for $\lambda_t$ similar relationships hold. Specifically, for a GARCH($p, q$) processes, i.e., $q = p$, when $\alpha_i = \beta_i = 0$ for $i = 1, \ldots, p - 1$ we have that $\lambda_t = (\alpha_p Z_t^2 + \beta_p)^{1/p} = (\lambda_t^*)^{1/p}$. Lemma 1 shows that $\lambda_t$ and $\lambda_t^*$ have similar moment properties and related growth rates when multiplied over different $t$.

**Lemma 1** Suppose that $\{A_t\}$ are given by (3) where $Z_t$ are as defined by (1), $\lambda_t^*$ is defined by (20), and let $\lambda_t$ be the largest magnitude eigenvalue of the matrix $A_t$. Then for all GARCH($p, q$) processes with $p > 0$
\[
E(\ln \lambda_t^*) < \infty \quad \text{and} \quad E(\ln \lambda_t) < \infty, \tag{21}
\]
and
\[
\mu_t := \frac{1}{t} \ln \left( \prod_{i=1}^{t} \frac{\lambda_i^*}{\lambda_i} \right) \xrightarrow{a.s.} \mu \quad \text{as} \quad t \to \infty
\]
to a finite constant $\mu$ where $\mu = E(\ln \lambda^*) - E(\ln \lambda)$, with $-\infty < \mu < \infty$. When $p = 0$, i.e., an ARCH($q$) process, all convergences only hold if additionally we have that $E(\ln |Z_t|) < \infty$.

Note that $E(\ln |Z|) < \infty$ holds for all Gaussian and Student-$t$ distributions as the additional restriction that the condition implies is that the density of $Z$ is bounded at zero. In all examples we considered we found that $\mu \leq 0$.

There are pros and cons of the two approaches. As we have seen, $\lambda_t^*$ is analytically simple and so it is easy to derive closed form results and interpret, but we find in Sect. 6 it has limitations in terms of the numerical stability of its products and in Sect. 3.2 show that different derivations for GARCH and ARCH cases are required. In contrast, although $\lambda_t$ in general lacks a simple closed form, which makes analytical derivations more difficult as seen by the proof of Lemma 1, our numerical studies find that obtaining $\lambda_t$ is simple, fast and that there are no issues of numerical instability with its products. These properties influence our strategy. We first work with $\lambda_t^*$ to develop bounds (for GARCH in Sect. 3.2 and ARCH in Appendix A) and limits, then derive the equivalent results for $\lambda_t$. 
3.2 Bounds on matrix product norms for GARCH\((p, q)\) with \(p > 0\)

For a GARCH\((p, q)\) process \((p > 0)\), consider the \((p+q) \times (p+q)\) matrix \(A_t \cdots A_1\) with \((i, j)\)th entry \(a_{i,j}^{(t)}\). Given the form of \(A_t\) in expression (3), we have that \(a_{i,j}^{(t)} \geq 0\) for all \(t, i,\) and \(j\).

\[
\|A_{t+1} \cdots A_1\| = (Z_{t+1}^2 + 1) \left[ \sum_{i=1}^{q} \alpha_t \sum_{j=1}^{p+q} a_{i,j}^{(t)} \right] + \sum_{i=1}^{p} \beta_t \sum_{j=1}^{p+q} a_{i,j}^{(t)} + \sum_{i=1, i \neq q}^{p+q-1} \sum_{j=1}^{p+q} a_{i,j}^{(t)}.
\]  

(22)

We use expression (22) to develop bounds on \(\|A_t \cdots A_1\|\) that hold for all GARCH\((p, q)\) processes. As \(\phi_t \leq q_M(\cdot)\) for all \(i = 1, \ldots, p + q\) the term in the square brackets is bounded above by \(\phi(M)\|A_t \cdots A_1\|\) and as \(a_{i,j}^{(t)} \geq 0\) the last term is bounded above by \(\|A_t \cdots A_1\|\). By applying this combined bound iteratively, starting from \(t\), we have

\[
\|A_t \cdots A_1\| \leq \left( \prod_{i=2}^{t} \left( \phi(M) \left( Z_i^2 + 1 \right) + 1 \right) \right) \|A_1\|.
\]  

(23)

The lower bound follows from expression (22) by dropping all terms involving \(a_i, i = 1, \ldots, q - 1\) and \(\beta_t, j = 1, \ldots, p - 1\), taking a common lower bound on the coefficients of \(a_{i,j}^{(t)}\) terms for all \(i, j\), and setting \(c_{p,q} := \min(\alpha_t, \beta_t) > 0\), to give

\[
\|A_{t+1} \cdots A_1\| \geq \min(\{Z_i^2 + 1\}c_{p,q}, 1)\|A_t \cdots A_1\| \geq \left( \prod_{i=2}^{t} \min(\{Z_i^2 + 1\} \min(\alpha_t, \beta_t), 1) \right) \|A_1\|.
\]

The value of these bounds is not obvious at first sight given that they are stochastic and involve \(t\) products (where we are interested in \(t \to \infty\)) over unbounded variables. As the norm of the product \(\|A_t \cdots A_1\|\) is tending to zero it seems sensible to first normalise the size of the individual terms in the product to get more helpful bounds. We do this by scaling each matrix \(A_t\) by \(\lambda_t^*\). Let

\[
\Delta_t^* = \prod_{i=1}^{t} (A_{t+1-i}/\lambda_{t+1-i})^\prime \text{ and } \eta_t^* = \frac{1}{t} \ln \|\Delta_t^*\|.
\]  

(24)

then, using upper bound (23), gives that

\[
\eta_t^* \leq \frac{1}{t} \sum_{i=2}^{t} \ln \left( \left[ \phi(M) \left( Z_i^2 + 1 \right) + 1 \right] \right) + \frac{1}{t} \ln \left( \|A_1\|/\lambda_1^* \right).
\]

Using the property that \(\lambda_t^* \geq 0\) for any GARCH\((p, q)\) process, Lemma 2 in Appendix B for each term in expression (25), and that log is an increasing function, we have

\[
\eta_t^* \leq \frac{1}{t} \left[ \ln \left( \max \left( \phi(M) + 1, \phi(M) - 1 \right) \right) \right] + \frac{1}{t} \ln \left( \max \left( \frac{\phi(M) + 1}{\phi(M)}, \frac{\phi(M) - 1}{\phi(M)} \right) \right).
\]

Similarly, the lower bound for the norm of the product together with Lemma 2 gives

\[
\eta_t^* \geq \frac{1}{t} \left[ \ln \left( \min \left( \frac{\phi(M) + 1}{\phi(M)}, \frac{\phi(M) - 1}{\phi(M)} \right) \right) \right] + \frac{1}{t} \ln \left( \min \left( \frac{\phi(M) + 1}{\phi(M)}, \frac{\phi(M) - 1}{\phi(M)} \right) \right).
\]

The final term in each inequality bound converges almost surely to \(0\) as \(t \to \infty\). The second term of \(\min(\{c_{p,q}(Z_i^2 + 1), 1\})\) occurs with probability \(\Pr(Z_i^2 > (1 - c_{p,q})/c_{p,q})\) and, as \(t \to \infty\), the sample mean of these terms converge to \(-\ln(\alpha_q Z^2 + \beta_p) / Z^2 > (1 - c_{p,q})/c_{p,q}\), which is a finite constant as \(E(\ln|Z|) < \infty\). Similarly the sample mean of the first terms in the minimum converges to \(\ln(\alpha_q Z^2 + \beta_p) / (\alpha_q Z^2 + \beta_p)\), where the random variable concerned has finite bounds. These convergences follow by the strong law of large numbers and as \(E(\ln|\lambda_t^*|) < \infty\). If \(\lim_{t \to \infty} \eta_t^*\) exists, then this limit is in the finite interval \([c_L, c_U]\), where

\[
c_L = E \left( \ln \left( \min(\{c_{p,q}(Z^2 + 1), 1\})/(\alpha_q Z^2 + \beta_p) \right) \right), \quad c_U = \min((\phi(M) + 1)/\min(\alpha_q, \beta_p)).
\]

(26)

3.3 New formulations for \(\gamma\) in terms of \(\lambda^*\) and \(\lambda\)

**Theorem 1** Suppose that \(\{A_t\}\) are given by (3) with \(Z_t\) defined by (1), \(\lambda_t^*\) and \(\lambda^*\) are defined by (20), the limit (6) exists, and \(\eta_t^*\) is defined by (24). It follows that as \(t \to \infty\), \(\eta_t^* \xrightarrow{a.s.} \gamma = E(\ln \lambda^*) := \eta^*, \) for \(p > 0\), and for \(p = 0\) this holds when \(E(\ln|Z|) < \infty\), with the constant limit \(\eta^* \in [c_L, c_U]\), with these bounds defined by (26) and (35) for GARCH and ARCH processes respectively. Hence, for both processes the Lyapunov exponent satisfies \(\gamma = E(\ln \lambda^*) + \eta^*, \) where \(\eta^*\) is the almost sure limit of \(\eta_t^*\).

At first sight the value of Theorem 1 seems limited, as it simply expresses a limit \(\gamma\) as function of two terms. Whilst one is a simple expectation, the other, \(\eta^*\), is also a limit. The benefits though are for numerical evaluation and interpretation. Under definition (7), \(\gamma\) is the limit of \(\gamma_t\) as \(t \to \infty\).
where both the numerator and denominator diverge to infinity, with the numerator a product of \( t \) unbounded stochastic terms and is numerically unstable to evaluate as a limit. In contrast there are finite bounds for \( \eta^* \) for all \( t \).

An alternative version of Theorem 1 applies when the largest magnitude eigenvalue \( \lambda_1 \) of the set of eigenvalues of \( A_t \) replaces \( \lambda^* \). Similarly to the definition of \( \eta^* \), define

\[
\Delta_t = \prod_{i=1}^{t} \left( \frac{A_{t+1-i}}{\lambda_1 \cdot A_{t+1-i}} \right), \quad \eta_t = \frac{1}{t} \ln \| \Delta_t \| \quad \text{and} \quad \eta = \lim_{t \to \infty} \eta_t. \tag{27}
\]

Sect. 6.2 shows that \( \Delta_t \) and \( \eta_t \) are numerically stable to evaluate for large \( t \), unlike \( \Delta^*_t \) and \( \eta^*_t \).

