Department of Economics
FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Rasmus Søndergaard Pedersen

Inference and Testing in Multivariate GARCH Models

Academic advisor: Anders Rahbek

Date of submission: August 31, 2015

Acknowledgements

First and foremost I am deeply grateful to my thesis supervisor, Anders Rahbek. Since my final year as an undergraduate student, Anders has been an extremely supportive teacher, mentor, and supervisor. Needless to say, I am very much looking forward to conducting research with Anders in the future.

The Department of Economics at University of Copenhagen has provided me with a great environment for conducting research. In particular, the academic discussions and the lunch and coffee breaks with my colleagues have made life as a PhD student much more exciting.

In addition to my colleagues at the Department of Economics, I would like to thank, in particular, Dennis Kristensen, Thomas Mikosch, and Paolo Zaffaroni for some very helpful conversations related to my research.

I would also like to thank Paolo Zaffaroni and his staff for helping me arranging my research stay at Imperial College Business School (ICBS) in Spring 2014. I enjoyed being a visitor at ICBS, and the faculty and doctoral students were kind and welcoming during my stay. Financial support from the Niels Bohr Foundation during my visit was highly appreciated.

Lastly, I would like to thank Charlotte Ginsberg, my family, and my friends for their support for the last three years.

Summary

Most financial applications are, by nature, multivariate with estimates and forecasts of conditional covariance matrices as important components as in, for example, the rich asset pricing, portfolio choice, and value-at-risk literature. One way of obtaining such estimates and forecasts is by estimation of multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models - a class of models that, by now, is heavily used within the fields of financial econometrics and empirical finance. This thesis contains three self-contained parts (chapters) on estimation of and large-sample inference in multivariate GARCH models.

In the first chapter, "*Multivariate variance targeting in the BEKK-GARCH model*", we consider asymptotic inference in the multivariate BEKK-GARCH model based on (co)variance targeting (VT). By definition the VT estimator is a two-step estimator and the theory presented is based on expansions of the modified likelihood function, or estimating function, corresponding to these two steps. Strong consistency is established under weak moment conditions, while sixth-order moment restrictions are imposed to establish asymptotic normality. Included simulations indicate that the multivariately induced higher-order moment constraints are necessary.

Existing literature on VT estimation of multivariate GARCH models, including the first chapter of this thesis, relies on at least finite fourth-order moments of the data generating process in order to derive the large-sample distribution of the variance targeting estimator. Such moment conditions may not be a realistic assumption as financial return distributions are typically found to be heavy tailed. In the second chapter, "*Targeting estimation of CCC-GARCH models with infinite fourth moments*" , we consider the large-sample properties of the VT estimator for the multivariate extended constant conditional correlation (ECCC-)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, under suitable conditions, its limiting distribution is multivariate stable (different from a Gaussian distribution). The rate of consistency of the estimator is slower than $\sqrt{T}$ and depends on the tail shape of the data generating process. A simulation study illustrates the derived properties of the VT estimator.

Lastly, in the third chapter, "*Inference and testing on the boundary in extended constant conditional correlation GARCH models*", we consider testing for volatility spillovers (or interactions) in ECCC-GARCH models. The proposed tests imply that the parameter vector under the null hypothesis lies on the boundary of the maintained hypothesis, which leads to non-standard limiting distributions of the test statistics. The large-sample properties of the quasi-maximum likelihood estimator are derived together with limiting distributions of the related Lagrange multiplier, Wald, and quasi-likelihood ratio statistics. A simulation study investigates the size and power properties of the tests. As an empirical illustration, the proposed tests are applied to test for volatility spillovers between returns on foreign exchange rates.

Summary in Danish

De fleste anvendelser indenfor finansiering er af flerdimensionel natur med estimation og forecasts af betingede kovariansmatricer som vigtige komponenter, jf. felter som asset pricing, porteføljevalg, og visse områder indenfor risikoanalyse (såsom *value-at-risk*). En metode hvorpå sådanne estimater og forecasts kan opnås er via estimation af multivariate *generalized autoregressive conditional heteroskedasticity* (GARCH) modeller - en modelklasse der er hyppigt anvendt indenfor finansiel økonometri og empirisk finansiering. Denne afhandling indeholder tre selvstændige kapitler om estimation af og asymptotisk inferens i multivariate GARCH modeller.

I det første kapitel "*Multivariate variance targeting in the BEKK-GARCH model*", betragter vi asymptotisk inferens i multivariate BEKK-GARCH modeller baseret på såkaldt *variance targeting* (VT-)estimation. VT-estimatoren er pr. definition en to-trins-estimator, og vores teoretiske analyse er baseret på udviklinger af en modificeret estimationsfunktion, der svarer til disse to estimationstrin. Vi beviser, at estimatoren er stærk konsistent under milde momentantagelser, mens endelige 6.-ordens momenter er antaget for at bevise asymptotisk normalitet. Et simulationsstudie indikerer, at højere-ordens momentbetingelserne, der fremkommer via modellens flerdimensionelle natur, er nødvendige.

I den eksisterende litteratur indenfor VT-estimation af multivariate GARCH modeller, herunder det første kapitel i denne afhandling, antages det, at den datagenererende proces har mindst endelige 4.-ordens momenter, for at den asymptotiske fordeling af VT-estimatoren kan udledes. Sådanne momentbetingelser er ikke nødvendigvis en realistisk antagelse, eftersom afkastfordelinger typisk har tunge haler. I det andet kapitel, "*Targeting estimation of CCC-GARCH models with infinite fourth moments*", udleder vi de asymptotiske egenskaber for VT-estimatoren for den multivariate *extended constant conditional correlation* (ECCC-)GARCH model, i tilfældet hvor den datagenererende proces har uendelige 4.-ordens momenter. Ved brug af ikke-standard asymptotisk teori udleder vi nye resultater for estimatoren og viser, at estimatoren, under passende betingelser, har en asymptotisk multivariat stabil fordeling (forskellig fra en multivariat normalfordeling). Konsistensraten for estimatoren er langsommere end $\sqrt{T}$ og afhænger af halefordelingen for den datagenererende proces. De asymptotiske egenskaber for estimatoren illustreres i et simulationsstudie.

I det tredje kapitel, "*Inference and testing on the boundary in extended constant conditional correlation GARCH models*", betragter vi tests for volatilitets-*spillovers* (eller interaktioner) i ECCC-GARCH modeller. Vi udleder de asymptotiske egensk-

aber for quasi-maximum likelihood estimatoren samt for de tilhørende Lagrange multiplikator, Wald og quasi-likelihood ratio statistikker. De foreslåede tests medfører, at parameterværdien under nulhypotesen ligger på randen af parameterområdet, hvilket leder til ikke-konventionelle asymptotiske fordelinger af teststatistikkerne. I et simulationsstudie undersøges testenes empiriske egenskaber nærmere. Som en empirisk illustration anvender vi de foreslåede tests til at teste for volatilitets-*spillovers* mellem afkast på valutakurser.

Contents

Part I

Multivariate variance targeting in the BEKK-GARCH model

This chapter is joint research with Anders Rahbek (University of Copenhagen) and has been published in *The Econometrics Journal*, vol. 17, issue 1, February 2014.[1]

Abstract

In this paper we consider asymptotic inference in the multivariate BEKK model based on (co)variance targeting (VT). By definition the VT estimator is a two-step estimator and the theory presented is based on expansions of the modified likelihood function, or estimating function, corresponding to these two steps. Strong consistency is established under weak moment conditions, while sixth-order moment restrictions are imposed to establish asymptotic normality. Included simulations indicate that the multivariately induced higher-order moment constraints are necessary.

1 Introduction

As argued in Laurent et al. (2012) variance targeting (VT) estimation, or simply VT, is highly applicable when forecasting conditional covariance matrices. This paper derives large-sample properties of the variance targeting estimator (VTE) for the multivariate BEKK-GARCH model, establishing that asymptotic inference is feasible in the model when estimated by VT. To our knowledge, large-sample properties of the VTE have not been considered before for multivariate GARCH models, unlike for the univariate GARCH model where the properties have recently been considered by Francq et al. (2011), see also Kristensen and Linton (2004). We find that the VTE is strongly consistent if the observed process has finite second-order moments, and asymptotic normality applies if the observed process has finite sixth-order moments. These moment restrictions for large-sample inference in the

[1]We thank the editor Pierre Perron and three anonymous referees for valuable comments which have implied a much improved paper. Also we thank Paolo Paruolo, Andreas N. Jensen and Dennis Kristensen for comments. Finally, we thank for discussions at the 2013 Humboldt-Copenhagen (HUKU) and the Brunel University (MMF) conferences, and at the European Meeting of the Econometric Society 2013.

BEKK-GARCH model, when estimated by VT estimation, are in line with existing literature for large-sample inference based on quasi maximum likelihood (QML) estimation, see Hafner and Preminger (2009b). Included simulations indicate that our imposed sixth-order moment requirement (which are identical to Hafner and Preminger (2009b)) may not be relaxed for VT estimation. Thus our results point at that while VT estimation is simpler and even possible to implement for higher order systems, it requires no further moments when comparing to existing results for QML based estimation.

Most financial applications are by nature multivariate with forecasts of conditional covariance matrices as important components as in for example the rich portfolio choice and Value-at-Risk literature. Such forecasts may be based on estimation of multivariate conditionally heteroscedastic (GARCH) models such as the much applied BEKK model proposed by Engle and Kroner (1995), see e.g. Bauwens et al. (2006) and Laurent et al. (2012). However, a drawback of the BEKK model, despite the fact that it is a very simple extension of the popular univariate GARCH model in Bollerslev (1987), is that it contains a large number of parameters even for moderate dimensions. This implies that it is difficult, if not impossible, to estimate the model through classical QML estimation even for moderately sized series. At the same time, recent development in financial applications implies an increasing interest in conditional covariances and correlations based on vast, or high-dimensional models. To address this issue one may restrict, or simplify, further the BEKK model to reduce the number of parameters as is the case in for example diagonal-BEKK and scalar-BEKK models, see Bauwens et al. (2006). Alternatively – or additionally – one may consider a simplified estimation method such as VT estimation considered here.

VT estimation was originally proposed by Engle and Mezrich (1996) as a two-step estimation procedure, where the unconditional covariance matrix of the observed process is estimated by a moment estimator in a first step. Conditional on this, the remaining parameters are estimated in a second step by QML estimation. This two-step procedure reduces the number of parameters in the numerical optimization step which leads to optimization over fewer parameters, regardless of the model has a restricted or unrestricted BEKK representation. Recently, Noureldin et al. (2014) have proposed the so-called multivariate rotated ARCH (RARCH) model which is estimated in two steps closely related to VT estimation and thus saving the number of varying parameters in the optimization step.

High-order moment restrictions for the multivariate BEKK-ARCH model are extensively discussed by Avarucci et al. (2013). They argue that fourth-order moment restrictions for QML estimation cannot be relaxed even in the simple ARCH form of the BEKK model. Note also in this respect that the strong moment restrictions for asymptotic QML inference in the multivariate BEKK model are in contrast to the very mild conditions found for univariate GARCH models, see e.g. Jensen and Rahbek (2004) and Francq and Zakoïan (2012b) who find that asymptotic inference in the GARCH model is feasible even if the observed process is explosive.

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. If a matrix A is positive definite we write $A > 0$, and if A is positive semi-definite we write $A \geq 0$. The vector $\text{vec}(A)$ stacks the columns of a matrix A, and $\text{vech}(A)$ stacks the columns of a square matrix A from the principal diagonal downwards. The trace of a square matrix A is denoted $\text{tr}\,(A)$, and the determinant is denoted $\det(A)$. 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 $A \otimes B = \{a_{ij}B\}$. Moreover, for a matrix A we define $A^{\otimes 2} := (A \otimes A)$. With $\xi_1, ..., \xi_n$ the n eigenvalues of a matrix A, $\rho(A) = \max_{i \in \{1,...,n\}} |\xi_i|$ is the spectral radius of A. The matrix (Euclidean) norm of the matrix, or vector A, is defined as $\|A\| = \sqrt{\text{tr}\,(A'A)}$, and the spectral norm is defined as $\|A\|_{spec} = \sqrt{\rho\,(A'A)}$. For an $m \times n$ matrix A, the $mn \times mn$ commutation matrix C_{mn} has the property $C_{mn}\text{vec}\,(A) = \text{vec}\,(A')$. The zero matrix $0_{m \times n}$ is an $m \times n$ matrix with all elements equal to zero. The letters K and ϕ denote strictly positive generic constants with $\phi < 1$.

2 The variance targeting (VT) BEKK model

As in Hafner and Preminger (2009b) we focus on the BEKK(1,1,1) model, the BEKK model hereafter, which is the predominantly used version of the BEKK models in applications, see Silvennoinen and Teräsvirta (2009). The BEKK model is given by

$$X_t = H_t^{1/2} Z_t, \tag{2.1}$$

where $t = 1, ..., T$, and Z_t is an i.i.d.$(0, I_d)$ sequence of random variables. Moreover, $H_t^{1/2}$ is the symmetric square root of H_t given by

$$H_t = C + AX_{t-1}X_{t-1}'A' + BH_{t-1}B', \tag{2.2}$$

with parameters (C, A, B). The matrix C is $(d \times d)$-dimensional, $C > 0$, and A and B are $(d \times d)$-dimensional real matrices such that H_t is positive definite. With respect to initial values, we consider estimation conditional on the initial values X_0

and $H_0 := h > 0$.

We make the following assumptions throughout the text:

Assumption 1. *The distribution of Z_t is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, and zero is an interior point of the support of the distribution.*

Assumption 2. *The matrices A and B satisfy $\rho\left[A^{\otimes 2} + B^{\otimes 2}\right] < 1$.*

By Theorem 2.4 of Boussama et al. (2011), Assumptions 1 and 2 imply the existence of a unique stationary and ergodic solution to the model in (2.1) and (2.2). Moreover, the stationary solution has finite second order moments, $E\left\|X_t\right\|^2 < \infty$, and variance $V\left[X_t\right] = E\left[H_t\right] = \Gamma$ with $\Gamma > 0$, which is the solution to

$$\Gamma = C + A\Gamma A' + B\Gamma B'. \tag{2.3}$$

Remark 2.1. Note that Boussama et al. (2011, Lemma 4.2 and Proposition 4.3) show that (2.3) has a positive definite solution if and only if Assumption 2 applies.

VT can be presented by writing the model in terms of the parameters Γ, A and B rather than as in the original BEKK formulation C, A and B,

$$H_t = \Gamma - A\Gamma A' - B\Gamma B' + AX_{t-1}X_{t-1}'A' + BH_{t-1}B'. \tag{2.4}$$

With $\Gamma > 0$ and A, B $(d \times d)$-dimensional real matrices we refer to (2.4) as the VT BEKK model, or the VT BEKK representation of H_t. In other words, in the VT BEKK model, the covariance (of the stationary solution) appears explicitly in the formulation, thus generalizing the univariate VT GARCH formulation of Francq et al. (2011).

In the next section we consider estimation of the VT BEKK model.

3 Variance targeting (VT) estimation

With $\Gamma > 0$ and A, B $(d \times d)$-dimensional let θ, $\theta \in \mathbb{R}^{3d^2}$, denote the parameter vector of the VT BEKK model obtained as $\theta := (\gamma', \ \lambda')'$ with

$$\gamma := \operatorname{vec}(\Gamma) \quad \text{and} \quad \lambda := \left[\operatorname{vec}(A)', \operatorname{vec}(B)'\right]'. \tag{3.1}$$

Likewise, define the parameter space $\Theta := \Theta_\gamma \times \Theta_\lambda \subset \mathbb{R}^{d^2} \times \mathbb{R}^{2d^2}$. To emphasize dependence on the parameters γ and λ, we write $H_t(\gamma, \lambda)$ such that the VT BEKK model can be restated as

$$X_t = H_t^{1/2}(\gamma,\lambda)Z_t, \tag{3.2}$$

where

$$H_t(\gamma,\lambda) = \Gamma - A\Gamma A' - B\Gamma B' + AX_{t-1}X_{t-1}'A' + BH_{t-1}(\gamma,\lambda)B'. \tag{3.3}$$

Whereas classical QML estimation of the BEKK model has been considered by Comte and Lieberman (2003) and Hafner and Preminger (2009b) (as a special case of the VEC GARCH model), we consider as emphasized here VT. VT estimation studied here is a two-step estimation method where γ, see (3.1), is estimated by a the sample unconditional covariance matrix of X_t, and next λ is estimated by QML by optimizing the VT log-likelihood, given below, with respect to λ. This two-step estimation yields the VT estimator (VTE) of θ, denoted $\widehat{\theta}_{VT}$ and is detailed next.

For the VT BEKK model, the Gaussian log-likelihood function is given by

$$L_{T,h}(\gamma,\lambda) := \frac{1}{T}\sum_{t=1}^{T} l_{t,h}(\gamma,\lambda) \tag{3.4}$$

with likelihood contributions,

$$l_{t,h}(\gamma,\lambda) := \log\left\{\det\left[H_{t,h}\left(\gamma,\lambda\right)\right]\right\} + \operatorname{tr}\left\{X_t X_t' H_{t,h}^{-1}\left(\gamma,\lambda\right)\right\}, \tag{3.5}$$

where, as mentioned, the initial values X_0 and $H_{0,h}(\gamma,\lambda) := h > 0$ are conditioned upon in the statistical analysis. Observe that the subscript h in the conditional covariance $H_{t,h}(\gamma,\lambda)$ is used to emphasize that $H_{t,h}(\gamma,\lambda)$, defined recursively in (3.3), is initiated for $t = 0$ at $H_{0,h}(\gamma,\lambda) = h$.

As mentioned, in the first step of the VT estimation, γ is simply estimated by the sample covariance matrix[2],

$$\widehat{\gamma}_{VT} = \operatorname{vec}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right). \tag{3.6}$$

Next, inserting $\widehat{\gamma}_{VT}$ from (3.6), the VTE of λ is then defined as

$$\widehat{\lambda}_{VT} := \underset{\lambda\in\Theta_\lambda}{\arg\min} L_{T,h}\left(\widehat{\gamma}_{VT},\lambda\right), \tag{3.7}$$

[2]Note that one could sum from $t = 0$ and all results stated would still apply. However, for presentational purposes we sum from $t = 1$.

with $\Theta_\lambda \subset \mathbb{R}^{2d^2}$. In total, the two estimation steps yield the VTE of γ and λ,

$$\widehat{\theta}_{VT} := \left(\widehat{\gamma}'_{VT}, \widehat{\lambda}'_{VT}\right)'.$$

For estimation of C in the original BEKK model in (2.2), observe that

$$\widehat{C}_{VT} := \widehat{\Gamma}_{VT} - \widehat{A}_{VT}\widehat{\Gamma}'_{VT}\widehat{A}_{VT} - \widehat{B}_{VT}\widehat{\Gamma}'_{VT}\widehat{B}_{VT}. \tag{3.8}$$

Remark 3.1. Although Z_t is not assumed to be necessarily Gaussian, we choose to work with the Gaussian log-likelihood and hence, similar to the notion of QML, one could denote the estimator QVTE.

Remark 3.2. The $3d^2$-dimensional parameter vector θ has $2d^2 + d\,(d+1)\,/2$ unique elements as Γ is symmetric.

Remark 3.3. Compared to QML estimation, where all parameters are estimated in one step by numerical optimization, the VT estimation implies that there are less varying parameters in the optimization step. That is, in the first step $d\,(d+1)\,/2$ parameters are estimated by method of moments, and in the second step $2d^2$ parameters are estimated through optimization.

Remark 3.4. VT estimation may in particular be used to estimate simplified BEKK models, such as the diagonal BEKK, where A and B are further simplified. This combination of targeting the unconditional covariance matrix and reducing the structure of the matrices A and B decreases the proportion of varying parameters relative to the total number of model parameters additionally.

4 Large-sample properties of VT estimation

In this section we derive the asymptotic properties of the VTE defined by (3.6) and (3.7). Specifically, we establish that $\widehat{\theta}_{VT}$ converges almost surely to its true value, θ_0, and that asymptotic normality applies under the assumption of finite sixth-order moments of X_t. We discuss this, as well as additional assumptions for the asymptotic analysis below. As mentioned in the introduction our results are new and extend the univariate results of Francq et al. (2011). All proofs are stated in Appendix A.

We make the following classical assumptions for the asymptotic analysis.

Assumption 3. *The process $\{X_t\}$ is strictly stationary and ergodic.*

Assumption 4. *The true parameter $\theta_0 \in \Theta$ and Θ is compact.*

Assumption 5. *For $\lambda \in \Theta_\lambda$, if $\lambda \neq \lambda_0$ then $H_t\left(\gamma_0, \lambda\right) \neq H_t\left(\gamma_0, \lambda_0\right)$ almost surely, for all $t \geq 1$.*

Remark 4.1. Regarding Assumption 3, recall that Assumptions 1 and 2 imply the existence of a strictly stationary ergodic solution $\{X_t\}$ in the BEKK model. This assumption is in line with the existing literature on QML estimation of multivariate GARCH and VT estimation, see Comte and Lieberman (2003), Hafner and Preminger (2009b), Hafner and Preminger (2009a), Francq and Zakoïan (2012a), and Francq et al. (2011). It implies in particular that for the asymptotic analysis the process $\{X_t\}_{t=0,1,\ldots}$ is assumed to be initiated from the invariant distribution. To relax this, and allow for an arbitrary initial value X_0, one can use arguments similar to Jensen and Rahbek (2004) and Kristensen and Rahbek (2005) where univariate (G)ARCH models are considered.

Remark 4.2. Assumptions 4, and 5 are in line with Comte and Lieberman (2003) and Hafner and Preminger (2009b). Assumption 5 concerns identification, which is a high-level condition. However, note that Engle and Kroner (1995) state sufficient conditions for parameter identification in the model with the original unrestricted BEKK representation, (2.2). These conditions include that the first element in the matrices A and B should be strictly positive.

We are now able to state the following result regarding consistency:

Theorem 4.1. *Under Assumptions 1-5, as* $T \to \infty$

$$\widehat{\theta}_{VT} \overset{a.s.}{\to} \theta_0.$$

The relatively weak sufficient conditions of Theorem 4.1 suggest that consistency of the VTE applies for many practical purposes. Moreover, the finite second-order moments of X_t as implied by Assumptions 1 and 2, are in line with the moment restrictions for consistency of the VTE in the univariate case, see Francq et al. (2011). However, the moment restrictions are stronger than the ones that are sufficient for consistency of the QML estimator (QMLE) for the BEKK model of the form (2.2) where finite second-order moments of X_t are not necessary, see Hafner and Preminger (2009b).

Next, for asymptotic normality of the VTE, we make two further assumptions:

Assumption 6. $E\left\|X_t\right\|^6 < \infty$.

Assumption 7. θ_0 *is in the interior of* Θ.

Remark 4.3. Whereas only finite fourth-order moments are required in order to show the existence of the joint asymptotic covariance matrix of $\widehat{\gamma}_{VT}$ and the score (in the direction λ), see Lemmas B.8 and B.9, finite sixth-order moments are assumed in order to show that the second-order derivatives of the log-likelihood function converge uniformly on the parameter space, see the proof of Lemma B.6. As in Francq et al. (2011) this can be reduced to fourth-order moments in the univariate case. In the multivariate case the model structure is more complex, and as a result the derivation of sufficient conditions for asymptotic normality more involved. The implied requirement of finite further higher-order moments for multivariate GARCH models, when compared to univariate GARCH models, has, as mentioned, recently been discussed in Avarucci et al. (2013) for QML estimation of BEKK-ARCH models.

Remark 4.4. The moment conditions in Assumption 6 are identical to the ones found in existing literature on asymptotic normality of the QMLE, see Hafner and Preminger (2009b).

Remark 4.5. As mentioned in Section 2, under Assumptions 1 and 2, Boussama et al. (2011) have shown the existence of a strictly stationary solution of the BEKK process with finite second-order moments of X_t. To our knowledge, conditions on the matrices A and B and the innovation Z_t such that the BEKK model has a strictly stationary and ergodic solution with $E\left\|X_t\right\|^6 < \infty$, have not been derived. There are different approaches to derive such conditions, and one is to establish that the Markov chain $\{W_t\}$, $W_t := \left(\text{vech}\left(H_t\right)', X_t'\right)'$, as in Boussama et al. (2011), satisfies a drift criterion with a drift function that bounds the sixth-order moments of X_t. Choosing such a drift function is non-trivial for the BEKK-GARCH case. However, in Appendix C we establish sufficient conditions for geometric ergodicity and finite eighth, sixth, as well as lower order moments for the BEKK-ARCH model.

Theorem 4.2. *Under Assumptions 1-7, as $T \to \infty$*

$$\sqrt{T}\left(\widehat{\theta}_{VT} - \theta_0\right) \xrightarrow{D} N\left(0, \begin{pmatrix} I_{d^2} & 0_{d^2\times 2d^2} \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix} \Omega_0 \begin{pmatrix} I_{d^2} & 0_{d^2\times 2d^2} \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix}'\right),$$

where the nonsingular matrix J_0 and the matrix K_0 are stated in (A.8), and Ω_0 is stated in (B.40) below.

Remark 4.6. Since $\widehat{\theta}_{VT}$ is computed in two steps, with the moment-based estimator inserted, it is expected that the VTE is not efficient as also shown by Francq et al. (2011) for the univariate case. In particular Francq et al. (2011) show that the asymptotic covariance matrix of the VTE minus the asymptotic covariance matrix of the QMLE is positive semidefinite as would also be expected in the multivariate case. This property of the VTE is also confirmed by simulations in Section 5.4.

Remark 4.7. From Theorem 4.1 one may also consider taking a Newton step from the consistent estimate $\widehat{\theta}_{VT}$ in order to improve on the efficiency, see e.g. Robinson

(2005). That is θ is estimated in a first-step using VT, and from this estimate a Newton step is taken in the direction of the entire parameter vector θ in order to achieve a new estimate that might be more efficient than the one computed by VT.

Remark 4.8. As kindly raised by a referee, one may relax the assumption of compact Θ_γ, where $\Theta = \Theta_\gamma \times \Theta_\lambda$ and $\gamma \in \Theta_\gamma$, $\lambda \in \Theta_\lambda$. Indeed, for non-compact Θ_γ the consistency of $\hat{\theta}_{VT}$ still applies, as can be seen by using the strong consistency of the method of moment estimator $\hat{\gamma}_{VT}$, and modifying the proof in the appendix of consistency for $\hat{\lambda}_{VT}$ accordingly. However, for the derivation of the (joint) asymptotic distribution of $\hat{\gamma}_{VT}$ and $\hat{\lambda}_{VT}$ in Theorem 4.2 an expansion of the score is needed in both parameter directions. We employ here a mean-value expansion, using compactness of Θ and finite sixth-order moments of X_t. In the case of non-compact parameter space, as in Francq et al. (2011) and Jensen and Rahbek (2004), an alternative would be to employ an expansion in a (local and indeed compact) neighbourhood of the true parameter value θ_0. This approach may lead to requirements of higher order finite moments due to the multivariate complexity, as in Comte and Lieberman (2003, Theorem 3) where finite eighth-order moments are assumed for the QMLE of the BEKK-GARCH model.

Given the asymptotic distribution of $\widehat{\theta}_{VT}$, we can state the asymptotic distribution of the VTE for (C, A, B) in the original BEKK parametrization in (2.2):

Corollary 4.1. *Under the assumptions of Theorem 4.2, as $T \to \infty$*

$$\sqrt{T}\begin{pmatrix} \mathrm{vec}(\widehat{C}_{VT} - C_0) \\ \mathrm{vec}(\widehat{A}_{VT} - A_0) \\ \mathrm{vec}(\widehat{B}_{VT} - B_0) \end{pmatrix} \overset{D}{\to} N\left(0, \Sigma_0 \begin{pmatrix} I_{d^2} & 0_{d^2\times 2d^2} \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix} \Omega_0 \begin{pmatrix} I_{d^2} & 0_{d^2\times 2d^2} \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix}' \Sigma_0' \right),$$

where

$$\Sigma_0 = \begin{pmatrix} I_{d^2} - (A_0^{\otimes 2} + B_0^{\otimes 2}) & -(I_{d^2} + C_{dd})\left[(A_0\Gamma_0) \otimes I_d\right] & -(I_{d^2} + C_{dd})\left[(B_0\Gamma_0) \otimes I_d\right] \\ 0_{d^2\times d^2} & I_{d^2} & 0_{d^2\times d^2} \\ 0_{d^2\times d^2} & 0_{d^2\times d^2} & I_{d^2} \end{pmatrix}.$$

Remark 4.9. The asymptotic covariance matrix of $\widehat{\theta}_{VT}$, stated in Theorem 4.2, may in practice, using numerical derivatives, be estimated by

$$\begin{pmatrix} I_{d^2} & 0_{d^2\times 2d^2} \\ -\widehat{J}^{-1}\widehat{K} & -\widehat{J}^{-1} \end{pmatrix} \widehat{\Omega} \begin{pmatrix} I_{d^2} & 0_{d^2\times 2d^2} \\ -\widehat{J}^{-1}\widehat{K} & -\widehat{J}^{-1} \end{pmatrix}',$$

with

$$\widehat{J} := \frac{1}{T}\sum_{t=1}^{T}\widehat{J}_t \quad \text{and} \quad \widehat{K} := \frac{1}{T}\sum_{t=1}^{T}\widehat{K}_t,$$

where

$$\widehat{J}_t := \left.\frac{\partial^2 l_{t,h}(\theta)}{\partial\lambda\partial\lambda'}\right|_{\theta=\widehat{\theta}_{VT}} \quad \text{and} \quad \widehat{K}_t := \left.\frac{\partial^2 l_{t,h}(\theta)}{\partial\lambda\partial\gamma'}\right|_{\theta=\widehat{\theta}_{VT}},$$

and

$$\widehat{\Omega} := \frac{1}{T}\sum_{t=1}^{T}\widehat{\omega}_t\widehat{\omega}_t' \quad \text{with} \quad \widehat{\omega}_t := \begin{pmatrix} \operatorname{vec}(X_tX_t') - \widehat{\gamma}_{VT} \\ \left.\frac{\partial l_{t,h}(\theta)}{\partial\lambda}\right|_{\theta=\widehat{\theta}_{VT}} \end{pmatrix}.$$

The asymptotic covariance matrix of $\left[\operatorname{vec}(\widehat{C}_{VT})', \operatorname{vec}(\widehat{A}_{VT})', \operatorname{vec}(\widehat{B}_{VT})'\right]'$, stated in Corollary 4.1, may be estimated in a similar way with the matrices entering Σ_0 replaced with their estimated counterparts, $\widehat{\Gamma}_{VT}$, $\widehat{A}_{VT}$, and $\widehat{B}_{VT}$. Note that we can replace θ_0 by $\hat{\theta}_{VT}$ due to the established consistency together with the uniform law of large numbers as applied repeatedly in the appendix.

Remark 4.10. As an alternative to the method of moment estimator for γ one may consider other estimators for γ, as for example a sample covariance estimator based on shrinkage techniques, see Ledoit and Wolf (2004) and Hafner and Reznikova (2012). The asymptotic analysis of the VTE, we expect, will have to be modified in order to deal with other choices of estimators for γ.

5 Simulation study

In this section we illustrate the theoretical results of Section 4 through simulations. Specifically, we simulate the large-sample distribution of the VTE for three different cases. As data-generating process (DGP) we choose the bivariate diagonal-BEKK-ARCH with Gaussian noise, that is the process in (2.2) with $d = 2$, A diagonal, $B = 0$, and Z_t i.i.d.N$(0, I_2)$. In Appendix C we derive sufficient conditions for the matrix A in a BEKK-ARCH process such that $\{X_t\}$ is geometrically ergodic and such that certain moments of the stationary solution are finite. In the first simulation the sufficient moment restrictions for asymptotic normality, see Theorem 4.2, are satisfied - in particular the DGP has finite sixth-order moments. In the second case the sufficient conditions for finite fourth-order moments of the DGP are satisfied, but the derived sufficient conditions for finite sixth-order moments are not satisfied, which means that the conditions for asymptotic normality in Theorem 4.2 might be violated, and hence the VTE for the entire parameter vector may not be asymptotically normal. However, the moment restrictions for asymptotic normality of the VTE of γ are satisfied. In the last case the sufficient conditions for finite second-order moments of the DGP are satisfied, but the fourth-order moments are not finite. In this case the VTE of γ (and of λ) might not be asymptotically normally distributed. Moreover, we consider the relative efficiency between the VTE and the QMLE in the case where the DGP has finite sixth-order moments.

5.1 Case 1: The DGP satisfies the sufficient conditions for asymptotic normality

Consider the bivariate DGP for X_t given by (2.2) with $B = 0$. That is

$$X_t = H_t^{1/2} Z_t, \; Z_t \text{ i.i.d.N}(0, I_2), \text{ and } H_t = C + A X_{t-1} X_{t-1}' A', \tag{5.1}$$

$$\text{with } C = (C_{ij})_{i,j=1,2} = \begin{pmatrix} 0.8 & 0.5 \\ 0.5 & 0.7 \end{pmatrix}. \tag{5.2}$$

First we choose A such that $E\,\|X_t\|^6 < \infty$. Specifically, we set

$$A = (A_{ij})_{i,j=1,2} = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.5 \end{pmatrix}, \tag{5.3}$$

and observe that $\rho\,(A \otimes A) = 0.36$. By Theorem C.1 the stationary solution of the process has $E\,\|X_t\|^6 < \infty$, and hence the moment restrictions of Theorem 4.2 are satisfied.

For $N = 1000$ realizations of (5.1)-(5.3), $t = 1, ..., 10000$, $H_1 = C$, we estimate A and C by VT using the G@RCH Package version 6.1 for OxMetrics 6.1.

Figure 5.1 contains density and QQ plots of the estimates of A_{11} and C_{11} in the process (5.1)-(5.3). The figure suggests that the estimates seem to fit a normal distribution well, which is in line with Theorem 4.2. We now turn to the second case where the DGP does not meet the conditions of Theorem 4.2.

5.2 Case 2: The DGP satisfies sufficient conditions for $E\,\|X_t\|^4 < \infty$.

Next we consider the DGP (5.1)-(5.2) and choose A such that $E\,\|X_t\|^4 < \infty$, while the sufficient conditions for $E\,\|X_t\|^6 < \infty$ are not met. Specifically, we set

$$A = (A_{ij})_{i,j=1,2} = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.5 \end{pmatrix}, \tag{5.4}$$

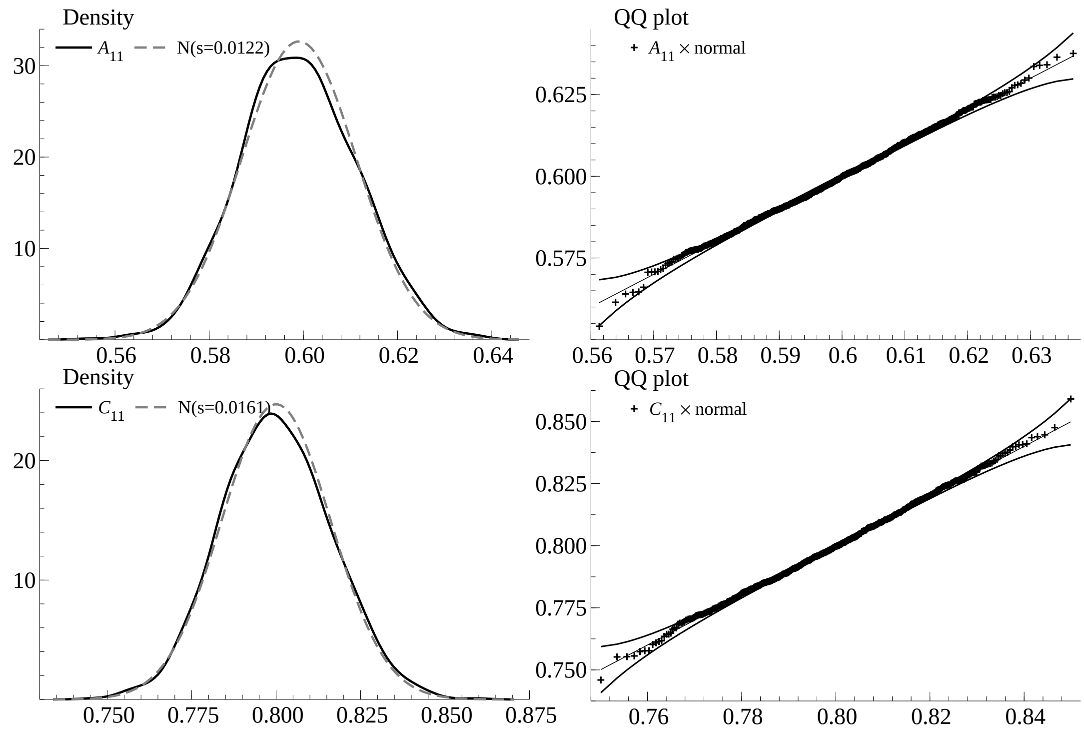

Figure 5.1: Density and QQ plots of $N = 1000$ VT estimates of A_{11} and C_{11} of the process (5.1)-(5.3). In the density plots the solid line is the plot of the estimated density of the VT estimates, and the dashed line is the plot for the normal distribution. The QQ plots compare the quantiles of the estimate with the ones of a normal distribution (crosses). The solid lines are the asymptotic 95% standard error bands of a normal distribution.

so that $\rho(A \otimes A) = 0.75^2 = 0.5625 < \frac{1}{\sqrt{3}}$ and the DGP is geometrically ergodic with $E\|X_t\|^4 < \infty$ for the stationary solution by Theorem C.1. As in Case 1 we consider $N = 1000$ realizations of the DGP and estimate A and C by VT.

Figure 5.2 contains density and QQ plots of the estimates of A_{11} and C_{11} in the process (5.1),(5.2),(5.4). The estimates of A_{11} do not seem to be well approximated by the Gaussian distribution: The density is skewed compared to a normal distribution, which can also be deduced by the s-shape of the points in the QQ plot. Contrary to this, the estimates of C_{11} do seem to fit a normal distribution, except for a few outliers (see QQ plot), which may be explained as follows. Recall from (3.8) that $\text{vec}(\widehat{C}_{VT}) = [I_{d^2} - (\widehat{A}_{VT} \otimes \widehat{A}_{VT})]\widehat{\gamma}_{VT}$, so the distribution of $\text{vec}(\widehat{C}_{VT})$ (or more correctly $\sqrt{T}[\text{vec}(\widehat{C}_{VT}) - \text{vec}(C)]$) depends on the distribution of $(\widehat{A}_{VT} \otimes \widehat{A}_{VT})$ and $\widehat{\gamma}_{VT} = \text{vec}(\widehat{\Gamma}_{VT})$. By Lemma B.8, the proof of Lemma B.9, and the Central Limit Theorem for martingale difference sequences, applied in the proof of Lemma B.10, it follows that $\text{vec}(\widehat{\Gamma}_{VT})$ is asymptotically Gaussian with asymptotic covariance matrix $E[\mathcal{A}_t]$ defined in (B.36), if $E\|X_t\|^4 < \infty$. This moment restriction holds for our

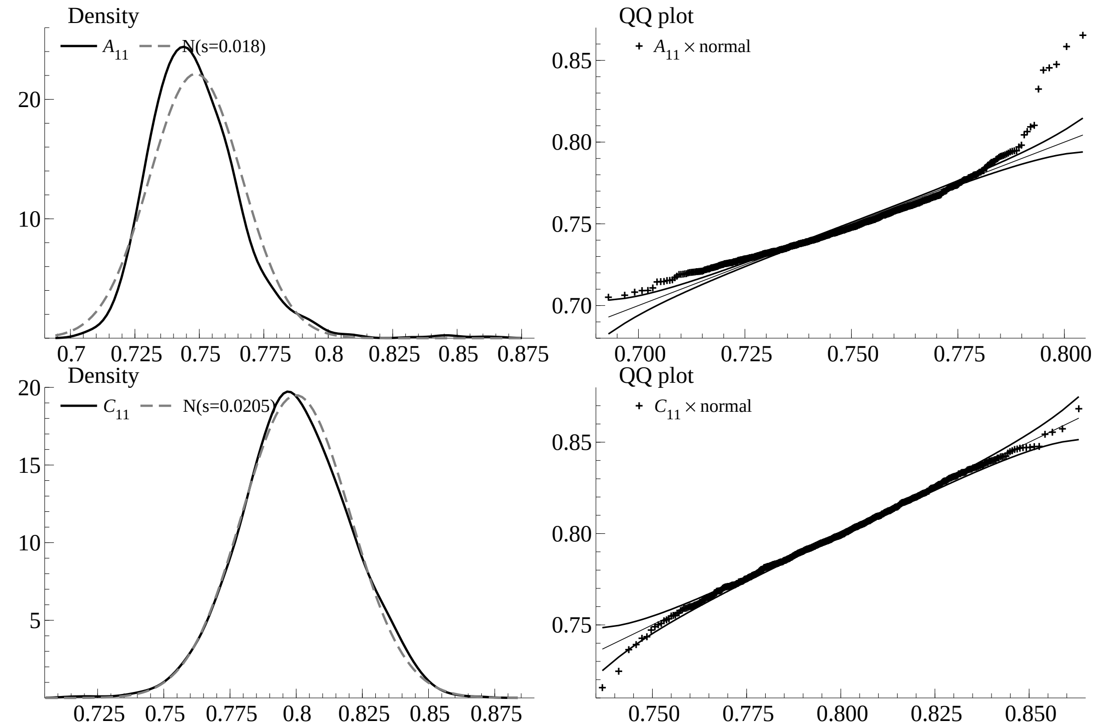

Figure 5.2: Density and QQ plots of $N = 1000$ VT estimates of A_{11} and C_{11} of the process (5.1),(5.2),(5.4). In the density plots the solid line is the plot of the estimated density of the VT estimates, and the dashed line is the plot for the normal distribution. The QQ plots compare the quantiles of the estimate with the ones of a normal distribution (crosses). The solid lines are the asymptotic 95% standard error bands of a normal distribution.

choice of DGP, so $\sqrt{T}[\text{vec}(\widehat{\Gamma}_{VT}) - \text{vec}\,(\Gamma_0)]$ is indeed asymptotically Gaussian. Next

$$\begin{aligned} \sqrt{T}\text{vec}\left(\widehat{C}_{VT} - C\right) &= \left[I_{d^2} - \left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right)\right] \sqrt{T}\text{vec}\left(\widehat{\Gamma}_{VT} - \Gamma\right) \\ &\quad -\sqrt{T}\left[\left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right) - (A \otimes A)\right] \text{vec}\,(\Gamma)\,. \end{aligned} \tag{5.5}$$

If $E\left\|X_t\right\|^4 < \infty$ the first term of the right hand side of (5.5) converges to a Gaussian distribution, and determines the distribution of $\sqrt{T}\text{vec}(\widehat{C}_{VT} - C)$, provided the last term tends to zero in probability. This would indeed be the case if $[(\widehat{A}_{VT} \otimes \widehat{A}_{VT}) - (A \otimes A)]$ converges to some (unknown) distribution with a rate higher than $\sqrt{T}$, say $T^{1/2+\delta}$ for some $\delta > 0$ which will be explored elsewhere.

Next, we turn to the case where sufficient conditions for $E\left\|X_t\right\|^2 < \infty$ are satisfied.

5.3 Case 3: The DGP has $E\|X_t\|^2 < \infty$, but $E\|X_t\|^4$ is not finite

We consider the DGP (5.1)-(5.2) and choose A such that $E\|X_t\|^2 < \infty$, but $E\|X_t\|^4$ is not finite as discussed below. We set

$$A = (A_{ij})_{i,j=1,2} = \begin{pmatrix} 0.95 & 0 \\ 0 & 0.8 \end{pmatrix}, \tag{5.6}$$

such that $\rho(A \otimes A) = 0.95^2 = 0.9025 < 1$ and the DGP is geometrically ergodic with $E\|X_t\|^2 < \infty$ by Theorem C.1. Note that the conditions in Theorem C.1 are sufficient, so based on that theorem we cannot say whether $E\|X_t\|^4$ is finite or not. However, Theorem 3 in Hafner (2003) provides necessary and sufficient conditions for $E\|X_t\|^4 < \infty$. The necessary and sufficient condition for our choice of DGP is that $\rho[(\widetilde{A} \otimes \widetilde{A})G_2] < 1$, where $\widetilde{A} := D_2^+ (A \otimes A) D_2$, with D_2 a (4×3)-dimensional duplication matrix, $D_2^+ = (D_2' D_2)^{-1} D_2'$, and G_2 is a (9×9)-dimensional matrix stated in Hafner (2003, equation (12)). For our choice of matrix A, $\rho[(\widetilde{A} \otimes \widetilde{A})G_2] \approx 2.44$, so $E\|X_t\|^4$ is not finite. As in Case 1 and 2 we consider $N = 1000$ realizations of the DGP and estimate A and C by VT.

Figure 5.3 contains density and QQ plots of the estimates of A_{11} and C_{11} in the process (5.1),(5.2),(5.6). None of the estimates seem to be well approximated by the normal distribution. In light of Case 2 this might be explained by the fact that $\sqrt{T}\text{vec}(\widehat{\Gamma}_{VT} - \Gamma)$ is not asymptotically normal as $E\|X_t\|^4$ is not finite.

5.4 Relative efficiency between VT and QML estimation

Similar to the simulation study of Francq et al. (2011), we now consider the relative efficiency between the VTE and the QMLE by a small simulation investigation. Specifically, we simulate the asymptotic covariance matrix of the two estimators for the parameter vector $\theta = (A_{11}, A_{22}, C_{11}, C_{12}, C_{22})'$ when the DGP is chosen as in Section 5.1. Based on 1000 realizations of the DGP with a sample size of 10000 observations, we estimate $\text{Avar}(\widehat{\theta}_{VT}) := \lim_{T\to\infty}\text{Var}\{\sqrt{T}(\widehat{\theta}_{VT} - \theta)\}$ and $\text{Avar}(\widehat{\theta}_{QML}) := \lim_{T\to\infty}\text{Var}\{\sqrt{T}(\widehat{\theta}_{QML} - \theta)\}$. In order to evaluate the relative efficiency of the two estimators, we consider the difference between the two asymptotic covariance matrices, $\text{Avar}(\widehat{\theta}_{VT}) - \text{Avar}(\widehat{\theta}_{QML})$. As the asymptotic distributions of $\widehat{\theta}_{VT}$ and $\widehat{\theta}_{QML}$ are unknown when $E\|X_t\|^6$ is not finite, we cannot compare estimated covariances in this case. Hence we consider the choice of DGP with $E\|X_t\|^6 < \infty$ in line with Theorem 4.2. Estimated values of $\text{Avar}(\widehat{\theta}_{VT}) - \text{Avar}(\widehat{\theta}_{QML})$ are presented in Table 5.1. Trying with different parameter combinations in the region where

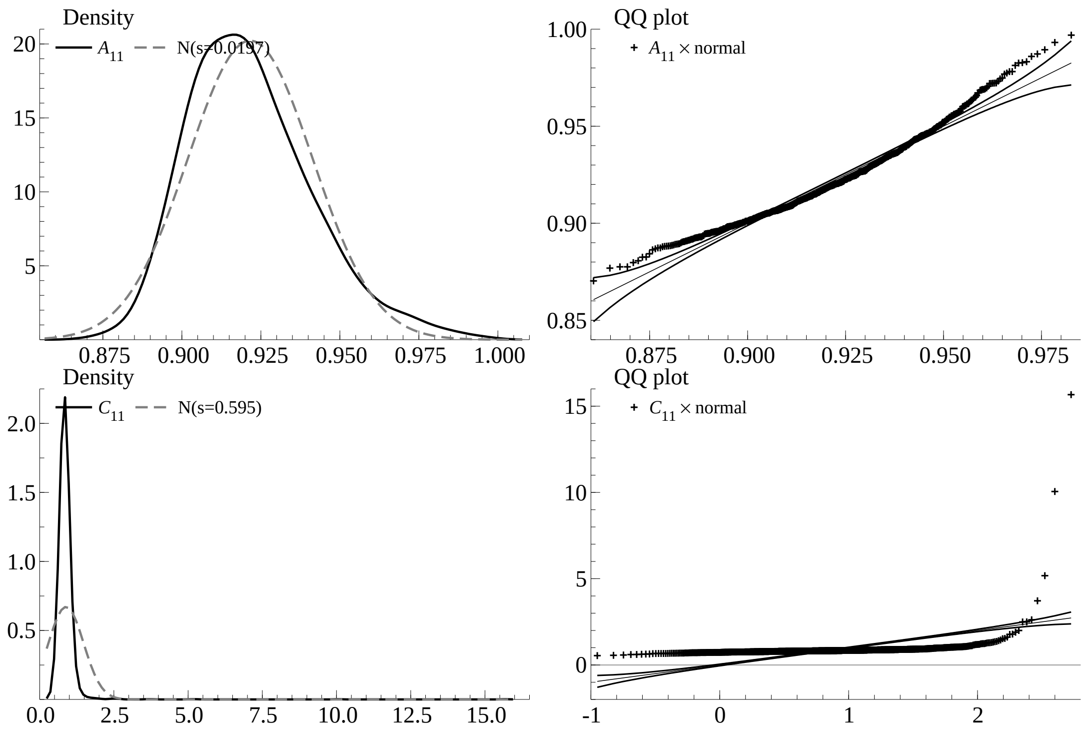

Figure 5.3: Density and QQ plots of $N = 1000$ VT estimates of A_{11} and C_{11} of the process (5.1),(5.2),(5.6). In the density plots the solid line is the plot of the estimated density of the VT estimates, and the dashed line is the plot for the normal distribution. The QQ plots compare the quantiles of the estimate with the ones of a normal distribution (crosses). The solid lines are the asymptotic 95% standard error bands of a normal distribution.

