

Evaluation of extremal properties of GARCH(p, q) processes

Fabrizio Laurini

Department of Economics and Management

University of Parma

Via J.F. Kennedy 6, 43125 Parma, Italy

fabrizio.laurini@unipr.it

Paul Fearnhead & Jonathan A. Tawn

Department of Mathematics and Statistics

Lancaster University

Lancaster LA1 4YF, UK,

p.fearnhead@lancaster.ac.uk, j.tawn@lancaster.ac.uk

August 20, 2019

Abstract

Generalized autoregressive conditionally heteroskedastic (GARCH) processes are widely used for modelling features commonly found in observed financial returns. The extremal properties of these processes are of considerable interest for market risk management. For the simplest GARCH(p, q) process, with $\max(p, q) = 1$, all extremal features have been fully characterised. Although the marginal features of extreme values of the process have been theoretically characterised when $\max(p, q) \geq 2$, much remains to be found about both marginal and dependence structure during extreme excursions. Specifically, a reliable method is required for evaluating the tail index, which regulates the marginal tail behaviour and there is a need for methods and algorithms for determining clustering. In particular, for the latter, the mean number of extreme values in a short-term cluster, i.e., the reciprocal of the extremal index, has only been characterised in special cases which exclude all GARCH(p, q) processes that are used in practice. Although recent research has identified the multivariate regular variation property of stationary GARCH(p, q) processes, currently there are no reliable methods for numerically evaluating key components of these characterisations. We overcome these issues and are able to generate the forward tail chain of the process to derive the extremal index and a range of other cluster functionals for all GARCH(p, q) processes including integrated GARCH processes and processes with unbounded and asymmetric innovations. The new theory and methods we present extend to assessing the strict stationarity and extremal properties for a much broader class of stochastic recurrence equations.

Keywords: *Cluster of extremes; extremal index; fixed point distributions; GARCH process; multivariate regular variation, particle filtering, stochastic recurrence equations*

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 , $t = 1, 2, \dots$, is the price of a generic asset. The behaviour of the extreme values of daily log-returns is critically important for market risk management. Isolated extreme values of daily log-return can often be managed, but there is

arXiv:1908.06835v1 [stat.CO] 19 Aug 2019

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 (SV) 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 with the leading decay rates the same for both models. Where they differ is in terms of their extremal dependence structure. One of the most common ways to measure this is through the lag τ tail dependence

$$\chi_X(\tau) = \lim_{x \rightarrow \infty} \Pr(X_{t+\tau} > x \mid X_t > x), \quad (1.1)$$

proposed by Ledford and Tawn (2003), with Davis and Mikosch (2009b) terming $\{\chi_\tau\}_{\tau \geq 0}$ the extremogram. For SV models Breidt and Davis (1998) show that there is no clustering of extreme values, so that $\chi_\tau = 0$ for all $\tau > 0$. Thus extreme values from SV processes occur in temporal isolation. In contrast, for any GARCH(p, q) process $\chi_\tau > 0$ for at least one value of $\tau > 0$. But, 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 this paper is to derive these extremal features for all GARCH(p, q) models used in typical financial applications and to present algorithms for their evaluation.

We consider GARCH(p, q) models, for $p \in \mathbb{N}$ and $q \in \mathbb{N}_+$, of the form

$$X_t = \sigma_t Z_t \quad (1.2)$$

where, for every fixed $t \in \mathbb{Z}$, the random variables Z_t and σ_t are independent. Furthermore, we assume that $\{Z_t\}$ are independent and identically distributed continuous random variables with $E(Z_t) = 0$ and $\text{Var}(Z_t) = 1$. The process $\{\sigma_t\}$, commonly referred to as the conditional volatility of $\{X_t\}$, is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \quad (1.3)$$

where $\alpha_0 > 0$ and the parameters $\alpha_i \geq 0, i = 1, \dots, q-1, \alpha_q > 0$ and $\beta_j \geq 0, j = 1, \dots, p-1, \beta_p > 0$ for $p \geq 1$, have to satisfy additional constraints for the process to be strictly stationary; see Section 2. When $p = 0$, equations (1.2) and (1.3) correspond to an ARCH(q) process and when

$$\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1 \quad (1.4)$$

they correspond to an integrated GARCH(p, q) process, denoted IGARCH(p, q), which is strictly stationary but not second-order stationary. 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 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) (Laurini and Tawn, 2012); and for asymmetric Z_t with $p = q = 1$ (Ehlert *et al.*, 2015). Additional results of other tails probabilities are derived in the two-dimensional case by Collamore *et al.* (2014) but are effective only for ARCH(1) as further complications arise for the GARCH(1,1).

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) and Basrak and Segers (2009), including the tails of marginal distributions and some results for the extremal clustering properties. At first sight it seems that these results give everything that is needed for numerical evaluation of the required extremal properties. But this is far from the case, as we will show.

Firstly consider the marginal tail behaviour of X_t and X_t^2 . Basrak *et al.* (2002) showed that for GARCH(p, q) processes there is an explicit theoretical expression for $\kappa \geq 0$ such that for fixed $x > 1$ as $u \rightarrow \infty$ then

$$\frac{\Pr(X_t > ux)}{\Pr(X_t > u)} \rightarrow x^{-2\kappa} \quad \text{and} \quad \frac{\Pr(X_t^2 > ux)}{\Pr(X_t^2 > u)} \rightarrow x^{-\kappa}.$$

These papers only give an asymptotic limiting expression for the evaluation of κ , but they do not illustrate its application. We find that direct computation using their expression gives very poor numerical performance. Janssen (2010) presents an alternative approach to evaluate κ , however that approach applies only under the assumption that the innovation Z_t has bounded support, ruling out many important distributions used by practitioners, e.g., Z_t being Gaussian or t -distributed. Furthermore, the associated numerical methods are very slow. We propose the first reliable and computational efficient numerical algorithms for the valuation of κ , which are valid irrespective of whether Z_t are unbounded or bounded. We also find a new formulation for κ which gives new insights and we show that for all IGARCH(p, q) processes $\kappa = 1$.

Now consider the extremal dependence/clustering features of the process. Basrak and Segers (2009) and Basrak and Segers (2011) propose computational algorithms for their evaluation. However, these methods have major limitations, which they identify, and only apply to some stochastic recurrence equations with bounded innovations, but do not hold for any GARCH(p, q) processes, see Section 2.2. 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 numerical algorithm to evaluate a range of cluster features for any GARCH(p, q) process, regardless of the values of p and q and without imposing any restrictive assumptions.

There are a range of extremal dependence features of GARCH process that are of interest to practitioners. The most standard features are summarised by the time-normalized point process N_n of exceedances of a level u_n , defined by

$$N_n(B) = \#\{k/n \in B : X_k > u_n\},$$

where u_n tends to the upper endpoint of the distribution of X_t as $n \rightarrow \infty$, such that $n\bar{F}_X(u_n) \rightarrow \psi$, for any finite $\psi > 0$, where F_X and \bar{F}_X are the marginal distribution and survivor functions of $\{X_t\}$ respectively. As $n \rightarrow \infty$, N_n converges to a compound Poisson process N , where events occur as in a homogeneous Poisson process with intensity $\psi\theta_X$, where $0 < \theta_X \leq 1$ is termed the extremal index and with multiplicities distribution denoted by $\pi_X(k)$ for $k \geq 1$ (Hsing *et al.*, 1988). We use the term clusters to describe the independent extreme events, with the associated multiplicities corresponding to the number/size of extreme values in each cluster. It follows that

$$\sum_{i=1}^{\infty} i\pi_X(i) = 1/\theta_X, \quad (1.5)$$

i.e., the extremal index θ_X is the reciprocal of the limiting mean cluster size of extreme values. The smaller the extremal index then the larger the average number of extreme values per cluster. The special case $\theta_X = 1$ corresponds to there being no clustering of extremes, so extreme values occur in isolation in time. Thus, minimally there is a need to derive $\{\chi_\tau\}_{\tau>0}$ and θ_X , and to get the latter we need $\pi_X(\cdot)$. Other functionals are also of interest to financial institutions for managing the duration of a stress period or predicting the total amount of losses that can be faced in such stress period, such as the aggregate of excesses over a cluster (total loss). These can also be derived using our methods.

All of these cluster functionals can be obtained from the tail chain of the process, which has been widely used for studying extremal clustering (Rootzén (1988), Smith *et al.* (1997), Segers (2003), Planinić and Soulier (2018)). The tail chain is defined for a heavy tailed process $\{X_t\}$ in the following way. When $u \rightarrow \infty$, if for any $t \in \mathbb{N}_+$

$$(X_0/u, X_1/u, \dots, X_t/u) \mid X_0 > u,$$

converges weakly to $(\hat{X}_0, \hat{X}_1, \dots, \hat{X}_t)$, with \hat{X}_0 non-degenerate, then the limit process $\{\hat{X}_t\}_{t \geq 0}$ is termed the forward tail chain. For SV models $\hat{X}_t = 0$ almost surely for all $t \geq 1$, so large values are not followed by large values for SV processes. In contrast, for any GARCH(p, q) process at least one \hat{X}_t , for $t \geq 1$, is non-degenerate, and every element of the tail chain is non-degenerate if $\min(\alpha_1, \dots, \alpha_q) > 0$ and $\min(\beta_1, \dots, \beta_p) > 0$, see Section 6.1 for further details.

Similarly, there is a backward tail chain $\{\hat{X}_t\}_{t \leq 0}$ with identical definition for negative t . Here \hat{X}_0 is the same value for both forward and backward chains. The connections between forward and backward chains were determined by Janssen and Segers (2014). However, in many cases only the forward chain is required to derive the functionals of interest. For example, the extremogram $\{\chi_X(\tau); \tau \geq 0\}$ and the extremal index of $\{X_t\}$ can be expressed, respectively, as

$$\chi_X(\tau) = \Pr(\hat{X}_\tau > 1 \mid \hat{X}_0 > 1) \text{ and } \theta_X = \lim_{t \rightarrow \infty} \Pr(\hat{X}_1 < 1, \dots, \hat{X}_t < 1 \mid \hat{X}_0 > 1).$$

In this paper we derive the theory for obtaining the forward tail chain, 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. We first obtain the forward tail chain for the squared GARCH(p, q) process and then use this in a filtering argument, similar to de Haan *et al.* (1989), to derive the features of interest.

The paper is structured in the following way. In Section 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 Section 3.1 we give new results for testing stationarity and a new formulation for the tail index κ . In Sections 3.3 and 4 we derive the tail chain for the series squared and original GARCH(p, q) processes respectively, with Section 3.2 containing the key particle filtering algorithm which enables us to sample from a $(p + q)$ -dimensional extreme state of the tail chain. Critical to the development of this algorithm is the theory of fixed point distributions. Section 5 discusses the novel numerical evaluation of all components of the method, including checking for stationarity and evaluating κ , and it also illustrates the rapid convergence of the particle filter algorithm. Section 6 has a study of a range of extremal dependence features of the GARCH(p, q) process over a variety of parameter values. In Section 7 we identify that the methods and algorithms that are developed here in the context of GARCH(p, q) processes immediately extend to a much broader class of stochastic recurrence equations, and so they are likely to have a much wider impact. The proofs of the theorems are given in the Appendix A.

2 Known properties of GARCH(p, q) processes

2.1 Strict Stationarity

Let us start by defining strict stationarity for GARCH(p, q) processes. We focus on the squared GARCH process, X_t^2 , and rewrite the process as a stochastic recurrence equation (SRE) as this enables the exploitation of a range of established results (Kesten, 1973) for such processes, e.g., the existence of results for the marginal distribution.

Let the $(p + q)$ vector \mathbf{Y}_t , the $(p + q) \times (p + q)$ matrix \mathbf{A}_t and the $(p + q)$ vector \mathbf{B}_t be

$$\mathbf{Y}_t = \begin{pmatrix} X_t^2 \\ \vdots \\ X_{t-q+1}^2 \\ \sigma_t^2 \\ \vdots \\ \sigma_{t-p+1}^2 \end{pmatrix} \quad \mathbf{A}_t = \begin{pmatrix} \alpha^{(q-1)} Z_t^2 & \alpha_q Z_t^2 & \beta^{(p-1)} Z_t^2 & \beta_p Z_t^2 \\ \mathbf{I}_{q-1} & \mathbf{0}_{q-1} & \mathbf{0}_{(q-1) \times (p-1)} & \mathbf{0}_{q-1} \\ \alpha^{(q-1)} & \alpha_q & \beta^{(p-1)} & \beta_p \\ \mathbf{0}_{(p-1) \times (q-1)} & \mathbf{0}_{p-1} & \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{pmatrix} \quad \mathbf{B}_t = \begin{pmatrix} \alpha_0 Z_t^2 \\ \mathbf{0}_{q-1} \\ \alpha_0 \\ \mathbf{0}_{q-1} \end{pmatrix} \quad (2.1)$$

where $\alpha^{(s)} = (\alpha_1, \dots, \alpha_s) \in \mathbb{R}^s$, $\beta^{(s)} = (\beta_1, \dots, \beta_s) \in \mathbb{R}^s$, \mathbf{I}_s is the identity matrix of size s , $\mathbf{0}_{(r \times s)}$ is a matrix of zeros with r rows and s columns and $\mathbf{0}_s$ is a column vector of zeros having length s . In each case here if $s < 0$ then these terms are to be interpreted as being dimensionless. Then it follows that the squared GARCH(p, q) processes satisfies the SRE

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{Z}, \quad (2.2)$$

where $\{\mathbf{A}_t\}$ and $\{\mathbf{B}_t\}$ are each sequences of independent and identically distributed stochastic matrices and vectors.