**Theorem 2** Suppose that \( \{A_t\} \) are given by (3) with \( Z_t \) defined by (1) and the limit \( \gamma \) of expression (6) exists. Let \( \lambda^* \) be defined by (20), \( \lambda \) be the largest magnitude eigenvalue of a matrix \( A \), which is identically distributed to \( A_t \), and let both \( \eta_t \) and \( \eta_t \) be defined by (27). We then have that \( \eta_t \xrightarrow{a.s.} \eta \) as \( t \to \infty \) where \( \eta \) is a finite constant, \( \eta \equiv \eta^* + \mu \), where \( \mu = E(\ln \lambda^*) - E(\ln \lambda) \), with \( -\infty < \mu < \infty \) and \( \gamma \equiv E(\ln \lambda) + \eta \).

To assess Theorems 1 and 2 in terms of what is already known we first compare with the GARCH(1,1) process which is known to satisfy \( A_t \cdot A_t = A_t \prod_{i=1}^{t} (\alpha_1 Z_t^2 + \beta_1) = A_t \prod_{i=1}^{t} \lambda_i^* \), and \( \lambda_i^* = \lambda_i \). So

\[
\nu_t = \frac{1}{t} \ln \| \prod_{i=1}^{t} A_{t+1-i} \| = \frac{1}{t} \ln \| A_t \| + \frac{1}{t} \sum_{i=1}^{t-1} \ln \lambda_i^* \xrightarrow{a.s.} E(\ln \lambda^*). \]

Hence Mikosch and Stărică (2000) conclude that \( \gamma = E(\ln(\alpha_1 Z^2 + \beta_1)) = E(\ln \lambda^*) \). Theorems 1 and 2 agree with this result, which implies that \( \eta^* = \eta = 0 \) for all GARCH(1,1) processes.

### 4 Evaluating the spectral measure and the tail index

This section gives the details of our algorithm for sampling from the limit distribution \( H_{\hat{\Theta}_0} \) and then uses this algorithm repeatedly to find \( \kappa \). The algorithm requires no assumptions on the support for \( Z_t \). Throughout we take \( t = 0 \) as we start the tail process at that time in Sect. 5.1. We will first assume that \( \kappa \) is known and present Algorithm 1 for generating from \( H_{\hat{\Theta}_0} \) and then discuss the case when \( \kappa \) is unknown.

To simulate from the spectral measure \( H_{\hat{\Theta}_0} \), defined via (15), our approach is to introduce a series of distributions related via a recursion, and whose invariant distribution is \( H_{\hat{\Theta}_0} \). We will then use sequential importance sampling (Doucet et al. 2000, ) to generate approximate samples from these distributions at consecutive iterations. We simulate this process until convergence, and then take the samples after this time as draws from \( H_{\hat{\Theta}_0} \). The recursion can be viewed as that for distributions associated with Feynman-Kac models, and our algorithm is similar to the use of sequential Monte Carlo for sampling from Feynman-Kac distribution (e.g. Del Moral and Miclo 2000) and fixed point distributions (e.g. Del Moral and Miclo 2003).

Denote the series of random variables whose distribution we are simulating from by \( \hat{\Theta}_s \), for iteration \( s \geq 0 \). The recursion relating the distributions of these random variables for \( s \geq 1 \) is

\[
Pr(\hat{\Theta}_s \in \cdot) = \frac{E(\|A\hat{\Theta}_s-1\|^\kappa; A\hat{\Theta}_s-1/\|A\hat{\Theta}_s-1\| \in \cdot)}{E(\|A\hat{\Theta}_0\|^\kappa) \cdot E(\|A\hat{\Theta}_0\|^\kappa)}, \tag{28}
\]

where as above \( E(X; Y) := E(X|Y) \). By construction, the invariant distribution of this process is \( H_{\hat{\Theta}_0} \), since if \( \hat{\Theta}_s-1 \) is drawn from \( H_{\hat{\Theta}_0} \), then the right-hand side of (28) is equal to

\[
\frac{E(\|A\hat{\Theta}_0\|^\kappa; A\hat{\Theta}_0/\|A\hat{\Theta}_0\| \in \cdot)}{E(\|A\hat{\Theta}_0\|^\kappa)}.
\]

As \( E(\|A\hat{\Theta}_0\|^\kappa) = 1 \), this distribution is equal to the definition of \( H_{\hat{\Theta}_0} \) given by expression (15).

Algorithm 1 uses importance sampling to generate a sample from \( Pr(\hat{\Theta}_s \in \cdot) \) based on a sample from \( Pr(\hat{\Theta}_s \in \cdot) \). This involves first simulating a value for \( \hat{\Theta}_s \) via \( \hat{\Theta}_s = A\hat{\Theta}_{s-1}/\|A\hat{\Theta}_{s-1}\| \), and assigning this value a weight proportional to \( \|A\hat{\Theta}_{s-1}\|^\kappa \). Thus we can use sequential importance sampling to generate samples of \( \hat{\Theta}_s \) for \( s \geq 1 \) from an initial sample of \( \hat{\Theta}_0 \). The assessment of the algorithm’s convergence is discussed in Sect. 6.3.

Our approach is similar to the MCMC algorithm Basrak and Segers (2009) propose for simulating from \( H_{\hat{\Theta}_0} \). Both algorithms simulate from the same Markov process. The difference is that they simulate a single realisation of a path of the process, and simulate the dynamics at each time point using rejection sampling. By comparison we use sequential importance sampling to simulate a sample of values from expression (28) at each time-point. The main advantage of our approach is that we do not need the distribution of \( Z_t \) to be bounded. Algorithm 1 is also computationally more efficient than that of Basrak and Segers (2009) in situations where a bound exists, but is large, as their probability of rejection is close to 1.

As it stands neither Algorithm 1, nor the algorithm Basrak and Segers (2009), have guaranteed unique convergence.
to the true limit. For Feynman-Kac models, which include $H_{\Theta_0}$, Del Moral and Guionnet (2001) provide conditions under which the convergence of Algorithm 1 is unique. Laurini et al. (2022) show that these conditions amount to a subtle, but relatively weak, condition on the mixing of the process, but we have not been able to prove that condition is satisfied by all GARCH processes. Instead, Sect. 6.3 and Laurini et al. (2022) presents extensive empirical evidence to show Algorithm 1 converges to the correct limit each time it has been used, with the convergence achieved after very few steps whatever the initial distribution of $\Theta_0$. The results obtained there are in very strong agreement with empirical extremal properties of the GARCH($p$, $q$) process when estimated using very long simulations of the true process.

Now consider the situation when $\kappa$ is unknown. For a trial value of $k$ (for $\kappa$), apply Algorithm 1 until convergence, giving a sample of weighted particles $\{J^{(j)}(k), m^{(j)}(k)\}_{j=1}^J$ after the chain is deemed to have converged. Using these particles we approximate the expectation $E(\|A\Theta\|^k)$ to give $\tilde{\rho}_k$, by the Monte Carlo approximation of $\rho_k$, as

$$\tilde{\rho}_k = \int_{\mathbb{R}} \sum_{j=1}^J \|A\Theta_{(j)}^{(k)}(k)\|^k \frac{m^{(j)}(k)}{\sum_{n=1}^J m^{(n)}(k)} F_Z(dz),$$  (29)

where $F_Z$ is the distribution function of $Z$. We repeat this evaluation over $k > 0$ until we find the unique value of $k$ which gives $\tilde{\rho}_k = 1$. This value is $k = \kappa$. See Sect. 6.4 for how we solve this equation whilst accounting for the Monte Carlo noise in $\tilde{\rho}_k$.

**Algorithm 1: Sampling from $H_{\Theta_0}(w)$**

1. Generate a sample of $\Theta_0$ from a distribution of $\mathbb{S}^{p+q}$. Set $s = 1$.
2. Generate $J$ independent copies of $A$, denote these as $A^{(j)}_{s-1}$ for $j = 1, \ldots, J$.
3. Generate $J$ equally weighted particles at time $s - 1$ by sampling independently from our approximation to the distribution of $\Theta_{s-1}$. Denote these particles as $\Theta_{s-1}^{(j)}$ for $j = 1, \ldots, J$.
4. Generate $J$ particles at time $s$,

$$\Theta^{(s)}_{(j)} = A^{(j)}_{s-1} \Theta_{s-1}^{(j)} / \|A^{(j)}_{s-1} \Theta_{s-1}^{(j)}\|, \quad j = 1, \ldots, J,$$

5. Assign each particle a weight, $m^{(s)}_{(j)} = \|A^{(j)}_{s-1} \Theta_{s-1}^{(j)}\|^p$ for $j = 1, \ldots, J$, and normalise these via

$$m^{(s)}_{(j)} = \frac{m^{(s)}_{(j)}}{\sum_{j=1}^J m^{(s)}_{(j)}}.$$

The weighted particles, $\{\Theta_{s}^{(j)}, m_{s}^{(j)}\}_{j=1}^J$ gives our approximation to the distribution of $\Theta_s$.
6. If converged to stationarity, output the set of weighted particles. Otherwise set $s = s + 1$ and go to step 2.

To use Algorithm 1 we need to be able to sample from a suitable random variable $\Theta_0$ on $\mathbb{S}^{p+q}$. Ideally the distribution of this random variable should be as close as possible to the target limit distribution function $H_{\Theta_0}(w)$, so that the rate of convergence of $\Theta_s \to^d \Theta_0$ as $s \to \infty$ is maximised. From limit (11) we have that $\text{Pr}(\Theta_0 \leq w \mid R_0 > x) \to H_{\Theta_0}(w)$, as $x \to \infty$. So for large enough $x$, i.e., $x \geq u$ for some high threshold $u$, if we treat this limiting representation as an equality, this gives us an initial estimate $H_{\Theta_0}^{(0)}$ of $H_{\Theta_0}$.

First, select $u$ as a high quantile of the sample of $R_t$ values, such that limit property (11) appears to be well represented, i.e., radial values appear to have a Pareto tail and radial and angular values appear to be independently distributed.