$E\left\|X_t\right\|^6 < \infty$, we found little differences, and hence we report a representative result here.

Based on the eigenvalues of $\widehat{\text{Avar}(\hat{\theta}_{VT})} - \widehat{\text{Avar}(\hat{\theta}_{QML})}$ the matrix is positive semidefinite, and hence indicating (as clearly expected) that the VTE cannot be more efficient than the QMLE. We see that although the elements of the covariance matrix for the VTE, in general, are larger than the ones of the QMLE, the method of VT does not seem to imply a substantial loss of efficiency. Our conclusion is directly comparable with the simulation study in Francq et al. (2011) for the univariate ARCH model, where Francq et al. (2011) find that the asymptotic covariance matrices of the two estimators likewise are similar. In fact, their conclusion holds even when the true parameter vector is moderately away from the region where $E\left\|X_t\right\|^4$ is close to not being finite. This region, as mentioned, we do not consider for the BEKK-ARCH model since in the multivariate case the asymptotic distributions of the VT and the QML estimators are unknown at present when $E\left\|X_t\right\|^6$ is not finite
.

$\widehat{\text{Avar}}(\widehat{\theta}_{VT}) - \widehat{\text{Avar}}(\widehat{\theta}_{QML})$	Eigenvalues
$\begin{pmatrix} 0.17 & 0.06 & 0.09 & 0.00 & -0.06 \\ 0.06 & 0.05 & 0.03 & 0.02 & 0.02 \\ 0.09 & 0.03 & 0.45 & 0.45 & 0.15 \\ 0.00 & 0.02 & 0.45 & 0.52 & 0.23 \\ -0.06 & 0.02 & 0.15 & 0.23 & 0.16 \end{pmatrix}$	1.03 0.25 0.07 0.00 0.00

Table 5.1: Difference between estimated covariance matrices of the VTE and the QML estimator for $\theta = (A_{11}, A_{22}, C_{11}, C_{12}, C_{22})'$ for the DGP (5.1)-(5.3)

5.5 Brief summary of simulation-based conclusions

The simulation study suggests that asymptotic normality of the VTE applies when X_t has finite sixth-order moments, which is in line with the theory derived in Section 4. Case 2 indicated that when relaxing the moment restrictions, $\widehat{A}_{VT}$ did no longer seem to be asymptotically normally distributed. This indicates that $E\|X_t\|^6 < \infty$ may be a necessary moment restriction for doing standard large-sample inference in the VT BEKK model. Case 2 also indicated that $\widehat{C}_{VT}$ might be asymptotically normal even in the case where the sufficient condition for $E\|X_t\|^6 < \infty$ is not satisfied, but $E\|X_t\|^4 < \infty$. Case 3 indicated that when $E\|X_t\|^2 < \infty$ but $E\|X_t\|^4$ not finite, neither $\widehat{A}_{VT}$ nor $\widehat{C}_{VT}$ seemed to be asymptotically normal. Moreover, simulations showed that the VT does not imply a substantial loss in efficiency compared to QML in the case where the DGP has finite sixth-order moments and both the VTE and the QMLE are known to be asymptotically normal.

6 Extensions and concluding remarks

We derive the asymptotic properties of the variance targeting estimator (VTE) for the multivariate BEKK-GARCH model. Variance targeting estimation relies on reparametrizing the BEKK model in (2.1)-(2.2) such that the unconditional covariance matrix of the observed process appears explicitly in the model equation. This yields a reparametrized (variance targeting) model given by (3.2)-(3.3). The parameters of the model are estimated in two steps yielding the VTE: The unconditional covariance matrix of the observed process is estimated by method of moments, and conditional on this, the rest of the parameters are estimated by QMLE. We establish that the VTE is consistent when the observed process has finite second-order moments, and is asymptotically Gaussian when the process has finite sixth-order moments. Our simulations indicate that these moment restrictions cannot be relaxed, and that the VTE does not seem to be substantially less efficient compared to the QMLE in the region of the parameter space where both estimators are known to be asymptotically normal.

With regards to extensions it seems interesting to apply our results for the analysis of the multivariate Rotated GARCH (RARCH) model recently proposed in Noureldin et al. (2014). The model and the proposed two-step estimation procedure has some similarities to VTE, and it may indeed be possible to exploit some of our theoretical results when investigating the asymptotic properties of the two-step estimator for the RARCH. Furthermore, with respect to the possibility of weakening the finite moment restrictions imposed for asymptotic normality, it would be interesting to consider the results on tail-trimming in Hill and Renault (2012) and maybe combine these with the idea of modified likelihood considered in Lange et al. (2011).

Appendix A Proofs of Theorems 4.1 and 4.2 and Corollary 4.1

In the asymptotic analysis we assume that the observed process $\{X_t\}$ is strictly stationary and ergodic, see Assumption 3. Recall from (3.4) that the log-likelihood is given by

$$L_{T,h}(\gamma,\lambda) = \frac{1}{T}\sum_{t=1}^{T} l_{t,h}(\gamma,\lambda) \tag{A.1}$$

with

$$l_{t,h}(\gamma,\lambda) = \log\left\{\det\left[H_{t,h}\left(\gamma,\lambda\right)\right]\right\} + \operatorname{tr}\left\{X_t X_t' H_{t,h}^{-1}\left(\gamma,\lambda\right)\right\}, \tag{A.2}$$

where $H_{t,h}\left(\gamma,\lambda\right)$ is given by the recursions

$$H_{t,h}(\gamma,\lambda) = \Gamma - A\Gamma A' - B\Gamma B' + AX_{t-1}X_{t-1}'A' + BH_{t-1,h}(\gamma,\lambda)B',$$

with fixed initial value $H_{0,h}\left(\gamma,\lambda\right) = h > 0$. For technical reasons, it is convenient to introduce the strictly stationary and ergodic solution to (3.3), $\{H_t(\gamma,\lambda)\}$. To distinguish between $H_{t,h}(\gamma,\lambda)$ and $H_t(\gamma,\lambda)$ we introduce correspondingly

$$L_T(\gamma,\lambda) = \frac{1}{T}\sum_{t=1}^{T} l_t(\gamma,\lambda), \tag{A.3}$$

with

$$l_t(\gamma,\lambda) = \log\left\{\det\left[H_t\left(\gamma,\lambda\right)\right]\right\} + \operatorname{tr}\left\{X_t X_t' H_t^{-1}\left(\gamma,\lambda\right)\right\}. \tag{A.4}$$

Observe that as both $H_{t,h}(\gamma,\lambda)$ and $H_t(\gamma,\lambda)$ are defined for the same strictly stationary and ergodic $\{X_t\}$,

$$\operatorname{vec}\left[H_t(\gamma,\lambda)-H_{t,h}(\gamma,\lambda)\right]=(B\otimes B)\operatorname{vec}\left[H_{t-1}(\gamma,\lambda)-H_{t-1,h}(\gamma,\lambda)\right],\ t\geq 1. \quad \text{(A.5)}$$

A.1 Proof of Theorem 4.1

For presentational purposes most steps of the proof rely on lemmas from Section B.1 below.

Observe initially that by Assumption 3 and as $E\left\|X_t\right\|^2<\infty$, it follows by the ergodic theorem that as $T\to\infty$,

$$\widehat{\gamma}_{VT}\overset{a.s.}{\to}\gamma_0. \quad \text{(A.6)}$$

It now remains to verify that $\widehat{\lambda}_{VT}$ is consistent. The proof follows the technique from the proof of Theorem 2.1 in Newey and McFadden (1994). We have that for any $\varepsilon>0$ almost surely for large enough T

$$\begin{aligned}
E\left[l_t\left(\gamma_0,\widehat{\lambda}_{VT}\right)\right] &< L_T\left(\gamma_0,\widehat{\lambda}_{VT}\right)+\varepsilon/5 \text{ by Lemma B.2}\\
L_T\left(\gamma_0,\widehat{\lambda}_{VT}\right) &< L_{T,h}\left(\widehat{\gamma}_{VT},\widehat{\lambda}_{VT}\right)+\varepsilon/5 \text{ by Lemma B.1}\\
L_{T,h}\left(\widehat{\gamma}_{VT},\widehat{\lambda}_{VT}\right) &< L_{T,h}\left(\widehat{\gamma}_{VT},\lambda_0\right)+\varepsilon/5 \text{ by the definition of } \widehat{\lambda}_{VT}\text{, see (3.7)}\\
L_{T,h}\left(\widehat{\gamma}_{VT},\lambda_0\right) &< L_T\left(\gamma_0,\lambda_0\right)+\varepsilon/5 \text{ by Lemma B.1}\\
L_T\left(\gamma_0,\lambda_0\right) &< E\left[l_t\left(\gamma_0,\lambda_0\right)\right]+\varepsilon/5 \text{ by Lemma B.2.}
\end{aligned}$$

Hence for any $\varepsilon>0$,

$$E\left[l_t\left(\gamma_0,\widehat{\lambda}_{VT}\right)\right]<E\left[l_t\left(\gamma_0,\lambda_0\right)\right]+\varepsilon.$$

By Lemma B.3 and standard arguments as in Newey and McFadden (1994), it follows that as $T\to\infty$, $\widehat{\lambda}_{VT}\overset{a.s.}{\to}\lambda_0$. Combined with (A.6), we conclude that as $T\to\infty$, $\widehat{\theta}_{VT}\overset{a.s.}{\to}\theta_0$.

We now turn to the proof of asymptotic normality of the VTE.

A.2 Proof of Theorem 4.2

Again, for presentational purposes most steps of the proof rely on lemmas stated in Section B.2. By Assumption 7, the definition of $\widehat{\lambda}_{VT}$ in (3.7), and the mean-value theorem

$$0=\frac{\partial L_{T,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda}+K_{T,h}\left(\theta^*\right)\left(\widehat{\gamma}_{VT}-\gamma_0\right)+J_{T,h}\left(\theta^*\right)\left(\widehat{\lambda}_{VT}-\lambda_0\right)\tag{A.7}$$

where

$$\frac{\partial L_{T,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda}:=\left.\frac{\partial L_{T,h}\left(\gamma,\lambda\right)}{\partial\lambda}\right|_{\theta=\theta_0},\quad K_{T,h}\left(\theta^*\right):=\left.\frac{\partial^2 L_{T,x}\left(\gamma,\lambda\right)}{\partial\lambda\partial\gamma'}\right|_{\theta=\theta^*}$$
$$\text{and}\quad J_{T,h}\left(\theta^*\right):=\left.\frac{\partial^2 L_{T,h}\left(\gamma,\lambda\right)}{\partial\lambda\partial\lambda'}\right|_{\theta=\theta^*},$$

with θ^* between θ_0 and $\widehat{\theta}_{VT}$, as in Lemma 1 of Jensen and Rahbek (2004). Let

$$\frac{\partial L_T\left(\gamma_0,\lambda_0\right)}{\partial\lambda}:=\left.\frac{\partial L_T\left(\gamma,\lambda\right)}{\partial\lambda}\right|_{\theta=\theta_0},\ K_T\left(\theta^*\right):=\left.\frac{\partial^2 L_T\left(\gamma,\lambda\right)}{\partial\lambda\partial\gamma'}\right|_{\theta=\theta^*},\ J_T\left(\theta^*\right):=\left.\frac{\partial^2 L_T\left(\gamma,\lambda\right)}{\partial\lambda\partial\lambda'}\right|_{\theta=\theta^*}.$$

Moreover, define

$$J_0:=E\left[\left.\frac{\partial^2 l_t\left(\gamma,\lambda\right)}{\partial\lambda\partial\lambda'}\right|_{\theta=\theta_0}\right]\quad\text{and}\quad K_0:=E\left[\left.\frac{\partial^2 l_t\left(\gamma,\lambda\right)}{\partial\lambda\partial\gamma'}\right|_{\theta=\theta_0}\right].\tag{A.8}$$

Observe that by Lemma B.6, Lemma B.7, and the consistency of $\hat{\theta}_{VT}$, $J_T\left(\theta^*\right)$ is invertible with probability approaching one. Hence, by Lemma B.11

$$\sqrt{T}\mathrm{vec}\left(\widehat{\lambda}_{VT}-\lambda_0\right)=-J_T\left(\theta^*\right)^{-1}\sqrt{T}\frac{\partial L_T\left(\gamma_0,\lambda_0\right)}{\partial\lambda}-J_T\left(\theta^*\right)^{-1}K_T\left(\theta^*\right)\sqrt{T}\left(\widehat{\gamma}_{VT}-\gamma_0\right)+o_P\left(1\right).$$

Collecting terms we get

$$\sqrt{T}\left(\widehat{\theta}_{VT}-\theta_0\right)=\begin{pmatrix}I_{d^2} & 0_{d^2\times 2d^2}\\ -J_T\left(\theta^*\right)^{-1}K_T\left(\theta^*\right) & -J_T\left(\theta^*\right)^{-1}\end{pmatrix}\sqrt{T}\begin{pmatrix}\left(\widehat{\gamma}_{VT}-\gamma_0\right)\\ \frac{\partial L_T\left(\gamma_0,\lambda_0\right)}{\partial\lambda}\end{pmatrix}+o_P\left(1\right).$$

By Lemma B.6 and Theorem 4.1,

$$\begin{pmatrix}I_{d^2} & 0_{d^2\times 2d^2}\\ -J_T\left(\theta^*\right)^{-1}K_T\left(\theta^*\right) & -J_T\left(\theta^*\right)^{-1}\end{pmatrix}\overset{P}{\to}\begin{pmatrix}I_{d^2} & 0_{d^2\times 2d^2}\\ -J_0^{-1}K_0 & -J_0^{-1}\end{pmatrix}.$$

The asymptotic normality of the VTE now follows from Lemmas B.8 and B.10 and Slutzky's theorem.

A.3 Proof of Corollary 4.1

Note that $\mathrm{vec}\left[C\left(\theta\right)\right]=\left[I_{d^2}-\left(A\otimes A\right)-\left(B\otimes B\right)\right]\gamma$. Since $\theta=\begin{bmatrix}\gamma' & \lambda'\end{bmatrix}'$,

$$\frac{\partial \mathrm{vec}\left[C\left(\theta\right)\right]}{\partial\theta'}=\begin{bmatrix}\frac{\partial \mathrm{vec}[C(\theta)]}{\partial\gamma'} & \frac{\partial \mathrm{vec}[C(\theta)]}{\partial \mathrm{vec}(A)'} & \frac{\partial \mathrm{vec}[C(\theta)]}{\partial \mathrm{vec}(B)'}\end{bmatrix}'.$$

We have that

$$\frac{\partial \mathrm{vec}\left[C\left(\theta\right)\right]}{\partial\gamma'}=\left[I_{d^2}-\left(A\otimes A\right)-\left(B\otimes B\right)\right],$$

and

$$\frac{\partial \mathrm{vec}\left[C\left(\theta\right)\right]}{\partial \mathrm{vec}\left(A\right)'}=-\frac{\partial \mathrm{vec}\left(A\Gamma A'\right)}{\partial \mathrm{vec}\left(A\right)'}.$$

Since Γ is symmetric

$$\frac{\partial \mathrm{vec}\left(A\Gamma A'\right)}{\partial \mathrm{vec}\left(A\right)'}=\left[I_{d^2}+C_{dd}\right]\left[\left(A\Gamma\right)\otimes I_d\right],$$

which follows by Result 7 in Section 10.5.1 of Lütkepohl (1996). Likewise,

$$\frac{\partial \mathrm{vec}\left(B\Gamma B'\right)}{\partial \mathrm{vec}\left(B\right)'}=\left[I_{d^2}+C_{dd}\right]\left[\left(B\Gamma\right)\otimes I_d\right].$$

Moreover, observe that

$$\frac{\partial \mathrm{vec}\left(A\right)}{\partial\theta'}=\left(0_{d^2\times d^2},I_{d^2},0_{d^2\times d^2}\right)\quad\text{and}\quad\frac{\partial \mathrm{vec}\left(B\right)}{\partial\theta'}=\left(0_{d^2\times d^2},0_{d^2\times d^2},I_{d^2}\right).$$

The asymptotic distribution of $\left[\mathrm{vec}(\widehat{C}_{VT})',\mathrm{vec}(\widehat{A}_{VT})',\mathrm{vec}(\widehat{B}_{VT})'\right]'$ now follows by the delta method using Theorems 4.1 and 4.2.

Appendix B Lemmas

The following section contains the lemmas that were used for establishing consistency and asymptotic normality of the VTE in Section 4. Before we turn to the lemmas we introduce some definitions and useful matrix analysis results for the proofs.

For matrices A, B, C, and D, suppose $ABCD$ is defined and square. Then

$$\mathrm{tr}\left\{ABCD\right\}=\left(\mathrm{vec}\left(D'\right)\right)'\left(C'\otimes A\right)\mathrm{vec}\left(B\right)=\left(\mathrm{vec}\left(D\right)\right)'\left(A\otimes C'\right)\mathrm{vec}\left(B'\right). \quad (B.1)$$

For matrices A and B, if AB is well-defined,

$$\left|\mathrm{tr}\left(AB\right)\right|\leq\left\|A\right\|\left\|B\right\|, \quad (B.2)$$

$$\|AB\| \leq \|A\|_{spec} \|B\|, \quad \|AB\| \leq \|A\| \|B\|_{spec}, \quad \|A+B\|_{spec} \leq \|A\|_{spec} + \|B\|_{spec}. \tag{B.3}$$

If A is $n \times n$, then

$$\|A\|_{spec} \leq \|A\| \leq \sqrt{n} \|A\|_{spec}. \tag{B.4}$$

For an $n \times n$ matrix $A > 0$ with eigenvalues $\xi_1(A), ..., \xi_n(A)$, it holds that

$$\log\det(A) = \sum_{i=1}^{n} \log \xi_i(A) \leq \sum_{i=1}^{n} \xi_i(A) = \mathrm{tr}(A). \tag{B.5}$$

Moreover,

$$\log\det(A) = \log\left(\det(A'A)\right)^{1/2} \leq n \log\left(\rho(A'A)\right)^{1/2} = n \log \|A\|_{spec}, \tag{B.6}$$

where the inequality follows from the fact that $\det(A) \leq \rho(A)^n$. For two square matrices A and B it holds that

$$\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\,\mathrm{tr}(B), \tag{B.7}$$

by Result 11(b) in Lütkepohl (1996) Section 2.4. Consider an $n \times n$ matrix $A \geq 0$ and an $n \times n$ matrix $B > 0$ with eigenvalues $\xi_1(B) \leq \cdots \leq \xi_n(B)$. Let $\xi_1(A+B) \leq \cdots \leq \xi_n(A+B)$ denote the eigenvalues of $(A+B)$, Then

$$\xi_i(A+B) \geq \xi_i(B), \ i = 1, ..., n$$

by Result 4 in Section 5.3.2 of Lütkepohl (1996). Moreover, by Result 4(g) in Section 3.5.1 in Lütkepohl (1996), if $\xi_i(B)$ is an eigenvalue of B, then $1/\xi_i(B)$ is an eigenvalue of B^{-1}. This implies that $0 < \xi_i\left((A+B)^{-1}\right) \leq \xi_i(B^{-1})$, $i = 1, ..., n$, and hence

$$0 < \mathrm{tr}\left[(A+B)^{-1}\right] \leq \mathrm{tr}\left(B^{-1}\right). \tag{B.8}$$

For an $n \times n$ matrix A and an $n \times n$ matrix $B \geq 0$, it holds that

$$\det(A+B) \geq \det(A), \tag{B.9}$$

by Result 11 in Section 4.2.6 of Lütkepohl (1996).

B.1 Lemmas for the proof of consistency

Lemma B.1. *Under Assumptions 1-5, as* $T \to \infty$

$$\sup_{\lambda \in \Theta_\lambda} \left| L_T(\gamma_0, \lambda) - L_{T,h}(\widehat{\gamma}_{VT}, \lambda) \right| \overset{a.s.}{\to} 0 \tag{B.10}$$

where $L_T(\gamma, \lambda)$ *is stated in (A.3) and* $L_{T,h}(\widehat{\gamma}_{VT}, \lambda)$ *is stated in (A.1).*

Proof. We have that

$$\begin{aligned}
& \sup_{\lambda \in \Theta_\lambda} \left| L_T(\gamma_0, \lambda) - L_{T,h}(\widehat{\gamma}_{VT}, \lambda) \right| \\
= & \sup_{\lambda \in \Theta_\lambda} \left| \frac{1}{T} \sum_{t=1}^{T} \left(\log \left\{ \frac{\det [H_t(\gamma_0, \lambda)]}{\det [H_{t,h}(\widehat{\gamma}_{VT}, \lambda)]} \right\} + \operatorname{tr} \left\{ X_t X_t' \left[H_t^{-1}(\gamma_0, \lambda) - H_{t,h}^{-1}(\widehat{\gamma}_{VT}, \lambda) \right] \right\} \right) \right| \\
\leq & \frac{1}{T} \sum_{t=1}^{T} \sup_{\lambda \in \Theta_\lambda} \left| \log \left\{ \frac{\det [H_t(\gamma_0, \lambda)]}{\det [H_{t,h}(\widehat{\gamma}_{VT}, \lambda)]} \right\} \right| \\
& + \frac{1}{T} \sum_{t=1}^{T} \sup_{\lambda \in \Theta_\lambda} \left| \operatorname{tr} \left\{ X_t X_t' \left[H_t^{-1}(\gamma_0, \lambda) - H_{t,h}^{-1}(\widehat{\gamma}_{VT}, \lambda) \right] \right\} \right|,
\end{aligned} \tag{B.11}$$

and we want to show that each of the averages in (B.11) converges to zero almost surely.

By definition of $H_t(\gamma, \lambda)$ in (3.3), $\Gamma - A\Gamma A' - B\Gamma B' > 0$ on Θ and $AX_{t-1}X_{t-1}'A' + BH_{t-1}B' \geq 0$ for all t and for all $\theta \in \Theta$, so applying (B.9) yields

$$\det [H_t(\gamma, \lambda)] \geq \det (\Gamma - A\Gamma A' - B\Gamma B') > 0.$$

In particular, $H_t(\gamma, \lambda)$, and similarly for $H_{t,h}(\gamma, \lambda)$, is invertible for all t and all $\theta \in \Theta$. Moreover,

$$\left\| H_t^{-1}(\gamma, \lambda) \right\| \leq \left\| H_t^{-1/2}(\gamma, \lambda) \right\|^2 = \operatorname{tr} \left[H_t^{-1}(\gamma, \lambda) \right] \leq \operatorname{tr} \left[(\Gamma - A\Gamma A' - B\Gamma B')^{-1} \right],$$

where the second inequality follows by (B.8). By Assumption 4, $\sup_{\lambda \in \Theta_\lambda} \left\| H_t^{-1}(\gamma_0, \lambda) \right\| \leq \sup_{\theta \in \Theta} \left\| H_t^{-1}(\gamma, \lambda) \right\|$, and by (A.6), for T sufficiently large almost surely,

$$\sup_{\lambda \in \Theta_\lambda} \left\| H_{t,h}^{-1}(\widehat{\gamma}_{VT}, \lambda) \right\| \leq \sup_{\theta \in \Theta} \left\| H_{t,h}^{-1}(\gamma, \lambda) \right\|.$$

By Assumption 4

$$\sup_{\theta \in \Theta} \left\| H_t^{-1}(\gamma, \lambda) \right\| \leq \sup_{\theta \in \Theta} \left| \operatorname{tr} \left[(\Gamma - A\Gamma A' - B\Gamma B')^{-1} \right] \right| \leq K, \tag{B.12}$$

and, hence,

$$\sup_{\lambda\in\Theta_\lambda}\left\|H_{t,h}^{-1}(\widehat{\gamma}_{VT},\lambda)\right\|\leq\sup_{\theta\in\Theta}\left\|H_{t,h}^{-1}\left(\gamma,\lambda\right)\right\|\leq K,\ \sup_{\lambda\in\Theta_\lambda}\left\|H_t^{-1}(\gamma_0,\lambda)\right\|\leq\sup_{\theta\in\Theta}\left\|H_t^{-1}\left(\gamma,\lambda\right)\right\|\leq K. \tag{B.13}$$

Next, by simple recursions

$$\begin{aligned}
&\operatorname{vec}\left[H_t\left(\gamma_0,\lambda\right)\right]-\operatorname{vec}\left[H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right] \qquad \text{(B.14)}\\
=\ &\sum_{i=0}^{t-1}\left(B^{\otimes 2}\right)^i\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right)\left(\gamma_0-\widehat{\gamma}_{VT}\right)+\left(B^{\otimes 2}\right)^t\operatorname{vec}\left[H_0\left(\gamma_0,\lambda\right)-h\right].
\end{aligned}$$

As $\rho(A^{\otimes 2}+B^{\otimes 2})<1$ on Θ, it follows from Boussama et al. (2011, Proposition 4.5) that $\rho(B^{\otimes 2})<1$ on Θ. Hence for any i and for some $0<\phi<1$,

$$\sup_{\lambda\in\Theta_\lambda}\left\|\left(B^{\otimes 2}\right)^i\right\|\leq K\phi^i. \tag{B.15}$$

As in Francq et al. (2011, p.644), (B.14), the compactness of Θ, (A.6), and (B.15) imply that as $T\to\infty$

$$\sup_{\lambda\in\Theta_\lambda}\left\|\operatorname{vec}\left[H_t\left(\gamma_0,\lambda\right)\right]-\operatorname{vec}\left[H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]\right\|\leq K\phi^t+o\left(1\right)\quad\text{a.s.} \tag{B.16}$$

Considering the terms in (B.11), we have that for T sufficiently large

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left|\log\left\{\frac{\det\left[H_t\left(\gamma_0,\lambda\right)\right]}{\det\left[H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]}\right\}\right|\\
=\ &\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left|\log\det\left[H_t\left(\gamma_0,\lambda\right)H_{t,h}^{-1}\left(\widehat{\gamma}_{VT},\lambda\right)\right]\right|\\
=\ &\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left|\log\det\left\{I_d+\left[H_t(\gamma_0,\lambda)-H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]H_{t,h}^{-1}\left(\widehat{\gamma}_{VT},\lambda\right)\right\}\right|\\
\leq\ &d\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left|\log\left\|I_d+\left[H_t(\gamma_0,\lambda)-H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]H_{t,h}^{-1}\left(\widehat{\gamma}_{VT},\lambda\right)\right\|_{spec}\right|\\
\leq\ &d\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left|\log\left(\left\|I_d\right\|_{spec}+\left\|\left[H_t(\gamma_0,\lambda)-H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]H_{t,h}^{-1}(\widehat{\gamma}_{VT},\lambda)\right\|\right)\right|\\
=\ &d\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left|\log\left(1+\left\|\left[H_t(\gamma_0,\lambda)-H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]H_{t,h}^{-1}\left(\widehat{\gamma}_{VT},\lambda\right)\right\|\right)\right|\\
\leq\ &d\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left\|\left[H_t(\gamma_0,\lambda)-H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right]H_{t,h}^{-1}\left(\widehat{\gamma}_{VT},\lambda\right)\right\|\\
\leq\ &dK\frac{1}{T}\sum_{t=1}^T\sup_{\lambda\in\Theta_\lambda}\left\|H_t(\gamma_0,\lambda)-H_{t,h}\left(\widehat{\gamma}_{VT},\lambda\right)\right\|,
\end{aligned}$$

where the first inequality follows from (B.6), the second from (B.3) and (B.4), and the third follows from the fact that $\log(x) \leq x-1$ for $x \geq 1$. Likewise,

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\sup_{\lambda\in\Theta_\lambda}\left|\operatorname{tr}\left\{X_tX_t'\left[H_t^{-1}(\gamma_0,\lambda)-H_{t,h}^{-1}(\widehat{\gamma}_{VT},\lambda)\right]\right\}\right| \\
= & \frac{1}{T}\sum_{t=1}^{T}\sup_{\lambda\in\Theta_\lambda}\left|\operatorname{tr}\left\{H_{t,h}^{-1}(\widehat{\gamma}_{VT},\lambda)\left[H_{t,h}(\widehat{\gamma}_{VT},\lambda)-H_t(\gamma_0,\lambda)\right]H_t^{-1}(\gamma_0,\lambda)X_tX_t'\right\}\right| \\
\leq & \frac{1}{T}\sum_{t=1}^{T}\sup_{\lambda\in\Theta_\lambda}\left\|H_{t,h}^{-1}(\widehat{\gamma}_{VT},\lambda)\right\|\left\|H_{t,h}(\widehat{\gamma}_{VT},\lambda)-H_t(\gamma_0,\lambda)\right\|\left\|H_t^{-1}(\gamma_0,\lambda)\right\|\left\|X_tX_t'\right\| \\
\leq & K\frac{1}{T}\sum_{t=1}^{T}\sup_{\lambda\in\Theta_\lambda}\left\|H_{t,h}(\widehat{\gamma}_{VT},\lambda)-H_t(\gamma_0,\lambda)\right\|\left\|X_t\right\|^2,
\end{aligned}$$

where the inequalities follow by (B.2) and (B.13) respectively. By (B.16) we conclude that

$$\sup_{\lambda\in\Theta_\lambda}\left|L_T(\gamma_0,\lambda)-L_{T,h}(\widehat{\gamma}_{VT},\lambda)\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\phi^t+K\frac{1}{T}\sum_{t=1}^{T}\phi^t\left\|X_t\right\|^2+o(1) \quad \text{a.s.}$$

For any $\varepsilon>0$,

$$\sum_{t=1}^{\infty}P\left(\phi^t\left\|X_t\right\|^2>\varepsilon\right)\leq\sum_{t=1}^{\infty}\frac{\phi^tE\left\|X_t\right\|^2}{\varepsilon}<\infty,$$

which follows by Markov's inequality and since $E\left\|X_t\right\|^2<\infty$. By the Borel-Cantelli lemma $\phi^t\left\|X_t\right\|^2 \overset{a.s.}{\to} 0$ as $t\to\infty$. It now follows by Cesàro's mean theorem that

$$\frac{1}{T}\sum_{t=1}^{T}\phi^t\left\|X_t\right\|^2 \overset{a.s.}{\to} 0,$$

and we conclude that (B.10) holds. □

Lemma B.2. *Under Assumptions 1-5, as $T\to\infty$*

$$\sup_{\theta\in\Theta}\left|L_T(\gamma,\lambda)-E\left[l_t(\gamma,\lambda)\right]\right| \overset{a.s.}{\to} 0$$

where $L_T(\theta)$ is the log-likelihood and $l_t(\theta)$ is the log-likelihood contribution (at time t) stated in (A.3) and (A.4), respectively.

Proof. The result follows by Lemma B.4 and the Uniform Law of Large Numbers for stationary ergodic processes, see White (1994, Theorem A.2.2). □

Lemma B.3. *Under Assumptions 1-5, for $l_t(\gamma,\lambda)$ defined in (A.4),*

$$E\left|l_t(\gamma_0,\lambda_0)\right|<\infty,$$

and if $\lambda \neq \lambda_0$ *then*

$$E\left[l_t\left(\gamma_0,\lambda\right)\right] > E\left[l_t\left(\gamma_0,\lambda_0\right)\right].$$

Proof. Observe that $E\left|l_t\left(\gamma_0,\lambda_0\right)\right| < \infty$ follows from Lemma B.4.
Following the steps from Francq and Zakoïan (2010, pp.298-299), suppose $\lambda \neq \lambda_0$ and let $\{\xi_{it} : i = 1,..,d\}$ be the (positive) eigenvalues of $H_t\left(\gamma_0,\lambda_0\right) H_t^{-1}\left(\gamma_0,\lambda\right)$ for a fixed t. Note that

$$\begin{aligned} & \operatorname{tr}\left\{X_t X_t'\left[H_t^{-1}\left(\gamma_0,\lambda\right) - H_t^{-1}\left(\gamma_0,\lambda_0\right)\right]\right\} \\ = \; & \operatorname{tr}\left\{\left[H_t^{1/2}\left(\gamma_0,\lambda_0\right) H_t^{-1}\left(\gamma_0,\lambda\right) H_t^{1/2}\left(\gamma_0,\lambda_0\right) - I_d\right] Z_t Z_t'\right\}. \end{aligned}$$

Let $\mathcal{F}_t \coloneqq \sigma\left(X_t, X_{t-1,\ldots}\right)$, the σ-field generated by X_t. By the law of iterated expectations and since Z_t is independent of $\mathcal{F}_t$,

$$\begin{aligned} & E\left(\operatorname{tr}\left\{X_t X_t'\left[H_t^{-1}\left(\gamma_0,\lambda\right) - H_t^{-1}\left(\gamma_0,\lambda_0\right)\right]\right\}\right) \\ = \; & E\left(\operatorname{tr}\left\{\left[H_t^{1/2}\left(\gamma_0,\lambda_0\right) H_t^{-1}\left(\gamma_0,\lambda\right) H_t^{1/2}\left(\gamma_0,\lambda_0\right) - I_d\right]\right\}\right) \\ = \; & E\left(\operatorname{tr}\left\{\left[H_t\left(\gamma_0,\lambda_0\right) H_t^{-1}\left(\gamma_0,\lambda\right) - I_d\right]\right\}\right) = E\left[\sum_{i=1}^{d}\left(\xi_{it} - 1\right)\right]. \end{aligned}$$

Moreover,

$$\begin{aligned} \log\det\left[H_t\left(\gamma_0,\lambda\right) H_t^{-1}\left(\gamma_0,\lambda_0\right)\right] &= -\log\det\left[H_t\left(\gamma_0,\lambda_0\right) H_t^{-1}\left(\gamma_0,\lambda\right)\right] = -\log\prod_{i=1}^{d}\xi_{it} \\ &= -\sum_{i=1}^{d}\log\xi_{it}. \end{aligned}$$

Hence

$$\begin{aligned} & E\left[l_t\left(\gamma_0,\lambda\right)\right] - E\left[l_t\left(\gamma_0,\lambda_0\right)\right] = E\left\{\log\det\left[H_t\left(\gamma_0,\lambda\right) H_t^{-1}\left(\gamma_0,\lambda_0\right)\right]\right\} \\ & + E\left(\operatorname{tr}\left\{X_t X_t'\left[H_t^{-1}\left(\gamma_0,\lambda\right) - H_t^{-1}\left(\gamma_0,\lambda_0\right)\right]\right\}\right) \\ = \; & E\left[\sum_{i=1}^{d}\left(\xi_{it} - 1 - \log\xi_{it}\right)\right] \geq 0 \end{aligned}$$

as $\log(x) \leq x - 1$ for all $x > 0$. Since $\log(x) = x - 1$ if and only if $x = 1$, the inequality is strict unless $\xi_{it} = 1$ for all i almost surely. $e_{it} = 1$ for all i almost surely is equivalent to $H_t\left(\gamma_0,\lambda\right) = H_t\left(\gamma_0,\lambda_0\right)$ almost surely, but this cannot be the case in light of Assumption 5. Hence the inequality must be strict, and we conclude that if $\lambda \neq \lambda_0$ then $E\left[l_t\left(\gamma_0,\lambda\right)\right] > E\left[l_t\left(\gamma_0,\lambda_0\right)\right]$. □

Lemma B.4. *Under Assumptions 1-5,*

$$E \sup_{\theta\in\Theta} |l_t(\gamma,\lambda)| < \infty,$$

where $l_t(\gamma,\lambda)$ is defined in (A.4).

Proof. We note that

$$\begin{aligned} \mathrm{vec}\,[H_t(\gamma,\lambda)] &= \left(I_{d^2} - A^{\otimes 2} - B^{\otimes 2}\right)\gamma + A^{\otimes 2}\mathrm{vec}\left(X_{t-1}X_{t-1}'\right) + B^{\otimes 2}\mathrm{vec}\,[H_{t-1}(\gamma,\lambda)] \\ &= \sum_{i=0}^{\infty}\left(B^{\otimes 2}\right)^i \left[\left(I_{d^2} - A^{\otimes 2} - B^{\otimes 2}\right)\gamma + A^{\otimes 2}\mathrm{vec}\left(X_{t-1-i}X_{t-1-i}'\right)\right], \end{aligned} \tag{B.17}$$

so by the triangle inequality

$$\sup_{\theta\in\Theta}\|\mathrm{vec}\,[H_t(\gamma,\lambda)]\| \le \sum_{i=0}^{\infty}\sup_{\theta\in\Theta}\left\|\left(B^{\otimes 2}\right)^i\left[\left(I_{d^2} - A^{\otimes 2} - B^{\otimes 2}\right)\gamma + A^{\otimes 2}\mathrm{vec}\left(X_{t-1-i}X_{t-1-i}'\right)\right]\right\|.$$

By (B.15),

$$\begin{aligned} &E\left(\sup_{\theta\in\Theta}\left\|\left(B^{\otimes 2}\right)^i\left[\left(I_{d^2} - A^{\otimes 2} - B^{\otimes 2}\right)\gamma + A^{\otimes 2}\mathrm{vec}\left(X_{t-1-i}X_{t-1-i}'\right)\right]\right\|\right) \\ &\le K\phi^i\left(K + KE\|X_t\|^2\right), \end{aligned}$$

and we conclude that

$$E\left[\sup_{\theta\in\Theta}\|H_t(\gamma,\lambda)\|\right] < \infty. \tag{B.18}$$

Now

$$\begin{aligned} E\left[\sup_{\theta\in\Theta}|l_t(\gamma,\lambda)|\right] &= E\left[\sup_{\theta\in\Theta}\left|\log\det[H_t(\gamma,\lambda)] + \mathrm{tr}\left[X_tX_t'H_t^{-1}(\gamma,\lambda)\right]\right|\right] \\ &\le E\left[\sup_{\theta\in\Theta}\left|\mathrm{tr}\,[H_t(\gamma,\lambda)] + \mathrm{tr}\left[X_tX_t'H_t^{-1}(\gamma,\lambda)\right]\right|\right] \\ &\le E\left\{\sup_{\theta\in\Theta}\left|K\left[\|H_t(\gamma,\lambda)\| + \left\|X_tX_t'H_t^{-1}(\gamma,\lambda)\right\|\right]\right|\right\} \\ &\le K\left[E\sup_{\theta\in\Theta}\|H_t(\gamma,\lambda)\|\right] + KE\left[\sup_{\theta\in\Theta}\|X_t\|^2\left\|H_t^{-1}(\gamma,\lambda)\right\|\right] < \infty, \end{aligned}$$

where the first inequality follows from (B.5), the second from (B.2), and the fourth from (B.18), (B.12), and $E\|X_t\|^2 < \infty$. □

B.2 Lemmas for the proof of asymptotic normality

Let θ_i, $i = 1, ..., 3d^2$, denote the i^{th} element of θ. Let $H_{0t} := H_t(\gamma_0, \lambda_0)$.

Lemma B.5. *Under Assumptions 1-7,* $E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right|\right] < \infty$ *for all* $i, j = 1, ..., 3d^2$, *where* $l_t(\gamma,\lambda)$ *is defined in (A.4).*

Proof. Note that

$$\begin{aligned}\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j} = & \operatorname{tr}\left(H_t^{-1}(\gamma,\lambda)\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right) \\ & -\operatorname{tr}\left(H_t^{-1}(\gamma,\lambda)\frac{\partial H_t(\gamma,\lambda)}{\partial\theta_j}H_t^{-1}(\gamma,\lambda)\frac{\partial H_t(\gamma,\lambda)}{\partial\theta_i}\right) \\ & +2\operatorname{tr}\left(H_t^{-1}(\gamma,\lambda)X_tX_t'H_t^{-1}(\gamma,\lambda)\frac{\partial H_t(\gamma,\lambda)}{\partial\theta_j}H_t^{-1}(\gamma,\lambda)\frac{\partial H_t(\gamma,\lambda)}{\partial\theta_i}\right) \\ & -\operatorname{tr}\left(H_t^{-1}(\gamma,\lambda)X_tX_t'H_t^{-1}(\gamma,\lambda)\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right). \end{aligned} \quad \text{(B.19)}$$

By (B.17), Minkowski's inequality, and Assumption 6,

$$E\left(\sup_{\theta\in\Theta}\|H_t(\gamma,\lambda)\|\right)^3 < \infty. \quad \text{(B.20)}$$

Moreover, using Minkowski's inequality repeatedly (see also Hafner and Preminger, 2009b, Proof of Lemma 3), and Assumption 6 one can show that

$$E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial H_t(\gamma,\lambda)}{\partial\theta_i}\right\|\right)^3 < \infty \text{ and } E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right\|\right)^3 < \infty. \quad \text{(B.21)}$$

By (B.2), (B.12), Hölder's inequality, and (B.21),

$$\begin{aligned} & E\left[\sup_{\theta\in\Theta}\left|\operatorname{tr}\left(H_t^{-1}X_tX_t'H_t^{-1}\frac{\partial H_t}{\partial\theta_j}H_t^{-1}\frac{\partial H_t}{\partial\theta_i}\right)\right|\right] \\ \leq\ & K\left[E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial H_t}{\partial\theta_j}\right\|\right)^3\right]^{1/3}\left[E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial H_t}{\partial\theta_i}\right\|\right)^3\right]^{1/3}\left[E\|X_t\|^6\right]^{1/3} < \infty. \end{aligned}$$

By similar arguments we conclude that $E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right|\right] < \infty$ for all $i, j = 1, ..., 3d^2$. □

Lemma B.6. *Under Assumptions 1-7* $\sup_{\theta\in\Theta}\left|\frac{\partial^2 L_T(\gamma,\lambda)}{\partial\theta_i\partial\theta_j} - E\left(\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right)\right| \overset{a.s.}{\to} 0$ *for all* $i, j = 1, ..., 3d^2$, *where* $L_T(\gamma,\lambda)$ *and* $l_t(\gamma,\lambda)$ *are defined in (A.3) and (A.4), respectively.*

Proof. Note that $\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}$ is a function of $(X_t, X_{t-1,\ldots})$ and thereby strictly stationary and ergodic. Hence the result follows by Lemma B.5 and the Uniform Law of Large Numbers for stationary ergodic processes, see Theorem A.2.2 of White (1994). □

Lemma B.7. *Under Assumptions 1-7, J_0 defined in (A.8) is non-singular.*

Proof. We prove this lemma arguing in line with Francq and Zakoïan (2010, pp.303-305), see also Comte and Lieberman (2003, pp.77-78). By definition

$$J_0 = E\left[\frac{\partial^2 l_t(\gamma_0,\lambda_0)}{\partial\lambda\partial\lambda'}\right],$$

with $\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\lambda_i\partial\lambda_j}$ given by (B.19). Hence,

$$\begin{aligned} E\left[\left.\frac{\partial^2 l_t(\gamma_0,\lambda_0)}{\partial\lambda_i\partial\lambda_j}\right|\mathcal{F}_{t-1}\right] &= \operatorname{tr}\left(H_{0t}^{-1}\frac{\partial H_{0t}}{\partial\lambda_j}H_{0t}^{-1}\frac{\partial H_{0t}}{\partial\lambda_i}\right) \\ &= h_{tj}'h_{ti}, \end{aligned} \tag{B.22}$$

where

$$h_{ti} := \left(H_{0t}^{-1/2}\right)^{\otimes 2}k_{ti}, \text{ and } k_{ti} := \operatorname{vec}\left(\frac{\partial H_{0t}}{\partial\lambda_i}\right),$$

noting that $\frac{\partial H_{0t}}{\partial\lambda_i}$ is symmetric. We now define the $d^2\times 2d^2$ matrices

$$h_t := (h_{t1},\cdots,h_{t2d^2}) \text{ and } k_t := (k_{t1},\cdots,k_{t2d^2}).$$

Let $\mathcal{H}_t := \left(H_{0t}^{-1/2}\right)^{\otimes 2}$, and note that $h_t = \mathcal{H}_t k_t$ and $J_0 = E[h_t'h_t]$. Suppose J_0 is singular. Then there exists a non-zero $c\in\mathbb{R}^{2d^2}$ such that $c'J_0c = E[c'h_t'h_tc] = 0$. As $c'h_t'h_tc \geq 0$, then almost surely

$$c'h_t'h_tc = c'k_t'\mathcal{H}_t^2 k_tc = 0.$$

Since $\mathcal{H}_t^2$ is positive definite a.s.,

$$k_tc = \sum_{i=1}^{d^2} c_i\frac{\partial}{\partial\lambda_i}\operatorname{vec}(H_{0t}) = 0 \quad \text{a.s. for all } t. \tag{B.23}$$

Let $\omega = (I_{d^2} - A^{\otimes 2} - B^{\otimes 2})\gamma$, then (B.23) gives

$$\widetilde{\omega} + \widetilde{A}\operatorname{vec}\left(X_{t-1}X_{t-1}'\right) + \widetilde{B}\operatorname{vec}(H_{0t-1}) + B^{\otimes 2}\sum_{i=1}^{2d^2} c_i\frac{\partial}{\partial\lambda_i}\operatorname{vec}(H_{0t-1}) = 0 \quad \text{a.s.} \tag{B.24}$$

where

$$\widetilde{\omega} := \sum_{i=1}^{2d^2} c_i \frac{\partial}{\partial \lambda_i} \omega \bigg|_{\theta=\theta_0}, \; \widetilde{A} := \sum_{i=1}^{d^2} c_i \frac{\partial}{\partial \lambda_i} A^{\otimes 2} \bigg|_{\theta=\theta_0}, \; \widetilde{B} := \sum_{i=d^2}^{2d^2} c_i \frac{\partial}{\partial \lambda_i} B^{\otimes 2} \bigg|_{\theta=\theta_0}.$$

By (B.23), (B.24) reduces to

$$\widetilde{\omega} + \widetilde{A} \text{vec}\left(X_{t-1} X_{t-1}'\right) + \widetilde{B} \text{vec}\left(H_{0t-1}\right) = 0 \quad \text{a.s.} \tag{B.25}$$

Subtracting (B.25) from $\text{vec}\left(H_{0t}\right)$ yields

$$\text{vec}\left(H_{0t}\right) = \left(\omega_0 - \widetilde{\omega}\right) + \left(A_0^{\otimes 2} - \widetilde{A}\right) \text{vec}\left(X_{t-1} X_{t-1}'\right) + \left(B_0^{\otimes 2} - \widetilde{B}\right) \text{vec}\left(H_{0t-1}\right).$$

Since $c \neq 0$, we have found another representation of $\text{vec}\left(H_{0t}\right)$, which contradicts Assumption 5 that ensures that $\text{vec}\left(H_{0t}\right)$ has a unique representation. Hence J_0 must be non-singular. □

Lemma B.8. *Under Assumptions 1-7, as $T \to \infty$,*

$$\sqrt{T} \begin{pmatrix} \widehat{\gamma}_{VT} - \gamma_0 \\ \frac{\partial L_T(\gamma_0, \lambda_0)}{\partial \lambda} \end{pmatrix} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t\left(\gamma_0, \lambda_0\right) \text{vec}\left(Z_t Z_t' - I_d\right) + o_P(1) \tag{B.26}$$

where $L_T(\gamma, \lambda)$ is defined in (A.3),

$$Y_t\left(\gamma_0, \lambda_0\right) := \begin{pmatrix} \left(I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right)^{-1} \left(I_{d^2} - B_0^{\otimes 2}\right) \left(H_{0t}^{1/2}\right)^{\otimes 2} \\ \left[-\sum_{i=0}^{\infty} \left(B_0^{\otimes 2}\right)^i M_{t-1-i}\left(\gamma_0, \lambda_0\right)\right]' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \\ \left[-\sum_{i=0}^{\infty} \left(B_0^{\otimes 2}\right)^i \widetilde{M}_{t-1-i}\left(\gamma_0, \lambda_0\right)\right]' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \end{pmatrix}, \tag{B.27}$$

with

$$M_t(\gamma, \lambda) := \left(\left\{\left[A\left(X_t X_t' - \Gamma\right)\right] \otimes I_d\right\} + \left\{I_d \otimes \left[A\left(X_t X_t' - \Gamma\right)\right]\right\} C_{dd}\right), \tag{B.28}$$

and

$$\widetilde{M}_t(\gamma, \lambda) := \left[\left(\left\{B\left[H_t(\gamma, \lambda) - \Gamma\right]\right\} \otimes I_d\right) + \left(I_d \otimes \left\{B\left[H_t(\gamma, \lambda) - \Gamma\right]\right\}\right) C_{dd}\right]. \tag{B.29}$$

Proof. In the next we make use of matrix differentials and apply the following notation. Let f_t be a function of the non-stochastic matrices A and B. Then $\text{d}\left\{f_t\left(A_0, B_0\right), \text{d}A\right\}$ denotes the first-order differential of f_t in the direction $\text{d}A$ and evaluated at (A_0, B_0). One can identify the Jacobian from the first-order differential, see e.g. Magnus and Neudecker (2007, p.199), from the fact that if, $\mathrm{d}\left\{f_t\left(A_0, B_0\right), \mathrm{d}A\right\} = \mathrm{vec}\left(D_t\right)' \mathrm{vec}\left(\mathrm{d}A\right)$ for some matrix D_t, then the first-derivative of f_t with respect to $\mathrm{vec}\left(A\right)$ and evaluated at (A_0, B_0) is $\partial f_t\left(A_0, B_0\right) / \partial \mathrm{vec}\left(A\right) = \mathrm{vec}\left(D_t\right)$.

