UNIVERSITY OF COPENHAGEN

Targeting estimation of CCC-GARCH models with infinite fourth moments

Pedersen, Rasmus Søndergaard

Publication date:
2014

Document version
Early version, also known as pre-print

Citation for published version (APA):
Pedersen, R. S. (2014). *Targeting estimation of CCC-GARCH models with infinite fourth moments*. Kbh.: Økonomisk institut, Københavns Universitet. University of Copenhagen. Institute of Economics. Discussion Papers (Online), No. 04, Vol.. 2014

Download date: 18. sep.. 2019

Discussion Papers
Department of Economics
University of Copenhagen

No. 14-04

Targeting estimation of CCC-Garch models with infinite fourth moments

Rasmus Søndergaard Pedersen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

TARGETING ESTIMATION OF CCC-GARCH MODELS WITH INFINITE FOURTH MOMENTS

RASMUS SØNDERGAARD PEDERSEN
UNIVERSITY OF COPENHAGEN

Abstract. As an alternative to quasi-maximum likelihood, targeting estimation is a much applied estimation method for univariate and multivariate GARCH models. In terms of variance targeting estimation recent research has pointed out that at least finite fourth-order moments of the data generating process is required if one wants to perform inference in GARCH models relying on asymptotic normality of the estimator, see Pedersen and Rahbek (2014) and Francq et al. (2011). Such moment conditions may not be satisfied in practice for financial returns highlighting a large drawback of variance targeting estimation. In this paper we consider the large-sample properties of the variance targeting estimator for the multivariate extended constant conditional correlation GARCH model when the distribution of the data generating process has infinite fourth moments. Using non-standard limit theory we derive new results for the estimator stating that its limiting distribution is multivariate stable. The rate of consistency of the estimator is slower than $\sqrt{T}$ (as obtained by the quasi-maximum likelihood estimator) and depends on the tails of the data generating process.

Keywords: Targeting; variance targeting; multivariate GARCH; constant conditional correlation; asymptotic theory; time series, multivariate regular variation, stable distributions.

JEL Classification: C32, C51, C58.

1. Introduction

Targeting estimation is, by now, a much applied tool when estimating multivariate volatility models, see e.g. Laurent et al. (2012), Hafner and Reznikova (2012), and Caporin and McAleer (2012) for recent applications and discussions of targeting estimation in multivariate GARCH models. Recently, Pedersen and Rahbek (2014) have considered the asymptotic properties of the (covariance) targeting estimator for multivariate BEKK-GARCH models, and Kristensen and Linton (2004) and Francq et al. (2011) have developed the large-sample properties of the (variance) targeting estimator for univariate GARCH models. As established in these papers, at least finite fourth-order moments of the observed process are required in order to establish asymptotic normality of the estimators. In practice such moment restrictions may not be satisfied for asset returns casting serious doubt on the validity of the inference performed in GARCH models based on targeting estimation. In the present paper we derive the limiting distribution of the targeting estimator for multivariate extended constant conditional correlation (CCC-)GARCH models

Date: February 20, 2014.

I thank Iliyan Giorgiev, Thomas Mikosch, and Anders Rahbek for helpful comments.

Address: University of Copenhagen, Department of Economics, Oester Farimagsgade 5, Building 26, 1353, Copenhagen K, Denmark. Phone: (+45)35323074. Fax: (+45)35323000. E-mail: `rsp@econ.ku.dk`.

when the data generating process does not have finite fourth-order moments but still finite moments of second order (implying consistency of the estimator). We show that under certain conditions, the limiting distribution of the targeting estimator is multivariate stable. The rate of consistency is slower than $\sqrt{T}$, the rate of consistency for the quasi-maximum likelihood estimator (QMLE), and is determined by the tails of the distribution of the observed vector process. Our conclusions are in line with the ones in a recent paper by Vaynman and Beare (2013) who consider the limiting distribution of the variance targeting estimator for univariate GARCH models in a similar setting.

Forecasts of conditional covariance matrices play an important role in a vast amount of financial applications as in for example the fields of (dynamic) portfolio allocation and (conditional) Value-at-Risk. Such forecasts can be based on multivariate GARCH models, see Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009) for surveys on such models. A classical multivariate GARCH model is indeed the CCC-GARCH model proposed by Bollerslev (1990) and its extended version proposed by Jeantheau (1998). The asymptotic properties of the QMLE for this model have been considered by Ling and McAleer (2003) and recently by Francq and Zakoïan (2010, Chapter 11) and Francq and Zakoïan (2012). A drawback of the model, and especially of its extended version, is the large number of model parameters, which makes classical quasi-maximum likelihood (QML) estimation difficult, if not infeasible, for a large dimension of the time series. One can address this curse of dimensionality issue by applying simplified versions of the model, and/or by considering an alternative estimation method, such as targeting estimation proposed by Engle and Mezrich (1996).

Targeting estimation is a two-step estimation procedure where, for the CCC-GARCH model, the model is reparametrized such that the vector of long-run variances enters explicitly in the equation for the vector of conditional variances. The long-run variances are estimated by a moment estimator in a first step and conditional on this, the remaining parameters are estimated in a second step using QML estimation. Regardless of the model has a simplified representation or not, the two-step estimation leads to optimization over fewer parameters in the numerical optimization step. Moreover, the targeting estimator yields consistent estimates of the unconditional variances (given that such exist) under model misspecification which is an advantage of the estimation method, if e.g. the focus is to perform long-horizon forecasts, see Francq et al. (2011) for a comprehensive treatment of this aspect for univariate GARCH models.

Recently Francq et al. (2013) have developed sufficient conditions for consistency and asymptotic normality of the targeting estimator for CCC-GARCH models. In particular and similar to the univariate case, see Francq et al. (2011) and Kristensen and Linton (2004), finite fourth-order moments of the observed process are required in order to establish asymptotic normality. Such moment restrictions for the observed process may not be a realistic assumption in practice, and hence it is highly relevant to consider the limiting distribution of the targeting estimator in the case where the moment restrictions are relaxed. This has recently been done by Vaynman and Beare (2013) for univariate GARCH models, and we extend this to CCC-GARCH models in the present paper. Specifically, we consider the case where the second-order moments are finite (implying consistency of the estimator), but the fourth moments are infinite. In such case a central limit theorem does not apply to the vector of sample variances. As established in e.g. Davis and Mikosch (1998), Mikosch and Stărică (2000), and Basrak et al. (2002b) for univariate (G)ARCH models and in Stărică (1999), Fernández and Muriel (2009), and Basrak and Segers (2009) for multivariate GARCH models, one can exploit that the data generating process has multivariate regularly varying marginal distributions to show that the limiting distribution of sample (autoco)variances is multivariate stable. Using that under certain conditions the CCC-GARCH process is multivariate regularly varying, we establish that the limiting distribution of the vector of sample variances is multivariate stable. Moreover, since the score (in the direction of all other parameters) tends to zero in probability when multiplied by the rate of consistency for the vector of sample variances, the joint targeting estimator has a singular multivariate stable distribution.

The rest of the paper is organized as follows. In Section 2 we introduce the notion of multivariate regular variation, which is a key ingredient for the derivation of the limiting distribution of the targeting estimator. In Section 3 we introduce the targeting CCC-GARCH model, and Section 4 considers the two-step targeting estimation of the model. The large-sample theory for the targeting estimator is presented in Section 5. In particular, we state sufficient conditions for strong consistency, asymptotic normality, and for the estimator to have a multivariate stable limiting distribution. Section 6 concludes the paper. All proofs can be found in the appendix that also includes brief summaries of the notion of vague convergence and point processes (see Appendix D).

Some notation throughout the paper: The absolute value of $a \in \mathbb{R}$ is denoted $|a|$. For $n \in \mathbb{N}$, I_n is the $(n \times n)$ identity matrix, and the zero matrix $O_{m\times n}$ is an $(m \times n)$ matrix with all elements equal to zero. For an $(m \times n)$ matrix $A = [a_{ij}]$ and an $(m \times m)$ matrix $B = [b_{ij}]$, the operators vec, vech, and vech^0 are defined as follows.

$$\begin{aligned}
\underset{(mn\times 1)}{\text{vec}\,(A)} &:= (a_{11}, a_{21}, ..., a_{m1}, a_{12}, ..., a_{m2}, a_{13}, ..., a_{mn})' \\
\underset{(m(m+1)/2\times 1)}{\text{vech}\,(B)} &:= (b_{11}, b_{21}, ...b_{m1}, b_{22}, ..., b_{m2}, b_{33}, ..., b_{mm})' \\
\underset{(m(m-1)/2\times 1)}{\text{vech}^0\,(B)} &:= (b_{21}, b_{31}, ...b_{m1}, b_{32}, ..., b_{m2}, b_{43}, ..., b_{m,m-1})',
\end{aligned}$$

i.e. vec stacks the columns of a matrix, vech stacks the columns from the principal diagonal downwards of a square matrix, and vech^0 stacks the columns below the principal diagonal downwards of a square matrix. The trace of a square matrix A is denoted $\text{tr}\,(A)$, and the determinant is denoted $\det(A)$. The operator diag transforms a vector $a = (a_1, ..., a_m)'$ into a diagonal matrix,

$$\text{diag}\,(a) = \begin{bmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_m \end{bmatrix}.$$

For a $(k \times l)$ matrix $A = [a_{ij}]$ and an $(m \times n)$ matrix B, the Kronecker product of A and B is the $(km \times ln)$ matrix defined by $A \otimes B = [a_{ij}B]$. The Hadamard product of two matrices of the same dimension, $A = [a_{ij}]$ and $B = [b_{ij}]$, is defined as $A \odot B = [a_{ij}b_{ij}]$, and we introduce the non-standard notation $A^{\odot 2} := A \odot A$. With $\xi_1(A), ..., \xi_n(A)$ the n eigenvalues of a matrix A, $\rho(A) := \max_{i\in\{1,...,n\}} |\xi_i(A)|$ is the spectral radius of A. The Euclidean norm of a matrix or vector, A, is defined as $\|A\| = \sqrt{\mathrm{tr}\{A'A\}}$, where A' is the transpose of A. We denote the compactified real line by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. With $\mathbb{1}()$ an indicator function, we denote the point measure concentrated at $x \in \mathbb{R}^d$ as $\delta_x(\cdot)$, i.e. for any Borel set A in $\mathbb{R}^d$, $\delta_x(A) = \mathbb{1}(x \in A)$. Moreover, "$\xrightarrow{p}$", "$\xrightarrow{a.s.}$", "$\xrightarrow{w}$", and "$\xrightarrow{v}$" denote convergence in probability, almost sure convergence, weak convergence, and vague convergence, respectively. (As mentioned vague convergence is introduced in Appendix D.) For two real-valued functions f and g, $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$. For a set, A, we denote its boundary by ∂A. The letters c and ϕ denote positive, finite generic constants always with $\phi < 1$.

2. Multivariate Regular Variation

In this section we give a brief introduction to multivariate regular variation which will show up to be a key tool when we derive the limiting distribution of the targeting estimator in the case where the data generating process does not have finite fourth-order moments. Regular variation is a standard notion for describing heavy-tailed distributions and tail dependence. A classical treatment of univariate regular variation can be found in Bingham et al. (1987), and a fairly recent treatment that includes multivariate regular variation can be found in Resnick (2007), see also Resnick (1986, 1987, 2004) and Mikosch (2004).

A measurable function $f : [0, \infty) \to [0, \infty)$ is said to be regularly varying (at ∞) with index $\kappa \in \mathbb{R}$, if for any $t > 0$ and $x \to \infty$

$$\frac{f(tx)}{f(x)} \to t^{\kappa},$$

see e.g. Resnick (2007, pp.20-21). The parameter κ is called the exponent of variation, and if $\kappa = 0$ f is said to be slowly varying. By definition, if f is regularly varying with index κ one can always write $f(x) = x^{\kappa}L(x)$ where $L(x)$ is slowly varying. In the case of multivariate distributions it is common to define regular variation in terms of convergence of measures, and in particular we will make use of the notion of vague convergence, introduced briefly in Appendix D.1. In order to make use of vague convergence, we consider the space $\overline{\mathbb{R}}^d \setminus \{0\}$ instead of $\mathbb{R}^d$. The reason is that sets that are bounded away from zero in $\mathbb{R}^d$ become relatively compact (i.e. they have compact closures) in $\overline{\mathbb{R}}^d \setminus \{0\}$ with respect to the relative topology, as described in e.g. Resnick (2007, pp.172-175). With $\mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$ the Borel σ-field of $\overline{\mathbb{R}}^d \setminus \{0\}$, recall that a measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$ is a Radon measure if $\mu(K) < \infty$ for all relatively compact $K \in \mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$. We are now ready to define multivariate regular variation.[1]

Definition 2.1 (Mikosch (2004, p.218))**.** A random vector $X \in \mathbb{R}^d$ and its distribution are said to be regularly varying if for a non-null Radon measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$

$$\mu_x(A) := \frac{\mathbb{P}(x^{-1}X \in A)}{\mathbb{P}(\|X\| > x)} \to \mu(A), \tag{2.1}$$

[1]The following definitions hold for any choice of norm, $\|\cdot\|$.

as $x \to \infty$ for any $A \in \mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$ bounded away from the origin with $\mu(\partial A) = 0$, i.e. $\mu_x(\cdot) \xrightarrow{v} \mu(\cdot)$ as $x \to \infty$. The measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$ satisfies the homogeneity property $\mu(tA) = t^{-\kappa}\mu(A)$, $\kappa \geq 0$, for all $t > 0$, and we say that X is multivariate regularly varying with index κ.