The formulation of the SRE via (2.1) is due to Francq and Zakoian (2010, Section 2.2.2, page 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 does not include the case $p = q = 1$. In contrast here when $p = q = 1$ we have that the terms in expression (2.1) simplify to

$$\mathbf{Y}_t = \begin{pmatrix} X_t^2 \\ \sigma_t^2 \end{pmatrix} \quad \mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha_1 & \beta_1 \end{pmatrix} \quad \mathbf{B}_t = \begin{pmatrix} \alpha_0 Z_t^2 \\ \alpha_0 \end{pmatrix}, \quad (2.3)$$

where $\alpha^{(s)}$ and $\beta^{(s)}$ are scalar and none of the \mathbf{I}_s , $\mathbf{0}_{r \times s}$ or $\mathbf{0}_s$ are included.

For general SRE of the form (2.2), but without the specific specification of \mathbf{A}_t and \mathbf{B}_t corresponding to a GARCH(p, q) process, Francq and Zakoian (2010, Theorem 2.4 - page 30/32) show that it is necessary and sufficient that there is a negative top Lyapunov exponent of \mathbf{A}_t for the existence of a unique, strictly stationary solution. Before defining the Lyapunov exponent of a stochastic matrix first consider the spectral radius of a deterministic square matrix \mathbf{A} , denoted $\rho(\mathbf{A})$. Here $\rho(\mathbf{A})$ is the greatest modulus of its eigenvalues, and an important algebraic result is $\lim_{t \rightarrow \infty} t^{-1} \log \|\mathbf{A}^t\| = \log \rho(\mathbf{A})$, where $\|\cdot\|$ is any norm on the space of \mathbf{A} . The extension to a sequence of strictly stationary and ergodic random matrices $\{\mathbf{A}_t, t \in \mathbb{Z}\}$, for which $E \ln^+ \|\mathbf{A}_t\| < \infty$ (here $\ln^+ x = \ln x$, if $x \geq 1$ and 0 otherwise), is such that the top Lyapunov exponent is

$$\gamma = \lim_{t \rightarrow \infty} \frac{1}{t} E (\ln \|\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1\|), \quad (2.4)$$

and $\exp(\gamma)$ is the spectral radius of the sequence $\{\mathbf{A}_t, t \in \mathbb{Z}\}$. Hence if $E \ln^+ \|\mathbf{A}_t\| < \infty$ for all t , it is necessary and sufficient that $\gamma < 0$ for a strictly stationary process \mathbf{Y}_t .

So for strict stationarity of the squared and original GARCH(p, q) processes X_t^2 and X_t we need $E \ln^+ \|\mathbf{A}_t\| < \infty$ for \mathbf{A}_t given by expression (2.1) and $\gamma < 0$. The finite moment condition holds for GARCH(p, q) processes as $\|\mathbf{A}_t\| < CZ_t^2 + D$, for constants $C > 0$ and $D > 0$ and so $E \ln^+ \|\mathbf{A}_t\| < E \ln^+(CZ_t^2 + D) = E(\ln(CZ_t^2 + D) \mid CZ_t^2 + D > 1) \Pr(CZ_t^2 + D > 1) < KE(\ln(Z_t) \mid CZ_t^2 + D > 1) < \infty$, for a suitable constant K and where the last inequality holds as $E(Z_t)$ is finite by the model definition (1.2). Consequently we only need to check if $\gamma < 0$.

Unfortunately, expression (2.4) is not an ideal starting point for evaluating γ . Instead we also have

$$\gamma_t = \frac{1}{t} \ln \|\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1\| \quad \text{and} \quad \gamma = \lim_{t \rightarrow \infty} \gamma_t, \quad (2.5)$$

see Francq and Zakoian (2010, Theorem 2.3 - page 30). So, via expression (2.5), it would appear a relatively simple simulation can be performed to obtain a reliable Monte Carlo estimate of γ . However, as we will show in Section 5.2, this is far from the case and a mix of careful asymptotic approximation analysis and numerical evaluation is required to evaluate γ .

In some special cases we do not need to evaluate γ to find if the process is strictly stationary, e.g., GARCH(p, q) processes are always strictly stationary when $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j \leq 1$; this includes all IGARCH(p, q) processes. It is also known that $\sum_{j=1}^p \beta_j < 1$ is necessary but not sufficient for strict stationarity. Therefore, numerical evaluation of γ is required whenever $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j > 1$ when $\sum_{j=1}^p \beta_j < 1$.

2.2 Existing Results

Basrak *et al.* (2002) show that there exists a unique stationary solution to the SRE (2.2) and that this solution exhibits a multivariate regular variation property, i.e., for any t , any norm $\|\cdot\|$ and all $r > 0$,

$$\frac{\Pr(\|\mathbf{Y}_t\| > rx, \mathbf{Y}_t / \|\mathbf{Y}_t\| \in \cdot)}{\Pr(\|\mathbf{Y}_t\| > x)} \xrightarrow{v} r^{-\kappa} \Pr(\hat{\Theta}_t \in \cdot), \quad \text{as } x \rightarrow \infty, \quad (2.6)$$

where \xrightarrow{v} denotes vague convergence (Kalleberg, 1983), $\kappa \geq 0$, and $\hat{\Theta}_t$ is a $(p+q)$ -dimensional random vector in the unit sphere (with respect to a norm $\|\cdot\|$) defined by $\mathbb{S}^{p+q} \subset \mathbb{R}^{p+q}$, and their $(p+q)$ elements will be denoted by $\hat{\Theta}_t = (\hat{\vartheta}_t^{(1)}, \dots, \hat{\vartheta}_t^{(p+q)})$. If condition (2.6) holds then \mathbf{Y}_t is said to exhibit multivariate regular variation with index κ and the distribution of $\hat{\Theta}_t$ is termed the spectral measure of the vector \mathbf{Y}_t . See Resnick (1987) for further details on multivariate regular variation. A consequence of the multivariate regular variation property (2.6) for GARCH(p, q) processes is that all the marginal variables of \mathbf{Y}_t have regularly varying tails with index κ , so in particular for $r \geq 1$ and all t

$$\Pr(X_t^2 > rx \mid X_t^2 > x) \rightarrow r^{-\kappa}, \quad \text{and} \quad \Pr(\sigma_t^2 > rx \mid \sigma_t^2 > x) \rightarrow r^{-\kappa}, \quad \text{as } x \rightarrow \infty. \quad (2.7)$$

So both the squared GARCH(p, q) process and its variance have regularly varying tails of index κ .

It is insightful to consider a slightly rearranged version of limit (2.6) and to be specific about which norm we will use. We take the L_1 norm, and define radial, R_t , and angular (two variants

Θ_t and Θ_t^-) random variables by

$$\begin{aligned} R_t &= \|\mathbf{Y}_t\| = X_t^2 + \dots + X_{t-q+1}^2 + \sigma_t^2 + \dots + \sigma_{t-p+1}^2, \\ \Theta_t &= \mathbf{Y}_t/\|\mathbf{Y}_t\| = (X_t^2, \dots, X_{t-q+1}^2, \sigma_t^2, \dots, \sigma_{t-p}^2, \sigma_{t-p+1}^2)/R_t \\ \Theta_t^- &= (X_t^2, \dots, X_{t-q+1}^2, \sigma_t^2, \dots, \sigma_{t-p}^2)/R_t, \end{aligned} \quad (2.8)$$

with \mathbb{S}^{p+q} the $(p+q)$ dimensional unit simplex. We have two angular variables as the $p+q$ dimensional variable Θ_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 Θ_t^- , which is related to Θ_t by $\Theta_t = (\Theta_t^-, 1 - \|\Theta_t^-\|)$ and Θ_t^- being Θ_t without its last component. We use this \mathbf{W}^- notation to create a $(p+q-1)$ dimensional vector from any $(p+q)$ dimensional vector \mathbf{W} on the simplex \mathbb{S}^{p+q} throughout. Furthermore, for $\mathbf{w} \in \mathbb{R}^{p+q-1}$, we use the notation

$$H_{\Theta_t}(\mathbf{w}) = \Pr(\Theta_t^- \leq \mathbf{w}), \quad (2.9)$$

with vector algebra, here and elsewhere, interpreted as being componentwise.

We will denote the limit random variables, that arise in limit (2.6), 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 $H_{\hat{\Theta}_t}$, defined similarly to distribution (2.9). Subsequently $H_{\hat{\Theta}_t}$ is referred to as the spectral measure, the term coming from multivariate regular variation terminology (Resnick, 1987). Then, for $r \geq 1$, as $x \rightarrow \infty$, the limit (2.6) becomes,

$$\Pr(R_t > rx, \Theta_t^- \leq \mathbf{w} \mid R_t > x) \xrightarrow{v} \Pr(\hat{R}_t > r) \Pr(\hat{\Theta}_t^- \leq \mathbf{w}) = r^{-\kappa} H_{\hat{\Theta}_t}(\mathbf{w}). \quad (2.10)$$

From the first expression for the asymptotic form in limit (2.10) we see that the radial variable R_t and the angular variables Θ_t become asymptotically independent, as the radial variable R_t grows due to $x \rightarrow \infty$, 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 Pareto random variable with tail index κ , i.e.,

$$\Pr(\hat{R}_t > r) = r^{-\kappa} \quad \text{for } r \geq 1. \quad (2.11)$$

There is additional structure imposed on both $H_{\hat{\Theta}_t}$ and κ by the GARCH(p, q) process, which is identified by Basrak and Segers (2009) and Janssen (2010) respectively. We discuss this structure in each case 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 (2.8), for all t , their result translates to

$$H_{\hat{\Theta}_t}(w) = \frac{2}{E(|Z|^{2\kappa})} \int_0^{(\frac{w}{1-w})^{1/2}} (1+s^2)^\kappa F_Z(ds), \quad \text{for } 0 \leq w \leq 1, \quad (2.12)$$

where F_Z is the distribution function of the innovations Z_t . When $\max(p, q) \geq 2$, through highly skilled use of the multivariate regular variation structure, Basrak and Segers (2009, Propositions 3.3, 5.1) show, that when \mathbf{A} is independent and identically distributed to \mathbf{A}_t , that

$$E(\|\mathbf{A}\hat{\Theta}_t\|^\kappa) = 1 \quad (2.13)$$

and uniquely

$$\Pr(\hat{\Theta}_t \in \cdot) = E(\|\mathbf{A}\hat{\Theta}_t\|^\kappa; \mathbf{A}\hat{\Theta}_t/\|\mathbf{A}\hat{\Theta}_t\| \in \cdot)$$

where the notation $E(X; Y) := E(X\mathbf{1}_Y)$ where $\mathbf{1}_Y$ is the indicator of the event Y . Thus

$$H_{\hat{\Theta}_t}(\mathbf{w}) = E(\|\mathbf{A}\hat{\Theta}_t\|^\kappa; (\mathbf{A}\hat{\Theta}_t/\|\mathbf{A}\hat{\Theta}_t\|)^- \leq \mathbf{w}). \quad (2.14)$$

Basrak and Segers (2009, p. 1075) propose an approach to simulate from $H_{\hat{\Theta}_t}(\mathbf{w})$ for an SRE of the form (2.2), with the required distribution $H_{\hat{\Theta}_t}$ being the invariant distribution of a Markov chain and MCMC methods used for its evaluation. However, this method cannot be used for general GARCH(p, q) processes for the following reasons. Firstly, they make an assumption that \mathbf{A}_t is bounded, which excludes the possibility of Z_t being, for example, Gaussian or t_ν distributed. Much more critically though, Basrak and Segers (2011) note that the proof that $H_{\hat{\Theta}_t}(\mathbf{w})$ is the stationary distribution of the Markov Chain that they proposed was flawed, and their claimed results only hold under one of two very specific conditions on the matrix \mathbf{A}_t in the

SRE framework (2.2). Neither of these conditions are satisfied by the form of \mathbf{A}_t for GARCH(p, q) processes, when $\max(p, q) \geq 2$, even with a bounded Z_t . Thus the algorithm proposed by Basrak and Segers (2009) cannot be used for obtaining $H_{\hat{\Theta}_t}(\mathbf{w})$ for a GARCH(p, q) process. Our approach in Section 3.2 overcomes both of these restrictions.

Next we focus on how κ is determined. In particular, Basrak *et al.* (2002) showed that there exist a $\kappa > 0$ which is the unique positive solution of the equation

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln E (\|\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1\|^\kappa) = 0 \quad (2.15)$$

where here and throughout the matrix norm we use will use is $\|\mathbf{A}\| = \sum |a_{ij}|$, where a_{ij} is (i, j) th element of matrix \mathbf{A} . For the GARCH(1, 1) process Mikosch and Stărică (2000) show that κ is simple to evaluate using expression (2.15). Specifically, taking \mathbf{A}_t as in expression (2.3) we have that

$$\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1 = \mathbf{A}_t \prod_{i=1}^{t-1} (\alpha_1 Z_i^2 + \beta_1),$$

from which it simply follows that expression (2.15) holds and κ satisfies

$$E \left[(\alpha_1 Z_t^2 + \beta_1)^\kappa \right] = 1. \quad (2.16)$$

Setting $\beta_1 = 0$ for the GARCH(1, 1) process gives the same result for κ derived by de Haan *et al.* (1989) for the ARCH(1) process. For general GARCH(p, q) processes no such existing simplification of equation (2.15) gives an easier expression for κ . So it is natural to try to find κ by a numerical solution of the limit equation (2.15). However, direct numerical solution is non-trivial due to numerical instabilities.

The only existing feasible method to evaluate κ was proposed by Janssen (2010, Proposition 4.3.1), which exploits Kesten (1973, Proof of Theorem 3). With \mathbf{A} specified in (2.2) the conditions required for the results of Kesten (1973) apply and the equality

$$\int_{\mathbb{S}^{p+q}} E \left[\|\mathbf{A}\mathbf{w}\|^k g \left(\frac{\mathbf{A}\mathbf{w}}{\|\mathbf{A}\mathbf{w}\|} \right) \right] H_k(d\mathbf{w}) = \rho_k \int_{\mathbb{S}^{p+q}} g(\mathbf{w}) H_k(d\mathbf{w}), \quad (2.17)$$

holds for all continuous functions g , all unit measures H_k on the space \mathbb{S}^{p+q} , and where ρ_k is a constant. The special case of $g \equiv 1$ in (2.17) gives the simplification

$$\int_{\mathbb{S}^{p+q}} E [\|\mathbf{A}\mathbf{w}\|^k] H_k(d\mathbf{w}) = \rho_k \int_{\mathbb{S}^{p+q}} H_k(d\mathbf{w}) = \rho_k. \quad (2.18)$$

For any given $k \in (0, \infty)$, whatever the chosen unit measure H_k , if the pair (ρ_k, H_k) satisfies equality (2.17), then ρ_k is determined solely by k , i.e., not by the choice of H_k .