To obtain $H_{\Theta_0}^{(0)}$ we generate a sample of length $n$ from the GARCH($p$, $q$) process and take the empirical distribution of simulated values of $\Theta$, given that $R_t > u$ after a burn in period of $n_b$ i.e.,

$$H_{\Theta_0}^{(0)}(w) = \frac{\sum_{j=n_b+1}^n 1(R_j > u, \Theta_j \leq w)}{\sum_{j=n_b+1}^n 1(R_j > u)}.$$

As initial particles for Algorithm 1 we use all the realisations of $\Theta_0$, given that $R_t > u$, for $t = 1, \ldots, n$. We take $n = 1.1 \times 10^7$, with $u$ being the 99.99% quantile of $R_t$ giving $J = 10^3$ particles, each with equal weight $J^{-1}$.

### 5 Tail processes and associated properties for GARCH($p$, $q$)

#### 5.1 Generation of different tail processes for GARCH($p$, $q$)

The tail process $\{\hat{X}_t\}_{t \geq 0}$ is evaluated using Algorithm 2. There are two stages to the algorithm, initialisation and propagation. However, in most cases we are interested into the extremal properties of both the upper and lower tails for $X_t$, and denote their respective tail processes by $\{\hat{X}^U_t\}_{t \geq 0}$ and $\{\hat{X}^L_t\}_{t \geq 0}$, defined in expression (32). These additional tail processes are evaluated using Algorithm 3, which is driven by outputs from Algorithm 2 and exploits the GARCH($p$, $q$) formulation (1). In the special case where $p = 0$, i.e., the ARCH($q$) process, an adaption to Algorithm 3 is required.

For the initialisation stage for Algorithm 2, we consider the behaviour of the squared GARCH process conditional on it being in an extreme state at time $t = 0$, so we require that $\hat{X}_0 > 1$. Focusing on limit (11) when $t = 0$, we have that the radial and angular limit variables are independent with distributions (12) and (15) respectively. We denote these limit variables by $\hat{R}_0$ and $\hat{\Theta}_0 = (\hat{\phi}_0^{(0)}, \ldots, \hat{\phi}_0^{(p+q)}) \sim H_{\Theta_0}$, with $\hat{R}_0 > 1$ and $\hat{\Theta}_0 \in \mathbb{S}^{p+q}$. Direct simulation of $(\hat{R}_0, \hat{\Theta}_0)$
Algorithm 2: Obtaining the tail process of squared GARCH\((p, q)\) 

1. Simulate \(\hat{\Theta}_0\) using Algorithm 1 and independently set \(\hat{R}_0 = U^{-1/s}\) where \(U\) is uniform\((0,1)\).
2. Repeat step 1 until \(\hat{R}_0\hat{\delta}_0^{(1)} > 1\) and set \(\hat{\Theta}_0 \equiv \hat{\Theta}_0\).
3. Evaluate \(\hat{\Theta}_t^{TC} = A_t\hat{\Theta}_{t-1}\) for \(t = 1, \ldots, T\), for large \(T\).
4. Extract the tail process \(\hat{X}_t^2 = \hat{R}_0\hat{\delta}_t^{(1)}\) and associated \(\hat{\sigma}_t^2 = \hat{R}_0\hat{\delta}_t^{(q+1)}\) for \(t = 0, \ldots, T\).
5. Repeat steps 1-4 to evaluate independent replicates of the tail process \(\{\hat{X}_t^2\}_{t \geq 0}\) of \(X_t^2\), and of the associated volatilities \(\{\hat{\sigma}_t^2\}_{t \geq 0}\), with each tail process having the property that \(\hat{X}_0^2 > 1\).

produces \(\hat{X}_0^2 = \hat{R}_0\hat{\delta}_0^{(1)} > 0\), not automatically \(\hat{X}_0^2 > 1\). So we use the rejection simulation method where replications of \((\hat{R}_0, \hat{\Theta}_0)\) are generated until a realisation with \(\hat{X}_0^2 > 1\) is achieved, which is guaranteed to happen as \(Pr(\hat{X}_0^2 > 1) > 0\). For a generated pair \((\hat{R}_0, \hat{\Theta}_0)\), with \(\hat{X}_0^2 > 1\), we have most interest in \((\hat{X}_0^2, \hat{\sigma}_0^2) = (\hat{R}_0\hat{\delta}_0^{(1)}, \hat{R}_0\hat{\delta}_0^{(q+1)})\), but all elements of \((\hat{R}_0, \hat{\Theta}_0)\) are needed for Algorithms 2 and 3.

The propagation stage uses results of Basrak et al. (2002), with the tail chain for \(\hat{\Theta}\), denoted \(\{\hat{\Theta}_t^{TC}\}_{t \geq 0}\), given by

\[
\hat{\Theta}_t^{TC} = A_t\hat{\Theta}_{t-1}^{TC} \quad \text{for } t \geq 1 \quad \text{with } \hat{\Theta}_0^{TC} = \hat{\Theta}_0
\]

with \(\hat{\Theta}_0\) being the vector generated in the initialisation step. From property (31), for \(t \geq 0\) we have \(\hat{\Theta}_t^{TC} = A_tA_{t-1} \cdots A_1\hat{\Theta}_0\). We denote the components of the vector variable in the tail chain by \(\hat{\Theta}_t^{TC} = (\hat{\delta}_t^{(1)}, \ldots, \hat{\delta}_t^{(p+q)})\) for \(t \geq 1\). From this tail chain for \(t = 1, 2, \ldots\) and from the initialisation variable \(\hat{R}_0\), we determine the tail process for \(X_t^2\) and associated conditional variances, by setting \(\hat{X}_t^2 = \hat{R}_0\hat{\delta}_t^{(1)}\) and \(\hat{\sigma}_t^2 = \hat{R}_0\hat{\delta}_t^{(q+1)}\) for all \(t \geq 1\).

Ideally tail chain (31) is run until \(t = T\), where \(T\) is such that \(\hat{X}_T^2 < 1\) for all \(t > T\) with probability 1. A good approximation of \(T\) is achievable as all components of \(\hat{\Theta}_t^{TC}\) have negative drift and converge almost surely to 0 as \(t \to \infty\). So \(T\) is taken as large as possible subject to limits of storage and computational time. Following sensitivity checks we took \(T = 1000\), but for processes with weak extremal dependence \(T = 50\) is more than sufficient.

For the class of GARCH\((p, q)\) processes (1) we are interested in both the upper and lower tails of \(X_t\), and denote their respective tail processes \(\{\hat{X}_t^U\}_{t \geq 0}\) and \(\{\hat{X}_t^L\}_{t \geq 0}\), with the latter the negated lower tail process of \(X_t\). Here the constituent variables of these processes are defined by

\[
\hat{X}_t^U = \lim_{u \to \infty} \{X_t/u | X_0 > u\}, \quad \text{and} \quad \hat{X}_t^L = \lim_{u \to \infty} \{-X_t/u | X_0 < -u\} \quad t \geq 0,
\]

so \(\hat{X}_t^U > 1\) and \(\hat{X}_t^L > 1\), with no constraints on the subsequent values/signs of the subsequent terms in the tail processes. When the distribution of the innovations \(Z_t\) is symmetric about 0 these two processes are identical stochastically. When deriving the upper and lower tail processes for \(X_t\) it may appear best to evolve the strategy of de Haan et al. (1989) by starting directly from the \(\{\hat{X}_t^2\}_{t \geq 0}\) tail process and to square root this process and filter the series retaining positive values by using independent Bernoulli(\(\delta\)) variables, with \(0 \leq \delta \leq 1\) termed the tail balance, given by Breiman’s lemma to be

\[
\delta = \text{Pr}(X_t > x | |X_t| > x) = \lim_{x \to \infty} \text{Pr}(\sigma_t(Z_t) > x)/\text{Pr}(\sigma_t|Z_t| > x) = E(|Z_t|^{2\delta})/E(|Z_t|^{2\delta})
\]

where \((Z_t)_+ = \max(Z_t, 0)\). Then \(\delta = 1/2\) when the tails are symmetric (this is the case considered by de Haan et al. (1989)) and if \(\delta = 0\) or 1 the lower and upper tail are dominant respectively. Unfortunately \(\{\hat{X}_t^2\}\) and the Bernoulli(\(\delta\)) variable associated with this are not independent, which we are grateful for a referee for identifying. Although \(\delta\) is not required for our approach we do report its value as it gives a good indication of the general asymmetry in the tails of GARCH processes.

A more subtle approach is required. Critically, in addition to the \(\{\hat{X}_t^2\}_{t \geq 0}\) tail process, Algorithm 2 also provides the associated values of \(\{\hat{\sigma}_t^2\}_{t \geq 0}\) when \(p > 0\). For GARCH\((p, q)\) processes we have the relationship \(Z_t^2 = X_t^2/\sigma_t^2\) for all \(t\) for the innovations. Hence for realisations of \((\hat{X}_t^2, \hat{\sigma}_t^2; \ t = 0, \ldots, T)\) we let \(s_t := \hat{X}_t^2/\hat{\sigma}_t^2\) denote the corresponding value of \(Z_t^2\). These \(s_t\) values are observations from the i.i.d. variables \(Z_t^2\) for \(t = 0, \ldots, T\), with \(Z_t\) having probability density \(f_Z\). Given \(\{s_t\}_{0 \leq t \leq T}\), the signs of the associated \(Z_0, \ldots, Z_T\) variables are determined by independent Bernoulli(\(\delta_t\)) variables with
Taking account of whether $Z_0$ is positive or negative determines the tail processes $\{X_t^U\}_{t \geq 0}$ and $\{X_t^L\}_{t \geq 0}$, see Algorithm 3.

When $p = 0$, i.e., with ARCH(q) processes, then $Y_t$ in the relationship (3) does not involve $\hat{\sigma}^2_t$ in $\Theta_0 \in \mathcal{D}$. As we need $\hat{\sigma}^2_t$ for $t = 0, \ldots, T$ in Algorithm 3, the methods used for the GARCH(p, q) processes do not immediately apply. However, with a slight of hand, we can derive the results for this case using the GARCH(1, q) process formulation (3) with $\beta_1 = 0$, and then applying Algorithms 1-3. This works as $\hat{\sigma}^2_t$ is then part of the initial simulation and propagation of $\{\hat{\Theta}_t\}_{t \geq 0}$.

