On the tail behavior of a class of multivariate conditionally heteroskedastic processes

Rasmus Søndergaard Pedersen, Olivier Wintenberger

On the tail behavior of a class of multivariate conditionally heteroskedastic processes

Rasmus Søndergaard Pedersen* and Olivier Wintenberger†

January 31, 2017

Abstract

Conditions for geometric ergodicity of multivariate autoregressive conditional heteroskedasticity (ARCH) processes, with the so-called BEKK (Baba, Engle, Kraft, and Kroner) parametrization, are considered. We show for a class of BEKK-ARCH processes that the invariant distribution is regularly varying. In order to account for the possibility of different tail indices of the marginals, we consider the notion of vector scaling regular variation, in the spirit of Perfekt (1997, *Advances in Applied Probability*, 29, pp. 138-164). The characterization of the tail behavior of the processes is used for deriving the asymptotic properties of the sample covariance matrices.

AMS 2010 subject classifications: 60G70, 60G10, 60H25, 39A50.

Keywords and phrases: Stochastic recurrence equations, Markov processes, regular variation, multivariate ARCH, asymptotic properties, geometric ergodicity.

1 Introduction

We investigate the tail behavior of a class of multivariate conditionally heteroskedastic processes. Specifically, we consider the BEKK-ARCH (or BEKK(1,0,l)) process,

*Department of Economics, University of Copenhagen. Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark. Correspondence to rsp@econ.ku.dk. Phone: +45 35 32 30 74.

†Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark & Sorbonne Universités, UPMC Univ Paris 06, LSTA, Case 158 4 place Jussieu, 75005 Paris, France.

introduced by Engle and Kroner (1995), satisfying

$$X_t = H_t^{1/2} Z_t, \quad t \in \mathbb{N} \tag{1.1}$$

$$H_t = C + \sum_{i=1}^{l} A_i X_{t-1} X_{t-1}^{\intercal} A_i^{\intercal}, \tag{1.2}$$

with $(Z_t : t \in \mathbb{N})$ *i.i.d.*, $Z_t \sim N(0, I_d)$, C a $d \times d$ positive definite matrix, $A_1, ..., A_l \in M(d, \mathbb{R})$ (the set of $d \times d$ real matrices), and some initial value X_0. Due to the assumption that Z_t is Gaussian, it holds that X_t can be written as the stochastic recurrence equation (SRE)

$$X_t = \tilde{M}_t X_{t-1} + Q_t, \tag{1.3}$$

with

$$\tilde{M}_t = \sum_{i=1}^{l} m_{it} A_i \tag{1.4}$$

and $(m_{it} : t \in \mathbb{N})$ is an *i.i.d.* process mutually independent of $(m_{jt} : t \in \mathbb{N})$ for $i \neq j$, with $m_{it} \sim N(0, 1)$. Moreover $(Q_t : t \in \mathbb{N})$ is an *i.i.d.* process with $Q_t \sim N(0, C)$ mutually independent of $(m_{it} : t \in \mathbb{N})$ for all $i = 1, ..., l$.

The merit of the representation in (1.3) is two-fold. Firstly, based on the representation, we find a new sufficient condition in terms of the matrices $A_1, ..., A_l$ in order for $(X_t : t \geq 0)$ to be geometrically ergodic. In particular, for the case $l = 1$, we derive a condition directly related to the eigenvalues of A_1, in line with the strict stationarity condition found by Nelson (1990) for the univariate ARCH(1) process. Secondly, the representation is used to characterize the tails of the stationary solution to X_t.

Whereas the tail behavior of univariate GARCH processes is well-established, see e.g. Basrak et al. (2002b), few results on the tail behavior of multivariate GARCH processes exist. Some exceptions are the multivariate constant conditional correlation (CCC) GARCH processes, see e.g. Stărică (1999), Pedersen (2016), and Matsui and Mikosch (2016), and a class of factor GARCH processes, see Basrak and Segers (2009). This existing body of literature relies on rewriting the (transformed) process on companion form that obeys a non-negative multivariate SRE. The characterization of the tails of the processes then follows by an application of Kesten's Theorem (Kesten (1973)) for non-negative SREs. Such approach is not feasible when analyzing BEKK-ARCH processes, as these are stated in terms of an $\mathbb{R}^d$-valued SRE in (1.3). Importantly, we show that the classical results of Kesten (1973, Theorem 6),

see also Alsmeyer and Mentemeier (2012), for SREs in $\mathbb{R}^d$ rely on rather restrictive conditions that can be shown not to hold for certain types of BEKK-ARCH processes. Hence alternative arguments are needed. Moreover, the classical results of Kesten would imply that all marginals of X_t have the same index of regular variation (or tail index). On the contrary, we show that the BEKK-ARCH process with diagonal matrices (called Diagonal BEKK-ARCH below) exhibits different marginal tail indices. Thus, most of the paper will focus on the study of the extremal properties of this subclass of the BEKK-ARCH model. In order to analyze this class of BEKK-ARCH processes, where the tail indices are allowed to differ among the elements of X_t, we introduce a new notion of vector-scaling regular variation in the spirit of Perfekt (1997) based on element-wise scaling of X_t instead of scaling by an arbitrary norm of X_t. We show that under the crucial Assumption **M** one can still express the regular variation properties thanks to a specific "pseudo-norm" denoted $\|\cdot\|_\alpha$, see Section 2.3 for details.

The rest of the paper is organized as follows. In Section 2, we state sufficient conditions for geometric ergodicity of the BEKK-ARCH process and introduce the notion of vector-scaling regular varying distributions. We show that the distribution of X_t satisfies this type of tail-behavior, under suitable conditions. Section 3 introduces the notion of vector scaling regularly varying time series and states that certain BEKK-ARCH processes satisfy this property. Moreover, the section considers the extremal behavior of the process, in terms of the asymptotic behavior of maxima and extremal indices. In Section 4, we consider the limiting distribution of the sample covariance matrix of X_t, which relies on characterizing the tail-behavior of cross-products of the entries of X_t. Section 5 contains some concluding remarks on future research directions.

Notation: Let $GL(d,\mathbb{R})$ denote the set of $d \times d$ invertible real matrices. With $M(d,\mathbb{R})$ the set of $d \times d$ real matrices and $A \in M(d,\mathbb{R})$, let $\rho(A)$ denote the spectral radius of A. With $\otimes$ denoting the Kronecker product, for any real matrix A let $A^{\otimes p} = A \otimes A \otimes \cdots \otimes A$ (p factors). For two matrices, A and B, of the same dimension, $A \odot B$ denotes the Hadamard product of A and B. Unless stated otherwise, $\|\cdot\|$ denotes an arbitrary matrix norm. For two positive functions f and g, $f(x) \sim g(x)$, if $\lim_{x\to\infty} f(x)/g(x) = 1$. Let $\mathcal{L}(X)$ refer to the distribution of X. By default, the mode of convergence for distributions is weak convergence.

2 Stationary solution of the BEKK-ARCH model

2.1 Existence and geometric ergodicity

We start out by stating the following theorem that provides a sufficient condition for geometric ergodicity of the BEKK-ARCH process. To our knowledge, this result is new.

Theorem 2.1. *Let X_t satisfy* (1.1)-(1.2). *With $\tilde{M}_t$ defined in* (1.4), *suppose that*

$$\inf_{n\in\mathbb{N}}\left\{\frac{1}{n}\mathbb{E}\left[\log\left(\left\|\prod_{t=1}^{n}\tilde{M}_t\right\|\right)\right]\right\}<0. \tag{2.1}$$

Then $(X_t : t = 0, 1, ...)$ is geometrically ergodic, and for the associated stationary solution, $\mathbb{E}[\|X_t\|^s] < \infty$ for some $s > 0$.

Proof. Exploiting that the process can be written as the SRE (1.3)-(1.4), the proof is fairly standard, and we recall it here for completeness. We rely on the approach of Alsmeyer (2003). The Markov chain (X_t) constitutes a sequence of iterated random Lipschitz functions of the form

$$X_t = \Phi_t(X_{t-1}), \qquad t \in \mathbb{N},$$

where Φ_t are i.i.d. maps of $\mathbb{R}^d$ that are Lipschitz with respect to the Euclidian norm $\|\cdot\|$ with coefficient $\|\tilde{M}_t\|$. As $\mathbb{E}[\log^+(\|\tilde{M}_t\|)] < \infty$, a sub-additive argument implies the existence of the top-Lyapunov exponent

$$\lim_{n\to\infty}\frac{1}{n}\log(\|\tilde{M}_1\cdots\tilde{M}_n\|)=\inf_{n\in\mathbb{N}}\left\{\frac{1}{n}\mathbb{E}\left[\log\left(\left\|\prod_{t=1}^{n}\tilde{M}_t\right\|\right)\right]\right\}, \qquad a.s.$$

Under (2.1), the top-Lyapunov coefficient is negative. Under the additional condition $\mathbb{E}[\log^+(\|Q_1\|)] < \infty$, that is automatically satisfied in our context, there exists a unique stationary solution (X_t) to (1.3). Moreover, as the conditional distributions are Gaussian and so continuous, we apply Theorems 2.1 and 2.2 of Alsmeyer (2003) to assert that (X_t) is an aperiodic positive Harris chain on $\mathbb{R}^d$. Moreover, since $\mathbb{E}[\|\tilde{M}_t\|^p] < \infty$ and $\mathbb{E}[\|Q_1\|^p] < \infty$ for any $p > 0$, we apply Theorem 3.2 of Alsmeyer (2003) to conclude that (X_t) is geometrically ergodic and that $\mathbb{E}[\|X_t\|^s] < \infty$ for some $s > 0$. □

Remark 2.2. Sufficient conditions for higher-order moments of X_t can be obtained from Theorem 5 of Feigin and Tweedie (1985). In particular, if $\rho(E[\tilde{M}_t^{\otimes 2n}]) < 1$ for some $n \in \mathbb{N}$, then, for the strictly stationary solution, $E[\|X_t\|^{2n}] < \infty$. For

example, $\rho(\sum_{i=1}^{l} A_i^{\otimes 2}) < 1$ implies that $E[\|X_t\|^2] < \infty$. This result complements Theorem C.1 of Pedersen and Rahbek (2014) that contains conditions for finite higher-order moments for the case $l = 1$.

For the case where $\tilde{M}_t$ contains only one term, i.e. $l = 1$, the condition in (2.1) simplifies and a condition for geometric ergodicity can be stated explicitly in terms of the eigenvalues of the matrix A_1:

Proposition 2.3. *Let X_t satisfy* (1.1)-(1.2) *with $l = 1$ and let $A := A_1$. Then a necessary and sufficient condition for* (2.1) *is that*

$$\rho(A) < \exp\left\{\frac{1}{2}\left[-\psi(1) + \log(2)\right]\right\} = 1.88736..., \tag{2.2}$$

where $\psi(\cdot)$ is the digamma function.