As we know from condition (2.13) that the κ moment of $\|\mathbf{A}\hat{\Theta}_t\|$ is equal to 1, where $\hat{\Theta}_t \sim H_{\hat{\Theta}_t}$ on the space \mathbb{S}^{p+q} then it follows from property (2.18) that $\rho_k = 1$ when $k = \kappa$. Kesten (1973) and Janssen (2010) shows that there is only one solution to equation (2.13), so κ is the unique solution of

$$\int_{\mathbb{S}^{p+q}} E [\|\mathbf{A}\mathbf{w}\|^\kappa] H_\kappa(d\mathbf{w}) = 1, \quad (2.19)$$

and that the unit measure H_κ must correspond to the distribution function $H_{\hat{\Theta}_t}$. Thus if we can find, or simulate from, $H_{\hat{\Theta}_t}$ we can find κ . To evaluate κ 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 $H_{\hat{\Theta}_t}$, and then vary k until property (2.19) is found.

Janssen (2010) proposes an algorithm to simulate from a class of functions H_k which adapts the invalid algorithm of Basrak and Segers (2009). This gives a valid method for finding κ and $H_{\hat{\Theta}_t}$ but critically it only applies when the innovations Z_t have bounded support. Therefore there is a need for an algorithm to cover cases where Z_t have unbounded support. In Section 3.2 we describe such an algorithm that applies whatever the support of Z_t . Furthermore, our methods for calculating κ , have substantial computational performance and efficiency gains compared to the algorithm of Janssen (2010); see Section 5.6.

3 Extremal properties of squared GARCH processes

3.1 New formulations for γ and κ

Expression (2.5) suggests using Monte Carlo for the evaluation of γ , by taking t to be very large. This approach suffers from serious numerical instabilities. As the norm of the product is tending to zero it seems sensible to first normalise the size of the individual terms in the product. We do this by scaling each matrix \mathbf{A}_i by its largest (in magnitude) eigenvalue, which we denote by λ_i , with $\lambda_i > 0$ for all i . From Kesten and Spitzer (1984) for an arbitrary \mathbf{A}_i , satisfying conditions of Section 2.1, it is guaranteed that λ_i is simple and exceeds all other eigenvalues in absolute value. Let

$$\Delta_t = \prod_{i=1}^t \left(\frac{\mathbf{A}_{t+1-i}}{\lambda_{t+1-i}} \right)$$

so that the product (2.5) can be re-written as

$$\gamma_t = \frac{1}{t} \ln \|\Delta_t\| + \frac{1}{t} \sum_{i=1}^t \ln \lambda_i, \quad (3.1)$$

and let

$$\eta_t = \frac{1}{t} \ln \|\Delta_t\| \quad \text{and} \quad \eta = \lim_{t \rightarrow \infty} \eta_t. \quad (3.2)$$

We study the limit behaviour of both components in equation (3.1) in Theorem 3.1, whose proof is postponed to Appendix A.

Theorem 3.1. *If $\mathbf{A}_t, t \in \mathbb{Z}$, is a sequence of independent and identically distributed random matrices, with non-negative entries, and*

$$\mathbf{C}_t := \prod_{i=1}^t \frac{\mathbf{A}_{t+1-i}}{(\lambda_{t+1-i} \exp(\eta))},$$

where λ_i is the magnitude of the largest eigenvalue of \mathbf{A}_i and $\eta \in \mathbb{R}$ is such that

$$\ln \|\Delta_t\|/t \rightarrow \eta, \quad \text{or equivalently} \quad \ln \|\mathbf{C}_t\|/t \rightarrow 0, \quad \text{almost surely as } t \rightarrow \infty, \quad (3.3)$$

then $\gamma = \lim_{t \rightarrow \infty} \gamma_t$ if and only if

$$\gamma = E(\ln \lambda) + \eta.$$

To assess this result in terms of what is already known we first compare with the GARCH(1, 1) process. In that case Mikosch and Stărică (2000) show that $\gamma = E[\ln(\alpha_1 Z_t^2 + \beta_1)]$, but the only, and hence largest in magnitude, eigenvalue of \mathbf{A}_t is $\alpha_1 Z_t^2 + \beta_1$, thus Theorem 3.1 is identical to their result when $\eta = 0$. To show that $\eta = 0$ for all GARCH(1, 1) processes note that

$$\begin{aligned} \prod_{i=1}^t \mathbf{A}_{t+1-i} &= \mathbf{A}_t \prod_{i=1}^{t-1} (\alpha_1 Z_i^2 + \beta_1) \\ &= \mathbf{A}_t \prod_{i=1}^{t-1} \lambda_i \end{aligned}$$

so

$$\begin{aligned} \frac{1}{t} \ln \left\| \prod_{i=1}^t \mathbf{A}_{t+1-i} \right\| &= \frac{1}{t} \ln \|\mathbf{A}_t\| + \frac{1}{t} \sum_{i=1}^{t-1} \ln \lambda_i \\ &\rightarrow 0 + E(\ln \lambda). \end{aligned}$$

Hence $\eta = 0$ from Theorem 3.1.

The practical evaluation of the tail index κ is not discussed by Basrak and Segers (2009) or subsequent authors. Using representation (2.15) as the basis for numerical evaluation of κ for a GARCH(p, q) process turns out to be trivial only for very ARCH(1) and GARCH(1, 1) processes. Monte Carlo is essential, but solving the limiting equation (2.15) is non-trivial when $\max(p, q) \geq 2$ due to major numerical instabilities. In Theorem 3.2 we present a new representation which provides both insight into which factors determine κ as well as a basis for a method of evaluation with greater numerical stability.

Theorem 3.2. *Under the same notation and conditions of Theorem 3.1 and that $\gamma < 0$, the unique solution $\kappa > 0$, of the limiting equation (2.15), satisfies*

$$E([\lambda \exp(\eta)]^\kappa) = 1.$$

Furthermore, it follows that

$$\eta = -\frac{1}{\kappa} \ln(E(\lambda^\kappa)). \quad (3.4)$$

For all strictly stationary GARCH(1,1) processes, Mikosch and Stărică (2000) show that κ must satisfy $E[(\alpha_1 Z_t^2 + \beta_1)^\kappa] = 1$, but this is simply $E[\lambda^\kappa] = 1$, so Theorem 3.2 gives that $\eta = 0$ for all GARCH(1,1) processes, as shown directly above.

The results of Theorem 3.2 are particularly powerful as they allow the simple evaluation of κ , if η is known, or vice-versa. Theorem 3.1 gives a limiting expression from which to approximate η and hence κ can be found approximately, but better still, Section 3.2 gives a reliable numerical method to calculate κ and then Theorem 3.2 provides the ideal way to find η . These approaches are illustrated in Section 5. Finally as an immediate consequence of Theorems 3.1 and 3.2 if we know the process is stationary, and we know κ , we can directly calculate γ using the following result.

Theorem 3.3. *Under the same notation and conditions of Theorem 3.1, if $\gamma < 0$ and limiting equation (2.13) gives $\kappa > 0$, then*

$$\gamma = E(\ln \lambda) - \frac{1}{\kappa} \ln(E(\lambda^\kappa)).$$

Unlike Theorem 3.1, we cannot use Theorem 3.3 to test for stationarity of the process as this result only provides an expression for γ given that $\gamma < 0$, i.e., stationarity needs to be confirmed prior to its use. The only comparable existing result to Theorem 3.3 is by Kesten and Spitzer (1984) where it is shown that $\gamma \leq \ln(E(\lambda))$ for general random matrix Markov processes. Finally, Theorem 3.4 shows that for all IGARCH(p, q) processes $\kappa = 1$, in which case expression (3.4) gives that $\eta = -\ln(E(\lambda))$ and Theorem 3.3 gives that $\gamma = E(\ln \lambda) - \ln(E(\lambda))$.

Theorem 3.4. *For all IGARCH(p, q) with $\gamma < 0$, we have that $\kappa = 1$. Furthermore, if we have a stationary GARCH(p, q) process with $\kappa = 1$ then it must be an IGARCH(p, q) process.*

This seems to be the first time that it has been claimed that any IGARCH(p, q) process with $\max(p, q) \geq 2$ has $\kappa = 1$, although it was proved for $p = q = 1$ by Mikosch and Stărică (2000). The finite mean and infinite variance of the IGARCH(p, q) process implies that $0.5 < \kappa \leq 1$. So our result gives much more, e.g., all IGARCH(p, q) processes have $E(|X_t|^{2-\epsilon}) < \infty$ for any $\epsilon \in (0, 2]$. This finding for κ is not too surprising though as the variance of X_t is infinite when $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1$ but was finite when this sum is less than 1, suggesting κ for the IGARCH(p, q) was a critical boundary point for having a finite variance.

3.2 Evaluating the Spectral Measure and the Tail Index

This section gives the details of our algorithm for sampling from the limit distribution $H_{\hat{\Theta}_t}$ and then uses this algorithm repeatedly to find κ . The algorithm requires no assumptions on the support for Z_t . Throughout we take $t = 0$, both to help simplify notation here and as it will be from time $t = 0$ that we start the tail chains in Section 3.3. We will first assume that κ is known and present Algorithm 1 for generating from $H_{\hat{\Theta}_0}$ and then discuss the case when κ is unknown.

To simulate from the spectral measure $H_{\hat{\Theta}_0}$, defined via (2.14), our approach is to introduce a stochastic process whose invariant distribution is $H_{\hat{\Theta}_0}$. We will then use sequential importance sampling (Doucet *et al.*, 2000) to generate approximate samples from the state of this stochastic process at consecutive time-steps. We denote the stochastic process by $\tilde{\Theta}_s$, for $s = 0, 1, \dots$, and denote its joint distribution function at iteration s by

$$\tilde{\Theta}_s \sim H_{\tilde{\Theta}}^{(s)}. \quad (3.5)$$

This stochastic process is constructed such that as $s \rightarrow \infty$, $\tilde{\Theta}_s \xrightarrow{d} \hat{\Theta}_0 \sim H_{\hat{\Theta}_0}$. We perform sequential updates until it appears that the distribution of the state of the stochastic process has converged to the invariant distribution of the process, $H_{\hat{\Theta}_0}(\mathbf{w})$. The samples at this final iteration are then taken as samples from $H_{\hat{\Theta}_0}$. This 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).

Let $\tilde{\Theta}_s$, for $s = 0, 1, \dots$, be a Markov process, with initial state an arbitrary distribution on \mathbb{S}^{p+q} and whose transitions for $s \geq 1$ are given by

$$\Pr(\tilde{\Theta}_s \in \cdot) = \frac{E(\|\mathbf{A}\tilde{\Theta}_{s-1}\|^\kappa; \mathbf{A}\tilde{\Theta}_{s-1}/\|\mathbf{A}\tilde{\Theta}_{s-1}\| \in \cdot)}{E(\|\mathbf{A}\tilde{\Theta}_{s-1}\|^\kappa)}, \quad (3.6)$$

where, as above, $E(X; Y) := E(X\mathbf{1}_Y)$. By construction, the invariant distribution of this process is $H_{\hat{\Theta}_0}$. To see this notice that if $\tilde{\Theta}_{s-1}$ is drawn from $H_{\hat{\Theta}_0}$, then the right-hand side of (3.6) is equal to

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

As $E(\|\mathbf{A}\hat{\Theta}_0\|^\kappa) = 1$, this is equal to the definition of $H_{\hat{\Theta}_0}$ given by expression (2.14).

Furthermore, if we have a sample from $\Pr(\tilde{\Theta}_{s-1} \in \cdot)$, we can use importance sampling to generate a sample from $\Pr(\tilde{\Theta}_s \in \cdot)$. This would involve first simulating a value for $\tilde{\Theta}_s$ via

$$\tilde{\Theta}_s = \mathbf{A}\tilde{\Theta}_{s-1}/\|\mathbf{A}\tilde{\Theta}_{s-1}\|,$$

and assigning this value an importance sampling weight proportional to $\|\mathbf{A}\tilde{\Theta}_{s-1}\|^\kappa$. Thus we can use sequential importance sampling to generate samples of $\tilde{\Theta}_s$ values for $s \geq 1$. Specifically, we implement Algorithm 1, with our choice of initial distribution, in step 1, being chosen to be close to $H_{\hat{\Theta}_0}(\mathbf{w})$, so as to speed up convergence, see Section 5.3 for details. For details of how we determine convergence in step 6 of Algorithm 1 see Section 5.4.

Our approach is closely related to approaches for sampling from quasi-stationary distributions (see Griffin *et al.*, 2017, and references therein). This can be most clearly seen in situations where $\|\mathbf{A}\|$ is bounded. In this case we can define a stochastic process with transitions given by (3.6) and with killing at each iteration, with the probability of survival being proportional to $\|\mathbf{A}\tilde{\Theta}_s\|^\kappa$. In this case the spectral measure, $H_{\hat{\Theta}_0}$, is the quasi-stationary distribution of the process, i.e., the stationary distribution of the process conditional on survival.