Algorithm 3: Obtaining the upper and lower tail processes of GARCH(p, q)

1. Input $\hat{X}_t^U$ and $\hat{X}_t^L$ for $t = 0, \ldots, T$ from the output of Algorithm 2.
2. Define $\hat{Z}_t^U = \hat{X}_t^U/\hat{\sigma}_t$ and evaluate $\delta_t$, for $t = 0, \ldots, T$, using the expression (33).
3. Simulate a sequence of independent variables $\{I_0, \ldots, I_T\}$, where $I_t \sim \text{Bernoulli}(\delta_t)$ is random variable which given $\delta_t$ is independent of $(\hat{X}_t^U, \hat{\sigma}_t)$ for all $t = 0, \ldots, T$.
4. If $I_0 = 1$ then $X_0^U = (X_0^C)^{1/2}$ and $X_t^U = (2I_t - 1)(\hat{X}_t^2)^{1/2}$ for $t = 1, \ldots, T$.
5. If $I_0 = 0$ then $X_0^L = (X_0^C)^{1/2}$ and $X_t^L = (1 - 2I_t)(\hat{X}_t^2)^{1/2}$ for $t = 1, \ldots, T$.
6. A single iteration produces a single tail process iteration of either $\{\hat{X}_t^U\}_{t \geq 0}$ or $\{\hat{X}_t^L\}_{t \geq 0}$ with probability $\delta_0$ and $1 - \delta_0$ respectively.

5.2 The evaluation of cluster functionals for squared and original GARCH processes

From repeated realisations of the tail processes $\{\hat{X}_t^U\}_{t \geq 0}$, $\{\hat{X}_t^L\}_{t \geq 0}$ and $\{\hat{X}_t\}_{t \geq 0}$ we can derive the properties of key cluster functionals for the $\hat{X}_t^2$, $X_t$ and $-X_t$ processes respectively. Here we illustrate this for the extremogram, the extremal index and the cluster size distribution.

The extremogram for the squared GARCH(p, q) process is

$$\chi_{X^2}(\tau) = \text{Pr}(\hat{X}_t^2 > 1|\hat{X}_0^2 > 1) = \text{Pr}(\hat{X}_t^2 > 1).$$

Similarly the extremogram for the upper and lower tails of the GARCH(p, q) process are given by $\chi_{X^U}(\tau) = \text{Pr}(\hat{X}_t^U > 1|\hat{X}_0^U > 1) = \text{Pr}(\hat{X}_t^U > 1)$ and $\chi_{X^L}(\tau) = \text{Pr}(\hat{X}_t^L > 1|\hat{X}_0^L > 1) = \text{Pr}(\hat{X}_t^L > 1)$. Thus, in each case, we can estimate the extremogram as the proportion of simulated tail processes, over independent replicates, with an exceedance of 1 at $t = \tau$.

Rootzén (1988, Theorem 4.1) and de Haan et al. (1989) derive both the extremal index and the cluster size distribution for a process by first deriving the probability that the forward tail process has $i$ exceedances, so that there are at least $i$ values being in a cluster. Expressed in terms of the tail process $\{\hat{X}_t^i\}_{t \geq 0}$ for the squared GARCH this gives $\theta_\tau^{(i)} = \text{Pr}(\#(t = 1, 2, \ldots : \hat{X}_t^2 > 1) = i - 1 | \hat{X}_0^2 > 1)$. Then the extremal index $0 < \theta_{X^2} < 1$ of the squared GARCH process, is given by $\theta_{X^2} = \theta_\tau^{(1)}$ and the probability that a cluster is of size $i$ is given by $\pi_{X^2}(i) = [\theta_\tau^{(i)} - \theta_\tau^{(i+1)}]/\theta_\tau^{(1)}$ for $i = 1, 2, \ldots$, with the reciprocal of the mean of this cluster size distribution being $\theta_{X^L}$. Identical results hold for extremal indices $\theta_{X^V}$ and $\theta_{X^L}$ for $X_t$ and $-X_t$ processes respectively.

6 Investigation of the performance of the algorithms

6.1 Introduction

Throughout this section a range of GARCH(p, q) models will be illustrated. The details of these models are given here and will be referenced subsequently as GARCH models A-E, where

A: $p = q = 2$ with $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.3, 0.15, 0.2, 0.1)$ so $\phi = 0.75$.
B: $p = q = 2$ with $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.07, 0.04, 0.8, 0.08)$ so $\phi = 0.99$.
C: $p = q = 1$ with $(\alpha_1, \beta_1) = (0.1, 0.9)$ so $\phi = 1$.
D: $p = q = 2$ with $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.07, 0.03, 0.8, 0.1)$ so $\phi = 1$.
E: $p = 0, q = 2$ with $(\alpha_1, \alpha_2) = (1.2, 0.5)$ so $\phi = 1.7$.

We selected these volatility models to give a variety of extremal behaviours, with all of these models known to be strictly stationary. Models A and B are also weak stationary. Models C and D, which are IGARCH models, and model E (Francq and Zakoian 2010, p. 35) are not weak stationary as they have $E(X_t^2) = \infty$. With the exception of model C all models have max$(p, q) \geq 2$, so extremal properties cannot be derived using existing methods of Laurini and Tawn (2012). Model C, with max$(p, q) = 1$, is an IGARCH(1,1) process, is not covered by previous results but it helps to illustrate the new methods in a case where some analytical solutions are possible. Model E is a ARCH(2) process. The parameter $\phi$ increases from model A to E, with $\phi = 1$ for models C and D.

To illustrate the effect of different choices for the distribution of the innovation $Z_t$, throughout the remainder of the paper we focus on the special cases [1] a scaled Student-$t$, [2] a normal distribution, [3] an ARMA-GARCH process, and [4] an exponential GARCH.
distribution, [2] the skew Student-$t_v$ distribution introduced in Azzalini and Capitanio (2003) and [3] the standard normal distribution. In each case the innovation distribution has zero mean and unit variance. First consider the univariate skew-$t_v$ distribution, denoted by $St(\mu, \omega, \xi, v)$, where $(\mu, \omega, \xi, v) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times (2, \infty)$ are location, scale, skewness and degrees of freedom parameters respectively. For the existence of the variance of $Z$ we require that $v > 2$. The distribution of $Z$ has density

$$f_Z(z; \mu, \omega, \xi, v) = \frac{2}{\omega} f_T(z_S; v) F_T\left(z_S \xi \sqrt{\frac{v+1}{v+z_S^2}}; v+1\right), \quad z \in \mathbb{R}$$

where $z_S = (z - \mu)/\omega$, and $f_T(\cdot; v)$ and $F_T(\cdot; v)$ denote, respectively, the density and distribution function of a Student-$t$ random variable with location and scale parameters $0$ and $1$ respectively and with $v$ degrees of freedom. The moment conditions on $Z$ of model (1) require that

$$\omega = \left[\frac{v}{v-2} - b_{v,\xi}^2\right]^{-1/2} \quad \text{and} \quad \mu = -\omega b_{v,\xi},$$

where $b_{v,\xi} = \frac{\xi}{(1+\xi^2)^{1/2}} \left[\frac{v}{\pi} \Gamma\left(\frac{v}{2} - 1/2\right)\Gamma\left(\frac{v}{2}\right)\right].$

with $\Gamma$ being the gamma function and where $v$ and $\xi$ are constrained further to ensure that $\omega$ is a positive real number. The parameter $\xi \in \mathbb{R}$ controls the skewness: $\xi > 0$ and $\xi < 0$ correspond to right and left skew respectively, and $\xi = 0$ to a symmetric distribution. Important special cases arise when: $\xi = 0$, $Z$ is the (scaled) Student-$t_v$ distribution; $v \to \infty$, $Z$ is a skew-Normal distribution; $\xi = 0$ and $v \to \infty$, $Z$ is the standard normal distribution.

Table 1 presents values of $\gamma$, $\eta$, $\kappa$, $\theta_{X^2}$, $\theta_{X^v}$, $\theta_{X^L}$ and $\delta$ for each of models A–E and for the three innovation distributions, where here, and subsequently, use of volatility model $A$ with innovation distribution [1] is denoted as model A1 and similarly for all possible 15 combinations. These values are derived using our numerical methods and evidence for their validity is provided in Sects. 6.2–6.6 for some of the models, and for all the rest in Laurini et al. (2022). Although the Monte Carlo evaluation of these results are subject to noise, it is possible to obtain any desired level of accuracy by running sufficient replicates and assessing the variability using central limit results. Throughout, we have ensured that all numerical values reported are accurate to the stated number of decimal places with a very high probability.

### 6.2 Evaluation of $\gamma$ and $\eta$

Using expression (7) for the evaluation of $\gamma$, suggests using Monte Carlo methods taking $t$ to be very large. Through Theorems 1 and 2, we have two new ways to evaluate $\gamma$ using $(\lambda^*, \eta^*)$ and $(\lambda, \eta)$ respectively. Of our methods, we prefer Theorem 2 as we found it had superior computational stability and reliability. We illustrate the performance of these methods for model A3, with similar results found for the other models.

Figure 1 (left) shows that $\gamma_t$, evaluated using expression (7), has serious numerical instabilities for large $t$. The plotted traces of ten independent realisations of $\gamma_t$ appear to have little variability, each with roughly the same negative value for $500 < t < 2000$. Critically though, the plotting of each trace terminates at a random time point $t_i > 2000$ for each of the $i = 1, \ldots, 10$, with the numerically derived values of $\gamma_t = -\infty$ for $t \geq t_i$ not plotted in Fig. 1 (left). For large $t$ all entries of the matrix $A_t \cdots A_1$ are non-negative, but at $t_i$ they are all calculated as being equal to 0 to computer machine precision, and hence $\gamma_{t_i} = -\infty$. For $t > t_i$ it follows that $\gamma_t = -\infty$. The value of $t_i$ varies over replications as it depends on the set of realisations of $\{Z_t; t \leq t_i\}$ in replicate $i$. These numerical findings suggest we cannot draw conclusions about the convergence of $\gamma_t$, as $t \to \infty$, from direct Monte Carlo evaluation of $\gamma$ using expression (7). Similar numerical degenerate features were found using $\eta^*_t$, although the associated times of degeneracy, $t_i$, were found to be larger than in Fig.1.

