Discussion Papers
Department of Economics
University of Copenhagen

No. 15-10

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

Rasmus Søndergaard Pedersen

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

ISSN: 1601-2461 (E)

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

Rasmus Søndergaard Pedersen†

September 4, 2015

**Abstract**

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

*Keywords:* ECCC-GARCH, QML, boundary, spillovers.

*JEL Classification:* C32, C51, C58.

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 I thank Donald W. K. Andrews for providing his unpublished manuscript "Estimation When a Parameter is on a Boundary: Part II" and Tomoaki Nakatani for providing the data from the paper Nakatani and Teräsvirta (2009).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

†Department of Economics, University of Copenhagen, Oester Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark. Tel.: +45 35 32 30 74. Fax: +45 35 32 35 32. E-mail: rsp@econ.ku.dk.

# 1 Introduction

Testing for volatility spillovers between time series has become an important tool in empirical finance. Following the simple arguments of Ross (1989) that the (conditional) variance of asset price changes is directly related to the rate of information flow, volatility spillovers may be viewed as a way of measuring information transmissions in and between markets and thereby their connectedness (Conrad and Weber, 2013). Typically, volatility spillovers are defined in relation to multivariate conditional volatility models, such as multivariate GARCH, for price changes. As an example, Conrad et al. (1991) applied bivariate GARCH models to conclude that volatility surprises to large market value firms are important to the future dynamics of the returns of smaller firms (but not conversely). Another example can be found in Bali and Hovakimian (2009) who applied a similar technique to conclude that there exist spillovers from option to equity markets. For other applications of multivariate GARCH models for assessing spillovers we refer to Conrad and Weber (2013) and the references therein. A multivariate GARCH model well suited for quantifying spillovers is the extended constant conditional correlation (ECCC-) GARCH model of Jeantheau (1998), considered in this paper. In the ECCC-GARCH model the matrices governing the ARCH and GARCH dynamics - respectively, the matrices $A$ and $B$ introduced in the following section - are allowed to be nondiagonal, and with the off-diagonal elements directly related to the volatility spillovers. Specifically, testing for no volatility spillovers relies on testing for whether the off-diagonal elements of the matrices are equal to zero.

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

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

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

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

## 2 The ECCC-GARCH model and its properties

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

## 2.1 The model

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

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

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

*Remark* 2.1. When the matrices $A$ and $B$ are restricted to be diagonal, the ECCC-GARCH model simplifies to the CCC-GARCH model proposed by Bollerslev (1990).

## 2.2 Properties of the ECCC-GARCH process

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

### 2.3 Volatility spillovers and Granger noncausality

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

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

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

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

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

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

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

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

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

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

It holds that $\mathbb{E}[X_{t,1}|\mathcal{F}_{t-1}^X] = 0$ a.s., so that $\mathbb{E}\{(X_{t,1} - \mathbb{E}[X_{t,1}|\mathcal{F}_{t-1}^X])^2|\mathcal{F}_{t-1}^X\} = h_{t,1}$ a.s. Hence, in light of the above definition, $X_{t,2}$ does not second-order Granger cause $X_{t,1}$ if

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

## 3 Estimation and large-sample properties of the QMLE

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

Given a realization $(X_t : t = 0, 1, ..., T)$ of the ECCC-GARCH model, the QMLE, $\hat{\theta}_T$, of $\theta$ is defined as

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

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

$$\begin{aligned}
\hat{L}_T(\theta) &:= \frac{1}{T}\sum_{t=1}^{T} \hat{l}_t(\theta), && (3.1)\\
\hat{l}_t(\theta) &:= \log\left\{\det\left[\hat{H}_t(\theta)\right]\right\} + X_t'\hat{H}_t^{-1}(\theta) X_t, && (3.2)\\
\hat{H}_t(\theta) &:= \hat{D}_t(\theta)R(\theta)\hat{D}_t(\theta), &&\\
\hat{D}_t^2(\theta) &:= \operatorname{diag}\left[\hat{h}_t(\theta)\right], && (3.3)\\
\hat{h}_t(\theta) &:= \kappa + AX_{t-1}^{\odot 2} + B\hat{h}_{t-1}(\theta), && (3.4)
\end{aligned}$$

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

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

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

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

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

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

$$\begin{aligned} L_T(\theta) &:= \frac{1}{T}\sum_{t=1}^{T} l_t(\theta) \\ l_t(\theta) &:= \log\left[\det\left(H_t(\theta)\right)\right] + X_t' H_t^{-1}(\theta) X_t \qquad (3.5) \\ H_t(\theta) &:= D_t(\theta)R(\theta)D_t(\theta) \\ D_t^2(\theta) &:= \operatorname{diag}\left(h_t(\theta)\right) \qquad (3.6) \\ h_t(\theta) &:= \kappa + AX_{t-1}^{\odot 2} + Bh_{t-1}(\theta). \qquad (3.7) \end{aligned}$$

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

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

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

As mentioned, we are interested in the case where some of the elements of $A_0$ and $B_0$ are equal to zero, implying that $\theta_0$ is not an interior point of $\Theta$. Let $\beta$ denote the $(s_1 \times 1)$ vector containing the $s_1 \geq 0$ elements of $A$ and $B$ that take value zero under the null, i.e. with true parameter value equal to zero, and let $\delta$ denote the $(s_2 \times 1)$ vector of the remaining $s_2 := (s_0 - s_1)$ parameters of $\theta$. Without loss of generality we consider throughout the remainder of the paper a reparametrized version of the ECCC-GARCH model such that

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

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

First, we make the following assumption about $\theta_0$ and $\Theta$.

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

*Remark* 3.3. Assumption 5 is essentially a special case of Assumption $2^{2*}$ in Andrews (1999, 2001) and has several purposes. First, it prevents the true parameter value $\delta_0$ from reaching the bounds of $\Theta$, which keeps things as simple as possible, as our main interest is to consider hypotheses where elements of $\beta$ are equal to zero (i.e. take value at the lower bound of $\Theta$). Second, this assumption allows us to make a Taylor-type expansion based on left/right partial derivatives of the log-likelihood function around $\theta_0$, see Andrews (1999, Appendix A) for details. Moreover, the assumption is important for approximating the quantity $\sqrt{T}(\hat{\theta}_T - \theta_0)$, see specifically the proof of Theorem 3.1 in the appendix. Although the assumption imposes additional structure on the parameter space it is compatible with the parameter restrictions given in Assumption 3. As in Francq and Zakoïan (2007), let $\theta_0(\epsilon)$ be defined as the vector obtained by replacing all zero elements of $\theta_0$ by $\epsilon > 0$. For some sufficiently small $\epsilon$, $\theta_0(\epsilon)$ belongs to the interior of $\Theta$. Consider the case where $B_0$ is diagonal. Provided that Assumptions 1 and 3 hold, $\rho(B_0) < 1$. For a real $m \times m$ matrix with nonnegative entries, it holds that $C = [C_{ij}] \geq 0$, $\rho(C) \leq \min\left\{\max_{i=1,\dots,m} \sum_{j=1}^m C_{ij}, \max_{j=1,\dots,m} \sum_{i=1}^m C_{ij}\right\}$. Hence for a sufficiently small $\epsilon > 0$, $\rho(B_{0\epsilon}) < 1$, where $B_{0\epsilon}$ is $B$ evaluated at $\theta_0(\epsilon)$.
Another example is the bivariate case where

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

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

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

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

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

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

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

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

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

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

for all $i_1 = 1,..,d$ and all $i = 1,...,s_0$, and a similar property with $\partial h_{t,i_1}(\theta)/\partial\theta_i$ replaced with $\partial^2 h_{t,i_1}(\theta)/\partial\theta_i\partial\theta_j$ and $\partial^3 h_{t,i_1}(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$. Consider, for simplicity, the case with $B = 0_{2\times 2}$ on $\Theta$, i.e. with no GARCH effects. Then

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

and hence

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

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

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

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

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

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

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

**Theorem 3.1.** *Under Assumptions 1-6,*

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

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

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

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

$\Lambda$ with respect to the metric induced by the inner product $\langle \cdot, \cdot \rangle_J$, where we recall that for $x, y \in \mathbb{R}^{s_0}$, $\langle x, y \rangle_J = x' J y$. Since $\Lambda$ is convex according to Assumption 5, it holds that $\lambda^\Lambda$ is unique. In the case where $\theta_0$ is not a boundary point, $s_1 = 0$, such that $\Lambda = \mathbb{R}^{s_0}$ and the limiting distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is $Z$, as mentioned in Remark 3.2. Notice that the matrices $J$ and $\Sigma$ are stated in terms of left/right-derivatives, as discussed in Andrews (1999, Appendix A). Moreover, Andrews (1999, pp.1367-1370) provides closed-form expressions for $\lambda^\Lambda$, and gives an outline of how to make draws of the distribution of $\lambda^\Lambda$ based on numerical methods. The next section is devoted to testing hypotheses about the parameters in $A$ and $B$.