Now consider the situation when κ is not known. For a trial value of k (for κ), apply Algorithm 1 until convergence, giving a sample of weighted particles $\{\tilde{\Theta}^{(j)}(k), m^{(j)}(k)\}_{j=1}^J$ after the chain is deemed to have converged. Using these particles approximate the expectation $E(\|\mathbf{A}\tilde{\Theta}_0\|^k)$ using the Monte Carlo sample by

$$\tilde{\rho}_k = \int_{\mathbb{R}} \sum_{j=1}^J \|\mathbf{A}\tilde{\Theta}^{(j)}(k)\|^k \frac{m^{(j)}(k)}{\sum_{n=1}^J m^{(n)}(k)} F_Z(dz), \quad (3.7)$$

where $\tilde{\rho}_k$ is the Monte Carlo approximation of ρ_k and F_Z is the distribution function of the GARCH(p, q) process. We repeat this evaluation over $k > 0$ until we find the unique value of k which gives this weighted mean to be equal to 1. This value is $k = \kappa$.

Algorithm 1: Sampling from $H_{\hat{\Theta}_0}(\mathbf{w})$

- 1 Generate a sample of $\tilde{\Theta}_0$ from any distribution of \mathbb{S}^{p+q} . See Section 5.3 for discussion on optimisation this choice. Set $s = 1$.
- 2 Generate J independent copies of \mathbf{A} , denote these as $\mathbf{A}_s^{(j)}$ for $j = 1, \dots, J$.
- 3 Generate J equally weighted particles at time $s - 1$ by sampling independently from our approximation to the distribution of $\tilde{\Theta}_{s-1}$. Denote these particles as $\Theta_{s-1}^{*(j)}$ for $j = 1, \dots, J$.
- 4 Generate J particles at time s ,

$$\tilde{\Theta}_s^{(j)} = \frac{\mathbf{A}_s^{(j)} \Theta_{s-1}^{*(j)}}{\|\mathbf{A}_s^{(j)} \Theta_{s-1}^{*(j)}\|}, \quad j = 1, \dots, J. \quad (3.8)$$

- 5 Assign each particle a weight,

$$m_s^{*(j)} = \|\mathbf{A}_s^{(j)} \Theta_{s-1}^{*(j)}\|^\kappa,$$

for $j = 1, \dots, J$, and then normalise these via

$$m_s^{(j)} = \frac{m_s^{*(j)}}{\sum_{j=1}^J m_s^{*(j)}}. \quad (3.9)$$

The resulting set of weighted particles, $\{\tilde{\Theta}_s^{(j)}, m_s^{(j)}\}_{j=1}^J$ is our empirical approximation to the distribution of $\hat{\Theta}_s$.

- 6 If we have converged to stationarity, output the set of weighted particles. Otherwise set $s = s + 1$ and go to step 2.
-

3.3 Generation of the tail chain of the squared process

The tail chain $\{\hat{X}_t^2\}_{t \geq 0}$ can be evaluated using Algorithm 2. There are two stages to the algorithm, initialisation and propagation of the chain. Key to getting the tail chain is finding the joint behaviour of $\hat{\Theta}_t = (\hat{v}_t^{(1)}, \dots, \hat{v}_t^{(p+q)})$ over time.

For initialisation we first need to consider the behaviour of the process conditional on it being in an extreme state, and we take the time of this, for convenience, to be $t = 0$. Focusing on limit (2.10) when $t = 0$, we have that the limit variables $(\hat{R}_0, \hat{\Theta}_0)$, are independent with distributions (2.11) and $H_{\hat{\Theta}_0}$, given by expression (2.14), respectively. We initialise the chain when $\hat{X}_0^2 > 1$, which is equivalent to $\hat{R}_0 \hat{v}_0^{(1)} > 1$. For propagation of the chain we can use the established results of Basrak *et al.* (2002), which give that for $t \geq 1$ that $\hat{\Theta}_t = \mathbf{A}_t \hat{\Theta}_{t-1}$ and hence

$$\hat{\Theta}_t = \mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1 \hat{\Theta}_0. \quad (3.10)$$

We extract the tail chain $\{\hat{X}_t^2\}_{t \geq 0}$ from the product of \hat{R}_0 and $\hat{v}_t^{(1)}$ for all t .

If interest is in the tail chain of the $\hat{\sigma}_t^2$ process instead, i.e., $\{\hat{\sigma}_t^2\}_{t \geq 0}$ given that $\hat{\sigma}_0^2 > 1$, then exactly the same approach can be taken as in Algorithm 2 but with the condition in step 2 changed to $\hat{R}_0 \hat{v}_0^{(q+1)} > 1$.

Algorithm 2: Obtaining the tail chain of GARCH(p, q)

- 1 Simulate $\hat{\Theta}_0$ using Algorithm 1 and independently set $\hat{R}_0 = U^{-1/\kappa}$ where U is uniform(0,1).
 - 2 Repeat step 1 until $\hat{R}_0 \hat{v}_0^{(1)} > 1$.
 - 3 Evaluate $\hat{\Theta}_t = \mathbf{A}_t \hat{\Theta}_{t-1}$ for $t = 1, \dots, T$, for large T .
 - 4 The tail chain $\hat{X}_t^2 = \hat{R}_0 \hat{v}_t^{(1)}$ and associated volatilities are $\hat{\sigma}_t^2 = \hat{R}_0 \hat{v}_t^{(q+1)}$, for $t = 0, \dots, T$.
 - 5 Repeat steps 1-4 to evaluate properties of the tail chain of \hat{X}_t^2 given $\hat{X}_0^2 > 1$.
-

Here T is selected so that $\hat{X}_t^2 = \hat{R}_0 \hat{v}_t^{(1)} < 1$ for all $t > T$ with probability as close to 1 as possible. This is achievable as all components of $\hat{\Theta}_t$ have negative drift and converge to 0 almost

surely. In practice, T is taken as large as possible subject to limits of storage and computational time, we took $T = 1000$ to save running chains unnecessarily long, but for processes with weak extremal dependence $T = 50$ is more than sufficient.

3.4 The evaluation of cluster functionals

From repeated realisations of the tail chain $\{\hat{X}_t^2\}_{t \geq 0}$ for the squared GARCH process we can derive the properties of key cluster functionals for the X_t^2 process, e.g., the extremogram, the extremal index and the cluster size distribution.

First note that the extremogram for the squared GARCH(p, q) process is

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

Thus we can numerically determine $\chi_{X^2}(\tau)$ as the proportion of tail chains starting above 1 at $t = 0$ with an exceedance at $t = \tau$ over different replicate tail chains. Any required precision of this value can be achieved by a suitable selection of the number of Monte Carlo replicate tail chains. To derive both the extremal index and the cluster size distribution we first define the measure introduced by Rootzén (1988), namely

$$\theta_{X^2}^{(i)} = \Pr(\#\{t = 0, 1, \dots : \hat{X}_t^2 > 1\} = i | \hat{X}_0^2 > 1),$$

i.e., the probability that there are at least i values in a cluster given that we look at a cluster only forwards in time from an arbitrary exceedance. O'Brien (1987) showed that the extremal index $0 < \theta_{X^2} \leq 1$, is given by $\theta_{X^2} = \theta_{X^2}^{(1)}$ and Rootzén (1988) showed that the cluster size distribution is given by

$$\pi_{X^2}(i) = \frac{\theta_{X^2}^{(i)} - \theta_{X^2}^{(i+1)}}{\theta_{X^2}^{(1)}} \text{ for } i = 1, 2, \dots, \quad (3.11)$$

with the reciprocal of the mean of this distribution being θ_{X^2} .

4 Tail chain properties for GARCH processes

First note that if X_t^2 is regularly varying with index $\kappa > 0$ and if

$$\Pr(X_t > x | |X_t| > x) \rightarrow \delta \quad \text{as } x \rightarrow \infty, \quad (4.1)$$

where $0 < \delta < 1$ then it follows that X_t is a regularly varying random variable, with index 2κ , in both its upper and lower tails. Details of the evaluation of δ are given in Section 5.7.

To translate results about the tail chain \hat{X}_t^2 of the squared GARCH process, into properties for the tail chain \hat{X}_t of the GARCH process we adopt a similar strategy to de Haan *et al.* (1989) and Ehlert *et al.* (2015). It is key to recognise that there are two tails chains for X_t , an upper and a lower tail chain \hat{X}_t^L and \hat{X}_t^U respectively, with

$$\hat{X}_t^U = I_t(\hat{X}_t^2)^{1/2} \text{ and } \hat{X}_t^L = -(1 - I_t)(\hat{X}_t^2)^{1/2} \quad (4.2)$$

where I_t is a sequence of independent and identically distributed Bernoulli(δ) variables, with $I_t = \{0, 1\}$ with respective probabilities $\{1 - \delta, \delta\}$ and where δ is given by limit (4.1). The sequence I_t is also independent of $\{\hat{Y}_t\}$.

Many properties of the cluster functions for \hat{X}_t^U and \hat{X}_t^L chains can simply be derived using Monte Carlo methods from the \hat{X}_t^2 tail chains by using Bernoulli thinning, implied by expression (4.2), but some functionals can be explicitly determined, we study a few of these below.

First, we focus on the \hat{X}_t^U tail chain, corresponding to positive events in the GARCH process. The extremogram for \hat{X}_t^U is given by

$$\chi_{X^U}(\tau) = \Pr(\hat{X}_\tau^U > 1 | \hat{X}_0^U > 1) = \Pr(I_t \hat{X}_t^2 > 1) = \delta \chi_{X^2}(\tau)$$

where $\chi_{X^2}(\tau)$ is the lag τ extremogram for the X_t^2 process. An event in the tail chain for $\{\hat{X}_t^2\}$ with i exceedances of the level 1 does not occur in the tail chain of $\{\hat{X}_t^U\}$ with probability δ^i . Therefore summing over all event lengths, the probability of no exceedances of level 1 from an event in the tail chain of $\{\hat{X}_t^2\}$ is given by

$$\Pi^U = \sum_{i=1}^{\infty} \pi_{X^2}(i)(1 - \delta)^i$$

where $\pi_{X^2}(i)$ is the probability that a cluster of length i in the $\{X_t^2\}$ series, see Section 3.4. The probability of a cluster of length j in the \hat{X}_t^U tail chain is

$$\pi_{X^U}(j) = \sum_{k \geq j} \pi_{X^2}(k) \binom{k}{j} \delta^j (1 - \delta)^{k-j} / (1 - \Pi^U),$$

where the denominator corresponds to conditioning on the cluster for $\{\hat{X}_t^2\}$ process being retained for the $\{\hat{X}_t^U\}$ series. Then the mean cluster size, $1/\theta_{X^U}$ for $\{\hat{X}_t^U\}$ is given by

$$\begin{aligned} 1/\theta_{X^U} &= \sum_{j=1}^{\infty} j \sum_{k \geq j} \pi_{X^2}(k) \binom{k}{j} \delta^j (1 - \delta)^{k-j} / (1 - \Pi^U) \\ &= \sum_{k=1}^{\infty} \pi_{X^2}(k) \sum_{j \leq k} j \pi_{X^2}(k) \binom{k}{j} \delta^j (1 - \delta)^{k-j} / (1 - \Pi^U) \\ &= \sum_{k=1}^{\infty} \pi_{X^2}(k) k \delta / (1 - \Pi^U) \\ &= \frac{\delta}{\theta_{X^2}(1 - \Pi^U)}, \end{aligned}$$

where θ_{X^2} is the extremal index of the squared GARCH process, see Section 3.4. So the extremal index of $\{X_t^U\}$ is

$$\theta_{X^U} = \theta_{X^2}(1 - \Pi^U) / \delta.$$

Similarly, for the lower tail behaviour of $\{X_t\}$ it follows that the extremogram is

$$\chi_{X^L}(\tau) = \Pr(\hat{X}_\tau^L > 1 | \hat{X}_0^L > 1) = (1 - \delta)\chi_{X^2}(\tau),$$

the probability of no values below the level -1 by X_t^L from an event in the tail chain of $\{\hat{X}_t^2\}$ is given by

$$\Pi^L = \sum_{i=1}^{\infty} \pi_{X^2}(i) \delta^i,$$

the probability of a cluster of length j in the \hat{X}_t^L tail chain is

$$\pi_{X^L}(j) = \sum_{k \geq j} \pi_{X^2}(k) \binom{k}{j} (1 - \delta)^j \delta^{k-j} / (1 - \Pi^L),$$

and extremal index of $\{X_t^L\}$ is $\theta_{X^L} = \theta_{X^2}(1 - \Pi^L) / (1 - \delta)$.

5 Numerical Methods

5.1 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 $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.3, 0.15, 0.2, 0.1)$

B : $p = q = 2$ with $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.07, 0.04, 0.8, 0.08)$

C : $p = q = 1$ with $(\alpha_1, \beta_1) = (0.1, 0.9)$

D : $p = q = 2$ with $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.07, 0.03, 0.8, 0.1)$

E : $p = 2, q = 0$ with $(\alpha_1, \alpha_2) = (1.2, 0.5)$.

We selected these models to give a variety of stationarity and extremal behaviours. Models A and B are second order stationary. Models C and D, which are IGARCH models, and model E are not second order stationary (Francq and Zakoian, 2010, p. 35).

Let $\phi = \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i$ be the sum of the meaningful GARCH parameters. The parameter ϕ is increasing from model A to E with $\phi = 1$ for models C and D. We will show that all these models are strictly stationary and that the marginal tail index κ decreases with increasing ϕ for these models. From Theorem 3.4 we have that when $\phi = 1$ then $\kappa = 1$, we will also

illustrate this numerically for models C and D. Model B corresponds to the model studied by Mikosch and Střičá (2000). Case C, though being an IGARCH(1,1) process, is not covered by previous results of Laurini and Tawn (2012) given its IGARCH form, but is of interest here as it helps to illustrate the new methods in a case where analytical solutions are possible.