This numerical instability for evaluating $\gamma$ does not appear to have been reported. For example, estimates of $\gamma$ using this approach are presented for ARCH(2) processes in Francq and
Zakoïan (2010, p. 34/35), however they stop evaluating $\gamma_t$, when $t = 1000$, which is before we see the critical failure of numerical evaluation in Fig. 1. By increasing $t$, for the cases they study, we find identical numerical problems to those experienced for model A3 in Fig. 1.

From Theorem 2, $\gamma = E(\ln \lambda) + \lim_{t \to \infty} \eta_t$, with $\eta_t$ defined by (27). Critically, here each of the $\Lambda_t$ terms in the product are first divided by the associated $\lambda_t$ before the matrix multiplication, thus avoiding any computer precision problems. Figure 1 (right) shows that $\eta_t$ is converging for each of the same 10 replicates shown (left) and does not experience any numerical instability; a finding that holds whatever the value of $t$. The value we report for $\eta = \lim_{t \to \infty} \eta_t$ is evaluated as the mean over the ten different realisations using $t = 3000$, giving $\eta = 0.019$. We evaluated $E(\ln \lambda)$ by both numerical integration and using the Monte Carlo approximation $\sum_{i=1}^J \ln(\lambda_i)/t$ for large $t$ and obtained identical results of $-0.359$ once Monte Carlo noise is accounted for. Combining results we find that $\gamma = -0.359 + 0.019 = -0.34 < 0$, showing that model A3 satisfies the strict stationarity condition.

6.3 Convergence of Algorithm 1

We illustrate the convergence of Algorithm 1 for models C3 and D3. First consider model C3 where the true distribution of $\Theta_0$, here a scalar, is given by expression (13). Hence we can compare the $s$ iteration estimate $H_{\Theta_0}^{(s)}(w)$ against the truth $H_{\Theta_0}(w)$. Figure 2 illustrates this distributional convergence as well as that of the distribution of the particle weights, $m_{s, j}^{(s)} (j = 1, \ldots, J)$ of expression (30) on iteration $s$, where we take $J = 10^6$. First note that the 95% pointwise confidence intervals of the initial estimate $H_{\Theta_0}^{(0)}$ given in Sect.4 do not contain the true target distribution $H_{\Theta_0}(w)$. Despite having a good initial guess for Algorithm 1, there is a statistically significant difference here due to the slow convergence of the distribution of $\Pr(\Theta_0 < w|R_0 > u)$ to $H_{\Theta_0}(w)$ as $u \to \infty$. 

Fig. 1 Monte Carlo properties for evaluation of $\gamma$ for GARCH model A3 against iteration $t$: left shows $\gamma_t$; right $\eta_t$. Both panels have the same 10 replicates of $A_1, \ldots, A_{10}$ displayed by greyscale lines. In the left panel each trace terminates at a random time $t_i$ for the $i$th replicate, with $\gamma_t = -\infty$ for all $t \geq t_i$. 

Fig. 2 Illustrations of Algorithm 1 convergence for model C3 at iterations $s = \{0, 1, 2, 100\}$. Left, spectral distribution function shown by thick grey solid line being $H_{\Theta_0}^{(0)}(w)$ and the true limit distribution $H_{\Theta_0}(w)$ is shown by thick black line. For $s = 0$ the 95% confidence intervals are given by light grey lines. Right, kernel density estimate for the particle mass. Left panel has $s = 0, 1, 2$ and $s = 100$ (only $s = 0$ is visible as the other values are equal to the true limit); Right panel has $s = \{1, 2, 100\}$; $s = 1$—dashed grey line, $s = 2$—dotted dark grey line and $s = 100$ is the black thick solid line.
After one step of Algorithm 1 we have that $H_{\hat{\Theta}_0}^{(s)}(w)$ for $s \geq 1$ is equal to $H_{\hat{\Theta}_0}^{-}(w)$ to within visible detection, so convergence is achieved in one step. Initially, i.e., for $s = 0$, all particle weights are equal to $J^{-1}$, but, as Fig. 2 (right) shows, within an iteration they have quite a different distribution of weights and that this distribution essentially has converged at $s = 2$ to its limit form. Thus Algorithm 1 works exceptionally well in this case where we know the answer. Similar tests over other GARCH(1,1) processes gave identical convergence performances. We found almost perfect numerical instabilities in evaluating $\kappa$ for any $p$ and $q$, using Algorithm 1. Here we illustrate this method and discuss the values of $\kappa$ that are obtained for models A-E.

Next we assess the convergence of $H_{\Theta}^{(s)}(w)$, for model D3 where $\hat{\Theta}_0$ has three dimensional distribution we do not know. We cannot easily show graphically the full joint distribution convergence and even for lower dimensional summaries we can only show the algorithm converges to some limit. Figure 3 illustrates convergence for each of the marginal distributions of $H_{\Theta}^{(s)}(w)$ and the distribution of the particle weights over iterations, for $s = 0, 1, 2$ and 100. We also assessed (not shown) the convergence of the dependence structure of $H_{\Theta}^{(s)}(w)$ through monitoring how $\text{corr}(\tilde{F}_j^{(i)}, \tilde{F}_j^{(j)})$ converges as $s \to \infty$, for all $1 \leq i < j \leq p + q - 1$, where $\tilde{F}_s = (\tilde{F}_s^{(1)}, \ldots, \tilde{F}_s^{(p+q-1)})$. In all the cases we explored, visual convergence was obtained before $s = 10$ and identical values found for $s > 100$.

The same limit values were achieved for a wide range of starting values which reassure us that there is a unique limit. Although in our examples the selection of a good initialisation to the algorithm (Sect. 4) was not found to be important on the convergence speed, we anticipate it could be when the complexity of the GARCH($p, q$) model increases. Our extensive empirical findings in Sect. 7 and Laurini et al. (2022) show that the limit we find using Algorithm 1 leads to values of $\kappa$ and the extremogram and extremal index all fully in accord with long run simulations of the GARCH process, for each of models A-E, which gives us extra confidence in the convergence.

### 6.4 Evaluation of $\kappa$

Basrak and Segers (2009), and subsequent authors, imply that the way to evaluate $\kappa$ is by numerical solution of the limiting equation (16), although they do not illustrate this. In Fig. 1 (left panel) we showed that there are major numerical instabilities in evaluating $\|A_t \cdots A_t\|$ for large $t$; so in practice it is impossible to solve equation (16) directly. We have developed a numerically robust approach for evaluating $\kappa$ for any $p$ and $q$, using Algorithm 1. Here we illustrate this method and discuss the values of $\kappa$ that are obtained for models A-E.

To calculate $\kappa$ we apply Algorithm 1, iterating over different values of $k$. Key to the solution is the evaluation of the Monte Carlo estimate $\tilde{\rho}_k$ of $\rho_k$ in expression (29). Figure 4 shows $\tilde{\rho}_k$ against $k$ for each of the models A3-E3. There is clearly a unique solution for $k > 0$ to the equation $\tilde{\rho}_k = 1$, with the values of $k = \kappa$ that solve this equation given in Table 1. We restricted the Monte Carlo noise in the estimates $\tilde{\rho}_k$ of $\rho_k$, for each value of $k$ shown in Fig. 4, by taking $J = 10^6$ and evaluated the Monte Carlo integral (29) with $10^4$ replicates on $Z$ to get $\kappa$ to the required precision. To find $\kappa$, from the curve of $\tilde{\rho}_k$, we used an initial grid search coupled with a bisection method.

Table 1 illustrates that $\phi$ has an impact on the value of $\kappa$. When $\phi \neq 1$ no explicit relationship appears to hold between $\phi$ and $\kappa$, as $\kappa$ changes markedly with the innovation distribution. When $\phi < 1$ the shorter the tail of the innovation distribution gives the larger $\kappa$ and hence shorter tails of the GARCH($p, q$) marginal distribution; whereas the reverse holds when $\phi > 1$; and when $\phi = 1$ then $\kappa$ is invariant to the innovation distribution, with $\kappa = 1$. The case when $\phi > 1$ is somewhat surprising, as it might be expected that hav-
Fig. 4 Plots of \((k, \tilde{\rho}_k)\): left, for models A3 (—) and B3 (⋯) which are weak stationary; right for models C3 (black dashed), D3 (grey solid) and E3 (black dotted), which are not weak stationary. In all panels grey dotted lines represent horizontal and vertical lines set at 1.

Fig. 5 Diagnostic QQ plot for the marginal tail of the squared GARCH models A3 and B3, left and right respectively, comparing empirical and limit distributions. Results are based on 1000 simulations of \(5 \times 10^7\) GARCH processes with threshold \(x\) the 0.99998 marginal quantile. The solid line has a gradient \(\kappa\) and the conditional quantiles of empirical estimators are shown for 2.5% - 97.5% as the shaded region and for 25%, 50% and 75% quantiles as grey lines.

6.5 Extremogram

Figure 6 gives the extremogram \(\chi^2(\tau)\) for the squared of GARCH process for models A3, A1, D3 and D1, with equivalent plots for the other models given in Laurini et al. (2022). In all cases the heavier tailed innovation distribution leads to weaker extremal dependence at all lags. Models D3 and D1, both IGARCH processes, exhibit much slower decay rates in extremal dependence as lag \(\tau\) increases than for models A3 and A1 with \(\phi < 1\). The level of extremal dependence appears to be strongly related to \(\phi\). Furthermore, we see for models D3 and D1 that \(\chi^2(2) > \chi^2(1)\), with \(\chi^2(\tau)\) decaying monotonically for \(\tau \geq 2\). We believe that the reason for this feature is that \(\beta_2 > \max(\alpha_1, \alpha_2)\), which makes the variance \(\sigma^2_t\) more likely to be larger at observations with lag 2 than lag 1, and hence this induces stronger clustering of extreme values two time points apart.
An empirical estimate \( \tilde{\chi}_{\chi^2}(\tau, u) \), where \( u \) is a threshold, of the extremogram of \( \chi_{\chi^2}(\tau) \) based on a sample of length \( n \) from a GARCH(\( p, q \)) process is given by
\[
\tilde{\chi}_{\chi^2}(\tau, u) = \frac{\sum_{j=1}^{n-\tau} 1(X^2_j > u, X^2_{j+\tau} > u)}{\sum_{j=1}^{n-\tau} 1(X^2_j > u)}.
\]

Figure 6 shows \( \tilde{\chi}_{\chi^2}(\tau, u) \) for large \( n \) and for three threshold choices \( u \) corresponding quantiles of \( \chi^2_t \). This gives strong evidence that our evaluation of \( \chi_{\chi^2}(\tau) \) is accurate. Specifically, the agreement with limit values \( \chi_{\chi^2}(\tau) \) is very good generally, with the empirical estimates suffering from bias and variance trade-off, as with all threshold based estimates. Model A has the slowest convergence of the empirical estimators, but even here at the highest threshold there is almost perfect overlap between empirical estimates and the true values for all lags. In contrast, for model D the highest threshold produces the least good estimate, presumably due to its high variance. Additional comparisons are given in Laurini et al. (2022) for the other models.