Remark 2.1. An equivalent to Definition 2.1 is the following, see Mikosch (2004, p.218). A random vector $X \in \mathbb{R}^d$ and its distribution are said to be regularly varying with index $\kappa \geq 0$ if there exists a probability measure $\varphi(\cdot)$ on the Borel σ-field of $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ (the unit sphere in $\mathbb{R}^d$ with respect to the norm $\|\cdot\|$) such that for all $t > 0$, as $x \to \infty$,

$$\frac{\mathbb{P}\left(\|X\| > tx, \frac{1}{\|X\|}X \in U\right)}{\mathbb{P}(\|X\| > x)} \to t^{-\kappa}\varphi(U), \tag{2.2}$$

for any $U \in \mathcal{B}(\mathbb{S}^{d-1})$, i.e. $\mathbb{P}(\|X\| > tx, \frac{1}{\|X\|}X \in \cdot)/\,\mathbb{P}(\|X\| > x) \xrightarrow{w} t^{-\kappa}\varphi(\cdot)$ as $x \to \infty$. Here $\varphi(\cdot)$ is the spectral measure of X, and one can interpret multivariate regular variation as being characterized by (i) a radial part, $t^{-\kappa}$, stating that the tails are power laws, and (ii) a spherical part, $\varphi(\cdot)$, that quantifies the tail dependence. Observe that if we set $U = \mathbb{S}^{d-1}$ in (2.2),

$$\frac{\mathbb{P}(\|X\| > tx)}{\mathbb{P}(\|X\| > x)} \to t^{-\kappa} \tag{2.3}$$

as $x \to \infty$, i.e. $\mathbb{P}(\|X\| > x)$ is regularly varying with index $-\kappa$. Moreover, if we set $t = 1$, we have that for any $U \in \mathcal{B}(\mathbb{S}^{d-1})$,

$$\frac{\mathbb{P}\left(\|X\| > x, \frac{1}{\|X\|}X \in U\right)}{\mathbb{P}(\|X\| > x)} = \mathbb{P}\left(\frac{1}{\|X\|}X \in U \middle| \|X\| > x\right) \to \varphi(U) \tag{2.4}$$

as $x \to \infty$. So one can interpret the spherical measure of X as the measure determining the distribution of the direction of $\frac{1}{\|X\|}X$, given that $\|X\|$ is large. Moreover, by relations (2.3) and (2.4) we see that the radial and spherical parts characterizing the multivariate regular variation of X become independent for $\|X\|$ sufficiently large.

Remark 2.2. Another way of characterizing the tails of a multivariate distribution is the following considered in e.g. Kesten (1973) and Basrak et al. (2002a,b). Every linear combination of the d-dimensional random vector X is regularly varying if there exists a $\kappa > 0$ and a slowly varying function L such that for all $u \in \mathbb{R}^d \setminus \{0\}$,

$$\lim_{x\to\infty} \frac{\mathbb{P}(u'X > x)}{x^{-\kappa}L(x)} = w(u) \text{ exists.} \tag{2.5}$$

The function w takes finite values and there exists an $u_0 \neq 0$ with $w(u_0) > 0$.
Basrak et al. (2002a) show that if X is multivariate regularly varying, then any linear combination is regularly varying, and under certain conditions one also has that regular variation of all linear combinations of X implies multivariate regular variation.

In Section 5 we exploit that under certain conditions the CCC-GARCH process is multivariate regularly varying and we use this fact, relying on point process techniques, to derive the limiting distribution of the targeting estimator.

3. The Targeting CCC-GARCH Model

Consider the extended CCC-GARCH model of Jeantheau (1998) for $t \in \mathbb{N}$ given by

$$X_t(\theta_C) = \Sigma_t^{1/2}(\theta_C)Z_t, \tag{3.1}$$

where $\{Z_t\}_{t\in\mathbb{N}}$ is an i.i.d.$(0, I_d)$ sequence of random variables. As in Hafner and Preminger (2009b) and Pedersen and Rahbek (2014) we consider a GARCH(1,1)-type model, where the $(d \times d)$ matrix $\Sigma_t^{1/2}(\theta_C)$ is the square-root of $\Sigma_t(\theta_C)$ given by the equations

$$\Sigma_t(\theta_C) = \tilde{D}_t(\theta_C)R\tilde{D}_t(\theta_C), \tag{3.2}$$

$$\tilde{D}_t^2(\theta_C) = \operatorname{diag}\left(\sigma_t^2(\theta_C)\right), \tag{3.3}$$

$$\sigma_t^2(\theta_C) = \omega + AX_{t-1}^{\odot 2}(\theta_C) + B\sigma_{t-1}^2(\theta_C), \tag{3.4}$$

where $X_t^{\odot 2}(\theta_C) := X_t(\theta_C) \odot X_t(\theta_C)$, i.e. the vector of the squared elements of X_t, ω is a vector with strictly positive entries, and R is a $(d \times d)$ positive definite conditional correlation matrix. The $(d \times d)$ matrices A and B have non-negative entries. Moreover, θ_C is the vector of model parameters defined as $\theta_C := [\omega', \operatorname{vec}(A)', \operatorname{vec}(B)', \operatorname{vech}^0(R)']'$, where the subscript C indicates the the model has the classical CCC-GARCH representation. We consider estimation conditional on the initial values X_0 and σ_0^2.

Throughout the text we assume that $\rho(A + B) < 1$ (stated formally in Assumption 3 in Section 5), which by He and Teräsvirta (2004, Section 3) implies that there exists a second-order stationary solution to the CCC-GARCH model. In particular the vector of unconditional variances of X_t exists and is finite, and is given by

$$\gamma := \mathbb{E}\left[X_t^{\odot 2}\right] = \mathbb{E}\left[\sigma_t^2\right] = (I_d - A - B)^{-1}\omega, \tag{3.5}$$

where we have omitted the dependence on θ_C. Targeting can be represented as rewriting the model so that the vector of unconditional variances (of the second-order stationary solution) appears explicitly in the equation for σ_t^2, which gives

$$\sigma_t^2 = (I_d - A - B)\gamma + AX_{t-1}^{\odot 2} + B\sigma_{t-1}^2,$$

and we say that σ_t^2 has the targeting CCC-GARCH representation.

In the next section we discuss estimation of the targeting CCC-GARCH model.

4. Targeting Estimation

With R a positive definite correlation matrix, γ $(d \times 1)$-dimensional with strictly positive entries, and A and B $(d \times d)$-dimensional with non-negative entries, let $\theta := (\gamma', \lambda')'$ denote the vector of parameters, where $\lambda := (\operatorname{vec}(A)', \operatorname{vec}(B)', \operatorname{vech}^0(R)')'$. In terms of the parameters the targeting CCC-GARCH model is given by the equations

$$X_t(\gamma,\lambda) = \Sigma_t^{1/2}(\gamma,\lambda)Z_t, \tag{4.1}$$

$$\Sigma_t(\gamma,\lambda) = \tilde{D}_t(\gamma,\lambda)R\tilde{D}_t(\gamma,\lambda), \tag{4.2}$$

$$\tilde{D}_t^2(\gamma,\lambda) = \operatorname{diag}\left(\sigma_t^2(\gamma,\lambda)\right), \tag{4.3}$$

$$\sigma_t^2(\gamma,\lambda) = (I_d - A - B)\gamma + AX_{t-1}^{\odot 2}(\gamma,\lambda) + B\sigma_{t-1}^2(\gamma,\lambda). \tag{4.4}$$

Note that $\theta \in \Theta := \Theta_\gamma \times \Theta_\lambda \subset (0,\infty)^d \times [0,\infty)^{2d^2} \times (-1,1)^{d(d-1)/2}$, and that the model contains $d + 2d^2 + d\,(d-1)\,/2 =: s_2$ parameters. Let $\theta_0 := (\gamma_0', \lambda_0')'$ denote the vector of true parameters. We consider the estimation method of targeting proposed by Engle and Mezrich (1996) where γ_0 is estimated by method of moments, and λ_0 is estimated by QML estimation conditional on the method of moments estimates.

For a realization $(X_t)_{t\in\{0,\dots,T\}}$ of the targeting CCC-GARCH process with parameter vector θ_0, i.e. $X_t := X_t(\gamma_0, \lambda_0)$, the Gaussian log-likelihood function is given by

$$\begin{aligned} \hat{L}_T(\gamma,\lambda) &:= \frac{1}{T}\sum_{t=1}^{T}\hat{l}_t(\gamma,\lambda), \qquad \text{with} \\ \hat{l}_t(\gamma,\lambda) &:= \log\Big\{\det\Big[\hat{H}_t(\gamma,\lambda)\Big]\Big\} + X_t'\hat{H}_t(\gamma,\lambda)X_t, \end{aligned} \tag{4.5}$$

where the matrix $\hat{H}_t(\gamma,\lambda)$ is given by the equations

$$\hat{H}_t(\gamma,\lambda) = \hat{D}_t(\gamma,\lambda)R\,(\lambda)\,\hat{D}_t(\gamma,\lambda), \tag{4.6}$$

$$\hat{D}_t^2(\gamma,\lambda) = \operatorname{diag}\Big(\hat{h}_t\,(\gamma,\lambda)\Big), \tag{4.7}$$

$$\hat{h}_t\,(\gamma,\lambda) = (I_d - A - B)\,\gamma + AX_{t-1}^{\odot 2} + B\hat{h}_{t-1}\,(\gamma,\lambda)\,. \tag{4.8}$$

In the statistical analysis, the initial value X_0 is, as mentioned, conditioned upon and $\hat{h}_0\,(\gamma,\lambda) := \hat{h} \in (0,\infty)^d$ is fixed.

Targeting estimation relies on estimating the vector of unconditional variances, γ_0, given in (3.5), by method of moments,

$$\hat{\gamma}_T := \frac{1}{T}\sum_{t=1}^{T}X_t^{\odot 2}.$$

Substituting this estimator into the log-likelihood function and minimizing yield the targeting estimator for λ_0,

$$\hat{\lambda}_T := \arg\min_{\lambda\in\Theta_\lambda}\hat{L}_T(\hat{\gamma}_T,\lambda). \tag{4.9}$$

The two steps yield the targeting estimator, a hybrid of method of moments and QML, of θ_0, $\hat{\theta}_T := (\hat{\gamma}_T', \hat{\lambda}_T')'$.

Remark 4.1. The targeting procedure reduces the number of parameters computed by numerical optimization of the log-likelihood compared to classical QML estimation. In the first step d parameters are estimated by method of moments, and in the second step $d\,(d-1)\,/2 + 2d^2$ parameters are estimated by optimization of the likelihood. Targeting estimation may in particular be used to estimate simplified CCC models such as diagonal and scalar models, where A and B are diagonal matrices and scalars, respectively. Combining targeting estimation with a simplified model decreases the number of varying parameters (in the second step) relative to the total number of parameters additionally.

In the next section we consider the large-sample properties of the targeting estimator.

5. Large-Sample Theory for the Targeting Estimator

Sufficient conditions for consistency and asymptotic normality of targeting estimators have been considered by Kristensen and Linton (2004) and Francq et al. (2011) for univariate GARCH models and by Pedersen and Rahbek (2014) for multivariate BEKK-GARCH models, and, as mentioned, in recent work by Francq et al. (2013) for the CCC-GARCH model. At least finite fourth-order moments are needed in order to derive asymptotic normality of the targeting estimator for these models. Such moment conditions may not be satisfied in practice, and in line with Vaynman and Beare (2013) we consider the limiting distribution of the targeting estimator in the case where the estimator is consistent but asymptotic normality is infeasible. Our results are new and extend the existing literature on targeting estimation. The proofs are given in Appendix A.

For completeness we start out by stating sufficient conditions for strong consistency of the targeting estimator for the true parameter vector θ_0. Observe that X_t can be written as $X_t = \tilde{D}_t(\gamma_0, \lambda_0)\varepsilon_t$, where

$$\varepsilon_t := R_0^{1/2} Z_t, \tag{5.1}$$

i.e. (ε_t) is i.i.d.$(0, R_0)$ with $R_0 := R(\theta_0)$.

Assumption 1. *The process (X_t) is strictly stationary and ergodic.*

Assumption 2. *The distribution of ε_t, defined in (5.1), admits a density strictly positive on $\mathbb{R}^d$. Moreover, with $\varepsilon_{t,j}$* the j-th element of ε_t, *$j = 1, ..., d$, either $\mathbb{E}[|\varepsilon_{t,j}|^{\beta}] < \infty$ for all $\beta > 0$ or there exists a $\beta_0 > 1$ such that $\mathbb{E}[|\varepsilon_{t,j}|^{\beta}] < \infty$ for $\beta < \beta_0$ and $\mathbb{E}[|\varepsilon_{t,j}|^{\beta_0}] = \infty$.*

Assumption 3. *For all $\lambda \in \Theta_\lambda$, R is a positive definite correlation matrix and $\rho\,(A + B) < 1$.*

Assumption 4. *The true parameters $\lambda_0 \in \Theta_\lambda$ and Θ_λ is compact.*

Moreover, in light of Assumption 1 it is convenient to introduce the strictly stationary and ergodic process $\{h_t\,(\gamma, \lambda)\}$ given recursively by

$$h_t\,(\gamma, \lambda) = (I_d - A - B)\,\gamma + AX_{t-1}^{\odot 2} + Bh_{t-1}\,(\gamma, \lambda)\,. \tag{5.2}$$

Assumption 5. *There exists a $t \in \mathbb{Z}$ such that if for $\lambda \in \Theta_\lambda$, $h_t(\gamma_0, \lambda) = h_t(\gamma_0, \lambda_0)$ a.s. and $R\,(\lambda) = R(\lambda_0)$, then $\lambda = \lambda_0$.*

Remark 5.1. Assumption 1 is in line with existing literature on QML estimation of multivariate GARCH and targeting estimation, see e.g. Comte and Lieberman (2003), Hafner and Preminger (2009a), Francq and Zakoïan (2012), Pedersen and Rahbek (2014), and Francq et al. (2011). Necessary and sufficient conditions for the existence of a unique strictly stationary and ergodic solution to the model are stated in Francq and Zakoïan (2010, Theorem 11.6) and Boussama (1998, Section 5.4). In particular Assumption 1 implies that for the asymptotic analysis the process $(X_t)_{t=0,1,...}$ is assumed to be initiated from the invariant distribution. As discussed in Pedersen and Rahbek (2014) one might relax this assumption and establish consistency and asymptotic normality when allowing for an arbitrary initial value X_0 of the data generating process.

Remark 5.2. The restrictions in Assumption 3 are natural as we only want to consider parameters where $\hat{H}_t(\gamma, \lambda)$ is positive definite and the elements of $\hat{h}_t$ are strictly positive.

Remark 5.3. Assumption 5 is a high-level identification condition, primitive identification conditions are discussed in Francq and Zakoïan (2010, Section 11.4.1).

As stated in the following theorem, the assumptions above are sufficient for strong consistency of the estimator. Strong consistency does apply under milder conditions, and, in particular, ε_t does not need to have a strictly positive density on $\mathbb{R}^d$ (Assumption 2), see e.g. Francq et al. (2011, Theorem 2.1). However, Assumptions 1 and 2 are used in order to establish multivariate regular variation of the CCC-GARCH process, see Lemma B.2, and hence needed later when we consider the limiting distribution of the targeting estimator in the case of infinite fourth moments.

Theorem 5.1. *Suppose that Assumptions 1-5 hold. Then $\hat{\theta}_T \overset{a.s.}{\to} \theta_0$ as $T \to \infty$.*

The assumption of finite second-order moments of X_t, implied by Assumption 3, is in line with the moment restrictions assumed in Francq et al. (2011) for consistency of the targeting estimator in the univariate case. The moment restrictions are stronger than the ones required for consistency of the QMLE of the classical CCC representation in (3.4), where only a fractional moment of X_t is required to be finite, see e.g. Francq and Zakoïan (2010, Theorem 11.7).

