Discussion Papers

Department of Economics

University of Copenhagen

No. 12-23

Multivariate Variance Targeting in the BEKK-GARCH Model

Rasmus Søndergaard Pedersen and Anders Rahbek

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

ISSN: 1601-2461 (E)

Multivariate Variance Targeting in the BEKK-GARCH Model

Rasmus Søndergaard Pedersen and Anders Rahbek*
Department of Economics, University of Copenhagen

November 14, 2012

Abstract

This paper considers 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 indeed necessary.

1 Introduction

As shown in Laurent, Rombouts, and Violante (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. Whereas large-sample properties of the VTE have recently been considered by Francq, Horváth, and Zakoïan (2011) for the univariate GARCH model, the properties have, to our knowledge, not been investigated before for the multivariate case. 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 BEKK-GARCH model, when estimated by VT estimation, are in line with existing literature for large-sample inference with quasi-maximum likelihood estimation (QMLE), see Hafner and Preminger (2009b). Included simulations indicate that our imposed sixth order moment restrictions cannot be relaxed for VT estimation, see also Avarucci, Beutner, and Zaffaroni (2012) where it is argued that at least fourth order moments are needed for QMLE. Thus our results points at that while VT estimation is simpler and even possible to implement for higher order systems, it requires no further moments than for QML based estimation.

*Rahbek is affliated with CREATES funded by The Danish National Research Foundation. Funding from The Danish Council for Independent Research | Social Sciences (grant no: 10-07974) is gratefully acknowledged.

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 BEKK model proposed by Engle and Kroner (1995), see e.g. Bauwens, Laurent, and Rombouts (2006) and Laurent, Rombouts, and Violante (2012). This is by now a well-known and much applied multivariate GARCH model; 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 a small number of series. This implies that it is difficult, if not impossible, to estimate the model through classical QMLE. 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. In light of this, one may reparametrize, or modify the BEKK model to obtain fewer parameters, while at the same time one may wish to consider a different estimation method from the usual Gaussian QMLE of all parameters. Examples of reducing the number of varying parameters in the optimization procedure include, for the BEKK model, diagonal-BEKK and scalar-BEKK, see Bauwens, Laurent, and Rombouts (2006).

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 QMLE. This two-step procedure saves the number of parameters in the optimization step which yields an optimization over fewer parameters regardless of the model has a restricted or unrestricted BEKK representation. Recently, Noureldin, Shephard, and Sheppard (2012) have proposed the so-called multivariate rotated ARCH (RARCH) model that 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 model – as contrary to the univariate GARCH model – is extensively discussed in Avarucci, Beutner, and Zaffaroni (2012), which argues that the high-order moment restrictions for QMLE cannot be relaxed. As mentioned simulations are included which support this view for the VT based estimation. Note also in this respect that the strong moment restrictions for asymptotic QML inference in the multivariate BEKK model are similarly in contrast to the very mild conditions found for univariate GARCH models, see e.g. Jensen and Rahbek (2004) and Francq and Zakoïan (2012) who find that asymptotic inference in the GARCH model is feasible even if the observed process is explosive.

The theoretical parts of this paper make extensive use of linear algebra and matrix differential calculus, see Lütkepohl (1996) and Magnus and Neudecker (2007) respectively.

Some notation throughout the paper: For $n \in \mathbb{N}$, I_n is the $n \times n$ identity matrix. The vector $\text{vec}(A)$ stacks the columns of a matrix A, and $\text{vech}(A)$ stacks the columns 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 $A \otimes B = \{a_{ij}B\}$. The matrix (Euclidean) norm of the matrix, or vector A, is defined as $\|A\| = (\text{tr}\{A'A\})^{1/2}$. With $\lambda_1, ..., \lambda_n$ the n distinct eigenvalues of a matrix A, $\rho(A) = \max_{i \in \{1,...,n\}} |\lambda_i|$ is the spectral radius of A. For an $n \times n$ matrix A, the $n^2 \times n^2$ commutation matrix K_{nn} has the property $K_{nn}\text{vec}(A) = \text{vec}(A')$. 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 IID$(0, I_d)$ sequence of random variables. $H_t^{1/2}$ 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 X_0 and H_0 fixed, and H_0 positive definite. Moreover, $C \in \mathbb{R}^d \times \mathbb{R}^d$ is positive definite and $A, B \in \mathbb{R}^d \times \mathbb{R}^d$, and hence H_t in (2.2) is positive definite.

Theorem 2.1 below states that, under certain assumptions, there exists a covariance stationary solution of the BEKK model. More precisely, if $\{X_t\}_{t=1,\dots,T}$ is covariance stationary, then $V[X_t] = E[H_t] = \Gamma$ where Γ is positive definite and solves the equation

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

Boussama, Fuchs, and Stelzer (2011, Lemma 4.2 and Proposition 4.3) establish that such solution exists if $\rho\left[(A \otimes A) + (B \otimes B)\right] < 1$. Variance targeting can be presented as rewriting the model so that the unconditional covariance matrix of X_t appears explicitly in the equation for H_t. Substituting (2.3) into (2.2) yields

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

and we say that H_t has the variance targeting BEKK representation, see also Noureldin, Shephard, and Sheppard (2012).

Define γ =vec(Γ) and

$$\lambda = \left[\mathrm{vec}\,(A)', \mathrm{vec}\,(B)'\right]', \tag{2.5}$$

and let θ denote the parameter vector of the model containing all the elements of Γ, A, and B, so that $\theta = [\gamma',\ \lambda']'$. Throughout the text we will use the notation $H_t(\gamma, \lambda)$, indicating that H_t depends on the parameters in γ and λ. Then the variance targeting BEKK model with parameter vector $[\gamma',\ \lambda']'$ is given by

$$X_t = H_t^{1/2}(\gamma, \lambda) Z_t, \tag{2.6}$$

where $t = 1, ..., T$, and Z_t is IID$(0, I_d)$, and

$$H_t(\gamma, \lambda) = \Gamma - A\Gamma A' - B\Gamma B' + AX_{t-1}X_{t-1}'A' + BH_{t-1}(\gamma, \lambda)B', \tag{2.7}$$

where γ =vec(Γ) and $\lambda = \left[\mathrm{vec}\,(A)'\ \ \mathrm{vec}\,(B)'\right]$. Note that the parameters in λ are restricted such that $\rho\left[(A \otimes A) + (B \otimes B)\right] < 1$ on the parameter space $\Theta \subset \mathbb{R}^{3d^2}$. Moreover, X_0 and H_0 are fixed, and H_0 and Γ are positive definite.

Some properties of a BEKK process have recently been investigated by Boussama, Fuchs, and Stelzer (2011) and may be summarized in the following theorem.

Theorem 2.1 *(Corollary to Theorem 2.4 of Boussama, Fuchs, and Stelzer (2011)) Let $\{X_t\}_{t=1,\ldots,T}$ be a process generated by a variance targeting BEKK process and define*

$$W_t = \left[vech\left(H_t\right)', X_t'\right]'. \tag{2.8}$$

Suppose that the distribution of Z_t is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, and that zero is an interior point of the support of the density. Then the Markov chain $\{W_t\}_{t=1,\ldots,T}$ is geometrically ergodic. Moreover, the strictly stationary and ergodic solution of the model associated with $\{W_t\}_{t=1,\ldots,T}$ has $E\left\|X_t\right\|^2 < \infty$ and $E\left\|H_t\right\| < \infty$ for all t.

Remark 2.1 *The geometric ergodicity of $\{W_t\}_{t=1,\ldots,T}$ implies that there exists a unique invariant distribution for W_t and that the marginal distribution of $\{W_t\}_{t=1,\ldots,T}$ converges to this stationary distribution when the chain is not initialized from its stationary distribution.*

Remark 2.2 *By initiating $\{W_t\}_{t=1,\ldots,T}$ from the invariant distribution, X_t is covariance stationary.*

Remark 2.3 *In Section 3 we show that asymptotic normality of the variance-targeting estimator can be established when $E\left\|X_t\right\|^6 < \infty$. Choosing a drift function for W_t in (2.8) which implies $E\left\|X_t\right\|^6 < \infty$ has, to our knowledge, not been considered anywhere in the literature. In Appendix C we establish conditions for geometric ergodicity and finite second, fourth, sixth, and eighth-order moments for the simpler BEKK-ARCH(1) model as in (2.2) with $B = 0$ and Z_t Gaussian.*

3 Variance targeting (VT) estimation

Whereas classical QMLE 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 the estimation method of variance targeting. VT estimation is a two-step estimation method where γ is estimated by a the sample unconditional covariance matrix of X_t, and next λ is estimated by QMLE by optimizing the VT log-likelihood with respect to λ. The two-step procedure yields the VTE of θ denoted $\widehat{\theta}_{VT}$. This will be explained in detail below.

Let Λ be a space of the same dimension as λ in (2.5). Note that the parameter $\theta \in \mathbb{R}^{3d^2}$ only contains $2d^2 + d\left(d+1\right)/2$ unique elements since Γ is symmetric. The VT procedure suggests that Γ is estimated by the sample covariance, so that

$$\widehat{\gamma}_{VT} = \text{vec}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right). \tag{3.1}$$

We observe that if X_t is strictly stationary and ergodic with $E\left\|X_t\right\|^2 < \infty$, $\widehat{\gamma}_{VT}$ is a (strongly) consistent estimator for $\gamma =$vec(Γ), by the ergodic theorem.