The first-order differential of the log-likelihood contribution at time t with respect to A and evaluated in (γ_0, λ_0) is given by

$$\begin{aligned}
\mathrm{d}\left\{l_t\left(\gamma_0, \lambda_0\right), \mathrm{d}A\right\} &= \mathrm{tr}\left\{H_{0t}^{-1/2}\left[\mathrm{d}\left\{H_t\left(\gamma_0, \lambda_0\right), \mathrm{d}A\right\}\right] H_{0t}^{-1/2}\right\} \\
&\quad - \mathrm{tr}\left\{H_{0t}^{-1/2} X_t X_t' H_{0t}^{-1/2} H_{0t}^{-1/2}\left[\mathrm{d}\left\{H_t\left(\gamma_0, \lambda_0\right), \mathrm{d}A\right\}\right] H_{0t}^{-1/2}\right\} \\
&= \mathrm{tr}\left\{H_{0t}^{-1/2}\left[\mathrm{d}\left\{H_t\left(\gamma_0, \lambda_0\right), \mathrm{d}A\right\}\right] H_{0t}^{-1/2}\right\} \\
&\quad - \mathrm{tr}\left\{Z_t Z_t' H_{0t}^{-1/2}\left[\mathrm{d}\left\{H_t\left(\gamma_0, \lambda_0\right), \mathrm{d}A\right\}\right] H_{0t}^{-1/2}\right\} \\
&= \mathrm{vec}(I_d - Z_t Z_t')'\left(H_{0t}^{-1/2}\right)^{\otimes 2} \mathrm{vec}\left[\mathrm{d}\left\{H_t\left(\gamma_0, \lambda_0\right), \mathrm{d}A\right\}\right],
\end{aligned}$$

where the last equality follows by (B.1). Likewise,

$$\mathrm{d}\left\{l_t\left(\gamma_0, \lambda_0\right), \mathrm{d}B\right\} = \mathrm{vec}(I_d - Z_t Z_t')'\left(H_{0t}^{-1/2}\right)^{\otimes 2} \mathrm{vec}\left[\mathrm{d}\left\{H_t\left(\gamma_0, \lambda_0\right), \mathrm{d}B\right\}\right].$$

Note that

$$H_t\left(\gamma, \lambda\right) = \Gamma + A\left(X_{t-1}X_{t-1}' - \Gamma\right)A' - B\left[H_{t-1}\left(\gamma, \lambda\right) - \Gamma\right]B'.$$

The first-order differential of $H_t\left(\gamma, \lambda\right)$ with respect to A is

$$\begin{aligned}
\mathrm{d}\left\{H_t\left(\gamma, \lambda\right), \mathrm{d}A\right\} &= (\mathrm{d}A)\left(X_{t-1}X_{t-1}' - \Gamma\right)A' \\
&\quad + A\left(X_{t-1}X_{t-1}' - \Gamma\right)(\mathrm{d}A)' + B\left[\mathrm{d}\left\{H_{t-1}\left(\gamma, \lambda\right), \mathrm{d}A\right\}\right]B',
\end{aligned}$$

implying directly

$$\begin{aligned}
\mathrm{vec}\left[\mathrm{d}\left\{H_t\left(\gamma, \lambda\right), \mathrm{d}A\right\}\right] &= B^{\otimes 2}\mathrm{vec}\left[\mathrm{d}\left\{H_{t-1}\left(\gamma, \lambda\right), \mathrm{d}A\right\}\right] \\
&\quad + \mathrm{vec}\left[(\mathrm{d}A)\left(X_{t-1}X_{t-1}' - \Gamma\right)A' + A\left(X_{t-1}X_{t-1}' - \Gamma\right)(\mathrm{d}A)'\right].
\end{aligned} \tag{B.30}$$

We note that

$$
\begin{aligned}
&\operatorname{vec}\left[(\mathrm{d}A)\left(X_{t-1}X_{t-1}'-\Gamma\right)A'+A\left(X_{t-1}X_{t-1}'-\Gamma\right)(\mathrm{d}A)'\right]\\
&=\operatorname{vec}\left[(\mathrm{d}A)\left(X_{t-1}X_{t-1}'-\Gamma\right)A'\right]+\operatorname{vec}\left[A(X_{t-1}X_{t-1}'-\Gamma)(\mathrm{d}A)'\right]\\
&=\left\{\left[A\left(X_{t-1}X_{t-1}'-\Gamma\right)\right]\otimes I_d\right\}\operatorname{vec}(\mathrm{d}A)+\left\{I_d\otimes\left[A\left(X_{t-1}X_{t-1}'-\Gamma\right)\right]\right\}\operatorname{vec}(\mathrm{d}A')\\
&=\left\{\left[A\left(X_{t-1}X_{t-1}'-\Gamma\right)\right]\otimes I_d\right\}\operatorname{vec}(\mathrm{d}A)+\left\{I_d\otimes\left[A\left(X_{t-1}X_{t-1}'-\Gamma\right)\right]\right\}C_{dd}\operatorname{vec}(\mathrm{d}A)\\
&=\left(\left\{\left[A\left(X_{t-1}X_{t-1}'-\Gamma\right)\right]\otimes I_d\right\}+\left\{I_d\otimes\left[A\left(X_{t-1}X_{t-1}'-\Gamma\right)\right]\right\}C_{dd}\right)\operatorname{vec}(\mathrm{d}A).
\end{aligned}
$$

With $M_t(\gamma,\lambda)$ defined in (B.28), recursions yield

$$\operatorname{vec}\left[\mathrm{d}\left\{H_t(\gamma,\lambda),\mathrm{d}A\right\}\right]=\sum_{i=0}^{\infty}\left(B^{\otimes 2}\right)^i M_{t-1-i}(\gamma,\lambda)\operatorname{vec}(\mathrm{d}A), \tag{B.31}$$

and we conclude that

$$\mathrm{d}\left\{l_t(\gamma_0,\lambda_0),\mathrm{d}A\right\}=\operatorname{vec}(Z_tZ_t'-I_d)'\left(H_{0t}^{-1/2}\right)^{\otimes 2}\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{t-1-i}(\gamma_0,\lambda_0)\right]\operatorname{vec}(\mathrm{d}A).$$

We now have that the first derivative of the log-likelihood function with respect to $\operatorname{vec}(A)$ and evaluated at $\theta=\theta_0$ is given by

$$\frac{\partial L_T(\gamma_0,\lambda_0)}{\partial\operatorname{vec}(A)}=\frac{1}{T}\sum_{t=1}^{T}\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{t-1-i}(\gamma_0,\lambda_0)\right]'\left(H_{0t}^{-1/2}\right)^{\otimes 2}\operatorname{vec}(Z_tZ_t'-I_d).$$

By similar arguments

$$\operatorname{vec}\left[\mathrm{d}\left\{H_t(\gamma,\lambda),\mathrm{d}B\right\}\right]=\sum_{i=0}^{\infty}\left(B^{\otimes 2}\right)^i \widetilde{M}_{t-1-i}(\gamma,\lambda)\operatorname{vec}(\mathrm{d}B), \tag{B.32}$$

and

$$\frac{\partial L_T(\gamma_0,\lambda_0)}{\partial\operatorname{vec}(B)}=\frac{1}{T}\sum_{t=1}^{T}\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i \widetilde{M}_{t-1-i}(\gamma_0,\lambda_0)\right]'\left(H_{0t}^{-1/2}\right)^{\otimes 2}\operatorname{vec}(Z_tZ_t'-I_d),$$

with $\widetilde{M}_t(\gamma,\lambda)$ defined in (B.29).

Consider the sample covariance matrix on vec form:

$$\widehat{\gamma}_{VT}=\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\operatorname{vec}\left(Z_tZ_t'-I_d\right)+\operatorname{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t}\right). \tag{B.33}$$

Moreover,

$$
\begin{aligned}
\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t}\right) &= \left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)\gamma_0 \\
&\quad +A_0^{\otimes 2}\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X_{t-1}'\right)+B_0^{\otimes 2}\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t-1}\right) \\
&= \left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)\gamma_0+A_0^{\otimes 2}\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}X_tX_t'\right)+B_0^{\otimes 2}\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t}\right) \\
&\quad +A_0^{\otimes 2}\frac{1}{T}\text{vec}\left(X_0X_0'-X_TX_T'\right)+B_0^{\otimes 2}\frac{1}{T}\text{vec}\left(H_{00}-H_{0T}\right),
\end{aligned}
$$

and collecting terms

$$
\begin{aligned}
\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t}\right) &= \left(I_{d^2}-B_0^{\otimes 2}\right)^{-1}\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)\gamma_0+\left[I_{d^2}-B_0^{\otimes 2}\right]^{-1}A_0^{\otimes 2}\widehat{\gamma}_{VT} \qquad \text{(B.34)} \\
&\quad +\left(I_{d^2}-B_0^{\otimes 2}\right)^{-1}\left[A_0^{\otimes 2}\frac{1}{T}\text{vec}\left(X_0X_0'-X_TX_T'\right)+B_0^{\otimes 2}\frac{1}{T}\text{vec}\left(H_{00}-H_{0T}\right)\right].
\end{aligned}
$$

Note that since $\rho(B_0^{\otimes 2})<1$, $(I_{d^2}-B_0^{\otimes 2})$ is invertible. Next, inserting (B.33) in (B.34) and isolating $\widehat{\gamma}_{VT}$ yields

$$
\begin{aligned}
\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)\widehat{\gamma}_{VT} &= \left(I_{d^2}-B_0^{\otimes 2}\right)\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\text{vec}\left(Z_tZ_t'-I_d\right)+\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)\gamma_0 \\
&\quad +\left[A_0^{\otimes 2}\frac{1}{T}\text{vec}\left(X_0X_0'-X_TX_T'\right)+B_0^{\otimes 2}\frac{1}{T}\text{vec}\left(H_{00}-H_{0T}\right)\right].
\end{aligned}
$$

Hence

$$
\begin{aligned}
\widehat{\gamma}_{VT}-\gamma_0 &= \left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)^{-1}\left(I_{d^2}-B_0^{\otimes 2}\right)\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\text{vec}\left(Z_tZ_t'-I_d\right) \\
&\quad +\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)^{-1}\left[A_0^{\otimes 2}\frac{1}{T}\text{vec}\left(X_0X_0'-X_TX_T'\right)+B_0^{\otimes 2}\frac{1}{T}\text{vec}\left(H_{00}-H_{0T}\right)\right].
\end{aligned}
$$

For any $\varepsilon>0$, by Markov's inequality,

$$
P\left(\left\|A_0^{\otimes 2}\frac{1}{\sqrt{T}}\text{vec}\left(X_0X_0'-X_TX_T'\right)+B_0^{\otimes 2}\frac{1}{\sqrt{T}}\text{vec}\left(H_{00}-H_{0T}\right)\right\|>\varepsilon\right)\leq\frac{KE\left\|X_t\right\|^2}{\sqrt{T}\varepsilon}\to 0
$$

as $T\to\infty$, which yields

$$
\widehat{\gamma}_{VT}-\gamma_0=\left[I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right]^{-1}\left(I_{d^2}-B_0^{\otimes 2}\right)\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\text{vec}\left(Z_tZ_t'-I_d\right)+o_P\left(T^{-1/2}\right). \qquad \text{(B.35)}
$$

We conclude that (B.26) holds. □

Lemma B.9. *Under Assumptions 1-7*

$$E\left\|Y_t(\gamma_0,\lambda_0)\operatorname{vec}(Z_tZ_t'-I_d)\right\|^2<\infty,$$

where $Y_t(\gamma_0,\lambda_0)$is given by (B.27).

Proof. By definition

$$Y_t(\gamma_0,\lambda_0)=\begin{pmatrix}\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)^{-1}\left(I_{d^2}-B_0^{\otimes 2}\right)\left(H_{0t}^{1/2}\right)^{\otimes 2}\\ \left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{t-1-i}(\gamma_0,\lambda_0)\right]'\left(H_{0t}^{-1/2}\right)^{\otimes 2}\\ \left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i \widetilde{M}_{t-1-i}(\gamma_0,\lambda_0)\right]'\left(H_{0t}^{-1/2}\right)^{\otimes 2}\end{pmatrix},$$

where M_t and $\widetilde{M}_t$ are defined in (B.28) and (B.29), respectively. Define

$$\varepsilon_t:=\operatorname{vec}(Z_tZ_t'-I_d),\ \eta_t^M:=\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i M_{t-1-i}(\gamma_0,\lambda_0)\right],$$
$$\eta_t^{\widetilde{M}}:=\left[-\sum_{i=0}^{\infty}\left(B_0^{\otimes 2}\right)^i \widetilde{M}_{t-1-i}(\gamma_0,\lambda_0)\right],\quad\text{and}\quad \Phi:=\left(I_{d^2}-A_0^{\otimes 2}-B_0^{\otimes 2}\right)^{-1}\left(I_{d^2}-B_0^{\otimes 2}\right),$$

and observe that

$$Y_t(\gamma_0,\lambda_0)\operatorname{vec}(Z_tZ_t'-I_d)\left[\operatorname{vec}(Z_tZ_t'-I_d)\right]'Y_t(\gamma_0,\lambda_0)'=\begin{pmatrix}\mathcal{A}_t & \mathcal{B}_t & \mathcal{C}_t\\ \mathcal{B}_t' & \mathcal{D}_t & \mathcal{E}_t\\ \mathcal{C}_t' & \mathcal{E}_t' & \mathcal{G}_t\end{pmatrix}$$

where

$$\begin{aligned}\mathcal{A}_t&:=\Phi(H_{0t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(H_{0t}^{1/2})^{\otimes 2}\Phi',\\ \mathcal{B}_t&:=\Phi(H_{0t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(H_{0t}^{-1/2})^{\otimes 2}\eta_t^M,\\ \mathcal{C}_t&:=\Phi(H_{0t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(H_{0t}^{-1/2})^{\otimes 2}\eta_t^{\widetilde{M}},\\ \mathcal{D}_t&:=\eta_t^{M\prime}(H_{0t}^{-1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(H_{0t}^{-1/2})^{\otimes 2}\eta_t^M,\\ \mathcal{E}_t&:=\eta_t^{M\prime}(H_{0t}^{-1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(H_{0t}^{-1/2})^{\otimes 2}\eta_t^{\widetilde{M}},\\ \mathcal{G}_t&:=\eta_t^{\widetilde{M}\prime}(H_{0t}^{-1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(H_{0t}^{-1/2})^{\otimes 2}\eta_t^{\widetilde{M}}.\end{aligned}\tag{B.36}$$

Hence $Y_t(\gamma_0,\lambda_0)\operatorname{vec}(Z_tZ_t'-I_d)$ is square-integrable if $E\|\mathcal{A}_t\|$, $E\|\mathcal{B}_t\|$, $E\|\mathcal{C}_t\|$, $E\|\mathcal{D}_t\|$, $E\|\mathcal{E}_t\|$, and $E\|\mathcal{G}_t\|$ are finite.

Using Minkowski's inequality,

$$E\left\|\eta_t^M\right\|^3\le\left\{\sum_{i=1}^{\infty}\phi^i\left(K+KE\|X_t\|^6\right)^{1/3}\right\}^3<\infty.\tag{B.37}$$

Likewise, by Minkowski's inequality and (B.20)

$$E\left\|\eta_t^{\widetilde{M}}\right\|^3 \leq \left\{\sum_{i=1}^{\infty}\phi^i\left(K+KE\left\|H_{0t}\right\|^3\right)^{1/3}\right\}^3 < \infty. \tag{B.38}$$

We note that

$$E\left\|\mathcal{A}_t\right\| \leq KE\left\|\left(H_{0t}^{1/2}\right)^{\otimes 2}\right\|^2 E\left\|\varepsilon_t\right\|^2$$

by the independence between Z_t and H_{0t}. Moreover,

$$E\left\|\left(H_{0t}^{1/2}\right)^{\otimes 2}\right\|^2 = E\left|\operatorname{tr}\left[\left(H_{0t}^{1/2}\right)^{\otimes 2}\left(H_{0t}^{1/2}\right)^{\otimes 2}\right]\right| = E\operatorname{tr}^2\left(H_{0t}\right) \leq KE\left\|H_{0t}\right\|^2 < \infty,$$

by (B.7) and (B.20). Moreover,

$$E\left\|\varepsilon_t\right\|^2 \leq E\left\|Z_t\right\|^4 + K < \infty,$$

as $E\left\|Z_t\right\|^4 \leq KE\left\|X_t\right\|^4$. Hence $E\left\|\mathcal{A}_t\right\| < \infty$. Next,

$$\begin{aligned} E\left\|\mathcal{B}_t\right\| &\leq KE\left(\left\|\left(H_{0t}^{1/2}\right)^{\otimes 2}\right\|\left\|\left(H_{0t}^{-1/2}\right)^{\otimes 2}\right\|\left\|\eta_t^M\right\|\left\|\varepsilon_t\right\|^2\right) \\ &\leq KE\left(\left\|\left(H_{0t}^{1/2}\right)^{\otimes 2}\right\|\left\|\left(H_{0t}^{-1/2}\right)^{\otimes 2}\right\|\left\|\eta_t^M\right\|\right)E\left\|\varepsilon_t\right\|^2, \end{aligned}$$

where the second inequality follows by the fact that ε_t and $\mathcal{F}_{t-1}$ are independent. Note that

$$\left\|\left(H_{0t}^{-1/2}\right)^{\otimes 2}\right\| = \sqrt{\operatorname{tr}\left(H_{0t}^{-1}\otimes H_{0t}^{-1}\right)} = \operatorname{tr}\left(H_{0t}^{-1}\right) \leq K\left\|H_{0t}^{-1}\right\| \leq K,$$

by (B.7) and (B.12). Hence by Hölder's inequality and (B.37)

$$\begin{aligned} E\left\|\mathcal{B}_t\right\| &\leq KE\left[\left\|\left(H_{0t}^{1/2}\right)^{\otimes 2}\right\|\left\|\eta_t^M\right\|\right]E\left[\left\|\varepsilon_t\right\|^2\right] \\ &\leq K\left\{E\left\|\left(H_{0t}^{1/2}\right)^{\otimes 2}\right\|^2\right\}^{1/2}\left\{E\left\|\eta_t^M\right\|^2\right\}^{1/2}E\left[\left\|\varepsilon_t\right\|^2\right] < \infty. \end{aligned}$$

By similar arguments $E\left\|\mathcal{C}_t\right\|$, $E\left\|\mathcal{D}_t\right\|$, $E\left\|\mathcal{E}_t\right\|$, and $E\left\|\mathcal{G}_t\right\|$ are finite. □

Lemma B.10. *Under Assumptions 1-7, as $T\to\infty$*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t\left(\gamma_0,\lambda_0\right)\operatorname{vec}\left(Z_tZ_t'-I_d\right) \overset{D}{\to} N\left(0,\Omega_0\right), \tag{B.39}$$

where

$$\Omega_0 := E\left\{Y_t\left(\gamma_0,\lambda_0\right)\mathrm{vec}\left(Z_tZ_t'-I_d\right)\left[\mathrm{vec}\left(Z_tZ_t'-I_d\right)\right]'Y_t\left(\gamma_0,\lambda_0\right)'\right\}, \tag{B.40}$$

and $Y_t\left(\gamma_0,\lambda_0\right)$ *is given by (B.27).*

Proof. Since $Y_t\left(\gamma_0,\lambda_0\right)$ is $\mathcal{F}_{t-1}$-measurable and $\mathrm{vec}\left(Z_tZ_t'-I_d\right)$ and $\mathcal{F}_{t-1}$ are independent, $\left\{Y_t\left(\gamma_0,\lambda_0\right)\mathrm{vec}\left(Z_tZ_t'-I_d\right),\mathcal{F}_t\right\}$ is an ergodic martingale difference sequence. Moreover, from Lemma B.9 the sequence is square-integrable, and the regularity conditions of Brown (1971) are satisfied by the ergodic theorem. This implies that (B.39) holds. □

Lemma B.11. *Under Assumptions 1-7, as* $T\to\infty$,

$$\left|\sqrt{T}\left[\frac{\partial L_T\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial L_{T,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right]\right|\xrightarrow{P}0, \tag{B.41}$$

for $i=1,...,2d^2$, *and*

$$\sup_{\theta\in\Theta}\left|\frac{\partial^2L_T\left(\gamma,\lambda\right)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2L_{T,h}\left(\gamma,\lambda\right)}{\partial\theta_i\partial\theta_j}\right|\xrightarrow{a.s.}0 \tag{B.42}$$

for $i,j=1,...,3d^2$, *where* $L_T\left(\gamma,\lambda\right)$ *and* $L_{T,h}\left(\gamma,\lambda\right)$ *are defined in (A.3) and (A.1), respectively.*

Proof. We proceed as in the proof of Lemma 4 in Hafner and Preminger (2009a). First, we establish that for some $r>0$,

$$E\left|\frac{\partial l_t\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial l_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right|^r=O\left(t\rho^t\right). \tag{B.43}$$

By Hafner and Preminger (2009a, (B.32)), for $0<r<1$,

$$\begin{aligned}
& E\left|\frac{\partial l_t\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial l_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right|^r \\
\leq\; & KE\left[\left(K+K\left\|X_t\right\|^{2r}\right)\left\|\frac{\partial H_t\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial H_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right\|^r\right] \\
& +KE\left[\left(K+K\left\|X_t\right\|^{2r}\right)\left\|\frac{\partial H_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}H_{t,h}\left(\gamma_0,\lambda_0\right)^{-1}\right\|\left\|H_{t,h}\left(\gamma_0,\lambda_0\right)-H_t\left(\gamma_0,\lambda_0\right)\right\|^r\right] \\
& +KE\left[\left\|X_t\right\|^{2r}\left\|H_t\left(\gamma_0,\lambda_0\right)-H_{t,h}\left(\gamma_0,\lambda_0\right)\right\|^r\left\|\frac{\partial H_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}H_{t,h}\left(\gamma_0,\lambda_0\right)^{-1}\right\|^r\right].
\end{aligned}$$

With $r=1/4$, it follows by Assumption 6, (B.13), and Hölder's inequality, that it is sufficient to establish that

$$E\left\|\text{vec}\left[H_{t,h}\left(\gamma_0,\lambda_0\right)-H_t\left(\gamma_0,\lambda_0\right)\right]\right\|=O\left(\rho^t\right), \tag{B.44}$$

$$E\left\|\text{vec}\left[\frac{\partial H_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial H_t\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right]\right\|=O\left(t\rho^t\right), \tag{B.45}$$

and

$$E\left\|\frac{\partial H_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right\|<\infty. \tag{B.46}$$

By iterating (A.5), (B.20), and since h is constant, it follows that

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[H_{t,h}\left(\gamma,\lambda\right)-H_t\left(\gamma,\lambda\right)\right]\right\|=O\left(\rho^t\right), \tag{B.47}$$

and thereby that (B.44) holds. Likewise, iterating (A.5) yields

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[\frac{\partial H_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_i}-\frac{\partial H_t\left(\gamma,\lambda\right)}{\partial\theta_i}\right]\right\|=E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_i}\left\{\left(B\otimes B\right)^t\text{vec}\left[h-H_0\left(\gamma,\lambda\right)\right]\right\}\right\|,$$

for $i=1,...,3d^2$. It now follows by (B.20) and (B.21) that

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[\frac{\partial H_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_i}-\frac{\partial H_t\left(\gamma,\lambda\right)}{\partial\theta_i}\right]\right\|=O\left(t\rho^t\right),\quad i=1,...,3d^2, \tag{B.48}$$

which implies (B.45). Observe that

$$\begin{aligned}\text{vec}\left[H_{t,h}\left(\gamma,\lambda\right)\right] &= \sum_{i=0}^{t-1}\left\{\left(B^{\otimes 2}\right)^i\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right)\gamma\right\}\\ &\quad+\sum_{i=0}^{t-1}\left(B^{\otimes 2}\right)^i A^{\otimes 2}\text{vec}\left(X_{t-1-i}X_{t-1-i}'\right)+\left(B^{\otimes 2}\right)^t\text{vec}\left(h\right)\end{aligned} \tag{B.49}$$

By simple differentiation of (B.49) and using that h is constant, we conclude that

$$E\sup_{\theta\in\Theta}\left\|\frac{\partial H_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_i}\right\|<\infty,\quad i=1,...,3d^2, \tag{B.50}$$

and hence that (B.46) holds. Thereby (B.43) holds with $r=1/4$. By the generalized Chebyshev inequality, the triangle inequality, and (B.43) for any $\varepsilon>0$,

$$\begin{aligned}&P\left(\left|\sqrt{T}\left[\frac{\partial L_T\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial L_{T,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right]\right|\geq\varepsilon\right)\\ \leq\ & T^{-1/8}\varepsilon^{-1/4}\sum_{t=1}^{T}E\left|\frac{\partial l_t\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}-\frac{\partial l_{t,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda_i}\right|^{1/4}\to 0\end{aligned}$$

as $T \to \infty$, and we conclude that (B.41) holds.

Next, we turn to (B.42). As before, from Hafner and Preminger (2009a, Proof of Lemma 4(ii)), we need to establish that for some $r > 0$,

$$E \sup_{\theta \in \Theta} \left| \frac{\partial^2 l_t(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 l_{t,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right|^r = O\left(t^2 \rho^t\right). \tag{B.51}$$

Again we can choose $r = 1/4$. From Hafner and Preminger (2009a, (B.36)), it follows, using Assumption 6, (B.47), (B.48), (B.50), (B.13), and Hölder's inequality, that it is sufficient to verify that

$$E \sup_{\theta \in \Theta} \left\| \text{vec} \left[\frac{\partial^2 H_t(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 H_{t,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right] \right\| = O\left(t^2 \rho^t\right), \tag{B.52}$$

and

$$E \sup_{\theta \in \Theta} \left\| \frac{\partial^2 H_{t,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right\| < \infty. \tag{B.53}$$

Equation (B.52) follows by iterating (A.5), and using that h is constant together with (B.20) and (B.21). By differentiating (B.49) twice and using that h is constant, we have that (B.53) holds. We conclude that (B.51) holds for $r = 1/4$. By (B.51) and the generalized Chebyshev inequality for any $\varepsilon > 0$,

$$\begin{aligned} &\sum_{t=0}^{\infty} P\left(\sup_{\theta \in \Theta} \left| \frac{\partial^2 l_t(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 l_{t,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right| > \varepsilon \right) \\ &\leq \sum_{t=0}^{\infty} \frac{1}{\varepsilon^{1/4}} E\left[\sup_{\theta \in \Theta} \left| \frac{\partial^2 l_t(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 l_{t,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right|^{1/4} \right] < \infty, \end{aligned}$$

so by the Borel-Cantelli lemma as $t \to \infty$,

$$\sup_{\theta \in \Theta} \left| \frac{\partial^2 l_t(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 l_{t,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right| \overset{a.s.}{\to} 0. \tag{B.54}$$

Now Cesàro's mean theorem and (B.54) imply that (B.42) holds. □

Appendix C Drift criteria for the BEKK-ARCH model

In order to find conditions for which the BEKK-ARCH model with Gaussian noise is geometrically ergodic with high-order moments we will make use of the following lemma.

Lemma C.1 ((Ghazal, 1996, Corollary 1.i-iii)). *Let $Q = x'\Omega x$ be a quadratic form where Ω is a $d \times d$ symmetric non-stochastic matrix and x is $N(0, I_d)$ distributed. Then*

$$\begin{aligned}
E\left[(x'\Omega x)^2\right] &= \operatorname{tr}^2\{\Omega\} + 2\operatorname{tr}\left\{\Omega^2\right\} \\
E\left[(x'\Omega x)^3\right] &= \operatorname{tr}^3\{\Omega\} + 6\operatorname{tr}\{\Omega\}\operatorname{tr}\left\{\Omega^2\right\} + 8\operatorname{tr}\left\{\Omega^3\right\} \\
E\left[(x'\Omega x)^4\right] &= \operatorname{tr}^4\{\Omega\} + 12\operatorname{tr}^2\{\Omega\}\operatorname{tr}\left\{\Omega^2\right\} + 12\operatorname{tr}^2\left\{\Omega^2\right\} + 32\operatorname{tr}\{\Omega\}\operatorname{tr}\left\{\Omega^3\right\} + 48\operatorname{tr}\left\{\Omega^4\right\}.
\end{aligned}$$

We are now able to prove the following theorem.

Theorem C.1. *Consider $\{X_t\}$ following the BEKK-GARCH process in (2.2) with $B = 0$ and Z_t i.i.d.$N(0, I_d)$. Then X_t is geometrically ergodic and the strictly stationary solution has (i) $E\|X_t\|^2 < \infty$ if $\rho(A \otimes A) < 1$, (ii) $E\|X_t\|^4 < \infty$ if $\rho(A \otimes A) < \frac{1}{\sqrt{3}} \approx 0.5774$, (iii) $E\|X_t\|^6 < \infty$ if $\rho(A \otimes A) < \frac{1}{15^{1/3}} \approx 0.4055$, and (iv) $E\|X_t\|^8 < \infty$ if $\rho(A \otimes A) < \frac{1}{105^{1/4}} \approx 0.3124$.*

Proof. Results *(i)* and *(ii)* are established in Rahbek (2004), see also Rahbek et al. (2002), where it is shown that the time-homogeneous Markov chain X_t is aperiodic, irreducible with respect to the Lebesgue measure, and compact sets in $\mathbb{R}^d$ are "small". This implies that we can use a k-step drift criterion, see also Tjøstheim (1990). Define the drift function

$$v(x) := 1 + (x'x)^3 = 1 + \|x\|^6 = 1 + \operatorname{tr}^3(xx').$$

Define $\Omega_x := C + Axx'A'$, then

$$\begin{aligned}
&E(v(X_t)\,|X_{t-1} = x) = 1 + E\left((X_t'X_t)^3\,|X_{t-1} = x\right) = 1 + E\left((Z_t'H_tZ_t)^3\,|X_{t-1} = x\right) \\
= \;& 1 + E\left[(Z_t'\Omega_x Z_t)^3\right] = 1 + \operatorname{tr}^3\{\Omega_x\} + 6\operatorname{tr}\{\Omega_x\}\operatorname{tr}\left\{\Omega_x^2\right\} + 8\operatorname{tr}\left\{\Omega_x^3\right\},
\end{aligned}$$

where the fourth equality follows by Lemma C.1. Ignoring terms of lower order than $\|x\|^6$, the right-hand side equals $15(x'A'Ax)^3$.
Let $L(\mathbb{R}^d)$ denote the space of linear mappings from $\mathbb{R}^d \to \mathbb{R}^d$. For linear mappings $\phi : L(\mathbb{R}^d) \to L(\mathbb{R}^d)$ we use the operator norm defined by $\|\phi\|_{op} := \sup_{\|x\|\neq 0} \frac{\|\phi(x)\|}{\|x\|}$, for which we observe that

$$\lim_{k\to\infty} \left\|\phi^k\right\|_{op}^{1/k} = \rho(\phi). \tag{C.1}$$

Let X be a $d \times d$ matrix in $L(\mathbb{R}^d)$, and define the mapping $\phi = (A \otimes A)$ from $L(\mathbb{R}^d) \to L(\mathbb{R}^d)$ by $\phi(X) := (A \otimes A)(X) = AXA'$. Note that $\phi^k(X) = A^kXA^{k\prime}$ and $\Omega_x = C_0 + \phi(xx')$.

Recursions give that $E\left(v\left(X_{t+k}\right)|X_t=x\right)$, apart from the lower-order terms, equals

$$15^k\left(x'A^{k\prime}A^kx\right)^3=\left(15^{k/3}x'A^{k\prime}A^kx\right)^3=\left\|\left(15^{1/3}\right)^k\phi^k\left(xx'\right)\right\|^3\leq\left\|\left(15^{1/3}\phi\right)^k\right\|_{op}^3\|x\|^6.$$

In light of (C.1), by choosing k large enough, we have that the drift condition is satisfied, if $\rho\left(15^{1/3}\phi\right)<1$,which means that $\rho\left(\phi\right)=\rho\left(A\otimes A\right)<1/15^{1/3}\approx 0.4055$. Result *(iv)* follows by similar arguments. □

Remark C.1. Theorem C.1 can be extended in order to establish conditions on $\rho\left(A\otimes A\right)$ for bounding other higher-order moments of X_t. If one seeks to verify that X_t is geometrically ergodic and $E\left\|X_t\right\|^n<\infty$, $n=2k$, $k\in\mathbb{N}$, one can define the drift function $v\left(x\right)=1+\left(x'x\right)^{n/2}$ and use general results for n^{th}-order moments of quadratic forms, see e.g. Corollary 2 of Bao and Ullah (2010).

Part II

Targeting estimation of CCC-GARCH models with infinite fourth moments

This chapter has been accepted for publication in *Econometric Theory.* It will appear in a revised form subsequent to editorial input by Cambridge University Press.[1]

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 moments of the data generating process is required, if one wants to perform inference in GARCH models by relying on asymptotic normality of the estimator. Such moment conditions may not be satisfied in practice for financial returns, highlighting a potential 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, under suitable conditions, its limiting distribution is multivariate stable. The rate of consistency of the estimator is slower than $\sqrt{T}$ and depends on the tail shape of the data generating process. A simulation study illustrates the derived properties of the targeting estimator.

[1] http://journals.cambridge.org/action/displayJournal?jid=ECT

I am grateful to Dennis Kristensen (co-editor) and the referees for their helpful comments and suggestions which have improved the paper substantially. I thank Brendan Beare, Iliyan Georgiev, Søren Johansen, Thomas Mikosch, Heino B. Nielsen, Anders Rahbek, participants at the INET/SoFiE Workshop on Skewness, Heavy Tails, Market Crashes, and Dynamics (Cambridge University, 2014), participants at the 10th BMRC-DEMS Conference (Brunel University, 2014), and participants at the Econometric Society World Congress (Montréal, 2015) for helpful comments. I also thank Farid Boussama for providing his Ph.D. thesis.

1 Introduction

In order to reduce the number of parameters in the numerical optimization when estimating multivariate GARCH models, targeting estimation has, by now, become a much applied tool among practitioners. The idea behind the method is to estimate the model in two steps. Initially, the model is reparametrized such that the unconditional (co)variances enter explicitly in the equation for the conditional (co)variances. In the first step, the unconditional variances are then estimated by a moment estimator and, conditional on this, the remaining parameters are estimated in a second step by quasi-maximum likelihood (QML). Recently, Pedersen and Rahbek (2014) have considered the asymptotic properties of the (covariance) targeting estimator for multivariate BEKK-GARCH models, whereas Francq et al. (2014) have considered similar properties for the (variance) targeting estimator for extended constant conditional correlation (CCC-) GARCH models. As established in both papers, and similar to the studies of the univariate GARCH models in Kristensen and Linton (2004) and Francq et al. (2011), at least finite fourth moments of the observed process is required in order to obtain asymptotic normality of the estimator. In practice, such moment restrictions may not be satisfied for asset returns, casting doubt on the validity of the inference performed in GARCH models based on targeting estimation. In this paper we derive the limiting distribution of the targeting estimator for CCC-GARCH models when the data generating process does not have finite fourth moments. By exploiting that, under certain conditions, the observed vector process has a multivariate regularly varying distribution, we show that the targeting estimator has a singular multivariate stable limiting distribution. The rate of consistency is slower than $\sqrt{T}$, as obtained in the presence of finite fourth moments, and is determined by the tail index (the index of regular variation) of the distribution of the observed process. Our conclusions are in line with the ones in a recent paper by Vaynman and Beare (2014) who consider the limiting distribution of the variance targeting estimator for univariate GARCH models.

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, as the classical CCC-GARCH model proposed by Bollerslev (1990) and its extended version by Jeantheau (1998). The asymptotic properties of the QML estimator for this model have been considered by Jeantheau (1998), Ling and McAleer (2003), and recently by Francq and Zakoïan (2012a). A drawback of the model, and especially of its extended version, is the large number of model parameters, which makes classical QML estimation difficult, if not infeasible, when the dimension of the time series is large. One can address this curse of dimensionality by applying simplified versions of the model, and/or by considering an alternative estimation method, such as targeting estimation originally proposed by Engle and Mezrich (1996). For the CCC-GARCH model, targeting estimation relies on estimating the vector of long-run variances in the first step. 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. Specifically, Francq et al. (2014) find that targeting estimation provides a decrease in computation time relative to QML estimation. 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. We refer to Francq et al. (2011) for a comprehensive treatment of that aspect for univariate GARCH models.

Existing literature on targeting estimation of multivariate GARCH models relies on at least finite fourth moments of the observed process in order to establish asymptotic normality. Such moment restrictions for the observed process may not be a realistic assumption, as for instance investigated by Loretan and Phillips (1994). We consider the case where the second moments are finite, implying consistency of the estimator, but the fourth moments are infinite. For this case, a central limit theorem does not apply to the vector of sample variances. The tail behavior of the CCC-GARCH process has investigated by Stărică (1999), whereas Fernández and Muriel (2009) derived the limiting distribution of the sample (auto-co)variances for the process. In line with their results, we exploit that a CCC-GARCH process can be represented by a stochastic recurrence equation (SRE). Under suitable conditions the SRE can be shown to have a multivariate regularly varying distribution, which allows us to characterize the tails of the distribution. Moreover, this property enables us to characterize the limiting distribution of the vector of sample variances by relying on theory for convergence of point processes. In particular, we show that this limiting distribution is multivariate stable, and 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 limiting distribution.

The rest of the paper is organized as follows. In Section 2 we introduce the targeting CCC-GARCH model, and Section 3 considers the two-step targeting estimation of the model. The large-sample theory for the targeting estimator is devoted to Section 4. Specifically, we introduce the notion of multivariate regular variation (Subsection 4.1) and convergence of point processes generated by multivariate regularly varying stochastic recurrence equations (Subsection 4.2). This provides a road map to the derivation of the limiting distribution of the targeting estimator in Subsection 4.3. Section 5 contains a simulation study that illustrates the derived properties of the estimator. Section 6 concludes the paper. All technical proofs can be found in the appendix.

Some notation and definitions throughout the paper: For $m, n \in \mathbb{N}$, I_n is the $(n \times n)$ identity matrix, and $O_{m \times n}$ is the $(m \times n)$ zero-matrix. Let $\|A\|$ denote the Euclidean norm of any scalar, vector, or matrix, A. For any positive definite matrix A, $A^{1/2}$ denotes the square-root of A in the Choleski sense. All limits are taken as the sample size $T \to \infty$, unless stated otherwise. Moreover, $\overset{p}{\to}$, $\overset{a.s.}{\to}$, and $\overset{w}{\to}$ denote convergence in probability, almost sure convergence, and weak convergence, respectively. For two real-valued functions f and g, $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$. 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 $t > 0$, $f(tx)/f(x) \to t^\kappa$ as $x \to \infty$, see e.g. Bingham et al. (1987, Chapter 1). The parameter κ is called the exponent of variation, and f is said to be slowly varying in the case where $\kappa = 0$. By definition, if f is regularly varying with exponent κ one can always write $f(x) = x^\kappa L(x)$ where $L(x)$ is slowly varying. Moreover, for any slowly varying function $L(x)$,

$$L(x)x^\alpha \underset{x\to\infty}{\to} 0 \;\; \forall \alpha < 0, \quad \text{and} \quad (L(x))^{\tilde{\alpha}} \text{ is slowly varying } \; \forall \tilde{\alpha} \in \mathbb{R} \,. \tag{1.1}$$

2 The targeting CCC-GARCH model

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

$$X_t(\theta_C) = \Sigma_t^{1/2}(\theta_C) Z_t, \tag{2.1}$$

where $\{Z_t : t \in \mathbb{N}\}$ is an i.i.d.$(0, I_d)$ sequence of random variables and the $(d \times d)$ matrix $\Sigma_t^{1/2}(\theta_C)$ is the square root of $\Sigma_t(\theta_C)$ given by the equations

$$\Sigma_t(\theta_C) = \tilde{D}_t(\theta_C) R \tilde{D}_t(\theta_C), \tag{2.2}$$

$$\tilde{D}_t^2(\theta_C) = \operatorname{diag}\left(\sigma_t^2(\theta_C)\right), \tag{2.3}$$

$$\sigma_t^2(\theta_C) = \omega + A X_{t-1}^{\odot 2}(\theta_C) + B \sigma_{t-1}^2(\theta_C). \tag{2.4}$$

Here $\operatorname{diag}(\sigma_t^2(\theta_C))$ is a diagonal matrix with the $(d \times 1)$ vector $\sigma_t^2(\theta_C)$ on the diagonal, ω is a $(d \times 1)$ vector with strictly positive entries, and A, B, and R are $(d \times d)$ matrices satisfying that A and B have nonnegative entries and R is a positive definite conditional correlation matrix. Moreover, $X_{t-1}^{\odot 2}(\theta_C) := X_{t-1}(\theta_C) \odot X_{t-1}(\theta_C)$, where

$\odot$ denotes the Hadamard product. The vector θ_C is the model parameters defined as $\theta_C := [\omega', \text{vec}(A)', \text{vec}(B)', \text{vech}^0(R)']'$, where $\text{vech}^0(R)$ stacks the columns below the principal diagonal downwards of R. The subscript C indicates that the model has the classical CCC-GARCH representation. We consider estimation conditional on the initial values X_0 and σ_0^2.

Necessary and sufficient conditions for the existence of a strictly stationary solution to the CCC-GARCH model are given in e.g. Boussama (1998, Section 5.4) and Francq and Zakoïan (2010, Theorem 11.6). In order to ensure that the second-order moments are finite, we assume throughout the paper that $\rho(A+B) < 1$ (stated formally in Assumption 2 in Section 4). By Jeantheau (1998, Proposition 3.1) this condition implies that the solution is also second-order stationary. In particular, the vector of unconditional variances of X_t is finite and given by

$$\gamma := \mathbb{E}[X_t^{\odot 2}] = \mathbb{E}[\sigma_t^2] = (I_d - A - B)^{-1} \omega \in (0, \infty)^d. \tag{2.5}$$

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 model.

3 Targeting estimation

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

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

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

$$\tilde{D}_t^2(\gamma,\lambda) = \text{diag}\left(\sigma_t^2(\gamma,\lambda)\right), \tag{3.3}$$

$$\sigma_t^2(\gamma,\lambda) = (I_d - A - B)\gamma + AX_{t-1}^{\odot 2}(\gamma,\lambda) + B\sigma_{t-1}^2(\gamma,\lambda), \tag{3.4}$$

such that $(I_d - A - B)\gamma \in (0,\infty)^d$ in order to ensure that $\sigma_t^2(\gamma,\lambda) \in (0,\infty)^d$ for all t. Notice 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. We now consider the targeting estimation method where first γ is estimated by method of moments, and λ is estimated by QML in a second step.

For a realization $\{X_t : t = 0, 1, ..., T\}$ of the model, the Gaussian log-likelihood function is given by

$$\hat{L}_T(\gamma,\lambda) := \frac{1}{T}\sum_{t=1}^{T}\hat{l}_t(\gamma,\lambda), \qquad \text{with} \tag{3.5}$$
$$\hat{l}_t(\gamma,\lambda) := \log\left\{\det[\hat{H}_t(\gamma,\lambda)]\right\} + X_t'\hat{H}_t^{-1}(\gamma,\lambda)X_t,$$

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

$$\hat{H}_t(\gamma,\lambda) = \hat{D}_t(\gamma,\lambda)R(\lambda)\hat{D}_t(\gamma,\lambda), \tag{3.6}$$
$$\hat{D}_t^2(\gamma,\lambda) = \text{diag}\left(\hat{h}_t(\gamma,\lambda)\right), \tag{3.7}$$
$$\hat{h}_t(\gamma,\lambda) = (I_d - A - B)\gamma + AX_{t-1}^{\odot 2} + B\hat{h}_{t-1}(\gamma,\lambda). \tag{3.8}$$

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, given in (2.5), by method of moments, $\hat{\gamma}_T := T^{-1}\sum_{t=1}^{T} X_t^{\odot 2}$. Substituting this estimator into the log-likelihood function and minimizing yield the targeting estimator for λ,

$$\hat{\lambda}_T := \arg\min_{\lambda\in\Theta_\lambda} \hat{L}_T(\hat{\gamma}_T,\lambda). \tag{3.9}$$

The two steps yield the targeting estimator of θ, $\hat{\theta}_T := (\hat{\gamma}_T', \hat{\lambda}_T')'$.