Next, we turn to the limiting distribution of the targeting estimator. The idea is to assume that the vector $Y_t := [X_t^{\odot 2\prime}, \sigma_t^2(\theta_0)']'$ is multivariate regularly varying with index $\kappa \in (1, 2)$, see Definition 2.1, implying that $\mathbb{E}[\|Y_t\|^2] = \infty$, i.e. X_t has infinite fourth moments. This assumption is reasonable because, as stated in Lemma B.2 in Appendix B, Assumptions 1-5 imply that any linear combination of $\sigma_t^2(\theta_0)$ is regularly varying with some index $\kappa > 0$, see Remark 2.2. This property holds by an application of Kesten's theorem, see Kesten (1973) and Basrak et al. (2002b, Theorem 2.4), and has been verified for CCC-GARCH processes in Fernández and Muriel (2009) and Stărică (1999). As also stated in Lemma B.2, when $\kappa \in (1, 2)$ it holds that Y_t is multivariate regularly varying. We limit ourselves to the case where $\kappa \in (1, 2)$, because the case where $\kappa = 2$ leads to very complicated derivations, see e.g. Basrak et al. (2002b), and we leave such analysis to elsewhere. Moreover, the case where $\kappa \in (0, 1]$, implying that the second-order moments of X_t are infinite, is ruled out by Assumption 2 and that $\rho(A_0 + B_0) < 1$ (Assumption 3), and it would not be of much use to do targeting estimation when the variance of X_t is infinite. In the case where $\kappa > 2$ we have that X_t has finite fourth-order moments and the limiting distribution of the targeting estimator is Gaussian as stated in Remark 5.5 below.

We emphasize that a sufficient condition for X_t having infinite fourth moments is that ε_t has infinite fourth moments. Indeed such condition is not necessary, and in contrast to e.g. Hall and Yao (2003) and Mikosch and Straumann (2006) who (for univariate GARCH processes) introduce heavy tails to X_t through heavy tails of the noise process, we here do only assume that the noise process has at least finite second-order moments, but may not be heavy-tailed. I.e. we can have that X_t has infinite fourth-order moments even if the noise process is Gaussian. A necessary and sufficient condition for finite fourth-moments of X_t is given in Appendix C.

In order to derive the limiting distribution of the targeting estimator, we make the following standard assumption that enables us to make a mean-value expansion of the log-likelihood function.

Assumption 6. *The true parameter vector, λ_0, belongs to the interior of Θ_λ.*

We are now able to state the following result.

Theorem 5.2. *Under Assumptions 1-6 suppose that $Y_t := [X_t^{\odot 2\prime}, \sigma_t^2(\theta_0)']'$ is multivariate regularly varying with index $\kappa \in (1,2)$. Then there exists a sequence (a_T), $0 < a_T \to \infty$ as $T \to \infty$, such that $T\,\mathbb{P}(\|Y_t\| > a_T) \to 1$ as $T \to \infty$, and*

$$Ta_T^{-1}\begin{pmatrix}\hat{\gamma}_T - \gamma_0 \\ \hat{\lambda}_T - \lambda_0\end{pmatrix} \xrightarrow{w} \begin{pmatrix} I_d \\ \left(J_0^\lambda\right)^{-1} J_0^\gamma\end{pmatrix} S,$$

where the matrices J_0^λ and J_0^γ are stated in (A.21) in Appendix A, and S has a d-dimensional multivariate κ-stable distribution.

The proof of the above theorem is somewhat involved, and we refer to the appendix for details. Briefly, the proof relies on exploiting the multivariate regular variation of Y_t and showing that the point process $\sum_{t=1}^T \delta_{a_T^{-1}Y_t}(\cdot)$ has a limiting distribution. Next, loosely speaking, one can show that $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$ is approximately a continuous mapping of $\sum_{t=1}^T \delta_{a_T^{-1}Y_t}(\cdot)$, so by the continuous mapping theorem we can find the limiting distribution of $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$ that shows up to be multivariate stable. The conclusion of the theorem now follows by showing that the score in the direction of all the other parameters tends to zero in probability when multiplied by Ta_T^{-1} and that the Hessians converge properly. For an introduction to multivariate stable distributions we refer to Samorodnitsky and Taqqu (1994, Chapter 2).

Remark 5.4. The limiting distribution in Theorem 5.2 is singular, and the reason is that the score in the direction of all the other parameters tends to zero in probability when multiplied by Ta_T^{-1}, see the proof of the theorem for additional details.

Remark 5.5. If we in addition to Assumptions 1-6 have that $\mathbb{E}[\|X_t\|^4] < \infty$, it follows that as $T \to \infty$

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{w} \mathcal{N}\left(0, \Sigma_0\right)$$

for some matrix Σ_0, which has been shown by Francq et al. (2013). As mentioned, a necessary and sufficient condition for $\mathbb{E}[\|X_t\|^4] < \infty$ is given in Appendix C.

Theorem 5.2 states that in the case where X_t has finite second-order moments but infinite fourth moments, the targeting estimator obeys consistency at rate Ta_T^{-1}, and the limiting distribution is multivariate stable. Observing that $a_T \sim (cT)^{1/\kappa}$, see Basrak et al. (2002b, Remark 2.11), the rate of convergence is slower than $\sqrt{T}$, which is the rate of consistency in the case with finite fourth-order moments. Specifically, one should not construct confidence intervals based on the Gaussian distribution. As pointed out by Vaynman and

Beare (2013), one can make use of subsampling techniques to construct confidence intervals for the targeting estimator. For the univariate GARCH model, Vaynman and Beare (2013) perform numerical simulations based on the subsampling techniques proposed by Politis et al. (1999, Ch.11), and find that such techniques do not perform well. We expect the same conclusion to hold for the even more complicated CCC-GARCH model.

6. Concluding remarks and extensions

In this paper we have considered the targeting estimator for the extended CCC-GARCH model. In particular we have investigated the limiting distribution of the estimator in the case where the fourth moments of the observed process are infinite. By exploiting that under certain conditions the CCC-GARCH process is multivariate regularly varying, we have shown that the rate of consistency is slower than $\sqrt{T}$ and that the limiting distribution of the estimator is multivariate stable. Hence one should not make use of confidence intervals based on the Gaussian distribution when heavy tails are present.

One obvious way of extending the theory derived in this paper is to consider tail trimming, as recently proposed by Hill and Renault (2012) for the variance targeting estimator in univariate GARCH models, where asymptotic normality may apply in the case of heavy tails. Moreover, one could consider different ways of obtaining confidence intervals for the targeting estimator than the subsampling method of Politis et al. (1999, Ch.11) considered by Vaynman and Beare (2013). One potential problem with that particular method is that the rate of convergence, Ta_T^{-1}, has to be estimated which may imply additional uncertainty. Alternatively, one could consider the subsampling method of Sherman and Carlstein (2004) where unknown rates of convergence are allowed for. We leave such studies to future research.

Appendix A. Proofs of theorems

Recall from (4.5) that the log-likelihood function is given by

$$\hat{L}_T(\gamma,\lambda) \quad = \quad \frac{1}{T}\sum_{t=1}^{T}\hat{l}_t(\gamma,\lambda),$$

with

$$\hat{l}_t(\gamma,\lambda) \quad = \quad \log\left\{\det\left[\hat{H}_t(\gamma,\lambda)\right]\right\} + X_t'\hat{H}_t^{-1}(\gamma,\lambda)X_t,$$

where $\hat{H}_t(\gamma,\rho,\lambda)$ is given by the recursions

$$\begin{aligned}
\hat{H}_t(\gamma,\rho,\lambda) &= \hat{D}_t(\gamma,\lambda)R(\lambda)\hat{D}_t(\gamma,\lambda)\\
\hat{D}_t^2(\gamma,\lambda) &= \operatorname{diag}\left(\hat{h}_t(\gamma,\lambda)\right)\\
\hat{h}_t(\gamma,\lambda) &= (I_d - A - B)\,\gamma + AX_{t-1}^{\odot 2} + B\hat{h}_{t-1}(\gamma,\lambda),
\end{aligned}$$

and $\hat{h}_0(\gamma,\lambda) := \hat{h}$ fixed with strictly positive entries. For technical reasons, we have also in Section 5 introduced the strictly stationary and ergodic process $\{h_t(\gamma,\lambda)\}$ given by the recursion (5.2). To distinguish between $\hat{h}_t(\gamma,\lambda)$ and $h_t(\gamma,\lambda)$ we introduce correspondingly

$$\begin{aligned} H_t(\gamma,\lambda) &= D_t(\gamma,\lambda)R(\lambda)D_t(\gamma,\lambda), \\ D_t^2(\gamma,\lambda) &= \operatorname{diag}\left(h_t(\gamma,\lambda)\right) \end{aligned}$$

and

$$L_T(\gamma,\lambda) := \frac{1}{T}\sum_{t=1}^{T} l_t(\gamma,\lambda) \tag{A.1}$$

where

$$l_t(\gamma,\lambda) := \log\{\det[H_t(\gamma,\lambda)]\} + X_t' H_t^{-1}(\gamma,\lambda)X_t. \tag{A.2}$$

Observe that both $\hat{h}_t(\gamma,\lambda)$ and $h_t(\gamma,\lambda)$ are defined for the same (by Assumption 1) strictly stationary and ergodic (X_t) generated with θ_0. Moreover, we have that by definition

$$\Sigma_t(\gamma_0,\lambda_0) = H_t(\gamma_0,\lambda_0), \quad \tilde{D}_t(\gamma_0,\lambda_0) = D_t(\gamma_0,\lambda_0), \quad \sigma_t^2(\gamma_0,\lambda_0) = h_t(\gamma_0,\lambda_0) \tag{A.3}$$

for all t.

Proof of Theorem 5.1. By $E[\|X_t\|^2] < \infty$ and the ergodic theorem, see e.g. Billingsley (1995, Theorem 24.1), we have that $\widehat{\gamma}_T$ is strongly consistent for γ_0. Hence it remains to show that $\widehat{\lambda}_T$ is strongly consistent for λ_0. Following Francq et al. (2011, Appendix A.1), and due to the compactness of Θ_λ, it suffices to verify the following three conditions:

(i) As $T \to \infty$,

$$\sup_{\lambda\in\Theta_\lambda} \left| L_T(\gamma_0,\lambda) - \hat{L}_T(\widehat{\gamma}_T,\lambda)\right| \overset{a.s.}{\to} 0$$

(ii) $\mathbb{E}[|l_t(\gamma_0,\lambda_0)|] < \infty$ and for $\lambda \in \Theta_\lambda$, if $\lambda \neq \lambda_0$ then $\mathbb{E}[l_t(\gamma_0,\lambda)] > \mathbb{E}[l_t(\gamma_0,\lambda_0)]$.

(iii) Any $\lambda \in \Theta_\lambda$, $\lambda \neq \lambda_0$, has a neighborhood $\mathcal{V}(\lambda)$ such that almost surely,

$$\liminf_{T\to\infty} \inf_{\lambda^\star\in\mathcal{V}(\lambda)} \hat{L}_T(\widehat{\gamma}_T,\lambda^\star) > \mathbb{E}[l_t(\gamma_0,\lambda_0)].$$

First, by the triangle inequality,

$$\begin{aligned} \sup_{\lambda\in\Theta_\lambda} \left| L_T(\gamma_0,\lambda) - \hat{L}_T(\widehat{\gamma}_T,\lambda)\right| &\leq \sup_{\lambda\in\Theta_\lambda} \left| L_T(\gamma_0,\lambda) - L_T(\widehat{\gamma}_T,\lambda)\right|, \\ &\quad + \sup_{\lambda\in\Theta_\lambda} \left| L_T(\widehat{\gamma}_T,\lambda) - \hat{L}_T(\widehat{\gamma}_T,\lambda)\right|. \end{aligned} \tag{A.4}$$

Next, observe that

$$\sup_{\lambda\in\Theta_\lambda} \left| L_T(\gamma_0,\lambda) - L_T(\widehat{\gamma}_T,\lambda)\right| \leq \frac{1}{T}\sum_{t=1}^{T} \sup_{\lambda\in\Theta_\lambda} \left| l_t(\gamma_0,\lambda) - l_t(\widehat{\gamma}_T,\lambda)\right|.$$

By the mean-value theorem,

$$l_t(\gamma_0,\lambda) - l_t(\widehat{\gamma}_T,\lambda) = \frac{\partial l_t(\gamma^\star,\lambda)}{\partial\gamma'}(\widehat{\gamma}_T - \gamma_0)$$

where $\gamma^\star$ lies between $\widehat{\gamma}_T$ and γ_0, as in Jensen and Rahbek (2004, Proof of Lemma 1). For T sufficiently large, by the consistency of $\widehat{\gamma}_T$,

$$\sup_{\lambda\in\Theta_\lambda} \left| l_t(\gamma_0,\lambda) - l_t(\widehat{\gamma}_T,\lambda)\right| \leq \sup\nolimits_{\theta\in\Theta_\gamma\times\Theta_\lambda} \left| \frac{\partial l_t(\gamma,\lambda)}{\partial\gamma'}(\widehat{\gamma}_T - \gamma_0)\right|, \tag{A.5}$$

where Θ_γ is chosen to be a compact subset of $(0,\infty)^d$ containing γ_0 and $\hat{\gamma}_T$ for T sufficiently large. Next observe that

$$\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left|\frac{\partial l_t(\gamma,\lambda)}{\partial\gamma'}\left(\widehat{\gamma}_T-\gamma_0\right)\right|\leq\sum_{i=1}^d\left|\widehat{\gamma}_{T,i}-\gamma_{0,i}\right|\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left|\frac{\partial l_t(\gamma,\lambda)}{\partial\gamma_i}\right|,\tag{A.6}$$

where $\widehat{\gamma}_{T,i}$ and $\gamma_{0,i}$ denote the i-th elements of $\widehat{\gamma}_T$ and γ_0, respectively, and $\partial l_t(\gamma,\lambda)/\partial\gamma_i$ denotes the derivative l_t with respect to the i-th element of γ. From Francq and Zakoïan (2010, equation (11.67)), suppressing the dependence on θ,

$$\begin{aligned}\frac{\partial l_t(\gamma,\lambda)}{\partial\gamma_i} &= -\operatorname{tr}\left\{\left(X_tX_t'D_t^{-1}R^{-1}+R^{-1}D_t^{-1}X_tX_t'\right)D_t^{-1}\frac{\partial D_t}{\partial\gamma_i}D_t^{-1}\right\}\\ &\quad +2\operatorname{tr}\left(D_t^{-1}\frac{\partial D_t}{\partial\gamma_i}\right),\end{aligned}\tag{A.7}$$

with

$$\frac{\partial D_t}{\partial\gamma_i}=\frac{1}{2}D_t^{-1}\operatorname{diag}\left(\frac{\partial h_t}{\gamma_i}\right).\tag{A.8}$$