For the variance targeting BEKK model, the profiled quasi log-likelihood is given by

$$L_T(\gamma, \lambda) = \frac{1}{T}\sum_{t=1}^{T} l_t(\gamma, \lambda) \tag{3.2}$$

with

$$l_t(\gamma, \lambda) = \log\{\det[H_t(\gamma, \lambda)]\} + \operatorname{tr}\{X_t X_t' H_t^{-1}(\gamma, \lambda)\}. \tag{3.3}$$

Given an estimate (3.1) of γ, the VTE of λ is defined as

$$\widehat{\lambda}_{VT} = \underset{\lambda \in \Lambda}{\arg\min} L_T(\widehat{\gamma}_{VT}, \lambda). \tag{3.4}$$

and the two-step procedure yields the VTE of γ and λ

$$\widehat{\theta}_{VT} = \left[\widehat{\gamma}_{VT}', \widehat{\lambda}_{VT}'\right]'.$$

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 QMLE, one could denote the estimator QVTE.*

Compared to QMLE the VT procedure saves the number of varying parameters in the optimization step: 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. If A and B are diagonal matrices, which is a restriction that is often imposed in practice, the proportion of varying parameters, relative to the total number of parameters to be estimated, is small for a moderate dimension of the observed process. This suggests that the combination of a restricted BEKK model, say the diagonal, and VT allows for estimating high-dimensional systems.

For estimation of C in the original BEKK model in Definition 2.2, recall that

$$\widehat{C}_{VT} = \widehat{\Gamma}_{VT} - \widehat{A}_{VT}\widehat{\Gamma}_{VT}'\widehat{A}_{VT} - \widehat{B}_{VT}\widehat{\Gamma}_{VT}'\widehat{B}_{VT}. \tag{3.5}$$

Its asymptotic distribution is stated in Proposition 4.1 below.

4 Large-sample properties of VT estimation

In this section we establish the consistency and asymptotic normality of the VTE. The proofs are stated in Appendix A.

As in Comte and Lieberman (2003), Hafner and Preminger (2009b), and Francq, Horváth, and Zakoïan (2011), we assume that$\{X_t\}_{t=0,\dots,T}$ is strictly stationary and ergodic:

Assumption 4.1 *The assumptions of Theorem 2.1 are satisfied, and the observed process $\{X_t\}_{t=0,\dots,T}$ is generated by the strictly stationary and ergodic solution of a variance-targeting BEKK process.*

Note that one could weaken this assumption so that $\{X_t\}_{t=0,\dots,T}$ is initiated from a fixed value, see Jensen and Rahbek (2004).

In addition to Assumption (4.1) we make the following assumptions:

Assumption 4.2 *The true parameter $\theta_0 \in \Theta$ and Θ is compact.*

Assumption 4.3 *For $\lambda \in \Lambda$, if $\lambda \neq \lambda_0$ then $H_t(\gamma_0, \lambda) \neq H_t(\gamma_0, \lambda_0)$ almost surely, for all $t \geq 1$.*

We are now able to state the following theorem.

Theorem 4.1 *Under Assumptions 4.1, 4.2, and 4.3, as $T \to \infty$ the VTE satisfies*

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

Remark 4.1 *Assumptions 4.2, and 4.3 are in line with Comte and Lieberman (2003) and Hafner and Preminger (2009b).*

Remark 4.2 *The finite second-order moments of X_t, implied by Assumption 4.1, are in line with the moment restrictions for consistency of the VTE in the univariate case, see Francq, Horváth, and Zakoïan (2011). The relatively weak sufficient conditions of Theorem 4.1 suggest that consistency of the VTE applies for many practical purposes. Notice that the moment restrictions are stronger than the ones that are sufficient for consistency of the 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).*

In order to show that the VTE is asymptotically normal, we make two additional assumptions:

Assumption 4.4 $E\left\|X_t\right\|^6 < \infty$.

Assumption 4.5 *θ_0 is in the interior of Θ.*

Theorem 4.2 *Under Assumptions 4.1-4.5, as $T \to \infty$*

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

where the matrices J_0 and K_0 are stated in (A.10) and Ω_0 is stated in (B.36) below.

Remark 4.3 *Assumption 4.4 states that the observed process X_t is required to have finite sixth-order moments. The moment restrictions are required in order to show that the second-order derivatives of the log-likelihood function converges uniformly on the parameter space, see the proof of Lemma B.5 below. Notice that the requirement of sixth-order moments is stronger than the requirement of finite fourth-order moments found by Francq, Horváth, and Zakoïan (2011) for the univariate case. However, notice that if we choose $d = 1$, our model corresponds to the one considered by Francq, Horváth, and Zakoïan (2011) and Assumption 4.4 can be weakened such that only finite fourth-order moments of X_t are required. In the case where the dimension is greater than one, the structure of the BEKK model implies that high-order moments of the X_t are required to be finite. This issue is discussed extensively in Avarucci, Beutner, and Zaffaroni (2012). Notice that the moment conditions are just as weak as the ones found in existing literature on asymptotic normality of the QMLE, see Hafner and Preminger (2009b). Assumption 4.4 is a strong assumption that is rarely satisfied in practice, and illustrates the main drawback of the BEKK models: Standard large-sample inference requires moment conditions that are rarely satisfied in real-world applications. Assumption 4.5 is a standard assumption in the literature.*

Given the asymptotic distribution of $\widehat{\theta}_{VT}$, we may derive the asymptotic distribution of the VTE for C in the original BEKK model in (2.2):

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

$$\sqrt{T}\left(vec\left(\widehat{C}_{VT}\right) - vec\left(C_0\right)\right) \xrightarrow{D} N\left(0, \Sigma_0 \begin{pmatrix} I_{d^2} & 0 \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix} \Omega_0 \begin{pmatrix} I_{d^2} & 0 \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix}' \Sigma_0' \right),$$

where

$$\underset{(d^2 \times 3d^2)}{\Sigma_0} = \left([I_{d^2} - (A_0 \otimes A_0) - (B_0 \otimes B_0)] \quad - [I_{d^2} + K_{dd}] \left[(A_0 \Gamma_0) \otimes I_d \right] \quad - [I_{d^2} + K_{dd}] \left[(B_0 \Gamma_0) \otimes I_d \right] \right).$$

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. In the first case the sufficient moment restrictions for asymptotic normality, see Theorem 4.2, are satisfied - in particular the data-generating process (DGP) has finite sixth-order moments. In the second case the DGP does not have finite sixth-order moments, but finite fourth-order moments. Hence the conditions of Theorem 4.2 are violated, so the VTE for the entire parameter vector may not be asymptotically normal. However, the moment restrictions for asymptotic normality of the VTE for γ are satisfied. In the last case the DGP has only finite second-order moments which suggests that even the VTE of γ cannot be asymptotically normally distributed. In order to keep things simple we focus on the bivariate diagonal-BEKK-ARCH(1) with Gaussian noise , that is the process in (2.2) with $d = 2$, A diagonal, $B = 0$, and Z_t IIDN$(0, I_2)$. In Appendix C we establish conditions for the matrix A in a BEKK-ARCH(1) process such that $\{X_t\}_{t=1,...,T}$ is geometrically ergodic and such that certain moments of the stationary solution are finite.

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

$$\begin{aligned} X_t &= H_t^{1/2} Z_t, \; Z_t \text{ IIDN}(0, I_2), \text{ and } H_t = C + A X_{t-1} X_{t-1}' A', && (5.1) \\ \text{with } C &= (C_{ij})_{i,j=1,2} = \begin{pmatrix} 0.8 & 0.5 \\ 0.5 & 0.7 \end{pmatrix}. && (5.2) \end{aligned}$$

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

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

and observe that $\rho(A \otimes A) = 0.36$. By Theorem C.1 the stationary solution of the process has $E\left\|X_t\right\|^6 < \infty$, 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 VTE using the G@RCH Package version 6.1 for OxMetrics 6.1.

Figure 5.1 contains density and Q-Q 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.

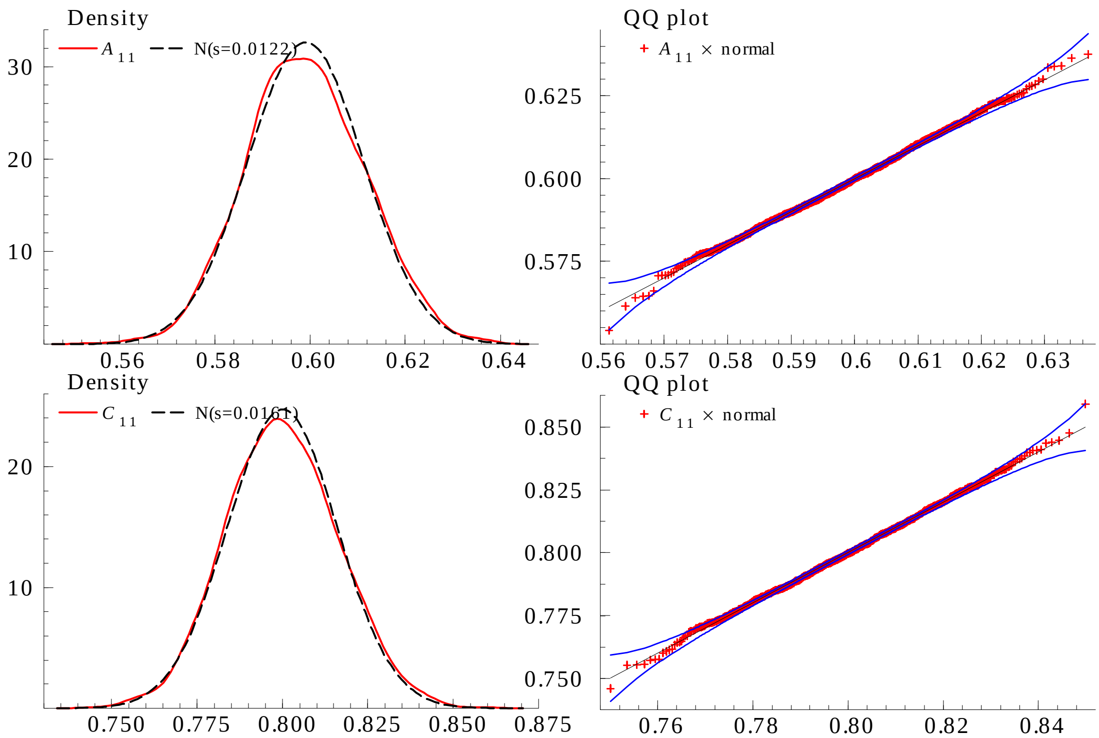

Figure 5.1: Density and Q-Q plots of $N = 1000$ VT estimates of A_{11} and C_{11} of the process (5.1)-(5.3). In the density plots the red line is the plot of the estimated density of the VT estimates, and the black dashed line is the plot for the normal distribution. The Q-Q plots compare the quantiles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

We now turn to the second case where the DGP does not satisfy the conditions in Theorem 4.2.