Proof. The condition (2.1) holds if and only if there exists $n \in \mathbb{N}$ such that

$$\mathbb{E}\left[\log\left(\left\|\prod_{t=1}^{n} M_t\right\|\right)\right] < 0. \tag{2.3}$$

Let $m_t := m_{1t}$. It holds that

$$\begin{aligned}
\mathbb{E}\left[\log\left(\left\|\prod_{t=1}^{n} M_t\right\|\right)\right] &= \mathbb{E}\left[\log\left(\left\|A^n \prod_{t=1}^{n} m_t\right\|\right)\right] \\
&= \log\left(\|A^n\|\right) - n\mathbb{E}\left[-\log(|m_t|)\right] \\
&= \log\left(\|A^n\|\right) - n\left\{\frac{1}{2}\left[-\psi(1) + \log(2)\right]\right\},
\end{aligned}$$

and hence (2.3) is satisfied if

$$\log\left(\|A^n\|^{1/n}\right) < \frac{1}{2}\left[-\psi(1) + \log(2)\right].$$

The result now follows by observing that $\|A^n\|^{1/n} \to \rho(A)$ as $n \to \infty$. □

Remark 2.4. It holds that $\rho(A \otimes A) = (\rho(A))^2$. Hence the condition in (2.2) is equivalent to

$$\rho(A \otimes A) \;<\; \exp\left\{-\psi(1) + \log(2)\right\} = \frac{1}{2}\exp\left[-\psi\left(\frac{1}{2}\right)\right] = 3.56...,$$

which is similar to the strict stationary condition found for the ARCH coefficient of the univariate ARCH(1) process with Gaussian innovations; see Nelson (1990).

Below we provide some examples of BEKK-ARCH processes that are geometrically ergodic and that will be studied in detail throughout this paper.

Example 2.5 (**ID BEKK-ARCH**). Following Alsmeyer and Mentemeier (2012), we consider BEKK processes with corresponding SRE's satisfying certain irreducibility and density conditions (ID), that is conditions **(A4)-(A5)** in Section A.1 in the appendix. Specifically, we consider the *bivariate* BEKK-ARCH process in (1.1)-(1.2) with

$$H_t \;=\; C + \sum_{i=1}^{4} A_i X_{t-1} X_{t-1}^{\intercal} A_i^{\intercal},$$

where

$$A_1 = \begin{pmatrix} a_1 & 0 \\ 0 & 0 \end{pmatrix} \quad A_2 = \begin{pmatrix} 0 & 0 \\ a_2 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & a_3 \\ 0 & 0 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 0 & 0 \\ 0 & a_4 \end{pmatrix} \tag{2.4}$$

and

$$a_1, a_2, a_3, a_4 \neq 0. \tag{2.5}$$

Writing X_t as an SRE, we obtain

$$X_t = \tilde{M}_t X_{t-1} + Q_t, \tag{2.6}$$

with

$$\tilde{M}_t = \sum_{i=1}^{4} A_i m_{it} \tag{2.7}$$

where $(m_{1t}), (m_{2t}), (m_{3t}), (m_{4t})$ are mutually independent i.i.d. processes with $m_{it} \sim N(0,1)$. Assuming that a_1, a_2, a_3, a_4 are such that the top Lyapunov exponent of $(\tilde{M}_t)$ is strictly negative, we have that the process is geometrically ergodic.

Notice that one could consider a more general d-dimensional process with the same structure as in (2.4)-(2.7), but with $\tilde{M}_t$ containing d^2 terms such that $\tilde{M}_t$ has a Lebesgue density on $M(d, \mathbb{R})$, as clarified in Example 2.10 below. Moreover, one could include additional terms to $\tilde{M}_t$, say a term containing a full matrix A or an autoregressive term, as presented in Remark 2.8 below. We will focus on the simple bivariate process, but emphasize that our results apply to more general processes. □

Example 2.6 (**Similarity BEKK-ARCH**). Consider the BEKK process in (1.1)-(1.2) with $l = 1$ and $A := A_1 = aO$, where a is a positive scalar and O is an orthogonal matrix. This implies that the SRE (1.3) has $\tilde{M}_t = a|m_t|\text{sign}(m_t)O$. By

definition, $\tilde{M}_t$ is a *similarity* with probability one, where we recall that a matrix is a similarity if it can be written as a product of a positive scalar and an orthogonal matrix. From Proposition 2.3, we have that if $a < \exp\{(1/2)[-\psi(1) + \log(2)]\} = 1.88736...$, then the process is geometrically ergodic.

An important process satisfying the similarity property is the well-known scalar BEKK-ARCH process, where $H_t = C + aX_{t-1}X_{t-1}^{\intercal}$, $a > 0$. Here $A = \sqrt{a}I_d$, with I_d the identity matrix. □

Example 2.7 (**Diagonal BEKK-ARCH**)**.** Consider the BEKK-ARCH process in (1.1)-(1.2) with $l = 1$ such that $A := A_1$ is diagonal. We refer to this process as the Diagonal BEKK-ARCH process. Relying on Proposition 2.3, we have that the process is geometrically ergodic, if each diagonal element of A is less than $\exp\{(1/2)[-\psi(1) + \log(2)]\} = 1.88736...$ in modulus.

As discussed in Bauwens et al. (2006), diagonal BEKK models are typically used in practice, e.g. within empirical finance, due to their relatively simple parametrization. As will be shown below, even though the parametrization is simple, the tail behavior is rather rich in the sense that each marginal of X_t has different tail indices, in general. □

Remark 2.8. As an extension to (1.1)-(1.2), one may consider the autoregressive BEKK-ARCH (AR BEKK-ARCH) process

$$\begin{aligned} X_t &= A_0 X_{t-1} + H_t^{1/2} Z_t, \quad t \in \mathbb{N} \\ H_t &= C + \sum_{i=1}^{l} A_i X_{t-1} X_{t-1}^{\intercal} A_i^{\intercal}, \end{aligned}$$

with $A_0 \in M(\mathbb{R}, d)$. This process has recently been studied and applied by Nielsen and Rahbek (2014) for modelling the term structure of interest rates. Notice that the process has the SRE representation

$$X_t = \tilde{M}_t X_{t-1} + Q_t, \quad \tilde{M}_t = A_0 + \sum_{i=1}^{l} m_{it} A_i.$$

Following the arguments used for proving Theorem 2.1, it holds that the (AR BEKK-ARCH) process is geometrically ergodic if condition (2.1) is satisfied. Interestingly, as verified by simulations in Nielsen and Rahbek (2014) the Lyapunov condition may hold even if the autoregressive polynomial has unit roots, i.e. if $A_0 = I_d + \Pi$, where $\Pi \in M(\mathbb{R}, d)$ has reduced rank.

2.2 Multivariate regularly varying distributions

The stationary solution of the BEKK-ARCH process (see Theorem 2.1) can be written as

$$X_t = \sum_{i=0}^{\infty} \prod_{j=1}^{i} \tilde{M}_{t-j+1} Q_{t-i}, \quad t \in \mathbb{Z}. \tag{2.8}$$

Even if the random matrices $\tilde{M}_t$ are light-tailed under the Gaussian assumption, the maximum of the products $(\prod_{t=1}^{T} \tilde{M}_t)_{T\geq 0}$ may exhibit heavy tails when $T \to \infty$. More precisely, the tails of the stationary distribution are suspected to have an extremal behavior as a power law function: for any $u \in \mathbb{S}^{d-1}$

$$\mathbb{P}(u^\intercal X_0 > x) \sim C(u) x^{-\alpha(u)}, \qquad x \to \infty, \tag{2.9}$$

for $C(u) > 0$ and $\alpha(u) > 0$. Cases $\alpha(u) = \alpha$, $u \in \mathbb{S}^{d-1}$ are referred as Kesten's cases, because of the seminal paper Kesten (1973), and are the subject of the monograph by Buraczewski et al. (2016). A class of multivariate distributions satisfying this property is the class of multivariate regularly varying distributions (de Haan and Resnick (1977)):

Definition 2.9. Let $\bar{\mathbb{R}}_0^d := \bar{\mathbb{R}}^d \setminus \{0\}$, $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, and $\bar{\mathcal{B}}_0^d$ be the Borel σ-field of $\bar{\mathbb{R}}_0^d$. For an $\mathbb{R}^d$-valued random variable X and some constant scalar $x > 0$, define $\mu_x(\cdot) := \mathbb{P}(x^{-1}X \in \cdot)/\mathbb{P}(\|X\| > x)$. Then X and its distribution are *multivariate regularly varying* if there exists a non-null Radon measure μ on $\bar{\mathcal{B}}_0^d$ which satisfies

$$\mu_x(\cdot) \to \mu(\cdot) \qquad \text{vaguely, as } x \to \infty. \tag{2.10}$$

For any μ-continuity set C and $t > 0$, $\mu(tC) = t^{-\alpha}\mu(C)$, and we refer to α as the index of regular variation.

We refer to de Haan and Resnick (1977) for the notion of vague convergence and additional details. Below, we provide two examples of multivariate regularly varying BEKK processes.

Example 2.10 (**ID BEKK-ARCH**, continued)**.** Consider the ID BEKK-ARCH process (2.4)-(2.7) from Example 2.5. By verifying the conditions of Theorem 1.1 of Alsmeyer and Mentemeier (2012) (stated in Section A.1 in the appendix), we establish that the process is multivariate regularly varying.

Since $(m_{1t}, m_{2t}, m_{3t}, m_{4t})$ and Q_t are Gaussian, we have that **(A1)**-**(A2)** hold.

Moreover,

$$\tilde{M}_t = \begin{pmatrix} a_1 m_{1t} & a_3 m_{3t} \\ a_2 m_{2t} & a_4 m_{4t} \end{pmatrix} \tag{2.11}$$

is invertible with probability one, which ensures that **(A3)** is satisfied. From (2.11) we also notice that the distribution of $\tilde{M}_t$ has a Lebesgue density on $M(d, \mathbb{R})$ which is strictly positive in a neighborhood of I_2. This ensures that the irreducibility and density conditions **(A4)**-**(A5)** are satisfied. The fact that $Q_t \sim N(0, C)$ and independent of $\tilde{M}_t$ implies that condition **(A6)** holds. Lastly, condition **(A7)** holds by the fact that $(m_{1t}, m_{2t}, m_{3t}, m_{4t})$ and Q_t are Gaussian. By Theorem 1.1 of Alsmeyer and Mentemeier (2012) we have established the following proposition:

Proposition 2.11. *Let X_t satisfy* (2.4)-(2.7) *such that the top Lyapunov exponent of $(\tilde{M}_t)$ is strictly negative. Then for the stationary solution (X_t), there exists $\alpha > 0$ such that*

$$\lim_{t\to\infty} t^{\alpha} \mathbb{P}(x^{\intercal} X_0 > t) = K(x), \quad x \in \mathbb{S}^1, \tag{2.12}$$

for some finite, positive, and continuous function K on $\mathbb{S}^1$.

The proposition implies that each marginal of the distribution of X_0 is regularly varying of order α. By Theorem 1.1.(ii) of Basrak et al. (2002a), we conclude that X_0 is multivariate regularly varying whenever α is a non-integer. Moreover, since X_0 is symmetric, the multivariate regular variation does also hold if α is an odd integer, see Remark 4.4.17 in Buraczewski et al. (2016).