Furthermore, from Francq and Zakoïan (2010, p. 297)

$$\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left\|D_t^{-1}(\gamma,\lambda)\right\|\leq c\quad\text{and}\quad\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left\|H_t^{-1}(\gamma,\lambda)\right\|\leq c.\tag{A.9}$$

Assumption 3 and Lemma B.8 imply that

$$\text{for any }\lambda\in\Theta_\lambda,\quad\rho\left(B\right)<1.\tag{A.10}$$

Recursions give that for any θ

$$\frac{\partial h_t}{\partial\gamma_i}=\sum_{j=0}^\infty B^j\left(I_d-A-B\right)\frac{\partial\gamma}{\partial\gamma_i},$$

which is finite by (A.10) and the compactness of Θ_λ. Hence,

$$\sup_{\theta\in\Theta_\lambda\times\Theta_\gamma}\left\|\frac{\partial h_t}{\partial\gamma_i}\right\|<\infty.\tag{A.11}$$

Now, in light of (A.7), (A.8), (A.9), (A.11), Assumption 4, and $E\left\|X_t\right\|^2<\infty$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left|\frac{\partial l_t(\gamma,\lambda)}{\partial\gamma_i}\right|\right]<\infty,\quad i=1,...,d.\tag{A.12}$$

Using (A.5), (A.6), (A.12), together with the ergodic theorem and the consistency of $\widehat{\gamma}_T$, we have that $\sup_{\lambda\in\Theta_\lambda}\left|L_T\left(\gamma_0,\lambda\right)-L_T\left(\widehat{\gamma}_T,\lambda\right)\right|\overset{a.s.}{\to}0$ as $T\to\infty$. Turning to the second part of (A.4),

$$\sup_{\lambda\in\Theta_\lambda}\left|L_T(\widehat{\gamma}_T,\lambda)-\hat{L}_T(\widehat{\gamma}_T,\lambda)\right|\leq\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left|L_T(\gamma,\lambda)-\hat{L}_T(\gamma,\lambda)\right|,$$

where again Θ_γ is chosen to be a compact subset of $(0,\infty)^d$ containing γ_0 and $\hat{\gamma}_T$ for T sufficiently large. From Francq and Zakoïan (2010, p.298),

$$\sup_{\theta\in\Theta_\gamma\times\Theta_\lambda}\left|L_T(\gamma,\lambda)-\hat{L}_T(\gamma,\lambda)\right|\overset{a.s}{\to}0$$

as $T\to\infty$, and we conclude that (i) holds.

(ii) follows by arguments similar to the ones stated in Francq and Zakoïan (2010, pp.298-299).

Turning to (iii), as in Francq et al. (2011, Appendix A.1) for all $\lambda \in \Theta_\lambda, \lambda \neq \lambda_0$, as $T \to \infty$, almost surely,

$$\begin{aligned}
\liminf_{T\to\infty} \inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} \hat{L}_T(\widehat{\gamma}_T, \lambda^\star) &\geq \liminf_{T\to\infty} \inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} L_T(\gamma_0, \lambda^\star) \\
&\quad + \liminf_{T\to\infty} \inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} \left(\hat{L}_T(\widehat{\gamma}_T, \lambda^\star) - L_T(\gamma_0, \lambda^\star) \right) \\
&= \liminf_{T\to\infty} \inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} L_T(\gamma_0, \lambda^\star) \\
&\geq \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} l_t(\gamma_0, \lambda^\star) \\
&= \mathbb{E}\left[\inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} l_t(\gamma_0, \lambda^\star) \right],
\end{aligned}$$

where the first equality follows from (i), and the last equality follows by applying the strong law of large numbers to the ergodic process $\left\{ \inf_{\lambda^\star \in \mathcal{V}(\lambda)\cap\Theta_\lambda} l_t(\gamma_0, \lambda) \right\}$ using that for all $\lambda \in \Theta_\lambda$, $\mathbb{E}[\max\{-l_t(\gamma_0, \lambda), 0\}] < \infty$, as in Francq and Zakoïan (2010, p.298). In light of (ii), we conclude that (iii) holds. □

Proof of Theorem 5.2. First observe that (a_T) exists by Lemma B.4. Define $s_1 := s_2 - d$, where s_2 is the dimension of θ. Using the definition of $\hat{\lambda}_T$ in (4.9) and Assumption 6, consider a mean-value expansion of the first derivative of the log-likelihood function around θ_0,

$$O_{s_1 \times 1} = \frac{\partial \hat{L}_T(\theta_0)}{\partial \lambda} + \hat{J}_T^\lambda \left(\hat{\lambda}_T - \lambda_0 \right) + \hat{J}_T^\gamma \left(\hat{\gamma}_T - \gamma_0 \right), \tag{A.13}$$

where

$$\hat{J}_T^\lambda := \frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \lambda \partial \lambda'}, \quad \text{and} \quad \hat{J}_T^\gamma := \frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \lambda \partial \gamma'},$$

for some $\theta^\star$ between $\hat{\theta}_T$ and θ_0, as in Jensen and Rahbek (2004, Proof of Lemma 1).

Initially, we want to verify the following points.

(iv) The matrix $J_0^\lambda := \mathbb{E}\left[\frac{\partial^2 l_t(\theta_0)}{\partial \lambda \partial \lambda'}\right]$ is non-singular and

$$\mathbb{E}\left[\left\| \frac{\partial l_t^2(\theta_0)}{\partial \theta \partial \theta'} \right\| \right] < \infty.$$

(v) For $i, j = 1, \ldots, s_2$, as $T \to \infty$,

$$\frac{\partial^2 L_T(\theta^\star)}{\partial \theta_i \partial \theta_j} \xrightarrow{p} \mathbb{E}\left[\frac{\partial l_t^2(\theta_0)}{\partial \theta_i \partial \theta_j'} \right].$$

(vi) There exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that for $i, j = 1, \ldots, s_2$, as $T \to \infty$,

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 L_T(\theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \hat{L}_T(\theta)}{\partial \theta_i \partial \theta_j} \right| \xrightarrow{p} 0.$$

(vii) For $i = d+1, ..., s_2$, as $T \to \infty$,

$$\left| Ta_T^{-1} \left[\frac{\partial L_T(\theta_0)}{\partial \theta_i} - \frac{\partial \hat{L}_T(\theta_0)}{\partial \theta_i} \right] \right| \overset{p}{\to} 0.$$

We choose $\mathcal{V}(\theta_0)$ sufficiently small such that all parameters in A, B, and γ are bounded away from zero on $\mathcal{V}(\theta_0)$. All the points (iv)-(vii) can be verified by arguments similar to the ones in Francq and Zakoïan (2012, Section A.4.2) and Francq and Zakoïan (2010, Section 11.4.3). In particular, point (iv) follows by Lemma B.1 and Assumption 5, and (v) follows by Lemma B.1 and Theorem 5.1. The points (vi)-(vii) do not depend on the parametrization and are verified along the lines of Francq and Zakoïan (2012, pp.204-206), with the latter point following by observing that $a_T \sim (cT)^{1/\kappa}$, see Basrak et al. (2002b, Remark 2.11).

Let $s_0 := (d + 2d^2)$. From Francq and Zakoïan (2012, p.198) we have that for $i = d+1, ..., s_0$,

$$\frac{\partial l_t(\theta_0)}{\partial \theta_i} = \operatorname{tr}\left\{ \left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right) \left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right) \right\} + \operatorname{tr}\left\{ \left(I_d - R_0^{-1} \varepsilon_t \varepsilon_t' \right) \left(\frac{\partial D_{0t}}{\partial \theta_i} D_{0t}^{-1} \right) \right\}, \tag{A.14}$$

and for $i = s_0 + 1, ..., s_2$,

$$\frac{\partial l_t(\theta_0)}{\partial \theta_i} = \operatorname{tr}\left\{ \left(I_d - R_0^{-1} \varepsilon_t \varepsilon_t' \right) \left(R_0^{-1} \frac{\partial R_0}{\partial \theta_i} \right) \right\}. \tag{A.15}$$

Observe that

$$\operatorname{tr}\left\{ \left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right) \left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right) \right\} = \sum_{j=1}^{d^2} \operatorname{vec}\left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right)_j \operatorname{vec}\left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right)_j, \tag{A.16}$$

where $\operatorname{vec}(I_d - \varepsilon_t \varepsilon_t' R_0^{-1})_j$ is the j-th element of $\operatorname{vec}(I_d - \varepsilon_t \varepsilon_t' R_0^{-1})$. As in Vaynman and Beare (2013, Proof of Theorem 3.3), with β_0 introduced in Assumption 2, let $\alpha \in (\kappa, \min\{\beta_0, 2\})$. Then for any $\delta > 0$

$$\begin{aligned}
& \mathbb{P}\left(\left| a_T^{-1} \sum_{t=1}^{T} \operatorname{tr}\left\{ \left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right) \left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right) \right\} \right| > \delta \right) \\
\leq\ & \sum_{j=1}^{d^2} \mathbb{P}\left(\left| a_T^{-1} \sum_{t=1}^{T} \operatorname{vec}\left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right)_j \operatorname{vec}\left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right)_j \right| > \delta / d^2 \right) \\
\leq\ & \sum_{j=1}^{d^2} \left(\delta / d^2 \right)^{-\alpha} \mathbb{E}\left[\left| a_T^{-1} \sum_{t=1}^{T} \operatorname{vec}\left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right)_j \operatorname{vec}\left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right)_j \right|^{\alpha} \right] \\
\leq\ & c \sum_{j=1}^{d^2} 2T a_T^{-\alpha} \mathbb{E}\left[\left| \operatorname{vec}\left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right)_j \operatorname{vec}\left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right)_j \right|^{\alpha} \right] \\
=\ & 2cT a_T^{-\alpha} \sum_{j=1}^{d^2} \mathbb{E}\left[\left| \operatorname{vec}\left(I_d - \varepsilon_t \varepsilon_t' R_0^{-1} \right)_j \right|^{\alpha} \right] \mathbb{E}\left[\left| \operatorname{vec}\left(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i} \right)_j \right|^{\alpha} \right] \\
\leq\ & cT a_T^{-\alpha} \to 0,
\end{aligned} \tag{A.17}$$

where the first inequality follows by (A.16) and the triangle inequality, the second inequality follows by the generalized Chebyshev inequality, the third inequality follows by the fact that $\{\operatorname{vec}(I_d - \varepsilon_t \varepsilon_t' R_0^{-1})_j \operatorname{vec}(D_{0t}^{-1} \frac{\partial D_{0t}}{\partial \theta_i})_j, \mathcal{F}_t\}$ is a martingale difference sequence together with von Bahr and Esseen (1965, Theorem 2), and the fourth inequality follows by Lemma B.1. By similar arguments, we have that for any $\delta > 0$

$$\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{\left(I_d - R_0^{-1}\varepsilon_t\varepsilon_t'\right)\left(\frac{\partial D_{0t}}{\partial\theta_i}D_{0t}^{-1}\right)\right\}\right| > \delta\right) \to 0, \tag{A.18}$$

and

$$\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{\left(I_d - R_0^{-1}\varepsilon_t\varepsilon_t'\right)\left(R_0^{-1}\frac{\partial R_0}{\partial\theta_i}\right)\right\}\right| > \delta\right) \to 0. \tag{A.19}$$

Considering (A.14) and (A.15), in light of (A.17), (A.18), and (A.19) we have that for any $i = d+1, ..., s_2$

$$a_T^{-1}T\frac{\partial L_T(\theta_0)}{\partial\theta_i} \xrightarrow{p} 0. \tag{A.20}$$

By (iv)-(vi) we have that $\hat{J}_T^{\lambda}$ is invertible with probability approaching one for T sufficiently large, so with probability approaching

$$Ta_T^{-1}\begin{pmatrix}\hat{\gamma}_T - \gamma_0\\ \hat{\lambda}_T - \lambda_0\end{pmatrix} = \begin{pmatrix} I_d & O_{d\times s_1}\\ -\left(\hat{J}_T^{\lambda}\right)^{-1}\hat{J}_T^{\gamma} & -\left(\hat{J}_T^{\lambda}\right)^{-1}\end{pmatrix} Ta_T^{-1}\begin{pmatrix}\hat{\gamma}_T - \gamma_0\\ \frac{\partial\hat{L}_T(\theta_0)}{\partial\lambda}\end{pmatrix}.$$

Now (iv)-(vii), (A.20), and Slutsky's lemma yield

$$Ta_T^{-1}\begin{pmatrix}\hat{\gamma}_T - \gamma_0\\ \hat{\lambda}_T - \lambda_0\end{pmatrix} \xrightarrow{w} \begin{pmatrix} I_d \\ \left(J_0^{\lambda}\right)^{-1} J_0^{\gamma}\end{pmatrix} S,$$

where we have used Lemma B.3, and where J_0^{λ} and J_0^{γ} are given by

$$J_0^{\lambda} := \mathbb{E}\left[\frac{\partial^2 l\left(\theta_0\right)}{\partial\lambda\partial\lambda'}\right], \quad \text{and} \quad J_0^{\gamma} := \mathbb{E}\left[\frac{\partial^2 l_t\left(\theta_0\right)}{\partial\lambda\partial\gamma'}\right]. \tag{A.21}$$

Observe that $(I_d, [(J_0^{\lambda})^{-1}J_0^{\gamma}]')'S$ has a multivariate stable distribution with index $\kappa \in (1,2)$ due to Samorodnitsky and Taqqu (1994, Theorems 2.1.2 and 2.5.1(c)). □

Appendix B. Lemmata

Lemma B.1. *Under Assumptions 1-6, there exists a neighborhood of θ_0, $\mathcal{V}(\theta_0)$, such that for all $i_1 = 1, ..., d$, all $i, j, k = 1, ..., s_2 - d(d-1)/2$ and any $r_0 \geq 1$,*

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{1}{h_{t,i_1}}\frac{\partial h_{t,i_1}}{\partial\theta_i}(\theta)\right|^{r_0}\right] < \infty, \tag{B.1}$$

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{1}{h_{t,i_1}}\frac{\partial^2 h_{t,i_1}}{\partial\theta_i\partial\theta_j}(\theta)\right|^{r_0}\right] < \infty, \tag{B.2}$$

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{1}{h_{t,i_1}}\frac{\partial^3 h_{t,i_1}}{\partial\theta_i\partial\theta_j\partial\theta_k}(\theta)\right|^{r_0}\right] < \infty, \tag{B.3}$$

and

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{h_{0t,i_1}}{h_{t,i_1}}\right|^{r_0}\right] < \infty \tag{B.4}$$

where h_{t,i_1} and h_{0t,i_1} denote element i_1 of $h_t(\theta)$ and $h_t(\theta_0)$, respectively.