In Sections 5 and 6 we take the the distribution of the innovation process Z_t , to be standard Gaussian, a scaled Student- t_ν distribution, and the skew Student- t_ν distribution introduced in Azzalini and Capitanio (2003). In each case the innovation distribution has zero mean and unit variance. First consider the univariate skew- t distribution, denoted by $\text{St}(\mu, \omega, \xi, \nu)$, where $(\mu, \omega, \xi, \nu) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times (2, \infty)$ are location, scale, skewness and degree of freedom parameters respectively. For the existence of the variance of Z we require that $\nu > 2$. The distribution of Z has density

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

where $z_S = (z - \mu)/\omega$, and f_T and F_T denote, respectively, the density and distribution function of a Student- t random variable with location and scale parameters 0 and 1 respectively and with ν degrees of freedom. With this notation we have

$$E(Z) = \mu + \omega b_{\nu, \xi} \quad \text{and} \quad \text{var}(Z) = \omega^2 \left[\frac{\nu}{\nu-2} - b_{\nu, \xi}^2 \right]$$

where

$$b_{\nu, \xi} = \frac{\xi}{(1 + \xi^2)^{1/2}} \left(\frac{\nu}{\pi} \right)^{1/2} \frac{\Gamma(\nu/2 + 1/2)}{\Gamma(\nu/2)},$$

with Γ being the gamma function. To impose the moment conditions on Z of model (1.2) we also require that

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

where ν and ξ are constrained further to ensure that ω 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_ν distribution; $\nu \rightarrow \infty$, Z is a skew-Normal distribution; $\xi = 0$ and $\nu \rightarrow \infty$, a standard normal distribution. If not stated below we assume that Z_t follows a standard normal distribution.

Table 1 presents values for the key stationarity and extremal properties, $\gamma, \eta, \kappa, \theta_{X^2}, \theta_{X^U}, \theta_{X^L}$ and δ for each of models A-E and for Student- t , asymmetric Student- t and Gaussian innovations. These values are derived using the numerical methods in the rest of Section 5, and their values are discussed in Sections 5 and 6.

5.2 Evaluation of γ and η

Expression (2.5) suggests using Monte Carlo for the evaluation of γ , by taking t to be very large. We have also introduced, through Theorems 3.1 and 3.3, two new ways to evaluate γ , with the latter only applicable once it is known that the process is stationary, and hence it is known that $\gamma < 0$. Here we compare these approaches to illustrate the superior computational stability and reliability of our proposed approaches. First we illustrate the methods in Figure 1 for model A which is known, from the discussions in Section 2.1, to be stationary without the requirement of the evaluation of γ .

Figure 1 (left) shows that γ_t , evaluated using expression (2.5), has serious numerical instabilities for large t . All ten independent realisations of γ_t appear to be converging to roughly the same negative value as t increases. This finding suggests that the limit γ is negative and so the process is correctly found to be strictly stationary. But at different random values of large t each replicate stops at a time when the norm for that replicate is calculated as 0 to machine precision. In these cases $\gamma_t = -\infty$ for all subsequent t even though γ is known to be finite. Hence wrong conclusions about strict stationarity can be reached for model A using this method. This numerical instability for evaluating γ does not appear to have been reported. For example, estimates of γ using this approach are presented for ARCH(2) processes in Francq and Zakoian (2010, p. 34/35), however they stop evaluating γ_t , when $t = 1000$, which is before we see is the critical failure of numerical evaluation in Figure 1. By increasing t we find similar numerical problems to those experienced for model A in Figure 1.

Model	γ	η	κ	θ_{X^2}	θ_{X^U}	θ_{X^L}	δ
A - 1	-0.4186	0.071	1.27	0.64	0.76		0.5
A - 2	-0.4611	0.041	1.23	0.69	0.72	0.89	0.80
A - 3	-0.3358	0.023	2.37	0.59	0.72		0.5
B - 1	-0.0400	0.003	1.26	0.38	0.49		0.5
B - 2	-0.0305	0.016	1.09	0.37	0.41	0.57	0.74
B - 3	-0.0155	0.005	1.92	0.16	0.24		0.5
C - 1	-0.0300	0	1	0.21	0.29		0.5
C - 2	-0.0335	0	1	0.29	0.33	0.45	0.71
C - 3	-0.0082	0	1	0.03	0.05		0.5
D - 1	-0.0208	0.007	1	0.21	0.29		0.5
D - 2	-0.0234	0.007	1	0.27	0.31	0.44	0.72
D - 3	-0.0062	0.002	1	0.03	0.05		0.5
E - 1	-0.7461	0.137	0.65	0.27	0.40		0.5
E - 2	-0.7595	0.129	0.68	0.29	0.40	0.45	0.56
E - 3	-0.2411	0.152	0.25	0.13	0.22		0.5

Table 1: Values of key stationarity and extremal properties for models A-E for three innovation distributions: 1 t_3 ; 2 skew t_3 with $\xi = 1$; and 3 Gaussian, with Model A - 3 denoting GARCH model formulation A with innovation distribution 3.

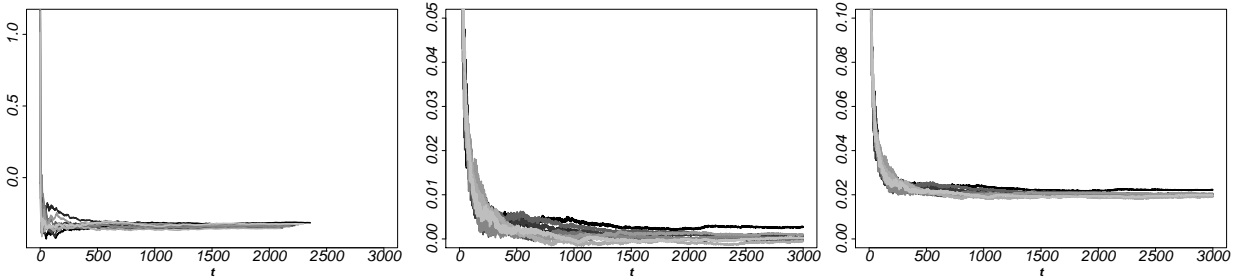


Figure 1: Monte Carlo properties for evaluation of γ for GARCH model A against iteration t : left shows γ_t ; middle $\frac{1}{t} \ln \|C_t\|$; right η_t . All panels have the same 10 replicates of $\mathbf{A}_1, \dots, \mathbf{A}_t$ displayed by greyscale lines.

Theorem 3.1 shows that actually $\gamma = E(\ln \lambda) + \lim_{t \rightarrow \infty} \eta_t$ with η_t is defined by expression (3.2), if condition (3.3) holds. To use this result first we check that condition (3.3) holds. For model A, Figure 1, shows $\frac{1}{t} \ln \|C_t\|$ for the same replicates as in the left panel. This figure indicates that this quantity appears to be converging to 0 as $t \rightarrow \infty$, i.e., condition (3.3) appears to hold.

For model A using the formulation of Theorem 3.1 we find that $\gamma = -0.359 + 0.019 = -0.34 < 0$, which clearly satisfies the strict stationarity condition. We have evaluated $E(\ln \lambda)$ by both numerical integration and using the Monte Carlo approximation $\sum_{i=1}^t \ln(\lambda_i)/t$ for large t where λ_i is the largest magnitude of the eigenvalue of \mathbf{A}_i , and obtained identical and stable results once Monte Carlo errors are accounted for. Evaluations of the η_t term are shown in Figure 1 right panel, for the same 10 replicates as before. The numerical problems are now resolved with convergence similar in all cases, with greater variation between realisations than between iterations over large ranges of t . The value we present for $\eta = \lim_{t \rightarrow \infty} \eta_t$ is evaluated as the mean over the ten different realisations using when $t = 30000$. Although Monte Carlo results are subject to noise, it is possible to obtain any desired level of accuracy by running sufficient replicates when assessing the variability using central limit results (Goldsheid, 1991). Using the results of Theorem 3.2, we can also evaluate η , but this requires knowledge of the value of κ . Using the methods of Section 5.6, based on Algorithm 1, for model A we have that $\kappa = 2.37$ and so it follows from Theorem 3.2 that $\eta = 0.0233$. Thus this shows that trying to evaluate η using the Monte Carlo limit, as in Figure 1, is liable to a small numerical error.

As stationarity of model A has been derived we can also evaluate γ using the even more

numerically reliable method given by Theorem 3.3, which comes from the expectations of two functions of the largest eigenvalue of the matrix \mathbf{A} . This evaluation requires the knowledge of κ . As for the evaluation of η above, for model A we use $\kappa = 2.37$, and we then obtain that $\gamma = -0.336$.

All of the values of both η and γ reported in Table 1 are evaluated using the methods based on Theorems 3.2 and 3.3 respectively, but of course they can only be used once stationarity has been determined, or at least there is strong evidence that γ may be negative based on using the method based on Theorem 3.1. Table 1 shows that although $\eta = 0$ for all GARCH(1, 1) processes (confirming Theorem 3.2) this does not hold for any of our GARCH(p, q) processes with $\max(p, q) \geq 2$. Furthermore, in all models, we have γ closest to 0 with the Gaussian innovation, then the symmetric t_3 distribution. There is no obvious pattern in the behaviour of η over the factors we explore in Table 1.

5.3 Initialising Algorithm 1

To be able generate realisations from the tail chain (3.10), through Algorithm 2, we first need to generate samples for $\hat{\Theta}_0$ using Algorithm 1. However, to use Algorithm 1 we need to be able to sample from a suitable random variable $\tilde{\Theta}_0$ on \mathbb{S}^{p+q} with a distribution which is as close as possible to the target limit distribution function $H_{\hat{\Theta}_0}(\mathbf{w})$, so that the rate of convergence of $\tilde{\Theta}_s \rightarrow^d \hat{\Theta}_0$ as $s \rightarrow \infty$ is maximised. From limit (2.10) we have that

$$\Pr(\Theta_0^- \leq \mathbf{w} \mid R_0 > x) \rightarrow H_{\hat{\Theta}_0}(\mathbf{w}), \quad \text{as } x \rightarrow \infty.$$

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 $H_{\hat{\Theta}_0}^{(0)}$ of $H_{\hat{\Theta}_0}$. We select u as a high threshold of R_t such that limit property (2.10) appears to be well represented, i.e., radial values appear to have a Pareto tail and radial and angular values appear independent.

In practice to obtain $H_{\hat{\Theta}_0}^{(0)}$ we generate a sample of length n from the required GARCH(p, q) process and take the empirical distribution of simulated values of Θ_t given that $R_t > u$ after a burn in period of n_b i.e.,

$$H_{\hat{\Theta}_0}^{(0)}(\mathbf{w}) = \frac{\sum_{j=n_b+1}^n \mathbf{1}(R_j > u, \Theta_j^- \leq \mathbf{w})}{n_u},$$

where $\mathbf{1}(F)$ is the indicator function of event F and with $n_u = \sum_{j=n_b+1}^n \mathbf{1}(R_j > u)$. As initial particles for Algorithm 1 we use all the realisations of Θ_t given that $R_t > u$, for $t = 1, \dots, n$. We used $n = 1.1 \times 10^7$, and u to be the 99.99% quantile of R_t giving $J = 10^3$ particles, each with equal weight J^{-1} .

5.4 Investigation into convergence of Algorithm 1

We illustrate the convergence of Algorithm 1 for models A and C. First consider model C where the true distribution of Θ_0 , here a scalar, is given by expression (2.12). Hence we can compare the s iteration estimate $\hat{H}_{\hat{\Theta}_0}^{(s)}(w)$ against the truth $\hat{H}_{\hat{\Theta}_0}(w)$. Figure 2 illustrates this distributional convergence as well as that of the distribution of the J particle weights, $\mathbf{m}^{(s)}$ of expression (3.9) on iteration s . First note that the 95% pointwise confidence intervals of the initial estimate $\hat{H}_{\hat{\Theta}_0}^{(0)}$ given in Section 5.3 does not contain the true target distribution $\hat{H}_{\hat{\Theta}_0}(w)$. Despite our efforts to obtain a good initial guess for Algorithm 1, the statistically significant difference between them is due to the slow convergence of the distribution of $\Pr(\Theta_0 < w \mid R_0 > u)$ to $H_{\hat{\Theta}_0}(w)$ as $u \rightarrow \infty$.

After one step of Algorithm 1 we have that $H_{\hat{\Theta}_0}^{(1)}(w)$ is equal to $H_{\hat{\Theta}_0}(w)$ to within visible detection. Further iterations of the algorithm lead to no visible changes in $\hat{H}_{\hat{\Theta}_0}^{(s)}(w)$ for $s > 1$. In fact, in this example, we found essentially a perfect convergence after one iteration whatever the initial distribution estimate indicating a unique solution with the algorithm being robust and highly efficient in converging to it. Now focus on the particle weights obtained in the algorithm. Initially, i.e., for $s = 0$, all weights are equal J^{-1} , but, as Figure 2 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.

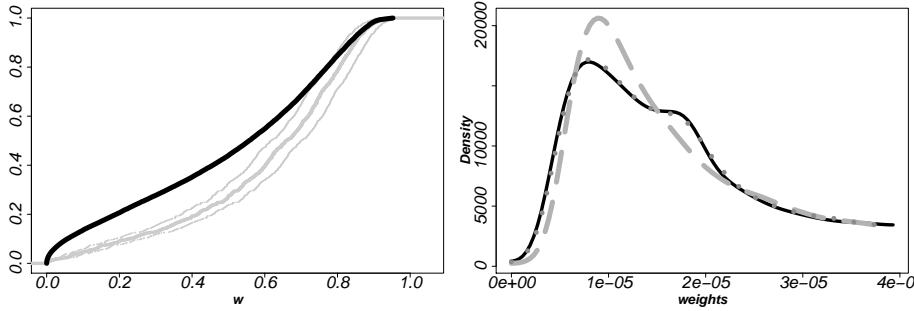


Figure 2: Illustrations of Algorithm 1 convergence for model C at iterations $s = \{0, 1, 2, 100\}$. Left, thick grey solid line is $H_{\hat{\Theta}}^{(0)}(\mathbf{w})$ and the true limit distribution $H_{\Theta}(\mathbf{w})$ is shown by thick black line. For $s = 0$ the 95% confidence intervals are given by light grey lines. Right, kernel density estimate for the particle mass. Line types are identical in each panel: $s = 1$ - dashed grey line, $s = 2$ - dotted dark grey line and $s = 100$ black thick solid line.