The proposition does also apply if $a_1 = 0$ or $a_4 = 0$. This can be seen by observing that $\prod_{k=1}^{n} \tilde{M}_k$ has a strictly positive density on $M(d, \mathbb{R})$ for n sufficiently large, which is sufficient for establishing conditions **(A4)**-**(A5)**. □

Example 2.12 (**Similarity BEKK-ARCH**, continued)**.** The Similarity BEKK-ARCH, introduced in Example 2.6, fits into the setting of Buraczewski et al. (2009), see also Section 4.4.10 of Buraczewski et al. (2016). Specifically, using the representation $\tilde{M}_t = a|m_t|\text{sign}(m_t)O$, we have that

(i) $\mathbb{E}[\log(|m_t a|)] < 0$ if $a < \exp\{(1/2)[-\psi(1) + \log(2)]\}$,

(ii) $\mathbb{P}(\tilde{M}_t x + Q_t = x) < 1$ for any $x \in \mathbb{R}^d$, and

(iii) $\log(|am_t|)$ has a non-arithmetic distribution.

Then, due to Theorem 1.6 of Buraczewski et al. (2009), we have the following proposition:

Proposition 2.13. *Let X_t satisfy* (1.1)-(1.2) *with $l = 1$ such that $A := A_1 = aO$, where $a > 0$ and O is an orthogonal matrix. If $a < \exp\{(1/2)[-\psi(1) + \log(2)]\} = 1.88736...$, then the process has a unique strictly stationary solution (X_t) with X_t multivariate regularly varying with index $\alpha > 0$ satisfying $E[(|m_t|a)^\alpha] = 1$.* □

In the following example, we clarify that the Diagonal BEKK-ARCH process, introduced in Example 2.7, does not satisfy the conditions of Theorem 1.1 of Alsmeyer and Mentemeier (2012). Moreover, we argue that the marginals may have different tail indices, which motivates the notion of *vector scaling regular variation*, introduced in the next section.

Example 2.14 (**Diagonal BEKK-ARCH**, continued)**.** Consider the BEKK-ARCH process in (1.1)-(1.2) with $l = 1$ such that $A := A_1$, $m_t := m_{1t}$, and $M_t := \tilde{M}_t$. Suppose that $A \in GL(d, \mathbb{R})$. For this process, the distribution of M_t is too restricted in order to apply the results by Alsmeyer and Mentemeier (2012), as in Example 2.10. Specifically, the irreducibility condition **(A4)** in Appendix A.1 can be shown not to hold, as clarified next. It holds that

$$\begin{aligned}\mathbb{P}\left\{\|x^\intercal \prod_{k=1}^{n} M_k\|^{-1}\left(x^\intercal \prod_{k=1}^{n} M_k\right) \in U\right\} &= \mathbb{P}\left\{|\prod_{k=1}^{n} m_k|^{-1}\|x^\intercal A^n\|^{-1}\left(\prod_{k=1}^{n} m_k\right) x^\intercal A^n \in U\right\} \\ &= \mathbb{P}\left\{\operatorname{sign}\left(\prod_{k=1}^{n} m_k\right)\|x^\intercal A^n\|^{-1} x^\intercal A^n \in U\right\}.\end{aligned}$$

Hence for any $x \in \mathbb{S}^{d-1}$ we can always find a non-empty open $U \subset \mathbb{S}^{d-1}$ such that

$$\max_{n\in\mathbb{N}} \mathbb{P}\left\{\operatorname{sign}\left(\prod_{k=1}^{n} m_k\right)\|x^\intercal A^n\|^{-1} x^\intercal A^n \in U\right\} = 0. \tag{2.13}$$

As an example, for $d = 2$ and A diagonal, choose $x = (1, 0)^\intercal$. Then $\|x^\intercal A^n\|^{-1} x^\intercal A^n \in \{(-1, 0)\} \cup \{(1, 0)\}$ for any $n \in \mathbb{N}$. We conclude that condition **(A4)** does not hold for the diagonal BEKK-ARCH process.

When A is diagonal, each element of $X_t = (X_{t,1}, ..., X_{t,d})'$ can be written as an SRE, given by

$$X_{t,i} = A_{ii} m_t X_{t-1,i} + Q_{t,i}, \qquad t \in Z, \qquad i = 1, \ldots, d.$$

By Theorem 4.1 of Goldie (1991), the stationary solution of the marginal equation exists if and only if $\mathbb{E}[\log(|A_{ii} m_0|)] < 0$. In that case there exists a unique $\alpha_i > 0$

such that $\mathbb{E}[|m_0|^{\alpha_i}] = |A_{ii}|^{-\alpha_i}$ and

$$\mathbb{P}(\pm X_{0,i} > x) \sim c_i x^{-\alpha_i} \qquad \text{where} \qquad c_i = \frac{\mathbb{E}[|X_{1,i}|^{\alpha_i} - |A_{ii}m_1 X_{0,i}|^{\alpha_i}]}{2\alpha_i \mathbb{E}[|A_{ii}m_1|^{\alpha_i}\log(|A_{ii}m_1|)]}.$$

Hence each marginal of X_0 may in general have different tail indices. More precisely, the tail indices are different if the diagonal elements of A, i.e. the A_{ii}s, are, and the heaviest marginal tail index α_{i_0} corresponds to the largest diagonal coefficient $A_{i_0 i_0}$. When i_0 is unique, the distribution X_0 can be considered as multivariate regularly varying with index α_{i_0} and with a limit measure μ with degenerate marginals $i \neq i_0$.□

2.3 Vector scaling regularly varying distributions

The previous Example 2.14 shows that the Diagonal BEKK-ARCH process fits into the case where $\alpha(u)$ in (2.9) is non-constant. Such cases have not attracted much attention in the existing body of literature. However, recent empirical studies, such as Matsui and Mikosch (2016), may suggest that it is more realistic to consider different marginal tail behaviors when modelling multidimensional financial observations. The idea is to use a vector scaling instead of the scaling $\mathbb{P}(\|X\| > x)$ in Definition 2.9 that reduced the regular variation properties of the vector X to the regular variation properties of the norm $\|X\|$ only. More precisely, let (X_t) be a stationary process in $\mathbb{R}^d$ and let $x = (x_1, \ldots, x_d)^\intercal \in \mathbb{R}^d$. Denote also $x^{-1} = (x_1^{-1}, \ldots, x_d^{-1})^\intercal$. We parameterize x by t such that $x_i(t) \to \infty$ as $t \to \infty$ for any $1 \leq i \leq d$ as in de Haan and Resnick (1977):

Definition 2.15. The distribution of X_0 is vector scaling regularly varying if and only if there exists a normalizing sequence $x(t)$ and a Radon measure μ with non-null marginals such that

$$t\mathbb{P}(x(t)^{-1} \odot X_0 \in \cdot) \to \mu(\cdot), \qquad \text{vaguely.} \tag{2.14}$$

If the vague convergence holds, necessarily the normalizing sequences $x_i(t)$, $1 \leq i \leq d$, are regularly varying of order α_i:

$$x_i(t) \sim t^{1/\alpha_i}\ell_i(t), \qquad t \to \infty,$$

where ℓ_i is a slowly varying function satisfying $\ell_i(bt)/\ell_i(t) \to 1$ as $t \to \infty$ for any $b > 0$. The limiting measure μ is also α_i-homogeneous with respect to the ith coordinate.

Remark 2.16. A vector scaling regularly varying distribution is also multivariate

regularly varying of index $\alpha = \min_{1\leq k\leq d} \alpha_k$. Then the limiting measure μ in (2.10) is degenerate at any marginal for which $\alpha_k > \alpha$.

Remark 2.17. In light of the above remark and Examples 2.10 and 2.12, the stationary solution of the ID and Similarity BEKK-ARCH processes satisfy that X_0 is vector scaling regularly varying with the same index of regular variation and the same slowly varying function for each marginal.

In order to extend the vector scaling regular variation definition to stationary sequences it is useful to rephrase the definition in terms of convergence in distribution. It is usually done thanks to the polar decomposition; see Theorem 3 of de Haan and Resnick (1977). In the vector scaling regularly varying setting the natural radius notion should be the function $N : \mathbb{R}^d \to \mathbb{R}_+$ having its ith marginal α_i-regularly varying so that $N(x(t)) = t$. However, this function is in general not marginally homogenous due to the slowly varying function $\ell_i(t)$. To circumvent this issue, we make the following assumption

Assumption M Each marginal of X_0 is regularly varying of order $\alpha_i > 0, i = 1, ..., d$. If the α_is are not equal, the slowly varying functions $\ell_i(t) \to c_i$ as $t \to \infty$, for $c_i > 0$ and $i = 1, ..., d$.

Remark 2.18. As explained above, it is reasonable in our setting to assume that $\ell_i(t) \sim c_i$, $1 \leq i \leq d$. The Diagonal BEKK-ARCH process introduced in Example 2.14 satisfies Assumption **M**. Moreover, any regularly varying distribution satisfying the Kesten property (2.9) satisfies Assumption **M**. Finally, any multivariate regularly varying distribution with non-degenerate marginal satisfies Assumption **M**. In particular, the ID and Similarity BEKK-ARCH processes, introduced in Examples 2.5 and 2.6 respectively, satisfy Assumption **M**.

Under Assumption **M**, the natural radius notion can be chosen as $N = \|\cdot\|_\alpha$, where

$$\|x\|_\alpha := c_i^{-1} \max_{1\leq k\leq d} |x_k|^{\alpha_k}, \tag{2.15}$$

will play a crucial role in the following. Notice that the homogeneity of its marginals will be essential for the proof.

Proposition 2.19. *Under Assumption* **M**, *the random vector X_0 is vector scaling regularly varying if and only if there exists some $Y_0 \in \mathbb{R}^d$ with non degenerate marginals such that*

$$\mathcal{L}(((c_i t)^{-1/\alpha_i})_{1\leq i\leq d} \odot X_0 \mid \|X_0\|_\alpha > t) \to_{t\to\infty} \mathcal{L}(Y_0), \tag{2.16}$$

where $\|\cdot\|_\alpha$ is defined in (2.15). *Moreover $\|Y_0\|_\alpha$ is standard Pareto distributed.*

Proof. Adapting Theorem 4 of de Haan and Resnick (1977), the definition of vector scaling regularly varying distribution of X_0 in (2.14) implies (2.16). Conversely, under Assumption **M**, we have that $|X_{0,k}|^{\alpha_k}$ is regularly varying of order 1 for all $1 \leq k \leq d$ with slowly varying functions $\ell_i(t) \sim c_i$. Moreover $\|X_0\|_\alpha$ is regularly varying from the weak convergence in (2.16) applied on the Borel sets $\{\|X_0\|_\alpha > ty\}$, $y \geq 1$. Thus, $\|X_0\|_\alpha$ is regularly varying of order 1 with slowly varying function $\ell(t)$. One can rewrite (2.16) as

$$\ell(t)^{-1} t \mathbb{P}(x(t)^{-1} \odot X_0 \in \cdot, \|X_0\|_\alpha > t) \to \mathbb{P}(Y_0 \in \cdot).$$

Using the slowly varying property of ℓ, we obtain, for any $\epsilon > 0$,

$$\ell(t)^{-1} t \mathbb{P}(x(t)^{-1} \odot X_0 \in \cdot, \|X_0\|_\alpha > t\epsilon) \to \epsilon^{-1} \mathbb{P}(Y_0 \in \cdot).$$

By marginal homogeneity of $\|\cdot\|_\alpha$ one can rewrite it as

$$\ell(t)^{-1} t \mathbb{P}(x(t)^{-1} \odot X_0 \in \cdot, \|x(t)^{-1} \odot X_0\|_\alpha > \epsilon) \to \epsilon^{-1} \mathbb{P}(Y_0 \in \cdot).$$