# 4 Testing

In this section we introduce Lagrange multiplier, Wald, and likelihood ratio statistics suitable for testing hypotheses about the matrices $A$ and $B$. In particular, these tests allow us to test for volatility and second-order Granger noncausality, as discussed in Subsection 2.3. Subsection 4.1 states the test statistics and their limiting distributions. In Subsection 4.2 we provide an algorithm for determining critical values for the proposed tests.

## 4.1 Test statistics

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

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

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

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

We test $\mathcal{H}_0$ against the alternative $\beta_{1,0} \neq 0_{\tilde{s}_1 \times 1}$ and with the maintained hypothesis that $\theta_0 \in \Theta$. Notice that under $\mathcal{H}_0$ it might be that some of the remaining parameters of $A$ and $B$ are equal to zero, which is the case when $\tilde{s}_2 = s_1 - \tilde{s}_1 > 0$, and we may consider $\beta_2$ as nuisance parameters attaining the zero bound of $\Theta$ under $\mathcal{H}_0$.

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

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

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

(QLR) statistic,

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

Next, let

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

Moreover, with $s_0$ the dimension of the parameter vector $\theta$, $s_1$ the dimension $\beta$ given in (3.8), and $\tilde{s}_1$ the dimension of the vector $\beta_1$ defined in (4.1), let

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

The second statistic is the Wald statistic,

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

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

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

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

In order to derive the limiting distribution of these statistics, we assume, similar to Assumption 3, that $\theta_0$ and $\Theta_0$ satisfy the following conditions.

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

Similar to $\lambda^\Lambda$ defined in Theorem 3.1, we consider $\lambda^{\Lambda_0}$ as the projection of random vector $Z$ with distribution $N(0, J^{-1}\Sigma J^{-1})$ onto $\Lambda_0$, i.e.

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

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

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

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

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

*(4.5).*
*Moreover,*

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

*Suppose in addition that $\Sigma$, defined in (3.13), is positive definite. Then*

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

*where $\chi^2_{\tilde{s}_1}$ is a chi-squared random variable with $\tilde{s}_1$ degrees of freedom, with $\tilde{s}_1$ the dimension of $\beta_1$.*

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

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

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

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

This implies that the limiting distributions of $W_T$ and $QLR_T$ do not depend on $\Lambda_2$ and thereby not on whether the nuisance parameters take zero value. In particular we have that

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

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

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

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

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

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

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

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

## 4.2 Calculating critical values

Following Andrews (1999, pp.1367-1370), we can obtain draws from the limiting distribution of the $W_T$ and $QLR_T$ statistics according to the following algorithm.[1]

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

1. *Draw $\varepsilon^\star$ randomly from $N(0, I_{s_1})$ and compute $Z_\beta^\star = [K\bar{J}_T^{-1}\bar{\Sigma}_T\bar{J}_T^{-1}K']^{1/2}\varepsilon^\star$.*
2. *Find $\tilde{\lambda}_\beta^\star$ that minimizes $\|Z_\beta^\star - \lambda_\beta\|^2_{(K\bar{J}_T^{-1}K')^{-1}} = (Z_\beta^\star - \lambda_\beta)'(K\bar{J}_T^{-1}K')^{-1}(Z_\beta^\star - \lambda_\beta)$ over $\lambda_\beta \in \Lambda_\beta = \mathbb{R}_+^{s_1}$, and compute $\|\tilde{\lambda}_\beta^\star\|^2_{(K\bar{J}_T^{-1}K')^{-1}}$.*
3. *Repeat steps 1.-2. $N$ times (with $N$ very large), and let $\{x_i : i = 1, ..., N\}$ denote the sequence of the $N$ independent draws of $\|\tilde{\lambda}_\beta^\star\|^2_{(K\bar{J}_T^{-1}K')^{-1}}$. Then $c$ is given by the $(1-\alpha)$ percentile of $\{x_i : i = 1, ..., N\}$.*

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

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

# 5 Simulations

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

## 5.1 Size simulations

We consider the size properties of the proposed test statistics, including the $LM_{ECCC}$ mentioned in Remark 4.5, for the bivariate ECCC-GARCH model. Specifically, we consider tests where the matrices $A$ and $B$ are diagonal under the null. In order to keep things simple we consider cases where no nuisance parameters in $A$ and $B$ take zero value. We consider the data-generating processes (DGPs) stated in Table 1, where DGP 1-3 correspond to DGP 1,2, and 4 in Nakatani and Teräsvirta (2009), respectively. Recall from Theorem 3.1 that we imposed finite sixth-order moments of $X_t$ (Assumption 6) in order to derive the limiting distribution of the QMLE. For all the DGPs we impose, for simplicity, that the innovation $\eta_t$ is Gaussian. This condition implies that $\eta_t$ has a strictly positive density on $\mathbb{R}^d$ with $\mathbb{E}[\|\eta_t\|^6] < \infty$, and hence from Lemmas B.7 and B.8, $\mathbb{E}[\|X_t\|^6] < \infty$ if and only if

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

Using Monte Carlo integration we have computed the value of $\Psi_6$ for each DGP, as also stated in Table 1. Whereas DGP 3-5 satisfy condition (5.1), DGP 1-2 do not. Although our theoretical results are not expected to hold for DGP 1 and 2, we have included the simulations results in order to compare with the results for the DGPs that do satisfy the moment condition. Moreover, for all the DGPs for the empirical size and power simulations it holds that the conditions in Jeantheau (1998, Definition 3.1.3 and Assumptions B1-B2) are satisfied, which by Jeantheau (1998, Proposition 3.4) implies that the identification condition in Assumption 4 holds. These above restrictions on the DGPs imply together that Corollary 4.1 holds for the processes (up to the sixth-order moment condition of $X_t$ in the case of DGP 1-2).

Table 1: DGPs for size simulations

| | DGP 1 | DGP 2 | DGP 3 | DGP 4 | DGP 5 |
|---|---|---|---|---|---|
| $A_0$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}$ | $\begin{bmatrix} 0.04 & 0 \\ 0 & 0.05 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}$ | $\begin{bmatrix} 0.07 & 0 \\ 0 & 0.08 \end{bmatrix}$ |
| $B_0$ | $\begin{bmatrix} 0.8 & 0 \\ 0 & 0.7 \end{bmatrix}$ | $\begin{bmatrix} 0.95 & 0 \\ 0 & 0.9 \end{bmatrix}$ | $\begin{bmatrix} 0.45 & 0 \\ 0 & 0.6 \end{bmatrix}$ | $\begin{bmatrix} 0.70 & 0 \\ 0 & 0.75 \end{bmatrix}$ | $\begin{bmatrix} 0.80 & 0 \\ 0 & 0.85 \end{bmatrix}$ |
| $r_0$ | 0.3 | 0.9 | 0.9 | 0.9 | 0.9 |
| $\Psi_6$ | 1.223 | 1.337 | 0.387 | 0.944 | 0.953 |

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

Table 2 contains the actual rejection frequencies of our proposed tests based on the 5% nominal level and on empirically relevant sample sizes of 1,000, 5,000, and 10,000 observations. All simulations are based on 2,000 replications with a burn-in period of 1,000 observations. The critical value of the $QLR_T$ and $W_T$ tests are carried out according to Algorithm 1 and Remark 4.6. For each replication the critical value is based on 100,000 draws from $\|\lambda_\beta^\Lambda\|^2_{(KJ^{-1}K')^{-1}}$. The critical values for the $LM_T$ and $LM_{ECCC}$ are based on a $\chi^2_4$-distribution, in line with Theorem 4.1 and Proposition C.1. We refer to Appendix D for additional technical details about the simulations.