Remark 3.1. In contrast to the work of Pedersen and Rahbek (2014) on the BEKK-GARCH model, we only target the vector of unconditional variances and not the entire unconditional covariance matrix of X_t. The reason is, as pointed out by Bauwens et al. (2006, p.89), that for the CCC-GARCH model, Σ_t depends non-linearly on the past values of X_tX_t', i.e. there is no direct relation between model parameters and the off-diagonal elements of the unconditional covariance matrix of X_t. In particular, it does not seem possible to estimate the elements in the conditional correlation matrix, R, in a first step.

4 Large-sample theory for the targeting estimator

In this section we consider the asymptotic properties of the targeting estimator. Before turning to the results for the estimator, we introduce some important definitions and properties for stochastic recurrence equations (SREs). These properties are important for understanding the derivation for our main result in Theorem 4.2 below, stating the limiting distribution of the estimator.

By definition, a d-dimensional process $\{W_t : t \in \mathbb{Z}\}$ is given by an SRE if

$$W_t = \tilde{A}_t W_{t-1} + \tilde{B}_t, \tag{4.1}$$

where $\{(\tilde{A}_t, \tilde{B}_t) : t \in \mathbb{Z}\}$ is an i.i.d. sequence containing $(d \times d)$ matrices $\tilde{A}_t$ and $(d \times 1)$ vectors $\tilde{B}_t$. Importantly, in terms of the CCC-GARCH process, let $\theta_0 = (\gamma_0', \lambda_0')'$ denote the vector of true parameters such that $X_t := X_t(\theta_0)$ and $\sigma_t^2 := \sigma_t^2(\theta_0)$. Then notice that $Y_t := (X_t^{\odot 2\prime}, \sigma_t^{2\prime})'$, satisfies the SRE given by $Y_t = K_t Y_{t-1} + M_t$ with

$$K_t := \begin{bmatrix} \mathrm{diag}(\varepsilon_t^{\odot 2}) A_0 & \mathrm{diag}(\varepsilon_t^{\odot 2}) B_0 \\ A_0 & B_0 \end{bmatrix} \quad \text{and} \quad M_t := \begin{bmatrix} \mathrm{diag}(\varepsilon_t^{\odot 2})(I_d - A_0 - B_0)\gamma_0 \\ (I_d - A_0 - B_0)\gamma_0 \end{bmatrix}, \tag{4.2}$$

where $\varepsilon_t := R_0^{1/2} Z_t$.

Subsection 4.1 defines multivariate regular variation, which is a way of characterizing the tails of a random vector. Subsection 4.2 is devoted to introducing the notion of point processes and to considering properties of such processes generated by multivariate regularly varying SREs. In the following, we assume that the elements of $(\tilde{A}_t, \tilde{B}_t)$ in (4.1) are almost surely nonnegative for all t and that the SRE has a strictly stationary solution. Similar conditions hold for Y_t for the derivation of the limiting distribution of the targeting estimator in Subsection 4.3.

4.1 Multivariate regular variation

Multivariate regular variation is most commonly defined in terms of convergence of measures, and in particular we will make use of the notion of vague convergence defined next. Define $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, let $\mathbb{F}$ be a subset of $\overline{\mathbb{R}}^d$, and let $\mathcal{B}(\mathbb{F})$ be the Borel σ-field generated by the open sets of $\mathbb{F}$.[2] A measure μ on $\mathcal{B}(\mathbb{F})$ is called Radon if $\mu(F) < \infty$ for all $F \in \mathcal{B}(\mathbb{F})$ that are relatively compact, i.e. the closure of F is compact. Let $C_K^+(\mathbb{F}) := \{f : \mathbb{F} \to [0, \infty) : f$ is continuous with compact support$\}$, and let $M_+(\mathbb{F})$ denote the space of Radon measures on $\mathcal{B}(\mathbb{F})$. 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 $0 \le s \le t$, where $\mu(f) := \int_{\mathbb{F}} f(x)\,\mu(dx)$.

[2] Following Resnick (1987, p.123) we may 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 $(G_n)_{n \ge 1}$ such that any open G can be written as $G = \cup_{\alpha \in J} G_\alpha$ for a finite and countable index set J, and that distinct points in $\mathbb{F}$ may be separated by disjoint neighborhoods.

This topology is called the vague topology and is metrizable as a complete metric space, see Resnick (1987, Proposition 3.17). 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 $f \in C_K^+(\mathbb{F})$, $\mu_n(f) \to \mu(f)$ as $n \to \infty$. This defining condition of vague convergence is equivalent to $\mu_n(F) \to \mu(F)$ for all relatively compact $F \in \mathcal{B}(\mathbb{F})$ satisfying $\mu(\partial F) = 0$, with ∂F the boundary of F. For a detailed introduction to vague convergence we refer to Resnick (1987, pp.139-149).

In order to define multivariate regular variation through vague convergence of measures, it is natural to 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 in $\overline{\mathbb{R}}^d \setminus \{0\}$ with respect to the relative topology, as described in e.g. Resnick (2007, pp.172-175). We are now ready to define multivariate regular variation.[3]

Definition 4.1 (Mikosch, 2004, p.218)**.** A random vector $V \in \mathbb{R}^d$ and its distribution are said to be regularly varying if for a non-null Radon measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$,

$$\mu_x(\cdot) := \frac{\mathbb{P}(x^{-1}V \in \cdot)}{\mathbb{P}(\|V\| > x)} \xrightarrow{v} \mu(\cdot) \quad \text{as } x \to \infty. \tag{4.3}$$

The measure μ satisfies the homogeneity property $\mu(tA) = t^{-\kappa}\mu(A)$, $\kappa \geq 0$, for all $t > 0$, for any $A \in \mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$ bounded away from the origin with $\mu(\partial A) = 0$. If V satisfies (4.3), we say that V is multivariate regularly varying with index κ.

Remark 4.1. An important property of a multivariate regularly varying vector V with index κ is that for any $t > 0$, $\mathbb{P}(\|V\| > tx)/\mathbb{P}(\|V\| > x) \to t^{-\kappa}$ as $x \to \infty$, i.e. $\mathbb{P}(\|V\| > x)$ is regularly varying with exponent $-\kappa$. This property determines which moments of $\|V\|$ that are finite. In particular, $\mathbb{E}[\|V\|^\alpha] < \infty$ for all $\alpha \in [0, \kappa)$ and $\mathbb{E}[\|V\|^\alpha] = \infty$ for all $\alpha > \kappa$. Whether or not $\mathbb{E}[\|V\|^\kappa]$ is finite depends on the slowly varying function L from the representation $\mathbb{P}(\|V\| > x) = x^{-\kappa}L(x)$.

Basrak et al. (2002b) used Kesten's theorem, see Kesten (1973), to show that the SRE of the type (4.1) is multivariate regularly varying, and, likewise, we establish in Subsection 4.3 that the CCC-GARCH process has a similar tail property.

4.2 Convergence of point processes

In this subsection we consider the weak convergence of point processes generated by the stationary solution $\{W_t\}$ to the SRE in (4.1). We start out with some important definitions.

[3]The following definition holds for any choice of norm, $\|\cdot\|$.

With $\mathbb{1}()$ the indicator function, for any $x \in \mathbb{F}$ and any $A \in \mathcal{B}(\mathbb{F})$, let $\delta_x(A) = \mathbb{1}(x \in A)$. Let $\{w_i : i \geq 1\}$ be a countable collection of points of $\mathbb{F}$. A point measure μ on $\mathbb{F}$ is an element of $M_+(\mathbb{F})$ such that $\mu : \mathcal{B}(\mathbb{F}) \to \mathbb{N} \cup \{0, \infty\}$ and $\mu(\cdot) = \sum_{i=1}^{\infty} \delta_{w_i}(\cdot)$. Next, let $M_p(\mathbb{F}) := \{\mu \in M_+(\mathbb{F}) : \mu \text{ is a point measure}\}$ and let $\mathcal{M}_p(\mathbb{F})$ denote the Borel σ-field of $M_p(\mathbb{F})$ induced by the vague topology. 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 $M_p(\mathbb{F})$.

A special type of point process 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 $k \in \mathbb{N}_0$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$,

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

   and

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

Suppose that any finite dimensional distribution of $\{W_t\}$ is multivariate regularly varying with index $\kappa > 0$. The multivariate regular variation of W_t implies that there exists a deterministic sequence $\{a_T : T \in \mathbb{N}\}$, $0 < a_T \to \infty$, such that $T\mathbb{P}(\|W_t\| > a_T) \to 1$. Basrak et al. (2002b) then showed that under suitable conditions,

$$N_T(\cdot) := \sum_{t=1}^{T} \delta_{a_T^{-1} W_t}(\cdot) \xrightarrow{w} N(\cdot) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}(\cdot) \quad \text{in } M_p(\mathbb{F}), \tag{4.4}$$

where $(P_i : i \in \mathbb{N})$ are the points of a Poisson process on $(0, \infty)$ with intensity measure $\nu(\mathrm{d}y) = \varphi\kappa y^{-\kappa-1} \mathbb{1}\{y \in [0, \infty)\} \,\mathrm{d}y$, $\varphi \in (0, 1]$. The process $(P_i : i \in \mathbb{N})$ is independent of the i.i.d. sequence of point processes, $\{\sum_{j=1}^{\infty} \delta_{Q_{ij}}(\cdot) : i \in \mathbb{N}\}$, on $[0, \infty)^d \backslash \{0\}$, see Davis and Mikosch (1998, p.2054) for additional details. The result in (4.4) will show up to be important as it allows us, due to the continuous mapping theorem, to characterize the limiting distribution of the suitably normalized sum $\sum_{t=1}^{T} W_t$, and thereby, in terms of the CCC-GARCH process, the limiting distribution of the sample variances of X_t.

4.3 Properties of the targeting estimator

We now consider the large-sample properties of the targeting estimator. For the asymptotic theory, it is assumed that the sample $\{X_t : t = 0, 1, ..., T\}$, $X_t := X_t(\theta_0)$, is generated by the strictly and second-order stationary, ergodic solution to the targeting CCC-GARCH process. Such a solution is ensured to exist under Assumptions 1-3 below, as stated in Jeantheau (1998, Proposition 3.1). It is assumed that X_t has infinite fourth-moments, which implies that the targeting estimator is consistent but not asymptotically normally distributed. Our results are new and extend the existing literature on targeting estimation of multivariate GARCH models. The derivation of the limiting distribution of the estimator, stated in Theorem 4.2, is to some extent based on the recent results of Vaynman and Beare (2014) together with arguments used i Davis and Hsing (1995) and Davis and Mikosch (1998). However, due to the multivariate nature of the CCC-GARCH model and, especially, the multidimensional method of moments estimator for γ, additional arguments are needed here compared to these papers, see in particular the technical Lemmas B.5-B.7 in the appendix for details.

We start out by stating sufficient conditions for strong consistency of the targeting estimator.

Assumption 1. *The distribution of Z_t admits a probability density strictly positive on $\mathbb{R}^d$. Moreover, with $Z_{t,j}$ the j-th element of Z_t, $j = 1, ..., d$, there exists a $\beta_{0,j} \in (1, \infty]$ such that $\mathbb{E}[|Z_{t,j}|^{2\beta_{0,j}}] = \infty$ and $\mathbb{E}[|Z_{t,j}|^{2\beta}] < \infty$ for all $\beta < \beta_{0,j}$.*

Assumption 2. *For all $\lambda \in \Theta_\lambda$, $\rho(A + B) < 1$ and R is a positive definite correlation matrix. Moreover, each element of $(I_d - A - B)\gamma_0$ is positive and bounded away from zero on Θ_λ.*

Assumption 3. *The true parameters $\lambda_0 \in \Theta_\lambda$ and Θ_λ is compact.*

In light of the strict stationarity and ergodicity of $\{X_t\}$, it is convenient to introduce the strictly stationary and ergodic process $\{h_t(\gamma, \lambda) : t \in \mathbb{Z}\}$ given recursively by

$$h_t(\gamma, \lambda) = (I_d - A - B)\gamma + AX_{t-1}^{\odot 2} + Bh_{t-1}(\gamma, \lambda) \tag{4.5}$$

for $\rho(A + B) < 1$ and $(I_d - A - B)\gamma \in (0, \infty)^d$. For later purpose, we introduce correspondingly $D_t^2(\gamma, \lambda) = \text{diag}(h_t(\gamma, \lambda))$, $H_t(\gamma, \lambda) = D_t(\gamma, \lambda)R(\lambda)D_t(\gamma, \lambda)$, and

$$L_T(\gamma, \lambda) := \frac{1}{T}\sum_{t=1}^{T} l_t(\gamma, \lambda), \tag{4.6}$$

where

$$l_t(\gamma,\lambda) \ := \ \log\{\det[H_t(\gamma,\lambda)]\} + X_t' H_t^{-1}(\gamma,\lambda) X_t. \tag{4.7}$$

Notice that, by definition,

$$\Sigma_t(\gamma_0,\lambda_0) = H_t(\gamma_0,\lambda_0), \quad \tilde{D}_t(\gamma_0,\lambda_0) = D_t(\gamma_0,\lambda_0), \quad \sigma_t^2(\gamma_0,\lambda_0) = h_t(\gamma_0,\lambda_0) \quad \forall t. \tag{4.8}$$

Assumption 4. *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 4.2. Assumption 2 ensures that $\hat{h}_t(\gamma_0,\lambda) \in (0,\infty)^d$ and $h_t(\gamma_0,\lambda) \in (0,\infty)^d$ for all t and all $\lambda \in \Theta_\lambda$, and hence, since R is positive definite, that $\hat{L}_T(\gamma_0,\lambda)$ and $L_T(\gamma_0,\lambda)$ are well-defined on Θ_λ. This, together with the fact that $\hat{\gamma}_T$ is strongly consistent for $\gamma_0 \in (0,\infty)^d$ (as explained below) is, of course, necessary for deriving the asymptotic properties of the targeting estimator. Notice that for the univariate case, $\rho(A+B) < 1$ means that $A+B < 1$, which implies that $(1-A-B)\gamma_0 \in (0,\infty)$ is automatically satisfied, since $\gamma_0 \in (0,\infty)$. Hence the condition $(I_d - A - B)\gamma_0 \in (0,\infty)^d$ arise due to the multivariate nature of the model.

Remark 4.3. Assumption 4 is a high-level identification condition. Primitive identification conditions are discussed in Jeantheau (1998) and 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. We emphasize that strong consistency does apply under milder conditions, and, in particular, Z_t does not need to have a strictly positive density on $\mathbb{R}^d$ (Assumption 1), see e.g. Francq et al. (2014). Likewise, it might be possible to derive consistency in the case where $\{Z_t\}$ is not i.i.d. but only a martingale difference sequence, see e.g. Escanciano (2009). However, the assumption that $\{Z_t\}$ is i.i.d. is needed later when we consider the limiting distribution of the targeting estimator, where the i.i.d. assumption is used to show that the CCC-GARCH process is multivariate regularly varying. The proof of the following theorem can be found in Appendix A.

Theorem 4.1. *Suppose that Assumptions 1-4 hold. Then $\hat{\theta}_T \overset{a.s.}{\to} \theta_0$.*

Remark 4.4. The moment restrictions imposed are stronger than the ones required for consistency of the QMLE of the classical CCC representation in (2.4), where only a fractional moment of X_t is required to be finite, see e.g. Francq and Zakoïan (2010, Theorem 11.7). In particular, we need that $\mathbb{E}[\|X_t\|^2] < \infty$ in order to have that $\hat{\gamma}_T$ is strongly consistent, by a law of large numbers for ergodic processes, for $\gamma_0 \in (0, \infty)^d$.

Next, we turn to the limiting distribution of the targeting estimator and make the following assumption about the parameter space Θ_λ.

Assumption 5. *The true parameter vector, λ_0, belongs to the interior of Θ_λ.*

Assumption 5 is important for two reasons. First, it allows us to make a mean-value expansion of the first derivative of the log-likelihood function around θ_0. Second, it is crucial for showing that the CCC-GARCH process is multivariate regularly varying, as explained next. Under Assumptions 1-5, we show in Lemma B.3 that $Y_t := [X_t^{\odot 2\prime}, \sigma_t^{2\prime}]'$ is multivariate regularly varying with index $\kappa > 1$. This is done by first establishing that $\sigma_t^2 := \sigma_t^2(\theta_0)'$ is multivariate regularly varying (with the same κ). The proof relies on observing that σ_t^2 satisfies the SRE given by $\sigma_t^2 = \tilde{K}_t \sigma_{t-1}^2 + \tilde{M}_t$, where $\tilde{K}_t := [A_0 \operatorname{diag}(\varepsilon_{t-1}^{\odot 2}) + B_0]$, $\varepsilon_t := R_0^{1/2} Z_t$, and $\tilde{M}_t := (I_d - A_0 - B_0)\gamma_0$. By using Kesten's theorem it is shown that σ_t^2 is multivariate regularly varying. However, in order to make use of Kesten's theorem it is, among other things, required that $\prod_{i=1}^n \tilde{K}_i$ has strictly positive elements almost surely for some $n \in \mathbb{N}$. This condition is satisfied when λ_0 is in the interior of Θ_λ, implying that all elements of A_0 and B_0 are strictly positive.

Having established that Y_t is multivariate regularly varying with $\kappa > 1$, the idea is now to assume that $\kappa \in (1, 2)$. We limit ourselves to that specific case, since the case where $\kappa = 2$ leads to very complicated derivations, see e.g. Basrak et al. (2002b). If $\kappa > 2$ we have that X_t has finite fourth moments and the limiting distribution of the targeting estimator is Gaussian as stated in Remark 4.8 below.

We emphasize that a sufficient condition for X_t to have infinite fourth moments is that Z_t has infinite fourth moments. Indeed such condition is not necessary. In contrast to e.g. Berkes and Horváth (2003), 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 innovation Z_t, we do, according to Assumption 1, only assume that the noise process has at least finite second moments, but may not be heavy-tailed, i.e. we can have that X_t has infinite fourth 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.

The following theorem states the limiting distribution of the, suitably normalized, targeting estimator when X_t has infinite fourth moments.

Theorem 4.2. *Under Assumptions 1-5, 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 for a deterministic sequence*

$\{a_T : T \in \mathbb{N}\}$, satisfying $0 < a_T \to \infty$, $T\,\mathbb{P}(\|Y_t\| > a_T) \to 1$, and $a_T = T^{1/\kappa}L(T)$ with $L(T)$ slowly varying,

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

where the constant matrices J_0^{λ} and J_0^{γ} are stated in (4.18) below, and S has a d-dimensional multivariate κ-stable distribution.

We refer to Samorodnitsky and Taqqu (1994, Chapter 2) for a definition of and results for multivariate stable distributions.

Remark 4.5. Theorem 4.2 states that in the case where X_t has finite second moments but infinite fourth moments, the targeting estimator is consistent with rate Ta_T^{-1}. Since $a_T = T^{1/\kappa}L(T)$ with $L(T)$ slowly varying, it follows from (1.1) that the rate of consistency is slower than $\sqrt{T}$, which is the rate of consistency in the case of finite fourth moments, as stated in Remark 4.8. The limit of the (suitably scaled) targeting estimator, $(I_d, [(J_0^{\lambda})^{-1}J_0^{\gamma}]')'S$, 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)). The limiting distribution is concentrated on a d-dimensional subspace of $\mathbb{R}^{s_2}$ (with s_2 the dimension of θ) and is hence singular. Moreover, the parameters of the κ-stable distribution of S depend on the distribution of the point process $N(\cdot)$ introduced in the proof of Theorem 4.2 below. This fact makes it complicated to identify the parameters of the distribution of S, as recently pointed out by Bartkiewicz et al. (2011). Relying on blocking techniques instead of convergence of point processes, Bartkiewicz et al. (2011) derive the limiting distribution of $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$ for the univariate case and show that, under suitable conditions, the parameters of the limiting κ-stable distribution depend entirely on λ_0 and the distribution of Z_t.

Remark 4.6. In order to investigate the distribution of the targeting estimator further, we have included a simulation experiment in the next section. As investigated by Vaynman and Beare (2014), one can make use of subsampling techniques to construct confidence intervals for the targeting estimator. For the univariate GARCH model, in order to deal with the fact that a_T is unknown, Vaynman and Beare (2014) perform numerical simulations based on subsampling techniques for the self-normalized quantity $\sqrt{T}(\hat{\gamma}_T - \gamma_0)/\tau_T$ where $\tau_T^2 := \frac{1}{T}\sum_{t=1}^{T} X_t^4$. Although the validity of subsampling is shown to hold, Vaynman and Beare (2014) find that the technique does not perform well in practice for a reasonable sample size. We expect the same conclusion to hold for the CCC-GARCH model. As kindly pointed out by a referee, one should in general be careful using resampling techniques to construct confidence sets for the mean of a distribution that is in the domain of attraction of a κ-stable distribution $\kappa \in (0,2)$, see e.g. Berkes et al. (2010) and the references therein for a discussion.

Proof of Theorem 4.2. Define $s_1 := s_2 - d$, where s_2 is the dimension of θ. Using Assumption 5 and the strong consistency of $\hat{\theta}_T$ (Theorem 4.1), consider a mean-value expansion of the first derivative of the log-likelihood function around θ_0,

$$O_{s_1 \times 1} \quad = \quad \frac{\partial \hat{L}_T(\theta_0)}{\partial \lambda} + \hat{J}_T^{\lambda}\left(\hat{\lambda}_T - \lambda_0\right) + \hat{J}_T^{\gamma}\left(\hat{\gamma}_T - \gamma_0\right) \quad a.s., \tag{4.9}$$

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. Observe that $\{a_T : T \in \mathbb{N}\}$ exists due to the Lemma B.3.2. By Lemma B.1, we have that $\hat{J}_T^{\lambda}$ is invertible with probability approaching one for T sufficiently large, so reorganizing (4.9) yields that

$$Ta_T^{-1}\begin{pmatrix} \hat{\gamma}_T - \gamma_0 \\ \hat{\lambda}_T - \lambda_0 \end{pmatrix} = \begin{pmatrix} I_d & O_{d \times s_1} \\ -(\hat{J}_T^{\lambda})^{-1}\hat{J}_T^{\gamma} & -(\hat{J}_T^{\lambda})^{-1} \end{pmatrix} Ta_T^{-1}\begin{pmatrix} \hat{\gamma}_T - \gamma_0 \\ \frac{\partial \hat{L}_T(\theta_0)}{\partial \lambda} \end{pmatrix}. \tag{4.10}$$

The idea is to consider the limiting behavior of $Ta_T^{-1}\frac{\partial \hat{L}_T(\theta_0)}{\partial \lambda}$ and $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$. From Lemma B.4 we have that

$$Ta_T^{-1}\frac{\partial \hat{L}_T(\theta_0)}{\partial \lambda} \xrightarrow{p} 0. \tag{4.11}$$

With $\beta_{0,j}$, $j = 1, ..., d$, introduced in Assumption 1, notice that in the case where $\min\{\beta_{0,j} : j = 1, ..., d\} > 2$, it holds that $\mathbb{E}[\|Z_t\|^4] < \infty$, so similar to the usual QMLE case, see Francq and Zakoïan (2012a, Proof of Theorem 3.2), we have that $\sqrt{T}[\partial \hat{L}_T(\theta_0)/\partial \lambda] = O_p(1)$. Using that $a_T = T^{1/\kappa}L(T)$ and (1.1), it follows that $Ta_T^{-1}[\partial \hat{L}_T(\theta_0)/\partial \lambda] = o_p(1)$. In the case where $\min\{\beta_{0,j} : j = 1, ..., d\} \in (1, 2]$, a more sophisticated argument is needed, and we refer to the proof of Lemma B.4 for details.

We now consider the limit of $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$. Since the elements of A_0 and B_0 are non-negative and $\rho(A_0 + B_0) < 1$, it follows from Ling and McAleer (2003, Lemma 4.1) that $\rho(B_0) < 1$. Using arguments similar to the ones in Pedersen and Rahbek (2014, Proof of Lemma B.8), it holds that for any $\delta \in (0, 1)$

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

where $C_0 := (I_d - A_0 - B_0)^{-1}(I_d - B_0)$ and $\varepsilon_t := R_0^{1/2} Z_t$. Choosing $\delta \in (1 - 1/\kappa, 1)$, and using that $a_T = T^{1/\kappa} L(T)$ together with (1.1),

$$Ta_T^{-1}(\hat{\gamma}_T - \gamma_0) = C_0 a_T^{-1} \sum_{t=1}^{T} \left\{ \mathrm{diag}(\varepsilon_t^{\odot 2}) - I_d \right\} \sigma_t^2 + o_p(1) \,. \tag{4.12}$$

Consider $S_T := a_T^{-1} \sum_{t=1}^{T} \{\mathrm{diag}(\varepsilon_t^{\odot 2}) - I_d\} \sigma_t^2 = a_T^{-1} \sum_{t=1}^{T} \{X_t^{\odot 2} - \sigma_t^2\}$. Let $M_P(\mathbb{F})$ denote the collection of point processes on $\mathbb{F} := [0, \infty]^{2d} \backslash \{0\}$. Essentially due to the multivariate regular variation of Y_t and the fact that Y_t can be written as an SRE, it holds (Lemma B.3.3) that for the point process generated by $\{Y_t : t = 1, ..., T\}$, $N_T(\cdot) := \sum_{t=1}^{T} \delta_{a_T^{-1} Y_t}(\cdot)$,

$$N_T(\cdot) \xrightarrow{w} N(\cdot) \quad \text{in } M_P(\mathbb{F}) \,, \tag{4.13}$$

where N is specified in detail in Lemma B.3. The idea is to realize that S_T is essentially a mapping of $N_T(\cdot)$ and then exploit the convergence in (4.13). Specifically, it will be useful to define the mapping $V_\eta : M_P(\mathbb{F}) \to \mathbb{R}^d$, with the property

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

where $y_{t,i}$ denotes the i-th element of y_t. When evaluated at $N_T(\cdot)$, V_η yields an element-wise censoring of S_T. The censoring allows us to establish (see Lemma B.5) that V_η is continuous on a subset of $M_P(\mathbb{F})$ containing $N(\cdot)$ with probability one. The continuous mapping theorem and (4.13) then imply that

$$V_\eta(N_T) \xrightarrow{w} V_\eta(N). \tag{4.14}$$

Moreover, Lemma B.6 yields that

$$V_\eta(N) \xrightarrow{w} \tilde{S} \quad \text{as } \eta \to 0, \tag{4.15}$$

where $\tilde{S}$ is a d-dimensional random vector with a multivariate κ-stable distribution. From Lemma B.7 we have that

$$\lim_{\eta \to 0} \limsup_{T \to \infty} \mathbb{P} \left(\| S_T - V_\eta(N_T) \| \geq \delta \right) = 0. \tag{4.16}$$

Combining (4.14)-(4.16) with a variant of Slutsky's lemma (Billingsley, 1999, Theorem 3.2), we have that $S_T \overset{w}{\to} \tilde{S}$. In light of (4.12) we then conclude that

$$Ta_T^{-1}(\hat{\gamma}_T - \gamma_0) \overset{w}{\to} C_0\tilde{S} =: S. \tag{4.17}$$

Since $\kappa \in (1, 2)$, we have from Samorodnitsky and Taqqu (1994, Theorems 2.1.2 and 2.5.1(c)) that S has a multivariate κ-stable distribution. Combining (4.10), (4.11), (4.17), and Lemma B.1 with Slutsky's lemma yields

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

where J_0^{λ} and J_0^{γ} are given by

$$J_0^{\lambda} := \mathbb{E}\left[\frac{\partial^2 l_t(\theta_0)}{\partial\lambda\partial\lambda'}\right], \quad \text{and} \quad J_0^{\gamma} := \mathbb{E}\left[\frac{\partial^2 l_t(\theta_0)}{\partial\lambda\partial\gamma'}\right]. \tag{4.18}$$

□

Remark 4.7. As mentioned in the introduction, the limiting distribution of the (suitably normalized) sample auto-covariances of X_t has been studied by Fernández and Muriel (2009). Specifically, in the case where each element of $\varepsilon_t := R_0^{1/2}Z_t$ has a symmetric marginal distribution, the limit of $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$ is essentially stated in Fernández and Muriel (2009, Theorem 18.(2)), since $\hat{\gamma}_T$ is the diagonal of the sample auto-covariance of order $h = 0$. However, the result in Fernández and Muriel (2009) is only shown for the case $h \geq 1$, whereas the more complicated case where $h = 0$ is omitted from the proof. Hence the derivation of the limit of $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$ in the proof of Theorem 4.2 (partly) fills a gap in the existing literature on sample auto-covariances of CCC-GARCH processes.

Remark 4.8. As recently shown by Francq et al. (2014), if we in addition to Assumptions 1-5 have that $\mathbb{E}[\|X_t\|^4] < \infty$, then $\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{w}{\to} \mathcal{N}(0, \Sigma_0)$ for some matrix Σ_0. Notice that this moment condition is milder than what is assumed for establishing asymptotic normality of the QML estimator or targeting estimator for certain other multivariate GARCH models, see e.g. Hafner and Preminger (2009b) and Pedersen and Rahbek (2014) who require finite sixth moments.

Remark 4.9. For a given sample, it is difficult to know whether the fourth moments of the data-generating process are infinite. One way of testing for this moment condition would be to make inference about the maximal moment exponent of the process. Berkes et al. (2003) propose, and establish consistency and asymptotic normality of, an estimator for the maximal moment exponent for univariate GARCH(1, 1) processes. We conjecture that a similar estimator may be derived for CCC-GARCH processes, enabling one to test (under suitable conditions) for infinite fourth moments of X_t. Another way of testing for infinite fourth moments is the technique recently developed by Trapani (2014).

Remark 4.10. Assumption 1 only requires that the innovation, Z_t, has at least finite second moments (in addition to a strictly positive density). In particular, we allow for infinite fourth moments of the innovations. In such case, Berkes and Horváth (2004) showed that (for univariate GARCH models) one can obtain more efficient QML estimates by applying other likelihood functions than the Gaussian. A similar property would be interesting to investigate for the targeting estimator.

5 Simulation experiments

In this section we investigate the finite-sample properties of the targeting estimator via simulations. In particular, we illustrate that the distribution of the targeting estimator is poorly approximated by a Gaussian distribution (see Remark 4.8) when X_t has infinite fourth moments. Notice that for a given CCC-GARCH process (with dimension greater than one) it is not straightforward to estimate the index of regular variation κ. Instead we check the moment condition as follows. Obviously, X_t has infinite fourth moments if Z_t has. Next, provided that Z_t has finite fourth moments, it follows from Theorem C.1 in Appendix C that X_t has finite fourth moments if and only if

$$\eta_4 := \rho\left(\mathbb{E}\left\{\left[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0\right] \otimes \left[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0\right]\right\}\right) < 1,$$

where $\varepsilon_t := R_0^{1/2} Z_t$, and $\otimes$ denotes the Kronecker product. The condition is easy to check, by Monte Carlo integration, for a given distribution of Z_t and given parameter values θ_0.

In order to reduce the computational burden of the simulations, we consider a simplified version of the targeting CCC-GARCH model (3.1)-(3.4) with B a diagonal matrix. In that case, the vector of parameters, $\theta_\dagger$, coincide with θ except that $\theta_\dagger$ does not contain the off-diagonal elements of B. We define accordingly $\lambda_\dagger$, such that $\theta_\dagger = (\gamma', \lambda_\dagger')$, $\Theta_{\lambda_\dagger}$ (the parameter space of $\lambda_\dagger$), the true parameter vector $\theta_{\dagger,0} = (\gamma_0', \lambda_{\dagger,0}')'$, and the targeting estimator of $\theta_{\dagger,0}$, $\hat{\theta}_{\dagger,T} = (\hat{\gamma}_T', \hat{\lambda}_{\dagger,T}')'$. The following corollary states that $\hat{\theta}_{\dagger,T}$ has the same characteristics as $\hat{\theta}_T$. The proof can be found in Appendix A.

Corollary 5.1. *Suppose that the conditions of Theorem 4.2 hold with θ, λ, Θ_λ, and*

θ_0 replaced with $\theta_\dagger$, $\lambda_\dagger$, $\Theta_{\lambda_\dagger}$, and $\theta_{\dagger,0}$, respectively. Then $\hat{\theta}_{\dagger,T} \overset{a.s.}{\to} \theta_{\dagger,0}$, and

$$Ta_T^{-1}\begin{pmatrix} \hat{\gamma}_T - \gamma_0 \\ \hat{\lambda}_{\dagger,T} - \lambda_{\dagger,0} \end{pmatrix} \overset{w}{\to} \begin{pmatrix} I_d \\ -(J^{\lambda}_{\dagger,0})^{-1} J^{\gamma}_{\dagger,0} \end{pmatrix} S,$$

where S has a d-dimensional multivariate κ-stable distribution, and

$$J^{\lambda}_{\dagger,0} := \mathbb{E}\left[\frac{\partial^2 l_t(\theta_{\dagger,0})}{\partial\lambda_\dagger \partial\lambda_\dagger'}\right], \quad \textit{and} \quad J^{\gamma}_{\dagger,0} := \mathbb{E}\left[\frac{\partial^2 l_t(\theta_{\dagger,0})}{\partial\lambda_\dagger \partial\gamma'}\right].$$

The bivariate version of the simplified model has parameter vector

$$\theta_\dagger = (\gamma_1, \gamma_2, A_{11}, A_{12}, A_{21}, A_{22}, B_{11}, B_{22}, r)',$$

such that

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix}, \quad R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}.$$

For the simulations we consider the data-generating process (DGP) with true parameters

$$\gamma_0 = \begin{pmatrix} 18 \\ 7 \end{pmatrix}, \quad A_0 = \begin{pmatrix} A_{11,0} & 0.05 \\ 0.05 & 0.07 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 0.85 & 0 \\ 0 & 0.80 \end{pmatrix}, \quad r_0 = 0.4,$$

and $A_{11,0} \in \{0.06, 0.10, 0.115, 0.13\}$. Moreover, the distribution of Z_t can take four different forms: $N(0, I_2)$, t_3, t_5, and t_{10}, where t_i is a bivariate t-distribution with $i > 2$ degrees of freedom and suitably normalized such that $\mathbb{E}[Z_t] = O_{2\times 1}$ and $\mathbb{E}[Z_t Z_t'] = I_2$. For each simulation of the DGP we use a burn-in period of 1,000 observations, and all simulations are based on 1,000 replications. The simulations are carried out in OxMetrics 7.0.

First, we consider the case where Z_t is Gaussian and where $A_{11,0} \in \{0.06, 0.10, 0.115, 0.13\}$. We find that $\eta_4 < 1$ in the case $A_{11,0} \in \{0.06, 0.10, 0.115\}$ and $\eta_4 > 1$ when $A_{11,0} = 0.13$, i.e. we expect the distribution of the targeting estimator not to be well-approximated by a Gaussian distribution when $A_{11,0} = 0.13$, since X_t has infinite fourth moments for this choice of parameter values. Figure 5.1 contains kernel density estimates of the targeting estimates of A_{11}, where the estimates are based on $T = 1{,}000$ observations. For all four possible values of $A_{11,0}$ the associated estimates seem to be well-approximated by a Gaussian distribution. Even when $A_{11,0} = 0.13$, the Gaussian distribution seems to be a decent approximation, which may appear unexpected from the established theory. Figure 5.2 contains density

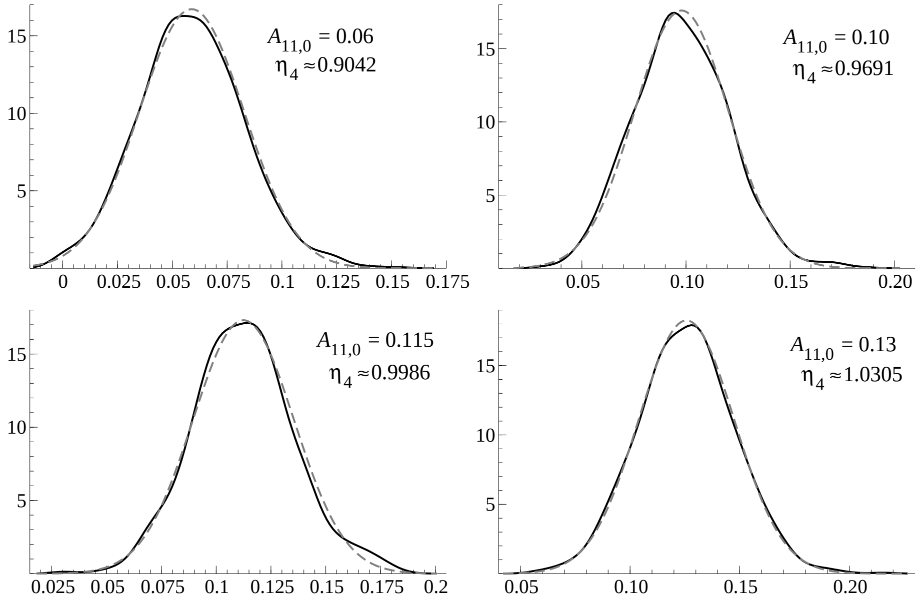

Figure 5.1: Kernel density estimates (full line) of the distribution of the targeting estimator of A_{11} and Gaussian density (dashed line) with same mean and variance. $A_{11,0} \in \{0.06, 0.10, 0.115, 0.13\}$, $Z_t \sim N(0, I_2)$, and $T = 1{,}000$.

plots similar to the ones in Figure 5.1, but for estimates based on $T = 50{,}000$ observations. Clearly, the targeting estimates for the case $A_{11,0} = 0.13$ do no longer seem to be well-approximated by a Gaussian distribution. The kernel density estimates for the targeting estimates of all other parameters (not reported here) show similar patterns, i.e. well-approximated by a Gaussian distribution for all choices of $A_{11,0}$ when $T = 1{,}000$ and not well-approximated when $A_{11,0} = 0.13$ and $T = 50{,}000$. The only exceptions are the estimates of γ that do not seem Gaussian distributed in the case $A_{11,0} = 0.13$ and $T = 1{,}000$, as can be seen from Figure 5.3 containing density plots of the estimates of the first element of γ, γ_1. Specifically, the distribution of the estimates for γ_1 seems to have a heavy right tail. The behavior of the estimates may be explained as follows. We have from Theorem 4.2 that the targeting estimator for all parameters, except for γ, has a singular limiting distribution. However, recall from the proof of Theorem 4.2, specifically (4.10), that for any finite T that $\hat{\lambda}_T$ depends on the first derivative of the log-likelihood function in the direction of λ, $\partial \hat{L}_T(\theta_0)/\partial\lambda$. Since Z_t has finite fourth moments, we know that $\sqrt{T}[\partial \hat{L}_T(\theta_0)/\partial\lambda]$ has a Gaussian limit, so the reason why the estimates of A_{11} seem to be well-approximated by a Gaussian distribution for $T = 1{,}000$ might be that $Ta_T^{-1}[\partial \hat{L}_T(\theta_0)/\partial\lambda]$ dominates $Ta_T^{-1}(\hat{\gamma}_T - \gamma_0)$ for "small" T. In contrast, the estimates of γ are not well-approximated by a Gaussian distribution due to the fact

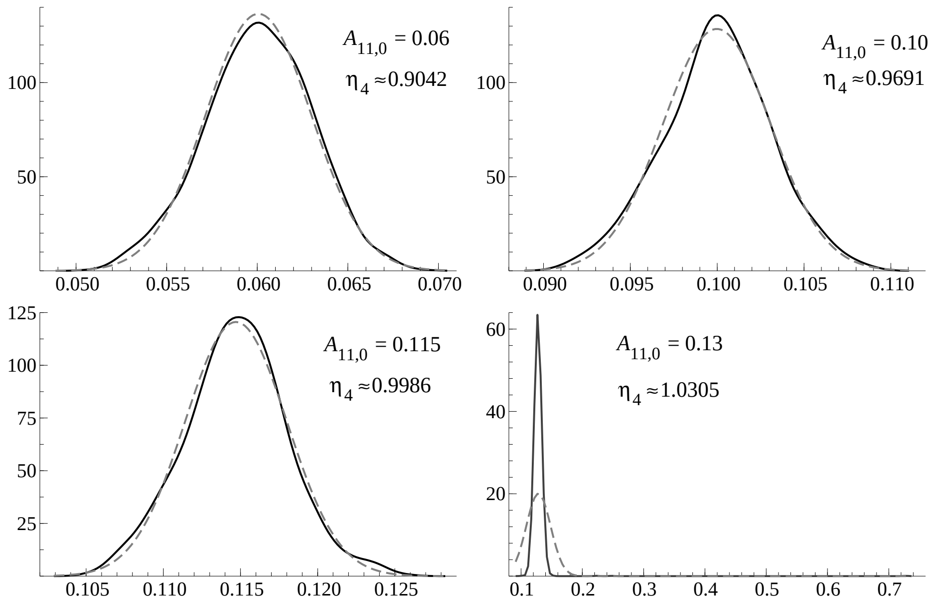

Figure 5.2: Kernel density estimates (full line) of the distribution of the targeting estimates of A_{11} and Gaussian density (dashed line) with same mean and variance. $A_{11,0} \in \{0.06, 0.10, 0.115, 0.13\}$, $Z_t \sim N(0, I_2)$, and $T = 50{,}000$.

that $\hat{\gamma}_T$ does not depend on any first derivatives of the log-likelihood function.

Next, we consider DGPs where $A_{11,0} = 0.06$ and Z_t has different distributions. Figure 5.4 contains density estimates for the targeting estimates for A_{11}. We see that the estimates are well-approximated by a Gaussian density for the cases where Z_t is $N(0, I_2)$-, t_{10}-, or t_5-distributed, which is in line with our theory, since η_4 is less than one in those situations. In the case where Z_t is t_3-distributed, X_t has infinite fourth moments, and we see that the targeting estimates do not fit a Gaussian density well. Similar observations hold for the estimates of all the other parameters (not reported here) and also for smaller sample sizes such as $T = 1{,}000$.

6 Concluding remarks and extensions

We have considered the targeting estimator for the extended CCC-GARCH model. In particular, we 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, by relying on results for convergence of point processes, shown that the rate of consistency is slower than $\sqrt{T}$ and that the limiting distribution of the estimator is singular multivariate stable. A simulation study illustrated the theoretical results.

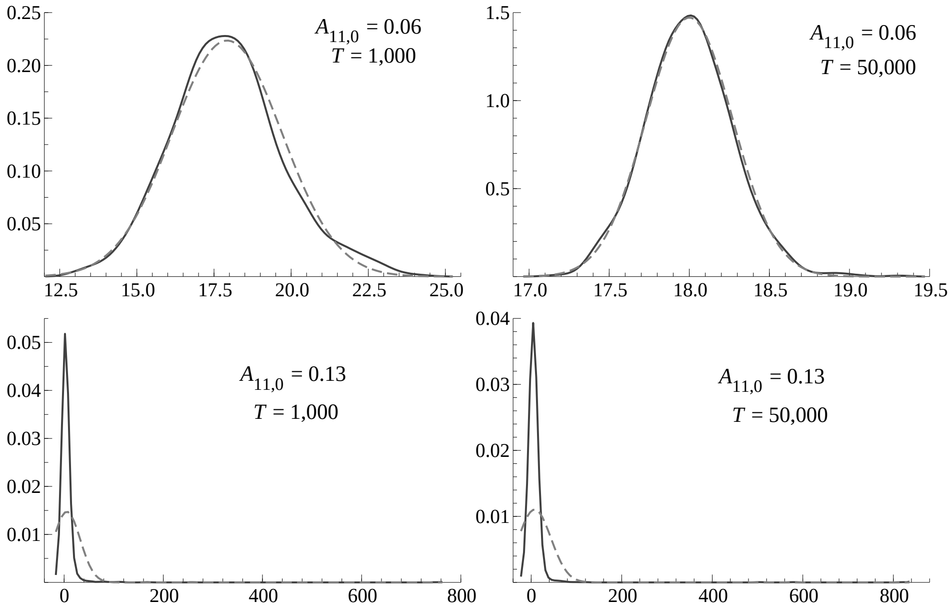

Figure 5.3: Kernel density estimates (full line) of the distribution of the targeting estimates of γ_1 and Gaussian density (dashed line) with same mean and variance. $A_{11,0} \in \{0.06, 0.13\}$, $Z_t \sim N(0, I_2)$, and $T \in \{1{,}000, 50{,}000\}$.

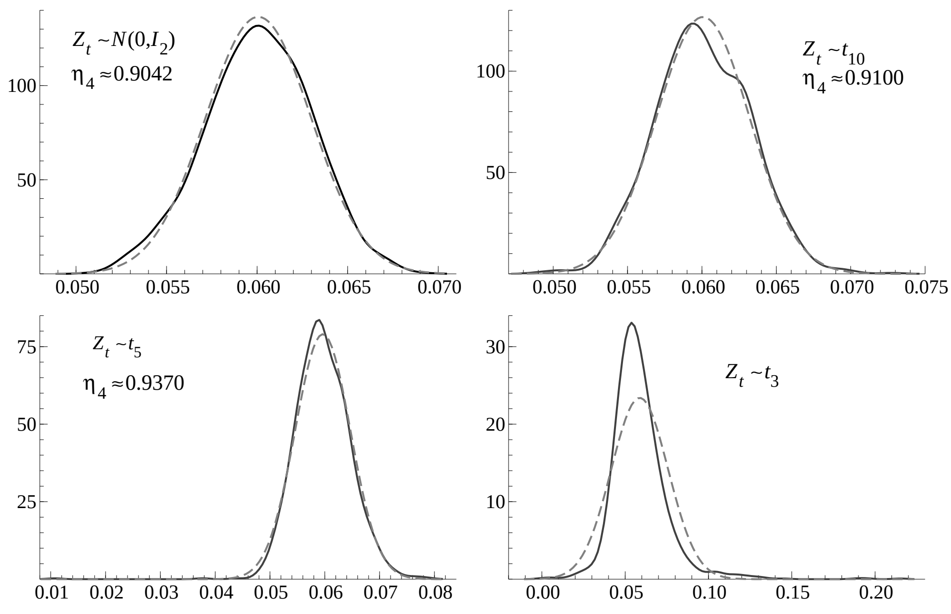

Figure 5.4: Kernel density estimates (full line) of the distribution of the targeting estimates of A_{11} and Gaussian density (dashed line) with same mean and variance. $A_{11,0} = 0.06$ and $T = 50{,}000$.

In particular, the simulations suggest that the distribution of the targeting estimator is potentially poorly approximated by a Gaussian distribution in the case where the CCC-GARCH process has infinite fourth moments. Interestingly, we find that for small sample sizes (T) the parameters estimated by QML in the second step of the estimation method seem to be well-approximated by a Gaussian distribution in the case where the CCC-GARCH process has infinite fourth moments, but the fourth moments of the noise process are finite. This might be explained by the fact for any finite T, the second step parameter estimates are determined by the method of moments estimates (the first step) together with the first derivative of the likelihood function, that has a Gaussian limiting distribution (when scaled with $\sqrt{T}$). For small T, the latter seems to dominate the distribution of the estimates from the second step, whereas the method of moments estimates seem to dominate for large T, as expected from our theoretical derivations.

An important area of future research is the development of techniques suitable for constructing precise confidence sets for the model parameters. Moreover, as recently proposed by Hill and Renault (2012) for univariate GARCH models, one could consider a so-called tail-trimmed version of the targeting estimator. The idea of tail-trimming is to robustify the targeting estimator against extreme observations, and Hill and Renault (2012) even find that the tail-trimmed estimator may have an asymptotic Gaussian distribution in the presence of heavy tails.

Appendix A Proofs of Theorem 4.1 and Corollary 5.1

Throughout the appendices, c and ϕ denote generic constants with $c \in [0, \infty)$ and $\phi \in [0, 1)$.