5.2 Case 2: The DGP does not satisfy the sufficient conditions for asymptotic normality

Next we consider the DGP (5.1)-(5.2) and choose A such that $E\left\|X_t\right\|^4 < \infty$, but $E\left\|X_t\right\|^6$ is not finite. We set

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

so that $\rho(A \otimes A) = 0.75^2 = 0.5625$. This implies that $E\left\|X_t\right\|^6$ is not finite, however $\rho(A \otimes A) < \frac{1}{\sqrt{3}} \approx 0.5774$, so we have that the DGP is geometrically ergodic with $E\left\|X_t\right\|^4 < \infty$ 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 VTE.

Figure 5.2 contains density and Q-Q 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 normally distributed. The density is skewed compared to normal distribution, which can also be deduced by the S-shape of the points in the Q-Q plot. The estimates of C_{11} do seem to fit a normal distribution, except for a few outliers (see Q-Q plot). In the following we explain why this can happen.

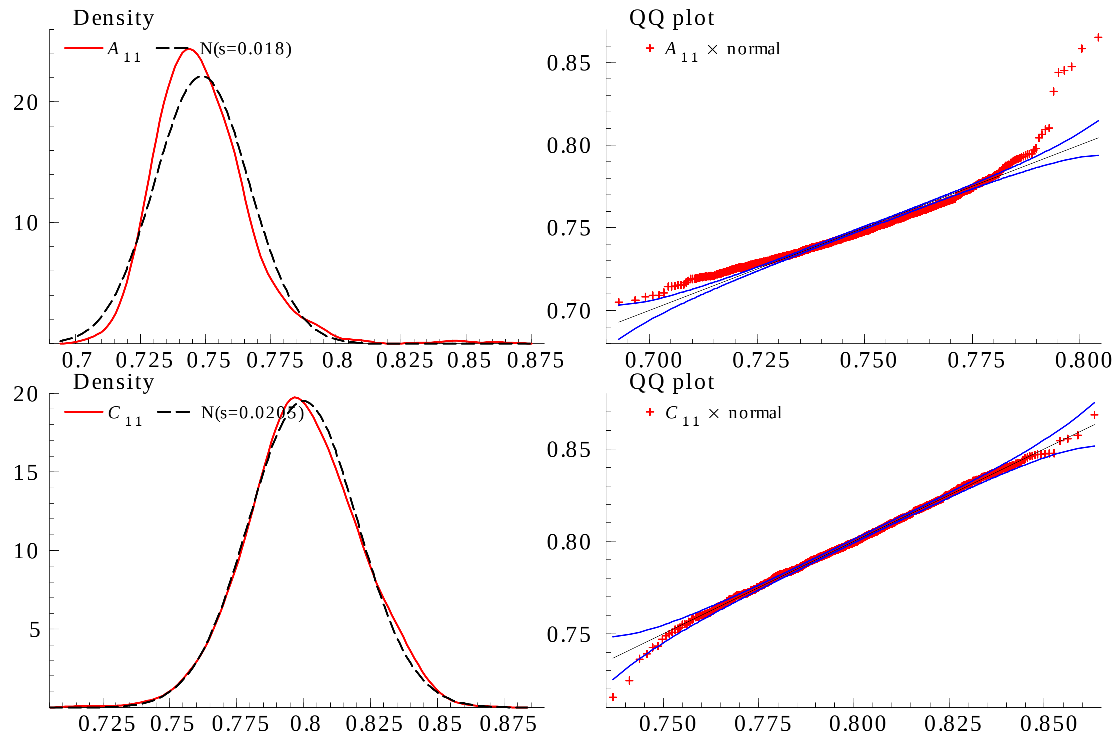

Figure 5.2: Density and Q-Q plots of $N = 1000$ VT estimates of A_{11} and C_{11} of the process (5.1),(5.2),(5.4). In the density plots the red line is the plot of the estimated density of the VT estimates, and the black dashed line is the plot for the normal distribution. The Q-Q plots compare the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

Recall that $\mathrm{vec}(\widehat{C}_{VT}) = [I_{d^2} - (\widehat{A}_{VT} \otimes \widehat{A}_{VT})]\widehat{\gamma}_{VT}$, so the distribution of $\mathrm{vec}(\widehat{C}_{VT})$ (or more correctly $\sqrt{T}[\mathrm{vec}(\widehat{C}_{VT}) - \mathrm{vec}(C_0)]$) depends on the distribution of $(\widehat{A}_{VT} \otimes \widehat{A}_{VT})$ and $\widehat{\gamma}_{VT}$. Recall that $\widehat{\gamma}_{VT}$ is asymptotically Gaussian if $E\left\|X_t\right\|^4 < \infty$ by the Central Limit Theorem, which can be verified by observing that $\mathrm{vec}(\widehat{\Gamma}_{VT})$ is given by (B.33), and that $\sqrt{T}[\mathrm{vec}(\widehat{\Gamma}_{VT}) - \mathrm{vec}(\Gamma_0)]$ is asymptotically Gaussian, if $E\left\|\mathcal{A}_t\right\| < \infty$, see proof of Lemma B.9. This is the case if $E\left\|X_t\right\|^4 < \infty$, which holds for our choice of DGP, so $\sqrt{T}[\mathrm{vec}(\widehat{\Gamma}_{VT}) - \mathrm{vec}(\Gamma_0)]$ is indeed asymptotically Gaussian. Next

$$\begin{aligned} \sqrt{T}\mathrm{vec}\left(\widehat{C}_{VT} - C_0\right) &= \left[I_{d^2} - \left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right)\right]\sqrt{T}\mathrm{vec}\left(\widehat{\Gamma}_{VT} - \Gamma_0\right) \\ &\quad - \sqrt{T}\left[\left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right) - (A \otimes A)\right]\mathrm{vec}\left(\Gamma_0\right). \end{aligned} \tag{5.5}$$

If $E\left\|X_t\right\|^4 < \infty$ the first term of the right hand side of (5.5) converges to a Gaussian variable, and determines the distribution of $\sqrt{T}\mathrm{vec}\left(\widehat{C}_{VT} - C\right)$ if $\left[\left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right) - (A \otimes A)\right]$ converges to some (unknown) distribution with a rate of $T^{1/2+\delta}$ for some $\delta > 0$. Suppose $\left[\left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right) - (A \otimes A)\right] = O_P\left(1/T^{1/2+\delta}\right)$, then $\sqrt{T}\left[\left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right) - (A \otimes A)\right] = o_P(1)$, and hence
$\sqrt{T}\mathrm{vec}\left(\widehat{C}_{VT} - C\right) = \left[I_{d^2} - \left(\widehat{A}_{VT} \otimes \widehat{A}_{VT}\right)\right]\sqrt{T}\mathrm{vec}\left(\widehat{\Gamma}_{VT} - \Gamma_0\right) + o_P(1)$, which ensures that $\sqrt{T}\mathrm{vec}\left(\widehat{C}_{VT} - C_0\right)$ is asymptotically normally distributed.

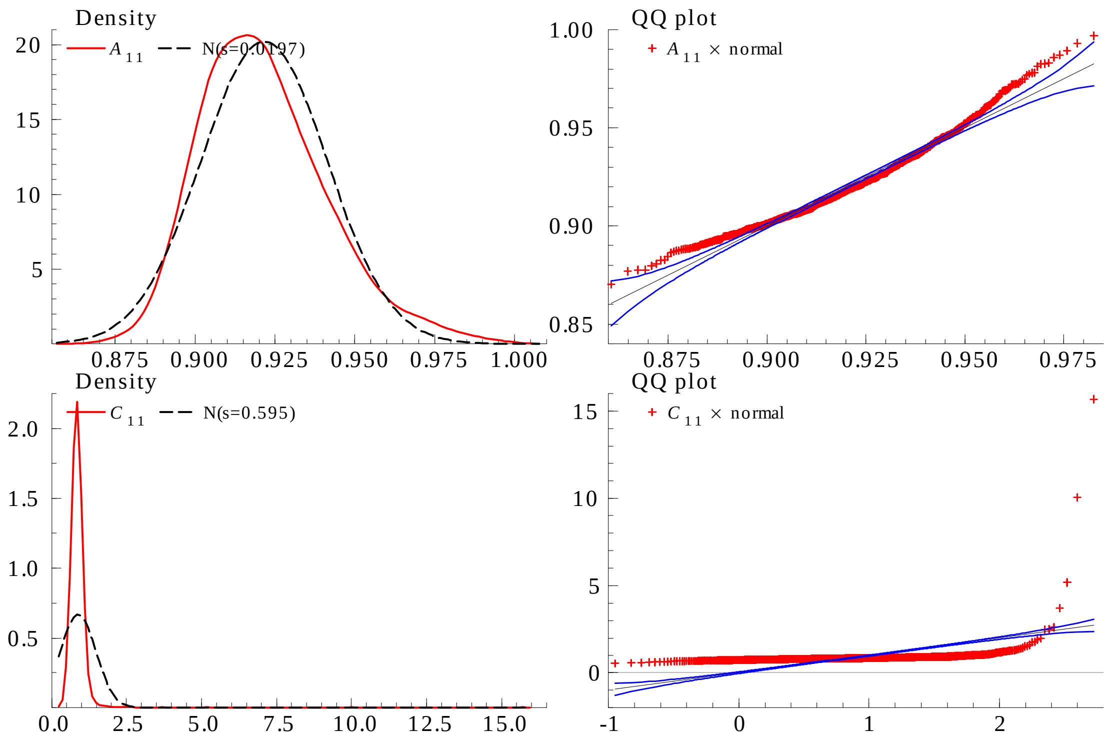

Figure 5.3: Density and Q-Q plots of $N = 1000$ VT estimates of A_{11} and C_{11} of the process (5.1),(5.2),(5.6). In the density plots the red line is the plot of the estimated density of the VT estimates, and the black dashed line is the plot for the normal distribution. The Q-Q plots compare the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

Next we turn to the case where $E\|X_t\|^2 < \infty$, but $E\|X_t\|^4$ is not finite.