Proof. We choose $\mathcal{V}(\theta_0) \subset \Theta$ such that all elements of γ, A, and B are bounded away from zero on $\mathcal{V}(\theta_0)$. Let $h_t := h_t(\theta)$. Considering (B.1), recursions give that $h_t = \sum_{i=0}^{\infty} B^i[(I_d -$

$A-B)\gamma+AX_{t-1-i}^{\odot 2}]$, where we have used that $\rho(B)<1$. For $i=1,...,d$ and any i_1 and $r_0>0$ we have that

$$E\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{\partial h_{t,i_1}}{\partial\theta_i}(\theta)\right|^{r_0}\right]<\infty,$$

see (A.11). For $i=d+1,...,d^2$ and any i_1

$$\theta_i\frac{\partial h_{t,i_1}}{\partial\theta_i}\leq h_{t,i_1},$$

so indeed (B.1) holds for $i=d+1,...,d+d^2$. Moreover, for $i=d+d^2+1,...,d+2d^2$

$$\begin{aligned}\frac{\partial h_t}{\partial\theta_i} &= \sum_{k=0}^{\infty}\left(\frac{\partial}{\partial\theta_i}B^k\right)\left[(I_d-A-B)\,\gamma+AX_{t-1-k}^{\odot 2}\right]\\ &\quad+\sum_{k=0}^{\infty}B^k\left[-\frac{\partial B}{\partial\theta_i}\gamma\right]\\ &=: W_t^{(1)}+W_t^{(2)}.\end{aligned}\tag{B.5}$$

First define $f_t:=(I_d-A-B)\,\gamma+AX_{t-1}^{\odot 2}$ and observe that

$$W_{t,i_1}^{(1)}=\sum_{k=1}^{\infty}\sum_{j_1=1}^{d}kB^k(i_1,j_1)f_{t-k,j_1},$$

where $B^k(i_1,j_1)$ denotes element (i_1,j_1) of B^k, and $W_{t,i_1}^{(1)}$ is element i_1 of $W_t^{(1)}$. Also for any $k\geq 0$ and any j_1

$$\begin{aligned}h_{t,i_1} &= \sum_{k=0}^{\infty}\sum_{j_1=1}^{d}B^k(i_1,j_1)f_{t-k,j_1}\\ &\geq \sum_{j_1=1}^{d}B^k(i_1,j_1)f_{t-k,j_1}\\ &\geq \zeta+B^k(i_1,j_1)f_{t-k,j_1},\end{aligned}\tag{B.6}$$

with $\zeta:=\inf_{k,j_1}\{B^k(i_1,j_1)\left[(I_d-A-B)\,\gamma\right]_{j_1}\}>0$, where $\left[(I_d-A-B)\,\gamma\right]_{j_1}$ is element j_1 of $[(I_d-A-B)\,\gamma]$. Hence for any $r_0\geq 1$

$$\begin{aligned}\frac{1}{h_{t,i_1}}W_{t,i_1}^{(1)} &= \sum_{k=1}^{\infty}\sum_{j_1=1}^{d}k\frac{B^k(i_1,j_1)f_{t-k,j_1}}{h_{t,i_1}}\\ &\leq \sum_{k=1}^{\infty}\sum_{j_1=1}^{d}k\frac{B^k(i_1,j_1)f_{t-k,j_1}}{\zeta+B^k(i_1,j_1)f_{t-k,j_1}}\\ &\leq \sum_{k=1}^{\infty}\sum_{j_1=1}^{d}k\left(\frac{B^k(i_1,j_1)f_{t-k,j_1}}{\zeta}\right)^{1/r_0}\\ &\leq c\sum_{k=1}^{\infty}\sum_{j_1=1}^{d}k\phi_{j_1}^k f_{t-k,j_1}^{1/r_0},\end{aligned}\tag{B.7}$$

where the first inequality follows from (B.6), the second follows from the fact that $x/(1+x)\leq x^s$ for all $x\geq 0$ and $s\in[0,1]$. Using that $\rho(B)<1$ which follows by Assumption 3 and Lemma B.8, $\phi_{j_1}\in[0,1)$ is a constant depending on j_1,i_1, and r_0. Considering (B.7), we have that for $j_1 = 1, ..., d$, $\mathbb{E}[\sup_{\theta\in\mathcal{V}(\theta_0)} |f_{t,j_1}|] < \infty$ since X_t has finite second-order moments. Hence for any $i_1 = 1, ..., d$ and any $r_0 \geq 1$, by Minkowski's inequality,

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{1}{h_{t,i_1}}W^{(1)}_{t,i_1}\right|^{r_0}\right] < \infty. \tag{B.8}$$

Next, each element of $W_t^{(2)}$ is bounded because $\rho(B) < 1$, so we conclude, using (B.5) and (B.8), that (B.1) is true for any $i_1 = 1, ..., d$ and $r_0 \geq 1$. Using similar arguments, one can establish that also (B.2) and (B.3) hold. Turning to (B.4), observe that $h_{0t,i_1} \leq c + \sum_{k=0}^{\infty}\sum_{j_1=1}^{d} B_0^k(i_1, j_1) f_{0t-k,j_1}$ where $f_{0t-k} := A_0 X_{t-1-k}^{\odot 2}$, and for any $k \geq 0$ and $j_1 = 1, ..., d$ $h_{t,i_1} \geq \zeta + B^k(i_1, j_1) f_{t-k,j_1}$ as above. The result now follows by arguments similar to the ones in Francq and Zakoïan (2012, p.202). □

Lemma B.2. *Let $\sigma_t^2 := \sigma_t^2(\theta_0)$. Under Assumptions 1-6 there exists a constant $\kappa > 0$ and a function $w(y)$ such that*

$$\forall y \in \mathbb{R}^d \setminus \{0\}, \quad \lim_{u\to\infty} u^{\kappa}\, \mathbb{P}\left(y'\sigma_t^2 > u\right) = w(y),$$

and $w(y) > 0$ for any non-negative $y \neq 0$.
Moreover, if κ is not an even integer the vector $Y_t := (X_t^{\odot 2\prime}, \sigma_t^{2\prime})'$ and all finite dimensional distributions of (Y_t) are multivariate regularly varying with index κ.

Proof. The first part of the lemma follows directly from Fernández and Muriel (2009, Theorem 5 and Remark 7) and is established using Kesten's theorem, see e.g. Basrak et al. (2002b, Theorem 2.4). So we have that any linear combination of σ_t^2 is regularly varying, see Remark 2.2, by observing that a constant is slowly varying. When κ is not an even integer the multivariate regular variation of σ_t^2 follows by Basrak et al. (2002a, Theorem 1.1 (iii)-(iv)). Next, since $Y_t = [\mathrm{diag}(\varepsilon_t^{\odot 2}), O_{d\times d}]'\sigma_t^2$ and using that $\mathbb{E}[\|\varepsilon_t\|^{2u}] < \infty$ for some $u > \kappa$, as in Basrak et al. (2002b, Proof of Corollary 3.5), Y_t is multivariate regularly varying with index κ by the multivariate version of Breiman's lemma, see Basrak et al. (2002b, Proposition A.1).

The regular variation of any finite dimensional distribution of (Y_t) follows by induction, using arguments similar to the ones given in Basrak et al. (2002b, Proof of Corollary 3.5). Define $Y_t(k) := (Y_t', ..., Y_{t+k-1}')'$. Indeed, $Y_t(1) = Y_t$ is regularly varying with index κ. Suppose that $Y_t(k)$ is regularly varying with index κ for some $k \geq 1$. Observe that

$$\begin{pmatrix} Y_t(k) \\ \sigma_{t+k}^2 \end{pmatrix} = \begin{pmatrix} O_{2kd\times 1} \\ (I_d - A_0 - B_0)\gamma_0 \end{pmatrix} + \begin{pmatrix} I_{2kd} \\ \left(O_{d\times(2(k-1)d)}, A_0, B_0\right) \end{pmatrix} Y_t(k),$$

which is regularly varying with index κ by Basrak et al. (2002b, Proposition A.1). This implies, again using Basrak et al. (2002b, Proposition A.1), that

$$Y_t(k+1) = \begin{pmatrix} I_{2kd} & O_{2kd\times d} \\ O_{d\times 2kd} & \mathrm{diag}(\varepsilon_{t+k}^{\odot 2}) \\ O_{d\times 2kd} & I_d \end{pmatrix} \begin{pmatrix} Y_t(k) \\ \sigma_{t+k}^2 \end{pmatrix}$$

is regularly varying with index κ. □

Lemma B.3. *Suppose that the assumptions of Theorem 5.2 are satisfied. Then as* $T \to \infty$

$$Ta_T^{-1}\left(\hat{\gamma}_T - \gamma_0\right) \xrightarrow{w} S,$$

where the sequence (a_T) *is defined in Theorem 5.2, and* S *has a* d*-dimensional* κ*-stable distribution.*

Proof. Following arguments similar to the ones in Pedersen and Rahbek (2014, Proof of Lemma B.8) and using that $\rho(B_0) < 1$, it can be shown that for any $\delta < 1$

$$\hat{\gamma}_T - \gamma_0 = \frac{1}{T}\sum_{t=1}^{T} C_0 \left\{\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) - I_d\right\} \sigma_t^2 + o_p\left(T^{-\delta}\right),$$

where $C_0 := (I_d - A_0 - B_0)^{-1}(I_d - B_0)$ and $\sigma_t^2 := \sigma_t^2(\gamma_0, \lambda_0)$. Choosing $\delta = (\kappa - 1)/\kappa$, and using that $a_T \sim (cT)^{1/\kappa}$, see Basrak et al. (2002b, Remark 2.11), we have that

$$Ta_T^{-1}(\hat{\gamma}_T - \gamma_0) = C_0 a_T^{-1} \sum_{t=1}^{T} \left\{\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) - I_d\right\} \sigma_t^2 + o_p\left(1\right). \tag{B.9}$$

With V_η the mapping defined in Lemma B.5 and N_T and N the point processes defined in Lemma B.4, observe that Lemmata B.4 and B.5 and the continuous mapping theorem imply that as $T \to \infty$,

$$V_\eta\left(N_T(\cdot)\right) \xrightarrow{w} V_\eta\left(N(\cdot)\right). \tag{B.10}$$

Combining (B.10) with Lemmata B.6 and B.7 and using Billingsley (1999, Theorem 3.2) it follows that $a_T^{-1}\sum_{t=1}^{T}\{\operatorname{diag}(\varepsilon_t^{\odot 2}) - I_d\}\sigma_t^2 \xrightarrow{w} \tilde{S}$ as $T \to \infty$ for some random vector $\tilde{S}$ that has a multivariate κ-stable distribution. We conclude that

$$Ta_T^{-1}(\hat{\gamma}_T - \gamma_0) \xrightarrow{w} C_0\tilde{S}.$$

Since $\kappa \in (1, 2)$, we have from Samorodnitsky and Taqqu (1994, Theorems 2.1.2 and 2.5.1(c)) that $S = C_0\tilde{S}$ has a multivariate κ-stable distribution. □

Lemma B.4. *Let* $\sigma_t^2 := \sigma_t^2(\theta_0)$ *and* $Y_t := (X_t^{\odot 2\prime}, \sigma_t^{2\prime})'$*, and, moreover, let* $M_P(\overline{\mathbb{R}}^{2d} \setminus \{0\})$ *denote the collection of point measures on* $\overline{\mathbb{R}}^{2d} \setminus \{0\}$*. Under the assumptions of Theorem 5.2 there exists a sequence* (a_T)*,* $0 < a_T \to \infty$ *as* $T \to \infty$*, such that* $T\,\mathbb{P}(\|Y_t\| > a_T) \to 1$ *as* $T \to \infty$*. Moreover, as* $T \to \infty$

$$N_T(\cdot) := \sum_{t=1}^{T} \delta_{a_T^{-1}Y_t}(\cdot) \xrightarrow{w} N\left(\cdot\right), \tag{B.11}$$

where N *is a point process on* $\overline{\mathbb{R}}^{2d} \setminus \{0\}$ *that can be represented as*

$$N(\cdot) \overset{D}{=} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}(\cdot),$$

consisting of

1. *a Poisson random measure* $\sum_{i=1}^{\infty} \delta_{P_i}(\cdot)$ *on* $(0, \infty)$ *with intensity measure* $\nu\left(\mathrm{d}y\right) = \psi\kappa y^{-\kappa-1}\mathbb{1}\left\{y \in [0,)\right\}\mathrm{d}y$*,* $\psi \in (0, 1]$*,*
2. *an i.i.d. sequence* $\{\sum_{j=1}^{\infty} \delta_{Q_{ij}}(\cdot)\}_{i\in\mathbb{N}}$ *of point processes in* $\tilde{M}_P(\overline{\mathbb{R}}^{2d} \setminus \{0\}) := \{\mu \in M_P(\overline{\mathbb{R}}^{2d} \setminus \{0\}) : \mu\left(\{y : \|y\| > 1\}\right) = 0$ *and* $\mu(\{y : y \in \mathbb{S}^{2d-1}\}) > 0\}$ *independent of* $\sum_{i=1}^{\infty} \delta_{P_i}(\cdot)$*.*

Proof. By Lemma B.2 and Resnick (2007, Theorem 3.6) we have that the sequence (a_T) exists. We establish (B.11) by verifying the conditions of Basrak and Segers (2009, Theorem 4.5) for (Y_t). Any finite-dimensional distribution of (Y_t) is regularly varying with index κ by Lemma B.2. The anti-clustering condition, Basrak and Segers (2009, Condition 4.1), can be shown to hold by arguments similar to the ones given in Basrak et al. (2002b, Proof of Theorem 2.10). First, observe that $Y_t = A_t Y_{t-1} + B_t$ with

$$A_t := \begin{bmatrix} \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) A_0 & \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) B_0 \\ A_0 & B_0 \end{bmatrix} \tag{B.12}$$

and

$$B_t := \begin{bmatrix} \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)(I_d - A_0 - B_0)\gamma_0 \\ (I_d - A_0 - B_0)\gamma_0 \end{bmatrix}.$$