Proof of Theorem 4.1. Since $\{X_t : t \in \mathbb{Z}\}$ is ergodic with $E[\|X_t\|^2] < \infty$, the ergodic theorem implies that $\widehat{\gamma}_T$ is strongly consistent for γ_0. 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)$\sup_{\lambda \in \Theta_\lambda} \left| L_T(\gamma_0, \lambda) - \hat{L}_T(\widehat{\gamma}_T, \lambda) \right| \overset{a.s.}{\to} 0$.

(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) For any $\lambda \in \Theta_\lambda$, $\lambda \neq \lambda_0$, there exists an open ball with center λ, B_λ, such that almost surely,

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

First, by the triangle inequality,

$$\sup_{\lambda\in\Theta_\lambda}\left|L_T(\gamma_0,\lambda)-\hat{L}_T(\widehat{\gamma}_T,\lambda)\right| \leq \sup_{\lambda\in\Theta_\lambda}\left|L_T(\gamma_0,\lambda)-L_T(\widehat{\gamma}_T,\lambda)\right|, + \sup_{\lambda\in\Theta_\lambda}\left|L_T(\widehat{\gamma}_T,\lambda)-\hat{L}_T(\widehat{\gamma}_T,\lambda)\right|. \quad \text{(A.1)}$$

Next, for T sufficiently large, the mean-value theorem yields that, almost surely,

$$\sup_{\lambda\in\Theta_\lambda}\left|L_T(\gamma_0,\lambda)-L_T(\widehat{\gamma}_T,\lambda)\right| \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{d}\left|\widehat{\gamma}_{T,i}-\gamma_{0,i}\right| \sup_{\theta\in\tilde{\Theta}_\gamma\times\Theta_\lambda}\left|\frac{\partial l_t(\gamma,\lambda)}{\partial\gamma_i}\right|.$$

where $\tilde{\Theta}_\gamma$ is chosen to be a compact subset of $(0,\infty)^d$ such that $(I_d-A-B)\gamma\in(0,\infty)^d$ and bounded away from zero on $\tilde{\Theta}_\gamma\times\Theta_\lambda$, and such that γ_0 lies in the interior of $\tilde{\Theta}_\gamma$. Here $\widehat{\gamma}_{T,i}$ 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 of $l_t(\gamma,\lambda)$ with respect to the i-th element of γ. An expression for $\partial l_t(\gamma,\lambda)/\partial\gamma_i$ can be found in Francq and Zakoïan (2010, equation (11.67)). Assumption 2 together with Ling and McAleer (2003, Lemma 4.1) imply that $\rho(B)<1$ on Θ_λ, which ensures (since $\tilde{\Theta}_\gamma\times\Theta_\lambda$ is compact) that $\sup_{\theta\in\tilde{\Theta}_\gamma\times\Theta_\lambda}\|\partial h_t(\theta)/\partial\gamma_i\|<\infty$. Combining this with the moment condition, $E[\|X_t\|^2]<\infty$ and the fact that $\sup_{\theta\in\tilde{\Theta}_\gamma\times\Theta_\lambda}\|D_t^{-1}(\gamma,\lambda)\|\leq c$ and $\sup_{\theta\in\tilde{\Theta}_\gamma\times\Theta_\lambda}\|H_t^{-1}(\gamma,\lambda)\|\leq c$ give that $\mathbb{E}[\sup_{\theta\in\tilde{\Theta}_\gamma\times\Theta_\lambda}|\partial l_t(\gamma,\lambda)/\partial\gamma_i|]<\infty$, $i=1,...,d$. By the ergodic theorem and the consistency of $\widehat{\gamma}_T$, we have that $\sup_{\lambda\in\Theta_\lambda}|L_T(\gamma_0,\lambda)-L_T(\widehat{\gamma}_T,\lambda)|\overset{a.s.}{\to}0$. Francq and Zakoïan (2010, p.298) showed that $\sup_{\theta\in\tilde{\Theta}_\gamma\times\Theta_\lambda}|L_T(\gamma,\lambda)-\hat{L}_T(\gamma,\lambda)|\overset{a.s}{\to}0$, so in light of (A.1) 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 $\lambda\in\Theta_\lambda,\lambda\neq\lambda_0$, $\liminf_{T\to\infty}\inf_{\lambda^\star\in B_\lambda\cap\Theta}\hat{L}_T(\widehat{\gamma}_T,\lambda^\star)\geq\mathbb{E}[\inf_{\lambda^\star\in B_\lambda\cap\Theta}l_t(\gamma_0,\lambda^\star)]$ *a.s.*, where we have used (i), that for all $\lambda\in\Theta_\lambda$, $\mathbb{E}[\max\{-l_t(\gamma_0,\lambda),0\}]<\infty$, and the ergodic theorem. In light of (ii), we conclude that (iii) holds. □

Proof of Corollary 5.1. It is straightforward to show that the proof of Theorem 4.1 holds when B is diagonal, and hence that $\hat{\theta}_{\dagger,T}$ is strongly consistent. Turning to the limiting distribution of $\hat{\theta}_{\dagger,T}$, observe that it is possible to show that Lemma B.3 holds for the new strictly stationary process $\{\sigma_t^2(\theta_{\dagger,0})\}$. Similar to the proof of Lemma B.3.2, this can be done by relying on the derivations from Fernández and Muriel (2009, Proof of Theorem 5) where in particular the matrix $\prod_{i=1}^{n}(A_0\operatorname{diag}(\varepsilon_i^{\odot 2})+B_0)$ has almost surely strictly positive elements for some $n\in\mathbb{N}$ even if B_0 is diagonal, which then enables us to make use of Kesten's theorem. The remaining parts in the proof of Theorem 4.2 are straightforward to verify for the new model. $\square$

Appendix B Lemmas

Lemma B.1. *Suppose that the assumptions of Theorem 4.2 hold. With s_2 the dimension of θ and $\{a_T : T \in \mathbb{N}\}$ the deterministic sequence introduced in Theorem 4.2, it holds that*

1. $\mathbb{E}[\|\partial l_t^2(\theta_0)/\partial\theta\partial\theta'\|] < \infty$ *and* $J_0^\lambda := \mathbb{E}[\partial^2 l_t(\theta_0)/\partial\lambda\partial\lambda']$ *is non-singular.*

2. *With $\theta^\star$ between $\hat{\theta}_T$ and θ_0 as in the proof of Theorem 4.2,*

$$\partial^2 L_T(\theta^\star)/\partial\theta_i\partial\theta_j \xrightarrow{p} \mathbb{E}[\partial l_t^2(\theta_0)/\partial\theta_i\partial\theta_j'],\ i,j = 1,\dots,s_2.$$

3. *There exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that*

$$\sup_{\theta\in\mathcal{V}(\theta_0)} |\partial^2 L_T(\theta)/\partial\theta_i\partial\theta_j - \partial^2\hat{L}_T(\theta)/\partial\theta_i\partial\theta_j| \xrightarrow{p} 0,\ i,j = 1,\dots,s_2$$

4. $|Ta_T^{-1}[\partial L_T(\theta_0)/\partial\theta_i - \partial\hat{L}_T(\theta_0)/\partial\theta_i]| \xrightarrow{p} 0$*, $i = d+1, ..., s_2$.*

Proof. We choose $\mathcal{V}(\theta_0)$ sufficiently small such that all parameters in A, B, and γ are bounded away from zero and such that $(I_d - A - B)\gamma \in (0,\infty)^d$ is bounded away from zero on $\mathcal{V}(\theta_0)$. The points (1)-(4) can be verified by arguments similar to the ones in Francq and Zakoïan (2012a, Section A.4.2) and Francq and Zakoïan (2010, Section 11.4.3). In particular, (1) follows from Lemma B.2 and Assumption 4, and (2) follows from Lemma B.2 and Theorem 4.1. Points (3)-(4) do not depend on the parametrization and are verified along the lines of Francq and Zakoïan (2012a, pp.204-206), with the latter point following by using that $a_T = L(T)T^{1/\kappa}$ and (1.1). $\square$

Lemma B.2. *Under Assumptions 1-5, 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 and such that $(I_d-A-B)\gamma\in(0,\infty)^d$ is 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 $\mathbb{E}\{\sup_{\theta\in\mathcal{V}(\theta_0)}|[\partial h_{t,i_1}(\theta)/\partial\theta_i]|^{r_0}\}<\infty$. For $i=d+1,...,d^2$ and any i_1, $\theta_i(\partial h_{t,i_1}/\partial\theta_i)\leq h_{t,i_1}$,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 $k\geq 1$ and any j_1,

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

with $\zeta:=\min_{j_1}\{[(I_d-A-B)\gamma]_{j_1}\}>0$, where $[(I_d-A-B)\gamma]_{j_1}$ 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 2 and Ling and McAleer (2003, Lemma 4.1), $\phi_{j_1}\in[0,1)$ 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,

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

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 $r_0 \geq 1$. Similar arguments yield (B.2) and (B.3). 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 $k \geq 1$ 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 (2012a, p.202). □

For establishing the multivariate regular variation of the CCC-GARCH process in the next lemma, it will be useful to introduce the notion of regular variation in the Kesten sense, which is a way of characterizing the tails of a random vector through linear combinations, see Kesten (1973).

Definition B.1 (Basrak et al., 2002b, p.98)**.** A d-dimensional random vector V is said to be regularly varying in the Kesten sense if there exists a $\kappa > 0$ and a slowly varying function L such that for all $u \in \mathbb{R}^d \setminus \{0\}$,

$$\lim_{x\to\infty}\frac{\mathbb{P}\left(u'V > x\right)}{x^{-\kappa}L\left(x\right)} = w\left(u\right) \text{ exists.} \tag{B.9}$$

The function w takes finite values and there exists a $u_0 \neq 0$ with $w(u_0) > 0$.

The relationship between regular variation in the Kesten sense and multivariate regular variation, as in Definition 4.1, is investigated in Basrak et al. (2002a) and Boman and Lindskog (2009).

Lemma B.3. *Let $\sigma_t^2 := \sigma_t^2(\theta_0)$ and $Y_t := (X_t^{\odot 2\prime}, \sigma_t^{2\prime})'$, and let $M_p(\mathbb{F})$ denote the collection of point measures on $\mathbb{F} := [0,\infty]^{2d}\backslash\{0\}$. Suppose that Assumptions 1-5 hold. Then*

1. *σ_t^2 is multivariate regularly varying with some $\kappa > 1$.*
2. *any finite dimensional distribution of $\{Y_t : t \in \mathbb{Z}\}$ is multivariate regularly varying with index κ. Moreover, there exists a deterministic sequence $\{a_T : T \in \mathbb{N}\}$ satisfying $0 < a_T \to \infty$, $T\,\mathbb{P}(\|Y_t\| > a_T) \to 1$, and $a_T = T^{1/\kappa}L(T)$ for a slowly varying function $L(T)$.*
3. *it holds that*

$$N_T(\cdot) := \sum_{t=1}^{T}\delta_{a_T^{-1}Y_t}(\cdot) \xrightarrow{w} N(\cdot) \quad \text{in } M_p(\mathbb{F}), \tag{B.10}$$

where N is a point process on $\mathbb{F}$ with the representation

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

consisting of

(a) a Poisson random measure $\sum_{i=1}^{\infty}\delta_{P_i}(\cdot)$ on $(0,\infty)$ with intensity measure $\nu\,(\mathrm{d}y) = \varphi\kappa y^{-\kappa-1}\mathbb{1}\{y \in [0,\infty)\}\,\mathrm{d}y$, $\varphi \in (0,1]$,

(b) an i.i.d. sequence $\{\sum_{j=1}^{\infty}\delta_{Q_{ij}}(\cdot) : i \in \mathbb{N}\}$ of point processes in $\tilde{M}_p(\mathbb{F}) := \{\mu \in M_p(\mathbb{F}) : \mu\,(\{y : \|y\| > 1\}) = 0$ and $\mu(\{y : y \in \mathbb{S}^{2d-1}\}) > 0\}$, independent of $\sum_{i=1}^{\infty}\delta_{P_i}(\cdot)$.

Proof. Notice that σ_t^2 satisfies the SRE $\sigma_t^2 = \tilde{K}_t\sigma_{t-1}^2 + \tilde{M}_t$, where $\tilde{K}_t := [A_0\,\mathrm{diag}(\varepsilon_{t-1}^{\odot 2}) + B_0]$, $\varepsilon_t := R_0^{1/2}Z_t$, and $\tilde{M}_t := (I_d - A_0 - B_0)\gamma_0$. From Fernández and Muriel (2009, Proof of Theorem 5 and Remark 7) it holds that σ_t^2 is regularly varying in the Kesten sense with some $\kappa > 0$. That result is established by Kesten's theorem, see e.g. Basrak et al. (2002b, Theorem 2.4). It is here used that λ_0 is an interior point such that $A_0, B_0 \in (0,\infty)^{d\times d}$ implying that $\prod_{i=1}^{n}\tilde{K}_i > 0$ almost surely for some $n \in \mathbb{N}$, and that the strict stationarity of $\{\sigma_t^2\}$ implies that the top Lyapunov exponent of $\{\tilde{K}_t\}$ is strictly negative (Bougerol and Picard, 1992, Theorem 2.5). In particular, it holds that σ_t^2 satisfies Definition B.1 with L constant and $w\,(u) > 0$ for all $u \in [0,\infty)^d \setminus \{0\}$ and $w(u) = 0$ for all $u \in (-\infty, 0]^d \backslash \{0\}$. It then follows from Boman and Lindskog (2009, Corollary 2) that σ_t^2 is also multivariate regularly varying with index κ. With $\sigma_{t,i}^2$ the i-th element of σ_t^2, Kesten's theorem implies that $\mathbb{P}(\sigma_{t,i}^2 > k) \sim ck^{-\kappa}$ as $k \to \infty$, where c potentially depends on i. This property implies that $\mathbb{E}[\|\sigma_t^2\|^{\kappa}] = \infty$. Using Jeantheau (1998, Proposition 3.1), we have that $\mathbb{E}[\|\sigma_t^2\|] < \infty$, and hence that $\kappa > 1$. We conclude that point (1) holds.
Next, since $Y_t = [\mathrm{diag}(\varepsilon_t^{\odot 2}), I_d]'$ 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 Basrak et al. (2002b, Proposition A.1). The multivariate regular variation of any finite dimensional distribution of $\{Y_t\}$ follows using arguments similar to the ones given in Basrak et al. (2002b, Proof of Corollary 3.5). The sequence $\{a_T : T \in \mathbb{N}\}$ exists due to the fact that $\mathbb{P}(\|Y_t\| > k)$ is regularly varying with index $-\kappa$ (see Remark 4.1) and Resnick (2007, Theorem 3.6)
Point (3) is established by relying on Davis and Mikosch (1998, Theorem 2.8). The mixing condition $\mathcal{A}(a_n)$, see Davis and Mikosch (1998, p.2052), holds for $\{Y_t\}$ since the process is strongly mixing due to Lemma B.8. For verifying the anti-clustering condition, Davis and Mikosch (1998, (2.10)), we notice that Y_t satisfies the SRE in

(4.2). The condition can then be shown to hold for Y_t using arguments similar to the ones given in Basrak et al. (2002b, Proof of Theorems 2.10 and 3.6). In particular, it is used that the strict stationarity of $\{Y_t\}$ implies that the top Lyapunov exponent of $\{K_t\}$ is strictly negative. The limit in Davis and Mikosch (1998, (2.11)) can be shown to be strictly positive by using a multivariate version of Davis and Hsing (1995, Lemma 2.9) together with Basrak and Segers (2009, Proposition 4.2). The characterization of $N(\cdot)$ follows from Davis and Mikosch (1998, Corollary 2.4). □

Lemma B.4. *Under the assumptions of Theorem 4.2,* $a_T^{-1}T[\partial\hat{L}_T(\theta_0)/\partial\theta_i] \xrightarrow{p} 0$.

Proof. From Lemma B.1.4 we have that $a_T^{-1}T[\partial\hat{L}_T(\theta_0)/\partial\theta_i] = a_T^{-1}T[\partial L_T(\theta_0)/\partial\theta_i] + o_p(1)$, so it suffices to consider the limit of $a_T^{-1}T[\partial L_T(\theta_0)/\partial\theta_i]$. Let $s_0 := (d+2d^2)$. From Francq and Zakoïan (2012a, p.198) we have that for $i = d+1, ..., s_0$,

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i} = \text{tr}\left\{\left(I_d - \varepsilon_t\varepsilon_t' R_0^{-1}\right)\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)\right\} + \text{tr}\left\{\left(I_d - R_0^{-1}\varepsilon_t\varepsilon_t'\right)\left(\frac{\partial D_{0t}}{\partial\theta_i}D_{0t}^{-1}\right)\right\}, \tag{B.11}$$

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

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i} = \text{tr}\left\{\left(I_d - R_0^{-1}\varepsilon_t\varepsilon_t'\right)\left(R_0^{-1}\frac{\partial R_0}{\partial\theta_i}\right)\right\}. \tag{B.12}$$

Observe that

$$\text{tr}\left\{\left(I_d - \varepsilon_t\varepsilon_t' R_0^{-1}\right)\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)\right\} = \sum_{j=1}^{d^2}\text{vec}\left(I_d - \varepsilon_t\varepsilon_t' R_0^{-1}\right)_j \text{vec}\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)_j, \tag{B.13}$$

where $\text{vec}(I_d - \varepsilon_t\varepsilon_t' R_0^{-1})_j$ is the j-th element of $\text{vec}(I_d - \varepsilon_t\varepsilon_t' R_0^{-1})$. With $\beta_{0,j}$, $j = 1, .., d$, introduced in Assumption 1, let $\beta_0 := \min\{\beta_{0,j} : j = 1, ..., d\}$ and choose $\alpha \in (\kappa, \min\{\beta_0, 2\})$. As in Vaynman and Beare (2014, Proof of Theorem 3.3), then for any $\delta > 0$

$$\begin{aligned}
&\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\text{tr}\left\{\left(I_d - \varepsilon_t\varepsilon_t' R_0^{-1}\right)\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)\right\}\right| > \delta\right) \\
\leq\ & \sum_{j=1}^{d^2}\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\text{vec}\left(I_d - \varepsilon_t\varepsilon_t' R_0^{-1}\right)_j \text{vec}\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)_j\right| > \delta/d^2\right) \\
\leq\ & \sum_{j=1}^{d^2}\left(\delta/d^2\right)^{-\alpha}\mathbb{E}\left[\left|a_T^{-1}\sum_{t=1}^{T}\text{vec}\left(I_d - \varepsilon_t\varepsilon_t' R_0^{-1}\right)_j \text{vec}\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)_j\right|^{\alpha}\right],
\end{aligned} \tag{B.14}$$

where the first inequality follows by (B.13) and the triangle inequality, and the second inequality follows by the generalized Chebyshev inequality. With $\mathcal{F}_t$ the filtration generated by $\{X_i : i \leq t\}$, observe that $\{\text{vec}(I_d - \varepsilon_t\varepsilon_t' R_0^{-1})_j \text{vec}[D_{0t}^{-1}(\partial D_{0t}/\partial\theta_i)]_j, \mathcal{F}_t\}$

is a martingale difference sequence. Then by von Bahr and Esseen (1965, Theorem 2) we have that

$$\sum_{j=1}^{d^2}\left(\delta/d^2\right)^{-\alpha}\mathbb{E}\left[\left|a_T^{-1}\sum_{t=1}^{T}\operatorname{vec}\left(I_d-\varepsilon_t\varepsilon_t' R_0^{-1}\right)_j\operatorname{vec}\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)_j\right|^{\alpha}\right]$$
$$\leq\ cTa_T^{-\alpha}\sum_{j=1}^{d^2}\mathbb{E}\left[\left|\operatorname{vec}\left(I_d-\varepsilon_t\varepsilon_t' R_0^{-1}\right)_j\operatorname{vec}\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)_j\right|^{\alpha}\right].$$

This combined with Lemma B.2 and (B.14) gives that

$$\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{\left(I_d-\varepsilon_t\varepsilon_t' R_0^{-1}\right)\left(D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right)\right\}\right|>\delta\right)\ \leq\ cTa_T^{-\alpha}\to 0,\quad \text{(B.15)}$$

where the limit follows from the fact that $a_T = T^{1/\kappa}L(T)$ and (1.1). By similar arguments, we have that for any $\delta > 0$,

$$\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{\left(I_d-R_0^{-1}\varepsilon_t\varepsilon_t'\right)\left(\frac{\partial D_{0t}}{\partial\theta_i}D_{0t}^{-1}\right)\right\}\right|>\delta\right)\to 0 \qquad \text{(B.16)}$$

and

$$\mathbb{P}\left(\left|a_T^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{\left(I_d-R_0^{-1}\varepsilon_t\varepsilon_t'\right)\left(R_0^{-1}\frac{\partial R_0}{\partial\theta_i}\right)\right\}\right|>\delta\right)\to 0. \qquad \text{(B.17)}$$

Considering (B.11) and (B.12), in light of (B.15)-(B.17) we have that for any $i = d+1, ..., s_2$, $a_T^{-1}T[\partial L_T(\theta_0)/\partial\theta_i] \xrightarrow{p} 0$. □

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

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

where $y_{t,i}$ denotes the i-th element of y_t. Under the assumptions of Theorem 4.2, V_η is continuous on a subset of $M_p(\mathbb{F})$ containing the point process $N(\cdot)$, defined in Lemma B.3, with probability one.

Proof. The proof follows by arguments similar to the ones in Vaynman and Beare (2014, Proof of Lemma A.2), see also (Resnick, 1986, pp.84-85) for considerations about continuity of functionals on $M_p(\mathbb{F})$. Define the sets $B_\eta = \{x \in [0,\infty]^{2d} : \max_{i=d+1,...,2d}(x_i) > \eta\}$ and $A_\eta = \{\mu \in M_p(\mathbb{F}) : \mu(\partial B_\eta) = 0\}$. Moreover, consider a sequence $\{\mu_T : T \in \mathbb{N}\}$, $\mu_T \in M_p(\mathbb{F})$, such that $\mu_T \xrightarrow{v} \mu \in A_\eta$. 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 $(x_{T,1},...,x_{T,k})$ and $(x_1,...,x_k)$, respectively, for some finite k. Moreover, for each $i = 1,...,k$

$$x_{T,i} \to x_i. \tag{B.18}$$

Hence for T sufficiently large $V_\eta(\mu_T)$ and $V_\eta(\mu)$ do only depend on $(x_{T,1},...,x_{T,k})$ and $(x_1,...,x_k)$, respectively. By (B.18), $V_\eta(\mu_T) \to V_\eta(\mu)$, so V_η is continuous on A_η. The point process N from Lemma B.3 has the representation $N(\cdot) \stackrel{D}{=} \sum_{i=1}^\infty \sum_{j=1}^\infty \delta_{P_iQ_{ij}}(\cdot)$. Observe that the event $(N(\cdot) \notin A_\eta)$ can only occur if $(P_iQ_{ij} \in \partial B_\eta$ for some $i,j)$. Hence $\mathbb{P}(N(\cdot) \notin A_\eta) = \mathbb{P}(P_iQ_{ij} \in \partial B_\eta$ for some $i,j) \le \sum_{i=1}^\infty \sum_{j=1}^\infty \mathbb{P}(P_iQ_{ij} \in \partial B_\eta)$. 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 $\mathbb{P}(P_iQ_{ij} \in \partial B_\eta) = 0$, and we conclude that $\mathbb{P}(N(\cdot) \notin A_\eta) = 0$. □

Lemma B.6. *With V_η the mapping defined in Lemma B.5 and $N(\cdot)$ the point process defined in Lemma B.3, suppose that the assumptions of Theorem 4.2 are satisfied. Then*

$$V_\eta(N(\cdot)) \xrightarrow{w} \tilde{S} \quad \text{as } \eta \to 0, \tag{B.19}$$

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

Proof. Similar to Davis and Hsing (1995, pp.897-898), see also Vaynman and Beare (2014, Proof of Lemma A.3), consider the characteristic function of $V_\eta(N(\cdot))$, $\Psi_\eta : \mathbb{R}^d \to \mathbb{C}$. We derive the weak convergence by showing that $\Psi_\eta(t)$ converges point-wise 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. First, we establish the point-wise 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 $\sup_{0<a<b\le\eta} |\Psi_a(t) - \Psi_b(t)| < \epsilon$. With $S^{(\eta)} := V_\eta(N(\cdot))$ and $S_j^{(\eta)}$ its j-th element, observe that for any $\delta > 0$

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

where we have used Jensen's inequality. Moreover,

$$|\exp(\imath t'S^{(b)}) - \exp(\imath t'S^{(a)})| = \sqrt{2 - 2\cos(t'(S^{(b)} - S^{(a)}))},$$

so we have that

$$|\Psi_b(t) - \Psi_a(t)| \leq \mathbb{E}[\sqrt{2-2\cos(t'(S^{(b)}-S^{(a)}))}\mathbb{1}\{\max_{j\in\{1,\dots,d\}}|S_j^{(b)}-S_j^{(a)}|\leq\delta\}] \quad (B.20)$$
$$+\mathbb{E}[\sqrt{2-2\cos(t'(S^{(b)}-S^{(a)}))}\mathbb{1}\{\max_{j\in\{1,\dots,d\}}|S_j^{(b)}-S_j^{(a)}|>\delta\}].$$

Since t is fixed, $\max_{j\in\{1,\dots,d\}}|S_j^{(b)}-S_j^{(a)}|\leq\delta$, $t'(S^{(b)}-S^{(a)})\to 0$ 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 $\max_{j\in\{1,\dots,d\}}|S_j^{(b)}-S_j^{(a)}|\leq\delta$. Thereby the first term of the right-hand side of (B.20) 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.20) 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

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

so we just have to find an $\eta>0$ such that $\mathbb{P}(\max_{j\in\{1,\dots,d\}}|S_j^{(b)}-S_j^{(a)}|>\delta)<\epsilon/4$. We define $\tilde{V}_{a,b}:=\max_{j\in\{1,\dots,d\}}|V_{b,j}-V_{a,j}|$, 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.3, with probability one. The same must then hold for $\tilde{V}_{a,b}$. Hence $\tilde{V}_{a,b}(N_T)\xrightarrow{w}\tilde{V}_{a,b}(N)$, and we have that

$$\mathbb{P}(\max_{j\in\{1,\dots,d\}}|S_j^{(b)}-S_j^{(a)}|>\delta)=\mathbb{P}[\tilde{V}_{a,b}(N)>\delta]=\lim_{T\to\infty}\mathbb{P}[\tilde{V}_{a,b}(N_T)>\delta]. \quad (B.21)$$

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.7. Then

$$\tilde{V}_{a,b}(N_T(\cdot)) = \max_{j\in\{1,\dots,d\}}|S_{T,j}^{(b)}-S_{T,j}^{(a)}|\leq\sum_{j=1}^d(|S_{T,j}-S_{T,j}^{(b)}|+|S_{T,j}-S_{T,j}^{(a)}|). \quad (B.22)$$

In light of (B.21), (B.22), and Lemma B.7, choosing $\eta>0$ small enough, we have that

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

Next, 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 $\sup_{0<a<b\leq\eta}\sup_{t\in\mathcal{A}}|\Psi_a(t)-\Psi_b(t)| < \epsilon$. 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 $\mathcal{A}$. This, combined with the fact that $\Psi_\eta(t)$ is continuous on $\mathcal{A}$, yields that $\Psi(t)$ is continuous on $\mathcal{A}$, and in particular at $t = 0$. We conclude that as $\eta \to 0$, (B.19) 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)) and arguing in line with Davis and Hsing (1995, p.898). □

Lemma B.7. *Define* $S_T := a_T^{-1}\sum_{t=1}^{T}\{\mathrm{diag}(\varepsilon_t^{\odot 2}) - I_d\}\sigma_t^2$, *where* $\sigma_t^2 := \sigma_t^2(\gamma_0,\lambda_0)$ *and* $\varepsilon_t := R_0^{1/2}Z_t$. *Under the Assumptions of Theorem 4.2, with* V_η *the mapping defined in Lemma B.5 and* $N_T(\cdot)$ *the point process defined in Lemma B.3, for any* $\delta > 0$

$$\lim_{\eta\to 0}\limsup_{T\to\infty}\mathbb{P}(\|S_T - V_\eta(N_T)\| \geq \delta) = 0. \tag{B.23}$$

Proof. First observe that

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

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,

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

with $\eta > 0$. With $\beta_{0,j}$, $j = 1,..,d$, introduced in Assumption 1, let $\beta_0 := \min\{\beta_{0,j} : j = 1,...,d\}$ and choose $\alpha \in (\kappa, \min\{\beta_0, 2\})$. As in Vaynman and Beare (2014, Proof of Lemma A.1), with $\mathcal{F}_t$ the filtration generated by $\{X_i : i \leq t\}$ it holds that $\{(\varepsilon_{t,i}^2 - 1)\sigma_{t,i}^2, \mathcal{F}_t\}$ is a martingale difference sequence. This together with von Bahr and Esseen (1965, Theorem 2) gives

$$\begin{aligned}\mathbb{E}[|a_T^{-1}\sum_{t=1}^{T}(\varepsilon_{t,i}^2 - 1)\sigma_{t,i}^2\mathbb{1}\{\sigma_{t,i}^2 \leq \eta a_T\}|^\alpha] \quad &\leq \quad c a_T^{-\alpha}T\,\mathbb{E}[|\varepsilon_{t,i}^2 - 1|^\alpha]\,\mathbb{E}[(\sigma_{t,i}^2)^\alpha\mathbb{1}\{\sigma_{t,i}^2 \leq \eta a_T\}] \\ &= \quad c a_T^{-\alpha}T\int_0^{\eta a_T} x^\alpha\,\mathbb{P}(\sigma_{t,i}^2 \leq dx).\end{aligned} \tag{B.24}$$

By a variant of Karamata's theorem (Resnick, 2007, p.36) we have that

$$ca_T^{-\alpha}T\int_0^{\eta a_T} x^\alpha\,\mathbb{P}(\sigma_{t,i}^2 \;\leq\; dx) \sim ca_T^{-\alpha}T(\eta a_T)^\alpha\,\mathbb{P}(\sigma_{t,i}^2>\eta a_T)\left(\frac{\kappa}{\alpha-\kappa}\right).$$

This combined with Resnick (2007, Theorem 3.6) and (B.24) yields

$$\mathbb{E}[|a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2\mathbb{1}\{\sigma_{t,i}^2\leq\eta a_T\}|^\alpha] \underset{T\to\infty}{\to} c\eta^{\alpha-\kappa}.$$

Hence $\lim_{\eta\to0}\limsup_{T\to\infty}\mathbb{E}[|a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2\mathbb{1}\{\sigma_{t,i}^2\leq\eta a_T\}|^\alpha]=0$. We conclude, using Chebyshev's inequality, that for any $\tilde{\delta}>0$ and any $i\in\{1,...,d\}$,

$$\lim_{\eta\to0}\limsup_{T\to\infty}\mathbb{P}\left[|a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2\mathbb{1}\{\sigma_{t,i}^2\leq\eta a_T\}|\geq\tilde{\delta}\right]=0,$$

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

$$\lim_{\eta\to0}\limsup_{T\to\infty}\mathbb{P}\left[|a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2-a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2\mathbb{1}\{\sigma_{t,i}^2>\eta a_T\}|\geq\tilde{\delta}\right]=0. \tag{B.25}$$

Using the triangle and Boole's inequalities, we have that $\mathbb{P}(\|S_T-V_\eta(N_T)\|\geq\tilde{\delta})$ is bounded by $\sum_{i=1}^d\mathbb{P}[|a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2-a_T^{-1}\sum_{t=1}^T(\varepsilon_{t,i}^2-1)\sigma_{t,i}^2\mathbb{1}\{\sigma_{t,i}^2>\eta a_T\}|\geq\tilde{\delta}/d]$, so in light of (B.25) we conclude that (B.23) holds. □

Lemma B.8. *Let $\{Y_t:t\in\mathbb{N}_0\}$, $Y_t=(X_t^{\odot2\prime},\sigma_t^{2\prime})'$, be a process generated by the CCC-GARCH model, (2.1)-(2.4), with fixed initial values $X_0:=x\in\mathbb{R}^d$ and $\sigma_0^2:=h\in(0,\infty)^d$, and with $\theta_C=[\omega_0',\mathrm{vec}(A_0)',\mathrm{vec}(B_0)',\mathrm{vech}^0(R_0)']'$ satisfying $\rho(A_0+B_0)<1$ and that the diagonal elements of A_0 are strictly positive. Suppose that the distribution, Γ, of $\varepsilon_t:=R_0^{1/2}Z_t$ admits a probability density strictly positive on $\mathbb{R}^d$. Then $\{Y_t:t\in\mathbb{N}_0\}$ is geometrically ergodic on $[0,\infty)^d\times(0,\infty)^d$, and the associated strictly stationary process $\{Y_t:t\in\mathbb{Z}\}$ is geometrically β-mixing.*

Proof. Consider the Markov chain $\{\sigma_t^2:t\in\mathbb{N}_0\}$ given by $\sigma_t^2=\omega_0+[A_0\,\mathrm{diag}(\varepsilon_{t-1}^{\odot2})+B_0]\sigma_{t-1}^2$, with $\sigma_0^2=h$. The proof is structured as follows. First, we use the theory of Boussama et al. (2011) to show that $\{\sigma_t^2:t\in\mathbb{N}_0\}$ is aperiodic and ψ-irreducible. We then use a k-step ($k\geq1$) drift criterion, due to Tjøstheim (1990), to establish that the chain is V_σ-geometrically ergodic for some suitable Lyapunov function V_σ. Lastly, we use a result by Meitz and Saikkonen (2008) to conclude that the process $\{Y_t:t\in\mathbb{N}_0\}$ is V_Y-geometrically ergodic for some other suitable Lyapunov function V_Y, and that the stationary process $\{Y_t:t\in\mathbb{Z}\}$ is geometrically β-mixing.

We start out by showing that $\{\sigma_t^2:t\in\mathbb{N}_0\}$ belongs to the class of semi-polynomial

Markov chains as defined in Boussama et al. (2011, p.2339). Using the notation of Boussama et al. (2011), let $\mathcal{V} := \mathbb{R}^d$ and $\mathcal{U} := (0,\infty)^d$, and define $F : \mathcal{U} \times \mathbb{R}^d \to \mathcal{U}$, $F(z,x) = \omega_0 + (A_0 \operatorname{diag}(x^{\odot 2}) + B_0)z$. Observe that $\sigma_t^2 = F(\sigma_{t-1}^2, \varepsilon_{t-1})$, $t \geq 1$. The function f can, in our case, be chosen as the identity, which in turn means that the functions F and φ coincide. Therefrom it is straightforward to conclude that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is a semi-polynomial Markov chain.

Next, we seek to show that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is aperiodic and ψ-irreducible by verifying that Boussama et al. (2011, Proposition 3.10) applies to the chain. Let $\operatorname{supp}(\Gamma)$ denote the support of Γ. Boussama et al. (2011, Condition (A1)) is satisfied due to the assumptions on Γ. Moreover, for $k \in \mathbb{N}$ let $F^k(z, x_1, ..., x_k) := F(F^{k-1}(z, x_1, ..., x_{k-1}), x_k)$, where $z \in \mathcal{U}$, $(x_1, ..., x_k) \in (\mathbb{R}^d)^k$, and $F^1 := F$. Recursions give that $\sigma_t^2 = \{\sum_{i=0}^{t-1} \prod_{j=1}^{i} [A_0 \operatorname{diag}(\varepsilon_{t-j}^{\odot 2}) + B_0]\}\omega_0 + \{\prod_{i=0}^{t-1} [A_0 \operatorname{diag}(\varepsilon_{t-1-i}^{\odot 2}) + B_0]\}h$. Define the sequence $\{\sigma_{t,z,a}^2 : t \in \mathbb{N}_0\}$ with $\sigma_{0,z,a}^2 := z \in \mathcal{U}$ and $\sigma_{t,z,a}^2 = F(\sigma_{t-1,z,a}^2, a)$ where a is some interior point of $\operatorname{supp}(\Gamma)$. Notice that $\rho(B_0) < 1$ since $\rho(A_0 + B_0) < 1$ (Ling and McAleer, 2003, Lemma 4.1). Since zero is an interior point of $\operatorname{supp}(\Gamma)$, it holds that for any $z \in (0,\infty)^d$,

$$\sigma_{t,z,0}^2 = F^t(z, 0, ..., 0) = \sum_{i=0}^{t-1} B_0^i \omega_0 + B_0^t z \underset{t\to\infty}{\to} (I_d - B_0)^{-1} \omega_0 =: \xi \in \mathcal{U} \cap \mathcal{V}.$$

Hence we have established that Boussama et al. (2011, Condition (A2)) is satisfied. For any $z \in \mathcal{V} \cap \mathcal{U}$ we define the orbit $\mathcal{S}_F(z) := \bigcup_{k \in \mathbb{N}} \{F^k(z, x_1, ..., x_k) : x_1, ..., x_k \in \operatorname{supp}(\Gamma)\}$. Due to the non-negativity constraints of the elements in A_0 and B_0, it holds that for any $k \in \mathbb{N}$ that $F^k(\xi, x_1, ..., x_k)$, with $x_1, ..., x_k \in \operatorname{supp}(\Gamma)$, cannot take a value less than ξ element-wise. Let ξ_i denote the i-th element of ξ. Since the diagonal elements of A_0 are strictly positive, and due to the assumptions on Γ, it holds that $\mathcal{S}_F(\xi) = \times_{i=1}^d [\xi_i, \infty)$. Let $\mathcal{W}$ denote the Zariski closure of $\mathcal{S}_F(\xi)$, i.e. the smallest algebraic set that contains $\mathcal{S}_F(\xi)$, see Boussama et al. (2011, Appendix A) for details. The only polynomial that vanishes on $\times_{i=1}^d [\xi_i, \infty)$ is the zero polynomial, so we have that $\mathcal{W} = \mathbb{R}^d$. From Boussama et al. (2011, Proposition 3.10) we conclude that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is aperiodic and ψ-irreducible on $\mathcal{W} \cap \mathcal{U} = (0,\infty)^d$. Those properties of the Markov chain allow us, due to Tjøstheim (1990), to consider a k-step drift criterion for the Markov chain for some $k \in \mathbb{N}$. Specifically, we want to show that there exists a small set $C \in \mathcal{B}((0,\infty)^d)$, positive constants $a < 1$ and $b < \infty$, and a Lyapunov function $V_\sigma : (0,\infty)^d \to [1,\infty)$ such that for some fixed $k \in \mathbb{N}$,

$$\mathbb{E}\left[V_\sigma(\sigma_k^2) | \sigma_0^2 = h\right] \leq a V_\sigma(h) + b \cdot \mathbb{1}(h \in C) \quad \forall h \in (0,\infty)^d.$$

With ι_d a $(d \times 1)$ vector of ones, consider the function $V_\sigma(h) := 1 + \iota_d' h$, and, for some constant m sufficiently large, the set $C := \{h \in (0,\infty)^d : \iota_d' h \leq m\}$. Observe that

$$\mathbb{E}\left[V_\sigma(\sigma_k^2)|\sigma_0^2 = h\right] = \frac{1 + \iota_d'[\sum_{i=0}^{k-1}(A_0+B_0)^i]\omega_0 + \iota_d'(A_0+B_0)^k h}{1+\iota_d' h} V_\sigma(h),$$

where we have used that $\{\varepsilon_t\}$ is i.i.d. with $\mathbb{E}[\mathrm{diag}(\varepsilon_t^{\odot 2})] = I_d$. Since $\rho(A_0 + B_0) < 1$ and choosing k sufficiently large, there exists an m large enough such that for $h \in C^{\complement}$, $V_\sigma(h) \geq 1 + \iota_d'[\sum_{i=0}^{k-1}(A_0+B_0)^i]\omega_0 + \iota_d'(A_0+B_0)^k h$. We conclude that suitable constants a and b exist. In line with Boussama et al. (2011, Section 4.6) it can be shown that C is small. It then holds that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is V_σ-geometrically ergodic. From Meitz and Saikkonen (2008, Proposition 1 and the comments immediately after) we conclude that $\{Y_t : t \in \mathbb{N}_0\}$ is V_Y-geometrically ergodic, for some suitable function $V_Y : [0,\infty)^d \times (0,\infty)^d \to [1,\infty)$, and that the associated strictly stationary process $\{Y_t : t \in \mathbb{Z}\}$ is geometrically β-mixing. □

Appendix C A necessary and sufficient condition for finite 4TH moments of a CCC-GARCH process

Theorem C.1. *Let $\{X_t\}$ be the strictly stationary solution to the CCC-GARCH model in (2.1)-(2.4) with $\theta_C = \theta_0$ satisfying $\rho(A_0 + B_0) < 1$. Define $\varepsilon_t := R_0^{1/2} Z_t$ and suppose that $\mathbb{E}[\mathrm{diag}(\varepsilon_t^{\odot 2}) \otimes \mathrm{diag}(\varepsilon_t^{\odot 2})]$ exists and is finite. Then 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*

$$\eta_4 := \rho(\mathbb{E}\{[A_0\,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0] \otimes [A_0\,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0]\}) < 1, \qquad \text{(C.1)}$$

where $\otimes$ denotes the Kronecker product.

Proof. The sufficiency follows from He and Teräsvirta (2004, Corollary 2), so it remains to verify the necessity. In the following, for $(n \times m)$ matrices A and B, "$A > B$" means that each element of A exceeds the corresponding element of B. Similarly, we define $A \geq B$, and, moreover, $A > (\geq) 0$ means that each element of A is strictly positive (nonnegative), and $A < \infty$ means that each element of A is finite. Notice that $\mathbb{E}[X_t^{\odot 2}(X_t^{\odot 2})'] < \infty$ implies that $\mathbb{E}[\sigma_t^2 \otimes \sigma_t^2] < \infty$, where $\sigma_t^2 := \sigma_t^2(\gamma_0, \lambda_0)$.

Since $X_t^{\odot 2} = \mathrm{diag}(\varepsilon_t^{\odot 2})\sigma_t^2$,

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

Let $C_\sigma := (\omega_0 \otimes \omega_0) + \mathbb{E}[\omega_0 \otimes \{(A_0 \,\mathrm{diag}(\varepsilon_{t-1}^{\odot 2}) + B_0)\sigma_{t-1}^2\}] + \mathbb{E}[\{(A_0 \,\mathrm{diag}(\varepsilon_{t-1}^{\odot 2}) + B_0)\sigma_{t-1}^2\} \otimes \omega_0]$. Using that ε_{t-1} and σ_{t-1}^2 are independent,

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

Recursions give that for any $\tau \geq 1$,

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

using that $(\mathbb{E}[\{(A_0 \,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0)\} \otimes (A_0 \,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0)])^\tau \, \mathbb{E}[\sigma_{t-\tau}^2 \otimes \sigma_{t-\tau}^2] > 0$. Hence,

$$\infty > \mathbb{E}[\sigma_t^2 \otimes \sigma_t^2] \geq \sum_{i=0}^{\infty} (\mathbb{E}[\{(A_0 \,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0)\} \otimes (A_0 \,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0)])^i C_\sigma. \quad \text{(C.2)}$$

Since

$$\sum_{i=0}^{\infty} (\mathbb{E}[\{(A_0 \,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0)\} \otimes (A_0 \,\mathrm{diag}(\varepsilon_t^{\odot 2}) + B_0)])^i \geq 0 \quad \text{(C.3)}$$

and $C_\sigma \geq (\omega_0 \otimes \omega_0) > 0$, we have, in light of (B.33), that (C.3) converges, which is necessary and sufficient for (C.1). □

Part III

Inference and testing on the boundary in extended constant conditional correlation GARCH models

Abstract

We consider inference and testing in extended constant conditional correlation GARCH models in the case where the true parameter vector is a boundary point of the parameter space. This is of particular importance when testing for no volatility spillovers in the model. The large-sample properties of the QMLE are derived together with the limiting distributions of the related LR, Wald, and LM statistics. Due to the boundary problem, these large-sample properties become non-standard. The size and power properties of the tests are investigated in a simulation study. As an empirical illustration we test for no volatility spillovers between foreign exchange rates.

1 Introduction[1]

Testing for volatility spillovers between time series has become an important tool in empirical finance. Following the simple arguments of Ross (1989) that the (conditional) variance of asset price changes is directly related to the rate of information flow, volatility spillovers may be viewed as a way of measuring information transmissions in and between markets and thereby their connectedness (Conrad and Weber, 2013). Typically, volatility spillovers are defined in relation to multivariate conditional volatility models, such as multivariate GARCH, for price changes. As an example, Conrad et al. (1991) applied bivariate GARCH models to conclude that volatility surprises to large market value firms are important to the future dynamics of the returns of smaller firms (but not conversely). Another example is Bali and

[1]I am grateful to Pasquale Della Corte, Heino B. Nielsen, Thor P. Nielsen, and Anders Rahbek for helpful comments and suggestions. I thank Donald W. K. Andrews for providing his unpublished manuscript "Estimation When a Parameter is on a Boundary: Part II" and Tomoaki Nakatani for providing the data from the paper Nakatani and Teräsvirta (2009). Parts of this research were carried out while I was visiting the Imperial College Business School in Spring 2014. I thank Paolo Zaffaroni for his hospitality and the Niels Bohr Foundation for financial support.

Hovakimian (2009) who applied a similar technique to individual stock and option returns. Their findings indicate that there exist spillovers from options to equity markets, suggesting that options markets are subject to trading by investors with private information. For other applications of multivariate GARCH models for assessing spillovers we refer to Conrad and Weber (2013) and the references therein. A multivariate GARCH model well suited for quantifying spillovers is the extended constant conditional correlation (ECCC-) GARCH model of Jeantheau (1998), considered in this paper. In the ECCC-GARCH model the matrices governing the ARCH and GARCH dynamics - respectively, the matrices A and B introduced in the following section - are allowed to be non-diagonal, and with the off-diagonal elements directly related to the volatility spillovers. Specifically, testing for no volatility spillovers relies on testing for whether the off-diagonal elements of the matrices are equal to zero.

In this paper we consider the properties of the quasi-maximum likelihood estimator (QMLE) for the parameters in the ECCC-GARCH model in the case where some of the elements of the A and B matrices are allowed to be zero under the null. For the ECCC-GARCH model, the parameter space is typically restricted such that all elements of A and B are nonnegative, which is assumed in the existing literature on the large-sample properties of the QMLE, as in Jeantheau (1998, Definition 3.1), Ling and McAleer (2003, Assumption 3), and Francq and Zakoïan (2012a, p.183). The constraints are convenient as they (partly) ensure that the conditional covariance matrix is positive definite, and hence that the log-likelihood function is well-defined. However, as will be the main message from this present paper, the constraints lead to complications if one wants to test for no spillovers, and in particular one cannot rely on standard large-sample theory for QML estimation. Technically, the parameter is on the boundary of the parameter space under the null hypothesis of no spillovers. This implies that the limiting distribution of the QMLE cannot be obtained by relying on arguments based on a Taylor expansion around a zero-valued score.

We make the following contributions. First, we consider the asymptotic properties of the QMLE in the case where the true parameter value is on the boundary of the parameter space. In contrast to the standard case where the parameter value is an interior point, the (suitably normalized) QMLE does not have a Gaussian limit, but instead its limiting distribution is the given by the projection of a Gaussian vector (that occurs in the interior case) onto a set that depends on the true parameter. Second, in order to avoid boundary issues when testing for spillovers in the ECCC-GARCH model, Nakatani and Teräsvirta (2009) proposed a Lagrange multiplier (LM) statistic. We consider a modified version of this statistic, that is based on left/right partial derivatives of the log-likelihood function with respect to the parameters on the boundary, and moreover the test is a QML-type that allows for an unknown distribution of the (independent) innovations, see White (1996, Chapter 8). We also consider QLR and Wald tests both taking into account that the true parameter is a boundary point. Whereas the limiting distribution of the QMLE for univariate GARCH models when the true parameter is on the boundary has been considered by Andrews (1998, 2001) and Francq and Zakoïan (2007, 2009), we are not aware of any other papers considering this for the QMLE for multivariate GARCH models. Some early considerations on testing when the null vector is a boundary point of the maintained hypothesis can be found in Chernoff (1954) and Perlman (1969), whereas Andrews (1999, 2001) provides a very general theory for extremum estimators when the null parameter vector is a boundary point of the parameter space.