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

Finally, 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. We set

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

and we have that $\rho(A \otimes A) = 0.95^2 = 0.9025$. This implies that $E\|X_t\|^4$ is not finite, however $\rho(A \otimes A) < 1$, so we have that the DGP is geometrically ergodic with $E\|X_t\|^2 < \infty$ by Theorem C.1. As in Case 1 and 2 we consider $N = 1000$ realizations of the DGP and estimate A and C by VTE.

Figure 5.3 contains density and Q-Q 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 normally distributed. In light of Case 2 this might be explained by the fact that $\sqrt{T}\text{vec}\left(\widehat{\Gamma}_{VT} - \Gamma\right)$ is not asymptotically normal as $E\|X_t\|^4$ is not finite.

Briefly, 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 showed that when relaxing the moment restrictions, $\widehat{A}_{VT}$ is no longer asymptotically normally distributed. This indicates that $E\|X_t\|^6 < \infty$ is a necessary moment restriction for doing standard large-sample inference in the BEKK-ARCH(1) model when estimated by VTE. Case 2 also showed that $\widehat{C}_{VT}$ is asymptotically normal even if $E\|X_t\|^6$ is not finite (but $E\|X_t\|^4 < \infty$) , which might be explained by the fact that asymptotic normality of $\widehat{\Gamma}_{VT}$ only requires that $E\|X_t\|^4 < \infty$. Case 3 showed that when $E\|X_t\|^2 < \infty$ but $E\|X_t\|^4$ is not finite, neither $\widehat{A}_{VT}$ nor $\widehat{C}_{VT}$ are 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 variance of the observed process appears explicitly in the model equation. This yields a reparametrized (variance-targeting) model given by (2.6)-(2.7). The parameters of the model are estimated in two steps yielding the VTE: The variance 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.

An obvious way to extend our results is to consider the general BEKK(p,q,k) model and the multivariate Rotated GARCH (RARCH) model recently proposed in Noureldin, Shephard, and Sheppard (2012). The model and the proposed two-step estimation procedure has some similarities to VTE, and it may be possible to exploit some of our theoretical results when investigating the asymptotic properties of the two-step estimator for the RARCH.

A Proofs of Theorems

In the asymptotic analysis we assume that the observed process $\{X_t\}_{t=0,\dots,T}$ is strictly stationary and ergodic, see Assumption 4.1. Throughout the text we use the probability measure where $W_t = \left[\text{vech}\left(H_t\right)', X_t'\right]'$ in (2.8) is strictly stationary and ergodic with appropriate moments finite. We define for $t \geq 1$

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

where $H_0(\gamma,\lambda)$ is strictly stationary. For the recursions defining $H_t(\gamma,\lambda)$ in (A.1) it is useful to introduce also $H_{t,h}(\gamma,\lambda)$ given by

$$H_{t,h}(\gamma,\lambda) = \Gamma - A\Gamma A' - B\Gamma B' + AX_{t-1}X_{t-1}'A' + BH_{t-1,h}(\gamma,\lambda)B', \tag{A.2}$$

where $H_{0,h}(\gamma,\lambda) = h$ is fixed and positive definite. We observe that as both recursions in (A.1) and (A.2) are defined for the same strictly stationary $\{X_t\}_{t=0,\dots,T}$,

$$\text{vec}\left[H_t(\gamma,\lambda) - H_{t,h}(\gamma,\lambda)\right] = (B\otimes B)\,\text{vec}\left[H_{t-1}(\gamma,\lambda) - H_{t-1,h}(\gamma,\lambda)\right], \; t\geq 1. \tag{A.3}$$

Recall that

$$L_T(\gamma,\lambda) = \frac{1}{T}\sum_{t=1}^{T} l_t(\gamma,\lambda), \tag{A.4}$$

with

$$l_t(\gamma,\lambda) = \log\left\{\det\left[H_t\left(\gamma,\lambda\right)\right]\right\} + \operatorname{tr}\left\{X_t X_t' H_t^{-1}\left(\gamma,\lambda\right)\right\}, \tag{A.5}$$

and $H_t(\gamma,\lambda)$ given by (A.1). To distinguish between $H_t(\gamma,\lambda)$ and $H_{t,h}(\gamma,\lambda)$ we introduce correspondingly

$$L_{T,h}(\gamma,\lambda) = \frac{1}{T}\sum_{t=1}^{T} l_{t,h}(\gamma,\lambda) \tag{A.6}$$

with

$$l_{t,h}(\gamma,\lambda) = \log\left\{\det\left[H_{t,h}\left(\gamma,\lambda\right)\right]\right\} + \operatorname{tr}\left\{X_t X_t' H_{t,h}^{-1}\left(\gamma,\lambda\right)\right\}, \tag{A.7}$$

with $H_{t,h}\left(\gamma,\lambda\right)$ given by (A.2).

A.1 Proof of Theorem 4.1

In order to make the proof readable, most of its steps rely on lemmas stated and proved in Section B.1 below.

Observe initially that by the ergodic theorem, as $T \to \infty$

$$\widehat{\gamma}_{VT} \overset{a.s.}{\to} \gamma_0. \tag{A.8}$$

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

$$\begin{aligned}
E\left[l_t\left(\gamma_0,\widehat{\lambda}_{VT}\right)\right] &< L_T\left(\gamma_0,\widehat{\lambda}_{VT}\right) + \varepsilon/5 \ \text{ by Lemma B.3} \\
L_T\left(\gamma_0,\widehat{\lambda}_{VT}\right) &< L_{T,h}\left(\widehat{\gamma}_{VT},\widehat{\lambda}_{VT}\right) + \varepsilon/5 \ \text{ by Lemma B.1} \\
L_{T,h}\left(\widehat{\gamma}_{VT},\widehat{\lambda}_{VT}\right) &< L_{T,h}\left(\widehat{\gamma}_{VT},\lambda_0\right) + \varepsilon/5 \ \text{ by (3.4)} \\
L_{T,h}\left(\widehat{\gamma}_{VT},\lambda_0\right) &< L_T\left(\gamma_0,\lambda_0\right) + \varepsilon/5 \ \text{ by Lemma B.1} \\
L_T\left(\gamma_0,\lambda_0\right) &< E\left[l_t\left(\gamma_0,\lambda_0\right)\right] + \varepsilon/5 \text{ by Lemma B.3.}
\end{aligned}$$

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 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.8), 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, in order to make the proof readable, most of the steps rely on lemmas stated in Section B.2. By Assumption 4.5, (3.4), and the mean-value theorem

$$0 = \frac{\partial L_{T,h}\left(\gamma_0,\lambda_0\right)}{\partial\lambda} + K_{T,h}\left(\theta^*\right)\left(\widehat{\gamma}_{VT} - \gamma_0\right) + J_{T,h}\left(\theta^*\right)\left(\widehat{\lambda}_{VT} - \lambda_0\right) \tag{A.9}$$

where

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

and θ^* on the line between θ_0 and $\widehat{\theta}_{VT}$, see also the proof of Lemma 1 in Jensen and Rahbek (2004). Let

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

By Lemma B.6, Lemma B.7, and Theorem 4.1, $J_T(\theta^*)$ is invertible with probability approaching one, so by Lemma B.11

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

Hence

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

Define

$$
J_0 := E\left[\left.\frac{\partial^2 l_t(\gamma,\lambda)}{\partial \lambda \partial \lambda'}\right|_{\theta=\theta_0}\right] \quad \text{and} \quad K_0 := E\left[\left.\frac{\partial^2 l_t(\gamma,\lambda)}{\partial \lambda \partial \gamma'}\right|_{\theta=\theta_0}\right]. \tag{A.10}
$$

By Lemma B.6 and Theorem 4.1

$$
\begin{pmatrix} I_{d^2} & 0 \\ -J_T^{-1}(\theta^*)K_T(\theta^*) & -J_T^{-1}(\theta^*) \end{pmatrix} \xrightarrow{P} \begin{pmatrix} I_{d^2} & 0 \\ -J_0^{-1}K_0 & -J_0^{-1} \end{pmatrix}.
$$

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

A.3 Proof of Proposition 4.1

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

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

We have that

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

and

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

Since Γ is symmetric

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

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

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

The distribution of $\sqrt{T}\left[\mathrm{vec}\left(\widehat{C}_{VT}\right) - \mathrm{vec}\left(C_0\right)\right]$ now follows by the delta method.

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, see also Lütkepohl (1996).