Table 2: Size simulations

| | $T$ | $LM_T$ | $W_T$ | $QLR_T$ | $LM_{ECCC}$ |
|---|---|---|---|---|---|
| **DGP 1** | 1,000 | 0.0277 | 0.00877 | 0.0411 | 0.1886 |
| | 5,000 | 0.0552 | 0.0326 | 0.0577 | 0.1068 |
| | 10,000 | 0.0436 | 0.0276 | 0.0461 | 0.0657 |
| **DGP 2** | 1,000 | 0.0194 | 0.0187 | 0.0406 | 0.2710 |
| | 5,000 | 0.0460 | 0.0490 | 0.0555 | 0.1090 |
| | 10,000 | 0.0505 | 0.0545 | 0.0610 | 0.0790 |
| **DGP 3** | 1,000 | 0.0277 | 0.0164 | 0.0507 | 0.1638 |
| | 5,000 | 0.0477 | 0.0271 | 0.0432 | 0.0974 |
| | 10,000 | 0.0455 | 0.0345 | 0.0460 | 0.0760 |
| **DGP 4** | 1,000 | 0.0353 | 0.0243 | 0.0487 | 0.1764 |
| | 5,000 | 0.0551 | 0.0406 | 0.0561 | 0.0966 |
| | 10,000 | 0.0455 | 0.0360 | 0.0495 | 0.0680 |
| **DGP 5** | 1,000 | 0.0137 | 0.0180 | 0.0416 | 0.2270 |
| | 5,000 | 0.0445 | 0.0310 | 0.0506 | 0.1036 |
| | 10,000 | 0.0390 | 0.0365 | 0.0445 | 0.0685 |

Actual rejection frequencies based on the 5% nominal level.

From Table 2 we notice that $LM_T$ seems to be slightly under-sized for a sample size of 1,000 observations, whereas the test seems to have very reasonable size properties for larger sample sizes. The $LM_{ECCC}$ test seems to be over-sized for sample sizes of 1,000 and 5,000 observations, but only slightly over-sized for 10,000 observations.[2] Moreover, the Wald test appears to be slightly conservative for most of the DGPs and in particular for sample sizes of 1,000 observations. The quasi-likelihood ratio test has very reasonable

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

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

## 5.2 Power simulations

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

Table 3: DGPs for power simulations

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

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

Table 4 states the rejection frequencies of the tests when the null is incorrect according to the DGPs given in Table 4. The simulations are based on 2,000 replications, a burn-in period of 1,000 observations, and the same seed values as the size simulations. The reported powers are size corrected in the sense that the critical value for the tests (at the 5% nominal level) is chosen as the 95 percentile of the simulated test values from the size simulations.

From Table 4 we see that the power of the tests is low whenever the off-diagonal elements of $A$ and $B$ are all close to zero. In particular, even for a sample size of 10,000 observations the power is not impressive for any of the test statistics for the DGPs 5.1, 5.5 and 5.7. For all other DGPs the test statistics seem to have great power as $T$ increases. Moreover, the proposed Wald and likelihood ratio tests have better power properties than the other tests for all choices of DGP and for all sample lengths.

Table 4: Empirical power

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

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

# 6 Empirical illustration

In this section we provide an empirical application of the proposed tests for volatility spillovers. We apply the same data set as in Nakatani and Teräsvirta (2009) and investigate the volatility spillovers between a pair of foreign exchange rates. The exchange rates are daily noon buying rates of the Japanese yen (JPY) and the Swiss franc (CHF) against the U.S. dollar (USD). The series go from 2 January 1975 to 2 December 2005, with a total of 7,766 observations in each series. Descriptive statistics of the data series are contained in Nakatani and Teräsvirta (2009, Tables 7 and 8).[3]

Table 5: Estimation results

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

Point estimates of parameters in the restricted ECCC-GARCH model (CCC) and in the unrestricted ECCC-GARCH model (ECCC). The estimates of the elements of $\kappa$ are multiplied by 1,000. The p-values of the $LM_T$, $W_T$, $QLR_T$, and $LM_{ECCC}$ test for diagonality of $A$ and $B$ are reported in parentheses. The p-values for $W_T$ and $QLR_T$ are obtained according to Algorithm 1 and Remark 4.6 based on 1,000,000 draws. The p-values for $LM_T$ and $LM_{ECCC}$ are based on a $\chi^2_4$-distribution.

We fit a bivariate ECCC-GARCH model to the return series and test whether the matrices $A$ and $B$ are diagonal. The tests are based on the assumption that the diagonal elements of $A$ and $B$ are strictly positive under the null, such that no nuisance parameters

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

take zero value. This enables us to determine the critical values of the tests according to Algorithm 1 and Remark 4.6. For each individual series of the standardized residuals, based on a Jarque-Bera test, we rejected the null of normality, suggesting that the $LM_{ECCC}$ test based on a $\chi_4^2$ limiting distribution, as performed in Nakatani and Teräsvirta (2009), may not be appropriate for testing for no spillovers in the return series, as mentioned in Remark 4.5. Table 5 contains the estimation results. First, we notice that the point estimates of the off-diagonal elements of $A$ and $B$ are fairly small. Second, based on the $LM_T$ statistic we fail to reject the null of no spillovers, whereas the null is rejected based on the $LM_{ECCC}$ test with the p-value based on a $\chi_4^2$-distribution. The latter is in line with the findings in Nakatani and Teräsvirta (2009), but, as the standardized residuals, as mentioned, did not appear to be normally distributed, the validity of the $LM_{ECCC}$ test is dubious. Based on the $QLR_T$ and $W_T$ tests, we reject the null of no spillovers. In light of the very reasonable size properties and superior power properties of these tests compared to $LM_T$, we find evidence for volatility spillovers between the JPY/USD and CHF/USD rates, in line with the findings in Nakatani and Teräsvirta (2009).

## 7 Concluding remarks and future research directions

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

A simulation study showed that, in particular, the QLR test has very reasonable empirical size properties. Moreover, simulations showed that the Wald and QLR tests have superior empirical power properties compared to the LM test.

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

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

# A Proofs of theorems

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

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

where $J_{\beta\beta}$ is $(s_1 \times s_1)$ and $G_\beta$ is $(s_1 \times 1)$ and so forth.

***Proof of Theorem 3.1.*** The asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is derived along the lines of Andrews (1999, Proof of Theorem 3) and Francq and Zakoïan (2007, Proof of Theorem 2). Initially, notice that $\hat{\theta}_T$ is strongly consistent for $\theta_0$, as mentioned in Section 3. Moreover, $\Theta - \theta_0$ is locally equal to a union of orthants, $\Lambda$, (Assumption 3) and $\hat{L}_T(\theta)$ has continuous left/right partial derivative of order 2 on $\Theta$. Due to Andrews (1999, Theorem 6) we are able to make a second-order Taylor-type expansion of $\hat{L}_T(\theta)$ around $\theta_0$ such that for any point $\theta \in \Theta$ there exists a point $\theta^\star$ between $\theta$ and $\theta_0$

$$\hat{L}_T(\theta) = \hat{L}_T(\theta_0) + \frac{\partial L_T(\theta_0)}{\partial \theta'}(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'}(\theta - \theta_0) + R_T(\theta) + R_T^\star(\theta),$$

where

$$R_T(\theta) = \left( \frac{\partial \hat{L}_T(\theta_0)}{\partial \theta'} - \frac{\partial L_T(\theta_0)}{\partial \theta'} \right)(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \left[ \frac{\partial^2 \hat{L}_T(\theta_0)}{\partial \theta \partial \theta'} - \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'} \right](\theta - \theta_0), \tag{A.2}$$

and

$$R_T^\star(\theta) = \frac{1}{2}(\theta - \theta_0)' \left[ \frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \theta \partial \theta'} - \frac{\partial^2 \hat{L}_T(\theta_0)}{\partial \theta \partial \theta'} \right](\theta - \theta_0). \tag{A.3}$$

Moreover, with

$$J_T := \frac{\partial^2 L_T(\theta_0)}{\partial \theta \partial \theta'}$$

and

$$Z_T := -J_T^{-1}\sqrt{T}\frac{\partial L_T(\theta_0)}{\partial \theta}, \tag{A.4}$$

we have, by definition, that

$$\hat{L}_T(\theta) = \hat{L}_T(\theta_0) - \frac{1}{2T}\|Z_T\|_{J_T}^2 + \frac{1}{2T}\|Z_T - \sqrt{T}(\theta - \theta_0)\|_{J_T}^2 + R_T(\theta) + R_T^\star(\theta). \tag{A.5}$$

Notice that in the definition of $Z_T$ in (A.4) we have used that $J_T = J + o(1)$ almost surely with $J$ nonsingular, as proved below. So, technically, $Z_T$ might only exist almost surely for $T$ sufficiently large. It suffices to establish the following points:

1. $\sqrt{T}\partial L_T(\theta_0)/\partial\theta \xrightarrow{w} G$ with $\mathcal{L}(G) = N(0,\Sigma)$ and $J_T = J + o_p(1)$, where the matrices $J \in \mathbb{S}^{s_0}_{++}$ and $\Sigma$ are given by (3.13).