The rest of the paper is structured as follows. In Section 2 we introduce the ECCC-GARCH model and state some important properties of ECCC-GARCH processes. Moreover, we introduce the notion of spillovers and their relation to Granger causality. Section 3 introduces the QMLE and states the large-sample properties of the estimator, whereas the associated likelihood ratio, Wald, and Lagrange multiplier tests (for no-spillovers) are presented in Section 4, which also contains an algorithm for determining critical values for the proposed tests. Section 5 contains simulation studies that investigate the empirical size and power properties of the proposed tests, whereas Section 6 is devoted to an empirical illustration where we test for no volatility spillovers between assets in foreign exchange markets. Section 7 concludes the paper. All technical derivations can be found in the appendix.

Some notation and definitions: Unless stated otherwise all limits are taken as $T \to \infty$. Let $\xrightarrow{w}$ denote convergence in distribution. For a random vector X, $\mathcal{L}(X)$ denotes the distribution of X. For $n \in \mathbb{N}$, I_n is the $(n \times n)$ identity matrix, and the zero matrix $0_{m\times n}$ is an $(m \times n)$ matrix with all elements equal to zero. With $\otimes$ denoting the Kronecker product and $\odot$ the Hadamard product, we introduce for a matrix A the notation $A^{\otimes p} := A \otimes A \otimes \cdots \otimes A$ and $A^{\odot p} := A \odot A \odot \cdots \odot A$ (p factors). The Euclidean norm of a vector or matrix is denoted $\|\cdot\|$. Let $\mathbb{R}_+$ denote the nonnegative real numbers, and let $\mathbb{S}^d_{++}$ denote the space of $(d \times d)$ positive definite matrices. For any $C \in \mathbb{S}^d_{++}$ and any $(d \times 1)$ vectors x and y let $\langle x, y\rangle_C := x'Cy$ and $\|x\|_C := \langle x, x\rangle_C^{1/2}$. Moreover, for $\Theta \subset \mathbb{R}^d$ and $\theta \in \Theta$, $\Theta - \theta := \{x - \theta : x \in \Theta\}$.

2 The ECCC-GARCH model and its properties

In this section we introduce the ECCC-GARCH model, state some important properties of the ECCC-GARCH process, and introduce the notion of volatility spillovers and its relation to Granger (non)causality.

2.1 The model

We consider the ECCC-GARCH$(1,1)$ model of Jeantheau (1998) for $t \in \mathbb{Z}$ given by

$$
\begin{aligned}
X_t(\theta) &= \Sigma_t^{1/2}(\theta)\eta_t, && (2.1)\\
\Sigma_t(\theta) &= \tilde{D}_t(\theta)R(\theta)\tilde{D}_t(\theta), && (2.2)\\
\tilde{D}_t^2(\theta) &= \operatorname{diag}[\sigma_t^2(\theta)], && (2.3)\\
\sigma_t^2(\theta) &= \kappa + AX_{t-1}^{\odot 2}(\theta) + B\sigma_{t-1}^2(\theta), && (2.4)
\end{aligned}
$$

with $(\eta_t : t \in \mathbb{Z})$ an i.i.d. sequence of d-dimensional random variables with $\mathbb{E}[\eta_t] = 0_{d\times 1}$ and $\mathbb{E}[\eta_t\eta_t'] = I_d$. Moreover, $\operatorname{diag}[\sigma_t^2(\theta)]$ is a diagonal matrix with the $(d \times 1)$ vector $\sigma_t^2(\theta)$ on the diagonal and $R(\theta)$ is a positive definite correlation matrix. The model is parametrized according to $\theta = (\kappa', \operatorname{vec}(A)', \operatorname{vec}(B)', \operatorname{vech}^0(R)')'$, where $\operatorname{vech}^0(R)$ stacks the columns below the principal diagonal downwards of R. The parameter space, Θ, is given by a subset of $(0,\infty)^d \times [0,\infty)^{2d^2} \times (-1,1)^{d(d-1)/2} \subset \mathbb{R}^{s_0}$ with $s_0 := d + 2d^2 + (d(d-1)/2$. Observe that the parameter space is defined such that the elements of A and B are nonnegative. This condition, together with the restriction $\kappa \in (0,\infty)^d$, ensures that $\sigma_t^2(\theta_c) \in (0,\infty)^d$ almost surely, which, combined with the fact that $R(\theta) \in \mathbb{S}_{++}^d$, implies that $\Sigma_t(\theta) \in \mathbb{S}_{++}^d$ almost surely for all $\theta \in \Theta$.

Remark 2.1. The ECCC-GARCH model is a generalized version of the CCC-GARCH model proposed by Bollerslev (1990) where the matrices A and B are restricted to be diagonal.

2.2 Properties of the ECCC-GARCH process

For a fixed $\theta \in \Theta$, equations (2.1)-(2.4) yield an ECCC-GARCH process $(X_t : t \in \mathbb{Z})$. The properties of such a process have been studied several places in the literature, including Jeantheau (1998), Boussama (1998, Chapter 5), Ling and McAleer (2003), He and Teräsvirta (2004), and Francq and Zakoïan (2010, Chapter 11). Importantly, by Francq and Zakoïan (2010, Theorem 11.6), under suitable conditions, it holds that the process has a unique strictly stationary and ergodic solution if and only if $\gamma := \inf\{\mathbb{E}[(n+1)^{-1}\log(\|\Xi_0\Xi_{-1}\cdots\Xi_{-n}\|)] : n \in \mathbb{N}\} < 0$, where $\Xi_t := \{A\,\mathrm{diag}[(R^{1/2}\eta_t)^{\odot 2}] + B\}$. Here γ is the so-called top Lyapunov exponent of the sequence $(\Xi_t : t \in \mathbb{Z})$. Notice that an ECCC-GARCH process satisfying this strict stationarity condition may not have any finite (high-order) moments. In Section 3 it will be assumed that X_t has finite sixth-order moments when the asymptotic distribution of the QMLE is derived, and hence it is useful to have conditions on the distribution of η_t and θ ensuring these moment restrictions. Such conditions can be found in Lemmas B.7 and B.8 in Appendix B containing novel results for the ECCC-GARCH process. Specifically, from Lemma B.7 if for some $p \in \mathbb{N}$ it holds that η_t has a strictly positive density on $\mathbb{R}^d$ with $\mathbb{E}[\|(\eta_t^{\odot 2})^{\otimes p}\|] < \infty$, if the diagonal elements of A_0 are strictly positive, and if $\rho(\mathbb{E}[(\Xi_t)^{\otimes p}]) < 1$, with $\rho(\cdot)$ denoting the spectral radius, then $(X_t : t \in \mathbb{Z})$ is geometrically β-mixing with $\mathbb{E}[\|(X_t^{\odot 2})^{\otimes p}\|] < \infty$. Moreover, from Lemma B.8, $\mathbb{E}[\|(X_t^{\odot 2})^{\otimes p}\|] < \infty$ implies that $\rho(\mathbb{E}[(\Xi_t)^{\otimes p}]) < 1$.

2.3 Volatility spillovers and Granger noncausality

The main objective of this paper is to consider tests concerning spillovers in ECCC-GARCH processes. As clarified below, volatility spillovers (or interactions) are quantified by the off-diagonal elements of the matrices A and B, and thereby testing for spillovers relies on testing if certain of the off-diagonal elements of A and B are equal to zero.

Consider, as an example, the bivariate process with $X_t := (X_{t,1}, X_{t,2})'$ and

$$h_t = \begin{pmatrix} h_{t,1} \\ h_{t,2} \end{pmatrix} = \begin{pmatrix} \kappa_1 + A_{11}X_{t-1,1}^2 + A_{12}X_{t-1,2}^2 + B_{11}h_{t-1,1} + B_{12}h_{t-1,2} \\ \kappa_2 + A_{21}X_{t-1,1}^2 + A_{22}X_{t-1,2}^2 + B_{21}h_{t-1,1} + B_{22}h_{t-1,2} \end{pmatrix}.$$

Here the coefficients A_{12} and A_{21} quantify the effects of the past squared shocks $X_{t-1,2}^2$ and $X_{t-1,1}^2$ on the conditional variances $h_{t,1}$ and $h_{t,2}$, respectively. These effects are often referred to as the ARCH spillovers, see e.g. Conrad and Weber (2013). Likewise, the coefficients B_{12} and B_{21} measure the GARCH spillovers from the conditional variances $h_{t-1,2}$ and $h_{t-1,1}$ to $h_{t,1}$ and $h_{t,2}$, respectively.

Remark 2.2. As discussed in Conrad and Karanasos (2010) and Nakatani and Teräsvirta (2008), when considering the ECCC-GARCH model one may allow some of the off-diagonal elements of A and B to be negative, and thereby introduce the notion of negative volatility spillovers, see also Section 2.3. To our knowledge the large-sample behavior of the QMLE is unknown when allowing for such negative parameter values, and we do not allow for such (milder) parameter restrictions in this paper.

Intuitively, the spillovers characterize some of the dependence between $X_{t,1}$ and

$X_{t,2}$, and, as explained next, the spillovers are closely related to Granger causality. With $\mathcal{F}_t^X := \sigma(X_s : s \leq t)$ and $\mathcal{F}_t^{X_1} := \sigma(X_{s,1} : s \leq t)$, we consider the following notion of second-order Granger noncausality, introduced by Granger et al. (1986): $X_{t,2}$ is said not to second-order Granger cause $X_{t,1}$ (with respect to $\mathcal{F}_{t-1}^X$) if

$$\mathbb{E}\{(X_{t,1} - \mathbb{E}[X_{t,1}|\mathcal{F}_{t-1}^X])^2|\mathcal{F}_{t-1}^X\} - \mathbb{E}\{(X_{t,1} - \mathbb{E}[X_{t,1}|\mathcal{F}_{t-1}^X])^2|\mathcal{F}_{t-1}^{X_1}\} \quad = \quad 0 \ a.s. \ \forall t \in \mathbb{Z}.$$

If the quantity on the left-hand side is nonzero (with strictly positive probability) then $X_{t,2}$ is said to second-order Granger cause $X_{t,1}$.

Suppose that $(X_t : t \in \mathbb{Z})$ is strictly stationary, which implies that $\rho(B) < 1$ (Francq and Zakoïan, 2010, pp.290-291). Then

$$h_t = (I_2 - B)^{-1}\kappa + \sum_{i=0}^{\infty}(B^i A)X_{t-1-i}^{\odot 2}.$$

It holds that $\mathbb{E}[X_{t,1}|\mathcal{F}_{t-1}^X] = 0$ a.s., so that $\mathbb{E}\{(X_{t,1} - \mathbb{E}[X_{t,1}|\mathcal{F}_{t-1}^X])^2|\mathcal{F}_{t-1}^X\} = h_{t,1}$ a.s. Hence, in light of the above definition, $X_{t,2}$ does not second-order Granger cause $X_{t,1}$ if $h_{t,1} = \mathbb{E}[h_{t,1}|\mathcal{F}_{t-1}^{X_1}]$ a.s. which is the case if $B_{12} = A_{12} = 0$. These restrictions on the matrices A and B thereby yield a sufficient condition for $X_{t,2}$ not to second-order Granger cause $X_{t,1}$. Likewise, $X_{t,1}$ does not second-order Granger cause $X_{t,2}$ if $B_{21} = A_{21} = 0$, and we have that there is no second-order causation in the process if A and B are diagonal. Notice that the above definition of Granger causality differs from, and is simpler than, the original notion of Granger causality stated in terms of the conditional distribution of $X_{t,1}$, see e.g. Granger (1969) and Engle et al. (1983). However, for practical purposes the above definition is much more operational, as discussed in e.g. Granger (1980, Section 3). We refer to Comte and Lieberman (2000), Hafner and Herwartz (2008), and Woźniak (2015) for additional considerations about Granger causality in multivariate GARCH processes.

3 Estimation and large-sample properties of the QMLE

In the following we consider inference in the ECCC-GARCH model where we allow elements of A and B to be equal to zero. Throughout the remainder of the paper, let $\partial f(\theta)/\partial\theta$ denote the vector of left/right partial derivatives of the function $f : \Theta \to \mathbb{R}$ with respect to the vector θ, and let $\partial^2 f(\theta)/\partial\theta\partial\theta'$ denote the matrix of left/right second-order partial derivatives as defined in Andrews (1999, pp.1350-1351).

Given a realization $(X_t : t = 0, 1, ..., T)$ of the ECCC-GARCH model, the QMLE,

$\hat{\theta}_T$, of θ is defined as

$$\hat{\theta}_T = \arg\inf_{\theta\in\Theta} \hat{L}_T(\theta),$$

with the feasible log-likelihood function, $\hat{L}_T(\theta)$, given by

$$\hat{L}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}\hat{l}_t(\theta), \tag{3.1}$$

$$\hat{l}_t(\theta) := \log\left[\det\left(\hat{H}_t(\theta)\right)\right] + X_t'\hat{H}_t^{-1}(\theta)X_t, \tag{3.2}$$

$$\hat{H}_t(\theta) := \hat{D}_t(\theta)R(\theta)\hat{D}_t(\theta),$$

$$\hat{D}_t^2(\theta) := \operatorname{diag}\left(\hat{h}_t(\theta)\right), \tag{3.3}$$

$$\hat{h}_t(\theta) := \kappa + AX_{t-1}^{\odot 2} + B\hat{h}_{t-1}(\theta), \tag{3.4}$$

with $\hat{h}_0(\theta) = \hat{h}_0 \in (0,\infty)^d$ fixed. Next, we consider the asymptotic properties of the QMLE.

For the probability analysis of the QMLE we let θ_0 denote the true parameter vector such that $X_t := X_t(\theta_0)$. The derivation of the limiting distribution of the QMLE relies on the following assumptions.

Assumption 1. *$\theta_0 \in \Theta$ and Θ is compact.*

Assumption 2. *The sequence $(X_t : t \in \mathbb{Z})$ is strictly stationary and ergodic.*

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

In light of Assumption 2, consider the (infeasible) ergodic version of the log-likelihood function, i.e. for the strictly stationary and ergodic sequence $(X_t : t \in \mathbb{Z})$ we define for $t \in \mathbb{Z}$ and $\theta \in \Theta$,

$$L_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} l_t(\theta)$$

$$l_t(\theta) := \log\left[\det\left(H_t(\theta)\right)\right] + X_t'H_t^{-1}(\theta)X_t \tag{3.5}$$

$$H_t(\theta) := D_t(\theta)R(\theta)D_t(\theta)$$

$$D_t^2(\theta) := \operatorname{diag}\left(h_t(\theta)\right) \tag{3.6}$$

$$h_t(\theta) := \kappa + AX_{t-1}^{\odot 2} + Bh_{t-1}(\theta). \tag{3.7}$$

Assumption 4. *For $\theta \in \Theta$, $\{h_t(\theta) = h_t(\theta_0)$ a.s. and $R = R_0\}$ implies that $\theta = \theta_0$.*

Remark 3.1. Assumption 4 is a high-level identification condition. Primitive conditions are discussed in e.g. Jeantheau (1998), Ling and McAleer (2003), and Francq and Zakoïan (2010, 2012a). In particular, for the simulation study in Section 5, all data generating processes can be shown to be minimal in the sense of Jeantheau (1998, Definition 3.3) which (under some additional mild regularity conditions) ensures identification.

Remark 3.2. The above assumptions are standard and imply that $\hat{\theta}_T = \theta_0 + o(1)$ almost surely. If one additionally assumes that $\theta_0 \in \overset{\circ}{\Theta}$, i.e. θ_0 is an interior point of Θ, and that η_t has finite fourth moments, then $\sqrt{T}(\hat{\theta}_T - \theta_0)$ has a Gaussian limit with zero mean and covariance $J^{-1}\Sigma J^{-1}$ with J and Σ given in (3.13) below. Both results are established in Francq and Zakoïan (2012a).

As mentioned, we are interested in the case where θ_0 is not an interior point of Θ, and our main interest is to allow for the case where some of the elements of A_0 and B_0 are equal to zero. Let β denote the $(s_1 \times 1)$ vector containing the $s_1 \geq 0$ elements of A and B that take value zero under the null, i.e. with true parameter value equal to zero, and let δ denote the $(s_2 \times 1)$ vector of the remaining $s_2 := (s_0 - s_1)$ parameters of θ. Without loss of generality we consider throughout the remainder of the paper a reparametrized version of the ECCC-GARCH model such that

$$\underset{(s_0\times 1)}{\theta} = \begin{pmatrix} \underset{(s_1\times 1)}{\beta} \\ \underset{(s_2\times 1)}{\delta} \end{pmatrix}, \tag{3.8}$$

and with Θ defined accordingly. Notice that for the case where $s_1 = 0$, we have that $\theta = \delta$. We also consider accordingly a partition of the true parameter value $\theta_0 = (\beta_0', \delta_0')'$, and by definition $\beta_0 = 0_{s_1 \times 1}$. For the case $s_1 > 0$, with the QMLE $\hat{\theta}_T = (\hat{\beta}_T', \hat{\delta}_T')'$, it holds that $\sqrt{T}(\hat{\beta}_T - \beta_0) = \sqrt{T}\hat{\beta}_T \in [0, \infty)^{s_1}$ which cannot have a Gaussian limit. Hence the theory for the QMLE for the case where θ_0 is an interior point, as described in Remark 3.2, is no longer applicable. We deal with the boundary problem by making two additional Assumptions 5 and 6.

First, we make the following assumption about θ_0 and Θ.

Assumption 5. *The set* $\Theta - \theta_0$ *is locally equal to* $\Lambda := \Lambda_\beta \times \Lambda_\delta = \mathbb{R}_+^{s_1} \times \mathbb{R}^{s_2}$*, i.e. there exists an* $\epsilon > 0$ *such that* $\Lambda \cap C(0, \epsilon) = \Theta \cap C(0, \epsilon)$*, where* $C(x, \epsilon) \subset \mathbb{R}^d$ *denotes an open cube centered at* $x \in \mathbb{R}^d$ *and with side length* 2ϵ*.*

Remark 3.3. Assumption 5 is essentially a special case of Assumption 2^{2*} in Andrews (1999, 2001) and has several purposes. First, it prevents the true parameter value δ_0 from reaching the bounds of Θ, which keeps things as simple as possible, as our main interest is to consider hypotheses where elements of β are equal to zero (i.e. take value at the lower bound of Θ). Second, this assumption allows us to make a Taylor-type expansion based on left/right partial derivatives of

the log-likelihood function around θ_0, see Andrews (1999, Appendix A) for details. Moreover, the assumption is important for approximating the quantity $\sqrt{T}(\hat{\theta}_T - \theta_0)$, see specifically the proof of Theorem 3.1 in the appendix. Although the assumption imposes additional structure on the parameter space it is compatible with the parameter restrictions given in Assumption 3. As in Francq and Zakoïan (2007), let $\theta_0(\epsilon)$ be defined as the vector obtained by replacing all zero elements of θ_0 by $\epsilon > 0$. For some sufficiently small ϵ, $\theta_0(\epsilon)$ belongs to the interior of Θ. Consider the case where B_0 is diagonal. Provided that Assumptions 1 and 3 hold, $\rho(B_0) < 1$. For a real $m \times m$ matrix with non-negative entries, it holds that $C = [C_{ij}] \geq 0$, $\rho(C) \leq \min \left\{ \max_{i=1,\ldots,m} \sum_{j=1}^m C_{ij}, \max_{j=1,\ldots,m} \sum_{i=1}^m C_{ij} \right\}$. Hence for a sufficiently small $\epsilon > 0$, $\rho(B_{0\epsilon}) < 1$, where $B_{0\epsilon}$ is B evaluated at $\theta_0(\epsilon)$.
Another example is the bivariate case where

$$B_0 = \begin{pmatrix} B_{11,0} & 0 \\ B_{21,0} & B_{22,0} \end{pmatrix},$$

and $B_{11,0}$ and $B_{22,0}$ are strictly positive. Here the eigenvalues of $B_{0\epsilon}$ are

$$\frac{1}{2}(B_{11,0} + B_{22,0}) \pm \frac{1}{2}\sqrt{(B_{11,0} - B_{22,0})^2 + 4B_{21,0}\epsilon}.$$

Since $\rho(B_0) < 1$ we know that $B_{11,0}$ and $B_{22,0}$ are strictly less than one, so for a sufficiently small $\epsilon > 0$, $\rho(B_{0\epsilon}) < 1$.

Second, deriving the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ typically relies on, among other things, verifying a condition such as

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left| \frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} \right| \right] < \infty \tag{3.9}$$

or, given that θ_0 is an interior point, i.e. $\theta_0 \in \mathring{\Theta}$,

$$\mathbb{E}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^3 l_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \right] < \infty$$

for all $i, j, k = 1, ..., s_0$ and for some neighborhood $\mathcal{V}(\theta_0)$ around θ_0. With $h_{t,i_1}(\theta)$ denoting element i_1 of $h_t(\theta)$, the latter condition it usually verified by showing that

$$\mathbb{E}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{1}{h_{t,i_1}(\theta)} \frac{\partial h_{t,i_1}(\theta)}{\partial \theta_i} \right|^3 \right] < \infty \tag{3.10}$$

for all $i_1 = 1, .., d$ and all $i = 1, ..., s_0$, and a similar property with $\partial h_{t,i_1}(\theta) / \partial \theta_i$

replaced with $\partial^2 h_{t,i_1}(\theta)/\partial\theta_i\partial\theta_j$ and $\partial^3 h_{t,i_1}(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$. Consider, for simplicity, the case with $B = 0_{2\times 2}$, i.e. with no GARCH effects. Then

$$h_t(\theta) = \begin{pmatrix} h_{t,1}(\theta) \\ h_{t,2}(\theta) \end{pmatrix} = \begin{pmatrix} \kappa_1 + A_{11}X_{t-1,1}^2 + A_{12}X_{t-1,2}^2 \\ \kappa_2 + A_{21}X_{t-1,1}^2 + A_{22}X_{t-1,2}^2 \end{pmatrix},$$

and hence

$$\frac{1}{h_{t,1}(\theta)}\frac{\partial h_{t,1}(\theta)}{\partial A_{12}} = \frac{X_{t-1,2}^2}{\kappa_1 + A_{11}X_{t-1,1}^2 + A_{12}X_{t-1,2}^2}. \tag{3.11}$$

For the case where $\theta_0 \in \overset{\circ}{\Theta}$, one can choose $\mathcal{V}(\theta_0)$ such that the elements of A are bounded away from zero on $\mathcal{V}(\theta_0)$, see Francq and Zakoïan (2012a, pp.199-202). This implies that the fraction in (3.11) is bounded by $\sup_{\theta\in\mathcal{V}(\theta_0)} A_{12}^{-1}$ on $\mathcal{V}(\theta_0)$ and hence that all moments are finite. However, such argument cannot be applied to bound the moments of the derivatives of the log-likelihood function in the case where some of the elements of A_0 can take zero value. Suppose additionally that A_0 is diagonal, then

$$\frac{1}{h_{t,1}(\theta_0)}\frac{\partial h_{t,1}(\theta_0)}{\partial A_{12}} = \frac{X_{t-1,2}^2}{\kappa_{1,0} + A_{11,0}X_{t-1,1}^2},$$

which is not bounded by a constant. The asymptotic properties derived in this paper rely on establishing condition (3.9), which is done by imposing the condition that $\mathbb{E}[\|X_t\|^6] < \infty$, similar to Francq and Zakoïan (2007, Assumption A7).

Assumption 6. $\mathbb{E}[\|X_t\|^6] < \infty$.

Remark 3.4. As mentioned in Subsection 2.2, Lemmas B.7-B.8 provide necessary and sufficient conditions for Assumption 6 to hold.

We are now able to state the limiting distribution of the QMLE.

Theorem 3.1. *Under Assumptions 1-6,*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{w}{\to} \lambda^\Lambda \tag{3.12}$$

where $\lambda^\Lambda = \arg\inf_{\lambda\in\Lambda} \|Z-\lambda\|_J^2$, *with* $\|Z-\lambda\|_J^2 := (Z-\lambda)'J(Z-\lambda)$, *and where* Λ *is defined in Assumption 5,* Z *is a random vector with distribution* $\mathcal{L}(Z) = N(0, J^{-1}\Sigma J^{-1})$, *and*

$$J := \mathbb{E}[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'] \in \mathbb{S}_{++}^{s_0}, \quad \Sigma := \mathbb{E}[(\partial l_t(\theta_0)/\partial\theta)(\partial l_t(\theta_0)/\partial\theta')]. \tag{3.13}$$

The theorem states that the limiting distribution of the normalized QMLE is given by λ^Λ which by definition is the projection of the $N(0, J^{-1}\Sigma J^{-1})$-distributed

Z onto the set Λ with respect to the metric induced by the inner product $\langle\cdot,\cdot\rangle_J$, where we recall that for $x, y \in \mathbb{R}^{s_0}$, $\langle x, y\rangle_J = x'Jy$. Since Λ is convex, it holds that λ^Λ is unique. In the case where θ_0 is not a boundary point, $s_1 = 0$, such that $\Lambda = \mathbb{R}^{s_0}$ and the limiting distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is Z, as mentioned in Remark 3.2. Notice that the matrices J and Σ are stated in terms of left/right-derivatives, as discussed in Andrews (1999, Appendix A). Moreover, Andrews (1999, pp.1367-1370) provides closed-form expressions for λ^Λ, and gives an outline of how to make draws of the distribution of λ^Λ based on numerical methods. The next section is devoted to testing hypotheses about the parameters in A and B.

4 Testing

In this section we introduce Lagrange multiplier, Wald, and likelihood ratio statistics suitable for testing hypotheses about the matrices A and B. In particular, we are interested in testing for no volatility spillovers, as discussed in Subsection 2.3. Subsection 4.1 states the test statistics and their limiting distributions. In Subsection 4.2 we provide an algorithm for determining critical values for the proposed tests.

4.1 Test statistics

We consider testing hypotheses where some of the parameters in the matrices A and B take zero value. With β defined according to the partition in (3.8), we consider the partition of β given by

$$\underset{(s_1\times 1)}{\beta} = \begin{pmatrix} \underset{(\tilde{s}_1\times 1)}{\beta_1} \\ \underset{(\tilde{s}_2\times 1)}{\beta_2} \end{pmatrix} \tag{4.1}$$

for some $\tilde{s}_1 \leq s_1$ and $\tilde{s}_2 := s_1 - \tilde{s}_1$. Note that, by convention, $\beta = \beta_1$ when $\tilde{s}_1 = s_1$. We are interested in testing whether β_1 takes value zero, i.e. in terms of the true parameter value $\theta_0 = (\beta_0', \delta_0')' = (\beta_{1,0}', \beta_{2,0}', \delta_0')'$, we want to test the hypothesis

$$\mathcal{H}_0: \quad \beta_{1,0} = 0_{\tilde{s}_1\times 1}.$$

We test $\mathcal{H}_0$ against the alternative $\beta_{1,0} \neq 0_{\tilde{s}_1\times 1}$ and with the maintained hypothesis that $\theta_0 \in \Theta$. Notice that under $\mathcal{H}_0$ it might be that some of the remaining parameters of A and B are equal to zero, which is the case when $\tilde{s}_2 = s_1 - \tilde{s}_1 > 0$, and we may consider β_2 as nuisance parameters attaining the zero bound of Θ under $\mathcal{H}_0$.

With $\hat{L}_T(\theta)$ the feasible log-likelihood function defined in (3.1), let $\tilde{\theta}_T$ be the constrained estimator given by

$$\tilde{\theta}_T = \arg\inf_{\theta\in\Theta_0} \hat{L}_T(\theta), \quad \text{with} \quad \Theta_0 := \{\theta = (\beta_1', \beta_2', \delta')' \in \Theta : \beta_1 = 0_{\tilde{s}_1\times 1}\}. \tag{4.2}$$

We propose three statistics for testing $\mathcal{H}_0$. The first statistic is a likelihood ratio statistic,

$$QLR_T := 2T[\hat{L}_T(\tilde{\theta}_T) - \hat{L}_T(\hat{\theta}_T)].$$

Next, let

$$\hat{J}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \hat{l}_t(\theta)}{\partial\theta\partial\theta'}, \quad \hat{\Sigma}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \hat{l}_t(\theta)}{\partial\theta}\frac{\partial \hat{l}_t(\theta)}{\partial\theta'}, \quad \hat{S}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \hat{l}_t(\theta)}{\partial\theta}. \tag{4.3}$$

Moreover, with s_0 the dimension of the parameter vector θ, s_1 the dimension β given in (3.8), and $\tilde{s}_1$ the dimension of the vector β_1 defined in (4.1), let

$$K := (I_{s_1}, 0_{s_1\times(s_0-s_1)}) \quad \text{and} \quad K_1 := (I_{\tilde{s}_1}, 0_{\tilde{s}_1\times(s_0+\tilde{s}_1)}). \tag{4.4}$$

The second statistic is the Wald statistic,

$$W_T := T\hat{\theta}_T' K_1'[K_1\hat{J}_T(\hat{\theta}_T)^{-1}K_1']^{-1}K_1\hat{\theta}_T,$$

and the last statistic is a Lagrange multiplier (LM) statistic,

$$LM_T := T\hat{S}_T(\tilde{\theta}_T)'\hat{J}_T(\tilde{\theta}_T)^{-1}K_1'[K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{\Sigma}_T(\tilde{\theta}_T)\hat{J}_T(\tilde{\theta}_T)^{-1}K_1']^{-1}K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{S}_T(\tilde{\theta}_T).$$

Remark 4.1. In addition to the QLR_T and W_T statistics, one could also consider a directed Lagrange multiplier statistic, that exploits that the true parameter is on the boundary under the null, similar to Andrews (2001, Section 7). We focus here on the first two statistics together with the "classical" Lagrange multiplier statistic, LM_T, that, although it is based on partial left/right derivatives, does not take any boundary issues into account.

In order to derive the limiting distribution of these statistics, we assume, similar to Assumption 3, that θ_0 and Θ_0 satisfy the following conditions.

Assumption 7. *$\theta_0 \in \Theta_0$ and $\Theta_0 - \theta_0$ is locally equal to* $\Lambda_0 := \Lambda_{0,\beta_1} \times \Lambda_{\beta_2} \times \Lambda_\delta = \{0_{\tilde{s}_1\times 1}\} \times \mathbb{R}_+^{\tilde{s}_2} \times \mathbb{R}^{s_2}$.

Similar to λ^Λ defined in Theorem 3.1, we consider λ^{Λ_0} as the projection of random vector Z with distribution $N(0, J^{-1}\Sigma J^{-1})$ onto Λ_0, i.e.

$$\lambda^{\Lambda_0} = (\lambda_\beta^{\Lambda_0 \prime}, \lambda_\delta^{\Lambda_0 \prime})' \in \Lambda_0 \text{ satisfies } \lambda^{\Lambda_0} = \arg\inf_{\lambda \in \Lambda_0} \|Z - \lambda\|_J^2 . \tag{4.5}$$

The following theorem states the limiting distributions of the proposed test statistics.

Theorem 4.1. *Let the matrices K and K_1 be given by (4.4), and let J be given by (3.13). Under Assumptions 1-7 and $\mathcal{H}_0$,*

$$QLR_T \xrightarrow{w} \|\lambda_\beta^\Lambda\|^2_{(KJ^{-1}K')^{-1}} - \|\lambda_\beta^{\Lambda_0}\|^2_{(KJ^{-1}K')^{-1}}, \tag{4.6}$$

where $\lambda^\Lambda = (\lambda_\beta^{\Lambda\prime}, \lambda_\delta^{\Lambda\prime})' = (\lambda_{\beta_1}^{\Lambda\prime}, \lambda_{\beta_2}^{\Lambda\prime}, \lambda_\delta^{\Lambda\prime})'$ is defined in Theorem 3.1, and $\lambda_\beta^{\Lambda_0}$ is defined in (4.5).
Moreover,

$$W_T \xrightarrow{w} \|\lambda_{\beta_1}^\Lambda\|^2_{(K_1 J^{-1} K_1')^{-1}}. \tag{4.7}$$

Suppose in addition that Σ, defined in (3.13), is positive definite. Then

$$LM_T \xrightarrow{w} \chi^2_{\tilde{s}_1}, \tag{4.8}$$

where $\chi^2_{\tilde{s}_1}$ is a chi-squared random variable with $\tilde{s}_1$ degrees of freedom, with $\tilde{s}_1$ the dimension of β_1.

Remark 4.2. Theorem 4.1 states that the limiting distribution of the QLR_T depends on the minimizer of the quadratic form $\|Z - \lambda\|_J^2$ over Λ and Λ_0, respectively. From Lemma B.6 it holds that $\inf_{\lambda \in \Lambda} \|Z - \lambda\|_J^2 = \inf_{\lambda_\beta \in \Lambda_{\beta_1} \times \Lambda_{\beta_2}} \|Z_\beta - \lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$, and by similar arguments $\inf_{\lambda \in \Lambda_0} \|Z - \lambda\|_J^2 = \inf_{\lambda_\beta \in \{0_{\tilde{s}_1 \times 1}\} \times \Lambda_{\beta_2}} \|Z_\beta - \lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$, where Z_β is defined from the partition $Z = (Z_\beta', Z_\delta')'$. Thereby the limiting distribution of QLR_T depends in general on the cone Λ_{β_2}, i.e. whether there are nuisance parameters (in A_0 and B_0) taking zero value. A similar observation applies to W_T, as $\lambda_{\beta_1}^\Lambda$ is a part of λ^Λ and hence requires knowledge about the shape of Λ. This issue appears to be an important topic within the field of testing at the boundary. We refer to Ketz (2014) for some recent considerations regarding hypothesis tests regarding a single parameter at the boundary with nuisance parameters potentially taking values on the boundary of the parameter space.

Remark 4.3. Unlike the QLR_T and W_T statistics, the limiting distribution of the LM_T statistic is pivotal and does not depend on nuisance parameters.

Remark 4.4. In the case where $(KJ^{-1}K')^{-1}$ is block diagonal, i.e. $K_2(KJ^{-1}K')^{-1}\bar{K}_2' = 0_{\tilde{s}_1 \times \tilde{s}_2}$ where $K_2 := (I_{\tilde{s}_1}, 0_{\tilde{s}_1 \times \tilde{s}_2})$ and $\bar{K}_2 := (0_{\tilde{s}_2 \times \tilde{s}_1}, I_{\tilde{s}_2})$, it can be shown (by applying the arguments from Remark 4.2 and the proof of Lemma B.6) that, with $Z_\beta = (Z'_{\beta_1}, Z'_{\beta_2})'$,

$$\inf_{\lambda_\beta \in \Lambda_{\beta_1} \times \Lambda_{\beta_2}} \|Z_\beta - \lambda_\beta\|^2_{(KJ^{-1}K')^{-1}} = \inf_{\lambda_{\beta_1} \in \Lambda_{\beta_1}} \|Z_{\beta_1} - \lambda_{\beta_1}\|^2_{K_2(KJ^{-1}K')^{-1}K_2'} + \inf_{\lambda_{\beta_2} \in \Lambda_{\beta_2}} \|Z_{\beta_2} - \lambda_{\beta_2}\|^2_{\bar{K}_2(KJ^{-1}K')^{-1}\bar{K}_2'}.$$

This implies that the limiting distributions of W_T and QLR_T do not depend on Λ_2 and thereby not on whether the nuisance parameters take zero value. In particular we have that

$$QLR_T \xrightarrow{w} \|\lambda^{\Lambda_{\beta_1}}\|^2_{K_2(KJ^{-1}K')^{-1}K_2'},$$

with $\lambda^{\Lambda_{\beta_1}} = \arg\inf_{\lambda_{\beta_1} \in \Lambda_{\beta_1}} \|Z_{\beta_1} - \lambda_{\beta_1}\|^2_{K_2(KJ^{-1}K')^{-1}K_2'}$. Moreover, for this case the limiting distribution of W_T is given by $\|\lambda^{\Lambda_{\beta_1}}\|^2_{(K_1J^{-1}K_1')^{-1}}$. Notice that the block diagonality property of $(KJ^{-1}K')^{-1}$ does not appear to hold in general.

The following corollary is immediate from Theorem 4.1 and states that the limiting distributions of QLR_T and W_T are the same in the case where there are no nuisance parameters (in A and B) taking zero value.

Corollary 4.1. *Under the same assumptions as in Theorem 4.1, suppose that $\tilde{s}_1 = s_1$ such that $\beta_{1,0} = \beta_0 = 0_{s_1 \times 1}$, i.e there are no nuisance parameters on the lower bound of Θ. Then the limiting distributions of QLR_T and W_T are both given by $\|\lambda^{\Lambda}_\beta\|^2_{(KJ^{-1}K')^{-1}}$.*

Remark 4.5. In the context of testing for diagonality of A_0 and B_0, and under the assumption that the innovations are Gaussian, i.e. $\mathcal{L}(\eta_t) = N(0, I_d)$, Nakatani and Teräsvirta (2009) propose the LM statistic,

$$LM_{ECCC} = \frac{1}{2} T \hat{S}_T(\tilde{\theta}_T)' K_1' [K_1 \hat{J}_T(\tilde{\theta}_T)^{-1} K_1'] K_1 \hat{S}_T(\tilde{\theta}_T).$$

Nakatani and Teräsvirta (2009) derive the limiting distribution of this statistic under the assumptions that the elements of A and B are nonnegative (Nakatani and Teräsvirta, 2009, footnote on p.149), similar to our assumption about the parameter space Θ, and that the true parameter vector is an interior point of the parameter space (Nakatani and Teräsvirta, 2009, Assumption 3.1). In Proposition C.1 in the appendix we state the limiting distribution of the LM_{ECCC} statistic under the same assumptions as in Theorem 4.1. Specifically, provided that $\mathcal{L}(\eta_t) = N(0, I_d)$, and that $\tilde{s}_1 = s_1$, the LM_{ECCC} statistic has an asymptotic $\chi^2_{\tilde{s}_1}$ distribution. In the more general cases where $s_1 - \tilde{s}_1 > 0$, i.e. with nuisance parameters attaining the zero bound of Θ, and where η_t may not be Gaussian, the limiting distribution will not be $\chi^2_{\tilde{s}_1}$, as also stated in Proposition C.1.

In the next section we provide an algorithm for calculating critical values for the proposed tests for the case with no nuisance parameters in A and B taking zero value.

4.2 Calculating critical values

Following Andrews (1999, pp.1367-1370), we can obtain draws from the limiting distribution of the W_T and QLR_T statistics according to the following algorithm.[2]

Algorithm 1. *Let $\bar{J}_T$ and $\bar{\Sigma}_T$ be consistent estimators for, respectively, the matrices J and Σ stated in (3.13). Suppose that $\tilde{s}_1 = s_1$, i.e. there are no nuisance parameters (in A and B) taking zero value, such that Corollary 4.1 applies. A critical value c for W_T and QLR_T yielding a test with asymptotic size α can be obtained as follows:*

1. *Draw $\varepsilon^\star$ randomly from $N(0, I_{s_1})$ and compute $Z^\star_\beta = [K\bar{J}_T^{-1}\bar{\Sigma}_T\bar{J}_T^{-1}K']^{1/2}\varepsilon^\star$.*

2. *Find $\tilde{\lambda}^\star_\beta$ that minimizes $\|Z^\star_\beta - \lambda_\beta\|^2_{(K\bar{J}_T^{-1}K')^{-1}} = (Z^\star_\beta - \lambda_\beta)'(K\bar{J}_T^{-1}K')^{-1}(Z^\star_\beta - \lambda_\beta)$ over $\lambda_\beta \in \Lambda_\beta = \mathbb{R}^{s_1}_+$, and compute $\|\tilde{\lambda}^\star_\beta\|^2_{(K\bar{J}_T^{-1}K')^{-1}}$.*

3. *Repeat steps 1.-2. N times (with N very large), and let $\{x_i : i = 1, ..., N\}$ denote the sequence of the N independent draws of $\|\tilde{\lambda}^\star_\beta\|^2_{(K\bar{J}_T^{-1}K')^{-1}}$. Then c is given by the $(1-\alpha)$ percentile of $\{x_i : i = 1, ..., N\}$.*

Remark 4.6. The minimization problem in point 2. of Algorithm 1 is a quadratic programming problem. Most programming languages have a build-in function that can deal with such problems, and for a fairly small amount of restrictions, i.e. for small s_1, the minimization is solved quickly. For the simulations and the empirical illustration in the following sections, the minimization problem is carried out by the `solveQP` function in OxMetrics 7.0. An alternative way of making draws of λ^Λ_β, and hence drawing from the distribution of $\|\lambda^\Lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$, is given by Andrews (1999, Section 6.3) where a closed-form expression for λ^Λ_β is provided. Moreover, throughout the simulations and the empirical illustration, we use $\hat{J}_T(\hat{\theta}_T)$ and $\hat{\Sigma}_T(\hat{\theta}_T)$ as estimators for J and Σ, respectively, where $\hat{\theta}_T$ is the QMLE and $\hat{J}_T(\theta)$ and $\hat{\Sigma}_T(\theta)$ are defined in (4.3). These estimators are consistent according to Lemma B.1.

[2] We here use Lemma B.6 stating that λ^Λ_β is equal to λ^{Λ_β}, $\lambda^{\Lambda_\beta} = \arg\inf_{\lambda_\beta \in \Lambda_\beta} \|Z_\beta - \lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$, where $\Lambda_\beta = \mathbb{R}^{s_1}_+$.

5 Simulations

In this section we investigate the empirical size and power properties of the proposed test statistics.

5.1 Size simulations

We consider the size properties of the proposed test statistics, including the LM_{ECCC} mentioned in Remark 4.5, for the bivariate ECCC-GARCH model. Specifically, we consider tests where the matrices A and B are diagonal under the null. In order to keep things simple we consider cases where no nuisance parameters attain the lower bound of Θ, i.e. none of the diagonal elements of A and B take zero value. We consider the data-generating processes (DGPs) stated in Table 5.1, where DGP 1-3 correspond to DGP 1,2, and 4 in Nakatani and Teräsvirta (2009), respectively. Recall from Theorem 3.1 that we imposed finite sixth-order moments of X_t (Assumption 6) in order to derive the limiting distribution of the QMLE. For all the DGPs we impose, for simplicity, that the innovation η_t is Gaussian. This condition implies that η_t has a strictly positive density on $\mathbb{R}^d$ with $\mathbb{E}[\|\eta_t\|^6] < \infty$, and hence from Lemmas B.7 and B.8, $\mathbb{E}[\|X_t\|^6] < \infty$ if and only if

$$\Psi_6 := \rho(\mathbb{E}\{[A_0 \operatorname{diag}((R_0^{1/2}\eta_t)^{\odot 2}) + B_0]^{\otimes 3}\}) < 1. \tag{5.1}$$

Using Monte Carlo integration we have computed the value of Ψ_6 for each DGP, as also stated in Table 5.1. Whereas DGP 3-5 satisfy condition (5.1), DGP 1-2 do not. Although our theoretical results are not expected to hold for DGP 1 and 2, we have included the simulations results in order to compare with the results for the DGPs that do satisfy the moment condition. Moreover, for all the DGPs for the empirical size and power simulations it holds that the conditions in Jeantheau (1998, Definition 3.1.3 and Assumptions B1-B2) are satisfied, which by Jeantheau (1998, Proposition 3.4) implies that the identification condition in Assumption 4 holds. These above restrictions on the DGPs imply together that Corollary 4.1 holds for the processes (up to the sixth-order moment condition of X_t in the case of DGP 1-2).

Table 5.2 contains the actual rejection frequencies of our proposed tests based on the 5% nominal level and on empirically relevant sample sizes of 1,000, 5,000, and 10,000 observations. All simulations are based on 2,000 replications with a burn-in period of 1,000 observations. The critical value of the QLR_T and W_T tests are carried out according to Algorithm 1 and Remark 4.6. For each replication the

Table 5.1: DGPs for size simulations

	DGP 1	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**
A_0	$\begin{bmatrix}0.1 & 0\\ 0 & 0.2\end{bmatrix}$	$\begin{bmatrix}0.04 & 0\\ 0 & 0.05\end{bmatrix}$	$\begin{bmatrix}0.1 & 0\\ 0 & 0.2\end{bmatrix}$	$\begin{bmatrix}0.1 & 0\\ 0 & 0.2\end{bmatrix}$	$\begin{bmatrix}0.07 & 0\\ 0 & 0.08\end{bmatrix}$
B_0	$\begin{bmatrix}0.8 & 0\\ 0 & 0.7\end{bmatrix}$	$\begin{bmatrix}0.95 & 0\\ 0 & 0.9\end{bmatrix}$	$\begin{bmatrix}0.45 & 0\\ 0 & 0.6\end{bmatrix}$	$\begin{bmatrix}0.70 & 0\\ 0 & 0.75\end{bmatrix}$	$\begin{bmatrix}0.80 & 0\\ 0 & 0.85\end{bmatrix}$
r_0	0.3	0.9	0.9	0.9	0.9
Ψ_6	1.223	1.337	0.387	0.944	0.953

For all DGPs $\kappa_0 = (0.1, 0.2)'$ and $\mathcal{L}(\eta_t) = N(0, I_2)$.

critical value is based on 100,000 draws from $\|\lambda_\beta^\Lambda\|^2_{(KJ^{-1}K')^{-1}}$. The critical values for the LM_T and LM_{ECCC} are based on a χ^2_4-distribution, in line with Theorem 4.1 and Proposition C.1. We refer to Appendix D for additional technical details about the simulations.

Table 5.2: Size simulations

	T	LM_T	W_T	QLR_T	LM_{ECCC}
DGP 1	1,000	0.0277	0.0080	0.0375	0.1886
	5,000	0.0552	0.0325	0.0575	0.1068
	10,000	0.0436	0.0275	0.0460	0.0657
DGP 2	1,000	0.0194	0.0145	0.0315	0.2710
	5,000	0.0460	0.0490	0.0550	0.1090
	10,000	0.0505	0.0545	0.0610	0.0790
DGP 3	1,000	0.0277	0.0125	0.0385	0.1638
	5,000	0.0477	0.0270	0.0430	0.0974
	10,000	0.0455	0.0345	0.0460	0.0760
DGP 4	1,000	0.0353	0.0215	0.0430	0.1764
	5,000	0.0551	0.0405	0.0550	0.0966
	10,000	0.0455	0.0360	0.0465	0.0680
DGP 5	1,000	0.0137	0.0130	0.0300	0.2270
	5,000	0.0445	0.0310	0.0505	0.1036
	10,000	0.0390	0.0365	0.0445	0.0685

Actual rejection frequencies based on the 5% nominal level.

From Table 5.2 we notice that LM_T seems to be slightly under-sized for a sample size of 1,000 observations, whereas the test seems to have very reasonable size properties for larger sample sizes. The LM_{ECCC} test seems to be over-sized for sample sizes of 1,000 and 5,000 observations, but only slightly over-sized for 10,000 observations.[3] Moreover, the Wald test appears to be slightly conservative for most of the DGPs and in particular for sample sizes of 1,000 observations. The quasi-likelihood ratio test has very reasonable size properties, in particular for the cases with sample

[3]The rejection frequencies for the LM_{ECCC} test reported in Nakatani and Teräsvirta (2009, Table 2) seem more favorable than the ones reported in Table 5.2. A correspondence with Tomoaki Nakatani and a careful inspection of the R code used to generate the results in Nakatani and Teräsvirta (2009) have, unfortunately, not enabled us to detect the source of the difference.

sizes of 5,000 and 10,000 observations. Notice that even though the DGPs 1 and 2 do not satisfy the moment condition in (5.1), and hence that our derived theory is not expected to apply for these processes, the violation of the condition does not seem to have any severe effect on the performance of the tests. Lastly, in similar studies (not reported here) we investigated the size properties of the tests for the case of 50,000 observations, and when testing for the single restriction $B_{12} = 0$. These studies yielded qualitatively the same conclusions as the simulations reported above.