If the matrix A is positive definite we write $A > 0$, and if A is positive semi-definite we write $A \geq 0$. For the matrices A, B, C, and D, suppose $ABCD$ is defined and square. Then

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

The spectral norm of the matrix A is defined as $\|A\|_{spec} = \sqrt{\rho\left(A'A\right)}$. For the matrices A and B, if AB is well-defined,

$$\left|\mathrm{tr}\left(AB\right)\right| \leq \|A\|\,\|B\|, \tag{B.1}$$

$$\|AB\| \leq \|A\|_{spec}\|B\|, \quad \|AB\| \leq \|A\|\,\|B\|_{spec}, \text{ and } \quad \|A+B\|_{spec} \leq \|A\|_{spec} + \|B\|_{spec}. \tag{B.2}$$

If A is $n\times n$, then

$$\|A\|_{spec} \leq \|A\| \leq \sqrt{n}\,\|A\|_{spec}. \tag{B.3}$$

For an $n\times n$ matrix $A > 0$ with eigenvalues $\lambda_1\left(A\right),...,\lambda_n\left(A\right)$, it holds that

$$\log\det(A) = \sum_{i=1}^{n}\log\lambda_i\left(A\right) \leq \sum_{i=1}^{n}\lambda_i\left(A\right) = \mathrm{tr}\left(A\right). \tag{B.4}$$

Moreover,

$$\log\det(A) = \log\left(\det(A'A)\right)^{1/2} \leq n\log\left(\rho(A'A)\right)^{1/2} = n\log\|A\|_{spec}, \tag{B.5}$$

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

$$\mathrm{tr}\left(A\otimes B\right) = \mathrm{tr}\left(A\right)\mathrm{tr}\left(B\right). \tag{B.6}$$

Consider an $n\times n$ matrix $A \geq 0$ and an $n\times n$ matrix $B > 0$ with eigenvalues $\lambda_1\left(B\right) \leq \cdots \leq \lambda_n\left(B\right)$. Let $\lambda_1\left(A+B\right) \leq \cdots \leq \lambda_n\left(A+B\right)$ denote the eigenvalues of $\left(A+B\right)$, Then

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

by Result 4 in Section 5.3.2 of Lütkepohl (1996). Moreover,

$$0 < \lambda_i\left((A+B)^{-1}\right) \leq \lambda_i\left(B^{-1}\right), \; i = 1, ..., n.$$

Hence

$$0 < \operatorname{tr}\left[(A+B)^{-1}\right] \leq \operatorname{tr}\left(B^{-1}\right). \tag{B.7}$$

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

$$\det(A+B) \geq \det(A), \tag{B.8}$$

by Result 11 in Section 4.2.6 of Lütkepohl (1996).
For two positive semi-definite $n \times n$ matrices A and B, it holds that

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

by Result 12 in Section 4.2.6 of Lütkepohl (1996).

For some matrix A we introduce the notation $A^{\otimes 2} := (A \otimes A)$.

B.1 Lemmas for the proof of consistency

Lemma B.1 *Under Assumptions 4.1-4.3, as* $T \to \infty$

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

where $L_T(\gamma, \lambda)$ *is stated in (A.4) and* $L_{T,h}(\widehat{\gamma}_{VT}, \lambda)$ *is stated in (A.6).*

Proof. We have that

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

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 (2.7), $\Gamma - A\Gamma A' - B\Gamma B' > 0$ 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.8) and (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.7). As the eigenvalues of $H_t(\gamma,\lambda)$ are continuous in γ and λ, and Θ is compact,

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

and, likewise, $\sup_{\theta\in\Theta}\left\|H_{t,h}^{-1}(\gamma,\lambda)\right\| \leq K$.
By (A.8) we have that for T sufficiently large almost surely

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

Next, we note that

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

As $\rho(A^{\otimes 2}+B^{\otimes 2}) < 1$ on Θ it follows from Proposition 4.5 of Boussama, Fuchs, and Stelzer (2011) that $\rho(B^{\otimes 2}) < 1$ on Θ. Hence for any i we have that

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

As in Francq, Horváth, and Zakoïan (2011, p.644), (B.14), the compactness of Θ, (A.8), and (B.15) imply that as $T \to \infty$

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

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

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

where the first inequality follows from (B.5), the second from (B.2) and (B.3), and the third follows from the fact that $\log(x)\leq x-1$ for $x\geq 1$. Likewise,

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

where the inequalities follow by (B.1) and (B.13) respectively. By (B.16) we conclude that

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

By Markov's inequality and $E\left\|X_t\right\|^2<\infty$, it follows that for any $\varepsilon>0$

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

By the Borel-Cantelli lemma $\phi^t\left\|X_t\right\|^2 \stackrel{a.s.}{\rightarrow} 0$ as $t \rightarrow \infty$. It now follows by Cesàro's mean theorem that $\frac{1}{T} \sum_{t=1}^T \phi^t\left\|X_t\right\|^2 \stackrel{a.s.}{\rightarrow} 0.$, and we conclude that (B.10) holds. ■

Lemma B.2 *Under Assumptions 4.1-4.3,*

$$E \sup_{\theta \in \Theta}\left|l_t(\gamma, \lambda)\right| \leq K .$$

Proof. We note that

$$\begin{aligned} \operatorname{vec}\left[H_t(\gamma, \lambda)\right] & =\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right) \gamma+A^{\otimes 2} \operatorname{vec}\left(X_{t-1} X_{t-1}^{\prime}\right)+B^{\otimes 2} \operatorname{vec}\left[H_{t-1}(\gamma, \lambda)\right] \\ & =\sum_{i=0}^{\infty}\left(B^{\otimes 2}\right)^i\left[\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right) \gamma+A^{\otimes 2} \operatorname{vec}\left(X_{t-1-i} X_{t-1-i}^{\prime}\right)\right] \quad \text{(B.17)}\end{aligned}$$

so

$$\sup_{\theta \in \Theta}\left\|\operatorname{vec}\left[H_t(\gamma, \lambda)\right]\right\| \leq \sum_{i=0}^{\infty} \sup_{\theta \in \Theta}\left\|\left(B^{\otimes 2}\right)^i\left[\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right) \gamma+A^{\otimes 2} \operatorname{vec}\left(X_{t-1-i} X_{t-1-i}^{\prime}\right)\right]\right\| .$$

Notice that

$$\begin{aligned} & \sum_{i=0}^{\infty} E\left\{\sup_{\theta \in \Theta}\left\|\left(B^{\otimes 2}\right)^i\left[\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right) \gamma+A^{\otimes 2} \operatorname{vec}\left(X_{t-1-i} X_{t-1-i}^{\prime}\right)\right]\right\|\right\} \\ \leq & \sum_{i=0}^{\infty}\left(K \phi^t+K \phi^t E\left\|X_t\right\|^2\right)<\infty .\end{aligned}$$

By Theorem 9.2 of Jacod and Protter (2003) we conclude that

$$E\left[\sup_{\theta \in \Theta}\left\|H_t(\gamma, \lambda)\right\|\right] \leq K . \quad \text{(B.18)}$$

Now

$$\begin{aligned} E\left[\sup_{\theta \in \Theta}\left|l_t(\gamma, \lambda)\right|\right] & =E\left[\sup_{\theta \in \Theta}\left|\log \operatorname{det}\left[H_t(\gamma, \lambda)\right]+\operatorname{tr}\left[X_t X_t^{\prime} H_t^{-1}(\gamma, \lambda)\right]\right|\right] \\ & \leq E\left[\sup_{\theta \in \Theta}\left|\operatorname{tr}\left[H_t(\gamma, \lambda)\right]+\operatorname{tr}\left[X_t X_t^{\prime} H_t^{-1}(\gamma, \lambda)\right]\right|\right] \\ & \leq E\left\{\sup_{\theta \in \Theta}\left|K\left[\left\|H_t(\gamma, \lambda)\right\|+\left\|X_t X_t^{\prime} H_t^{-1}(\gamma, \lambda)\right\|\right]\right|\right\} \\ & \leq K\left[E \sup_{\theta \in \Theta}\left\|H_t(\gamma, \lambda)\right\|\right]+K E\left[\sup_{\theta \in \Theta}\left\|X_t\right\|^2\left\|H_t^{-1}(\gamma, \lambda)\right\|\right] \\ & \leq K+K E\left\|X_t\right\|^2 \\ & \leq K,\end{aligned}$$

where the first inequality follows from (B.4), the second from (B.1), the fourth from (B.18) and (B.12), and the last inequality follows by the fact that $E\left\|X_t\right\|^2 < \infty$. ■

Lemma B.3 *Under Assumptions 4.1-4.3, 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.4) and (A.5), respectively.*

Proof. The result follows by Lemma B.2 and the Uniform Law of Large Numbers for stationary ergodic processes, see Theorem A.2.2 of White (1994). ■

Lemma B.4 *Under Assumptions 4.1-4.3,*

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

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

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

Proof. $E\left|l_t(\gamma_0, \lambda_0)\right| < \infty$ follows from Lemma B.2.
Following the steps from Section 3 in Comte and Lieberman (2003), suppose $\lambda \neq \lambda_0$ and let $\{e_{it} : i = 1, .., d\}$ be the (positive) eigenvalues of $H_t(\gamma_0, \lambda_0) H_t^{-1}(\gamma_0, \lambda)$ for a fixed t. Note that

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

By the law of iterated expectations and since Z_t is independent of $\mathcal{F}_{t-1} = \sigma\left(X_{t-1}, X_{t-2,...}\right)$,

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

Moreover,

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

Hence

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

as $\log x \leq x-1$ for all $x \geq 0$. Since $\log x = x-1$ if and only if $x=1$, the inequality is strict unless $e_{it}=1$ for all i almost surely. $e_{it}=1$ for all i almost surely is equivalent to $H_t(\gamma_0,\lambda) = H_t(\gamma_0,\lambda_0)$ almost surely, but this cannot be the case in light of Assumption 4.3. Hence the inequality must be strict, and we conclude that if $\lambda \neq \lambda_0$ then $E[l_t(\gamma_0,\lambda)] > E[l_t(\gamma_0,\lambda_0)]$. ■

B.2 Lemmas for the proof of asymptotic normality

In the following we will 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 $d\{f_t(A_0,B_0), dA\}$ denotes the first-order differential of f_t in the direction dA and evaluated at (A_0,B_0).