Notice that $\ell(t)t^{-1} > 0$ is non increasing as it is the tail of $N(X_0)$. So there exists a change of variable $t = h(t')$ so that $\ell(t)^{-1}t = t'$ and

$$t' \mathbb{P}(x(h(t'))^{-1} \odot X_0 \in \cdot, \|x(h(t'))^{-1} \odot X_0\|_\alpha > \epsilon) \to \epsilon^{-1} \mathbb{P}(Y_0 \in \cdot).$$

We obtain the existence of μ for $x' = x \circ h$ in (2.14) such that $\mu(\cdot, \|x\|_\alpha > \epsilon) = \mathbb{P}(\cdot)$, which is enough to characterize μ entirely, choosing $\epsilon > 0$ as small as necessary. □

Next, we consider the spectral properties of X_0, satisfying Proposition 2.19. Notice that for any $u \in \{+1, 0, -1\}^d$,

$$\lim_{t \to \infty} \mathbb{P}\left(\max_{1 \leq i \leq d} (c_i u_i X_{0,i})_+^{\alpha_k} > t \mid \|X_0\|_\alpha > t \right) = c_+(u) > 0.$$

Hence, by a continuous mapping argument, $(c \odot u \odot X_0)_+^\alpha$, with $c = (c_1, \ldots, c_d)^\intercal$, satisfies

$$\mathcal{L}(t^{-1}(c \odot u \odot X_0)_+^\alpha \mid \|(u \odot X_0)_+\|_\alpha > t) \to_{t \to \infty} \mathcal{L}(c_+(u)^{-1}(u \odot Y_0)_+^\alpha), \tag{2.17}$$

and $(c \odot u \odot X_0)_+^\alpha$ is regularly varying of index 1. By homogeneity of the limiting measure in the multivariate regular variation (2.10), we can also decompose the limit

as a product

$$\frac{\mathbb{P}((\|(c\odot u\odot X_0)_+\|_\alpha > ty, (c\odot u\odot X_0)_+^\alpha/\|(c\odot u\odot X_0)_+\|_\alpha \in \cdot)}{\mathbb{P}(\|(c\odot u\odot X_0)_+\|_\alpha > t)} \to y^{-\alpha}\mathbb{P}_{\Theta_u}(\cdot),$$

for any $y \geq 1$. Such limiting distribution is called a simple max-stable distribution and $\mathbb{P}_{\Theta_u}$, supported by the positive orthant, is called the spectral measure of $(u \odot X_0)_+^\alpha$, see de Haan and Resnick (1977) for more details. By identification of the two expressions of the same limit, we obtain the following proposition.

Proposition 2.20. *Under Assumption* **M**, *with* Y_0 *defined in Proposition 2.19, the distribution of* $(u \odot Y_0)_+^\alpha/\|(u \odot Y_0)_+\|_\alpha$, *if non-degenerate, is the spectral measure of* $(c \odot u \odot X_0)_+^\alpha \in [0,\infty)^d$. *Moreover, it is independent of* $\|(u \odot Y_0)_+\|_\alpha$, *and* $c_+(u)^{-1}\|(u \odot Y_0)_+\|_\alpha$ *is standard Pareto distributed.*

Proof. That $c_+(u)^{-1}\|(u \odot Y_0)_+\|_\alpha$ is standard Pareto distributed follows from the convergence in (2.17) associated with the regularly varying property, ensuring the homogeneity of the limiting measure. Then, using again the homogeneity in (2.17), it follows that $(u\odot Y_0)_+^\alpha/\|(u\odot Y_0)_+\|_\alpha$ and $c_+(u)^{-1}\|(u\odot Y_0)_+\|_\alpha$ are independent. □

Example 2.21 (**Diagonal BEKK-ARCH**, continued)**.** From Example 2.14 we have that each element of X_0, $X_{0,i}$, belongs to the domain of attraction of the Fréchet distribution with index α_i and slowly varying function $\ell_i(t) \sim c_i > 0$. Thus, it satisfies Assumption **M**. However, we have not been able to establish that X_0 belongs to the multivariate domain of attraction of the Fréchet distribution. In Section A.2 in the appendix we have included some estimates of the spectral measure of X_0 for the bivariate case. The plots suggest that the tails of the process are indeed dependent. They might be vector scaling regularly varying. □

3 Vector-scaling regularly varying time series and their extremal behavior

Proposition 2.19 from the previous section allows us to extend the asymptotic results of Perfekt (1997) to random vectors taking possibly negative values. In this section, we adapt the definition of a tail chain defined by Perfekt (1997) to the context of eventually different tail indices of the marginals.

3.1 Vector scaling regularly varying time series

In this section we introduce a new notion of multivariate regularly varying time series in the spirit of Basrak and Segers (2009) based on vector scaling of X_t. Let

$|X_0| = (|X_{0,1}|, ..., |X_{0,d}|)^\intercal$.

Definition 3.1. Under Assumption **M**, the stationary process (X_t) is vector scaling regularly varying if and only if there exists a process $(Y_t)_{t\geq 0}$ with Y_t non-degenerate for all t, such that

$$\mathcal{L}((c_i t)^{-1/\alpha_i})_{1\leq i\leq d} \odot (X_0, X_1, \ldots, X_k) \mid \|X_0\|_\alpha > t) \to_{t\to\infty} \mathcal{L}(Y_0, \ldots, Y_k),$$

for all $k \geq 0$. The sequence $(Y_t)_{t\geq 0}$ is called the *tail process*.

Remark 3.2. We recover the usual notion of multivariate regularly varying stationary processes of Basrak and Segers (2009) when $\alpha_i = \alpha$, for all $1 \leq i \leq d$.

Following Basrak and Segers (2009), we extend the notion of spectral measure to the one of *spectral processes* for any vector scaling regularly varying stationary processes:

Definition 3.3. Under Assumption **M**, the vector scaling regularly varying stationary process (X_t) admits the spectral process (Θ_t) if and only if

$$\mathcal{L}(\|X_0\|_\alpha^{-1}(X_0, X_1, \ldots, X_k) \mid \|X_0\|_\alpha > t) \to_{t\to\infty} \mathcal{L}(\Theta_0, \ldots, \Theta_k),$$

for all $k \geq 0$.

By arguments similar to the ones given above, it follows that the vector scaling regular variation properties also characterise the spectral process of $((c\odot u\odot X_t)_+^\alpha)_{t\geq 0}$, with X_0 following the stationary distribution, which has the distribution of $((u \odot Y_t)_+^\alpha / \|(u \odot Y_0)_+\|_\alpha)_{t\geq 0}$. We have the following proposition.

Proposition 3.4. *Under Assumption* **M***, the spectral process of any non-degenerate* $((c \odot u \odot X_t)_+^\alpha)_{t\geq 0}$ *is distributed as* $((u \odot Y_t)_+^\alpha / \|(u \odot Y_0)_+\|_\alpha)_{t\geq 0}$ *and independent of* $\|(u \odot Y_0)_+\|_\alpha$*. Moreover* $c_+(u)^{-1}\|(u \odot Y_0)_+\|_\alpha$ *is standard Pareto distributed.*

3.2 The tail chain

In the following, we will focus on the dynamics of the tail process $(Y_t)_{t\geq 1}$ in Definition 3.1, given the existence of Y_0. We will restrict ourselves to the case where (X_t) is a Markov chain, which implies that (Y_t) is also a Markov chain called the *tail chain*; see Perfekt (1997). We have the following proposition.

Proposition 3.5. *Let* (X_t) *satisfy* (1.1)-(1.2) *such that the stationary distribution is vector scaling regularly varying satisfying Assumption* **M***. With* $\tilde{M}_t$ *defined in* (1.4)*,*

the tail process, (Y_t), admits the multiplicative form

$$Y_{t+1} = \tilde{M}_{t+1} Y_t, \qquad t \geq 0. \tag{3.1}$$

Proof. Following the approach of Janssen and Segers (2014), one first notices that the existence of the kernel of the tail chain does not depend on the marginal distribution. Thus the characterization of the kernel extends automatically from the usual multivariate regular variation setting to the vector scaling regular variation one. It is straightforward to check Condition 2.2 of Janssen and Segers (2014). We conclude that the tail chain has multiplicative structure in (3.1). □

In light of the tail process dynamics in (3.1), under the existence of Y_0, we have that

$$Y_t^\intercal = Y_0^\intercal \prod_{i=1}^{t} \tilde{M}_i.$$

3.3 Asymptotic behavior of the maxima

From the previous section, we have that the tail chain (Y_t) quantifies the extremal behavior of (X_t) in (1.1)-(1.2). Let us consider the asymptotic behavior of the component-wise maxima

$$\max(X_1, \ldots, X_n) = \max(X_{1,k}, \ldots, X_{n,k})_{1 \leq k \leq d}.$$

Let $u = (1, \ldots, 1) = \mathbf{1} \in \mathbb{R}^{\mathbf{d}}$ and assume that $c_+(\mathbf{1}) = \lim_{t\to\infty} \mathbb{P}(X_0 \not\leq x(t) \mid |X_0| \not\leq x(t))$ is positive. Recall that for (X_t) *i.i.d.*, the suitably scaled maxima converge to the Fréchet distribution; see de Haan and Resnick (1977), i.e. for any $x = (x_1, \ldots, x_d)^\intercal \in \mathbb{R}_+^d$, defining $u_n(x)$ such that $n\mathbb{P}(X_{0,i} > u_{n,i}(x)) \sim x_i^{-1}$, $1 \leq i \leq d$, we have

$$\mathbb{P}(\max(X_1, \ldots, X_n) \leq u_n(x)) \to \exp(-A^*(x)),$$

if and only if $(X_0)_+$ is vector scaling regularly varying. In such case, due to Assumption **M**, one has the expression

$$A^*(x) = c_+(\mathbf{1}) \mathbb{E}\left[\frac{1}{\|(Y_0)_+\|_\alpha} \max_{1 \leq i \leq d} \frac{(Y_{0,i})_+^{\alpha_k}}{c_i x_i} \right]. \tag{3.2}$$

Theorem 3.6. *Let X_t satisfy* (1.1)-(1.2). *With $\tilde{M}_t$ defined in* (1.4), *suppose that condition* (2.1) *holds. Suppose that the stationary distribution is vector scaling regularly varying satisfying Assumption* **M**. *Assuming the existence of Y_0 in Definition*

3.1, we have that

$$\mathbb{P}(\max(X_m, \ldots, X_n) \leq u_n(x)) \to \exp(-A(x)),$$

where $A(x)$ *admits the expression*

$$c_+(\mathbf{1})\mathbb{E}\left[\max_{1\leq i\leq d} \frac{\max_{k\geq 0}\left(\left(Y_0^\intercal \prod_{1\leq j\leq k} \tilde{M}_j\right)_i\right)_+^{\alpha_k}}{\|(Y_0)_+\|_\alpha c_i x_i} - \max_{1\leq i\leq d} \frac{\max_{k\geq 1}\left(\left(Y_0^\intercal \prod_{1\leq j\leq k} \tilde{M}_j\right)_i\right)_+^{\alpha_k}}{\|(Y_0)_+\|_\alpha c_i x_i}\right] \quad (3.3)$$