Next, consider the skeleton $Y_t^{(n)} := Y_{tn}$ that satisfies the stochastic recurrence equation

$$Y_t^{(n)} = A_t^{(n)} Y_{t-1}^{(n)} + B_t^{(n)},$$

with $A_t^{(n)} := A_{tn} \cdots A_{(t-1)n+1}$ and $B_t^{(n)} := B_{tn} + \sum_{k=1}^{n-1} A_{tn} \cdots A_{tn-k+1} B_{tn-k}$. We have that $Y_0 = Y_0^{(n)}$, and since n is fixed, it follows that if $Y_t^{(n)}$ satisfies the anti-clustering condition, so does Y_t. Observe that

$$\mathbb{P}\left(\max_{\tau \le t \le r_T} \left\|Y_t^{(n)}\right\| > a_T u \middle| \left\|Y_0^{(n)}\right\| > a_T u\right) \le \sum_{\tau \le t \le r_T} \mathbb{P}\left(\left\|Y_t^{(n)}\right\| > a_T u \middle| \left\|Y_0^{(n)}\right\| > a_T u\right).$$

Moreover,

$$Y_t^{(n)} = \prod_{j=0}^{t-1} A_{t-j}^{(n)} Y_0^{(n)} + \sum_{k=0}^{t-1} \prod_{j=0}^{k-1} A_{t-j}^{(n)} B_{t-k}^{(n)} =: I_{t,1} Y_0^{(n)} + I_{t,2},$$

with $\prod_{j=0}^{-1} A_{t-j}^{(n)} = I_{2d}$ by convention. By Francq and Zakoïan (2010, Theorem 11.6) the strict stationarity of (Y_t) implies that the top Lyapunov exponent of (A_t), with A_t defined in (B.12), is strictly negative. Since ε_t has finite second-order moments, $\mathbb{E}[\log^+ \|A_t^{(n)}\|] < \infty$, it follows by Kingman (1973, Theorem 6) that

$$\lim_{n\to\infty} \frac{1}{n} \log \left\|A_t^{(n)}\right\| < 0 \quad a.s.$$

Hence for n sufficiently large it holds that $\mathbb{E}[\log \|A_t^{(n)}\|] < 0$. Moreover, $\mathbb{E}[\|A_t^{(n)}\|^\delta] < \infty$ for some $\delta > 0$, so in light of Basrak et al. (2002b, Remark 2.9), see also Ling (2007, p.167), it follows that for n sufficiently large, there exists an $\epsilon > 0$ such that

$$\mathbb{E}\left[\left\|A_{t-j}^{(n)}\right\|^\epsilon\right] < 1. \tag{B.13}$$

In particular we choose $\epsilon \in (0,1)$. By the triangle inequality,

$$\begin{aligned} \mathbb{P}\left(\left\|Y_t^{(n)}\right\| > a_T u \middle| \left\|Y_0^{(n)}\right\| > a_T u\right) \le\ & \mathbb{P}\left(\left\|I_{t,1} Y_0^{(n)}\right\| > a_T u/2 \middle| \left\|Y_0^{(n)}\right\| > a_T u\right) \\ & + \mathbb{P}\left(\|I_{t,2}\| > a_T u/2 \middle| \left\|Y_0^{(n)}\right\| > a_T u\right). \end{aligned} \tag{B.14}$$

For the first term in (B.14) it holds that

$$\begin{aligned}\mathbb{P}\left(\left\|I_{t,1}Y_0^{(n)}\right\| > a_T u/2\middle|\left\|Y_0^{(n)}\right\| > a_T u\right) &\leq 2^{\epsilon}\,\mathbb{E}\left(\|I_{t,1}\|^{\epsilon}\left\|Y_0^{(n)}\right\|^{\epsilon}\middle|\left\|Y_0^{(n)}\right\| > a_T u\right)\\ &\leq c\,\mathbb{E}\left[\|I_{t,1}\|^{\epsilon}\right](ua_T)^{-\epsilon}\,\mathbb{E}\left(\left\|Y_0^{(n)}\right\|^{\epsilon}\middle|\left\|Y_0^{(n)}\right\| > a_T u\right)\end{aligned}$$

where the first inequality follows by the conditional generalized Chebyshev inequality and the second inequality follows by the fact that $I_{t,1}$ and $Y_0^{(n)}$ are independent. By an application of Karamata's theorem (specifically, Pan et al. (2013, Proposition 4.1)) we have that

$$\begin{aligned}\limsup_{T\to\infty}\mathbb{P}\left(\left\|I_{t,1}Y_0^{(n)}\right\| > a_T u/2\middle|\left\|Y_0^{(n)}\right\| > a_T u\right) &\leq c\,\mathbb{E}\left[\|I_{t,1}\|^{\epsilon}\right]\\ &\leq c\left(\mathbb{E}\left[\left\|A_t^{(n)}\right\|^{\epsilon}\right]\right)^t\end{aligned}$$

where we have used the definition of $I_{t,1}$ and (B.13). Since $I_{t,2}$ and $Y_0^{(n)}$ are independent for $t \geq 1$, we have that $\mathbb{P}(\|I_{t,2}\| > a_T u/2|\,\|Y_0^{(n)}\| > a_T u) = \mathbb{P}(\|I_{t,2}\| > a_T u/2)$. Moreover, exploiting that $(A_t^{(n)}, B_t^{(n)})$ is i.i.d.,

$$\begin{aligned}\mathbb{P}(\|I_{t,2}\| > a_T u/2) &\leq a_T^{-\epsilon}(2/u)^{\epsilon}\,\mathbb{E}\left[\|I_{t,2}\|^{\epsilon}\right]\\ &= a_T^{-\epsilon}(2/u)^{\epsilon}\,\mathbb{E}\left[\left\|\sum_{k=0}^{t-1}\prod_{j=0}^{k-1}A_{t-j}^{(n)}B_{t-k}^{(n)}\right\|^{\epsilon}\right]\\ &\leq a_T^{-\epsilon}(2/u)^{\epsilon}\sum_{k=0}^{t-1}\mathbb{E}\left[\prod_{j=0}^{k-1}\left\|A_{t-j}^{(n)}\right\|^{\epsilon}\left\|B_{t-k}^{(n)}\right\|^{\epsilon}\right]\\ &= a_T^{-\epsilon}(2/u)^{\epsilon}\,\mathbb{E}\left[\left\|B_{t-k}^{(n)}\right\|^{\epsilon}\right]\sum_{k=0}^{t-1}\left(\mathbb{E}\left[\left\|A_{t-j}^{(n)}\right\|^{\epsilon}\right]\right)^k\\ &\leq a_T^{-\epsilon}(2/u)^{\epsilon}\,\mathbb{E}\left[\left\|B_{t-k}^{(n)}\right\|^{\epsilon}\right]\sum_{k=0}^{\infty}\left(\mathbb{E}\left[\left\|A_{t-j}^{(n)}\right\|^{\epsilon}\right]\right)^k\\ &\leq c a_T^{-\epsilon},\end{aligned}$$

where the first inequality follows by the generalized Chebyshev inequality, and the last inequality follows by (B.13). Now using arguments similar to the ones in Basrak et al. (2002b, p.104), we conclude that

$$\lim_{\tau\to\infty}\limsup_{T\to\infty}\sum_{\tau\leq t\leq r_T}\mathbb{P}\left(\left\|Y_t^{(n)}\right\| > a_T u\middle|\left\|Y_0^{(n)}\right\| > a_T u\right) = 0,$$

i.e. Basrak and Segers (2009, Condition 4.1) is satisfied.

Assumption 2 implies that zero is an interior point of the support of the density of ε_t. It then follows by Boussama (1998, Theorem 5.5.3) that (σ_t^2), and thereby (Y_t), is strongly mixing. This implies that the mixing condition, Basrak and Segers (2009, Condition 4.4), is satisfied as mentioned by Basrak and Segers (2009, p.1070). From Basrak and Segers (2009, Theorem 4.5) we now have that $N_T(\cdot) \xrightarrow{w} N(\cdot)$ as $T \to \infty$, and the characterization of the distribution of $N(\cdot)$ in terms of the point process $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_{P_iQ_{ij}}(\cdot)$ follows by Basrak and Segers (2009, Remark 4.6). □

Lemma B.5. *Let $M_P([0,\infty]^{2d} \setminus \{0\})$ denote the collection of point measures on $[0,\infty]^{2d} \setminus \{0\}$. For any $\eta > 0$ define the mapping $V_\eta : M_P([0,\infty]^{2d} \setminus \{0\}) \to \mathbb{R}^d$,*

$$V_\eta \left(\sum_{t=1}^{\infty} \delta_{y_t}(\cdot) \right) = \begin{bmatrix} \sum_{t=1}^{\infty} (y_{t,1} - y_{t,d+1}) \mathbb{1} \{ y_{t,d+1} > \eta \} \\ \vdots \\ \sum_{t=1}^{\infty} (y_{t,d} - y_{t,2d}) \mathbb{1} \{ y_{t,2d} > \eta \} \end{bmatrix},$$

where $y_{t,i}$ denotes the i-th element of y_t. Under the assumptions of Theorem 5.2, V_η is continuous on a subset of $M_P([0,\infty]^{2d} \setminus \{0\})$ containing the point process $N(\cdot)$, defined in Lemma B.4, with probability one.

Proof. The proof follows by arguments similar to the ones in Vaynman and Beare (2013, Proof of Lemma A.2). Define the sets $B_\eta = \{x \in [0,\infty]^{2d} \setminus \{0\} : \max_{i=d+1,\dots,2d}(x_i) > \eta\}$ and $A_\eta = \{\mu \in M_p([0,\infty]^{2d} \setminus \{0\}) : \mu(\partial B_\eta) = 0\}$. Moreover, consider a sequence (μ_T), $\mu_T \in M_p([0,\infty]^{2d} \setminus \{0\})$, such that $\mu_T \xrightarrow{v} \mu \in A_\eta$ as $T \to \infty$. Since B_η does not contain the origin it is relatively compact, so it follows by Resnick (1987, Proposition 3.13) that for T sufficiently large, we can label the points of μ_T and μ in B_η by $(x_{T,1}, \dots, x_{T,k})$ and $(x_1, \dots, x_k)$, respectively, for some finite k. Moreover, for each $i = 1, \dots, k$

$$x_{T,i} \to x_i \quad \text{as } T \to \infty. \tag{B.15}$$

Hence for T sufficiently large $V_\eta(\mu_T)$ and $V_\eta(\mu)$ do only depend on $(x_{T,1}, \dots, x_{T,k})$ and $(x_1, \dots, x_k)$, respectively. By (B.15), $V_\eta(\mu_T) \to V_\eta(\mu)$ as $T \to \infty$, so V_η is continuous on A_η. The point process N from Lemma B.4 has the representation

$$N(\cdot) \stackrel{D}{=} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}(\cdot).$$

Observe that the event $(N(\cdot) \notin A_\eta)$ can only occur if $(P_i Q_{ij} \in \partial B_\eta$ for some $i, j)$. Hence $\mathbb{P}(N(\cdot) \notin A_\eta) = \mathbb{P}(P_i Q_{ij} \in \partial B_\eta$ for some $i, j) \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathbb{P}(P_i Q_{ij} \in \partial B_\eta)$. The Poisson random measure $\sum_{i=1}^{\infty} \delta_{P_i}(\cdot)$ has intensity measure $\nu(\mathrm{d}y) = \psi \kappa y^{-\kappa-1} \mathbb{1}\{y \in \mathbb{R}_+\} \mathrm{d}y$, $\psi \in (0,1]$, which is absolutely continuous, so P_i must be a continuous random variable. Moreover, P_i is independent of Q_{ij}, so $\mathbb{P}(P_i Q_{ij} \in \partial B_\eta) = 0$, and we conclude that $\mathbb{P}(N(\cdot) \notin A_\eta) = 0$. $\square$

Lemma B.6. *Define*

$$S_T := a_T^{-1} \sum_{t=1}^{T} \left\{ \operatorname{diag}\left(\varepsilon_t^{\odot 2} \right) - I_d \right\} \sigma_t^2,$$

where $\sigma_t^2 := \sigma_t^2(\gamma_0, \lambda_0)$. Under the Assumptions of Theorem 5.2, with V_η the mapping defined in Lemma B.5 and $N_T(\cdot)$ the point process defined in Lemma B.4, for any $\delta > 0$

$$\lim_{\eta \to 0} \limsup_{T \to \infty} \mathbb{P}\left(\| S_T - V_\eta(N_T) \| \geq \delta \right) = 0. \tag{B.16}$$

Proof. First observe that

$$V_\eta(N_T(\cdot)) = \begin{bmatrix} a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,1}^2 - 1 \right) \sigma_{t,1}^2 \mathbb{1} \left\{ \sigma_{t,1}^2 > \eta a_T \right\} \\ \vdots \\ a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,d}^2 - 1 \right) \sigma_{t,d}^2 \mathbb{1} \left\{ \sigma_{t,d}^2 > \eta a_T \right\} \end{bmatrix},$$

where $\varepsilon_{t,i}^2$ and $\sigma_{t,i}^2$ are the i-th elements of $\varepsilon_t^{\odot 2}$ and σ_t^2, respectively. For $i \in \{1,..,d\}$ consider the i-th element of S_T,

$$\begin{aligned} a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 &= a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 \mathbb{1}\left\{\sigma_{t,i}^2 \leq \eta a_T\right\} \\ &\quad + a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 \mathbb{1}\left\{\sigma_{t,i}^2 > \eta a_T\right\}, \end{aligned}$$

with $\eta > 0$. As in Vaynman and Beare (2013, Proof of Lemma A.1), let $\alpha \in (\kappa, \min\{\beta_0, 2\})$, where β_0 is given in Assumption 2. Using that $\{(\varepsilon_{t,i}^2 - 1)\sigma_{t,i}^2, \mathcal{F}_t\}$ is a martingale difference sequence, together with von Bahr and Esseen (1965, Theorem 2), an application of Karamata's theorem, and Resnick (2007, Theorem 3.6),

$$\begin{aligned} \mathbb{E}\left[\left|a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 \mathbb{1}\left\{\sigma_{t,i}^2 \leq \eta a_T\right\}\right|^{\alpha}\right] &\leq c a_T^{-\alpha} T\, \mathbb{E}\left[\left|\varepsilon_{t,i}^2 - 1\right|^{\alpha}\right] \mathbb{E}\left[(\sigma_{t,i}^2)^{\alpha} \mathbb{1}\left\{\sigma_{t,i}^2 \leq \eta a_T\right\}\right] \\ &= c a_T^{-\alpha} T \int_0^{\eta a_T} x^{\alpha}\, \mathbb{P}\left(\sigma_{t,i}^2 \leq dx\right) \\ &\sim c a_T^{-\alpha} T \left(\eta a_T\right)^{\alpha} \mathbb{P}\left(\sigma_{t,i}^2 > \eta a_T\right) \left(\frac{\kappa}{\alpha - \kappa}\right) \\ &= c \eta^{\alpha} T\, \mathbb{P}\left(\sigma_{t,i}^2 > \eta a_T\right) \\ &\overset{T \to \infty}{\to} c \eta^{\alpha - \kappa} \\ &\overset{\eta \to 0}{\to} 0. \end{aligned}$$

So we conclude, using Chebyshev's inequality, that for any $\tilde{\delta} > 0$ and any $i \in \{1, ..., d\}$

$$\lim_{\eta \to 0} \limsup_{T \to \infty} \mathbb{P}\left[\left|a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 \mathbb{1}\left\{\sigma_{t,i}^2 \leq \eta a_T\right\}\right| \geq \tilde{\delta}\right] = 0,$$

and thus, for any $\tilde{\delta} > 0$, and any $i \in \{1, ..., d\}$

$$\lim_{\eta \to 0} \limsup_{T \to \infty} \mathbb{P}\left[\left|a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 - a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 \mathbb{1}\left\{\sigma_{t,i}^2 > \eta a_T\right\}\right| \geq \tilde{\delta}\right] = 0. \tag{B.17}$$

By the triangle and Boole's inequalities

$$\mathbb{P}\left(\|S_T - V_\eta(N_T)\| \geq \tilde{\delta}\right) \leq \sum_{i=1}^{d} \mathbb{P}\left[\left|a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 - a_T^{-1} \sum_{t=1}^{T} \left(\varepsilon_{t,i}^2 - 1\right) \sigma_{t,i}^2 \mathbb{1}\left\{\sigma_{t,i}^2 > \eta a_T\right\}\right| \geq \frac{\tilde{\delta}}{d}\right],$$

so in light of (B.17) we conclude that (B.16) holds. □

Lemma B.7. *With V_η the mapping defined in Lemma B.5 and $N(\cdot)$ the point process defined in Lemma B.4, suppose that the assumptions of Theorem 5.2 are satisfied. Then*

$$V_\eta(N(\cdot)) \overset{w}{\to} \tilde{S} \tag{B.18}$$

as $\eta \to 0$, where $\tilde{S}$ is a d-dimensional random vector with a multivariate κ-stable distribution, $\kappa \in (1, 2)$.