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 4.1-4.5* $E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right|\right] < \infty$ *for all* $i,j=1,...,3d^2$.

Proof. Notice that

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

By (B.17), Minkowski's inequality, and Assumption 4.4,

$$E\left(\sup_{\theta\in\Theta}\|H_t(\gamma,\lambda)\|\right)^3 \leq K. \tag{B.20}$$

Moreover, using Minkowski's inequality repeatedly (see also Hafner and Preminger, 2009b, Proof of Lemma 3), and Assumption 4.4 one can show that

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

By (B.1), (B.12), Hölder's inequality, and (B.21),

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

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 $i=1,...,3d^2$ and, $j=1,...,3d^2$. ■

Lemma B.6 *Under Assumptions 4.1-4.5* $\sup_{\theta\in\Theta}\left|\frac{\partial^2 L_T(\gamma,\lambda)}{\partial\theta_i\partial\theta_j} - E\left(\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right)\right| \xrightarrow{a.s.} 0$ *for all* $i,j=1,...,3d^2$.

Proof. Notice that $\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}$ is a function of $(X_t, X_{t-1,\dots})$ and θ 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 4.1-4.5 J_0 stated in (A.10) is non-singular.*

Proof. We prove this lemma arguing in line with the proof of Theorem 3.2 in Francq and Zakoïan (2010), see also p.77-78 in Comte and Lieberman (2003). 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, with $\mathcal{F}_{t-1} := \sigma(X_{t-1}, X_{t-2,\dots})$

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

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

$$h_t = \left(h_{t1}\vdots\cdots\vdots h_{t2d^2}\right) \text{ and } k_t = \left(k_{t1}\vdots\cdots\vdots k_{t2d^2}\right).$$

Let $\mathcal{H}_t = \left(H_{0t}^{-1/2}\right)^{\otimes 2}$, and 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 $c' J_0 c = E[c' h_t' h_t c] = 0$. As $c' h_t' h_t c \geq 0$, then almost surely

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

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

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

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

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

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

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

Subtracting (B.25) from vec(H_{0t}) 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 vec(H_{0t}), which contradicts Assumption 4.3 that ensures that vec(H_{0t}) has a unique representation. Hence J_0 must be non-singular. ■

Lemma B.8 *Under Assumptions 4.1-4.5, as $T\to\infty$,*

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

where

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

with

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

and

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

Proof. 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

$$\begin{aligned}
d\left\{l_t\left(\gamma_0,\lambda_0\right),dA\right\}&=\text{tr}\left\{H_{0t}^{-1/2}\left[d\left\{H_t\left(\gamma_0,\lambda_0\right),dA\right\}\right]H_{0t}^{-1/2}\right\}\\
&\quad-\text{tr}\left\{H_{0t}^{-1/2}X_tX_t'H_{0t}^{-1/2}H_{0t}^{-1/2}\left[d\left\{H_t\left(\gamma_0,\lambda_0\right),dA\right\}\right]H_{0t}^{-1/2}\right\}\\
&=\text{tr}\left\{H_{0t}^{-1/2}\left[d\left\{H_t\left(\gamma_0,\lambda_0\right),dA\right\}\right]H_{0t}^{-1/2}\right\}\\
&\quad-\text{tr}\left\{Z_tZ_t'H_{0t}^{-1/2}\left[d\left\{H_t\left(\gamma_0,\lambda_0\right),dA\right\}\right]H_{0t}^{-1/2}\right\}\\
&=\text{vec}(I_d-Z_tZ_t')'\left(H_{0t}^{-1/2}\right)^{\otimes 2}\text{vec}\left[d\left\{H_t\left(\gamma_0,\lambda_0\right),dA\right\}\right].
\end{aligned}$$

Likewise,

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

Notice 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 $H_t(\gamma,\lambda)$ with respect to A is

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

implying directly

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

We note that

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

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

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

We conclude that

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

Identifying the Jacobian from the first-order differential, see e.g. Magnus and Neudecker (2007, p. 199), we find that the score of the log-likelihood function with respect to 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

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

with $\widetilde{M}_t(\gamma,\lambda)$ defined in (B.29).
Consider the sample covariance matrix on vec form:

$$\widehat{\gamma}_{VT} = \frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\operatorname{vec}\left(Z_tZ_t'-I_d\right) + \operatorname{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t}\right). \tag{B.31}$$

Moreover,

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

and collecting terms

$$\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}H_{0t}\right) = \left[I_{d^2} - B_0^{\otimes 2}\right]^{-1}\left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right]\gamma_0 + \left[I_{d^2} - B_0^{\otimes 2}\right]^{-1}A_0^{\otimes 2}\widehat{\gamma}_{VT} \tag{B.32}$$
$$+ \left[I_{d^2} - B_0^{\otimes 2}\right]^{-1}\left[A_0^{\otimes 2}\frac{1}{T}\text{vec}\left(X_0X_0' - X_TX_T'\right) + B_0^{\otimes 2}\frac{1}{T}\text{vec}\left(H_{00} - H_{0T}\right)\right].$$

Notice that $(I_{d^2} - B_0^{\otimes 2})$ is invertible since $\rho(B_0^{\otimes 2}) < 1$, as already mentioned in the proof of Lemma B.1. Next, inserting (B.31) in (B.32) and isolating $\widehat{\gamma}_{VT}$ yields

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

Hence

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

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

$$\widehat{\gamma}_{VT} - \gamma_0 = \left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right]^{-1}\left[I_{d^2} - B_0^{\otimes 2}\right]\frac{1}{T}\sum_{t=1}^{T}\left(H_{0t}^{1/2}\right)^{\otimes 2}\text{vec}\left(Z_tZ_t' - I_d\right) + o_P\left(T^{-1/2}\right). \tag{B.33}$$

We conclude that (B.26) holds. ■

Lemma B.9 *Under Assumptions 4.1-4.5*

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

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

Proof. By definition

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

Define

$$\varepsilon_t := \mathrm{vec}\left(Z_t Z_t' - I_d\right), \; \eta_t^M := \left[-\sum_{i=0}^{\infty} \left(B_0^{\otimes 2}\right)^i M_{t-1-i}(\gamma_0, \lambda_0)\right], \; \eta_t^{\widetilde{M}} := \left[-\sum_{i=0}^{\infty} \left(B_0^{\otimes 2}\right)^i \widetilde{M}_{t-1-i}(\gamma_0, \lambda_0)\right],$$

and observe that

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

where

$$\begin{aligned}
\mathcal{A}_t &= \left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right]^{-1} \left(H_{0t}^{1/2}\right)^{\otimes 2} \varepsilon_t \varepsilon_t' \left(H_{0t}^{1/2}\right)^{\otimes 2} \left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right]^{-1}, \\
\mathcal{B}_t &= \left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right]^{-1} \left(H_{0t}^{1/2}\right)^{\otimes 2} \varepsilon_t \varepsilon_t' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \eta_t^M, \\
\mathcal{C}_t &= \left[I_{d^2} - A_0^{\otimes 2} - B_0^{\otimes 2}\right]^{-1} \left(H_{0t}^{1/2}\right)^{\otimes 2} \varepsilon_t \varepsilon_t' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \eta_t^{\widetilde{M}}, \\
\mathcal{D}_t &= \eta_t^{M\prime} \left(H_{0t}^{-1/2}\right)^{\otimes 2} \varepsilon_t \varepsilon_t' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \eta_t^M, \\
\mathcal{E}_t &= \eta_t^{M\prime} \left(H_{0t}^{-1/2}\right)^{\otimes 2} \varepsilon_t \varepsilon_t' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \eta_t^{\widetilde{M}}, \\
\mathcal{G}_t &= \eta_t^{\widetilde{M}\prime} \left(H_{0t}^{-1/2}\right)^{\otimes 2} \varepsilon_t \varepsilon_t' \left(H_{0t}^{-1/2}\right)^{\otimes 2} \eta_t^{\widetilde{M}}.
\end{aligned}$$

Hence $Y_t(\gamma_0, \lambda_0)\mathrm{vec}(Z_t Z_t' - I_d)$ is square-integrable if $E\|\mathcal{A}_t\| \leq K$, $E\|\mathcal{B}_t\| \leq K$, $E\|\mathcal{C}_t\| \leq K$, $E\|\mathcal{D}_t\| \leq K$, $E\|\mathcal{E}_t\| \leq K$, and $E\|\mathcal{G}_t\| \leq K$.
Using Minkowski's inequality,

$$E\left\|\eta_t^M\right\|^3 \leq \left\{\sum_{i=1}^{\infty} \phi^i \left(K + KE\|X_t\|^6\right)^{1/3}\right\}^3 \leq K. \tag{B.34}$$

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

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

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 W_{t-1}. Moreover,

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

by (B.6) and (B.20). Moreover,

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

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

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

Note that

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

by (B.6) and (B.12). Hence by Hölder's inequality and (B.34)

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

By similar arguments $E\left\|\mathcal{C}_t\right\| \leq K$, $E\left\|\mathcal{D}_t\right\| \leq K$, $E\left\|\mathcal{E}_t\right\| \leq K$, and $E\left\|\mathcal{G}_t\right\| \leq K$. ■

Lemma B.10 *Under Assumptions 4.1-4.5, as $T \to \infty$*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t\left(\gamma_0,\lambda_0\right) vec\left(Z_t Z_t' - I_d\right) \xrightarrow{D} N\left(0,\Omega_0\right), \tag{B.35}$$

where

$$\Omega_0 = E\left\{Y_t\left(\gamma_0,\lambda_0\right) vec\left(Z_t Z_t' - I_d\right)\left[vec\left(Z_t Z_t' - I_d\right)\right]' Y_t\left(\gamma_0,\lambda_0\right)'\right\}, \tag{B.36}$$