Proof. We verify the conditions of Theorem 4.5 of Perfekt (1997). Condition B2 of Perfekt (1997) is satisfied under the more tractable Condition 2.2 of Janssen and Segers (2014). Indeed, the tail chain depends only on the Markov kernel and one can apply Lemma 2.1 of Janssen and Segers (2014), because it extends immediately to the vector scaling regularly varying setting. Condition $\mathrm{D}(\mathrm{u_n})$ of Perfekt (1997) holds by geometric ergodicity of the Markov chain for a separate sequence $C \log n$, with $C > 0$ sufficiently large. Lastly, the finite clustering condition,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}[\max(|X_m|, \ldots, |X_{C\log n}|) \not\leq u_n(x) \mid |X_0| \not\leq u_n(x)] = 0, \quad (3.4)$$

holds for any $C > 0$ using the proof of Theorem 7.2 of Mikosch and Wintenberger (2014) under the drift condition (DC_p) for $p < \alpha = \min\{\alpha_i : 1 \leq i \leq d\}$ and as (X_t) is α regularly varying. Finally, as (3.4) is a special case of Condition $\mathrm{D}^\infty(\mathrm{c\log n})$ of Perfekt (1997), we obtain the desired result with the characterization given in Theorem 4.5 of Perfekt (1997)

$$A(x) = \int_{(0,\infty)^d \setminus (0,x)} \mathbb{P}\left(T_j \leq x, k \geq 1 \mid T_0 = y\right) \nu(dy),$$

where $(T_k)_{k\geq 0}$ is the tail chain of the standardized Markov chain $(c_i^{-1}(X_{k,i})_+^{\alpha_i})_{1\leq i\leq d}$, $k \geq 0$. As μ restricted to $(0,\infty)^d \setminus (0,1)^d$ is the distribution of the tail chain, we assume that $x_i \geq 1$ for all $1 \leq i \leq d$ so that we identify ν as the distribution of

$$(c_i^{-1}(Y_{0,i})_+^{\alpha_i})_{1\leq i\leq d} \qquad \text{under the constraint} \qquad \max_{1\leq i\leq d} c_i^{-1}(Y_{0,i})_+^{\alpha_i}/x_i > 1.$$

Thus we have

$$A(x) = \mathbb{P}\left(c_i^{-1}(Y_{k,i})_+^{\alpha_i}/x_i \leq 1, i \geq 1, 1 \leq i \leq d, \max_{1\leq i\leq d} c_i^{-1}(Y_{0,i})_+^{\alpha_i}/x_i > 1\right).$$

To obtain an expression that is valid for any $x_i > 0$, we exploit the homogeneity

property, and we obtain

$$A(x) = \mathbb{P}\left(\max_{k\geq 0}\max_{1\leq i\leq d}(c_i x_i)^{-1}Y_{k,i}^{\alpha_i} > 1\right) - \mathbb{P}\left(\max_{k\geq 1}\max_{1\leq i\leq d}(c_i x_i)^{-1}Y_{k,i}^{\alpha_i} > 1\right)$$
$$= c_+(\mathbf{1})\mathbb{E}\left[\frac{\max_{k\geq 0}\max_{1\leq i\leq d}(c_i x_i)^{-1}(Y_{k,i})_+^{\alpha_i}}{\|(Y_0)_+\|_\alpha} - \frac{\max_{k\geq 1}\max_{1\leq i\leq d}(c_i x_i)^{-1}(Y_{k,i})_+^{\alpha_i}}{\|(Y_0)_+\|_\alpha}\right]$$

because $c_+(\mathbf{1})^{-1}\|(Y_0)_+\|_\alpha$ is standard Pareto distributed and independent of the spectral process $(Y_k)_+^\alpha/\|(Y_0)_+\|_\alpha$. This expression is homogeneous and extends to any possible x by homogeneity. □

3.4 Extremal indices

As the random coefficients $\tilde{M}_t$ in (1.4) may be large, consecutive values of X_t can be large. In the univariate case, one says that the extremal values appear in clusters. An indicator of the average length of the cluster is the inverse of the extremal index, an indicator of extremal dependence; see Leadbetter et al. (1983).
Thus, the natural extension of the extremal index is the function $\theta(x) = A(x)/A^*(x)$, with $A^*(x)$ and $A(x)$ defined in (3.2) and (3.3), respectively. Notice that there is no reason why θ should not depend on x. When $x_i \geq c_+(\mathbf{1})$, for $1 \leq i \leq d$, we have the more explicit expression in terms of the spectral process,

$$\theta(x) = \mathbb{P}\left(Y_{k,i}^{\alpha_i} \leq c_i x_i, k \geq 1, 1 \leq i \leq d \mid Y_{0,i}^{\alpha_i} > c_i x_i, 1 \leq i \leq d\right). \tag{3.5}$$

However, the extremal index θ_i of the marginal index $(X_{t,i})$ is still well-defined. It depends on the complete dependence structure of the multivariate Markov chain thanks to the following proposition:

Proposition 3.7. *Let X_t satisfy* (1.1)-(1.2). *With $\tilde{M}_t$ defined in* (1.4) *satisfying* (2.1) *and assuming the existence of Y_0 in Definition 3.1, the extremal index θ, defined in* (3.5)*, is a positive continuous function bounded from above by* 1 *that can be extended to $(0,\infty]^d\backslash\{\infty,\ldots,\infty\}$. The extremal indices of the marginals are obtained due to the formula*

$$\theta_i = \theta(\infty,\ldots,\infty,x_i,\infty,\ldots,\infty)$$
$$= \frac{\mathbb{E}\left[\|(Y_0)_+\|_\alpha^{-1}\left(\max_{k\geq 0}\left(\left(Y_0^\intercal\prod_{1\leq j\leq k}\tilde{M}_j\right)_i\right)_+^{\alpha_k} - \max_{k\geq 1}\left(\left(Y_0^\intercal\prod_{1\leq j\leq k}\tilde{M}_j\right)_i\right)_+^{\alpha_i}\right)\right]}{\mathbb{E}\left[\|(Y_0)_+\|_\alpha^{-1}(Y_{0,i})_+^{\alpha_i}\right]}.$$

Proof. Except for the positivity of the extremal index, the result follows by Proposition 2.5 in Perfekt (1997). The positivity is ensured by applying Corollary 2 in

Segers (2005). □

Example 3.8 (**Diagonal BEKK-ARCH**, continued)**.** Assuming that X_0 is vector scaling regularly varying, it follows from the tail chain approach of Janssen and Segers (2014) that the stationary Markov chain (X_t) is regularly varying. Thanks to the diagonal structure of the matrices $\tilde{M}_k = Am_k$, one can factorize $\|(Y_0)_+\|_\alpha^{-1}(Y_{0,i})^{\alpha_i}$ in the expression of θ_i provided in Proposition 3.7. Thanks to the independence of $\|(Y_0)_+\|_\alpha^{-1}(Y_{0,i})^{\alpha_i}$ with m_k, $k \geq 1$, we recover a similar expression as in the remarks after Theorem 2.1 in de Haan et al. (1989):

$$\theta_i = \mathbb{E}\left[\max_{k\geq 0}\left(A_{ii}^k \prod_{1\leq j\leq k} m_j\right)_+^{\alpha_k} - \max_{k\geq 1}\left(A_{ii}^k \prod_{1\leq j\leq k} m_j\right)_+^{\alpha_i}\right].$$

We did not manage to provide a link between the θ_i and the extremal index $\theta(x)$ of the (multivariate) stationary solution (X_t) of the Diagonal BEKK-ARCH. Due to the different normalising sequences in the asymptotic extremal result given in Theorem 3.6, the extremal index $\theta(x)$ depends on the constants c_is. For $x_i^* = c_+(\mathbf{1})$, $1 \leq i \leq d$, the expression (3.5) gets more simple because $c_+(\mathbf{1})^{-1}\|(Y_0)_+\|_\alpha$ is standard Pareto distributed and supported on $[1, \infty)$:

$$\theta(x^*) = \mathbb{P}\left(A_{ii}^k \prod_{1\leq j\leq k} m_j Y_{0,i} \leq (c_i c_+(\mathbf{1}))^{1/\alpha_i}, k \geq 1, 1 \leq i \leq d\right).$$

One can check that $\theta(x^*) \geq \theta_{i_0}$ where $1 \leq i_0 \leq d$ satisfies $A_{i_0 i_0} \geq A_{ii}$, $1 \leq i \leq d$. Using similar expressions for θ_i, one can check that $\theta_{i_0} \leq \theta_{ii}$, $1 \leq i \leq d$, i.e that i_0 is the marginal containing the largest clusters. Thus the inverse of the extremal index of the multidimensional Diagonal BEKK-ARCH is not larger than the largest average length of the marginals clusters. However, it is not clear whether the inverse of the extremal index has a nice interpretation in terms of cluster lengths in the multivariate setting. □

4 Sample covariances

In this section, we derive the limiting distribution of the sample covariances for certain BEKK-ARCH processes. The derivation of the limiting distribution relies on considering the tail behavior of the cross-products of the entries of X_t.

4.1 Extension of Kesten's theory to matrices with null entries

Consider the general SRE

$$Y_t = A_t Y_{t-1} + B_t, \qquad t \in \mathbb{Z}, \tag{4.1}$$

where (A_t, B_t) are $i.i.d.$ copies with $A_t \in M(\mathbb{R}, d)$ and B_t taking values in $\mathbb{R}^d$. We are interested in the case where A_t has possible null-entries motivated by the following remark:

Remark 4.1. Let (X_t) satisfy the Diagonal BEKK-ARCH model in Example 2.7. In order to study the regularly varying properties of the products $X_{t,k}X_{t,k'}$, we write

$$\begin{pmatrix} X_{t,k}X_{t,k'} \\ X_{t,k} \\ X_{t,k'} \end{pmatrix} = \begin{pmatrix} \tilde{M}_{k,t}\tilde{M}_{k',t} & \tilde{M}_{k,t}Q_{t,k'} & \tilde{M}_{k't}Q_{t,k} \\ 0 & \tilde{M}_{kt} & 0 \\ 0 & 0 & \tilde{M}_{k't} \end{pmatrix} \begin{pmatrix} X_{t-1,k}X_{t-1,k'} \\ X_{t-1,k} \\ X_{t-1,k'} \end{pmatrix} + \tilde{Q}_t.$$

Hence the vector $(X_{t,k}X_{t,k'}, X_{t,k}, X_{t,k'})^\intercal$ satisfies a SRE (4.1) with an upper triangular multiplicative matrix A_t.

We seek to characterize the regular variation properties of the SRE in the above remark. Unfortunately, the conditions **(A3)**, **(A4)**, **(A5)** and **(A7)** of Theorem 1.1 of Alsmeyer and Mentemeier (2012) (see the Appendix A.1) are not satisfied because some entries of the matrix A_t are potentially null. On the other hand, the remaining conditions can be shown to hold: Condition **(A6)** follows since $\tilde{Q}_t$ is independent of the diagonal elements $(\tilde{M}_{k,t}\tilde{M}_{k',t}, \tilde{M}_{kt}, \tilde{M}_{k't})$, whereas the moment conditions, **(A1)** and **(A2)**, hold since $\tilde{M}_t$ and Q_t are Gaussian.