2. $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$.

3. For any $\bar{\theta}_T \in \Theta$ such that $\sqrt{T}(\bar{\theta}_T - \theta_0) = O_p(1)$, $R^\star_T(\bar{\theta}_T) = o_p(T^{-1})$ and $R_T(\bar{\theta}_T) = o_p(T^{-1})$.

4. $\sqrt{T}(\hat{\theta}_T - \theta_0) = \hat{\lambda}_T + o_p(1)$, where $\hat{\lambda}_T \in \mathrm{cl}(\Lambda)$ satisfies $\|Z_T - \hat{\lambda}_T\|^2_{J_T} = \inf_{\lambda\in\Lambda}\|Z_T - \lambda\|^2_{J_T}$ with $\Lambda$ defined in Assumption 5

5. $\hat{\lambda}_T \xrightarrow{w} \lambda^\Lambda$, where $\lambda^\Lambda \in \mathrm{cl}(\Lambda)$ satisfies $\|Z - \lambda^\Lambda\|^2_J = \inf_{\lambda\in\Lambda}\|Z - \lambda\|^2_J$, $Z := -J^{-1}G$.

First, it follows from Lemma B.3 that $\Sigma$ is finite. Expressions for $\partial l_t(\theta)/\partial\theta_i$, $i = 1, ..., s$, are given in the proof of Lemma B.4 below. As in Francq and Zakoïan (2012, p 200), by a central limit theorem for strictly stationary and ergodic martingale difference sequences, see e.g. Brown (1971), $\sqrt{T}\partial L_T(\theta_0)/\partial\theta \xrightarrow{w} G$. Moreover, the ergodic theorem implies that $J_T = J + o(1)$ almost surely. The positive definiteness of $J$ is established in Francq and Zakoïan (2012, pp.203-204), and we conclude that 1. holds.
From the derivation of 1. we have that $\|\cdot\|_{J_T}$ is almost surely a norm for $T$ sufficiently large due to the fact that $J$ is positive definite. With $R_T(\theta)$ defined in (A.2), it follows by Lemma B.5 that

$$R_T(\hat{\theta}_T) = o_p(T^{-1/2}\|\hat{\theta}_T - \theta_0\|) + o_p(\|\hat{\theta}_T - \theta_0\|^2) = o_p(T^{-1/2}\|\hat{\theta}_T - \theta_0\|_{J_T}) + o_p(\|\hat{\theta}_T - \theta_0\|^2_{J_T}). \tag{A.6}$$

For sufficiently large $T$, by Lemma B.5, $[\partial^2\hat{L}_T(\theta^\star)/\partial\theta\partial\theta' - \partial^2\hat{L}_T(\theta_0)/\partial\theta\partial\theta'] = [\partial^2 L_T(\theta^\star)/\partial\theta\partial\theta' - \partial^2 L_T(\theta_0)/\partial\theta\partial\theta'] + o_p(1)$. Also by Lemma B.5, $[\partial^2 L_T(\theta^\star)/\partial\theta\partial\theta' - \partial^2 L_T(\theta_0)/\partial\theta\partial\theta'] = \mathbb{E}[\partial^2 l_t(\theta^\star)/\partial\theta\partial\theta'] - \mathbb{E}[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'] + o_p(1)$, so by the continuity of $\mathbb{E}[\partial^2 l_t(\theta)/\partial\theta\partial\theta']$ on $\Theta$ and the consistency of $\hat{\theta}_T$,

$$R^\star_T(\hat{\theta}_T) = o_p(\|\hat{\theta}_T - \theta_0\|^2_{J_T}), \tag{A.7}$$

with $R^\star_T(\theta)$ defined in (A.3). Now from (A.5), (A.6)-(A.7), and the fact that $\hat{\theta}_T$ minimizes $\hat{L}_T(\theta)$,

$$\begin{aligned}
\hat{L}_T(\hat{\theta}_T) - \hat{L}_T(\theta_0) &= \frac{1}{2T}[\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} - \|Z_T\|^2_{J_T}] + R_T(\hat{\theta}_T) + R^\star_T(\hat{\theta}_T) \\
&= \frac{1}{2T}[\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} - \|Z_T\|^2_{J_T}] \\
&\quad + o_p(\|\hat{\theta}_T - \theta_0\|^2_{J_T}) + o_p(T^{-1/2}\|\hat{\theta}_T - \theta_0\|_{J_T}) \le 0. 
\end{aligned} \tag{A.8}$$

Since $\|\cdot\|_{J_T}$ is a norm for $T$ sufficiently large almost surely, it follows from 1. that

$\|Z_T\|_{J_T} = O_p(1)$. This fact together with (A.8) yields

$$\begin{aligned} \|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2 &\leq \|Z_T\|_{J_T}^2 + o_p(\|\sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2) + o_p(\sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T}) \\ &\leq (\|Z_T\|_{J_T} + o_p(\sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T}))^2. \qquad \text{(A.9)} \end{aligned}$$

The triangle inequality and (A.9) imply that

$$\begin{aligned} \sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T} &\leq \|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T} + \|Z_T\|_{J_T} \\ &\leq 2\|Z_T\|_{J_T} + o_p(\sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T}). \end{aligned}$$

We conclude that $\sqrt{T}\|\hat{\theta}_T - \theta_0\|_{J_T}[1 + o_p(1)] \leq O_p(1)$, and hence that 2. holds.
Result 3. is verified by arguments similar to the ones used to verify 2. together with Lemma B.5.
Turning to 4., notice that when $s_1 = 0$, i.e. when $\theta_0 \in \Theta$, it holds that $\hat{\lambda}_T = Z_T$, and the result follows immediately by the consistency of $\hat{\theta}_T$ and Lemma B.5. Let $\hat{\theta}_q$ satisfy $\|Z_T - \sqrt{T}(\hat{\theta}_q - \theta_0)\|_{J_T}^2 = \inf_{\theta \in \Theta} \|Z_T - \sqrt{T}(\theta - \theta_0)\|_{J_T}^2$. It holds that

$$\begin{aligned} \|\sqrt{T}(\hat{\theta}_q - \theta_0)\|_{J_T} &\leq \|Z_T - \sqrt{T}(\hat{\theta}_q - \theta_0)\|_{J_T} + \|Z_T\|_{J_T} \\ &= \inf_{\theta \in \Theta} \|Z_T - \sqrt{T}(\theta - \theta_0)\|_{J_T} + \|Z_T\|_{J_T} \\ &\leq 2\|Z_T\|_{J_T} = O_p(1), \end{aligned}$$

where the first inequality is due to the triangle inequality, the second inequality follows from the fact that $\theta_0 \in \Theta$, and the last equality follows from 1. Similar to the derivations above, we conclude that $\sqrt{T}(\hat{\theta}_q - \theta_0) = O_p(1)$. From (A.5), using that $\hat{\theta}_T$ minimizes $\hat{L}_T(\theta)$, and that $\hat{\theta}_q$ minimizes $\|Z_T - \sqrt{T}(\theta - \theta_0)\|_{J_T}^2$, together with results 2. and 3., we have that

$$\begin{aligned} 0 &\geq T[\hat{L}_T(\hat{\theta}_T) - \hat{L}_T(\hat{\theta}_q)] \\ &= \frac{1}{2}\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2 - \frac{1}{2}\|Z_T - \sqrt{T}(\hat{\theta}_q - \theta_0)\|_{J_T}^2 \\ &\quad + T[R_T^\star(\hat{\theta}_T) + R_T(\hat{\theta}_T) - R_T^\star(\hat{\theta}_q) - R_T(\hat{\theta}_q)] \\ &\geq T[R_T^\star(\hat{\theta}_T) + R_T(\hat{\theta}_T) - R_T^\star(\hat{\theta}_q) - R_T(\hat{\theta}_q)] = o_p(1). \qquad \text{(A.10)} \end{aligned}$$

Hence, using (A.5) and (A.10),

$$\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|_{J_T}^2 = \|Z_T - \sqrt{T}(\hat{\theta}_q - \theta_0)\|_{J_T}^2 + o_p(1). \qquad \text{(A.11)}$$