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

Proof. Let $\mathcal{F}_t = \sigma\left(W_t, W_{t-1}, ...\right)$ with $W_t = \left[\mathrm{vech}\left(H_{0t}\right)', X_t'\right]'$, see Theorem 2.1. Since $Y_t(\theta_0)$ is $\mathcal{F}_{t-1}$-measurable and $\mathrm{vec}(Z_t Z_t' - I_d)$ and $\mathcal{F}_{t-1}$ are independent, $\left\{Y_t\left(\gamma_0,\lambda_0\right)\mathrm{vec}\left(Z_t Z_t' - I_d\right), \mathcal{F}_t\right\}$ 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, which establishes (B.35). ■

Lemma B.11 *Under Assumptions 4.1-4.5, as $T \to \infty$,*

$$\left| \sqrt{T} \left[\frac{\partial L_T(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial L_{T,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right| \overset{a.s.}{\to} 0, \tag{B.37}$$

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

$$\sup_{\theta \in \Theta} \left| \frac{\partial^2 L_T(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 L_{T,h}(\gamma, \lambda)}{\partial \theta_i \partial \theta_j} \right| \overset{a.s.}{\to} 0 \tag{B.38}$$

for $i, j = 1, ..., 3d^2$.'

Proof. We have that

$$\begin{aligned}
& \left| \left[\frac{\partial l_t(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial l_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right| \\
\leq & \left| \operatorname{tr} \left\{ H_{0t}^{-1} \frac{\partial H_t(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial H_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} H_{0t,h}^{-1} \right\} \right| \\
& + \left| \operatorname{tr} \left\{ H_{0t,h}^{-1} X_t X_t H_{0t,h}^{-1} \frac{\partial H_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial H_t(\gamma_0, \lambda_0)}{\partial \lambda_i} H_{0t}^{-1} X_t X_t H_{0t}^{-1} \right\} \right| \\
\leq & K \left\| \operatorname{vec} \left[\frac{\partial H_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial H_t(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right\| \\
& + K \|X_t\|^4 \left\| \operatorname{vec} \left[\frac{\partial H_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial H_t(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right\|.
\end{aligned}$$

If $i = 1, ..., d^2$ (corresponding to the elements of A) using (B.14) repeatedly

$$\begin{aligned}
& \left\| \operatorname{vec} \left[\frac{\partial H_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial H_t(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right\| \\
= & \left\| \frac{\partial}{\partial \lambda_i} \left\{ \left(B_0^{\otimes 2} \right)^t \operatorname{vec} [H_{00,h} - H_{00}] \right\} \right\| \\
\leq & K \phi^t \left\| \operatorname{vec} \left[\frac{\partial H_0(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right\| \\
\leq & K \phi^t,
\end{aligned}$$

since $\|\operatorname{vec}[H_{00,h} - H_{00}]\|$ and $\left\| \operatorname{vec} \left[\frac{\partial H_0(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right\|$ can be treated as constants, as they do not depend on t. Hence

$$\left| \left[\frac{\partial l_t(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial l_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right| \leq K \phi^t \|X_t\|^4,$$

and

$$\begin{aligned}
\left| \sqrt{T} \left[\frac{\partial L_T(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial L_{T,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right| & = \frac{1}{\sqrt{T}} \left| \sum_{t=1}^{T} \left[\frac{\partial l_t(\gamma_0, \lambda_0)}{\partial \lambda_i} - \frac{\partial l_{t,h}(\gamma_0, \lambda_0)}{\partial \lambda_i} \right] \right| \\
& \leq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} K \phi^t \|X_t\|^4.
\end{aligned}$$

Since

$$\lim_{T\to\infty}\sum_{t=1}^{T}E\left[K\phi^{t}\left\|X_{t}\right\|^{4}\right]\leq K,$$

it follows by Theorem 9.2 of Jacod and Protter (2003) that

$$\sum_{t=1}^{T}K\phi^{t}\left\|X_{t}\right\|^{4}=O_{a.s.}\left(1\right),$$

Similarly, if $i=d^2+1,...,2d^2$ (corresponding to the elements of B),

$$\left|\left[\frac{\partial l_{t}\left(\gamma_{0},\lambda_{0}\right)}{\partial\lambda_{i}}-\frac{\partial l_{t,h}\left(\gamma_{0},\lambda_{0}\right)}{\partial\lambda_{i}}\right]\right|\leq K\left(t\phi^{t-1}+\phi^{t}\right)\left(1+\left\|X_{t}\right\|^{4}\right),$$

and

$$\begin{aligned}\left|\sqrt{T}\left[\frac{\partial L_{T}\left(\gamma_{0},\lambda_{0}\right)}{\partial\lambda_{i}}-\frac{\partial L_{T,h}\left(\gamma_{0},\lambda_{0}\right)}{\partial\lambda_{i}}\right]\right| &= \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\left[\frac{\partial l_{t}\left(\gamma_{0},\lambda_{0}\right)}{\partial\lambda_{i}}-\frac{\partial l_{t,h}\left(\gamma_{0},\lambda_{0}\right)}{\partial\lambda_{i}}\right]\right| \\ &\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{T}K\left(t\phi^{t-1}+\phi^{t}\right)\left(1+\left\|X_{t}\right\|^{4}\right),\end{aligned}$$

and we conclude that (B.37) holds.
Next, we turn to (B.38) which we establish along the lines of the proof of Lemma 4 in Hafner and Preminger (2009a). From (B.19) and suppressing notation for parameter dependence, we have that

$$\begin{aligned}&\frac{\partial^{2}l_{t}\left(\gamma,\lambda\right)}{\partial\theta_{i}\partial\theta_{j}}-\frac{\partial^{2}l_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_{i}\partial\theta_{j}}\\ =\;&-\mathrm{tr}\left\{H_{t}^{-1}\frac{\partial H_{t}}{\partial\theta_{j}}H_{t}^{-1}\frac{\partial^{2}H_{t}}{\partial\theta_{i}\partial\theta_{j}}-\frac{\partial^{2}H_{t,h}}{\partial\theta_{i}\partial\theta_{j}}H_{t,h}^{-1}\frac{\partial H_{t,h}}{\partial\theta_{j}}H_{t,h}^{-1}\right\}+\mathrm{tr}\left\{H_{t}^{-1}\frac{\partial^{2}H_{t}}{\partial\theta_{i}\partial\theta_{j}}-\frac{\partial^{2}H_{t,h}}{\partial\theta_{i}\partial\theta_{j}}H_{t,h}^{-1}\right\}\\ &+2\mathrm{tr}\left\{H_{t}^{-1}X_{t}X_{t}'H_{t}^{-1}\frac{\partial H_{t}}{\partial\theta_{j}}H_{t}^{-1}\frac{\partial H_{t}}{\partial\theta_{i}}-\frac{\partial H_{t,h}}{\partial\theta_{i}}H_{t,h}^{-1}X_{t}X_{t}'H_{t,h}^{-1}\frac{\partial H_{t,h}}{\partial\theta_{j}}H_{t,h}^{-1}\right\}\\ &-\mathrm{tr}\left\{H_{t}^{-1}X_{t}X_{t}'H_{t}^{-1}\frac{\partial^{2}H_{t}}{\partial\theta_{i}\partial\theta_{j}}-\frac{\partial^{2}H_{t,h}}{\partial\theta_{i}\partial\theta_{j}}H_{t,h}^{-1}X_{t}X_{t}'H_{t,h}^{-1}\right\}.\end{aligned}\tag{B.39}$$

Observe that by (A.3)

$$E\sup_{\theta\in\Theta}\left\|\mathrm{vec}\left[\frac{\partial H_{t}\left(\gamma,\lambda\right)}{\partial\theta_{i}}-\frac{\partial H_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_{i}}\right]\right\|=E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_{i}}\left\{\left(B^{\otimes2}\right)^{t}\mathrm{vec}\left[H_{0,h}-H_{0}\left(\gamma,\lambda\right)\right]\right\}\right\|.$$

Hence by (B.21), Hölder's inequality and as $H_{0,h}$ is fixed, we get

$$E\sup_{\theta\in\Theta}\left\|\mathrm{vec}\left[\frac{\partial H_{t}\left(\gamma,\lambda\right)}{\partial\theta_{i}}-\frac{\partial H_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_{i}}\right]\right\| = O\left(\phi^{t}\right),\ i=1,...,2d^{2},\text{ and}\tag{B.40}$$

$$E\sup_{\theta\in\Theta}\left\|\mathrm{vec}\left[\frac{\partial H_{t}\left(\gamma,\lambda\right)}{\partial\theta_{i}}-\frac{\partial H_{t,h}\left(\gamma,\lambda\right)}{\partial\theta_{i}}\right]\right\| = O\left(t\phi^{t-1}\right),\ i=2d^{2}+1,...,3d^{2}.\tag{B.41}$$

Likewise, using (A.3) as above

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 H_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right]\right\| = E\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left\{\left(B^{\otimes 2}\right)^t \text{vec}\left[H_{0,h}-H_0(\gamma,\lambda)\right]\right\}\right\|,$$

and by (B.21) and Hölder's inequality

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 H_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right]\right\| = O\left(\phi^t\right),\ i,j=1,...,2d^2 \tag{B.42}$$

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 H_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right]\right\| = O\left(t(t-1)\phi^{t-2}\right),\ i,j=2d^2+1,...,3d^2 \tag{B.43}$$

$$E\sup_{\theta\in\Theta}\left\|\text{vec}\left[\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 H_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right]\right\| = O\left(t\phi^{t-1}\right),\ \text{otherwise.} \tag{B.44}$$