Next we assess the convergence of $H_{\hat{\Theta}}^{(s)}(\mathbf{w})$, over s , for model A. Here $\hat{\Theta}_0$ is four dimensional and its distribution is not known, so we cannot easily show graphically the convergence of 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 $H_{\hat{\Theta}}^{(s)}(\mathbf{w})$ over s . Other than for the second component the marginals appear to stabilise to their limit after just one iteration, for that margin it occurs in two iterations. The final panel of Figure 3 similarly shows that the distribution of the particle weights of the particles also converges after two iterations. We also assessed (not shown) the convergence of the dependence structure of $H_{\hat{\Theta}}^{(s)}(\mathbf{w})$ through monitoring how $\text{corr}(\tilde{v}_s^{(i)}, \tilde{v}_s^{(j)})$ converges to $\text{corr}(\hat{v}_0^{(i)}, \hat{v}_0^{(j)})$, where $\tilde{\Theta} = (\tilde{v}^1, \dots, \tilde{v}^{p+q})$. In all cases we found rapid convergence.

We studied a number of other GARCH(p, q) processes and found excellent convergence of the algorithm, with convergence appearing to occur in $p + q - 1$ iterations in all cases with our initialisation method, and convergence to the same value over a range of other initialisation distributions. The theory behind Algorithm 1 indicates that there is a unique solution and that the algorithm will find this, hence our numerical studies support this and show that it works with very high efficiency.

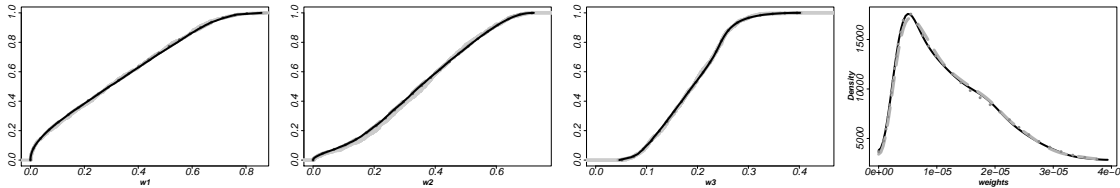


Figure 3: Convergence of Algorithm 1 for model A at iterations $s = \{0, 1, 2, 100\}$ with marginal distribution convergence for $H_{\tilde{v}_s^{(i)}}(w_i)$, for $i = 1, 2, 3$ and the kernel density for the particle mass. Values for $s > 100$ are identical to those for $s = 100$. Line types are as for Figure 2.

5.5 Assessment of the convergence for the Extremal Index

We show that our solution for $H_{\Theta_0}(w)$ gives values of functionals, such as the extremal index, which are consistent with estimates obtained from long-run simulations from the GARCH(p, q) process. The convergence of Algorithm 1 can be assessed for the cluster functionals by combining outputs from Algorithms 1 and 2. We illustrate this by finding θ_{X^2} , the extremal index of the squared process for model A. Using the tail chain $\{\hat{X}_t; t = 0, 1, \dots\}$ we have

$$\theta_{X^2} = \Pr(\hat{X}_t^2 < 1; t = 1, 2, \dots \mid \hat{X}_0^2 > 1).$$

We can estimate the extremal index using the runs estimator $\tilde{\theta}_{X^2}(u, m)$, proposed by Smith and Weissman (1994), based on a sample from the GARCH(p, q) process of length n , where

$$\begin{aligned}\tilde{\theta}_{X^2}(u, m) &= \hat{\text{Pr}}(X_t^2 < 0; t = 1, 2, \dots, m \mid X_0^2 > 1) \\ &= \frac{\sum_{j=1}^{n-m} \mathbf{1}(\max(X_{j+1}^2, \dots, X_{j+m}^2) < u, X_j^2 > u)}{\sum_{j=1}^{n-m} \mathbf{1}(X_j^2 > u)}.\end{aligned}$$

Figure 4 shows θ_{X^2} together with the runs estimate, based on $n = 10^7$, for a range of values of u and m . The limit value is within the estimated 95% confidence intervals for the runs estimate for all u and both values of m and the values of the runs estimate approach the true value θ_{X^2} as u increases for both m values, although the uncertainty in the estimators increases. This plot also goes to show why it is not possible to derive these extremal features of the GARCH(p, q) process simply from very long runs as the competing needs of large u , for convergence, and large numbers of exceedances of u , for numerical stability, makes getting numerically reliable values essentially impossible without taking n to be multiple orders of magnitude larger than here.

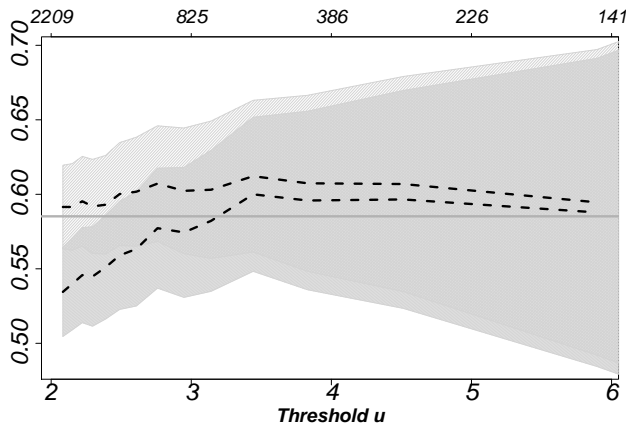


Figure 4: The squared GARCH process extremal index θ_{X^2} for the model A shown by the horizontal solid line and runs-method based estimates $\tilde{\theta}(u, m)$ obtained using a simulated process of length 10^7 . The runs estimates for different thresholds u and $m = 100$ and 1000 (top and bottom dashed lines respectively) and the associated pointwise 95% confidence intervals (grey shaded). The number of exceedances for each u is reported along the top axis.

5.6 Evaluation of κ

Basrak and Segers (2009), and subsequent authors, imply that the way to evaluate κ is to numerically solve the limiting equation (2.15), although they do not illustrate this. In Figure 1 (left panel) we showed that there are major numerical instabilities in evaluating $\|\mathbf{A}_t \cdots \mathbf{A}_1\|$ for large t ; so in practice it is impossible to solve equation (2.15) directly. In this paper we have discussed three alternative approaches for determining κ : the algorithm of Janssen (2010); using the formulation for κ given by Theorem 3.2; and exploiting Algorithm 1. Here we describe, and illustrate, the relative merits of these methods.

The algorithm of Janssen (2010) is only for bounded innovation variables. From a numerical efficiency perspective it suffers from the critical problem that as it is based on rejection sampling, meaning it can get seriously stuck. Finally, the routine was written in pure R and has a naive initialisation, so it is very slow (taking 2/3 days) to evaluate κ to an accuracy of three significant figures even when applied for a GARCH(2,1) model. The speed slows at a cubic rate as the number of the GARCH parameters grows. So this algorithm cannot be used for arbitrary GARCH(p, q) processes, even with bounded innovations.

Both the new approaches that we present for evaluating κ are not restricted by the choice of the GARCH dimensions p and q , they apply whether the innovations are bounded or unbounded, and they are relatively much faster as they are coded in C wrapped by R and run in parallel.

The first of our methods is based on the equivalent representation to limiting equation (2.15), i.e., that $\kappa > 0$ satisfies $E[(\lambda \exp(\eta))^\kappa] = 1$ as given by Theorem 3.2. As λ can be derived analytically (or, less efficiently, numerically) from \mathbf{A} it remains to find η and then κ can be found when

using either numerical integration or Monte Carlo methods to evaluate the required expectation. We derive an estimate of η using the methods presented in Section 5.2. Unfortunately, this approach is not ideal as is illustrated in Figure 1, which shows that the 10 replicates of η_t are not sufficiently stable and self-consistent in their values at large iterations to accurately deduce the precise value of the limit η .

We find that this approach only works well for calculating κ for models where $|\phi - 1| > 0.05$ as there is too much sensitivity to the uncertainty of η otherwise. Although this is not the ideal way to evaluate κ , its form gives helpful intuition into what influences κ . As seen in Section 5.2 we can actually evaluate η much more accurately, but that needs κ to be found, so that would lead to a circular argument.

Our preferred approach to calculating κ is to use Algorithm 1, iterating over k to give κ , as this provides no numerical problems whatever the dimension of the GARCH (p, q) process. Key to the solution is the evaluation of the Monte Carlo estimate $\tilde{\rho}_k$ of ρ_k in expression (3.7). Figure 5 shows $\tilde{\rho}_k$ against k for each of the models A-E. 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. However to achieve this we need to reduce the noise in the Monte Carlo estimates $\tilde{\rho}_k$ of ρ_k . For each value of k shown in Figure 5 we used $J = 10^6$ and evaluated the Monte Carlo integral (3.7) with 10^4 replicates on Z to get κ to the required precision. To find κ , from the curve of $\tilde{\rho}_k$, we used an initial grid search coupled with a bisection method.

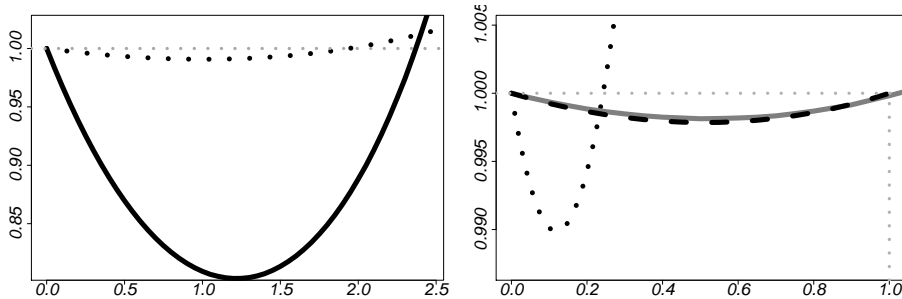


Figure 5: Plots of $(k, \tilde{\rho}_k)$: left, for models A (—) and B (···) which are second order stationary; right for models C (black dashed), D (grey solid) and E (black dotted), which are not second order stationary. In all panels grey dotted lines represent horizontal and vertical lines set at 1.

Figure 5 and Table 1 illustrate that ϕ , the sum of the meaningful GARCH parameters, has a substantial impact on the value of κ with for $\phi > 1$, we find $\kappa < 1$; when $\phi < 1$, then $\kappa > 1$; and for $\phi = 1$, $\kappa = 1$, with the latter consistent with Theorem 3.4. Unfortunately, when $\phi \neq 1$ no explicit relationship appears to hold between ϕ and κ , as κ changes markedly with the innovation distribution. From Table 1 it can be seen that when $\phi < 1$ we have that the shorter the tail of the innovation distribution gives the larger κ and hence shorter tails of the GARCH (p, q) marginal distribution; whereas the reverse holds when $\phi > 1$; and when $\phi = 1$ then κ is invariant to the innovation distribution. The case when $\phi > 1$ is somewhat surprising as at first thought you would expect that having a heavier tail innovation would result in a heavier tailed GARCH (p, q) process, whereas in fact the opposite occurs.

We finish with empirical diagnostic checks to illustrate that the derived value of κ is consistent with the observable tail of the GARCH (p, q) process. The observable tail can be derived from long run simulations. Specifically we compare the probabilities limiting $\Pr(\hat{X}_t^2 > r \mid \hat{X}_t^2 > 1) = r^{-\kappa}$ with the empirical estimate of the probabilities $\Pr(X_t^2 > rx \mid X_t^2 > x)$ for very large x , over a range of $r > 1$. Figure 6 shows this comparison on a log scale. With the choice of such scaling the true relationship between them has a gradient κ . The results show that at this far into the distributional tail, and subject to Monte Carlo noise, the empirical distribution is consistent with the limit formulation, and hence also is consistent with the value for κ that we have derived.

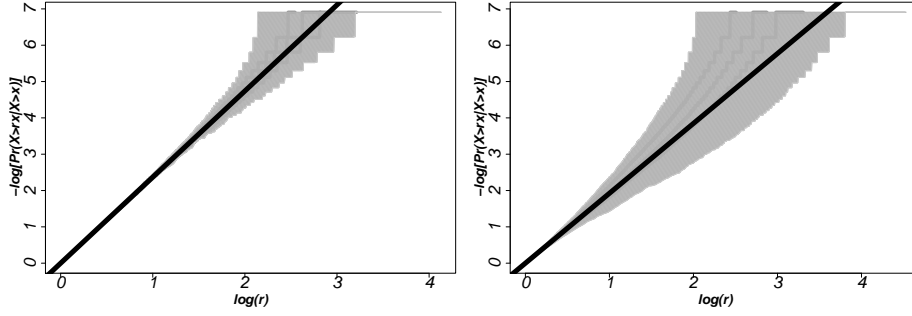


Figure 6: Diagnostic QQ plot for the marginal tail of the squared GARCH models A and B, left and right respectively, comparing empirical and limit distributions. Results are based on 1000 simulations of 5×10^7 GARCH processes with threshold x corresponding to the 0.99998 marginal quantile. The solid line has a gradient of κ 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.

5.7 Evaluation of δ

In Section 4 we introduced a tail-skewness parameter δ as a limiting conditional probability (4.1). We have not found any previous discussion on the evaluation of δ , which is an important parameter in the calculation of extremal features of GARCH(p, q) processes when the innovations are asymmetrically distributed, despite Ehlert *et al.* (2015) covering this class for GARCH(1,1) processes. A natural starting point to evaluate δ is to take a long-run simulation from the GARCH(p, q) process and simply estimate the probability (4.1) empirically for a large enough value of x . However, this is likely to be unreliable in practice.