Note that $\inf_{\theta \in \Theta} \|Z_T - \sqrt{T}(\theta - \theta_0)\|_{J_T}^2 = \inf_{\lambda \in \sqrt{T}(\Theta - \theta_0)} \|Z_T - \lambda\|_{J_T}^2$, where $\sqrt{T}(\Theta - \theta_0) := \{\lambda \in \mathbb{R}^{s_0} : \lambda = \sqrt{T}(\theta - \theta_0), \theta \in \Theta\}$. Moreover 1. and the fact that $\Lambda$ is a cone (Remark 3.3) imply, due to Andrews (1999, Lemma 2), that

$$\inf_{\lambda \in \sqrt{T}(\Theta - \theta_0)} \|Z_T - \lambda\|_{J_T}^2 = \inf_{\lambda \in \Lambda} \|Z_T - \lambda\|_{J_T}^2 + o_p(1). \qquad \text{(A.12)}$$

Let $\hat{\lambda}_T \in \Lambda$ satisfy $\|Z_T - \hat{\lambda}_T\|^2_{J_T} = \inf_{\lambda\in\Lambda}\|Z_T - \lambda\|^2_{J_T}$. Then combining (A.11) and (A.12) yields

$$\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} = \|Z_T - \hat{\lambda}_T\|^2_{J_T} + o_p(1). \tag{A.13}$$

Observe that

$$\begin{aligned}\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} &= \|\sqrt{T}(\hat{\theta}_T - \theta_0) - \hat{\lambda}_T\|^2_{J_T} + \|Z_T - \hat{\lambda}_T\|^2_{J_T} \\ &\quad +2\left\langle Z_T - \hat{\lambda}_T, \hat{\lambda}_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\right\rangle_{J_T}. \end{aligned}\tag{A.14}$$

Using that $\sqrt{T}(\hat{\theta}_T - \theta_0) \in \Lambda$ and that $\Lambda$ is closed for $s_1 > 0$, it follows from Zarantonello (1971, Lemma 1.1),

$$\left\langle Z_T - \hat{\lambda}_T, \hat{\lambda}_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\right\rangle_{J_T} \geq 0. \tag{A.15}$$

Combining (A.14) and (A.15) yields

$$\|Z_T - \sqrt{T}(\hat{\theta}_T - \theta_0)\|^2_{J_T} \geq \|\sqrt{T}(\hat{\theta}_T - \theta_0) - \hat{\lambda}_T\|^2_{J_T} + \|Z_T - \hat{\lambda}_T\|^2_{J_T}. \tag{A.16}$$

In light of (A.13) and (A.16), we conclude that 4. holds.

In line with Andrews (1999, p.1379), since $\Lambda$ is convex, $\hat{\lambda}_T$ is unique. Moreover, since $\hat{\lambda}_T$ satisfies $\|Z_T - \hat{\lambda}_T\|^2_{J_T} = \inf_{\lambda\in\Lambda}\|Z_T - \lambda\|^2_{J_T}$, $\hat{\lambda}_T = f(Z_T, J_T)$ with some implicitly given function $f$. The function $f$ is continuous at all points $(Z_T, J_T)$ where $J_T$ is nonsingular. Since $J$ is nonsingular, the continuous mapping theorem implies that $\hat{\lambda}_T = f(Z_T, J_T) \overset{w}{\to} f(Z, J) = \lambda^\Lambda$, and we conclude that 5. holds. □

***Proof of Theorem 4.1.*** From the proof of Theorem 3.1,

$$T[\hat{L}_T(\hat{\theta}_T) - \hat{L}_T(\theta_0)] = -\frac{1}{2}\|Z_T\|^2_{J_T} + \frac{1}{2}\|Z_T - \hat{\lambda}_T\|^2_{J_T} + o_p(1),$$

so the continuous mapping theorem together with points 1. and 5. from the proof of Theorem 3.1 imply that

$$2T\left[\hat{L}_T\left(\hat{\theta}_T\right) - \hat{L}_T(\theta_0)\right] \overset{w}{\to} -\|Z\|^2_J + \inf_{\lambda\in\Lambda}\|Z - \lambda\|^2_J. \tag{A.17}$$

Next, with $\lambda^\Lambda_\beta$ defined in Theorem 4.1, with $Z_\beta$, $G_\delta$, and $J_{\delta\delta}$ defined according to the partitions in (A.1), and with $\lambda^{\Lambda_\beta}$ satisfying $\|Z_\beta - \lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} = \inf_{\lambda_\beta\in\Lambda_\beta}\|Z_\beta - \lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$, it holds that

$$\begin{aligned}-\|Z\|^2_J + \inf_{\lambda\in\Lambda}\|Z - \lambda\|^2_J &= -\|Z_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J^{-1}_{\delta\delta}} + \inf_{\lambda\in\Lambda}\|Z - \lambda\|^2_J \\ &= -\|Z_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J^{-1}_{\delta\delta}} + \|Z_\beta\|^2_{(KJ^{-1}K')^{-1}} - \|\lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} \\ &= -\|G_\delta\|^2_{J^{-1}_{\delta\delta}} - \|\lambda^\Lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}, \end{aligned}\tag{A.18}$$

where the first equality follows from Lemma B.6.1. The second equality follows from Lemma B.6.2 and Perlman (1969, Lemma 4.1), and the third equality follows from Lemma

B.6.3. Combining (A.17) and (A.18) yields

$$2T[\hat{L}_T(\hat{\theta}_T) - \hat{L}_T(\theta_0)] \quad \overset{w}{\to} \quad -\|\lambda_\beta^\Lambda\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J_{\delta\delta}^{-1}}. \tag{A.19}$$

Notice that since $\theta_0 \in \Theta_0 \subset \Theta$ and $\Lambda_0 = \Lambda_{0,\beta_1} \times \Lambda_{\beta_2} \times \Lambda_\delta = \{0_{\tilde{s}_1 \times 1}\} \times \mathbb{R}_+^{\tilde{s}_2} \times \mathbb{R}^{s_2}$, it is possible, due to Assumption 7, to derive points 1.-6. in the proof of Theorem 3.1 with $\hat{\theta}_T$, $\hat{\lambda}_T$, $\lambda^\Lambda$, and $\Lambda$ replaced by $\tilde{\theta}_T$, $\tilde{\lambda}_T$, $\lambda^{\Lambda_0}$, and $\Lambda_0$, respectively. In particular, and similar to the derivations above,

$$2T[\hat{L}_T(\tilde{\theta}_T) - \hat{L}_T(\theta_0)] \quad \overset{w}{\to} \quad -\|\lambda_\beta^{\Lambda_0}\|^2_{(KJ^{-1}K')^{-1}} - \|G_\delta\|^2_{J_{\delta\delta}^{-1}}. \tag{A.20}$$

The convergence of (A.19) and (A.20) holds jointly, since the convergence of the two terms are due to point 1. in the proof of Theorem 3.1. This joint convergence and the Cramér-Wold theorem yield the limiting distribution of $QLR_T$.
Next, (4.7) follows by (3.12), Theorem B.1, and the continuous mapping theorem.
Lastly, we turn to the limiting distribution of $LM_T$. It holds, due to the consistency of $\tilde{\theta}_T$, Lemma B.1, and the invertibility of $J$ (Theorem 3.1), $\hat{J}_T(\tilde{\theta}_T)^{-1} = J^{-1} + o_p(1)$. By a Taylor-type expansion and Lemma B.5.1

$$\sqrt{T}\hat{S}_T(\tilde{\theta}_T) = \sqrt{T}S_T(\theta_0) + \hat{J}_T(\theta^\star)\sqrt{T}(\tilde{\theta}_T - \theta_0) + o_p(1),$$

where $\theta^\star$ is between $\tilde{\theta}_T$ and $\theta_0$ as in Jensen and Rahbek (2004, Proof of Lemma 1). By Lemma B.1 and by using that $\theta^\star = \theta_0 + o_p(1)$, it holds that $\hat{J}_T(\theta^\star) = J + o_p(1)$. Hence, using that $\sqrt{T}S_T(\theta_0)$ and $\sqrt{T}(\tilde{\theta}_T - \theta_0)$ are both $O_p(1)$,

$$\begin{aligned}
\sqrt{T}K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{S}_T(\tilde{\theta}_T) &= K_1J^{-1}\left[\sqrt{T}S_T(\theta_0) + J\sqrt{T}(\tilde{\theta}_T - \theta_0)\right] + o_p(1) \\
&= K_1J^{-1}\sqrt{T}S_T(\theta_0) + K_1\sqrt{T}(\tilde{\theta}_T - \theta_0) + o_p(1).
\end{aligned}$$

Since $K_1(\tilde{\theta}_T - \theta_0) = \beta_{1,0} = 0_{\tilde{s}_1 \times 1}$, by Slutsky's lemma and the fact that $\sqrt{T}\partial L_T(\theta_0)/\partial\theta \overset{w}{\to} G$,

$$\begin{aligned}
\sqrt{T}K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{S}_T(\tilde{\theta}_T) &= K_1J^{-1}\sqrt{T}S_T(\theta_0) + o_p(1) \\
&\overset{w}{\to} N\left(0, K_1J^{-1}\Sigma J^{-1}K_1'\right).
\end{aligned} \tag{A.21}$$