In the following we extend the result of Alsmeyer and Mentemeier (2012) for the general SRE (4.1) satisfying **(A1)**, **(A2)** and **(A6)**, but with $\prod_{t=1}^n A_t$ containing some null entries for all $n \geq 1$. For that, we replace the assumptions **(A3)**, **(A4)**, **(A5)** and **(A7)** of Alsmeyer and Mentemeier (2012) by the less restrictive assumptions **(A3')**, **(A4')**, **(A5')** and **(A7')** below. The main idea is to focus on a subspace of $\mathbb{R}^d$ stable under the multiplication of A_t. Let d' be an integer so that $1 \leq d' \leq d$:

- **(A3')** The subset $\mathbb{R}^{d'} \times \{0\} \times \cdots \times \{0\}$ is invariant under the multiplication by any matrix in the support of A_0. Moreover, the corresponding $d' \times d'$ sub-matrices are invertible.

- **(A4')** $\max_{n\in\mathbb{N}} \mathbb{P}\left\{\|x^\intercal \prod_{i=1}^n A_i\|^{-1} (x^\intercal \prod_{i=1}^n A_i) \in U\right\} > 0$, for any $x = (x'^\intercal, 0, \ldots, 0)^\intercal$ with $x' \in \mathbb{S}^{d'-1}$ and any $U = U' \times \{0\} \times \cdots \times \{0\}$ such that open $U' \in \mathbb{S}^{d'-1}$.

- **(A5')** Let $\mathcal{V}_\delta$ denote the open δ-ball in $GL(d, \mathbb{R})$ and let $\mathbb{LEB}$ denote the Lebesgue measure on $\mathbb{R}^{d^2}$. Then $\mathbb{P}\left\{\prod_{i=1}^{n_0} A_i \in A\right\} \geq \gamma_0 1_{\mathcal{V}_c(\Gamma_0)}(A)\mathbb{LEB}(A)$, for some $\Gamma_0 \in GL(d, \mathbb{R})$, $n_0 \in \mathbb{N}$, $c, \gamma_0 > 0$ and $A = A' \times \mathbb{R}^{d^2-d'^2}$ with $A' \subset \mathbb{R}^{d'^2}$.

- **(A7')** There exists $0 < \kappa_0'$ such that $\mathbb{E}[\inf_{\mathbb{S}^{d'-1}\times\{0\}\times\cdots\times\{0\}} \|A_0 x\|^{\kappa_0'}] \geq 1$ and $\mathbb{E}[\|Y_{0,k}\|^{\kappa_0'}] < \infty$ for $k = d'+1, \ldots, d$. Moreover $\|B_0\| > 0$ and $\|A_0\|$ have finite moments of any order.

We then have the following extension of Theorem 1.1 of Alsmeyer and Mentemeier (2012):

Theorem 4.2. *Consider the SRE in* (4.1)*. Under the assumptions* **(A1)**, **(A2)**, **(A3')**, **(A4')**, **(A5')**, **(A6)** *and* **(A7')** *there exists* $0 < \kappa' \leq \kappa_0'$ *such that*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{S}^{d'-1}\times\{0\}\times\cdots\times\{0\}} \mathbb{E}\left[\left\|x^\intercal \prod_{i=1}^n A_i\right\|^{\kappa'}\right]^{1/n} = 1,$$

and

$$\lim_{y\to\infty} y^{\kappa'} \mathbb{P}(u^\intercal Y_0 > y) = w_\alpha(u) \quad u = (u'^\intercal, 0, \ldots, 0)^\intercal \text{ with } u' \in \mathbb{S}^{d'-1},$$

for some finite, positive, and continuous function w_α *on* $\mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}$.

Proof. The existence of a unique stationary solution is already proved. The extension follows by the same steps as the ones of Theorem 1.1 of Alsmeyer and Mentemeier (2012). For instance, Lemma 2.1 of Alsmeyer and Mentemeier (2012) can be extended to the following Lemma. Its proof follows the lines of the proof of Lemma 2.1 of Alsmeyer and Mentemeier (2012) and is thus omitted. Define $P_C^n(x, \cdot) := \mathbb{P}(\|x^\intercal \prod_{i=1}^n A_i\|^{-1} x^\intercal \prod_{i=1}^n A_i \in \cdot, \prod_{i=1}^n A_i \in C)$. We have

Lemma 4.3. *Suppose that* **(A4')** *and* **(A5')** *hold. Then there exists a compact subset* C *of* $GL(d', \mathbb{R})$ *such that, for each* $x \in \mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}$*, there are* $\delta, p > 0$*,* $m \in \mathbb{N}$*, and a probability measure* ϕ *on* $B_\delta(x)$ *satisfying*

$$P^m(y, A) \geq P_C^m(y, A) \geq p\phi(A)$$

for all $y \in \mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}$ *and* $A = A' \times \mathbb{R}^{d^2-d'^2}$ *with* $A' \subset \mathbb{R}^{d'^2}$.

It implies that one can consider a Markov Random Walk as (U_n, V_n) with

$$U_n = \left\|x^\intercal \prod_{i=1}^n A_i\right\|^{-1} x^\intercal \prod_{i=1}^n A_i \qquad \text{and} \qquad V_n = \log\left\|x^\intercal \prod_{i=1}^n A_i\right\|,$$

for any initial value $x \in \mathbb{R}^{d'} \times \{0\} \times \cdots \times \{0\}$. The Markov chain (Y_t) is then Harris recurrent on $\mathbb{S}^d$ with a small set of the form $A' \times \mathbb{R}^{d^2-d'^2}$ with $A' \subset \mathbb{R}^{d'^2}$. One can then apply the Markov Renewal Theorem of Alsmeyer stated as Theorem 4.2 in Alsmeyer and Mentemeier (2012) to the MRW under a change of measure. We obtain an equivalent of Theorem 5.1 in Alsmeyer and Mentemeier (2012):

Theorem 4.4. *Under the assumptions of Theorem 4.2, there exists* $0 < \kappa' \leq \kappa'_0$ *and a positive function* $r : \mathbb{S}^{d'-1} \to (0, \infty)$ *such that for all bounded functions* f *and all* n

$$\frac{1}{r(x)} \mathbb{E}_x[r(Y_n) e^{\kappa' V_n} f(U_0, V_0, U_1, V_1, \cdots, U_n, V_n)]$$

defines a distribution $\mathbb{P}_x^\kappa$ *for any* $x \in \mathbb{R}^{d'} \times \{0\} \times \cdots \times \{0\}$. *Under this change of measure, the MRW* (U_n, V_n) *has a positive drift and satisfies the assumptions of Theorem 4.2 of Alsmeyer and Mentemeier (2012). The constant* κ' *is the unique value determined by*

$$\lim_{n\to\infty} \sup_{x \in \mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}} \mathbb{E}\left[\left\| x^\intercal \prod_{i=1}^n A_i \right\|^{\kappa'} \right]^{1/n} = 1.$$

The long proof follows the lines of the one of Theorem 5.1 in Alsmeyer and Mentemeier (2012). The strategy is to consider the Markov chain restricted on $\mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}$ rather than on $\mathbb{S}^{d'}$.

The conclusion follows from the implicit Markov renewal theory of Goldie (1991). It applies similarly to the restricted MRW as to the unrestricted one, following arguments similar to the ones in Section 6 of Alsmeyer and Mentemeier (2012). The crucial step is to check the integrability property

$$\sup_{x \in \mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}} \mathbb{E}|(x^\intercal A_0 Y_0 + x^\intercal B_0)_+^{\kappa'} - (x^\intercal A_0 Y_0)_+^{\kappa'}| < \infty.$$

One has to consider the d' first coordinates of the product $A_0 Y_0$ where Y_0 is the unique stationary solution of the SRE on $\mathbb{R}^d$ (and not $\mathbb{R}^{d'} \times \{0\} \times \cdots \times \{0\}$). The d' first coordinates of $A_0 Y_0$ may involve all the coordinates of Y_0. Consider, without loss of generality, the case where A_0 is blockwise upper-triangular. Denote by D_0 the first block diagonal of size $d' \times d'$ and T_0 the upper diagonal block. Then $x^\intercal A_0 Y_0 = x^\intercal D_0 (Y_{0,1}, \ldots, Y_{0,d'})^\intercal + x^\intercal U_0 (Y_{0,d'+1}, \ldots, Y_{0,d})^\intercal$. Then, one adapts Corollary 4.3 of Goldie (1991) and we obtain an upper bound

$$3^{\kappa'-1}(\mathbb{E}[|x^\intercal B_0|^{\kappa'}] + \mathbb{E}[|x^\intercal B_0||x^\intercal T_0 (Y_{0,d'+1}, \ldots, Y_{0,d})^\intercal|^{\kappa'}] + \mathbb{E}[|x^\intercal B_0|]\mathbb{E}[|x^\intercal D_0 Y_0|^{\kappa'-1}]).$$

The two first terms admit finite moments of order κ_0' and hence of order κ', due to Assumption **(A7')**. Turning to the last term, notice that $\mathbb{E}[|x^\intercal B_0|]$ is finite due to Assumption **(A7')**. For the second factor, we have that

$$\mathbb{E}[|x^\intercal D_0 Y_0|^{\kappa'-1}]^{1/(\kappa'-1)} \leq \frac{c_n}{1-\mathbb{E}[\|x^\intercal \prod_{i=1}^n A_i\|^{\kappa'-1}]^{1/(\kappa'-1)}},$$

by a recursive argument, such that for n sufficiently large, $-\mathbb{E}[\|x^\intercal \prod_{i=1}^n A_i\|^{\kappa'-1}] < 1$, and $c_n > 0$ is a constant that does only depend on moments of finite products of B_{-t} and A_{-t}, $1 \leq t \leq n$.

The reasoning in Sections 7 and 8 of Alsmeyer and Mentemeier (2012) also applies to the stationary solution restricted to its first d' coordinates in $u^\intercal Y_0$ for $u = (u'^\intercal, 0, \ldots, 0)^\intercal$ with $u' \in \mathbb{S}^{d'-1}$. More precisely, the key Lemma 8.1 of Alsmeyer and Mentemeier (2012) is valid when restricted to any $u \in \mathbb{S}^{d'-1} \times \{0\} \times \cdots \times \{0\}$: for any $t > 0$ we have

$$\mathbb{P}(u^\intercal Y_0 \leq t) < 1.$$

Then the support of $u^\intercal Y_0$ is unbounded and the limiting constant $w_\alpha(u)$ is positive. □

4.2 Regular variation properties of cross-products

We are now ready to apply this extension of Theorem 1.1 of Alsmeyer and Mentemeier (2012) in the context of products of BEKK-ARCH coordinates:

Proposition 4.5. *Let X_t satisfy* (1.1)-(1.2)*, and let $\tilde{M}_t$ be defined as in* (1.4) *satisfying* (2.1)*. Assume that Y_0 in Definition 3.1 exists (and hence that Assumption* **M** *holds). Assume moreover that $\mathbb{E}[|\tilde{M}_{k,t}\tilde{M}_{k',t}|^{\alpha_{kk'}}] = 1$ for some $\alpha_{kk'} < \alpha_i$, $1 \leq i \leq d$, and that $Z_t = (X_{t,k}X_{t,k'}, X_{t,1}, \ldots, X_{t,d})^\intercal \in \mathbb{R}^{d+1}$, $t \geq 0$, satisfies the assumptions of Theorem 4.2 with $d' = 1$. Then $(X_{t,k}X_{t,k'})$ is regularly varying of order $\alpha_{kk'}$ and its tail process is given by*

$$Y_{n,kk'}^\intercal = Y_{0,kk'}^\intercal \prod_{t=1}^n \tilde{M}_{k,t}\tilde{M}_{k',t},$$

with $Y_{0,kk'}$ is a symmetric r.v. such that $|Y_{0,kk'}|$ is Pareto distributed with index $\alpha_{kk'}$.