5.2 Power simulations

Next, we consider the power properties of the proposed tests. The power simulations are based on DGP 5 from the previous subsection, and we consider the data generating processes, deviating from the null of diagonality of the matrices A_0 and B_0, stated in Table 5.3. The DGPs are inspired by the ones used in Nakatani and Teräsvirta (2009, Table 3).

Table 5.3: DGPs for power simulations

	DGP 5.1	**DGP 5.2**	**DGP 5.3**	**DGP 5.4**
A_0	$\begin{bmatrix} 0.07 & 0.001 \\ 0.004 & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0.001 \\ 0.004 & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0.01 \\ 0.02 & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0.01 \\ 0.02 & 0.08 \end{bmatrix}$
B_0	$\begin{bmatrix} 0.80 & 0.004 \\ 0.002 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.80 & 0.04 \\ 0.03 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.80 & 0.004 \\ 0.002 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.80 & 0.04 \\ 0.03 & 0.85 \end{bmatrix}$
	DGP 5.5	**DGP 5.6**	**DGP 5.7**	**DGP 5.8**
A_0	$\begin{bmatrix} 0.07 & 0.001 \\ 0.004 & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0.01 \\ 0.02 & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0 \\ 0 & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0 \\ 0 & 0.08 \end{bmatrix}$
B_0	$\begin{bmatrix} 0.80 & 0 \\ 0 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.80 & 0 \\ 0 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.80 & 0.004 \\ 0.002 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.80 & 0.04 \\ 0.03 & 0.85 \end{bmatrix}$

For all DGPs $\kappa_0 = (0.1, 0.2)'$, $r_0 = 0.9$ and $\mathcal{L}(\eta_t) = N(0, I_2)$.

Table 5.4 states the rejection frequencies of the tests when the null is incorrect according to the DGPs given in Table 5.4. The simulations are based on 2,000 replications, a burn-in period of 1,000 observations, and the same seed values as the size simulations. The reported powers are size corrected in the sense that the critical value for the tests (at the 5% nominal level) is chosen as the 95 percentile of the simulated test values from the size simulations.

From Table 5.4 we see that the power of the tests is low whenever the off-diagonal elements of A and B are all close to zero. In particular, even for a sample size of 10,000 observations the power is not impressive for any of the test statistics for the DGPs 5.1, 5.5 and 5.7. For all other DGPs the test statistics seem to have great power as T increases. Moreover, the proposed Wald and likelihood ratio tests have

Table 5.4: Empirical power

	DGP 5.1				DGP 5.2			
T	LM_T	W_T	QLR_T	LM_{ECCC}	LM_T	W_T	QLR_T	LM_{ECCC}
1,000	.0493	.0898	.134	.0286	.254	.397	.619	.0654
5,000	.107	.278	.320	.0590	.983	.999	1.00	.954
10,000	.256	.514	.533	.173	1.00	1.00	1.00	1.00
	DGP 5.3				**DGP 5.4**			
T	LM_T	W_T	QLR_T	LM_{ECCC}	LM_T	W_T	QLR_T	LM_{ECCC}
1,000	.301	.484	.570	.0675	.639	.749	.895	.133
5,000	.971	.998	.997	.929	1.00	1.00	1.00	1.00
10,000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	DGP 5.5				**DGP 5.6**			
T	LM_T	W_T	QLR_T	LM_{ECCC}	LM_T	W_T	QLR_T	LM_{ECCC}
1,000	.0404	.0742	.0829	.0410	.288	.477	.522	.0675
5,000	.0605	.122	.130	.0520	.956	.995	.996	.891
10,000	.0875	.202	.198	.0730	1.00	1.00	1.00	.999
	DGP 5.7				**DGP 5.8**			
T	LM_T	W_T	QLR_T	LM_{ECCC}	LM_T	W_T	QLR_T	LM_{ECCC}
1,000	.0383	.0618	.0938	.0284	.206	.341	.542	.0704
5,000	.0665	.134	.174	.0465	.966	.996	.997	.911
10,000	.118	.266	.291	.0845	1.00	1.00	1.00	1.00

Actual rejection frequencies based on the size-corrected critical values at the 5% nominal level.

better power properties than the other tests for all choices of DGP and for all sample lengths.

6 Empirical illustration

In this section we provide an empirical application of the proposed tests for volatility spillovers. We apply the same data set as in Nakatani and Teräsvirta (2009) and investigate the volatility spillovers between a pair of foreign exchange rates. The exchange rates are daily noon buying rates of the Japanese yen (JPY) and the Swiss franc (CHF) against the U.S. dollar (USD). The series go from 2 January 1975 to 2 December 2005, with a total of 7,766 observations in each series. Descriptive statistics of the data series are contained in Nakatani and Teräsvirta (2009, Tables 7 and 8).[4]

We fit a bivariate ECCC-GARCH model to the return series and test whether the matrices A and B are diagonal. The tests are based on the assumption that the diagonal elements of A and B are strictly positive under the null, such that no nuisance parameters take zero value. This enables us to determine the critical values of the tests according to Algorithm 1 and Remark 4.6. For each individual series

[4] We have left out any empirical illustration containing the equity pairs investigated in Nakatani and Teräsvirta (2009), as standard Box-Pierce tests revealed significant auto-correlation of order 5 for these series, suggesting that a raw ECCC-GARCH model, i.e. with no VAR(MA) component, may not be suitable for capturing the dynamics of these return series.

of the standardized residuals, based on a Jarque-Bera test, we rejected the null of normality, suggesting that the LM_{ECCC} test based on a χ_4^2 limiting distribution, as performed in Nakatani and Teräsvirta (2009), may not be appropriate for testing for no spillovers in the return series, as mentioned in Remark 4.5. Table 6.1 contains the estimation results. First, we notice that the point estimates of the off-diagonal elements of A and B are fairly small. Second, based on the LM_T statistic we fail to reject the null of no spillovers, whereas the null is rejected based on the LM_{ECCC} test with the p-value based on a χ_4^2-distribution. The latter is in line with the findings in Nakatani and Teräsvirta (2009), but, as the standardized residuals, as mentioned, did not appear to be normally distributed, the validity of the LM_{ECCC} test is dubious. Based on the QLR_T and W_T tests, we reject the null of no spillovers. In light of the very reasonable size properties and superior power properties of these tests compared to LM_T, we find evidence for volatility spillovers between the JPY/USD and CHF/USD rates, in line with the findings in Nakatani and Teräsvirta (2009).

Table 6.1: Estimation results

Model		κ	A		B		r	LM_T	W_T	QLR_T	LM_{ECCC}
CCC	JPY	2.1	0.0513		0.9460		0.5416	8.87	52.57	76.21	40.23
	CHF	7.8		0.0574		0.9285		(0.0645)	(0.0285)	(0.0097)	(0.000)
ECCC	JPY	1.2	0.0449	0.0037	0.9493	0.0000	0.5417				
	CHF	6.7	0.0000	0.0588	0.0080	0.9229					

Point estimates of parameters in the restricted ECCC-GARCH model (CCC) and in the unrestricted ECCC-GARCH model (ECCC). The estimates of the elements of κ are multiplied by 1,000. The p-values of the LM_T, W_T, QLR_T, and LM_{ECCC} test for diagonality of A and B are reported in parentheses. The p-values for W_T and QLR_T are obtained according to Algorithm 1 and Remark 4.6 based on 1,000,000 draws. The p-values for LM_T and LM_{ECCC} are based on a χ_4^2-distribution.

7 Concluding remarks and future research directions

We have considered the large-sample properties of the quasi-maximum likelihood estimator (QMLE) for the extended constant conditional correlation GARCH model in the case where the true parameter is on the boundary of the parameter space. This case is of great importance in empirical finance where one is typically interested in testing for volatility spillovers between assets and markets. In contrast to the "standard" case, where the true parameter is an interior point, the limiting distribution is given by a projection of a Gaussian vector onto a set determined by the true parameter vector. Moreover, we proposed Lagrange multiplier (LM), Wald, and quasi-likelihood ratio statistics (QLR) suitable for testing for volatility spillovers. Similar to the QMLE, the Wald and QLR statistics do also have non-standard limiting distributions, however, as we demonstrate, these distributions are (under suitable conditions) straightforward to make draws from.

A simulation study showed that the three proposed tests have reasonable empirical size properties, in particular for samples with more than 5,000 observations. Moreover, simulations showed that the Wald and QLR tests have superior empirical power properties compared to the LM test.

Lastly, in an empirical illustration the proposed tests were applied to returns on foreign exchange rates. For the sample period from 2 January 1975 to 2 December 2005, based on the Wald and QLR tests we rejected the null of no volatility spillovers between the Japanese Yen/U.S. dollar and the Swiss Franc/U.S. dollar rates, in line with the findings of Nakatani and Teräsvirta (2009).

An important topic for future research is to investigate the limiting distributions of the proposed Wald and QLR statistics in more detail. Specifically, the limiting distributions appear, in general, to depend on nuisance parameters taking zero value, hence it is of particular interest to consider other tests, or corrections, that are pivotal to such boundary properties, as e.g. considered in recent work by Ketz (2014).

Appendix A Proofs of theorems

Throughout the proofs let $\mathcal{C}$ and ϕ denote positive, finite generic constants always with $\phi < 1$. Moreover, all Taylor-type expansions are based on partial left/right derivatives according to Andrews (1999, Appendix A), where all derivatives with respect to parameters at the boundary of Θ are right derivatives. Furthermore, for the proofs of the theorems as well as the lemmas stated in the next section, it will be convenient to consider the following partitions. With J and Σ the matrices defined in (3.13) and G and Z the random vectors given by $\mathcal{L}(G) = N(0, \Sigma)$ and $Z = J^{-1}G$, define according to the partition $\theta = (\beta', \delta')'$

$$J = \begin{bmatrix} J_{\beta\beta} & J_{\beta\delta} \\ J_{\delta\beta} & J_{\delta\delta} \end{bmatrix}, \quad G = \begin{bmatrix} G_\beta \\ G_\delta \end{bmatrix}, \quad \text{and} \quad Z = \begin{bmatrix} Z_\beta \\ Z_\delta \end{bmatrix}, \tag{A.1}$$

where $J_{\beta\beta}$ is $(s_1 \times s_1)$ and G_β is $(s_1 \times 1)$ and so forth.

Proof of Theorem 3.1. The asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is derived along the lines of Andrews (1999, Proof of Theorem 3) and Francq and Zakoïan

(2007, Proof of Theorem 2). Initially, notice that $\hat{\theta}_T$ is strongly consistent for θ_0, as mentioned in Section 3. Moreover, $\Theta - \theta_0$ is locally equal to a union of orthants, Λ, (Assumption 3) and $\hat{L}_T(\theta)$ has continuous left/right partial derivative of order 2 on Θ. Due to Andrews (1999, Theorem 6) we are able to make a second-order Taylor-type expansion of $\hat{L}_T(\theta)$ around θ_0 such that for any point $\theta \in \Theta$ there exists a point $\theta^\star$ between θ and θ_0

$$\hat{L}_T(\theta) = \hat{L}_T(\theta_0) + \frac{\partial L_T(\theta_0)}{\partial \theta'}(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'}(\theta - \theta_0) + R_T(\theta) + R_T^\star(\theta),$$

where

$$R_T(\theta) = \left(\frac{\partial \hat{L}_T(\theta_0)}{\partial \theta'} - \frac{\partial L_T(\theta_0)}{\partial \theta'} \right)(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \left[\frac{\partial^2 \hat{L}_T(\theta_0)}{\partial \theta \partial \theta'} - \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'} \right](\theta - \theta_0), \tag{A.2}$$

and

$$R_T^\star(\theta) = \frac{1}{2}(\theta - \theta_0)' \left[\frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \theta \partial \theta'} - \frac{\partial^2 \hat{L}_T(\theta_0)}{\partial \theta \partial \theta'} \right](\theta - \theta_0). \tag{A.3}$$

Moreover, with

$$J_T := \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'}$$

and

$$Z_T := -J_T^{-1} \sqrt{T} \frac{\partial L_T(\theta_0)}{\partial \theta}, \tag{A.4}$$

we have, by definition, that

$$\hat{L}_T(\theta) = \hat{L}_T(\theta_0) - \frac{1}{2T} \|Z_T\|_{J_T}^2 + \frac{1}{2T} \|Z_T - \sqrt{T}(\theta - \theta_0)\|_{J_T}^2 + R_T(\theta) + R_T^\star(\theta). \tag{A.5}$$

Notice that in the definition of Z_T in (A.4) we have used that $J_T = J + o(1)$ almost surely with J non-singular, as proved below. So, technically, Z_T might only exist almost surely for T sufficiently large. It suffices to establish the following points:

1. $\sqrt{T} \partial L_T(\theta_0)/\partial \theta \stackrel{w}{\rightarrow} G$ with $\mathcal{L}(G) = N(0, \Sigma)$ and $J_T = J + o_p(1)$, where the matrices $J \in \mathbb{S}_{++}^{s_0}$ and Σ are given by (3.13).

2. $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$.

3. For any $\bar{\theta}_T \in \Theta$ such that $\sqrt{T}(\bar{\theta}_T - \theta_0) = O_p(1)$, $R_T^\star(\bar{\theta}_T) = o_p(T^{-1})$ and $R_T(\bar{\theta}_T) = o_p(T^{-1})$.

4. $\sqrt{T}(\hat{\theta}_T - \theta_0) = \hat{\lambda}_T + o_p(1)$, where $\hat{\lambda}_T \in \mathrm{cl}(\Lambda)$ satisfies $\|Z_T - \hat{\lambda}_T\|_{J_T}^2 = \inf_{\lambda \in \Lambda} \|Z_T - \lambda\|_{J_T}^2$ with Λ defined in Assumption 5

5. $\hat{\lambda}_T \overset{w}{\to} \lambda^\Lambda$, where $\lambda^\Lambda \in \mathrm{cl}(\Lambda)$ satisfies $\|Z - \lambda^\Lambda\|_J^2 = \inf_{\lambda\in\Lambda} \|Z - \lambda\|_J^2$, $Z := -J^{-1}G$.

First, it follows from Lemma B.3 that Σ is finite. Expressions for $\partial l_t(\theta)/\partial\theta_i$, $i = 1, ..., s$, are given in the proof of Lemma B.4 below. As in Francq and Zakoïan (2012a, p 200), by a central limit theorem for strictly stationary and ergodic martingale difference sequences, see e.g. Brown (1971), $\sqrt{T}\partial L_T(\theta_0)/\partial\theta \overset{w}{\to} G$. Moreover, the ergodic theorem implies that $J_T = J + o(1)$ almost surely. The positive definiteness of J is established in Francq and Zakoïan (2012a, pp.203-204), and we conclude that 1. holds.

From the derivation of 1. we have that $\|\cdot\|_{J_T}$ is almost surely a norm for T sufficiently large due to the fact that J is positive definite. With $R_T(\theta)$ defined in (A.2), it follows by Lemma B.5 that

$$R_T(\hat{\theta}_T) = o_p(T^{-1/2}\|\hat{\theta}_T-\theta_0\|)+o_p(\|\hat{\theta}_T-\theta_0\|^2) = o_p(T^{-1/2}\|\hat{\theta}_T-\theta_0\|_{J_T})+o_p(\|\hat{\theta}_T-\theta_0\|_{J_T}^2). \tag{A.6}$$

For sufficiently large T, by Lemma B.5, $[\partial^2\hat{L}_T(\theta^\star)/\partial\theta\partial\theta' - \partial^2\hat{L}_T(\theta_0)/\partial\theta\partial\theta'] = [\partial^2 L_T(\theta^\star)/\partial\theta\partial\theta' - \partial^2 L_T(\theta_0)/\partial\theta\partial\theta'] + o_p(1)$. Also by Lemma B.5, $[\partial^2 L_T(\theta^\star)/\partial\theta\partial\theta' - \partial^2 L_T(\theta_0)/\partial\theta\partial\theta'] = \mathbb{E}[\partial^2 l_t(\theta^\star)/\partial\theta\partial\theta'] - \mathbb{E}[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'] + o_p(1)$, so by the continuity of $\mathbb{E}[\partial^2 l_t(\theta)/\partial\theta\partial\theta']$ on Θ and the consistency of $\hat{\theta}_T$,

$$R_T^\star(\hat{\theta}_T) = o_p(\|\hat{\theta}_T - \theta_0\|_{J_T}^2), \tag{A.7}$$

with $R_T^\star(\theta)$ defined in (A.3). Now from (A.5), (A.6)-(A.7), and the fact that $\hat{\theta}_T$ minimizes $\hat{L}_T(\theta)$,

$$\begin{aligned}\hat{L}_T(\hat{\theta}_T) - \hat{L}_T(\theta_0) &= \frac{1}{2T}[\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2 - \|Z_T\|_{J_T}^2] + R_T(\hat{\theta}_T) + R_T^\star(\hat{\theta}_T) \\ &= \frac{1}{2T}[\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2 - \|Z_T\|_{J_T}^2] \\ &\quad + o_p(\|\hat{\theta}_T - \theta_0\|_{J_T}^2) + o_p(T^{-1/2}\|\hat{\theta}_T - \theta_0\|_{J_T}) \le 0.\end{aligned} \tag{A.8}$$

Since $\|\cdot\|_{J_T}$ is a norm for T sufficiently large almost surely, it follows from 1. that $\|Z_T\|_{J_T} = O_p(1)$. This fact together with (A.8) yields

$$\begin{aligned}\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2 &\le \|Z_T\|_{J_T}^2 + o_p(\|\sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2) + o_p(\sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T}) \\ &\le (\|Z_T\|_{J_T} + o_p(\sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T}))^2.\end{aligned} \tag{A.9}$$

The triangle inequality and (A.9) imply that

$$\begin{aligned}\sqrt{T}\|\hat{\theta}_T-\theta_0\|_{J_T} &\leq \|Z_T-\sqrt{T}(\hat{\theta}_T-\theta_0)\|_{J_T}+\|Z_T\|_{J_T}\\ &\leq 2\|Z_T\|_{J_T}+o_p(\sqrt{T}\|\hat{\theta}_T-\theta_0\|_{J_T}).\end{aligned}$$

We conclude that $\sqrt{T}\|\hat{\theta}_T-\theta_0\|_{J_T}[1+o_p(1)]\leq O_p(1)$, and hence that 2. holds. Result 3. is verified by arguments similar to the ones used to verify 2. together with Lemma B.5.

Turning to 4., notice that when $s_1=0$, i.e. when $\theta_0\in\Theta$, it holds that $\hat{\lambda}_T=Z_T$, and the result follows immediately by the consistency of $\hat{\theta}_T$ and Lemma B.5. Let $\hat{\theta}_q$ satisfy $\|Z_T-\sqrt{T}(\hat{\theta}_q-\theta_0)\|^2_{J_T}=\inf_{\theta\in\Theta}\|Z_T-\sqrt{T}(\theta-\theta_0)\|^2_{J_T}$. It holds that

$$\begin{aligned}\|\sqrt{T}(\hat{\theta}_q-\theta_0)\|_{J_T} &\leq \|Z_T-\sqrt{T}(\hat{\theta}_q-\theta_0)\|_{J_T}+\|Z_T\|_{J_T}\\ &= \inf_{\theta\in\Theta}\|Z_T-\sqrt{T}(\theta-\theta_0)\|_{J_T}+\|Z_T\|_{J_T}\\ &\leq 2\|Z_T\|_{J_T}=O_p(1),\end{aligned}$$

where the first inequality is due to the triangle inequality, the second inequality follows from the fact that $\theta_0\in\Theta$, and the last equality follows from 1. Similar to the derivations above, we conclude that $\sqrt{T}(\hat{\theta}_q-\theta_0)=O_p(1)$. From (A.5), using that $\hat{\theta}_T$ minimizes $\hat{L}_T(\theta)$, and that $\hat{\theta}_q$ minimizes $\|Z_T-\sqrt{T}(\theta-\theta_0)\|^2_{J_T}$, together with results 2. and 3., we have that

$$\begin{aligned}0 &\geq T[\hat{L}_T(\hat{\theta}_T)-\hat{L}_T(\hat{\theta}_q)]\\ &= \frac{1}{2}\|Z_T-\sqrt{T}(\hat{\theta}_T-\theta_0)\|^2_{J_T}-\frac{1}{2}\|Z_T-\sqrt{T}(\hat{\theta}_q-\theta_0)\|^2_{J_T}\\ &\quad +T[R^\star_T(\hat{\theta}_T)+R_T(\hat{\theta}_T)-R^\star_T(\hat{\theta}_q)-R_T(\hat{\theta}_q)]\\ &\geq T[R^\star_T(\hat{\theta}_T)+R_T(\hat{\theta}_T)-R^\star_T(\hat{\theta}_q)-R_T(\hat{\theta}_q)]=o_p(1).\end{aligned} \tag{A.10}$$

Hence, using (A.5) and (A.10),

$$\|Z_T-\sqrt{T}(\hat{\theta}_T-\theta_0)\|^2_{J_T}=\|Z_T-\sqrt{T}(\hat{\theta}_q-\theta_0)\|^2_{J_T}+o_p(1). \tag{A.11}$$

Note that $\inf_{\theta\in\Theta}\|Z_T-\sqrt{T}(\theta-\theta_0)\|^2_{J_T}=\inf_{\lambda\in\sqrt{T}(\Theta-\theta_0)}\|Z_T-\lambda\|^2_{J_T}$, where $\sqrt{T}(\Theta-\theta_0):=\{\lambda\in\mathbb{R}^{s_0}:\lambda=\sqrt{T}(\theta-\theta_0),\theta\in\Theta\}$. Moreover 1. and the fact that Λ is a cone (Remark 3.3) imply, due to Andrews (1999, Lemma 2), that

$$\inf_{\lambda\in\sqrt{T}(\Theta-\theta_0)}\|Z_T-\lambda\|^2_{J_T}=\inf_{\lambda\in\Lambda}\|Z_T-\lambda\|^2_{J_T}+o_p(1). \tag{A.12}$$

Let $\hat{\lambda}_T \in \Lambda$ satisfy $\|Z_T - \hat{\lambda}_T\|^2_{J_T} = \inf_{\lambda\in\Lambda}\|Z_T - \lambda\|^2_{J_T}$. Then combining (A.11) and (A.12) yields

$$\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} = \|Z_T - \hat{\lambda}_T\|^2_{J_T} + o_p(1). \tag{A.13}$$

Observe that

$$\begin{aligned}\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} &= \|\sqrt{T}(\hat{\theta}_T - \theta_0) - \hat{\lambda}_T\|^2_{J_T} + \|Z_T - \hat{\lambda}_T\|^2_{J_T} \\ &\quad +2\left\langle Z_T - \hat{\lambda}_T, \hat{\lambda}_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\right\rangle_{J_T}. \end{aligned} \tag{A.14}$$

Using that $\sqrt{T}(\hat{\theta}_T - \theta_0) \in \Lambda$ and that Λ is closed for $s_1 > 0$, it follows from Zarantonello (1971, Lemma 1.1),

$$\left\langle Z_T - \hat{\lambda}_T, \hat{\lambda}_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\right\rangle_{J_T} \geq 0. \tag{A.15}$$

Combining (A.14) and (A.15) yields

$$\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} \geq \|\sqrt{T}(\hat{\theta}_T - \theta_0) - \hat{\lambda}_T\|^2_{J_T} + \|Z_T - \hat{\lambda}_T\|^2_{J_T}. \tag{A.16}$$

In light of (A.13) and (A.16), we conclude that 4. holds.
In line with Andrews (1999, p.1379), since Λ is convex, $\hat{\lambda}_T$ is unique. Moreover, since $\hat{\lambda}_T$ satisfies $\|Z_T - \hat{\lambda}_T\|^2_{J_T} = \inf_{\lambda\in\Lambda}\|Z_T - \lambda\|^2_{J_T}$, $\hat{\lambda}_T = f(Z_T, J_T)$ with some implicitly given function f. The function f is continuous at all points (Z_T, J_T) where J_T is nonsingular. Since J is nonsingular, the continuous mapping theorem implies that $\hat{\lambda}_T = f(Z_T, J_T) \overset{w}{\to} f(Z, J) = \lambda^\Lambda$, and we conclude that 5. holds. □

Proof of Theorem 4.1. From the proof of Theorem 3.1,

$$T[\hat{L}_T(\hat{\theta}_T) - \hat{L}_T(\theta_0)] = -\frac{1}{2}\|Z_T\|^2_{J_T} + \frac{1}{2}\|Z_T - \hat{\lambda}_T\|^2_{J_T} + o_p(1),$$

so the continuous mapping theorem together with points 1. and 5. from the proof of Theorem 3.1 imply that

$$2T\left[\hat{L}_T\left(\hat{\theta}_T\right) - \hat{L}_T(\theta_0)\right] \overset{w}{\to} -\|Z\|^2_J + \inf_{\lambda\in\Lambda}\|Z - \lambda\|^2_J. \tag{A.17}$$

Next, with λ^Λ_β defined in Theorem 4.1, with Z_β, G_δ, and $J_{\delta\delta}$ defined according to the partitions in (A.1), and with λ^{Λ_β} satisfying $\|Z_\beta - \lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} = \inf_{\lambda_\beta\in\Lambda_\beta}\|Z_\beta -$

$\lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$, it holds that

$$\begin{aligned} -\|Z\|^2_J + \inf_{\lambda\in\Lambda}\|Z-\lambda\|^2_J &= -\|Z_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J^{-1}_{\delta\delta}} + \inf_{\lambda\in\Lambda}\|Z-\lambda\|^2_J \\ &= -\|Z_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J^{-1}_{\delta\delta}} + \|Z_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|\lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} \\ &= -\|G_\delta\|^2_{J^{-1}_{\delta\delta}} - \|\lambda^\Lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}, \end{aligned} \tag{A.18}$$

where the first equality follows from Lemma B.6.1. The second equality follows from Lemma B.6.2 and Perlman (1969, Lemma 4.1), and the third equality follows from Lemma B.6.3. Combining (A.17) and (A.18) yields

$$2T[\hat{L}_T(\hat\theta_T) - \hat{L}_T(\theta_0)] \overset{w}{\to} -\|\lambda^\Lambda_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J^{-1}_{\delta\delta}}. \tag{A.19}$$

Notice that since $\theta_0 \in \Theta_0 \subset \Theta$ and $\Lambda_0 = \Lambda_{0,\beta_1} \times \Lambda_{\beta_2} \times \Lambda_\delta = \{0_{\tilde{s}_1\times 1}\} \times \mathbb{R}^{\tilde{s}_2}_+ \times \mathbb{R}^{s_2}$, it is possible, due to Assumption 7, to derive points 1.-6. in the proof of Theorem 3.1 with $\hat\theta_T$, $\hat\lambda_T$, λ^Λ, and Λ replaced by $\tilde\theta_T$, $\tilde\lambda_T$, λ^{Λ_0}, and Λ_0, respectively. In particular, and similar to the derivations above,

$$2T[\hat{L}_T(\tilde\theta_T) - \hat{L}_T(\theta_0)] \overset{w}{\to} -\|\lambda^{\Lambda_0}_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J^{-1}_{\delta\delta}}. \tag{A.20}$$

The convergence of (A.19) and (A.20) holds jointly, since the convergence of the two terms are due to point 1. in the proof of Theorem 3.1. This joint convergence and the Cramér-Wold theorem yield the limiting distribution of QLR_T.
Next, (4.7) follows by (3.12), Theorem B.1, and the continuous mapping theorem.
Lastly, we turn to the limiting distribution of LM_T. It holds, due to the consistency of $\tilde\theta_T$, Lemma B.1, and the invertibility of J (Theorem 3.1), $\hat{J}_T(\tilde\theta_T)^{-1} = J^{-1} + o_p(1)$. By a Taylor-type expansion and Lemma B.5.1

$$\sqrt{T}\hat{S}_T(\tilde\theta_T) = \sqrt{T}S_T(\theta_0) + \hat{J}_T(\theta^\star)\sqrt{T}(\tilde\theta_T - \theta_0) + o_p(1),$$

where $\theta^\star$ is between $\tilde\theta_T$ and θ_0 as in Jensen and Rahbek (2004, Proof of Lemma 1). By Lemma B.1 and by using that $\theta^\star = \theta_0 + o_p(1)$, it holds that $\hat{J}_T(\theta^\star) = J + o_p(1)$. Hence, using that $\sqrt{T}S_T(\theta_0)$ and $\sqrt{T}(\tilde\theta_T - \theta_0)$ are both $O_p(1)$,

$$\begin{aligned} \sqrt{T}K_1\hat{J}_T(\tilde\theta_T)^{-1}\hat{S}_T(\tilde\theta_T) &= K_1J^{-1}\left[\sqrt{T}S_T(\theta_0) + J\sqrt{T}(\tilde\theta_T - \theta_0)\right] + o_p(1) \\ &= K_1J^{-1}\sqrt{T}S_T(\theta_0) + K_1\sqrt{T}(\tilde\theta_T - \theta_0) + o_p(1). \end{aligned}$$

Since $K_1(\tilde\theta_T - \theta_0) = \beta_{1,0} = 0_{\tilde{s}_1\times 1}$, by Slutsky's lemma and the fact that $\sqrt{T}\partial L_T(\theta_0)/\partial\theta \overset{w}{\to}$

G,

$$\begin{aligned}\sqrt{T}K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{S}_T(\tilde{\theta}_T) &= K_1J^{-1}\sqrt{T}S_T(\theta_0)+o_p(1) \\ &\overset{w}{\to} N\left(0, K_1J^{-1}\Sigma J^{-1}K_1'\right). \end{aligned} \tag{A.21}$$

By Lemma B.1 and the fact that $\sqrt{T}(\tilde{\theta}_T-\theta_0)=O_p(1)$,

$$K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{\Sigma}_T(\tilde{\theta}_T)\hat{J}_T(\tilde{\theta}_T)^{-1}K_1' = K_1J^{-1}\Sigma J^{-1}K_1'+o_p(1). \tag{A.22}$$

Hence (4.8) follows by combining (A.21) and (A.22) and applying Slutzky's lemma and the continuous mapping theorem. □

Appendix B Lemmas

Lemma B.1. *With J and Σ given in (3.13) and $\hat{J}_T(\theta)$ and $\hat{\Sigma}_T(\theta)$ given in (4.3), let $\bar{\theta}_T\in\Theta$ satisfy $\bar{\theta}_T=\theta_0+o_p(1)$. Under the assumptions of Theorem 3.1,*

$$\hat{J}_T(\bar{\theta}_T)=J+o_p(1). \tag{B.1}$$

Additionally, suppose that $\sqrt{T}(\bar{\theta}_T-\theta_0)=O_p(1)$. Then

$$\hat{\Sigma}_T(\bar{\theta}_T)=\Sigma+o_p(1). \tag{B.2}$$

Proof. The proof is quite similar to the arguments given in Ling and McAleer (2010, p.100). Define, $J_T(\theta) := T^{-1}\sum_{t=1}^T \partial^2 l_t(\theta)/\partial\theta\partial\theta'$, where $l_t(\theta)$ is given by (3.5). Lemma B.5 implies that $\hat{J}_T(\bar{\theta}_T)=J_T(\bar{\theta}_T)+o_p(1)$, so in order to establish (B.1) it remains to show that $J_T(\bar{\theta}_T)=J+o_p(1)$. This property follows directly from Lemma B.5, the consistency of $\bar{\theta}_T$, and the fact that $\mathbb{E}[\partial^2 l_t(\theta)/\partial\theta\partial\theta']$ is continuous as θ_0.

Next, we seek to prove (B.2). Notice that with $\hat{l}_t(\theta)$ given by (3.2),

$$\begin{aligned}\hat{\Sigma}_T(\bar{\theta}_T) &= \frac{1}{T}\sum_{t=1}^T\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}+\frac{1}{T}\sum_{t=1}^T\frac{\partial l_t(\theta_0)}{\partial\theta}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta'}-\frac{\partial l_t(\theta_0)}{\partial\theta'}\right] \\ &+\frac{1}{T}\sum_{t=1}^T\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta}-\frac{\partial l_t(\theta_0)}{\partial\theta}\right]\frac{\partial l_t(\theta_0)}{\partial\theta'} \\ &+\frac{1}{T}\sum_{t=1}^T\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta}-\frac{\partial l_t(\theta_0)}{\partial\theta}\right]\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta'}-\frac{\partial l_t(\theta_0)}{\partial\theta'}\right].\end{aligned} \tag{B.3}$$

The ergodic theorem implies that $T^{-1}\sum_{t=1}^{T}[\partial l_t(\theta_0)/\partial\theta][\partial l_t(\theta_0)/\partial\theta'] = \Sigma + o_p(1)$, so it remains to show that the other terms in (B.3) vanish with probability approaching one. It suffices to establish that

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right]\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta'} = o_p(1) \tag{B.4}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right]\frac{\partial l_t(\theta_0)}{\partial\theta'} = o_p(1). \tag{B.5}$$

A Taylor-type expansion yields

$$T^{-1/2}\sum_{t=1}^{T}[\partial\hat{l}_t(\bar{\theta}_T)/\partial\theta] = T^{-1/2}\sum_{t=1}^{T}[\partial\hat{l}_t(\theta_0)/\partial\theta] + \hat{J}_T(\theta_T^{\star})\sqrt{T}(\bar{\theta}_T - \theta_0),$$

where $\theta_T^{\star}$ is between $\bar{\theta}_T$ and θ_0. By Lemma B.5.1 and point 1. in the proof of Theorem 3.1, $T^{-1/2}\sum_{t=1}^{T}\partial\hat{l}_t(\theta_0)/\partial\theta = O_p(1)$, and using arguments similar to the ones used to show (B.1), $\hat{J}_T(\theta_T^{\star}) = J + o_p(1)$. Hence, using that $\sqrt{T}(\bar{\theta}_T - \theta_0) = O_p\left(1\right)$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} = O_p(1). \tag{B.6}$$

Moreover, also by a Taylor-type expansion,

$$\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta} = \left[\frac{\partial\hat{l}_t(\theta_0)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right] + \left\{\left[\frac{\partial^2\hat{l}_t(\theta_T^{\star})}{\partial\theta\partial\theta'} - \frac{\partial^2 l_t(\theta_T^{\star})}{\partial\theta\partial\theta'}\right] + \frac{\partial^2 l_t(\theta_T^{\star})}{\partial\theta\partial\theta'}\right\}(\bar{\theta}_T - \theta_0). \tag{B.7}$$

For any $\epsilon > 0$ and some $r > 0$, by Boole's and the generalized Chebyshev inequalities,

$$\mathbb{P}\left(\max_{t\in\mathbb{N}}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta_0)}{\partial\theta}\right\| > \epsilon\sqrt{T}\right) \leq \frac{1}{\epsilon^r T^{(r/2)}}\sum_{t=1}^{\infty}\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right\|^r\right] = o(1), \tag{B.8}$$

where we have used Lemma B.4.1. Likewise, using Lemma B.4.3, we have that

$$\frac{1}{\sqrt{T}}\max_{t\in\mathbb{N}}\left\|\frac{\partial^2\hat{l}_t(\theta_T^{\star})}{\partial\theta\partial\theta'} - \frac{\partial^2 l_t(\theta_T^{\star})}{\partial\theta\partial\theta'}\right\| = o_p(1), \tag{B.9}$$

and using Lemma B.4.4,

$$\frac{1}{\sqrt{T}}\left\|\frac{\partial^2 l_t(\theta_T^{\star})}{\partial\theta\partial\theta'}\right\| = o_p(1). \tag{B.10}$$

Combining (B.6), (B.7), (B.8), (B.9), (B.10), and that $(\bar{\theta}_T - \theta_0) = o_p\left(1\right)$ yields (B.4).

Similar arguments yield (B.5). □

Lemma B.2. *Let $\hat{h}_t(\theta)$ and $h_t(\theta)$ be given by (3.4) and (3.7), respectively, and let $\hat{D}_t(\theta)$ and $D_t(\theta)$ be given by (3.3) and (3.6), respectively. Suppose that the assumptions of Theorem 3.1 are satisfied. It holds that for all $t \in \mathbb{N}_0$, $i, j = 1, ..., d + 2d^2$, and some $k \geq 0$,*

$$\mathbb{E}[\sup_{\theta\in\Theta} \|h_t(\theta)\|^3] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial h_t(\theta)}{\partial\theta_i}\right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|^3\right] < \infty,$$

$$\mathbb{E}[\sup_{\theta\in\Theta} \left\|\hat{h}_t(\theta)\right\|^3] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial \hat{h}_t(\theta)}{\partial\theta_i}\right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 \hat{h}_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|^3\right] < \infty,$$

$$\sup_{\theta\in\Theta}\|R^{-1}(\theta)\| \leq \mathcal{C}, \quad \sup_{\theta\in\Theta}\|D_t^{-1}(\theta)\| \leq \mathcal{C}, \quad \sup_{\theta\in\Theta}\|\hat{D}_t^{-1}(\theta)\| \leq \mathcal{C},$$

$$\mathbb{E}[\sup_{\theta\in\Theta}\|h_t(\theta) - \hat{h}_t(\theta)\|] = O(t^k\phi^t),$$

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial h_t(\theta)}{\partial\theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial\theta_i}\right\|\right] = O(t^k\phi^t), \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_t(\theta)}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 \hat{h}_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|\right] = O(t^k\phi^t).$$

Proof. Notice that since $\rho(B) < 1$ on Θ, and Θ is compact

$$\sup_{\theta\in\theta}\|B^t\| \leq \mathcal{C}\phi^t. \tag{B.11}$$

Since $\rho(B) < 1$, recursions give that $h_t(\theta) = \sum_{i=0}^{\infty} B^i(\kappa + AX_{t-1-i}^{\odot 2})$, so by repeated use of Minkowski's inequality, the compactness of Θ, (B.11), and the fact that $\mathbb{E}[\|X_t\|^6] < \infty$ yield that

$$\mathbb{E}[\sup_{\theta\in\Theta}\|h_t(\theta)\|^3] < \infty. \tag{B.12}$$

Moreover, $\hat{h}_t(\theta) = \sum_{i=0}^{t-1} B^i(\kappa + AX_{t-1-i}^{\odot 2}) + B^t\hat{h}_0$, so similar arguments and the fact that $\hat{h}_0$ is fixed yield that for all $t \in \mathbb{N}_0$, $\mathbb{E}[\sup_{\theta\in\Theta}\|\hat{h}_t(\theta)\|^3] < \infty$. Next, we consider the partial derivatives (potentially of the left/right type) of $h_t(\theta)$. For convenience, we differentiate with respect to the standard parametrization as introduced in subsection 2.1, i.e. without loss of generality we let $\theta = (\kappa', \text{vec}(A)', \text{vec}(B)', \text{vech}^0(R)')'$. Let $\tilde{r}_2 := d + d^2$ and $\tilde{r}_1 := d + 2d^2$. Using that $\rho(B) < 1$,

$$\frac{\partial h_t(\theta)}{\partial\theta_i} = \sum_{j=0}^{\infty} B^i \frac{\partial\kappa}{\partial\theta_i} \quad \text{for } i = 1, ..., d,$$

$$\frac{\partial h_t(\theta)}{\partial \theta_i} = \sum_{j=0}^{\infty} B^i \frac{\partial A}{\partial \theta_i} X_{t-1-i}^{\odot 2} \quad \text{for } i = d+1, ..., \tilde{r}_2,$$

$$\frac{\partial h_t(\theta)}{\partial \theta_i} = \sum_{j=0}^{\infty} B^i \frac{\partial B}{\partial \theta_i} h_{t-1-i} \quad \text{for } i = \tilde{r}_2+1, ..., \tilde{r}_1.$$

So repeated use of Minkowski's inequality, $\mathbb{E}[\|X_t\|^6] < \infty$, (B.12), (B.11), and the compactness of Θ yield that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial h_t(\theta)}{\partial \theta_i}\right\|^3\right] < \infty \tag{B.13}$$

for $i = 1, ..., \tilde{r}_1$. By similar arguments,

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_t(\theta)}{\partial \theta_i \partial \theta_j}\right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial \hat{h}_t(\theta)}{\partial \theta_i}\right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 \hat{h}_t(\theta)}{\partial \theta_i \partial \theta_j}\right\|^3\right] < \infty, \tag{B.14}$$

for all $i, j = 1, ..., \tilde{r}_1$. Moreover, $\sup_{\theta\in\Theta}\|R^{-1}(\theta)\| \leq \mathcal{C}$, $\sup_{\theta\in\Theta}\|D_t^{-1}(\theta)\| \leq \mathcal{C}$, and $\sup_{\theta\in\Theta}\|\hat{D}_t^{-1}(\theta)\| \leq \mathcal{C}$ follow by arguments given in Francq and Zakoïan (2012a, p.195). We have that $h_t(\theta) - \hat{h}_t(\theta) = B^t[h_0(\theta) - \hat{h}_0]$, so (B.11), (B.12) and the fact that $\hat{h}_0$ is fixed give that $\mathbb{E}[\sup_{\theta\in\Theta}\|h_t(\theta) - \hat{h}_t(\theta)\|] = O(\phi^t)$. Similarly,

$$\frac{\partial h_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i} = B^t\left[\frac{\partial h_0(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_0(\theta)}{\partial \theta_i}\right] \quad \text{for } i = 1, ..., \tilde{r}_2,$$

$$\frac{\partial h_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i} = \frac{\partial B^t}{\partial \theta_i}\left[h_0(\theta) - \hat{h}_0\right] + B^t\left[\frac{\partial h_0(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_0(\theta)}{\partial \theta_i}\right] \quad \text{for } i = \tilde{r}_2+1, ..., \tilde{r}_1,$$

and we conclude, using (B.13), that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial h_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i}\right\|\right] = O(t\phi^t) \quad \text{for } i = 1, ..., \tilde{r}_1. \tag{B.15}$$

Likewise, using (B.14),

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_t(\theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \hat{h}_t(\theta)}{\partial \theta_i \partial \theta_j}\right\|\right] = O(t^2\phi^t) \quad \text{for } i, j = 1, ..., \tilde{r}_1. \tag{B.16}$$

□

Lemma B.3. *Under the assumptions of Theorem 3.1, the matrix Σ defined in (3.13) is finite.*

Proof. Due to Hölder's inequality, it suffices to show that $\mathbb{E}\{[\partial l_t(\theta_0)/\partial\theta_i)]^2\} < \infty$ for all $i = 1, ..., s_0$, where s_0 is the dimension of θ. Similar to the proof of Lemma B.2, we consider (without loss of generality) differentiation with respect to the standard parametrization where $\theta = (\kappa', \text{vec}(A)', \text{vec}\,(B)', \text{vech}^0(R)')'$. We define $\tilde{r}_2 := d + d^2$ and $\tilde{r}_1 := d + 2d^2$. From Francq and Zakoïan (2012a, p.198), it holds that

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i} = \text{tr}\left\{(I_d - R_0^{-1}\varepsilon_t\varepsilon_t')\frac{\partial D_{0t}}{\partial\theta_i}D_{0t}^{-1} + (I_d - \varepsilon_t\varepsilon_t' R_0^{-1})D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right\}$$

for $i = 1, ..., \tilde{r}_1$, where the "0" indicates that the functions are evaluated at θ_0, and $\varepsilon_t := R_0^{1/2} Z_t$. It holds that for $i = 1, ..., \tilde{r}_1$,

$$\frac{\partial D_t}{\partial\theta_i} = \frac{1}{2}D_t^{-1}\text{diag}\left(\frac{\partial h_t(\theta)}{\partial\theta_i}\right), \tag{B.17}$$

so by Lemma B.2, it holds that $\mathbb{E}[\|\partial D_{0t}/\partial\theta_i\|^3] < \infty$. Since $\partial D_{0t}/\partial\theta_i$ and $\varepsilon_t\varepsilon_t'$ are independent, and $\mathbb{E}[\|\varepsilon_t\|^6] < \infty$, we conclude using Hölder's inequality that $\mathbb{E}\{[\partial l_t(\theta_0)/\partial\theta_i]^2\} < \infty$ for $i = 1, ..., \tilde{r}_1$. Moreover, from Francq and Zakoïan (2012a, p.198), it holds that

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i} = \text{tr}\left\{(I_d - R_0^{-1}\varepsilon_t\varepsilon_t')\left(R_0^{-1}\frac{\partial R_0}{\partial\theta_i}\right)\right\}$$

for $i = \tilde{r}_1{+}1, ..., s_0$. Using similar arguments as above, we conclude that $\mathbb{E}\{[\partial l_t(\theta_0)/\partial\theta_i]^2\} < \infty$ for $i = \tilde{r}_1 + 1, ..., s_0$. □

Lemma B.4. *Suppose that the assumptions of Theorem 3.1 hold. Then with $\hat{l}_t(\theta)$ and $l_t(\theta)$ given by (3.2) and (3.5), respectively, the following statements are true.*

1. *For some $k \geq 0$ and some $u > 0$,*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right\|^u\right] = O(t^k\phi^t) \quad \forall t \in \mathbb{N}.$$

2. *For some $k \geq 0$ and some $u > 0$,*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left|l_t(\theta) - \hat{l}_t(\theta)\right|^r\right] = O(t^k\phi^t) \quad \forall t \in \mathbb{N}.$$

3. *For some $k \geq 0$ and some $u > 0$,*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial\hat{l}_t^2(\theta)}{\partial\theta\partial\theta'}\right\|^u\right] = O(t^k\phi^t) \quad \forall t \in \mathbb{N}.$$

4. $\mathbb{E}[\sup_{\theta\in\Theta}\|\partial^2 l_t(\theta)/\partial\theta\partial\theta'\|]<\infty.$

Proof. Similar to the proof of Lemma B.2, we consider differentiation with respect to the standard parametrization where $\theta=(\kappa',\mathrm{vec}(A)',\mathrm{vec}(B)',\mathrm{vech}^0(R)')'$, and define $\tilde r_2 := d+d^2$ and $\tilde r_1 := d+2d^2$. From Francq and Zakoïan (2012a, p.198) it holds that for $i=1,...,\tilde r_1$,

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial\theta_i} &= \mathrm{tr}\left\{D_t^{-1}\left[2I_d-\left(X_tX_t'D_t^{-1}R^{-1}D_t^{-1}+D_t^{-1}X_tX_t'D_t^{-1}R^{-1}\right)\right]\frac{\partial D_t}{\partial\theta_i}\right\}\\ &= \xi_t'\left[D_t^{-1}\otimes I_d\right]\mathrm{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)\end{aligned}$$

with

$$\xi_t := \mathrm{vec}\left[2I_d-\left(X_tX_t'D_t^{-1}R^{-1}D_t^{-1}+D_t^{-1}X_tX_t'D_t^{-1}R^{-1}\right)\right]. \tag{B.18}$$

Similarly,

$$\frac{\partial\hat l_t(\theta)}{\partial\theta_i}=\hat\xi_t'[\hat D_t^{-1}\otimes I_d]\,\mathrm{vec}\left(\frac{\partial\hat D_t}{\partial\theta_i}\right),$$

with $\hat\xi_t := \mathrm{vec}[2I_d-(X_tX_t'\hat D_t^{-1}R^{-1}\hat D_t^{-1}+\hat D_t^{-1}X_tX_t'\hat D_t^{-1}R^{-1})]$. Hence,

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial\theta_i}-\frac{\partial\hat l_t(\theta)}{\partial\theta_i} &= \xi_t'\left[D_t^{-1}\otimes I_d\right]\mathrm{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)-\hat\xi_t'[\hat D_t^{-1}\otimes I_d]\,\mathrm{vec}\left(\frac{\partial\hat D_t}{\partial\theta_i}\right)\\ &= (\xi_t'-\hat\xi_t')\left[D_t^{-1}\otimes I_d\right]\mathrm{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)\\ &\quad+\hat\xi_t'[(D_t^{-1}-\hat D_t^{-1})\otimes I_d]\,\mathrm{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)\\ &\quad+\hat\xi_t'[\hat D_t^{-1}\otimes I_d]\left[\mathrm{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)-\mathrm{vec}\left(\frac{\partial\hat D_t}{\partial\theta_i}\right)\right].\end{aligned}$$

It holds that

$$\begin{aligned}\xi_t-\hat\xi_t &= \mathrm{vec}\left[X_tX_t'(\hat D_t^{-1}-D_t^{-1})R^{-1}\hat D_t^{-1}\right]-\mathrm{vec}\left[X_tX_t'D_t^{-1}R^{-1}(D_t^{-1}-\hat D_t^{-1})\right]\\ &\quad+\mathrm{vec}\left[\hat D_t^{-1}X_tX_t'(\hat D_t^{-1}-D_t^{-1})R^{-1}\right]-\mathrm{vec}\left[(D_t^{-1}-\hat D_t^{-1})X_tX_t'D_t^{-1}R^{-1}\right],\end{aligned}$$

and

$$\|\hat D_t^{-1}-D_t^{-1}\|=\|\hat D_t^{-1}(\hat D_t-D_t)D_t^{-1}\|\le\mathcal{C}\|\hat D-D_t\|,$$

where we have used Lemma B.2. By the same lemma for some $k\geq 0$,

$$\mathbb{E}[\sup_{\theta\in\Theta}\|h_t(\theta)-\hat h_t(\theta)\|]=O(t^k\phi^t),$$

so we have that for some for some $\tilde{u} > 0$ and some $k \geq 0$,

$$\mathbb{E}[\sup_{\theta \in \Theta} \|\hat{D}_t^{-1} - D_t^{-1}\|^{\tilde{u}}] = O(t^k \phi^t). \tag{B.19}$$

Consequently, by Hölder's inequality for some $u^\star > 0$,

$$\mathbb{E}[\sup_{\theta \in \Theta} \|\xi_t - \hat{\xi}_t\|^{u^\star}] = O(t^k \phi^t)$$

For $i = 1, ..., \tilde{r}_1$,

$$\frac{\partial D_t}{\partial \theta_i} = \frac{1}{2} D_t^{-1} \operatorname{diag}\left(\frac{\partial h_t}{\partial \theta_i}\right),$$

and due to (B.19) and

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial h_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i} \right\|\right] = O(t^k \phi^t),$$

by Lemma B.2, it holds that for some $u^\star > 0$,

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \operatorname{vec}\left(\frac{\partial D_t}{\partial \theta_i}\right) - \operatorname{vec}\left(\frac{\partial \hat{D}_t}{\partial \theta_i}\right) \right\|^{u^\star}\right] = O(t^k \phi^t).$$

Consequently, by Hölder's inequality we have that for $i = 1, ..., \tilde{r}_1$ and some $u > 0$

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left| \frac{\partial l_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{l}_t(\theta)}{\partial \theta_i} \right|^u\right] = O(t^k \phi^t)$$

For $i = \tilde{r}_1 + 1, ..., s_0$,

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = \operatorname{tr}\left(R^{-1} \frac{\partial R}{\partial \theta_i}\right) - \operatorname{vec}(D_t^{-1})' \left[(X_t X_t') \otimes \left(R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1}\right)\right] \operatorname{vec}(D_t^{-1})$$

and

$$\frac{\partial \hat{l}_t(\theta)}{\partial \theta_i} = \operatorname{tr}\left(R^{-1} \frac{\partial R}{\partial \theta_i}\right) - \operatorname{vec}(\hat{D}_t^{-1})' \left[(X_t X_t') \otimes \left(R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1}\right)\right] \operatorname{vec}(\hat{D}_t^{-1}),$$

so by similar arguments as above, using (B.19) and Lemma B.2,

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left| \frac{\partial l_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{l}_t(\theta)}{\partial \theta_i} \right|^u\right] = O(t^k \phi^t)$$

$i = \tilde{r}_1 + 1, ..., s_0$. Using the c_r-inequality, we conclude that 1. holds.