By Lemma B.1 and the fact that $\sqrt{T}(\tilde{\theta}_T - \theta_0) = O_p(1)$,

$$K_1\hat{J}_T(\tilde{\theta}_T)^{-1}\hat{\Sigma}_T(\tilde{\theta}_T)\hat{J}_T(\tilde{\theta}_T)^{-1}K_1' = K_1J^{-1}\Sigma J^{-1}K_1' + o_p(1). \tag{A.22}$$

Hence (4.8) follows by combining (A.21) and (A.22) and applying Slutsky's lemma and the continuous mapping theorem. $\square$

## B Lemmas

**Lemma B.1.** *With $J$ and $\Sigma$ given in (3.13) and $\hat{J}_T(\theta)$ and $\hat{\Sigma}_T(\theta)$ given in (4.3), let $\bar{\theta}_T \in \Theta$ satisfy $\bar{\theta}_T = \theta_0 + o_p(1)$. Under the assumptions of Theorem 3.1,*

$$\hat{J}_T(\bar{\theta}_T) = J + o_p(1). \tag{B.1}$$

*Additionally, suppose that $\sqrt{T}(\bar{\theta}_T - \theta_0) = O_p(1)$. Then*

$$\hat{\Sigma}_T(\bar{\theta}_T) = \Sigma + o_p(1). \tag{B.2}$$

*Proof.* The proof is quite similar to the arguments given in Ling and McAleer (2010, p.100). Define, $J_T(\theta) := T^{-1}\sum_{t=1}^{T} \partial^2 l_t(\theta)/\partial\theta\partial\theta'$, where $l_t(\theta)$ is given by (3.5). Lemma B.5 implies that $\hat{J}_T(\bar{\theta}_T) = J_T(\bar{\theta}_T) + o_p(1)$, so in order to establish (B.1) it remains to show that $J_T(\bar{\theta}_T) = J + o_p(1)$. This property follows directly from Lemma B.5, the consistency of $\bar{\theta}_T$, and the fact that $\mathbb{E}[\partial^2 l_t(\theta)/\partial\theta\partial\theta']$ is continuous as $\theta_0$.
Next, we seek to prove (B.2). Notice that with $\hat{l}_t(\theta)$ given by (3.2),

$$\begin{aligned}
\hat{\Sigma}_T(\bar{\theta}_T) \;=\; & \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'} + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\theta_0)}{\partial\theta}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta'} - \frac{\partial l_t(\theta_0)}{\partial\theta'}\right] \\
& + \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right]\frac{\partial l_t(\theta_0)}{\partial\theta'} \\
& + \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right]\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta'} - \frac{\partial l_t(\theta_0)}{\partial\theta'}\right].
\end{aligned} \tag{B.3}$$

The ergodic theorem implies that $T^{-1}\sum_{t=1}^{T}[\partial l_t(\theta_0)/\partial\theta][\partial l_t(\theta_0)/\partial\theta'] = \Sigma + o_p(1)$, so it remains to show that the other terms in (B.3) vanish with probability approaching one. It suffices to establish that

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right]\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta'} = o_p(1) \tag{B.4}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\hat{l}_t(\bar{\theta}_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right]\frac{\partial l_t(\theta_0)}{\partial\theta'} = o_p(1). \tag{B.5}$$

A Taylor-type expansion yields

$$T^{-1/2}\sum_{t=1}^{T}[\partial\hat{l}_t(\bar{\theta}_T)/\partial\theta] = T^{-1/2}\sum_{t=1}^{T}[\partial\hat{l}_t(\theta_0)/\partial\theta] + \hat{J}_T(\theta_T^\star)\sqrt{T}(\bar{\theta}_T - \theta_0),$$

where $\theta_T^\star$ is between $\bar{\theta}_T$ and $\theta_0$. By Lemma B.5.1 and point 1. in the proof of Theorem 3.1, $T^{-1/2}\sum_{t=1}^{T}\partial\hat{l}_t(\theta_0)/\partial\theta = O_p(1)$, and using arguments similar to the ones used to show

(B.1), $\hat{J}_T(\theta_T^\star) = J + o_p(1)$. Hence, using that $\sqrt{T}(\bar{\theta}_T - \theta_0) = O_p(1)$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \hat{l}_t(\bar{\theta}_T)}{\partial \theta} = O_p(1). \tag{B.6}$$

Moreover, also by a Taylor-type expansion,

$$\frac{\partial \hat{l}_t(\bar{\theta}_T)}{\partial \theta} - \frac{\partial l_t(\theta_0)}{\partial \theta} = \left[\frac{\partial \hat{l}_t(\theta_0)}{\partial \theta} - \frac{\partial l_t(\theta_0)}{\partial \theta}\right] + \left\{\left[\frac{\partial^2 \hat{l}_t(\theta_T^\star)}{\partial \theta \partial \theta'} - \frac{\partial^2 l_t(\theta_T^\star)}{\partial \theta \partial \theta'}\right] + \frac{\partial^2 l_t(\theta_T^\star)}{\partial \theta \partial \theta'}\right\}(\bar{\theta}_T - \theta_0). \tag{B.7}$$

For any $\epsilon > 0$ and some $r > 0$, by Boole's and the generalized Chebyshev inequalities,

$$\mathbb{P}\left(\max_{t \in \mathbb{N}} \left\| \frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial \hat{l}_t(\theta_0)}{\partial \theta} \right\| > \epsilon \sqrt{T}\right) \leq \frac{1}{\epsilon^r T^{(r/2)}} \sum_{t=1}^{\infty} \mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial l_t(\theta)}{\partial \theta} - \frac{\partial \hat{l}_t(\theta)}{\partial \theta} \right\|^r\right] = o(1), \tag{B.8}$$

where we have used Lemma B.4.1. Likewise, using Lemma B.4.3, we have that

$$\frac{1}{\sqrt{T}} \max_{t \in \mathbb{N}} \left\| \frac{\partial^2 \hat{l}_t(\theta_T^\star)}{\partial \theta \partial \theta'} - \frac{\partial^2 l_t(\theta_T^\star)}{\partial \theta \partial \theta'} \right\| = o_p(1), \tag{B.9}$$

and using Lemma B.4.4,

$$\frac{1}{\sqrt{T}} \left\| \frac{\partial^2 l_t(\theta_T^\star)}{\partial \theta \partial \theta'} \right\| = o_p(1). \tag{B.10}$$

Combining (B.6), (B.7), (B.8), (B.9), (B.10), and that $(\bar{\theta}_T - \theta_0) = o_p(1)$ yields (B.4). Similar arguments yield (B.5). □

**Lemma B.2.** *Let $\hat{h}_t(\theta)$ and $h_t(\theta)$ be given by (3.4) and (3.7), respectively, and let $\hat{D}_t(\theta)$ and $D_t(\theta)$ be given by (3.3) and (3.6), respectively. Suppose that the assumptions of Theorem 3.1 are satisfied. It holds that for all $t \in \mathbb{N}_0$, $i, j = 1, ..., d + 2d^2$, and some $k \geq 0$,*

$$\mathbb{E}[\sup_{\theta \in \Theta} \|h_t(\theta)\|^3] < \infty, \quad \mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial h_t(\theta)}{\partial \theta_i} \right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 h_t(\theta)}{\partial \theta_i \partial \theta_j} \right\|^3\right] < \infty,$$

$$\mathbb{E}[\sup_{\theta \in \Theta} \left\|\hat{h}_t(\theta)\right\|^3] < \infty, \quad \mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i} \right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 \hat{h}_t(\theta)}{\partial \theta_i \partial \theta_j} \right\|^3\right] < \infty,$$

$$\sup_{\theta \in \Theta} \|R^{-1}(\theta)\| \leq \mathcal{C}, \quad \sup_{\theta \in \Theta} \|D_t^{-1}(\theta)\| \leq \mathcal{C}, \quad \sup_{\theta \in \Theta} \|\hat{D}_t^{-1}(\theta)\| \leq \mathcal{C},$$

$$\mathbb{E}[\sup_{\theta \in \Theta} \|h_t(\theta) - \hat{h}_t(\theta)\|] = O(t^k \phi^t),$$