Key to the method we propose is the following expression

$$\begin{aligned}
\delta &= \lim_{x \rightarrow \infty} \Pr(X_t > x \mid |X_t| > x) = \Pr(\hat{X}_t > 1 \mid |\hat{X}_t| > 1) \\
&= \Pr(\hat{\sigma}_t Z_t > 1 \mid \hat{\sigma}_t^2 Z_t^2 > 1) \\
&= \Pr(Z_t > 0 \mid \hat{\sigma}_t^2 Z_t^2 > 1) \\
&= \int_0^\infty \Pr(Z_t > 0 \mid \hat{\sigma}_t^2 Z_t^2 > 1, \hat{\sigma}_t^2 = s) F_{\hat{\sigma}_t^2}(ds) \\
&= \int_0^\infty \Pr(Z_t > 0 \mid Z_t^2 > s^{-1}) F_{\hat{\sigma}_t^2}(ds) \\
&= \int_0^\infty \Pr(Z_t > s^{-1/2} \mid |Z_t| > s^{-1/2}) F_{\hat{\sigma}_t^2}(ds), \quad (5.1)
\end{aligned}$$

where $F_{\hat{\sigma}_t^2}$ is the distribution of $\hat{\sigma}_t^2$, the $(q+1)$ th component of the limit vector $\hat{\Theta}_t$ which is defined by limit (2.6).

For the $\text{St}(\mu, \omega, \xi, \nu)$ innovation distribution then expression (5.1) becomes

$$\delta = \int_0^\infty \frac{1 - F_T((s^{-1/2} - \mu)\xi\sqrt{\nu+1}/\omega; \nu+1)}{1 - F_T((s^{-1/2} - \mu)\xi\sqrt{\nu+1}/\omega; \nu+1) + F_T(-(s^{-1/2} + \mu)\xi\sqrt{\nu+1}/\omega; \nu+1)} F_{\hat{\sigma}_t^2}(ds)$$

With a sample of s_1, \dots, s_m from $\hat{\sigma}_t^2$ derived using Algorithm 1 we can get a Monte Carlo approximation, to any desired accuracy through the choice of m , as follows

$$\delta \approx \frac{1}{m} \sum_{i=1}^m \frac{1 - F_T((s_i^{-1/2} - \mu)\xi\sqrt{\nu+1}/\omega; \nu+1)}{1 - F_T((s_i^{-1/2} - \mu)\xi\sqrt{\nu+1}/\omega; \nu+1) + F_T(-(s_i^{-1/2} + \mu)\xi\sqrt{\nu+1}/\omega; \nu+1)}.$$

Intuitively it seems as though δ should be equal to $\delta_Z = \lim_{x \rightarrow z_F} \Pr(Z_t > x \mid |Z_t| > x)$, where z_F is the upper end point of $|Z_t|$. This is not the case though, as knowing that $|X_t|$ is large is quite different from knowing that $|Z_t|$ is large, as the former also can be achieved with large volatility and small innovations. For $\text{St}(\mu, \omega, \xi, \nu)$ we have that δ_Z is given by

$$\delta_Z = \frac{1 - F_T(\xi\sqrt{\nu+1}/\omega; \nu+1)}{1 - F_T(\xi\sqrt{\nu+1}/\omega; \nu+1) + F_T(-\xi\sqrt{\nu+1}/\omega; \nu+1)}.$$

Figure 7 shows how $\delta = \delta(\xi)$ varies with the skew- t_3 distribution parameter ξ for $\xi \geq 0$; the values of δ for $\xi < 0$ follow due to δ being symmetrical about 0.5, i.e., for $\xi < 0$ then $\delta(\xi)$ is equal to $1 - \delta(|\xi|)$. The figure shows that for a given level of ξ , i.e., skewness in the innovation distribution, as ϕ increases there is a diminishing level of skewness in the tails of the GARCH(p, q) process as measured by δ . When $\phi = 0$ then $\delta = \delta_Z$, and this value is seen to be an upper bound for δ in Figure 7. These results are consistent with intuition as the larger the value of ϕ the more the process is driven by the past values of the process and the less the new innovations (and their skewness) matter. Table 1 illustrates how δ changes over models when $\xi = 1$, with δ typically decreasing as ϕ increases, as the persistence of volatility is more important than the innovation structure as ϕ increases.

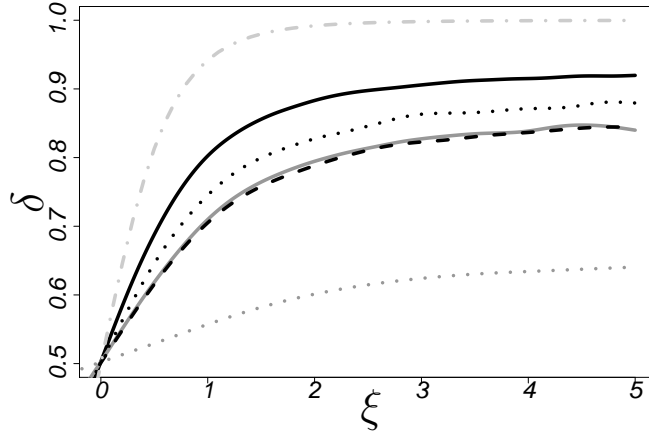


Figure 7: Plot of $\delta = \delta(\xi)$ against ξ , for $\xi \geq 0$, for models: A (black continuous), B (black dotted), C (black dashed), D (grey continuous) and E (grey dotted) for the skew t distribution with $\nu = 3$. Also plotted is δ_Z (light grey dashed and dotted). For $\xi < 0$ then $\delta(\xi)$ is equal to $1 - \delta(|\xi|)$. Line types are as consistent as possible with Figure 5.

6 Results for the GARCH(p, q) process

6.1 Extremogram

Figure 8 gives the extremogram $\chi_{X^2}(\tau)$ for the squared of GARCH process for models A-D with normal and t innovations with $\nu = 3$ degrees of freedom. Firstly notice the impact of the innovation distribution on $\chi_{X^2}(\tau)$. In all cases the heavier tailed innovation distribution leads to weaker extremal dependence at all lags. Models C and D, both IGARCH processes (with $\phi = 1$), exhibit much slower decay rates in extremal dependence as lag τ increases than for models A and B with $\phi < 1$, with the level of extremal dependence appearing to be strongly related to ϕ . Furthermore, we see for models B and D that $\chi_{X^2}(2) > \chi_{X^2}(1)$, with $\chi_{X^2}(\tau)$ decaying monotonically for $\tau \geq 2$. The reason for this seems to be that $\beta_2 > \max(\alpha_1, \alpha_2)$ here. From Section 4 we have that the evaluation of $\chi_{X^U}(\tau)$ and $\chi_{X^L}(\tau)$ is simple from $\chi_{X^2}(\tau)$ once δ is known, which we have from Section 5.7 and Table 1.

An empirical estimate $\tilde{\chi}_{X^2}(\tau, u)$ of the extremogram of $\chi_{X^2}(\tau)$ based on a sample of length n from a GARCH(p, q) process is given by

$$\tilde{\chi}_{X^2}(\tau, u) = \frac{\sum_{j=1}^{n-\tau} \mathbf{1}(X_j^2 > u, X_{j+\tau}^2 > u)}{\sum_{j=1}^{n-\tau} \mathbf{1}(X_j^2 > u)},$$

where u is a threshold. Figure 8 shows $\tilde{\chi}_{X^2}(\tau, u)$ for large n and for three threshold choices u corresponding to 0.99, 0.999 and 0.9999 quantiles of X_t^2 . The agreement with limit values $\chi_{X^2}(\tau)$ that we have evaluated is very good generally, with the empirical estimates suffering from bias and variance trade-off, as with all threshold based estimates. Model B has the slowest convergence of the empirical estimator, but even here at the highest threshold there is almost perfect overlap between empirical estimates and the true values for all lags τ . In contrast for model C the highest threshold produces the least good estimate, presumably due to its high variance. This gives strong evidence that our evaluation of $\chi_{X^2}(\tau)$ is accurate, but it also shows how difficult it is to get

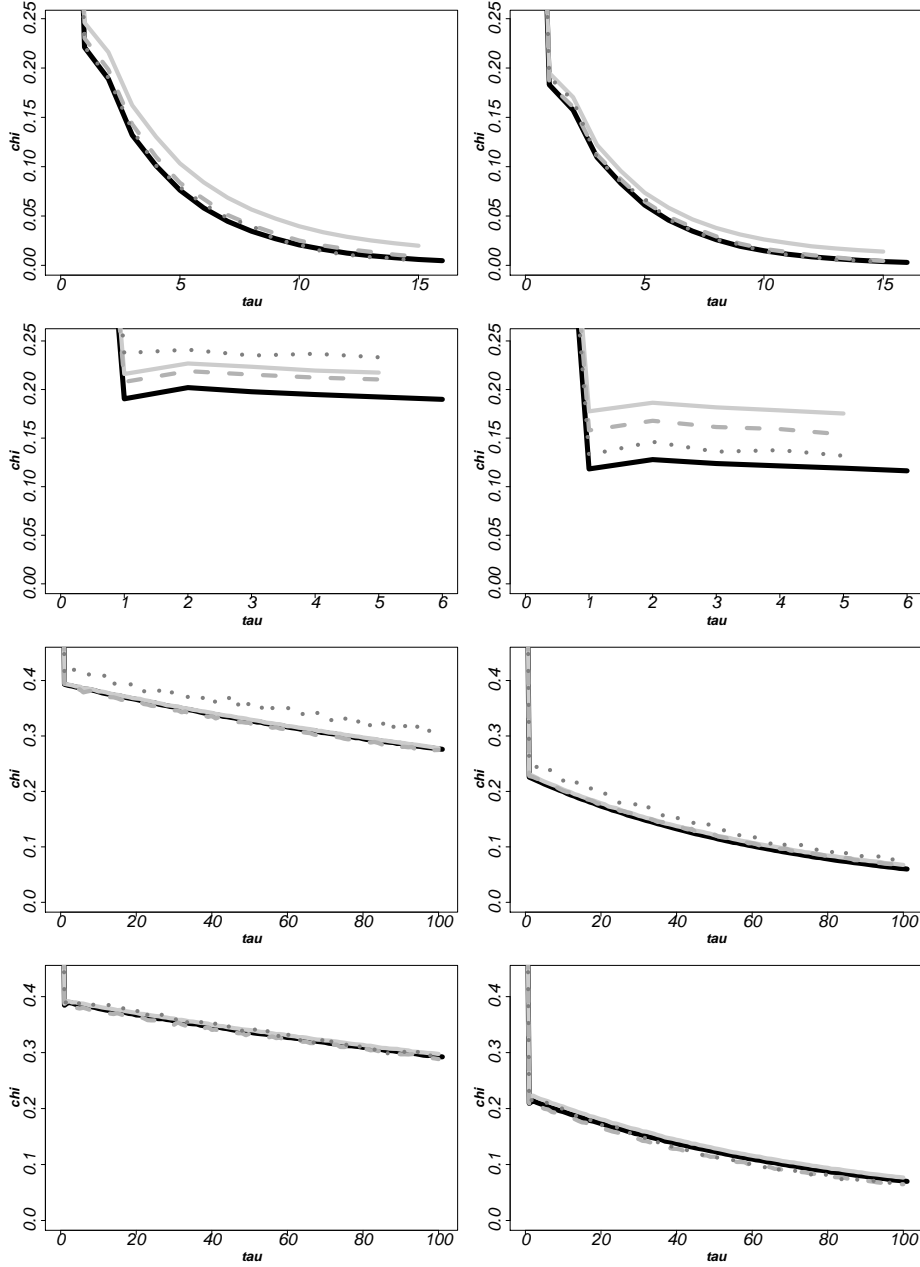


Figure 8: Extremogram (τ, χ_τ) for various squared GARCH processes with: $Z_t \sim N(0, 1)$ (left panels) and scaled $t(3)$ (right panels) and for models A-D from top to bottom rows respectively. Black lines are true limit values and the three grey lines are empirical extremogram estimates $\tilde{\chi}_\tau(u)$, based on a sample of size $n = 5 \times 10^7$, at u corresponding to 0.99 (continuous solid light grey), 0.999 (dashed grey) and 0.9999 (dotted dark grey) quantiles of X_t^2 .

accurate values from direct simulations due to different convergence rates from apparently rather similar models.

6.2 Extremal Index

We finish by looking at how the extremal index θ_{X^2} , θ_{X^U} and θ_{X^L} , of the upper tail of the series $\{X_t^2\}$, $\{X_t\}$ and $\{-X_t\}$, change over GARCH(p, q) processes. Results for each of these characteristics are given for models A-E and three innovation distributions are given in Table 1. Firstly consider the effect of the innovation distribution for a given model on the extremal index of the $\{X_t^2\}$ process, θ_{X^2} through the associated average size of clusters extremes values, i.e., $1/\theta_{X^2}$. With shorter tailed innovations clusters last longer on average and introducing skewness

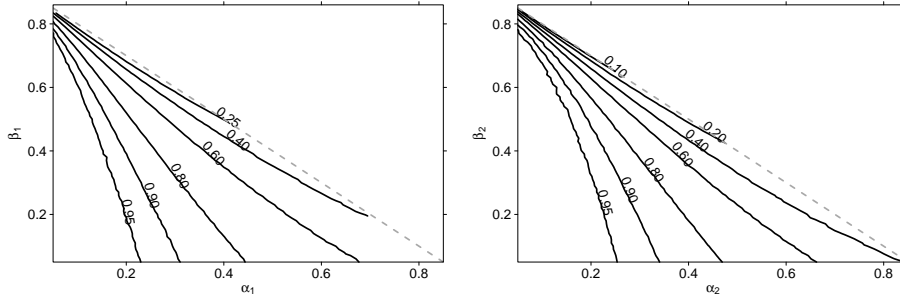


Figure 9: Contour plots for the extremal index θ_{X^U} for the GARCH(2, 2) process: left, as function of (α_1, β_1) with $\alpha_2 = \beta_2 = 0.05$; right, as function of (α_2, β_2) with $\alpha_1 = \beta_1 = 0.05$. In both panels the innovation Z_t is standard normal and the grey dashed line is the boundary of the IGARCH(2, 2).

further reduces the mean cluster size. For increasing ϕ , for $0 < \phi \leq 1$, we have increasing average cluster sizes, but that pattern does not follow when $\phi > 1$. In all cases, $\min(\theta_{X^U}, \theta_{X^V}) \geq \theta_{X^2}$, indicating the extremes of the processes $\{X_t\}$ and $\{-X_t\}$ exhibit less clustering on average than the $\{X_t^2\}$ process. We have equality, in this inequality, only when $\delta = 0$ or 1, and find that as δ tends to these limits one or other of the processes $\{X_t\}$ and $\{-X_t\}$ has similar cluster of extremes events as the $\{X_t^2\}$ process. For $\xi > 0$ more clustering occurs in the upper tail than in the lower tail of the GARCH(p, q) process, with the reverse happening with $\xi < 0$.