Proof. The proof follows the lines of Davis and Hsing (1995, pp.897-898), see also Vaynman and Beare (2013, Proof of Lemma A.3). Consider the characteristic function of $V_\eta(N(\cdot))$, $\Psi_\eta : \mathbb{R}^d \to \mathbb{C}$. We establish the weak convergence by showing that $\Psi_\eta(t)$ converges pointwise to a function $\Psi(t)$ as $\eta \to 0$, and that this function is continuous at $t = 0$. The weak convergence then follows by Lévy's Continuity Theorem. We establish the pointwise convergence by showing that $\Psi_\eta(t)$ is Cauchy as $\eta \to 0$, i.e. for any $\epsilon > 0$ there exists an $\eta > 0$ such that

$$\sup_{0<a<b\leq\eta} |\Psi_a(t) - \Psi_b(t)| < \epsilon.$$

With $S^{(\eta)} := V_\eta(N(\cdot))$ and $S_j^{(\eta)}$ its j-th element, observe that for any $\delta > 0$

$$\begin{aligned}
|\Psi_b(t) - \Psi_a(t)| &= \left|\mathbb{E}\left[\exp\left(\imath t'S^{(b)}\right) - \mathbb{E}\left[\exp\left(\imath t'S^{(a)}\right)\right]\right]\right| \\
&\leq \mathbb{E}\left[\left|\exp\left(\imath t'S^{(b)}\right) - \exp\left(\imath t'S^{(a)}\right)\right|\right] \\
&= \mathbb{E}\left[\left|\exp\left(\imath t'S^{(b)}\right) - \exp\left(\imath t'S^{(a)}\right)\right| \mathbb{1}\left\{\max_{j\in\{1,\ldots,d\}} \left|S_j^{(b)} - S_j^{(a)}\right| \leq \delta\right\}\right] \\
&\quad + \mathbb{E}\left[\left|\exp\left(\imath t'S^{(b)}\right) - \exp\left(\imath t'S^{(a)}\right)\right| \mathbb{1}\left\{\max_{j\in\{1,\ldots,d\}} \left|S_j^{(b)} - S_j^{(a)}\right| > \delta\right\}\right],
\end{aligned}$$

where we have used Jensen's inequality. Moreover,

$$\begin{aligned}
\left|\exp\left(\imath t'S^{(b)}\right) - \exp\left(\imath t'S^{(a)}\right)\right| &= \left|\cos\left(t'S^{(b)}\right) - \cos\left(t'S^{(a)}\right) + \imath\left[\sin\left(t'S^{(b)}\right) - \sin\left(t'S^{(a)}\right)\right]\right| \\
&= \sqrt{\left[\cos\left(t'S^{(b)}\right) - \cos\left(t'S^{(a)}\right)\right]^2 + \left[\sin\left(t'S^{(b)}\right) - \sin\left(t'S^{(a)}\right)\right]^2} \\
&= \sqrt{2 - 2\cos\left(t'S^{(b)}\right)\left(t'S^{(a)}\right) - 2\sin\left(t'S^{(b)}\right)\sin\left(t'S^{(a)}\right)} \\
&= \sqrt{2 - 2\cos\left(t'\left(S^{(b)} - S^{(a)}\right)\right)},
\end{aligned}$$

so we have that

$$\begin{aligned}
& |\Psi_b(t) - \Psi_a(t)| \\
\leq\ & \mathbb{E}\left[\sqrt{2 - 2\cos\left(t'\left(S^{(b)} - S^{(a)}\right)\right)}\mathbb{1}\left\{\max_{j\in\{1,\ldots,d\}} \left|S_j^{(b)} - S_j^{(a)}\right| \leq \delta\right\}\right] \qquad \text{(B.19)} \\
& + \mathbb{E}\left[\sqrt{2 - 2\cos\left(t'\left(S^{(b)} - S^{(a)}\right)\right)}\mathbb{1}\left\{\max_{j\in\{1,\ldots,d\}} \left|S_j^{(b)} - S_j^{(a)}\right| > \delta\right\}\right].
\end{aligned}$$

Since t is fixed, $\max_{j\in\{1,\ldots,d\}} |S_j^{(b)} - S_j^{(a)}| \leq \delta$, $t'(S^{(b)} - S^{(a)}) \to 0$ as $\delta \to 0$. Moreover, since $\sqrt{2 - 2\cos(x)} \to 0$ as $x \to 0$, we conclude that for any $\epsilon > 0$, choosing $\delta > 0$ small enough, we have that that $\sqrt{2 - 2\cos(t'(S^{(b)} - S^{(a)}))} < \epsilon/2$ when $\max_{j\in\{1,\ldots,d\}} |S_j^{(b)} - S_j^{(a)}| \leq \delta$. Thereby the first term of the right-hand side of (B.19) is less than $\epsilon/2$ for small enough δ. Next, we fix such δ, and we show that the second term of the right-hand side of (B.19) is less than $\epsilon/2$ for small enough $\eta > 0$ with $\eta \geq b > a > 0$. Since $\sqrt{2 - 2\cos(t'(S^{(b)} - S^{(a)}))} \in [0, 2]$, we have that

$$\mathbb{E}\left[\sqrt{2 - 2\cos\left(t'\left(S^{(b)} - S^{(a)}\right)\right)}\mathbb{1}\left\{\max_{j\in\{1,\ldots,d\}} \left|S_j^{(b)} - S_j^{(a)}\right| > \delta\right\}\right] \leq 2\,\mathbb{P}\left(\max_{j\in\{1,\ldots,d\}} \left|S_j^{(b)} - S_j^{(a)}\right| > \delta\right),$$

so we just have to find an $\eta > 0$ such that

$$\mathbb{P}\left(\max_{j\in\{1,\ldots,d\}}\left|S_j^{(b)}-S_j^{(a)}\right|>\delta\right)<\epsilon/4.$$

We define $\tilde{V}_{a,b} := \max_{j\in\{1,\ldots,d\}}|V_{b,j}-V_{a,j}|$, where $V_{\eta,j}$ denotes the the j-th element of V_η. According to Lemma B.5, V_η is continuous on a subset of $M_P([0,\infty]^{2d}\setminus\{0\})$ containing the point process $N(\cdot)$, defined in Lemma B.4, with probability one. The same must then hold for $\tilde{V}_{a,b}$. Hence $\tilde{V}_{a,b}(N_T)\overset{w}{\to}\tilde{V}_{a,b}(N)$ as $T\to\infty$, and we have that

$$\mathbb{P}\left(\max_{j\in\{1,\ldots,d\}}\left|S_j^{(b)}-S_j^{(a)}\right|>\delta\right)=\mathbb{P}\left(\tilde{V}_{a,b}(N)>\delta\right)=\lim_{T\to\infty}\mathbb{P}\left(\tilde{V}_{a,b}(N_T)>\delta\right). \tag{B.20}$$

Let $S_{T,j}^{(\eta)}$ denote the j-th element of $V_\eta(N_T(\cdot))$ and let $S_{T,j}$ denote the j-th element of S_T defined in Lemma B.6. Then

$$\begin{aligned}\tilde{V}_{a,b}(N_T(\cdot)) &= \max_{j\in\{1,\ldots,d\}}\left|S_{T,j}^{(b)}-S_{T,j}^{(a)}\right|\\ &\le \sum_{j=1}^{d}\left|S_{T,j}^{(b)}-S_{T,j}^{(a)}\right|\\ &\le \sum_{j=1}^{d}\left(\left|S_{T,j}-S_{T,j}^{(b)}\right|+\left|S_{T,j}-S_{T,j}^{(a)}\right|\right).\end{aligned} \tag{B.21}$$

In light of (B.20), (B.21), and Lemma B.6, choosing $\eta > 0$ small enough, we have that

$$\begin{aligned}\sup_{0<a<b\le\eta}\mathbb{P}\left(\max_{j\in\{1,\ldots,d\}}|S_{b,j}-S_{a,j}|>\delta\right) &= \sup_{0<a<b\le\eta}\lim_{T\to\infty}\mathbb{P}\left(\tilde{V}_{a,b}(N_T(\cdot))>\delta\right)\\ &\le \sup_{0<a<b\le\eta}\lim_{T\to\infty}\mathbb{P}\left[\sum_{j=1}^{d}\left(|S_{j,T}-S_{b,j,T}|+|S_{j,T}-S_{a,j,T}|\right)>\delta\right]\\ &< \epsilon/4.\end{aligned}$$

By arguments similar to the ones above, one can show that $\Psi_\eta(t)$ is uniformly Cauchy on a set, $\mathcal{A}$, containing the origin, i.e. for any $\epsilon > 0$ there exists an $\eta > 0$ such that

$$\sup_{0<a<b\le\eta}\sup_{t\in\mathcal{A}}|\Psi_a(t)-\Psi_b(t)|<\epsilon.$$

(In particular we can choose $\mathcal{A} := \{t\in\mathbb{R}^d : \max_{i\in\{1,..,d\}}|t_i|\le 1\}$ as in Vaynman and Beare (2013, Proof of Lemma A.3).) This implies that $\sup_{t\in\mathcal{A}}|\Psi_\eta(t)-\Psi(t)|\to 0$ as $\eta\to 0$, i.e. $\Psi_\eta(t)$ converges uniformly to $\Psi(t)$ on $\mathcal{A}$. Because $\Psi_\eta(t)$ is continuous on $\mathcal{A}$ we have that, using Rudin (1976, Theorem 7.12), that $\Psi(t)$ is continuous on $\mathcal{A}$, and in particular at $t = 0$. We conclude that as $\eta\to 0$, (B.18) holds for some d-dimensional random vector $\tilde{S}$ with characteristic function Ψ. As in Davis and Mikosch (1998, Proof of Proposition 3.3), one can show that the variable $\tilde{S}$ has a multivariate stable distribution with index $\kappa\in(1,2)$ by showing that every linear combination has a stable distribution (see Samorodnitsky and Taqqu (1994, Theorem 2.1.5)), arguing in line with Davis and Hsing (1995, p.898). □

Lemma B.8. *For any matrices A and B with non-negative entries, it holds that $\rho(A+B)<1\Rightarrow\rho(B)<1$.*

Proof. Suppose that $\rho(A+B)<1$, then for $\omega \in (0,\infty)^d$ there exists a $h \in (0,\infty)^d$ that solves the equation

$$h = \omega + Ah + Bh. \tag{B.22}$$

Setting $\tilde{\omega} := \omega + Ah \in (0,\infty)^d$, observe that h solves $h = \tilde{\omega} + Bh$, which is the case if and only if $\rho(B) < 1$. □

Appendix C. A Necessary and Sufficient Condition for Finite 4th-order Moments of a CCC-GARCH Process

Theorem C.1. *Let (X_t) be the strictly and second-order stationary solution to the CCC-GARCH model in (4.1)-(4.4), and define $\varepsilon_t := R^{1/2}Z_t$. Suppose that*

$$\mathbb{E}\left[\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) \otimes \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)\right]$$

exists and is finite. The fourth-order moment matrix of X_t, $\mathbb{E}[X_t^{\odot 2}(X_t^{\odot 2})']$, exists and is finite if and only if

$$\rho\left(\mathbb{E}\left\{\left[A\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)+B\right]\otimes\left[A\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)+B\right]\right\}\right) < 1. \tag{C.1}$$

Proof. Recall the following rules from matrix calculus. Lütkepohl (1996, 2.4.(3)): For $(m\times n)$-dimensional matrices A and B and $(p\times q)$-dimensional matrices C and D, it holds that

$$(A+B)\otimes(C+D) = A\otimes C + A\otimes D + B\otimes C + B\otimes D. \tag{C.2}$$