Turning to 2., from Francq and Zakoïan (2012a, pp.195-196),

$$\sup_{\theta\in\Theta}\left|l_t(\theta)-\hat{l}_t(\theta)\right| \leq \sup_{\theta\in\Theta}\left|\mathrm{tr}\left\{X_tX_t'(H_t^{-1}-\hat{H}_t^{-1})\right\}\right| + \sup_{\theta\in\Theta}\left|\log\left\{\det(H_t)\right\}-\log\left\{\det(\hat{H}_t)\right\}\right|. \tag{B.20}$$

It holds that

$$\sup_{\theta\in\Theta}\left|\mathrm{tr}\left\{X_tX_t'(H_t^{-1}-\hat{H}_t^{-1})\right\}\right| \leq \|X_tX_t'\|\sup_{\theta\in\Theta}\left\|H_t^{-1}-\hat{H}_t^{-1}\right\|,$$

Since

$$\begin{aligned}H_t^{-1}-\hat{H}_t^{-1} &= H_t^{-1}(\hat{H}_t-H_t)\hat{H}_t^{-1}\\ &= D_t^{-1}R^{-1}D_t^{-1}[(\hat{D}_t-D_t)R\hat{D}_t+D_tR(\hat{D}_t-D_t)]\hat{D}_t^{-1}R^{-1}\hat{D}_t^{-1},\end{aligned}$$

it follows from Lemma B.2 and Hölder's inequality that for some $u^\star>0$,

$$\mathbb{E}[\sup_{\theta\in\Theta}|\mathrm{tr}\{X_tX_t'(H_t^{-1}-\hat{H}_t^{-1})\}|^{u^\star}]=O(t^k\phi^t). \tag{B.21}$$

Next let h_{it} and $\hat{h}_{it}$ denote the ith element $(i=1,..,d)$ of $h_t(\theta)$ and $\hat{h}_t(\theta)$ respectively. Consider the second term in (B.20). From Ling and McAleer (2003, p.302),

$$\begin{aligned}|\log\{\det(H_t)\}-\log\{\det(\hat{H}_t)\}| &= |\log\{\det(D_t^2\hat{D}_t^{-2})\}|\\ &= |\log(\prod_{i=1}^d h_{it}/\hat{h}_{it})| = |\sum_{i=1}^d\log(h_{it}/\hat{h}_{it})|,\end{aligned}$$

where we have used that h_{it} and $\hat{h}_{it}$ have a positive lower bound for each i uniformly on Θ. Since $\log(1+x)\leq x$ for $x>-1$, we have that

$$|\log\{\det(H_t)\}-\log\{\det(\hat{H}_t)\}| \leq \sum_{i=1}^d|\log\{1+(h_{it}-\hat{h}_{it})\hat{h}_{it}^{-1}\}| \leq \sum_{i=1}^d|(h_{it}-\hat{h}_{it})\hat{h}_{it}^{-1}|,$$

so, using Lemma B.2, for some $u^\star>0$,

$$\mathbb{E}[\sup_{\theta\in\Theta}\left|\log\{\det(H_t)\}-\log\{\det(\hat{H}_t)\}\right|^{u^\star}]=O(t^k\phi^t). \tag{B.22}$$

By combining (B.20), (B.21), (B.22), and Hölder's inequality, we conclude that point 2. holds.

Turning to point 3., expressions for $\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j$ for different choices of i and j are stated in Francq and Zakoïan (2012a, pp.200-201) (note that in Francq and

Zakoïan (2012a) ϵ_t corresponds to X_t here). By similar arguments as above, relying on Lemma B.2, we conclude that 3. holds.
In order to establish 4. it suffices to show that $\mathbb{E}[\sup_{\theta\in\Theta}|\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j|] < \infty$ for all $i, j = 1, ..., s_0$. Again, by relying on expressions for $\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_i$ for different choices of i and j are stated in Francq and Zakoïan (2012a, pp.200-201), it is seen that this moment restriction holds due to Lemma B.2 and Hölder's inequality. □

Lemma B.5. *Under the assumptions of Theorem 3.1, with $\hat{l}_t(\theta)$ and $l_t(\theta)$ given by (3.2) and (3.5), respectively,*

1. $\sup_{\theta\in\Theta} \|\frac{1}{T}\sum_{t=1}^T \partial l_t(\theta)/\partial\theta - \frac{1}{T}\sum_{t=1}^T \partial\hat{l}_t(\theta)/\partial\theta\| = o_p(T^{-1/2})$.
2. $\sup_{\theta\in\Theta} |\frac{1}{T}\sum_{t=1}^T l_t(\theta) - \frac{1}{T}\sum_{t=1}^T \hat{l}_t(\theta)| = o_p(T^{-1})$.
3. $\sup_{\theta\in\Theta} \|\frac{1}{T}\sum_{t=1}^T \partial^2 l_t(\theta)/\partial\theta\partial\theta' - \frac{1}{T}\sum_{t=1}^T \partial^2\hat{l}_t(\theta)/\partial\theta\partial\theta'\| = o(1)$ *a.s.*
4. $\sup_{\theta\in\Theta} \|\frac{1}{T}\sum_{t=1}^T \partial^2 l_t(\theta)/\partial\theta\partial\theta' - \mathbb{E}[\partial^2 l_t(\theta)/\partial\theta\partial\theta']\| = o(1)$ *a.s.*

Proof. In order to show 1., we use arguments similar to the ones given in Pedersen and Rahbek (2014, Proof of Lemma B.11), see also Hafner and Preminger (2009a, Proof of Lemma 4). For any $\epsilon > 0$ and some $u > 0$, by the generalized Chebyshev inequality,

$$\mathbb{P}\left(\sqrt{T}\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^T\left[\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right]\right\| > \epsilon\right) \leq \frac{T^{(u-2)/2}}{\epsilon^r}\sum_{t=1}^T\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right\|^u\right] = o(1),$$

choosing $u < 2$, where we have used Lemma B.4.1.
Using similar arguments and Lemma B.4.2, we conclude that point 2. holds.
Turning to point 3., for any $\epsilon > 0$ and some $\tilde{u} > 0$, by the generalized Chebyshev inequality,

$$\sum_{t=0}^\infty\mathbb{P}\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\hat{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| > \epsilon\right) \leq \epsilon^{-\tilde{u}}\sum_{t=0}^\infty\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\hat{l}_t(\theta)}{\partial\theta\partial\theta'}\right\|^{\tilde{u}}\right] < \infty,$$

where we have used Lemma B.4.3. The Borel-Cantelli lemma then implies that almost surely

$$\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\hat{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| \to 0 \quad \text{as } t\to\infty,$$

and point 3. then follows by Cesàro's mean theorem.
The proof of 4. follows by Lemma B.4.4 and a uniform law of large numbers for ergodic processes, see e.g. Ranga Rao (1962). □

Lemma B.6. *Let Z_β, G_δ, and $J_{\delta\delta}$ be defined according to (A.1). Moreover, with $\Lambda = \Lambda_\beta \times \Lambda_\delta$ defined in Assumption 5, let $\lambda^\Lambda = (\lambda_\beta^{\Lambda\prime}, \lambda_\delta^{\Lambda\prime})'$satisfy $\|Z - \lambda^\Lambda\|_J^2 = \inf_{\lambda\in\Lambda} \|Z-\lambda\|_J^2$ and let λ^{Λ_β} satisfy $\|Z_\beta - \lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} = \inf_{\lambda_\beta\in\Lambda_\beta} \|Z_\beta-\lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$. Under Assumptions 1-6,*

1. $Z'JZ = Z_\beta'(KJ^{-1}K')^{-1}Z_\beta + G_\delta' J_{\delta\delta}^{-1} G_\delta$,
2. $\|Z-\lambda^\Lambda\|_J^2 = \|Z_\beta - \lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} = \|Z_\beta - \lambda_\beta^\Lambda\|^2_{(KJ^{-1}K')^{-1}}$,
3. $\lambda_\beta^\Lambda = \lambda^{\Lambda_\beta}$.

Proof. The proof follows the lines of Andrews (1999, Proof of Theorem 4). First, recall that for matrices $A\in\mathbb{R}^{m\times m}$, $B\in\mathbb{R}^{m\times n}$, $C\in\mathbb{R}^{n\times m}$, and $D\in\mathbb{R}^{n\times n}$ satisfying that $E := \begin{bmatrix} A & B \\ C & D\end{bmatrix}$, D, and $(A-BD^{-1}C)$ are nonsingular, then

$$E^{-1} = \begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1}\end{bmatrix} \tag{B.23}$$

Define the matrices

$$M := \begin{bmatrix} I_{s_1} \\ -J_{\delta\delta}^{-1}J_{\delta\beta}\end{bmatrix}, \quad P^\perp := MK, \quad P := I_{s_0} - P^\perp.$$

Observe that by orthogonality

$$(Px_1)'J(P^\perp x_2) = 0 \quad \forall x_1, x_2 \in \mathbb{R}^{s_0}. \tag{B.24}$$

By (B.23),

$$KJ^{-1}K' = (J_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta})^{-1}, \tag{B.25}$$

and, moreover,

$$M'JM = J_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta},$$

so

$$M'JM = (KJ^{-1}K')^{-1}. \tag{B.26}$$

Let $\bar K := (0_{s_2\times s_1}, I_{s_2})$. By definition $I_{s_0} = (K', \bar K')'$, so

$$PJ^{-1}G = \begin{bmatrix} KPJ^{-1}G \\ \bar K PJ^{-1}G\end{bmatrix}. \tag{B.27}$$

It holds that

$$
\begin{aligned}
KP &= K(I_{s_0} - MK) \\
&= K - KMK \\
&= K - [I_{s_1}, 0_{s_1\times s_2}] \begin{bmatrix} I_{s_1} \\ -J_{\delta\delta}^{-1} J_{\delta\beta} \end{bmatrix} K \\
&= 0_{s_1\times s_0}, \qquad \text{(B.28)}
\end{aligned}
$$

so

$$KPJ^{-1}G = 0_{s_1\times 1}.$$

Furthermore, make the following partition $J^{-1} = \begin{bmatrix} J^{(1)} & J^{(2)} \\ J^{(3)} & J^{(4)} \end{bmatrix}$ according to (B.23) such that $J^{(1)} := (J_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta})^{-1}$, $J^{(2)} := -J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}$, $J^{(3)} := -J_{\delta\delta}^{-1}J_{\delta\beta}J^{(1)}$, and $J^{(4)} := J_{\delta\delta}^{-1} + J_{\delta\delta}^{-1}J_{\delta\beta}J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}$, with $J_{\beta\beta}$, $J_{\beta\delta}$, $J_{\delta\beta}$, and $J_{\delta\delta}$ defined according to (A.1). Then

$$
\begin{aligned}
J_{\delta\delta}\overline{K}PJ^{-1}G &= J_{\delta\delta}\bar{K}(I_{s_1+s_2} - MK)J^{-1}G \\
&= J_{\delta\delta}([0_{s_2\times s_1}, I_{s_2}] - \bar{K}M[I_{s_1}, 0_{s_1\times s_2}]) \begin{bmatrix} J^{(1)} & J^{(2)} \\ J^{(3)} & J^{(4)} \end{bmatrix} G \\
&= J_{\delta\delta}([J^{(3)}, J^{(4)}] - \bar{K}M[J^{(1)}, J^{(2)}])G \\
&= J_{\delta\delta}\left([J^{(3)}, J^{(4)}] - \bar{K}\begin{bmatrix} I_{s_1} \\ -J_{\delta\delta}^{-1}J_{\delta\beta} \end{bmatrix}[J^{(1)}, J^{(2)}]\right) G \\
&= J_{\delta\delta}\left([J^{(3)}, J^{(4)}] - [0_{s_2\times s_1}, I_{s_2}]\begin{bmatrix} [J^{(1)}, J^{(2)}] \\ -J_{\delta\delta}^{-1}J_{\delta\beta}[J^{(1)}, J^{(2)}] \end{bmatrix}\right) G \\
&= J_{\delta\delta}([J^{(3)}, J^{(4)}] + J_{\delta\delta}^{-1}J_{\delta\beta}[J^{(1)}, J^{(2)}])G \\
&= ([J_{\delta\delta}J^{(3)}, J_{\delta\delta}J^{(4)}] + J_{\delta\beta}[J^{(1)}, J^{(2)}])G
\end{aligned}
$$

Hence,

$$
\begin{aligned}
J_{\delta\delta}\bar{K}PJ^{-1}G &= ([-J_{\delta\beta}J^{(1)}, I_r + J_{\delta\beta}J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}] + [J_{\delta\beta}J^{(1)}, -J_{\delta\beta}J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}])G \\
&= [0_{s_2\times s_1}, I_{s_2}]G \qquad \text{(B.29)} \\
&= G_\delta.
\end{aligned}
$$

Combining (B.27), (B.32), and (B.29) yields

$$PJ^{-1}G = \begin{bmatrix} 0_{s_1\times 1} \\ J_{\delta\delta}^{-1}G_\delta \end{bmatrix}. \qquad \text{(B.30)}$$

Now (B.24) implies that

$$Z'JZ = (PZ)'\,J\,(PZ) + (P^{\perp}Z)'J(P^{\perp}Z).$$

This combined with (B.26), (B.30), and that $Z = -J^{-1}G$ (by definition) proves 1. For $\lambda = (\lambda_\beta', \lambda_\delta') \in \Lambda_\beta \times \Lambda_\delta$ it holds that

$$P\lambda = \begin{bmatrix} 0_{s_1\times 1} \\ \lambda_\delta + J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta \end{bmatrix}. \tag{B.31}$$

Using (B.24), (B.31), (B.30), and (B.26) gives

$$\begin{aligned} \|Z-\lambda\|_J^2 &= \|P\,(Z-\lambda)\|_J^2 + \|P^{\perp}\,(Z-\lambda)\,\|_J^2 \\ &= \left\| \begin{bmatrix} 0_{s_1\times 1} \\ J_{\delta\delta}^{-1}G_\delta \end{bmatrix} - \begin{bmatrix} 0_{s_1\times 1} \\ \lambda_\delta + J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta \end{bmatrix} \right\|_J^2 + \|K\,(Z-\lambda)\|_{M'JM}^2 \\ &= \|J_{\delta\delta}^{-1}G_\delta - \lambda_\delta - J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta\|_{J_{\delta\delta}}^2 + \|Z_\beta - \lambda_\beta\|_{(KJ^{-1}K')^{-1}}^2. \end{aligned} \tag{B.32}$$

Since $\Lambda = \Lambda_\beta \times \Lambda_\delta$ and $\Lambda_\delta = \mathbb{R}^{s_2}$, for any $\lambda_\beta \in \Lambda_\beta$

$$\inf_{\lambda_\delta \in \Lambda_\delta} \|J_{\delta\delta}^{-1}G_\delta - \lambda_\delta - J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta\|_{J_{\delta\delta}}^2 = \inf_{\lambda_\delta \in \mathbb{R}^r} \|J_{\delta\delta}^{-1}G_\delta - \lambda_\delta - J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta\|_{J_{\delta\delta}}^2 = 0,$$

so

$$\inf_{\lambda\in\Lambda} \|Z-\lambda\|_J^2 = \inf_{\lambda_\beta \in \Lambda_\beta} \|Z_\beta - \lambda_\beta\|_{(KJ^{-1}K')^{-1}}^2,$$

which proves the first equality of 2. holds. Turning to the second equality of 2., notice that

$$\begin{aligned} 0 &\leq \|Z_\beta - \lambda_\beta^\Lambda\|_{(KJ^{-1}K')^{-1}}^2 - \|Z_\beta - \lambda^{\Lambda_\beta}\|_{(KJ^{-1}K')^{-1}}^2 \\ &\leq \|Z_\beta - \lambda_\beta^\Lambda\|_{(KJ^{-1}K')^{-1}}^2 + \|J_{\delta\delta}^{-1}G_\delta - \lambda_\delta^\Lambda - J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta^\Lambda\|_{J_{\delta\delta}}^2 - \|Z_\beta - \lambda^{\Lambda_\beta}\|_{(KJ^{-1}K')^{-1}}^2 \\ &= \|Z-\lambda^\Lambda\|_J^2 - \|Z_\beta - \lambda^{\Lambda_\beta}\|_{(KJ^{-1}K')^{-1}}^2 = 0, \end{aligned}$$

where we have used (B.32) and the first equality of 2.
Point 3. follows from 2, and the fact that λ^{Λ_β} is unique due to the convexity of Λ_β. □

Lemma B.7. *Let $\{Y_t : t \in \mathbb{N}_0\}$, $Y_t = (X_t^{\odot 2\prime}, \sigma_t^{2\prime})'$, be the Markov chain generated by the ECCC-GARCH model (2.1)-(2.4) for $t \geq 1$, with fixed initial values $X_0 := x \in \mathbb{R}^d$ and $\sigma_0^2 := h \in (0,\infty)^d$, and with fixed $\theta = [\kappa_0', \mathrm{vec}(A_0)', \mathrm{vec}(B_0)', \mathrm{vech}^0(R_0)']'$. Suppose that $\rho(B_0) < 1$ and that the diagonal elements of A_0 are strictly positive. Let $p \in \mathbb{N}$, and suppose that the distribution, Γ, of $\varepsilon_t := R_0^{1/2}\eta_t$ admits a probability den-*

sity strictly positive on $\mathbb{R}^d$ *with* $\mathbb{E}[(\varepsilon_t^{\odot 2})^{\otimes p}] < \infty$*, and* $\rho(\mathbb{E}\{[A_0 \text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$*. Then* $\{Y_t : t \in \mathbb{N}_0\}$ *is geometrically ergodic on* $[0,\infty)^d \times (0,\infty)^d$*, and the associated strictly stationary process* $\{Y_t : t \in \mathbb{Z}\}$ *is geometrically* β*-mixing with* $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$*.*

Proof. The proof is similar to Pedersen (2015, Proof of Lemma B.8). Consider the process $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ given by $\sigma_t^2 = \kappa_0 + [A_0 \,\text{diag}(\varepsilon_{t-1}^{\odot 2}) + B_0]\sigma_{t-1}^2$, with $\sigma_0^2 = h$. Relying on the theory of Boussama et al. (2011), it follows from Pedersen (2015, Proof of Lemma B.8) that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is a Markov chain which is aperiodic and ψ-irreducible on $(0,\infty)^d$, see Meyn and Tweedie (2009, Section 4.2). These properties of the Markov chain allow us, due to Tjøstheim (1990), to consider a k-step drift criterion for the Markov chain for some $k \in \mathbb{N}$. Specifically, with $\mathscr{B}((0,\infty)^d)$ the Borel σ-field of $(0,\infty)^d$, we want to show that there exists a small set $\mathcal{K} \in \mathscr{B}((0,\infty)^d)$, positive constants $a < 1$ and $b < \infty$, and a Lyapunov function $V_\sigma : (0,\infty)^d \to [1,\infty)$ such that for some fixed $k \in \mathbb{N}$,

$$\mathbb{E}\left[V_\sigma(\sigma_k^2)|\sigma_0^2 = h\right] \leq aV_\sigma(h) + b \cdot \mathbb{1}(h \in \mathcal{K}) \quad \forall h \in (0,\infty)^d.$$

With ι_{d^p} a $(d^p \times 1)$ vector of ones, consider the function $V_\sigma(h) := 1 + \iota_{d^p}' h^{\otimes p}$, and, for some constant m sufficiently large, the set $\mathcal{K} := \{h \in (0,\infty)^d : \iota_{d^p}' h^{\otimes p} \leq m\}$.
For $t \in \mathbb{N}$, it holds that $(\sigma_{t+1}^2)^{\otimes p} = C_{t,p} + [A_0 \,\text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}(\sigma_t^2)^{\otimes p}$, where for $p \geq 2$ $C_{t,p} := \{C_{t,p-1} \otimes \sigma_t^2 + [A_0 \,\text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p-1} \otimes \kappa_0\}$ and $C_{t,1} := \kappa_0$. Recursions give that

$$(\sigma_{t+k}^2)^{\otimes p} = \sum_{i=0}^{k-1}\prod_{j=1}^{i}[A_0 \,\text{diag}(\varepsilon_{t+k-j}^{\odot 2}) + B_0]^{\otimes p} C_{t+k-1-i,p} + \prod_{i=1}^{k}[A_0 \,\text{diag}(\varepsilon_{t+k-i}^{\odot 2}) + B_0]^{\otimes p}(\sigma_t^2)^{\otimes p}.$$

Observe that

$$\mathbb{E}\left[V_\sigma(\sigma_k^2)|\sigma_0^2 = h\right] = \frac{1 + \iota_{d^p}'\tilde{C} + \iota_{d^p}'(\mathbb{E}\{[A_0 \,\text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\})^k h^{\otimes p}}{1 + \iota_{d^p}' h^{\otimes p}} V_\sigma(h),$$

where we have used that $\{\varepsilon_t\}$ is i.i.d. and where $\tilde{C}$ contains terms of h of lower order than p. Since $\rho(\mathbb{E}\{[A_0 \,\text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$ and choosing k sufficiently large, there exists an m large enough such that for $h \in \mathcal{K}^{\complement}$, $V_\sigma(h) \geq 1 + \iota_{d^p}'\tilde{C} + \iota_{d^p}'(\mathbb{E}\{[A_0 \,\text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\})^k h^{\otimes p}$. We conclude that suitable constants a and b exist. In line with Boussama et al. (2011, Section 4.6) it can be shown that $\mathcal{K}$ is small. It then holds that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is V_σ-geometrically ergodic. From Meitz and Saikkonen (2008, Proposition 1 and the comments immediately after) we conclude that $\{Y_t : t \in \mathbb{N}_0\}$ is V_Y-geometrically ergodic, for some suitable function V_Y :

$[0,\infty)^d \times (0,\infty)^d \to [1,\infty)$, and that the associated strictly stationary process $\{Y_t : t \in \mathbb{Z}\}$ is geometrically β-mixing. Moreover, $\mathbb{E}[\|(\sigma_t^2)^{\otimes p}\|] \leq \mathcal{C}\mathbb{E}[V_\sigma(\sigma_t^2)] < \infty$, and by using that $\mathbb{E}[(\varepsilon_t^{\odot 2})^{\otimes p}] < \infty$, we have that $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$. □

Lemma B.8. *Let $\{X_t : t \in \mathbb{Z}\}$, be a strictly stationary process generated by the ECCC-GARCH model (2.1)-(2.4) with fixed $\theta = [\kappa_0', \operatorname{vec}(A_0)', \operatorname{vec}(B_0)', \operatorname{vech}^0(R_0)']' \in \Theta$. For $p \in \mathbb{N}$ suppose that $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$. Then $\rho(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$, where $\varepsilon_t := R_0^{1/2}\eta_t$.*

Proof. The proof is similar to that of Ling and McAleer (2002, Proof of Theorem 2.1). Notice that $\mathbb{E}[(\varepsilon_t^{\odot 2})^{\otimes p}] < \infty$ is necessary for $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$, and that $\mathbb{E}[(\sigma_t^2)^{\otimes p}] < \infty$. Similar to the proof of Lemma B.7, we obtain for $k \in \mathbb{N}$

$$(\sigma_t^2)^{\otimes p} = \sum_{i=0}^{k-1} \prod_{j=1}^{i} [A_0 \operatorname{diag}(\varepsilon_{t-j}^{\odot 2}) + B_0]^{\otimes p} C_{t-1-i,\otimes p} + \prod_{i=1}^{k} [A_0 \operatorname{diag}(\varepsilon_{t-i}^{\odot 2}) + B_0]^{\otimes p} (\sigma_{t-k}^2)^{\otimes p}.$$

Since $\prod_{i=1}^{k} [A_0 \operatorname{diag}(\varepsilon_{t-i}^{\odot 2}) + B_0]^{\otimes p} (\sigma_{t-k}^2)^{\otimes p} \geq 0 \ \forall k$ and $C_{t-1-i,\otimes p} \geq \kappa_0^{\otimes p}$, we obtain

$$\infty > \mathbb{E}[(\sigma_t^2)^{\otimes p}] \geq \sum_{i=0}^{\infty} \left(\mathbb{E}\left\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\right\}\right)^i \kappa_0^{\otimes p}. \tag{B.33}$$

Since $(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) \geq 0$ and $\kappa_0^{\otimes p} \in (0,\infty)^{d^p}$, we have, in light of (B.33), that $\sum_{i=0}^{\infty} (\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\})^i$ converges, which is necessary and sufficient for $\rho(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$. □

Appendix C The LM_{ECCC} statistic of Nakatani and Teräsvirta (2009)

Proposition C.1. *Let $\hat{J}_T(\theta)$ and $\hat{S}_T(\theta)$ be defined by (4.3), let K_1 be defined by (4.4), and let $\tilde{\theta}_T$ be the constrained estimator given in (4.2). Consider the test statistic given by*

$$LM_{ECCC} = \frac{1}{2} T \hat{S}_T(\tilde{\theta}_T)' K_1' [K_1 \hat{J}_T(\tilde{\theta}_T)^{-1} K_1'] K_1 \hat{S}_T(\tilde{\theta}_T).$$

With the matrices J and Σ defined in (3.13) and λ^{Λ_0} defined in (4.5), let $\mathcal{L}(G) = N(0,\Sigma)$ and consider the partitions of J, G, and λ^{Λ_0}, according to $\theta = (\beta_1', \beta_2', \delta')'$, *given by*

$$J = \begin{pmatrix} J_{\beta_1\beta_1} & J_{\beta_1\beta_2} & J_{\beta_1\delta} \\ J_{\beta_2\beta_1} & J_{\beta_2\beta_2} & J_{\beta_2\delta} \\ J_{\delta\beta_1} & J_{\delta\beta_2} & J_{\delta\delta} \end{pmatrix}, \quad G = \begin{pmatrix} G_{\beta_1} \\ G_{\beta_2} \\ G_{\delta} \end{pmatrix}, \quad \text{and} \quad \lambda^{\Lambda_0} = \begin{pmatrix} \lambda^{\Lambda_0}_{\beta_1} \\ \lambda^{\Lambda_0}_{\beta_2} \\ \lambda^{\Lambda_0}_{\delta} \end{pmatrix}.$$

Under Assumptions 1-7 and $\mathcal{H}_0$,

$$LM_{ECCC} \xrightarrow{w} \frac{1}{2}\|\zeta\|^2_{(K_1 J^{-1} K_1')}, \tag{C.1}$$

where $\zeta := G_{\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}G_\delta + (J_{\beta_1\beta_2} - J_{\beta_1\delta}J_{\delta\delta}^{-1}J_{\delta\beta_2})\lambda^{\Lambda_0}_{\beta_2}$.
Suppose in addition that $\tilde{s}_1 = s_1$ and that Σ is positive definite. Then

$$LM_{ECCC} \xrightarrow{w} \sum_{i=1}^{m} \xi_i \chi^2_{m_i}, \tag{C.2}$$

where ξ_i, $i = 1, ..., m$, are the m distinct eigenvalues of $(1/2)\Omega^{1/2}(K_1 J^{-1} K_1')\Omega^{1/2}$, with $\Omega^{1/2}$ the positive definite matrix square root of the $(s_1 \times s_1)$ matrix Ω, given by

$$\Omega = \Sigma_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}\Sigma_{\delta\beta} - \Sigma_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta} + J_{\delta\beta}J_{\delta\delta}^{-1}\Sigma_{\beta\beta}J_{\delta\delta}^{-1}J_{\delta\beta},$$

and $\chi^2_{m_i}$, $i = 1, ..., m$ are mutually independent, and m_i is the multiplicity of ξ_i.
Finally, suppose furthermore that $\mathcal{L}(\eta_t) = N(0, I_d)$, then

$$LM_{ECCC} \xrightarrow{w} \chi^2_{s_1}. \tag{C.3}$$

Proof. Similar to the derivations in proof of Theorem 4.1 we obtain from a Taylor-type expansion,

$$\begin{aligned} \sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{\hat{L}_T(\theta_0)}{\partial\beta_1} + \frac{\partial^2\hat{L}_T(\theta^\star)}{\partial\beta_1\partial\theta'}\sqrt{T}(\tilde{\theta}_T - \theta_0) \\ &= \sqrt{T}\frac{\hat{L}_T(\theta_0)}{\partial\beta_1} + \frac{\partial^2\hat{L}_T(\theta^\star)}{\partial\beta_1\partial\beta_2'}\sqrt{T}(\tilde{\beta}_{2,T} - \beta_{2,0}) + \frac{\partial^2\hat{L}_T(\theta^\star)}{\partial\beta_1\partial\delta'}\sqrt{T}(\tilde{\delta}_T - \delta_0) \end{aligned}$$

where $\theta^\star$ is between $\tilde{\theta}_T$ and θ_0, and where the second equality follows from the fact that $\tilde{\beta}_{1,T} - \beta_{1,0} = 0_{\tilde{s}_1\times 1}$. Since δ_0 does not attain the bounds of Θ, we have by a Taylor-type expansion that

$$0_{s_2\times 1} = \frac{\hat{L}_T(\theta_0)}{\partial\delta} + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\theta'}(\tilde{\theta}_T - \theta_0)$$

$$= \frac{\hat{L}_T(\theta_0)}{\partial\delta} + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\beta_2'}(\tilde{\beta}_{2,T}-\beta_{2,0}) + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}(\tilde{\delta}_T-\delta_0), \quad (C.4)$$

where $\theta^{\star\star}$ is between $\tilde{\theta}_T$ and θ_0. Hence

$$(\tilde{\delta}_T-\delta_0) = -\left(\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}\right)^{-1}\left[\frac{\hat{L}_T(\theta_0)}{\partial\delta} + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\beta_2'}(\tilde{\beta}_{2,T}-\beta_{2,0})\right], \quad (C.5)$$

and substituting (C.5) into (C.4) and rearranging yield

$$\begin{aligned}\sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{\hat{L}_T(\theta_0)}{\partial\beta_1} - \frac{\partial^2\hat{L}_T(\theta^{\star})}{\partial\beta_1\partial\delta'}\left(\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}\right)^{-1}\sqrt{T}\frac{\hat{L}_T(\theta_0)}{\partial\delta} \\ &+ \left[\frac{\partial^2\hat{L}_T(\theta^{\star})}{\partial\beta_1\partial\beta_2'} - \frac{\partial^2\hat{L}_T(\theta^{\star})}{\partial\beta_1\partial\delta'}\left(\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}\right)^{-1}\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\beta_2'}\right]\sqrt{T}(\tilde{\beta}_{2,T}-\beta_{2,0}).\end{aligned}$$

From Lemma B.5, using that $\sqrt{T}(\tilde{\theta}_T-\theta_0)$ and $\sqrt{T}S_T(\theta_0)$ are $O_p(1)$,

$$\begin{aligned}\sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{L_T(\theta_0)}{\partial\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}\sqrt{T}\frac{L_T(\theta_0)}{\partial\delta} \\ &+ \left[J_{\beta_1\beta_2} - J_{\beta_1\delta}J_{\delta\delta}^{-1}J_{\delta\beta_2}\right]\sqrt{T}(\tilde{\beta}_{2,T}-\beta_{2,0}) + o_p(1) \\ &\overset{w}{\to} G_{\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}G_\delta + (J_{\beta_1\beta_2} - J_{\beta_1\delta}J_{\delta\delta}^{-1}J_{\delta\beta_2})\lambda_{\beta_2}^{\Lambda_0}, \quad (C.6)\end{aligned}$$

where we have used that the terms converge jointly due to point 1. of the proof of Theorem 3.1, and that $\sqrt{T}(\tilde{\theta}_T-\theta_0)\overset{w}{\to}\lambda^{\Lambda_0}$. Moreover, Lemma B.5 and the consistency of $\tilde{\theta}_T$ imply that

$$K_1\hat{J}_T(\tilde{\theta}_T)^{-1}K_1' = K_1J^{-1}K_1' + o_p(1). \quad (C.7)$$

Hence by combining (C.6) and (C.7) and by applying the continuous mapping theorem, we have shown that (C.1) holds.
Next, for the case $\tilde{s}_1 = s_1$, we have that β_2 vanishes such that

$$\begin{aligned}\sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{L_T(\theta_0)}{\partial\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}\sqrt{T}\frac{L_T(\theta_0)}{\partial\delta} + o_p(1) \\ &= (I_{s_1}, -J_{\beta_1\delta}J_{\delta\delta}^{-1})\sqrt{T}\frac{L_T(\theta_0)}{\partial\theta} + o_p(1) \\ &\overset{w}{\to} (I_{s_1}, -J_{\beta_1\delta}J_{\delta\delta}^{-1})G, \quad (C.8)\end{aligned}$$

where we have used arguments similar to the ones given above. It holds that

$$\mathcal{L}[(I_{s_1}, -J_{\beta_1\delta}J_{\delta\delta}^{-1})G] = N(0,\Omega). \quad (C.9)$$

Combining (C.7)-(C.9) and using White (1996, Theorem 8.6), we conclude that (C.2) holds. In the case where $\tilde{s}_1 = s_1$ and $\mathcal{L}(\eta_t) = N(0, I_d)$, the information equality implies that $2\Sigma = J$ and it is straightforward, using (B.23) and the continuous mapping theorem, to establish that (C.3) holds. □

Appendix D Additional details about the simulations

This section contains some additional details about the simulations reported in Section 5.

- The simulations are carried out in OxMetrics 7.0.
- All replications are based on a burn-in period of 1,000 observations, and all simulations are based on the same seed value.
- The computation of the QMLE $\hat{\theta}_T$ and the constrained QMLE $\tilde{\theta}_T$ is based on maximization of the log-likelihood function according to the `MaxSQP` function. For the computation of $\hat{\theta}_T$ we use the starting values:

$$\kappa = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.10 & 0.05 \\ 0.05 & 0.11 \end{bmatrix}, \quad B = \begin{bmatrix} 0.85 & 0.05 \\ 0.05 & 0.80 \end{bmatrix}, \quad \rho = 0.5.$$

  For the computation of $\tilde{\theta}_T$ we use the starting values:

$$\kappa = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.10 & \\ & 0.11 \end{bmatrix}, \quad B = \begin{bmatrix} 0.85 & \\ & 0.80 \end{bmatrix}, \quad \rho = 0.5.$$

- For the computation of the log-likelihood function, we use as initial value $\hat{h}_0 = T^{-1} \sum_{t=1}^{T} X_t^{\odot 2}$.
- The following constraints are imposed on the parameters for the optimization: With $\kappa = (\kappa_1, \kappa_2)'$, $\kappa_1, \kappa_2 \geq 0.000001$, $\rho \in [-0.99999, 0.99999]$, and all elements of the matrices A and B are nonnegative.
- All derivatives of the log-likelihood function are obtained by numerical techniques.
- If a replication yields an estimate $\hat{J}_T(\hat{\theta}_T)$ or $\hat{J}_T(\tilde{\theta}_T)$ that is found to be (numerically) singular, this replication is discarded from the calculations. The singu-

larity of the matrices was mainly an issue for the replications with $T = 1{,}000$ observations.

References

Andrews, D. W. K. (1998): "Estimation when a parameter is on a boundary: Part II," Yale University, unpublished manscript.

——— (1999): "Estimation when a parameter is on a boundary," *Econometrica*, 67, 1341–1383.

——— (2001): "Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 69, 683–734.

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

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.

Bali, T. G. and A. Hovakimian (2009): "Volatility spreads and expected stock returns," *Management Science*, 55, 1797–1812.

Bao, Y. and A. Ullah (2010): "Expectation of quadratic forms in normal and nonnormal variables with applications," *Journal of Statistical Planning and Inference*, 140, 1193 – 1205.

Bartkiewicz, K., A. Jakubowski, T. Mikosch, and O. Wintenberger (2011): "Stable limits for sums of dependent infinite variance random variables," *Probability Theory and Related Fields*, 150, 337–372.

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.

Berkes, I. and L. Horváth (2003): "The rate of consistency of the quasi-maximum likelihood estimator," *Statistics & Probability Letters*, 61, 133–143.

——— (2004): "The efficiency of the estimators of the parameters in GARCH processes," *The Annals of Statistics*, 32, 633–655.

Berkes, I., L. Horváth, and P. Kokoszka (2003): "Estimation of the maximal moment exponent of a GARCH(1, 1) sequence," *Econometric Theory*, 19, 565–586.

Berkes, I., L. Horváth, and J. Schauer (2010): "Non-central limit theorems for random selections," *Probability Theory and Related Fields*, 147, 449–479.

Billingsley, P. (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. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return," *The Review of Economics and Statistics*, 69, 542–47.

——— (1990): "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model," *The Review of Economics and Statistics*, 72, 498–505.

Boman, J. and F. Lindskog (2009): "Support theorems for the Radon transform and Cramér-Wold theorems," *Journal of Theoretical Probability*, 22, 683–710.

Bougerol, P. and N. Picard (1992): "Strict stationarity of generalized autoregressive processes," *The Annals of Probability*, 20, 1714–1730.

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

Boussama, F., F. Fuchs, and R. Stelzer (2011): "Stationarity and geometric ergodicity of BEKK multivariate GARCH models," *Stochastic Processes and their Applications*, 121, 2331–2360.

Brown, B. M. (1971): "Martingale Central Limit Theorems," *The Annals of Mathematical Statistics*, 42, 59–66.

Chernoff, H. (1954): "On the distribution of the likelihood ratio," *The Annals of Mathematical Statistics*, 25, 573–578.

Comte, F. and O. Lieberman (2000): "Second-order noncausality in multivariate GARCH processes," *Journal of Time Series Analysis*, 21, 535–557.

——— (2003): "Asymptotic theory for multivariate GARCH processes," *Journal of Multivariate Analysis*, 84, 61–84.

Conrad, C. and M. Karanasos (2010): "Negative volatility spillovers in the unrestricted ECCC-GARCH model," *Econometric Theory*, 26, 838–862.

Conrad, C. and E. Weber (2013): "Measuring persistence in volatility spillovers," Discussion Paper 543, Department of Economics, University of Heidelberg.

Conrad, J., M. N. Gultekin, and G. Kaul (1991): "Asymmetric predictability of conditional variances," *The Review of Financial Studies*, 4, 597–622.

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.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

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

Escanciano, J. C. (2009): "Quasi-maximum likelihood estimation of semi-strong GARCH models," *Econometric Theory*, 25, 561–570.

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.

——— (2014): "Variance targeting estimation of multivariate GARCH models," *Journal of Financial Econometrics*, to appear.

Francq, C. and J.-M. Zakoïan (2007): "Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero," *Stochastic Processes and their Applications*, 117, 1265–1284.

——— (2009): "Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons," *Journal of the American Statistical Association*, 104, 313–324.

——— (2010): *GARCH Models: Structure, Statistical Inference and Financial Applications*, Chichester: Wiley.

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

——— (2012b): "Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models," *Econometrica*, 80, 821–861.

Ghazal, G. (1996): "Recurrence formula for expectations of products of quadratic forms," *Statistics and Probability Letters*, 27, 101 – 109.

Granger, C. W. J. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 37, 424–438.

——— (1980): "Testing for causality: A personal viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.

Granger, C. W. J., R. P. Robins, and R. F. Engle (1986): *Wholesale and retail prices: Bivariate time-series modeling with forecastable error variances*, Cambridge, MA: The MIT Press, 1–16.

Hafner, C. M. (2003): "Fourth moment structure of multivariate GARCH models," *Journal of Financial Econometrics*, 1, 26–54.

Hafner, C. M. and H. Herwartz (2008): "Testing for causality in variance using multivariate GARCH models," *Annales d'Économie et de Statistique*, 215–241.

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.

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

Ketz, P. (2014): "Testing near or at the boundary of the parameter space," Unpublished manuscript, Department of Economics, Brown University.

Kristensen, D. and O. Linton (2004): "Consistent standard errors for target variance approach to GARCH estimation," *Econometric Theory*, 20, 990–993.

Kristensen, D. and A. Rahbek (2005): "Asymptotics of the QMLE for a class of ARCH(q) models," *Econometric Theory*, 21, 946–961.

Lange, T., A. Rahbek, and S. Jensen (2011): "Estimation and asymptotic inference in the AR-ARCH model," *Econometric Reviews*, 30, 129–153.

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.

Ledoit, O. and M. Wolf (2004): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

Ling, S. and M. McAleer (2002): "Necessary and sufficient moment conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) models," *Econometric Theory*, 18, 722–729.

——— (2003): "Asymptotic theory for a vector ARMA-GARCH model," *Econometric Theory*, 19, 280–310.

——— (2010): "A general asymptotic theory for time-series models," *Statistica Neerlandica*, 64, 97–111.

Loretan, M. and P. C. B. Phillips (1994): "Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets," *Journal of Empirical Finance*, 1, 211–248.

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

Magnus, J. R. and H. Neudecker (2007): *Matrix differential calculus with applications in statistics and econometrics*, John Wiley & Sons, 3rd ed.

Meitz, M. and P. Saikkonen (2008): "Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models," *Econometric Theory*, 24, 1291–1320.

Meyn, S. and R. L. Tweedie (2009): *Markov Chains and Stochastic Stability*, Cambridge: Cambridge University Press, 2 ed.

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, Boca Raton, FL: Chapman & Hall/CRC, vol. 99 of *Monographs on Statistics and Applied Probability*, chap. 5, 185–286.

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.

Nakatani, T. and T. Teräsvirta (2008): "Positivity constraints on the conditional variances in the family of conditional correlation GARCH models," *Finance Research Letters*, 5, 88–95.

——— (2009): "Testing for volatility interactions in the constant conditional correlation GARCH model," *The Econometrics Journal*, 12, 147–163.

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. McFadden, Amsterdam: Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 36, 2111–2245.

Noureldin, D., N. Shephard, and K. Sheppard (2014): "Multivariate rotated ARCH models," *Journal of Econometrics*, 179, 16–30.

Pedersen, R. S. (2015): "Targeting estimation of CCC-GARCH models with infinite fourth moments," *Econometric Theory*, FirstView, 1–34.

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

Perlman, M. D. (1969): "One-sided testing problems in multivariate analysis," *The Annals of Mathematical Statistics*, 40, 549–567, Corrections in *The Annals of Mathematical Statistics* (1971), 42, 1777.

Rahbek, A. (2004): "Stochastic properties of multivariate time series equations with emphasis on ARCH," *System Identification 2003*, 227–232.

Rahbek, A., E. Hansen, and J. Dennis (2002): "ARCH innovations and their impact on cointegration rank testing," Tech. Rep. 22, Department of Statistics and Operations Research, University of Copenhagen.

Ranga Rao, R. (1962): "Relations between weak and uniform convergence of measures with applications," *The Annals of Mathematical Statistics*, 33, 659–680.

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.

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

Robinson, P. M. (2005): "Efficiency improvements in inference on stationary and nonstationary fractional time series," *The Annals of Statistics*, 33, 1800–1842.

Ross, S. A. (1989): "Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy," *The Journal of Finance*, 44, 1–17.

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

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, Berlin: Springer, 201–229.

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

Tjøstheim, D. (1990): "Non-linear time series and Markov chains," *Advances in Applied Probability*, 22, 587–611.

Trapani, L. (2014): "Testing for (in)finite moments," Working paper 19, Cass Business School, Faculty of Finance.

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

White, H. (1994): *Estimation, Inference and Specification Analysis*, New York, NY: Cambridge University Press.

——— (1996): *Estimation, Inference and Specification Analysis*, Econometric Society Monographs, New York, NY: Cambridge University Press.

Woźniak, T. (2015): "Testing causality between two vectors in multivariate GARCH models," *International Journal of Forecasting*, 31, 876–894.

Zarantonello, E. H. (1971): "Projections on convex sets in Hilbert spaces and spectral theory," in *Contributions to Nonlinear Functional Analysis*, ed. by E. H. Zarantonello, New York, NY: Academic Press, 237–341.