6.6 Evaluation of the extremal index

Algorithms 1-3 give output values than can be used to evaluate the extremal index, \( \theta_{\chi^2}, \theta_{\chi^U} \) and \( \theta_{\chi_L} \), for processes \( \chi^2_t, \chi_t \) and \( -\chi_t \) respectively by using the methods discussed in Sect. 5.2. Here we show that these values are consistent with empirical estimates obtained from long-run simulations from the GARCH(\( p, q \)) process.

We can estimate the extremal index for a stationary process \( \{V_t\} \) using the runs estimator, \( \hat{\theta}_V(u, m) \), proposed by Smith and Weissman (1994), based on a sample of size \( n \), a threshold \( u \) and a run length \( m \), where
\[
\hat{\theta}_V(u, m) = \hat{\Pr}(V_t < u; t = 1, 2, \ldots, m \mid V_0 > u) = \frac{\sum_{j=1}^{n-m} 1(\max(V_{j+1}, \ldots, V_{j+m}) < u, V_j > u)}{\sum_{j=1}^{n-m} 1(V_j > u)}.
\]

We take \( V_t = \chi^2_t, \chi_t \) and \( -\chi_t \) respectively, where \( \chi_t \) is a GARCH(\( p, q \)) process. We compare the runs estimator (for a range of \( u \) and \( m \)) with our values of the extremal indices. As the selection of the run length \( m \) in the runs estimator is subjective (we take \( m = 100 \) and 1000 to assess sensitivity to...
For model D2, Fig. 7 shows $\theta_{X^L}, \theta_{X^U}$ and $\theta_{XL}$ together with the runs estimator ($m = 100$ (+) and $m = 1000$ (x)) and the intervals estimator (o) are illustrated with their estimated 95% confidence intervals (shaded regions) based on 100 independent replicates of $n = 10^8$ sample data. The number of exceedances for each $u$ are reported on the top axis.

Now that we are confident that we have reliable estimates of the extremal index, we examine the implications for the three extremal indices for models A-E, and each innovation distribution, with these values given in Table 1. Recall that the extremal index, $\theta_V$, of the process $\{V_t\}$ determines the average size of clusters of extremes of $\{V_t\}$ via $1/\theta_V$. We find that shorter tailed innovations give larger clusters on average. For increasing $\phi$, for $0 < \phi \leq 1$, we have increasing average cluster sizes, but that pattern does not follow when $\phi > 1$. In all cases, $\min(\theta_{X^L}, \theta_{X^U}) \geq \theta_{X^2}$, indicating the extremes of the processes $\{X_t\}$ and $\{-X_t\}$ exhibit less clustering on average than the $\{X_t^2\}$ process. When $\xi = 0$ then $\theta_{X^L} = \theta_{X^U}$, whereas for $\xi > 0$ more clustering occurs in the upper than in the lower tail of the GARCH($p, q$) process, with the reverse when $\xi < 0$.

7 Study of extremal properties for subclasses of the GARCH(2,2) process

7.1 Subclass specification

Although Sect. 6 illustrates that our algorithms provide robust methods for the computational evaluation of extremal properties of a general GARCH($p, q$) process, a systematic study of the extremal properties of a general GARCH($p, q$) process is beyond the scope of this paper. However, we consider some investigation of these properties is justified for simple subclasses, to show how different features of the construction affect the various extremal characteristics.

We consider the following classes of GARCH(2, 2) with Gaussian innovations, Class 1: $\alpha_1 = \alpha_2 = \omega \phi/2$, $\beta_1 = \beta_2 = (1 - \omega)\phi/2$ and Class 2: $\alpha_1 = \beta_1 = \omega \phi/2$, $\alpha_2 = \beta_2 = (1 - \omega)\phi/2$, with $\phi > 0$ having its usual meaning (2). Here $0 < \omega < 1$ controls the relative importance of the past process values relative to the past volatility values (Class 1) and values of process and volatility at $t - 1$ to $t - 2$ (Class 2), with increasing $\omega$ giving more weight the former in each case. For $0 < \phi < 1$ the process is stationary, whereas the necessary condition for strict stationarity when $\phi > 1$, given in Sect. 2.2, implies $\phi < 1/(1-\omega)$ (Class 1) and $\phi < 2$ (Class 2). So for these cases we need to check for strict stationarity as well.

7.2 Tail index

Without restricting $\phi$ to give strict stationarity, i.e., not checking the value of $\gamma$ to assess if strict stationarity applies, we explored what $\kappa$ values were given by applying Algorithm 1 for Classes 1 and 2. Figure 8 plots the values of evaluated $\kappa$...
against \( \phi \), for \( 0.5 \leq \phi \leq 1.5 \). For both classes and all studied \( \omega \) values, the derived values for \( \kappa \) decrease monotonically with increasing \( \phi \). There are three conclusions we can draw from these results which have broader implications.

Firstly, for each of the examples in Fig. 8 the evaluated value is \( \kappa = 0 \) for large \( \phi \). In this plot we report a value \( \kappa = 0 \) when the only solution of \( k \) to the equation (19) is \( k = 0 \) (we realise that \( \kappa \) cannot really be 0), with these cases arising when \( \rho_k \) is found to be an increasing function for \( k \geq 0 \). The value of \( \phi \) for which \( \kappa = 0 \) is achieved appears to be different for each example. We denote this by \( \phi_S \), with \( \phi_S = \min(\phi > 0 : \kappa = 0) \), and note that \( \phi_S > 1 \) in all cases.

As \( \kappa = 0 \) for \( \phi > \phi_S \) is obtained using Algorithm 1 this suggests that for these \( \phi \) values the Class 1 and 2 examples are not like the processes studied in Fig. 4, which were strictly stationary processes. A possible reason for this could be that \( \kappa \) is found to be \( 0 \) by Algorithm 1 when the process is not strictly stationary. We investigated this supposition by testing for strict stationarity of the example classes for values \( (\phi_L, \phi_U) \), of \( \phi \), which are within 0.01 of \( \phi_S \), with \( 1 < \phi_L < \phi_S < \phi_U \). In all the examples we evaluated \( \gamma \) for \( (\phi_L, \phi_U) \) and found that \( \phi_L \) gave \( \gamma < 0 \) and all \( \phi_U \) gave \( \gamma > 0 \), with the odd case requiring longer runs to reduce Monte Carlo noise issues.

For this example it appears that \( 0 < \phi < \phi_S \) is required for strict stationarity. As there is nothing special about these classes of examples it appears that strict stationarity holds if and only if the only solution to equation (19) is \( k > 0 \). If this were found to hold for all \( \text{GARCH}(p,q) \) processes then this would offer a novel and computationally efficient way to test for strict stationarity, with some benefits over the approach we developed in Theorem 2.

Secondly, Fig. 8 shows that \( \phi \) strongly influences the value of \( \kappa \), though \( \kappa \) also varies over \( \omega \). Critically we have that, whatever the value of \( \omega \), for Classes 1 and 2 that \( 0 < \phi < 1, \phi = 1, \phi > 1 \) if and only if \( \kappa > 1, \kappa = 1, 0 \leq \kappa < 1 \), respectively. These findings about \( \kappa \) and \( \phi \) are consistent with our results in Table 1 for models A-E, and with extensive empirical evidence we have obtained for a range of \( \text{GARCH}(p,q) \) process with different \( p \) and \( q \) and different innovation tail decays, so we believe they are indicative of a general result. It is also noteworthy that \( \kappa \) is much more sensitive to \( \omega \) in Class 1 than Class 2, suggesting that the process term values are much more important than the volatility terms in determining \( \kappa \).

Thirdly, if our supposition that \( \phi = 1 \) if and only if \( \kappa = 1 \) for any \( \text{IGARCH}(p,q) \) process with \( \max(p,q) \geq 2 \) held generally, it would be a novel result. This result was proved for \( p = q = 1 \) by Mikosch and Stărică (2000) and \( \kappa = 1 \) implies \( \phi = 1 \) can be shown general using Breiman’s Lemma (as noted by a referee). However, the result should not be a surprise as the finite mean and infinite variance of the \( \text{IGARCH}(p,q) \) process implies that \( 0.5 < \kappa \leq 1 \). So if our claim is true it gives that all \( \text{IGARCH}(p,q) \) processes have \( E(|X_1|^{2-\epsilon}) < \infty \) for any \( \epsilon \in (0,2] \). This claim for \( \kappa \) is not too surprising though, as the variance of \( X_1 \) is infinite when \( \phi = 1 \) but is finite when \( \phi < 1 \), suggesting that \( \kappa \), for the \( \text{IGARCH}(p,q) \), is the boundary point for having a finite variance.