Lütkepohl (1996, 2.4.(4)): Let the matrices A, B, C, and D be respectively of dimension $(m\times n)$, $(p\times q)$, $(n\times r)$, and $(q\times s)$, then

$$(AC)\otimes(BD) = (A\otimes B)(C\otimes D). \tag{C.3}$$

Lütkepohl (1996, 7.2.(7)): Let the matrices A, B, and C have respectively dimension $(m\times n)$, $(n\times r)$, and $(r\times s)$, then

$$\operatorname{vec}(ABC) = (C'\otimes A)\operatorname{vec}(B). \tag{C.4}$$

First we show that $\mathbb{E}\left[\sigma_t^2\otimes\sigma_t^2\right]$ exists and is finite if and only if (C.1) holds, where $\sigma_t^2 := \sigma_t^2(\gamma_0,\lambda_0)$. Observe that $X_t^{\odot 2} = \operatorname{diag}(\varepsilon_t^{\odot 2})\sigma_t^2$, so with $\omega_0 := (I_d - A_0 - B_0)\gamma_0$ we have that

$$\begin{aligned}
\mathbb{E}\left[\sigma_t^2\otimes\sigma_t^2\right] &= \mathbb{E}\left[\left\{\omega_0 + \left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\otimes\left\{\omega_0 + \left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\right] \\
&= (\omega_0\otimes\omega_0) + \mathbb{E}\left[\omega_0\otimes\left\{\left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\right] \\
&\quad + \mathbb{E}\left[\left\{\left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\otimes\omega_0\right] \\
&\quad + \mathbb{E}\left[\left\{\left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\otimes\left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right],
\end{aligned}$$

where we have used (C.2). Let

$$\begin{aligned}
C_\sigma &:= (\omega_0\otimes\omega_0) + \mathbb{E}\left[\omega_0\otimes\left\{\left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\right] \\
&\quad + \mathbb{E}\left[\left\{\left(A_0\operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right)+B_0\right)\sigma_{t-1}^2\right\}\otimes\omega_0\right]
\end{aligned}$$

which exists and is finite since X_t is second-order stationary. Thus, using (C.3) and that ε_{t-1} and σ_{t-1}^2 are independent,

$$\mathbb{E}\left[\sigma_t^2 \otimes \sigma_t^2\right] = C_\sigma + \mathbb{E}\left[\left\{\left(A_0 \operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right) + B_0\right)\right\} \otimes \left(A_0 \operatorname{diag}\left(\varepsilon_{t-1}^{\odot 2}\right) + B_0\right)\right] \mathbb{E}\left[\sigma_{t-1}^2 \otimes \sigma_{t-1}^2\right].$$

Recursions give that

$$\begin{aligned}\mathbb{E}\left[\sigma_t^2 \otimes \sigma_t^2\right] &= \sum_{i=0}^{\tau-1}\left(\mathbb{E}\left[\left\{\left(A_0 \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) + B_0\right)\right\} \otimes \left(A_0 \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) + B_0\right)\right]\right)^i C_\sigma \\ &\quad + \left(\mathbb{E}\left[\left\{\left(A_0 \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) + B_0\right)\right\} \otimes \left(A_0 \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) + B_0\right)\right]\right)^\tau \mathbb{E}\left[\sigma_{t-\tau}^2 \otimes \sigma_{t-\tau}^2\right],\end{aligned}$$

and we conclude that $\mathbb{E}\left[\sigma_t^2 \otimes \sigma_t^2\right]$ converges as $\tau \to \infty$ if and only if (C.1) holds, see Lütkepohl (1996, Results 9.3.5(a) and 9.3.2). Observe that

$$\begin{aligned}\mathbb{E}\left[\operatorname{vec}\left\{X_t^{\odot 2}\left(X_t^{\odot 2}\right)'\right\}\right] &= \mathbb{E}\left[\operatorname{vec}\left\{\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)\sigma_t^2(\sigma_t^2)' \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)'\right\}\right] \\ &= \mathbb{E}\left[\operatorname{diag}\left(\varepsilon_t^{\odot 2}\right) \otimes \operatorname{diag}\left(\varepsilon_t^{\odot 2}\right)\right] \mathbb{E}\left[\sigma_t^2 \otimes \sigma_t^2\right] \operatorname{vec}\left(I_d\right),\end{aligned}$$

where the second equality follows by (C.4) and the independence between ε_t and σ_t^2. Observe that $\mathbb{E}[\operatorname{diag}(\varepsilon_t^{\odot 2}) \otimes \operatorname{diag}(\varepsilon_t^{\odot 2})]$ is finite by assumption, so $\mathbb{E}[\operatorname{vec}\{X_t^{\odot 2}(X_t^{\odot 2})'\}]$ exists and is finite if and only if $\mathbb{E}\left[\sigma_t^2 \otimes \sigma_t^2\right]$ exists and is finite. □

Appendix D. Vague convergence and point processes

This section contains brief introductions to the notions of vague convergence and point processes, based on Resnick (1987, Ch.3). In the following, let $\mathbb{F}$ be a subset of $\overline{\mathbb{R}}^d$ (following Resnick (1987, p.123), we only need that $\mathbb{F}$ is a locally compact second countable Hausdorff space, i.e. that every $x \in \mathbb{F}$ has a compact neighborhood, there exists open $(G_n)_{n\geq 1}$ such that any open G can be written as $G = \cup_{\alpha \in I} G_\alpha$ for a finite and countable index set I, and that distinct points in $\mathbb{F}$ may be separated by disjoint neighborhoods). Let $\mathcal{B}(\mathbb{F})$ be the Borel σ-field generated by the open sets of $\mathbb{F}$.

D.1. **Vague convergence.** A measure μ is called Radon (or locally finite) if $\mu(K) < \infty$ for all subsets K of $\mathbb{F}$ that are relatively compact, i.e. the closure of K is compact. Next, define the sets

$$C_K^+(\mathbb{F}) := \{f : \mathbb{F} \to [0, \infty) : f \text{ is continuous with compact support}\},$$

and

$$M_+(\mathbb{F}) := \{\mu : \mu \text{ is nonnegative on } \mathcal{B}(\mathbb{F}) \text{ and } \mu \text{ is Radon}\}.$$

A topology on $M_+(\mathbb{F})$ can be obtained by letting its subbasis consist of sets of the form

$$\{\mu \in M_+(\mathbb{F}) : s < \mu(f) < t\}$$

for $f \in C_K^+(\mathbb{F})$ and $0 \leq s \leq t$, where $\mu(f) := \int_{\mathbb{F}} f(x)\,\mu(dx)$. This topology is called the vague topology. If $\mu_n, \mu \in M_+(\mathbb{F})$ for all $n \geq 1$, then μ_n converges vaguely (converges in the vague topology) to μ, written $\mu_n \xrightarrow{v} \mu$, if and only if for all $f \in C_K^+(\mathbb{F})$,

$$\mu_n(f) \to \mu(f) \text{ as } n \to \infty. \tag{D.1}$$

Remark D.1. It holds that $\mu_n \xrightarrow{v} \mu$ if and only if $\mu_n(F) \to \mu(F)$ for all relatively compact $F \in \mathcal{B}(\mathbb{F})$ satisfying $\mu(\partial F) = 0$.
For a detailed treatment of this topic we refer to Resnick (1987, pp.139-149) and Kallenberg (1983, pp.168-171).

D.2. **Point processes.** Let $(x_i)_{i\geq 1}$ be a countable collection of points of $\mathbb{F}$. A point measure on $\mathbb{F}$ is a measure μ defined as $\mu(\cdot) := \sum_{i=1}^{\infty} \delta_{x_i}(\cdot)$ and μ is Radon. Next, define the set

$$M_p(\mathbb{F}) := \{\mu : \mu \text{ is a point measure on } \mathbb{F}\},$$

and define the σ-field $\mathcal{M}_p(\mathbb{F})$ of $M_p(\mathbb{F})$ to be the smallest σ-field containing all sets of the form $\{\mu \in M_p(\mathbb{F}) : \mu(F) \in B\}$ for $F \in \mathcal{B}(\mathbb{F})$ and $B \in \mathcal{B}([0,\infty])$. For a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a point process on $\mathbb{F}$ is a measurable map $N : (\Omega, \mathcal{A}, \mathbb{P}) \to (M_p(\mathbb{F}), \mathcal{M}_p(\mathbb{F}))$, i.e. a random element of $M_p(\mathbb{F})$.

Remark D.2. For a fixed $\omega \in \Omega$, $N(\omega, \cdot)$ is a point measure, and for $F \in \mathcal{B}(\mathbb{F})$ $N(\omega, F)$ is the number of points in F for the realization ω.

A special type of point processes, that is of particular interest in this paper, is the Poisson random measure defined as follows. Let μ be a Radon measure on $\mathcal{B}(\mathbb{F})$. A point process N is a Poisson random measure with mean (or intensity) measure μ if N satisfies that

(1) for any $F \in \mathcal{B}(\mathbb{F})$, and any non-negative integer k

$$\mathbb{P}[N(F) = k] = \begin{cases} \exp\{-\mu(F)\}\{\mu(F)\}^k/k! & \text{if } \mu(F) < \infty \\ 0 & \text{if } \mu(F) = \infty, \end{cases}$$

and

(2) for any $k \geq 1$, if $F_1, ..., F_k$ are mutually disjoint sets in $\mathcal{B}(\mathbb{F})$, then $\{N(F_i)\}_{i=1,\dots,k}$ are independent random variables.

References

von Bahr, B. and C.-G. Esseen (1965): "Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leq r \leq 2$," *The Annals of Mathematical Statistics*, 36, 299–303.

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.

Billingsley, P. (1995): *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, New York, NY: John Wiley & Sons, 3 ed.

——— (1999): *Convergence of Probability Measures*, Hoboken,NJ: John Wiley & Sons, 2 ed.

Bingham, N., C. Goldie, and J. Teugels (1987): *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge: Cambridge University Press.

Bollerslev, T. (1990): "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model," *The Review of Economics and Statistics*, 72, 498–505.

Boussama, F. (1998): "Ergodité, mélange et estimation dans les modèles GARCH," Ph.D. thesis, Université Paris 7.

Caporin, M. and M. McAleer (2012): "Do We Really Need Both BEKK and DCC? A Tale of Two Multivariate GARCH Models," *Journal of Economic Surveys*, 26, 736–751.

Comte, F. and O. Lieberman (2003): "Asymptotic Theory For Multivariate GARCH Processes," *Journal of Multivariate Analysis*, 84, 61–84.

Davis, R. A. and T. Hsing (1995): "Point Process and Partial Sum Convergence for Weakly Dependent Random Variables with Infinite Variance," *The Annals of Probability*, 23, 879–917.

Davis, R. A. and T. Mikosch (1998): "The Sample Autocorrelations of Heavy-Tailed Processes with Applications to ARCH," *The Annals of Statistics*, 26, 2049–2080.

Engle, R. and J. Mezrich (1996): "GARCH for Groups," *Risk Magazine*, 9, 36–40.

Fernández, B. and N. Muriel (2009): "Regular variation and related results for the multivariate GARCH(p,q) model with constant conditional correlations," *Journal of Multivariate Analysis*, 100, 1538–1550.

Francq, C., L. Horváth, and J.-M. Zakoïan (2011): "Merits and Drawbacks of Variance Targeting in GARCH Models," *Journal of Financial Econometrics*, 9, 619–656.

——— (2013): "Variance Targeting Estimation of Multivariate GARCH Models," CREST, Personal communication.

Francq, C. and J.-M. Zakoïan (2010): *GARCH Models: Structure, Statistical Inference and Financial Applications*, Chichester: Wiley.

——— (2012): "QML estimation of a class of multivariate asymmetric GARCH models," *Econometric Theory*, 28, 179–206.

Hafner, C. M. and A. Preminger (2009a): "Asymptotic Theory for a Factor GARCH Model," *Econometric Theory*, 25, 336–363.

——— (2009b): "On asymptotic theory for multivariate GARCH models," *Journal of Multivariate Analysis*, 100, 2044 – 2054.

Hafner, C. M. and O. Reznikova (2012): "On the estimation of dynamic conditional correlation models," *Computational Statistics & Data Analysis*, 56, 3533 – 3545.

Hall, P. and Q. Yao (2003): "Inference in ARCH and GARCH Models with Heavy-Tailed Errors," *Econometrica*, 71, 285–317.

He, C. and T. Teräsvirta (2004): "An Extended Constant Conditional Correlation GARCH Model and Its Fourth-Moment Structure," *Econometric Theory*, 20, 904–926.

Hill, J. B. and E. Renault (2012): "Variance targeting for heavy tailed time series," Unpublished manuscript, University of North Carolina at Chapel Hill.

Jeantheau, T. (1998): "Strong consistency of estimators for multivariate ARCH models," *Econometric Theory*, 14, 70–86.

Jensen, S. T. and A. Rahbek (2004): "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

Kallenberg, O. (1983): *Random measures*, Akademie-Verlag Berlin, 3 ed.

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

Kingman, J. F. C. (1973): "Subadditive Ergodic Theory," *The Annals of Probability*, 1, 883–909.

Kristensen, D. and O. Linton (2004): "Consistent Standard Errors for Target Variance Approach to GARCH Estimation," *Econometric Theory*, 20, 990–993.

Laurent, S., J. V. K. Rombouts, and F. Violante (2012): "On the forecasting accuracy of multivariate GARCH models," *Journal of Applied Econometrics*, 27, 934–955.

Ling, S. (2007): "A Double AR(p) Model: Structure and Estimation," *Statistica Sinica*, 17, 161–175.

Ling, S. and M. McAleer (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

Lütkepohl, H. (1996): *Handbook of Matrices*, John Wiley & Sons.

Mikosch, T. (2004): "Modeling Dependence and Tails of Financial Time Series," in *Extreme Values in Finance, Telecommunications, and the Environment*, ed. by B. Finkenstädt and H. Rootzén, Chapman & Hall/CRC, vol. 99 of *Monographs on Statistics and Applied Probability*, chap. 5, 185–286.

Mikosch, T. and C. Stărică (2000): "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1, 1) Process," *The Annals of Statistics*, 28, 1427–1451.

Mikosch, T. and D. Straumann (2006): "Stable Limits of Martingale Transforms with Application to the Estimation of Garch Parameters," *The Annals of Statistics*, 34, 493–522.

Pan, X., X. Leng, and T. Hu (2013): "The second-order version of Karamata's theorem with applications," *Statistics & Probability Letters*, 83, 1397–1403.

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

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*, Springer Series in Statistics, New York, NY: Springer.

Resnick, S. I. (1986): "Point Processes, Regular Variation and Weak Convergence," *Advances in Applied Probability*, 18, 66–138.

——— (1987): *Extreme Values, Regular Variation, and Point Processes*, Applied Probability, New York, NY: Springer-Verlag.

——— (2004): "On the Foundations of Multivariate Heavy-Tail Analysis," *Journal of Applied Probability*, 41, 191–212.

——— (2007): *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer Series in Operations Research, Springer.

Rudin, W. (1976): *Principles of Mathematical Analysis*, International series in pure and applied mathematics, Singapore: McGraw-Hill, 3 ed.

Samorodnitsky, G. and M. Taqqu (1994): *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Stochastic modeling, New York, NY: Chapman & Hall.

Sherman, M. and E. Carlstein (2004): “Confidence intervals based on estimators with unknown rates of convergence,” *Computational Statistics & Data Analysis*, 46, 123–139.

Silvennoinen, A. and T. Teräsvirta (2009): “Multivariate GARCH Models,” in *Handbook of Financial Time Series*, ed. by T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen, Springer Berlin Heidelberg, 201–229.

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

Vaynman, I. and B. K. Beare (2013): “Stable limit theory for the variance targeting estimator,” Working paper, University of California, San Diego, http://econweb.ucsd.edu/ bbeare/pdfs/stablevte.pdf.