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial h_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i} \right\|\right] = O(t^k \phi^t), \quad \mathbb{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 h_t(\theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \hat{h}_t(\theta)}{\partial \theta_i \partial \theta_j} \right\|\right] = O(t^k \phi^t).$$

*Proof.* Notice that since $\rho(B) < 1$ on $\Theta$, and $\Theta$ is compact

$$\sup_{\theta\in\theta} \|B^t\| \leq \mathcal{C}\phi^t. \tag{B.11}$$

Since $\rho(B) < 1$, recursions give that $h_t(\theta) = \sum_{i=0}^{\infty} B^i(\kappa + AX_{t-1-i}^{\odot 2})$, so by repeated use of Minkowski's inequality, the compactness of $\Theta$, (B.11), and the fact that $\mathbb{E}[\|X_t\|^6] < \infty$ yield that

$$\mathbb{E}[\sup_{\theta\in\Theta} \|h_t(\theta)\|^3] < \infty. \tag{B.12}$$

Moreover, $\hat{h}_t(\theta) = \sum_{i=0}^{t-1} B^i(\kappa + AX_{t-1-i}^{\odot 2}) + B^t\hat{h}_0$, so similar arguments and the fact that $\hat{h}_0$ is fixed yield that for all $t \in \mathbb{N}_0$, $\mathbb{E}[\sup_{\theta\in\Theta} \|\hat{h}_t(\theta)\|^3] < \infty$. Next, we consider the partial derivatives (potentially of the left/right type) of $h_t(\theta)$. For convenience, we differentiate with respect to the standard parametrization as introduced in subsection 2.1, i.e. without loss of generality we let $\theta = (\kappa', \mathrm{vec}(A)', \mathrm{vec}(B)', \mathrm{vech}^0(R)')'$. Let $\tilde{r}_2 := d + d^2$ and $\tilde{r}_1 := d + 2d^2$. Using that $\rho(B) < 1$,

$$\frac{\partial h_t(\theta)}{\partial\theta_i} = \sum_{j=0}^{\infty} B^i \frac{\partial\kappa}{\partial\theta_i} \quad \text{for } i = 1, ..., d,$$

$$\frac{\partial h_t(\theta)}{\partial\theta_i} = \sum_{j=0}^{\infty} B^i \frac{\partial A}{\partial\theta_i} X_{t-1-i}^{\odot 2} \quad \text{for } i = d+1, ..., \tilde{r}_2,$$

$$\frac{\partial h_t(\theta)}{\partial\theta_i} = \sum_{j=0}^{\infty} B^i \frac{\partial B}{\partial\theta_i} h_{t-1-i} \quad \text{for } i = \tilde{r}_2 + 1, ..., \tilde{r}_1.$$

So repeated use of Minkowski's inequality, $\mathbb{E}[\|X_t\|^6] < \infty$, (B.12), (B.11), and the compactness of $\Theta$ yield that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial h_t(\theta)}{\partial\theta_i}\right\|^3\right] < \infty \tag{B.13}$$

for $i = 1, ..., \tilde{r}_1$. By similar arguments,

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\hat{h}_t(\theta)}{\partial\theta_i}\right\|^3\right] < \infty, \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\hat{h}_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|^3\right] < \infty, \tag{B.14}$$

for all $i, j = 1, ..., \tilde{r}_1$. Moreover, $\sup_{\theta\in\Theta} \|R^{-1}(\theta)\| \leq \mathcal{C}$, $\sup_{\theta\in\Theta} \|D_t^{-1}(\theta)\| \leq \mathcal{C}$, and $\sup_{\theta\in\Theta} \|\hat{D}_t^{-1}(\theta)\| \leq \mathcal{C}$ follow by arguments given in Francq and Zakoïan (2012, p.195). We have that $h_t(\theta) - \hat{h}_t(\theta) = B^t[h_0(\theta) - \hat{h}_0]$, so (B.11), (B.12) and the fact that $\hat{h}_0$ is fixed give that $\mathbb{E}[\sup_{\theta\in\Theta} \|h_t(\theta) - \hat{h}_t(\theta)\|] = O(\phi^t)$. Similarly,

$$\frac{\partial h_t(\theta)}{\partial\theta_i} - \frac{\partial\hat{h}_t(\theta)}{\partial\theta_i} = B^t\left[\frac{\partial h_0(\theta)}{\partial\theta_i} - \frac{\partial\hat{h}_0(\theta)}{\partial\theta_i}\right] \quad \text{for } i = 1, ..., \tilde{r}_2,$$

$$\frac{\partial h_t(\theta)}{\partial\theta_i} - \frac{\partial\hat{h}_t(\theta)}{\partial\theta_i} = \frac{\partial B^t}{\partial\theta_i}\left[h_0(\theta) - \hat{h}_0\right] + B^t\left[\frac{\partial h_0(\theta)}{\partial\theta_i} - \frac{\partial\hat{h}_0(\theta)}{\partial\theta_i}\right] \quad \text{for } i = \tilde{r}_2 + 1, ..., \tilde{r}_1,$$

and we conclude, using (B.13), that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial h_t(\theta)}{\partial\theta_i}-\frac{\partial\hat{h}_t(\theta)}{\partial\theta_i}\right\|\right]=O(t\phi^t)\quad\text{for } i=1,...,\tilde{r}_1. \tag{B.15}$$

Likewise, using (B.14),

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_t(\theta)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\hat{h}_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|\right]=O(t^2\phi^t)\quad\text{for } i,j=1,...,\tilde{r}_1. \tag{B.16}$$

□

**Lemma B.3.** *Under the assumptions of Theorem 3.1, the matrix $\Sigma$ defined in (3.13) is finite.*

*Proof.* Due to Hölder's inequality, it suffices to show that $\mathbb{E}\{[\partial l_t(\theta_0)/\partial\theta_i)]^2\}<\infty$ for all $i=1,...,s_0$, where $s_0$ is the dimension of $\theta$. Similar to the proof of Lemma B.2, we consider (without loss of generality) differentiation with respect to the standard parametrization where $\theta=(\kappa',\mathrm{vec}(A)',\mathrm{vec}(B)',\mathrm{vech}^0(R)')'$. We define $\tilde{r}_2:=d+d^2$ and $\tilde{r}_1:=d+2d^2$. From Francq and Zakoïan (2012, p.198), it holds that

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i}=\mathrm{tr}\left\{(I_d-R_0^{-1}\varepsilon_t\varepsilon_t')\frac{\partial D_{0t}}{\partial\theta_i}D_{0t}^{-1}+(I_d-\varepsilon_t\varepsilon_t'R_0^{-1})D_{0t}^{-1}\frac{\partial D_{0t}}{\partial\theta_i}\right\}$$

for $i=1,...,\tilde{r}_1$, where the "0" indicates that the functions are evaluated at $\theta_0$, and $\varepsilon_t:=R_0^{1/2}Z_t$. It holds that for $i=1,...,\tilde{r}_1$,

$$\frac{\partial D_t}{\partial\theta_i}=\frac{1}{2}D_t^{-1}\mathrm{diag}\left(\frac{\partial h_t(\theta)}{\partial\theta_i}\right), \tag{B.17}$$

so by Lemma B.2, it holds that $\mathbb{E}[\|\partial D_{0t}/\partial\theta_i\|^3]<\infty$. Since $\partial D_{0t}/\partial\theta_i$ and $\varepsilon_t\varepsilon_t'$ are independent, and $\mathbb{E}[\|\varepsilon_t\|^6]<\infty$, we conclude using Hölder's inequality that $\mathbb{E}\{[\partial l_t(\theta_0)/\partial\theta_i]^2\}<\infty$ for $i=1,...,\tilde{r}_1$. Moreover, from Francq and Zakoïan (2012, p.198), it holds that

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i}=\mathrm{tr}\left\{(I_d-R_0^{-1}\varepsilon_t\varepsilon_t')\left(R_0^{-1}\frac{\partial R_0}{\partial\theta_i}\right)\right\}$$

for $i=\tilde{r}_1+1,...,s_0$. Using similar arguments as above, we conclude that $\mathbb{E}\{[\partial l_t(\theta_0)/\partial\theta_i]^2\}<\infty$ for $i=\tilde{r}_1+1,...,s_0$. □

**Lemma B.4.** *Suppose that the assumptions of Theorem 3.1 hold. Then with $\hat{l}_t(\theta)$ and $l_t(\theta)$ given by (3.2) and (3.5), respectively, the following statements are true.*

1. *For some $k\geq 0$ and some $u>0$,*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\theta)}{\partial\theta}-\frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right\|^u\right]=O(t^k\phi^t)\quad\forall t\in\mathbb{N}.$$