A weakness of Algorithm 1 is that it is rather black-box and requires considerable numerical exploration to gain insight into how key stochastic features like \((\lambda^*, \eta^*)\) or \((\lambda, \eta)\) influence the value of \( \kappa \). To provide some insight into the role of these features we have developed a relatively good approximation \( \kappa_A \) for \( \kappa \) which can be expressed transparently and trivially as a function of these features. We used a range of heuristic arguments and empirical investigations, to propose \( \kappa_A > 0 \) as the unique positive solution the equation

\[
E \left( [\lambda \exp(\eta)]^{\kappa_A} \right) = 1,
\]  

Fig. 8 Plot of \( \kappa \) (black) and \( \kappa_A \) (grey) against \( \phi \), for classes 1 (left) and 2 (right); for \( \omega = (0.25, 0.5, 0.75) \) corresponding to continuous, dashed and dotted line types respectively. The vertical and horizontal black dotted-and-dashed lines cross at \( \phi = \kappa = 1 \). The values of \( \phi \) which our algorithms produce a value of \( \kappa = 0 \) correspond to GARCH models that are not strictly stationary.
where $\lambda$ and $\eta$ are defined in Sect. 3.3. To explain a little about where $\kappa_A$ comes from consider the GARCH(1,1) processes, where in Sect. 3.3 we found that $\lambda^* = \lambda$ and $\eta^* = \eta = 0$, yet from Mikosch and Stårică (2000) we know that $\kappa > 0$ must satisfy $E[(\alpha_1 Z_t^2 + \beta_1)^\kappa] = 1$, which is simply equation (34) in this case, so here $\kappa_A = \kappa$. Figure 8 shows that $\kappa_A$ generally provides an excellent approximation to $\kappa$ in terms how it varies with $\phi$ and $\omega$ for examples Classes 1 and 2, typically being within 7% of the value of $\kappa$. Furthermore, for models A3-E3 we find $\kappa_A$ to be, approximately $+1\%$, $+8\%$, $0\%$, $+4\%$, and $-25\%$ different than $\kappa$, respectively.

7.3 Extremal index

To see how the GARCH(2,2) parameters in affect the extremal index, Fig. 9 presents a contour plot of $\theta_{X^U}$, with normal $Z_t$, over parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$. In each panel we hold fixed two parameters and contour over the others. Figure 9, shows that $\theta_{X^U}$ decreases, i.e., average cluster sizes increase, with increasing $\phi$, up to an IGARCH(2,2) model. Consequently, contours are near linear in the parameters. Unless $\max(\beta_1, \beta_2)$ is large then small values of $\alpha_1$ and $\alpha_2$ tend to lead to very limited clustering. This seems logical as small $\alpha_1$ and $\alpha_2$ reduce the effect of the large $X_t^2$ value on subsequent volatilities, so without $\beta_1$ and $\beta_2$ being large, to pick up the momentum of the evolution of the event, the large event is very likely to die out rapidly. We obtain stronger extremal dependence with larger values of the pair $(\alpha_2, \beta_2)$ than for $(\alpha_1, \beta_1)$ for equal values of $\phi$, as seen by values of $\theta_{X^U}$ being smaller in Fig. 9 right panel by comparison to the left panel.

8 Discussion

Naïvely, it might seem that for a GARCH($p, q$) process the extremal features can be obtained sufficiently accurately using empirical estimators based on simulated data from very long runs. However, Figs. 5, 6 and 7 illustrate that this is essentially impossible due to the needs for a very large threshold $u$, to be selected for convergence to be assured, and the need for large numbers of exceedances of $u$, for numerical stability; collectively these lead to $n$ needing to be too large for computation. This is seen through the confidence intervals and threshold sensitivity in such estimates given by Laurini et al. (2022).

It appears likely that the new theory and methods we present extend to the broader class of stochastic recurrence equations discussed by Collamore and Mentemeier (2018). Specifically, for any process $Y_t \in \mathbb{R}^d$, with $Y_t = A_t Y_{t-1} + B_t$, where $A_t$ and $B_t$ are stochastic, i.i.d. sequences of matrices and vectors respectively, satisfying the conditions of Kesten (1973), then our paper gives methods which help assess stationarity of the process; determine marginal distribution tail index of $Y_t$; determine ways to simulate from the spectral measure; and are able to generate a wide range of extremal properties.

None of these properties can currently be obtained due to a lack of appropriate numerically stable algorithms. The theory and methods presented here overcome these limitations and provide a broad toolbox of numerically robust approaches to derive the extremal properties of a wide class of stochastic recurrence equations. Extensions of our methods to allow $\Pr(Z_t = 0) > 0$, to cover liquidity issues, missing values and market closures, and to account for asymmetric volatility behaviour, both areas of active research (see Francq and Zakoian, 2013 and Francq and Sucarrat, 2021), will be particularly useful.

We focused on the limiting extremal properties, but there is also interest in sub-asymptotic properties, e.g., sub-asymptotic versions of the extremogram and the extremal index given by
\( \chi_X(t; x) = \Pr(X_{t + t} > x \mid X_t > x) \)
and \( \theta_X(x, m) = \Pr(X_t < x; t = 1, 2, \ldots, m \mid X_0 > x) \)
respectively, for large but finite \( x \) and a suitably large \( m \). These sub-asymptotic quantities might converge quite slowly to their respective limits \( \chi_X(t) \) and \( \theta_X \). So it is worth evaluating how this convergence behaves and obtaining values of these extremal properties for values of \( x \) which are of relevance to applications. We believe that it should be possible to get such results using our methods in combination with the techniques of Collamore and Mentemeier (2018).

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11222-022-10164-5.

Acknowledgements We would like to thank Feridun Turkman, Kevin Glazebrook and Dmitry Korshunov for discussions and the referees for valuable comments that substantially improved the work and its presentation. FL visits to Lancaster have financially been supported by the Programme "FIL-Quota Incentivante" of the University of Parma, Italy and co-sponsored by Fondazione Cariparma, Parma, Italy. This research benefits from the HPC (High Performance Computing) facility of the University of Parma, Italy.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: Bounds on matrix product norms for \( \text{ARCH}(q) \)

For the \( \text{ARCH}(q) \) process (i.e., a \( \text{ARCH} \) process with \( p = 0 \)), it is easiest to work from the \( q \times q \) matrix \( A_t \) in expression (3). Then, letting \( \alpha_0 = \max(\alpha_1, \ldots, \alpha_{q - 1}) \),

\[
\|A_{t+1} \cdots A_1\| = \sum_{i=1}^{q} \alpha_i \sum_{j=1}^{q} a_{i,j}^{(t)} + \alpha_q Z_{t+1}^2 \sum_{j=1}^{q} a_{i,j}^{(t)} + \sum_{i=1}^{q} \sum_{j=1}^{q} a_{i,j}^{(t)} \leq [\max(\alpha_0, \alpha_q) Z_{t+1}^2 + 1]\|A_t \cdots A_1\|,
\]

by combining upper bounds on the first two terms and adding in the \( a_{i,j}^{(t)} \) values for the fundamental term. Similarly, by dropping the first term, and joining lower bounds on the last two terms, gives

\[
\|A_{t+1} \cdots A_1\| \geq \min(1, \alpha_q Z_{t+1}^2)\|A_t \cdots A_1\|.
\]
By iteration of bounds, and defining \( c_i(Z_1) = t^{-1}\ln[\|A_1\|/(\alpha_q Z_1^2)] \), then \( \eta^*_t \), in (24), satisfies

\[
\frac{1}{t} \sum_{i=2}^{t} \ln[\min(1, 1/\alpha_q Z_i^2)] + c_i(Z_1) \leq \eta^*_t
\]
\[
\leq \frac{1}{t} \sum_{i=2}^{t} \ln[(\alpha_q Z_i^2 + 1)/(\alpha_q Z_i^2)] + c_i(Z_1).
\]

If \( \eta^* := \lim_{t \to \infty} \eta^*_t \) exists, then \( -\infty < c_L \leq \eta^* \leq c_U < \infty \), with

\[
c_L = E(-2\ln|Z| - \ln \alpha_q | Z^2 > 1/\alpha_q) \Pr(Z^2 > 1/\alpha_q),
\]
\[
c_U = E[\ln(\max(\alpha_0, \alpha_q) Z^2 + 1)/\alpha_q Z^2)],
\]

where this follows by the strong law of large numbers and that \( E(\|\ln \lambda^*_t\|) < \infty \).

Appendix B: Lemmas and proofs of theorems

Proof of Lemma 1 First we show that \( E(\|\ln \lambda^*_t\|) < \infty \) holds for all \( \text{ARCH}(p, q) \) processes with \( p > 0 \). We have that for \( z^2 > \max[1, (1 - \beta_p)/\alpha_q] \), then \( \ln(\alpha_q z^2 + \beta_p) = \ln(\alpha_q z^2 + \beta_p) < \ln((\alpha_q + \beta_p)z^2) = \ln(\alpha_q + \beta_p) + 2 \ln|z| < \ln(\alpha_q + \beta_p) + 2|z| \). By this inequality and definition (1), which requires \( E(Z_t^2) = 1 \), for the upper tail of \( |Z| \) there are no conditions required for convergence of \( E(\|\ln \lambda^*_t\|) \). Similarly, for the lower tail of \( |Z|, \lambda^*_t > \beta_p > 0 \) so there are convergence issues. However, when \( p = 0 \), i.e., an \( \text{ARCH}(q) \), \( \|\ln \lambda^*_t\| = |\ln \alpha_q| + 2 \ln|z|, \) so \( E(\|\ln Z\|) < \infty \) as required.