Proof. The proof follows from the application of Theorem 4.2 to the process Z_t with $d' = 1$. The regular variation of the linear combination $u^\intercal Z_0 = X_{0,k}X_{0,k'}$ with $u = (1, 0, \ldots, 0)$ and $u = (-1, 0, \ldots, 0)$ is equivalent to the regular variation of the marginal $X_{0,k}X_{0,k'}$ because the other marginals have larger tail indices $\alpha_i > \alpha_{kk'}$, $1 \leq i \leq d$, thanks to the existence of Y_0 in Definition 3.1. The tails are equally

balanced applying Theorem 2.3 (Case b) of Goldie (1991). The regular variation of the process and the determination of its tail process follow by similar arguments than in Basrak and Segers (2009). □

Under the assumptions of Proposition 4.5, it follows from an application of Theorem 2.3 in Goldie (1991) that the $\alpha_{kk'}$ moments of $X_{0,k}$ and $X_{0,k'}$ are finite. If $\alpha_{kk'} \geq \alpha_k \wedge \alpha_{k'}$ the proposition does not apply and we conjecture that in this case the product $X_{0,k}X_{0,k'}$ might be regularly varying of order $\alpha_k \wedge \alpha_{k'}$. The proposition does not provide any multi scaling regular variation properties on $(X_{0,k}X_{0,k'}, X_{0,k}, X_{0,k'})^\intercal$. However, it is still useful to assert the convergence of the empirical covariances, as clarified in the next section.

Example 4.6 (**Diagonal BEKK-ARCH**, continued)**.** Consider the case $d = 2$ with $\alpha_1 < \alpha_2 \leq 2\alpha_1$. Then $|A_{ii}|^{\alpha_i} = 1/E[|m_0|^{\alpha_i}]$ so that and

$$\mathbb{E}[|A_{11}m_0A_{22}m_0|^{\alpha_1}] = \frac{\mathbb{E}\left[m_0^{2\alpha_1}\right]}{\mathbb{E}[|m_0|^{\alpha_1}]\mathbb{E}[|m_0|^{\alpha_2}]^{\alpha_1/\alpha_2}}.$$

From Jensen's inequality we have $\mathbb{E}[|m_0|^{\alpha_2}] \leq \mathbb{E}[|m_0|^{2\alpha_1}]^{\alpha_2/(2\alpha_1)}$ so that

$$\mathbb{E}[|A_{11}m_0A_{22}m_0|^{\alpha_1}] = \frac{\mathbb{E}\left[m_0^{2\alpha_1}\right]}{\mathbb{E}[|m_0|^{\alpha_1}]\mathbb{E}[|m_0|^{\alpha_2}]^{\alpha_1/\alpha_2}} \geq \frac{\mathbb{E}\left[m_0^{2\alpha_1}\right]^{1/2}}{\mathbb{E}[|m_0|^{\alpha_1}]} > 1$$

by applying Jensen's inequality a second time. Then $\alpha_1 > \alpha_{12}$ the unique positive index so that $\mathbb{E}[|A_{11}m_0A_{22}m_0|^{\alpha_{12}}] = 1$. One can apply Proposition 4.5 and obtain that the cross products series $(X_{t,1}X_{t,2})$ is regularly varying with index $\alpha_{12} < \alpha_1 \wedge \alpha_2$. □

4.3 Asymptotic properties of the sample covariances

We now consider the asymptotic properties of the sample covariance matrix for the BEKK-ARCH process. Let

$$\hat{\Gamma} = \frac{1}{n}\sum_{t=1}^{n} X_t X_t^\intercal.$$

Applying Theorems 2.2 and 4.1 of Mikosch and Wintenberger (2014) yields the following result.

Theorem 4.7. *Assume that* (X_t) *is the unique stationary solution of* (1.1)-(1.2) *under* (2.1)*. Assume that* $(X_{0,k}X_{0,k'})_{k,k'}$ *is multivariate regularly varying with a*

single index $\alpha > 0$ with the associated slowly varying $\ell(t) \sim c$ as $t \to \infty$. Then

$$\begin{aligned} n^{1-2/\alpha}\text{vech}(\hat{\Gamma}) &\to S_\alpha, \qquad \textit{if} \qquad 0 < \alpha < 2, \\ n^{1-2/\alpha}\text{vech}\left(\hat{\Gamma} - \mathbb{E}[X_0 X_0^\intercal]\right) &\to S_\alpha, \qquad \textit{if} \qquad 2 < \alpha < 4, \\ n^{1/2}\text{vech}\left(\hat{\Gamma} - \mathbb{E}[X_0 X_0^\intercal]\right) &\to N, \qquad \textit{if} \qquad \alpha > 4, \end{aligned}$$

where S_α is a symmetric non-degenerate (but singular) α-stable random vector and N is a non-degenerate Gaussian random vector.

Notice that the multiple indices case is embedded into this framework when the smallest index $\alpha_{k,k'}$ is unique and equal to α. In that case, $S_{\alpha,i,j} = 0$ a.s. except for $i = k$, $j = k'$, where $S_{\alpha,k,k'}$ is the element of S_α that corresponds to $X_{0,k}X_{0,k'}$.

Example 4.8 (**Similarity BEKK-ARCH**, continued)**.** Consider the scalar BEKK-ARCH process introduced in Example 2.6. Here $A = \sqrt{a}I_d$, with I_d the identity matrix, such that $\tilde{M}_t$ is diagonal. This enables us to establish that the scalar BEKK process satisfies Proposition 4.5. Specifically, maintaining the assumption of Proposition 2.13 that $a < \exp\{(1/2)[-\psi(1) + \log(2)]\}$, notice that the index of regular variation for each marginal of X_t is given by α satisfying $\mathbb{E}[|\sqrt{a}m_t|^\alpha] = 1$. Moreover, it holds that $\tilde{M}_{k,t}\tilde{M}_{k',t} = am_t^2$, and hence for a suitable choice of a there exists α_x satisfying $\mathbb{E}[(am_t^2)^{\alpha_x}] = 1$ such that $\alpha_x < \alpha$. One example is when $a = 1$, implying that $\alpha_x = 1$ and $\alpha = 2$.

In the case $\alpha_x < \alpha$ with $1 < \alpha_x < 2$, we have from Theorem 4.7 that all sample covariances have a stable limiting distribution. □

Example 4.9 (**Diagonal BEKK-ARCH**, continued)**.** Consider the case $d = 2$ with $\alpha_1 < \alpha_2 \leq 2\alpha_1$. Then $|A_{11}| > |A_{22}|$ and

$$\mathbb{E}[|A_{11}m_0A_{22}m_0|^{\alpha_1/2}] < \mathbb{E}[|A_{11}m_0A_{11}m_0|^{\alpha_1/2}] = \mathbb{E}[|A_{11}m_0|^{\alpha_1}] = 1$$

so that $\alpha_1/2 < \alpha_{12}$. So $(X_{t,1}^2, X_{t,1}X_{t,2}, X_{t,2}^2)^\intercal$ is multivariate regularly varying with index $\alpha_1/2$ and degenerate tail marginal processes $(Y_{0,2}, Y_{0,3}) = (0, 0)$, $Y_{0,1}$ being a symmetric r.v. so that $|Y_{0,1}|$ is $\alpha_1/2$-Pareto distributed. As $\alpha = \alpha_1$, one marginal having infinite fourth moments makes the empirical covariances approximation of the order $n^{1-2/\alpha_1} = o(\sqrt{n})$ only. This rate of convergence does not provide any information about the more precise ($\alpha_1/2 < \alpha_{12}$) approximation of $\mathbb{E}[X_{t,1}X_{t,2}]$ by its empirical covariance. □

The two previous examples are important in relation to variance targeting estimation of the BEKK-ARCH model, as considered in Pedersen and Rahbek (2014).

For the univariate GARCH process, Vaynman and Beare (2014) have shown that the limiting distribution of the (suitably scaled) variance targeting estimator follows a singular stable distribution when the tail index of the process lies in $(2, 4)$. We expect a similar result to hold for the BEKK-ARCH process.

5 Concluding remarks

We exhibit the interesting special extremal behavior of the simple Diagonal BEKK-GARCH model among the BEKK-GARCH model class. We want to investigate further the use of the simple Diagonal BEKK-GARCH on financial data. The results contained in this paper are expected to be important for future research related to the statistical analysis of the Diagonal BEKK-ARCH model. As recently shown by Avarucci et al. (2013), the (suitably scaled) maximum likelihood estimator for the general BEKK-ARCH model (with $l = 1$) does only have a Gaussian limiting distribution, if the second-order moments of X_t is finite. In order to obtain the limiting distribution in the presence of very heavy tails, i.e. when $\mathbb{E}[\|X_t\|^2] = \infty$, we believe that non-standard arguments are needed, and in particular the knowledge of the tail-behavior is expected to be crucial for the analysis. We leave additional considerations in this direction to future research.

References

Alsmeyer, G. (2003): "On the Harris recurrence of iterated random Lipschitz functions and related convergence rate results," *Journal of Theoretical Probability*, 16, 217–247.

Alsmeyer, G. and S. Mentemeier (2012): "Tail behaviour of stationary solutions of random difference equations: The case of regular matrices," *Journal of Difference Equations and Applications*, 18, 1305–1332.

Avarucci, M., E. Beutner, and P. Zaffaroni (2013): "On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models," *Econometric Theory*, 29, 545–566.

Basrak, B., R. A. Davis, and T. Mikosch (2002a): "A characterization of multivariate regular variation," *The Annals of Applied Probability*, 12, 908–920.

——— (2002b): "Regular variation of GARCH processes," *Stochastic Processes and their Applications*, 99, 95–115.

Basrak, B. and J. Segers (2009): "Regularly varying multivariate time series," *Stochastic Processes and their Applications*, 119, 1055–1080.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79–109.

Buraczewski, D., E. Damek, Y. Guivarc'h, A. Hulanicki, and R. Urban (2009): "Tail-homogeneity of stationary measures for some multidimensional stochastic recursions," *Probability Theory and Related Fields*, 145, 385–420.

Buraczewski, D., E. Damek, and T. Mikosch (2016): *Stochastic Models with Power-Law Tails: The Equation $X = AX + B$*, Springer Series in Operations Research and Financial Engineering, Springer International Publishing.

de Haan, L. and S. I. Resnick (1977): "Limit theory for multivariate sample extremes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40, 317–337.

de Haan, L., S. I. Resnick, H. Rootzén, and C. G. de Vries (1989): "Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes," *Stochastic Processes and their Applications*, 32, 213–224.

Einmahl, J. H., L. de Haan, and V. I. Piterbarg (2001): "Nonparametric estimation of the spectral measure of an extreme value distribution," *The Annals of Statistics*, 29, 1401–1423.

Engle, R. F. and K. F. Kroner (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, 11, 122–150.

Feigin, P. and R. Tweedie (1985): "Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments," *Journal of Time Series Analysis*, 6, 1–14.

Goldie, C. M. (1991): "Implicit renewal theory and tails of solutions of random equations," *The Annals of Applied Probability*, 1, 126–166.