To give a better idea of how the GARCH(p, q) parameters affect the extremal index Figure 9 presents a contour plot of θ_{X^U} for GARCH(2,2) process over different key parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$. In particular, in each panel we hold fixed two parameters and contour over the other two key parameters of a GARCH(2,2) process. Figure 9, shows that broadly θ_{X^U} decreases, i.e., average cluster sizes increase, with increasing ϕ , up to the boundary case of 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 α_1 and α_2 tend to lead to very limited clustering. This seems logical given the GARCH formulation, as small α_1 and α_2 mean that the effect of the large X_t^2 value can have limited impact of the subsequent volatilities, so without β_1 and β_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 also have that we obtain stronger extremal dependence with larger values of the pair (α_2, β_2) than for (α_1, β_1) for equal values of ϕ , as seen by values of θ_{X^U} being smaller in Figure 9 right panel by comparison to the left panel. The reason for this is the stronger effect of β coefficients compared to that of the α as the persistency in the volatility is translated into a stronger extremal dependence.

7 Discussion

The new theory and methods we present extend to assessing the strict stationarity and extremal properties for a much broader class of stochastic recurrence equations. Specifically, for any process

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

where \mathbf{A}_t and \mathbf{B}_t are stochastic, independent and identically distributed, 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 its tail index of regular variation for the marginal distributions of \mathbf{Y}_t ; determine ways to simulate from the spectral measure from the limiting joint distribution of \mathbf{Y}_t given some norm, $\|\mathbf{Y}_t\|$, tends to infinity; and are able to generate a wide range of properties, such as the extremal index of any marginal of the process \mathbf{Y}_t from its tail chain. Previously none of these properties could be derived due to a combination of a lack of appropriate numerically stable algorithms. The methods presented here overcome these limitations and provide a broad toolbox of numerically robust approaches to derive the extremal analysis of a wide class of stochastic recurrence equations including all GARCH(p, q) processes with bounded and unbounded innovation variables.

Appendix A Proof of Theorems

Proof of Theorem 3.1. First rewrite the product of independent matrices as

$$\prod_{i=1}^t \mathbf{A}_{t+1-i} = \prod_{i=1}^t \left(\frac{\mathbf{A}_{t+1-i}}{\lambda_{t+1-i} \exp(\eta)} \right) \left(\prod_{i=1}^t \lambda_i \right) \exp(\eta t) = \mathbf{C}_t \left(\prod_{i=1}^t \lambda_i \right) \exp(\eta t). \quad (\text{A.1})$$

Thus

$$\left\| \prod_{i=1}^t \mathbf{A}_{t+1-i} \right\| = \left\| \mathbf{C}_t \left(\prod_{i=1}^t \lambda_i \right) \exp(\eta t) \right\| = \left\| \mathbf{C}_t \right\| \left(\prod_{i=1}^t \lambda_i \right) \exp(\eta t)$$

so as $t \rightarrow \infty$

$$\begin{aligned} \frac{1}{t} \ln \left\| \prod_{i=1}^t \mathbf{A}_{t+1-i} \right\| &= \ln(\|\mathbf{C}_t\|)/t + \sum_{i=1}^t \ln \lambda_i / t + \eta \\ &\rightarrow E(\ln \lambda) + \eta, \end{aligned}$$

hence $\gamma = E(\ln \lambda) + \eta$. □

Proof of Theorem 3.2. With the same notation as for Theorem 3.1 we have that

$$\begin{aligned} \|\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1\|^\kappa &= \left\| \mathbf{C}_t \left(\prod_{i=1}^t \lambda_i \right) \exp(\eta t) \right\|^\kappa \\ &= \left\| \mathbf{C}_t \right\|^\kappa \left(\prod_{i=1}^t \lambda_i^\kappa \right) \exp(\eta \kappa t). \end{aligned} \quad (\text{A.2})$$

Under condition (3.3) we have that there exists a sequence $s_t > 0$ such that $s_t \rightarrow 0$ and $s_t t \rightarrow \infty$ as $t \rightarrow \infty$

$$\exp(-s_t t) < \|\mathbf{C}_t\| < \exp(s_t t)$$

for all t . Combining these inequalities with expression (A.2) we obtain that

$$E \left(\exp(-s_t t) \left(\prod_{i=1}^t \lambda_i^\kappa \right) \right) \exp(\eta \kappa t) < E(\|\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1\|^\kappa) < E \left(\exp(s_t t) \left(\prod_{i=1}^t \lambda_i^\kappa \right) \right) \exp(\eta \kappa t).$$

Focusing first on the upper bound we have that

$$\begin{aligned} E \left(\exp(s_t t) \prod_{i=1}^t \lambda_i^\kappa \right) \exp(\eta \kappa t) &= \exp(s_t t) E \left(\prod_{i=1}^t \lambda_i^\kappa \right) \exp(\eta \kappa t) \\ &= \exp(s_t t) \{E(\lambda^\kappa)\}^t \exp(\eta \kappa t) \\ &= \exp(s_t t) \{E[(\lambda \exp(\eta))^\kappa]\}^t, \end{aligned}$$

where the second step comes from λ_i , $i = 1, \dots, t$ being independent and identically distributed. Hence, as $t \rightarrow \infty$

$$\begin{aligned} \frac{1}{t} \ln \left\{ E \left(\exp(s_t t) \left(\prod_{i=1}^t \lambda_i^\kappa \right) \right) \exp(\eta \kappa t) \right\} &= s_t + \ln \{E[(\lambda \exp(\eta))^\kappa]\} \\ &\rightarrow \ln \{E[(\lambda \exp(\eta))^\kappa]\}. \end{aligned}$$

By an identical argument, the lower bound is found to be asymptotically equal to the upper bound. Hence

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln E(\|\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_1\|^\kappa) = \ln \{E[(\lambda \exp(\eta))^\kappa]\},$$

so this limit is equal to 0, as required by condition (2.15), only when κ satisfies $E[(\lambda \exp(\eta))^\kappa] = 1$. □

Proof of Theorem 3.4. First suppose that $\kappa = 1$, then from property (2.13) we have that $E(\|\mathbf{A}\hat{\Theta}_t\|) = 1$, where \mathbf{A} is independent of $\hat{\Theta}_t$. Then, by considering the vector \mathbf{Y}_t , we see that marginally the first q and the last p components of $\hat{\Theta}_t$ each have identical marginal distributions. Furthermore, as $E(Z_t^2) = 1$, it follows that all $E(\hat{\vartheta}_t^{(i)}) = 1/(p+q)$ for all $i = 1, \dots, p+q$, where recall $\hat{\Theta}_t = (\hat{\vartheta}_t^{(1)}, \dots, \hat{\vartheta}_t^{(p+q)})$. We also have that

$$E(\mathbf{A}) = \begin{pmatrix} \alpha^{(q-1)} & \alpha_q & \beta^{(p-1)} & \beta_p \\ \mathbf{I}_{q-1} & 0_{q-1} & \mathbf{0}_{(q-1) \times (p-1)} & 0_{q-1} \\ \alpha^{(q-1)} & \alpha_q & \beta^{(p-1)} & \beta_p \\ \mathbf{0}_{(p-1) \times (q-1)} & 0_{p-1} & \mathbf{I}_{p-1} & 0_{p-1} \end{pmatrix}.$$

Let $a_{i,j}$ denote the (i,j) th element of \mathbf{A} , using the independence of $a_{i,j}$ from $\hat{\vartheta}_t^{(j)}$ then

$$\begin{aligned} E(\|\mathbf{A}\hat{\Theta}_t\|) &= E\left(\sum_{i=1}^{p+q} \sum_{j=1}^{p+q} a_{i,j} \hat{\vartheta}_t^{(j)}\right) \\ &= \sum_{i=1}^{p+q} \sum_{j=1}^{p+q} E(a_{i,j}) E(\hat{\vartheta}_t^{(j)}) \\ &= \frac{1}{p+q} \sum_{i=1}^{p+q} \sum_{j=1}^{p+q} E(a_{i,j}) \\ &= \frac{1}{p+q} \left(2 \sum_{i=1}^q \alpha_i + 2 \sum_{i=1}^p \beta_i + q - 1 + p - 1 \right) \\ &= 1 + \frac{2}{p+q} \left(\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i - 1 \right). \end{aligned}$$

Thus $E(\|\mathbf{A}\hat{\Theta}_t\|) = 1$ only when $\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i = 1$, i.e., when the process is IGARCH(p, q). The argument is simply reversed giving that $\kappa = 1$ for any IGARCH(p, q) process. \square

Acknowledgements

We would like to thank Feridun Turkman for encouraging and helpful discussions.

References

- Azzalini, A. and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B*, **65**, 367–389.
- Basrak, B., Davis, R. A. and Mikosch, T. (2002) Regular variation of GARCH processes. *Stochastic Processes and their Applications*, **99**, 95–115.
- Basrak, B. and Segers, J. (2009) Regularly varying multivariate time series. *Stochastic Processes and their Applications*, **119**, 1055–1080.
- Basrak, B. and Segers, J. (2011) Erratum to: “Regularly varying multivariate time series” [Stochastic Process. Appl. 119 (2009) 1055–1080]. *Stochastic Processes and their Applications*, **121**, 896 – 898.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.
- Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, **52**, 115–127.
- Breidt, J. F. and Davis, R. A. (1998) Extremes of stochastic volatility models. *The Annals of Applied Probability*, **8**, 664–675.

- Collamore, J. F., Diao, G. and Vidyashankar, A. N. (2014) Rare event simulation for processes generated via stochastic fixed point equations. *The Annals of Applied Probability*, **24**, 2143–2175.
- Davis, R. A. and Mikosch, T. (2009a) Extreme value theory for GARCH processes. In *Handbook of Financial Time Series* (eds. T. Andersen, R. Davis, J. Kreiss and T. Mikosch). New York: Springer, pp. 187–200.
- Davis, R. A. and Mikosch, T. (2009b) The extremogram: a correlogram for extreme events. *Bernoulli*, **4**, 977–1009.
- de Haan, L., Resnick, S. I., Rootzén, H. and de Vries, C. G. (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications*, **32**, 213–224.
- Del Moral, P. and Miclo, L. (2000) Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Seminaire de probabilités XXXIV*. Springer, pp. 1–145.
- Del Moral, P. and Miclo, L. (2003) Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups. *ESAIM: Probability and Statistics*, **7**, 171–208.
- Doucet, A., Godsill, S. and Andrieu, C. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197–208.
- Ehlert, A., Fiebig, U.-R., Janssen, A. and Schlather, M. (2015) Joint extremal behavior of hidden and observable time series with applications to GARCH processes. *Extremes*, **18**, 109–140.
- Francq, C. and Zakoïan, J.-M. (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester, United Kingdom: John Wiley & Sons, Ltd.
- Goldsheid, I. Y. (1991) Lyapunov exponents and asymptotic behaviour of the product of random matrices. *Lecture Notes in Mathematics*, **1486**, 23–37.
- Griffin, A., Jenkins, P. A., Roberts, G. O. and Spencer, S. E. (2017) Simulation from quasi-stationary distributions on reducible state spaces. *Advances in Applied Probability*, **49**, 960–980.
- Hsing, T., Hüsler, J. and Leadbetter, M. R. (1988) On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, **78**, 97–112.
- Janssen, A. (2010) *On Some Connections between Light Tails, Regular Variation and Extremes*. Ph.D. thesis, University of Gottingen.
- Janssen, A. and Segers, J. (2014) Markov tail chains. *Journal of Applied Probability*, **51**, 1133–1153.
- Kallenberg, O. (1983) *Random Measures*. Berlin: Akademic-Verlag, 3rd edn.
- Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, **131**, 207–248.
- Kesten, H. and Spitzer, F. (1984) Random difference equations and renewal theory for products of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **67**, 363–386.
- Laurini, F. and Tawn, J. A. (2012) The extremal index for GARCH(1, 1) processes. *Extremes*, **15**, 511–529.
- Ledford, A. W. and Tawn, J. A. (2003) Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society, Series B*, **65**, 521–543.
- Mikosch, T. and Stărică, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics*, **28**, 1427–1451.

- O'Brien, G. L. (1987) Extreme values for stationary and Markov sequences. *The Annals of Probability*, **15**, 281–291.
- Planinić, H. and Soulier, P. (2018) The tail process revisited. *Extremes*, **21**, 551–579. URL <https://doi.org/10.1007/s10687-018-0312-1>.
- Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.
- Rootzén, H. (1988) Maxima and exceedances of stationary Markov chains. *Advanced in Applied Probability*, **20**, 371–390.
- Segers, J. (2003) Functionals of clusters of extremes. *Advances in Applied Probability*, **35**, 1028–1045.
- Smith, R. L., Tawn, J. A. and Coles, S. G. (1997) Markov chain models for threshold exceedances. *Biometrika*, **84**, 249–268.
- Smith, R. L. and Weissman, I. (1994) Estimating the extremal index. *Journal of the Royal Statistical Society, Series B*, **56**, 515–528.
- Taylor, S. J. (1986) *Modelling Financial Time Series*. Chichester: Wiley.