Observe that

$$\begin{aligned}\text{vec}\left[H_{t,h}(\gamma,\lambda)\right] &= \sum_{i=0}^{t-1}\left\{\left(B^{\otimes 2}\right)^i\left(I_{d^2}-A^{\otimes 2}-B^{\otimes 2}\right)\gamma\right\}+\sum_{i=0}^{t-1}\left(B^{\otimes 2}\right)^i A^{\otimes 2}\text{vec}\left(X_{t-1-i}X'_{t-1-i}\right)\\ &\quad+\left(B^{\otimes 2}\right)^t\text{vec}\left(H_{0,h}\right).\end{aligned} \tag{B.45}$$

By simple differentiation of (B.45) and using that $H_{0,h}$ is fixed, we conclude that

$$E\sup_{\theta\in\Theta}\left\|\frac{\partial H_{t,h}(\gamma,\lambda)}{\partial\theta_i}\right\|\le K. \tag{B.46}$$

Considering the last term in (B.39) and using Hölder's inequality, Assumption 4.4, and (B.12),

$$\begin{aligned}&E\sup_{\theta\in\Theta}\left|\text{tr}\left\{H_t^{-1}X_tX_t'H_t^{-1}\frac{\partial^2 H_t}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 H_{t,h}}{\partial\theta_i\partial\theta_j}H_{t,h}^{-1}X_tX_t'H_{t,h}^{-1}\right\}\right|^{1/4} \\ \le\ & K\left(E\sup_{\theta\in\Theta}\left\|\frac{\partial^2 H_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 H_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right\|\right)^{1/4}\left(E\left\|X_t\right\|^4\right)^{1/4},\end{aligned} \tag{B.47}$$

which is either $O\left(\phi^t\right)$, $O\left(t(t-1)\phi^{t-2}\right)$, or $O\left(t\phi^{t-1}\right)$ in light of (B.42)-(B.44). By similar arguments together with (B.40)-(B.41) and (B.46), it follows by the C_r inequality (White, 2001, Proposition 3.8) that $E\sup_{\theta\in\Theta}\left|\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 l_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right|^{1/4}$ is $O\left(t^2\phi^t\right)$. Next, by the generalized Chebyshev inequality for any $\varepsilon>0$

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

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

$$\sup_{\theta\in\Theta}\left|\frac{\partial^2 l_t(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2 l_{t,h}(\gamma,\lambda)}{\partial\theta_i\partial\theta_j}\right|\overset{a.s.}{\to}0. \tag{B.48}$$

By (B.48) and Cesàro's mean theorem we conclude that (B.38) holds. ■

C Drift criteria for the BEKK-ARCH(1) model

In order to find conditions for which the BEKK-ARCH(1) model with Gaussian noise is geometrically ergodic with high-order moments we will make use of the following lemmas.

Lemma C.1 *(Bec and Rahbek, 2004, Proof of Theorem 1)*
Let $(X_t)_{t=0,1,\ldots}$ be a time-homogeneous Markov chain on the state space $\mathbb{R}^d$ endowed with the Borel σ-algebra, $\mathbb{B}^d$. Assume that for all sets $A \in \mathbb{B}^d$ and for some integer $m \geq 1$, that the m-step transition density with respect to the Lebesgue measure $f(\cdot|\cdot)$ as defined by

$$P(X_t \in A | X_{t-m} = x) = \int_A f(y|x)\, dy$$

is strictly positive and bounded on compact sets. Let $v : \mathbb{R}^d \mapsto [1, \infty)$ be some drift function. Assume there exists an integer $k \geq 1$, a compact set $B \subset \mathbb{R}^d$ and constants $0 < \gamma < 1$, $g > 0$ such that

$$E[v(X_{t+k}) | X_t = x] \leq \gamma v(x)$$

for $x \in B^c$, while $E[v(X_{t+k}) | X_t = x]$ is bounded by g on B.
Then X_t is geometrically ergodic and X_0 can be given an initial distribution such that X_t is stationary. Moreover, all moments bounded by $v(\cdot)$ exist.

Lemma C.2 *(Corollary 1 (i)-(iii) in Ghazal (1996))*
Let $Q = x'\Omega x$ be a quadratic form where Ω is a $d \times d$ symmetric non-stochastic matrix and x is IIDN$(0, I_d)$. Then

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

We are now able to prove the following theorem,

Theorem C.1 *Let $\{X_t\}_{t=1,\ldots,T}$ follow a BEKK-ARCH(1) process as in (2.2) with $B = 0$ and Z_t IIDN$(0, I_d)$. 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, Hansen, and Dennis (2002). Now consider *(iii)*: Clearly X_t is a Markov chain and, conditional on X_{t-1}, X_t is Gaussian with mean zero and covariance H_t. So indeed the one-step ($m = 1$) transition density of X_t conditional on X_{t-1} is continuous in both X_t and X_{t-1} and positive. Hence we can apply Lemma C.1. Define the drift function

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

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

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

where the fourth equality follows by Lemma C.2. Ignoring terms of lower order than $\|x\|^6$, the right-hand side equals $15\left(x'A'Ax\right)^3$.
Let $L\left(\mathbb{R}^d\right)$ denote the space of linear mappings from $\mathbb{R}^d \mapsto \mathbb{R}^d$. For linear mappings $\phi : L\left(\mathbb{R}^d\right) \mapsto L\left(\mathbb{R}^d\right)$ we use the operator norm defined by

$$\|\phi\|_{op} := \sup_{\|x\|\neq 0} \frac{\|\phi\left(x\right)\|}{\|x\|}.$$

It holds 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 $L\left(\mathbb{R}^d\right)$, and define the mapping $\phi = (A\otimes A)$ from $L\left(\mathbb{R}^d\right) \mapsto L\left(\mathbb{R}^d\right)$ by

$$\phi\left(X\right) := \left(A\otimes A\right)\left(X\right) = AXA'.$$

We notice that $\phi^k\left(X\right) = A^kXA^{k\prime}$ and $\Omega_x = C_0 + \ \phi\left(xx'\right)$.
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 adjusted 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 $E\|X_t\|^n < \infty$, $n = 2k$, $k\in\mathbb{N}$, 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).*

References

Avarucci, M., E. Beutner, and P. Zaffaroni (2012): "On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models," Working Paper DSS-E3 WP 2012/1, University of Rome La Sapienza, DSS Empirical Economics and Econometrics, Forthcoming in Econometric Theory.

Bao, Y., and A. Ullah (2010): "Expectation of quadratic forms in normal and nonnormal variables with applications," *Journal of Statistical Planning and Inference*, 140(5), 1193 – 1205.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: a survey," *Journal of Applied Econometrics*, 21(1), 79–109.

Bec, F., and A. Rahbek (2004): "Vector equilibrium correction models with non-linear discontinuous adjustments," *Econometrics Journal*, 7(2), 628–651.

Bollerslev, T. (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69(3), 542–47.

Boussama, F., F. Fuchs, and R. Stelzer (2011): "Stationarity and geometric ergodicity of BEKK multivariate GARCH models," *Stochastic Processes and their Applications*, 121(10), 2331–2360.

Brown, B. (1971): "Martingale central limit theorems," *The Annals of Mathematical Statistics*, pp. 59–66.

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

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

Engle, R. F., and K. F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11(01), 122–150.

Francq, C., L. Horváth, and J.-M. Zakoïan (2011): "Merits and Drawbacks of Variance Targeting in GARCH Models," *Journal of Financial Econometrics*, 9(4), 619–656.

Francq, C., and J.-M. Zakoïan (2010): "QML estimation of a class of multivariate GARCH models without moment conditions on the observed process," MPRA Paper 20779, University Library of Munich, Germany.

——— (2012): "Strict Stationarity Testing and Estimation of Explosive and Stationary Generalized Autoregressive Conditional Heteroscedasticity Models," *Econometrica*, 80(2), 821–861.

Ghazal, G. (1996): "Recurrence formula for expectations of products of quadratic forms," *Statistics and Probability Letters*, 27(2), 101 – 109.

Hafner, C. M., and A. Preminger (2009a): "Asymptotic Theory for a Factor GARCH Model," *Econometric Theory*, 25(02), 336–363.

——— (2009b): "On asymptotic theory for multivariate GARCH models," *Journal of Multivariate Analysis*, 100(9), 2044 – 2054.

Jacod, J., and P. Protter (2003): *Probability Essentials*, Universitext. Springer Berlin Heidelberg, 2nd edn.

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

Laurent, S., J. V. K. Rombouts, and F. Violante (2012): "On the forecasting accuracy of multivariate GARCH models," *Journal of Applied Econometrics*, 27(6), 934–955.

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

MAGNUS, J. R., AND H. NEUDECKER (2007): *Matrix differential calculus with applications in statistics and econometrics.* John Wiley & Sons, 3rd edn.

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, vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier.

NOURELDIN, D., N. SHEPHARD, AND K. SHEPPARD (2012): "Multivariate Rotated ARCH Models," Nuffield College Economics Working Paper 2012-W01, Economics Group, Nuffield College, University of Oxford.

RAHBEK, A. (2004): "Stochastic properties of multivariate time series equations with emphasis on ARCH," *System Identification 2003*, pp. 227–232.

RAHBEK, A., E. HANSEN, AND J. DENNIS (2002): "ARCH Innovations and their Impact on Cointegration Rank Testing," Discussion Paper 22, Department of Statistics and Operations Research, University of Copenhagen.

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, pp. 201–229. Springer Berlin Heidelberg.

WHITE, H. (1994): *Estimation, Inference and Specification Analysis.* Cambridge University Press, New York, NY.

——— (2001): *Asymptotic Theory for Econometricians (Revised Edition).* Academic Press.