Janssen, A. and J. Segers (2014): "Markov tail chains," *Journal of Applied Probability*, 51, 1133–1153.

Kesten, H. (1973): "Random difference equations and renewal theory for products of random matrices," *Acta Mathematica*, 131, 207–248.

Kulik, R., P. Soulier, and O. Wintenberger (2015): "The tail empirical process of regularly varying functions of geometrically ergodic Markov chains," Https://arxiv.org/abs/1511.04903.

Leadbetter, M., G. Lindgren, and H. Rootzén (1983): *Extremes and related properties of random sequences and processes*, Springer series in statistics, Springer-Verlag.

Matsui, M. and T. Mikosch (2016): "The extremogram and the cross-extremogram for a bivariate GARCH(1,1) process," *Advances in Applied Probability*, Forthcoming.

Mikosch, T. and O. Wintenberger (2014): "The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate Markov chains," *Probability Theory and Related Fields*, 159, 157–196.

Nelson, D. B. (1990): "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318–334.

Nielsen, H. B. and A. Rahbek (2014): "Unit root vector autoregression with volatility induced stationarity," *Journal of Empirical Finance*, 29, 144–167.

Pedersen, R. S. (2016): "Targeting estimation of CCC-GARCH models with infinite fourth moments," *Econometric Theory*, 32, 498–531.

Pedersen, R. S. and A. Rahbek (2014): "Multivariate variance targeting in the BEKK-GARCH model," *The Econometrics Journal*, 17, 24–55.

Perfekt, R. (1997): "Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$," *Advances in Applied Probability*, 29, 138–164.

Segers, J. (2005): "Generalized Pickands estimators for the extreme value index," *Journal of Statistical Planning and Inference*, 128, 381–396.

Stărică, C. (1999): "Multivariate extremes for models with constant conditional correlations," *Journal of Empirical Finance*, 6, 515–553.

Vaynman, I. and B. K. Beare (2014): "Stable limit theory for the variance targeting estimator," in *Essays in Honor of Peter C. B. Phillips*, ed. by Y. Chang, T. B. Fomby, and J. Y. Park, Emerald Group Publishing Limited, vol. 33 of *Advances in Econometrics*, chap. 24, 639–672.

A Appendix

A.1 Theorem 1.1 of Alsmeyer and Mentemeier (2012)

Consider the general SRE

$$Y_t = A_t Y_{t-1} + B_t \tag{A.1}$$

with (A_t, B_t) a sequence of i.i.d. random variables with generic copy (A, B) such that A is a $d \times d$ real matrix and B takes values in $\mathbb{R}^d$. Consider the following conditions of Alsmeyer and Mentemeier (2012):

- **(A1)** $\mathbb{E}[\log^+(\|A\|)] < \infty$, where $\|\cdot\|$ denotes the operator norm.
- **(A2)** $\mathbb{E}[\log^+(\|B\|)] < \infty$.
- **(A3)** $\mathbb{P}[A \in GL(d, \mathbb{R})] = 1$.
- **(A4)** $\max_{n\in\mathbb{N}} \mathbb{P}\left\{\|x^\intercal \prod_{i=1}^n A_i\|^{-1} (x^\intercal \prod_{i=1}^n A_i) \in U\right\} > 0$, for any $x \in \mathbb{S}^{d-1}$ and any non-empty open subset U of $\mathbb{S}^{d-1}$.
- **(A5)** Let $\mathcal{V}_\delta$ denote the open δ-ball in $GL(d, \mathbb{R})$ and let $\mathbb{LEB}$ denote the Lebesgue measure on $M(d, \mathbb{R})$. It holds that for any Borel set $A \in M(d, \mathbb{R})$, $\mathbb{P}(\prod_{i=1}^{n_0} A_i \in A) \geq \gamma_0 1_{\mathcal{V}_c(\Gamma_0)}(A)\mathbb{LEB}(A)$ for some $\Gamma_0 \in GL(d, \mathbb{R})$, $n_0 \in \mathbb{N}$, and $c, \gamma_0 > 0$.
- **(A6)** $\mathbb{P}(A_0 v + B_0 = v) < 1$ for any $v \in \mathbb{R}^d$.
- **(A7)** There exists $\kappa_0 > 0$ such that

$$\mathbb{E}[\inf_{x\in\mathbb{S}^{d-1}} \|A_0 x\|^{\kappa_0}] \geq 1, \quad \mathbb{E}[\|A_0\|^{\kappa_0} \log^+ \|A_0\|] < \infty, \quad \text{and} \quad 0 < \mathbb{E}[\|B_0\|^{\kappa_0}] < \infty.$$

Theorem A.1 (Alsmeyer and Mentemeier (2012, Theorem 1.1))**.** *Consider the SRE in* (A.1)*) suppose that* $\beta := \lim_{n\to\infty} n^{-1} \log(\|\prod_{i=1}^n A_i\|) < 0$ *and that* **(A1)**-**(A7)** *hold, then there exists a unique* $\kappa \in (0, \kappa_0]$ *such that*

$$\lim_{n\to\infty} n^{-1} \log(\|\prod_{i=1}^n A_i\|^\kappa) = 0.$$

Moreover, the SRE has a strictly stationary solution satisfying,

$$\lim_{t\to\infty} t^\kappa \mathbb{P}(x^\intercal Y_0 > t) = K(x) \quad \textit{for all } x \in \mathbb{S}^{d-1},$$

where K *is a finite positive and continuous function on* $\mathbb{S}^{d-1}$.

A.2 Estimation of the spectral measure for the bivariate diagonal BEKK-ARCH process

In this section we consider the estimation of the spectral measure of the diagonal BEKK-ARCH process presented in Example 2.14. Specifically, we consider a special case of the BEKK-ARCH process in (1.1)-(1.2), where $d = 2$:

$$X_t = m_t A X_{t-1} + Q_t,$$

with $\{Q_t : t \in \mathbb{N}\}$ an *i.i.d.* process with $Q_t \sim N(0, C)$ independent of $\{m_t : t \in \mathbb{N}\}$, and

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}.$$

Following the approach for *i.i.d.* sequences of vectors given in Einmahl et al. (2001), we consider the following estimator of the spectral measure of $X_t = (X_{t,1}, X_{t,2})^\intercal$:

$$\hat{\Phi}(\theta) = \frac{1}{k} \sum_{t=1}^{T} \mathbf{1}_{\{R_t^{(1)} \vee R_t^{(2)} \geq T+1-k, \arctan \frac{T+1-R_t^{(2)}}{T+1-R_t^{(1)}} \leq \theta\}}, \quad \theta \in [0, \pi/2],$$

where $R_t^{(j)}$ denotes the rank of $X_{t,j}$ among $X_{1,j}, ..., X_{T,j}$, $j = 1, 2$, i.e.

$$R_t^{(j)} := \sum_{i=1}^{T} \mathbf{1}_{\{X_{i,j} \geq X_{t,j}\}}.$$

Here k is a sequence satisfying $k(T) \to \infty$ and $k(T) = o(T)$. Einmahl et al. (2001) showed that this estimator is consistent for i.i.d. series. We expect a similar result to hold for geometrically ergodic processes. The reason is that the asymptotic behavior of the empirical tail process used in Einmahl et al. (2001) has been extended to such cases in Kulik et al. (2015).

We consider the estimation of the spectral measure for different values of C, A_{11}, and A_{22}. In particular, the matrix C is

$$C = 10^{-5} \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}, \quad c \in \{0, 0.5\},$$

and the values A_{11} and A_{22} are determined according to choices of the tail indices of $X_{t,1}$ and $X_{t,2}$, respectively. I.e. A_{11} and A_{22} satisfy $\mathbb{E}[|m_t|^{\alpha_i}] = |A_{ii}|^{-\alpha_i}$ and are determined by analytical integration. Specifically, with $\phi(\cdot)$ the pdf of the standard

normal distribution,

$$\begin{aligned}
\alpha_i &= 0.5 \Rightarrow A_{ii} = (\int_{-\infty}^{\infty} |m|^{0.5}\phi(m)dm)^{-1/0.5} \approx 1.479 \\
\alpha_i &= 2.0 \Rightarrow A_{ii} = 1 \\
\alpha_i &= 3.0 \Rightarrow A_{ii} = (8/\pi)^{-1/6} \approx 0.8557 \\
\alpha_i &= 4.0 \Rightarrow A_{ii} = 3^{-1/4} \approx 0.7598
\end{aligned}$$

Figure A.1 contains plots of the estimates of the spectral measure. The estimates $\hat{\Phi}(\theta)$ are based on one realization of the process with $T = 2{,}000$ and a burn-in period of 10,000 observations.

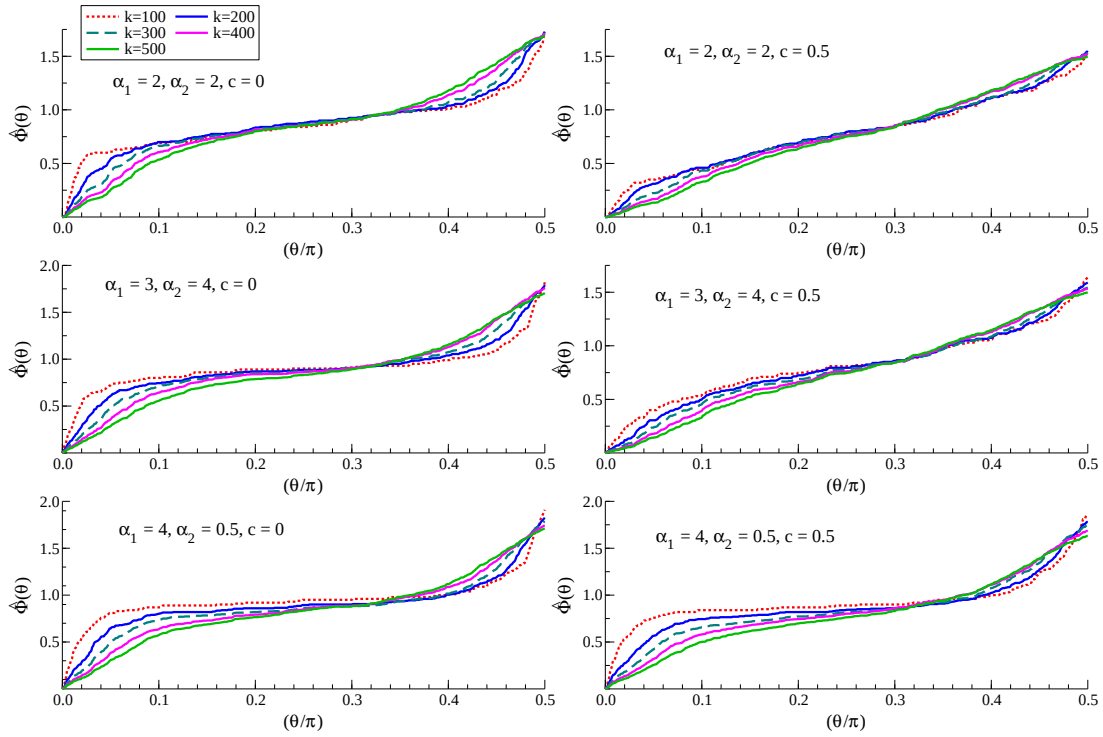

Figure A.1: Nonparametric estimates for $k = 100, 200, 300, 400, 500$ and for various choices of α_1, α_2, and c.