We now show that \( E(\|\ln \lambda_t\|) < \infty \) holds for all \( \text{ARCH}(p, q) \) processes. To do this we derive the tail behaviour of \( \lambda_t \) by studying the limiting properties of the solutions of the eigenvalue equation for values of \( Z_t^2 \) tending to \( \infty \) and \( 0 \) respectively. The eigenvalues \( s \) of matrix \( A_t \) are the solutions to the equation \( \det(A_t - s I) = 0 \), where \( I \) is the \((p + q) \times (p + q)\) identity matrix. Thus \( s \) satisfies

\[
s^{p+q} - \sum_{i=1}^{p+q} \lambda_i^+ s^{p+q-i} = 0, \] where \( \lambda_i^+ = \alpha_i Z_i^2 + \beta_i \), for \( i = 1, \ldots, \max(p, q) \) where we define \( \alpha_i = 0 \) for \( i > p \) and \( \beta_i = 0 \) for \( i > p \). It follows that \( s^{\min(p, q)}(s^m - \sum_{i=1}^{\max(p, q)} \lambda_i^+ s^{m-i}) = 0 \), there are at least \( \min(p, q) \) eigenvalues which are zero and for \( m = \max(p, q) \) the non-zero eigenvalues satisfy

\[
s^m - \sum_{i=1}^{m} \lambda_i^+ s^{m-i} = 0.
\]
Now consider the behaviour of the largest eigenvalue \( \lambda_1 \) for large \( Z_t^2 = z \). In this case the \( \beta_i \) terms are negligible in \( \lambda_1^2 \) for all \( i \). Let \( I \) denote the index of the first non-zero \( \alpha_i \), so by construction \( I \leq q \). Then \( s^m - z^2 \alpha_i s^{m-i} = o(1) \), where we can drop the terms \( i = 1, \ldots, I - 1 \) as these are zero. Substitute \( s = |z|^{2/2} \) and solve for \( y \). This gives that as \( |z| \to \infty \)

\[
|z|^{2/2} y^m - \alpha |z|^{2/2} y^{m-1} + o(|z|^{2/2}) = 0.
\]

So as \( |z| \to \infty \) we have that \( y = \alpha |z|^{1/(m+1-m)} + o(1) \) and so \( s = \alpha |z|^{1/(m+1-m)} |z|^{2/2} + o(\alpha |z|^{2/2}) \), and all other eigenvalues are smaller order. Thus there exists some constants \( 0 < C_1 < C_2 \) such that \( C_1 < \lambda_i < C_2 |z|^{2/2} \) for large \( z \). So as for \( \lambda_1^2 \), no additional conditions need to hold. Similar arguments produce a lower bound on \( \lambda_t \) as \( z \to 0 \), although we prove the result differently for the GARCH \( p > 0 \) and ARCH cases. First consider \( p > 0 \), then in the limit as \( |z| \to 0 \) then

\[
s^m - \sum_{i=1}^{p} \beta_i s^{m-i} = s^{m-p} \left( s^p - \sum_{i=1}^{p} \beta_i s^{p-i} \right) = 0.
\]

The \( p \) non-zero roots have product \( -\beta_p \), so the root (possible complex) with the largest magnitude is bounded away from zero. Hence there exists a constant \( c > 0 \), such that \( \lambda_t > c \). In the ARCH case then \( s = (\alpha_1)^{1/2} |z|^{2/2} (1 + o(1)) \), so \( \ln \lambda_t = \ln s \) is linear in \( \ln |z| \), so we require that \( E(\ln |Z_t|) < \infty \).

We now show that the products of \( \lambda_t \) and \( \lambda_t^2 \) grow in a related way

\[
\mu_t = \frac{1}{t} \ln \left( \prod_{i=1}^{t} \lambda_i \right) = \frac{1}{t} \sum_{i=1}^{t} \ln \lambda_i^* - \frac{1}{t} \sum_{i=1}^{t} \ln \lambda_i \xrightarrow{a.s.} E(\ln \lambda_t^*) - E(\ln \lambda) = \mu,
\]

as \( t \to \infty \). The convergence is given by the strong law of large numbers and property (21).

**Lemma 2** If \( \alpha, \beta \in \mathbb{R} \) with \( \min(\alpha, \beta) > 0 \) and there are constants \( c, d \) with \( c \leq d \) then for all \( s \geq 0 \)

\[
\frac{c}{\max(\alpha, \beta)} \leq \frac{c(s+1)}{\alpha s + \beta} < \frac{d(s+1) + d}{\alpha s + \beta} \leq \max \left( \frac{d+1}{\beta}, \frac{d}{\alpha} \right).
\]

**Proof of Lemma 2** For \( s \geq 0 \) note that

\[
g_1(s) = \frac{c(s+1)}{\alpha s + \beta} = \frac{c}{\alpha} + \frac{c(\alpha - \beta)}{\alpha(\alpha s + \beta)}
\]

then \( g_1(s) \) is monotone decreasing, increasing and a constant if \( \alpha > \beta, \alpha < \beta \) and \( \alpha = \beta \) respectively. So, for example, in the \( \alpha > \beta \) case it follows that the minimum of \( g_1(s) \) occurs as \( s \to \infty \) giving a minimum value of \( c/\alpha \). Similarly,

\[
g_2(s) = \frac{d(s+1) + 1}{\alpha s + \beta} = \frac{d}{\alpha} + \frac{(d+1)/\beta - d/\alpha}{\alpha s + \beta}
\]

with this being decreasing, increasing and constant depending on whether \( (d+1)/\beta - d/\alpha > 0, < 0 \) or \( = 0 \). The upper bound for \( g_2(s) \) follows similarly to the lower bound for \( g_1(s) \).

**Proof of Theorem 1** Under the conditions of Lemma 1 we have \( E(\ln \lambda_t^*) < \infty \), so as \( t \to \infty \)

\[
\eta_t^* = \frac{1}{t} \ln \left| \prod_{i=1}^{t} \left( \frac{A_{i+1-i}}{\lambda^*_{i+1-i}} \right) \right| = \frac{1}{t} \ln \left( \prod_{i=1}^{t} \left( \frac{A_{i+1-i}}{\lambda^*_{i+1-i}} \right) \right) = \gamma_t - \frac{1}{t} \sum_{i=1}^{t} \ln(\lambda_t^*) \xrightarrow{a.s.} \gamma - E(\ln \lambda_t^*)
\]

where in the first line the equality holds due to linearity of norms, and in the second line the convergence step \( \gamma_t \xrightarrow{a.s.} \gamma \) from limit (7) and \( \sum_{i=1}^{t} \ln \lambda_t^*/t \xrightarrow{a.s.} E(\ln \lambda_t^*) < \infty \) due to the strong law of large numbers (Grimmett and Stirzaker 2001). So \( \lim_{t \to \infty} \eta_t^* \) must converge almost surely to a constant \( \gamma = E(\ln \lambda_t^*) + \eta^* \) with the limit having deterministic finite bounds (26) and (35) for \( p > 0 \) and \( p = 0 \) cases respectively. Denoting \( \lim_{t \to \infty} \eta_t^* \) by \( \eta^* \) we have that \( \gamma = E(\ln \lambda_t^*) + \eta^* \).

**Proof of Theorem 2** The conditions of Lemma 1 ensure that both \( E(\ln \lambda_t^*) < \infty \) and \( E(\ln |Z_t|) < \infty \) hold, and with \( \mu_t \) and \( \mu \) as defined by Lemma 1, then as \( t \to \infty \) we have that

\[
\eta_t = \frac{1}{t} \ln \|\Delta_t\| = \frac{1}{t} \ln \left( \left\| \prod_{i=1}^{t} A_{i+1-i} \right\| \right) \xrightarrow{a.s.} \gamma - E(\ln(\lambda_t^*))
\]

where the almost sure limit follows from Theorem 1 and Lemma 1. So \( \eta = \eta^* + \mu \). From Theorem 1 and the definition of \( \mu \) we have that \( \gamma = \eta^* + E(\ln \lambda_t^*) = \eta - \mu + E(\ln \lambda_t^*) = \eta + E(\ln \lambda) \).
References

Azzalini, A., Capitanio, A.: Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J. Roy. Stat. Soc. B 65(2), 367–389 (2003)

Basrak, B., Segers, J.: Regularly varying multivariate time series. Stoch. Process. Appl. 119, 1055–1080 (2009)

Basrak, B., Segers, J.: Erratum to: Regularly varying multivariate time series. Stochastic Process. Appl. 119 (2009) 1055–1080. Stochastic Processes and their Applications 121(4), 896–898 (2011)

Basrak, B., Davis, R.A., Mikosch, T.: Regular variation of GARCH processes. Stoch. Process. Appl. 99(1), 95–115 (2002)

Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. 31, 307–327 (1986)

Bougerol, P., Picard, N.: Stationarity of GARCH processes and of some nonnegative time series. J. Econom. 52, 115–127 (1992)

Breidt, J.F., Davis, R.A.: Extreme values, regular variation, and extremes. PhD thesis, University of Goettingen (2010)

Breidt, J.F., Davis, R.A.: Extremes of stochastic volatility models. Ann. Appl. Probab. 8(3), 664–675 (1998)

Buraczewski, D., Damek, E., Mikosch, T.: Stochastic Models with Power-Law Tails. Springer, Cham (2016)

Collamore, J.F., Mentemeier, S.: Large excursions and conditioned laws for recursive sequences generated by random matrices. Ann. Probab. 46(4), 2064–2120 (2018)

Collamore, J.F., Diao, G., Vidyashankar, A.N.: Recent event simulation for processes generated via stochastic fixed point equations. Ann. Appl. Probab. 24(5), 2143–2175 (2014)

Davis, R.A., Mikosch, T.: Extreme value theory for GARCH processes. In: Andersen, T., Davis, R., Kreiss, J., Mikosch, T. (eds.) Handbook of Financial Time Series, pp. 187–200. Springer, New York (2009)

Davis, R.A., Mikosch, T.: The extremumogram: a correlogram for extreme events. Bernoulli 4, 977–1009 (2009)

Del Moral, P., Guionnet, A.: On the stability of interacting processes with applications to filtering and genetic algorithms. Annales de l’Institut Henri Poincaré Probability (B) and Statistics 37(2), 155–194 (2001)

Del Moral, P., Miclo, L.: Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In: Séminaire de probabilités XXXIV, pp. 1–145. Springer (2000)

Del Moral, P., Miclo, L.: Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman–Kac semigroups. ESAIM: Probab. Stat. 7, 171–208 (2003)

Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput. 10(3), 197–208 (2000)

Ehret, A., Fiebig, U.R., Janssen, A., Schlather, M.: Joint extremal behavior of hidden and observable time series with applications to GARCH processes. Extremes 18(1), 109–140 (2015)

Ferro, C., Segers, J.: Inference for clusters of extreme values. J. Roy. Stat. Soc. B 65, 545–556 (2003)

Françoiz, C., Suerrett, G.: Volatility estimation when the zero-process is nonstationary. J. Bus. Econ. Stat. 1, 1–14 (2021). https://doi.org/10.1080/07350015.2021.1999821

Françoiz, C., Zakoian, J.M.: GARCH Models: Structure. Statistical Inference and Financial Applications. John Wiley & Sons Ltd, Chichester, United Kingdom (2010)

Françoiz, C., Zakoian, J.M.: Inference in nonstationary asymmetric GARCH models. Ann. Stat. 41(4), 1970–1998 (2013). https://doi.org/10.1214/13-AOS1132

Grimmett, G., Stirzaker, D.: Probability and Random Processes, 3rd edn. Oxford University Press, Oxford (2001)