*2. For some $k \geq 0$ and some $u > 0$,*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left|l_t(\theta) - \hat{l}_t(\theta)\right|^r\right] = O(t^k\phi^t) \quad \forall t \in \mathbb{N}.$$

*3. For some $k \geq 0$ and some $u > 0$,*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial\hat{l}_t^2(\theta)}{\partial\theta\partial\theta'}\right\|^u\right] = O(t^k\phi^t) \quad \forall t \in \mathbb{N}.$$

*4. $\mathbb{E}[\sup_{\theta\in\Theta}\|\partial^2 l_t(\theta)/\partial\theta\partial\theta'\|] < \infty$.*

*Proof.* Similar to the proof of Lemma B.2, we consider differentiation with respect to the standard parametrization where $\theta = (\kappa', \text{vec}(A)', \text{vec}(B)', \text{vech}^0(R)')'$, and define $\tilde{r}_2 := d + d^2$ and $\tilde{r}_1 := d + 2d^2$. From Francq and Zakoïan (2012, p.198) it holds that for $i = 1, ..., \tilde{r}_1$,

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial\theta_i} &= \text{tr}\left\{D_t^{-1}\left[2I_d - \left(X_tX_t'D_t^{-1}R^{-1}D_t^{-1} + D_t^{-1}X_tX_t'D_t^{-1}R^{-1}\right)\right]\frac{\partial D_t}{\partial\theta_i}\right\}\\ &= \xi_t'\left[D_t^{-1}\otimes I_d\right]\text{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)\end{aligned}$$

with

$$\xi_t := \text{vec}\left[2I_d - \left(X_tX_t'D_t^{-1}R^{-1}D_t^{-1} + D_t^{-1}X_tX_t'D_t^{-1}R^{-1}\right)\right]. \tag{B.18}$$

Similarly,

$$\frac{\partial\hat{l}_t(\theta)}{\partial\theta_i} = \hat{\xi}_t'[\hat{D}_t^{-1}\otimes I_d]\,\text{vec}\left(\frac{\partial\hat{D}_t}{\partial\theta_i}\right),$$

with $\hat{\xi}_t := \text{vec}[2I_d - (X_tX_t'\hat{D}_t^{-1}R^{-1}\hat{D}_t^{-1} + \hat{D}_t^{-1}X_tX_t'\hat{D}_t^{-1}R^{-1})]$. Hence,

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial\theta_i} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta_i} &= \xi_t'\left[D_t^{-1}\otimes I_d\right]\text{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right) - \hat{\xi}_t'[\hat{D}_t^{-1}\otimes I_d]\,\text{vec}\left(\frac{\partial\hat{D}_t}{\partial\theta_i}\right)\\ &= (\xi_t' - \hat{\xi}_t')\left[D_t^{-1}\otimes I_d\right]\text{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)\\ &\quad + \hat{\xi}_t'[(D_t^{-1} - \hat{D}_t^{-1})\otimes I_d]\,\text{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right)\\ &\quad + \hat{\xi}_t'[\hat{D}_t^{-1}\otimes I_d]\left[\text{vec}\left(\frac{\partial D_t}{\partial\theta_i}\right) - \text{vec}\left(\frac{\partial\hat{D}_t}{\partial\theta_i}\right)\right].\end{aligned}$$

It holds that

$$\begin{aligned}\xi_t - \hat{\xi}_t &= \text{vec}\left[X_tX_t'(\hat{D}_t^{-1} - D_t^{-1})R^{-1}\hat{D}_t^{-1}\right] - \text{vec}\left[X_tX_t'D_t^{-1}R^{-1}(D_t^{-1} - \hat{D}_t^{-1})\right]\\ &\quad + \text{vec}\left[\hat{D}_t^{-1}X_tX_t'(\hat{D}_t^{-1} - D_t^{-1})R^{-1}\right] - \text{vec}\left[(D_t^{-1} - \hat{D}_t^{-1})X_tX_t'D_t^{-1}R^{-1}\right],\end{aligned}$$

and

$$\|\hat{D}_t^{-1} - D_t^{-1}\| = \|\hat{D}_t^{-1}(\hat{D}_t - D_t)D_t^{-1}\| \leq \mathcal{C}\|\hat{D} - D_t\|,$$

where we have used Lemma B.2. By the same lemma for some $k \geq 0$,

$$\mathbb{E}[\sup_{\theta \in \Theta} \| h_t(\theta) - \hat{h}_t(\theta) \|] = O(t^k \phi^t),$$

so we have that for some for some $\tilde{u} > 0$ and some $k \geq 0$,

$$\mathbb{E}[\sup_{\theta \in \Theta} \| \hat{D}_t^{-1} - D_t^{-1} \|^{\tilde{u}}] = O(t^k \phi^t). \tag{B.19}$$

Consequently, by Hölder's inequality for some $u^\star > 0$,

$$\mathbb{E}[\sup_{\theta \in \Theta} \| \xi_t - \hat{\xi}_t \|^{u^\star}] = O(t^k \phi^t)$$

For $i = 1, ..., \tilde{r}_1$,

$$\frac{\partial D_t}{\partial \theta_i} = \frac{1}{2} D_t^{-1} \operatorname{diag}\left( \frac{\partial h_t}{\partial \theta_i} \right),$$

and due to (B.19) and

$$\mathbb{E}\left[ \sup_{\theta \in \Theta} \left\| \frac{\partial h_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{h}_t(\theta)}{\partial \theta_i} \right\| \right] = O(t^k \phi^t),$$

by Lemma B.2, it holds that for some $u^\star > 0$,

$$\mathbb{E}\left[ \sup_{\theta \in \Theta} \left\| \operatorname{vec}\left( \frac{\partial D_t}{\partial \theta_i} \right) - \operatorname{vec}\left( \frac{\partial \hat{D}_t}{\partial \theta_i} \right) \right\|^{u^\star} \right] = O(t^k \phi^t).$$

Consequently, by Hölder's inequality we have that for $i = 1, ..., \tilde{r}_1$ and some $u > 0$

$$\mathbb{E}\left[ \sup_{\theta \in \Theta} \left| \frac{\partial l_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{l}_t(\theta)}{\partial \theta_i} \right|^u \right] = O(t^k \phi^t)$$

For $i = \tilde{r}_1 + 1, ..., s_0$,

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = \operatorname{tr}\left( R^{-1} \frac{\partial R}{\partial \theta_i} \right) - \operatorname{vec}(D_t^{-1})' \left[ (X_t X_t') \otimes \left( R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1} \right) \right] \operatorname{vec}(D_t^{-1})$$

and

$$\frac{\partial \hat{l}_t(\theta)}{\partial \theta_i} = \operatorname{tr}\left( R^{-1} \frac{\partial R}{\partial \theta_i} \right) - \operatorname{vec}(\hat{D}_t^{-1})' \left[ (X_t X_t') \otimes \left( R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1} \right) \right] \operatorname{vec}(\hat{D}_t^{-1}),$$

so by similar arguments as above, using (B.19) and Lemma B.2,

$$\mathbb{E}\left[ \sup_{\theta \in \Theta} \left| \frac{\partial l_t(\theta)}{\partial \theta_i} - \frac{\partial \hat{l}_t(\theta)}{\partial \theta_i} \right|^u \right] = O(t^k \phi^t)$$

$i = \tilde{r}_1 + 1, ..., s_0$. Using the $c_r$-inequality, we conclude that 1. holds.

Turning to 2., from Francq and Zakoïan (2012, pp.195-196),

$$\sup_{\theta\in\Theta}\left|l_t(\theta)-\hat{l}_t(\theta)\right| \leq \sup_{\theta\in\Theta}\left|\mathrm{tr}\left\{X_tX_t'(H_t^{-1}-\hat{H}_t^{-1})\right\}\right| + \sup_{\theta\in\Theta}\left|\log\{\det(H_t)\}-\log\left\{\det(\hat{H}_t)\right\}\right|. \tag{B.20}$$

It holds that

$$\sup_{\theta\in\Theta}\left|\mathrm{tr}\left\{X_tX_t'(H_t^{-1}-\hat{H}_t^{-1})\right\}\right| \leq \|X_tX_t'\|\sup_{\theta\in\Theta}\left\|H_t^{-1}-\hat{H}_t^{-1}\right\|,$$

Since

$$\begin{aligned} H_t^{-1}-\hat{H}_t^{-1} &= H_t^{-1}(\hat{H}_t-H_t)\hat{H}_t^{-1} \\ &= D_t^{-1}R^{-1}D_t^{-1}[(\hat{D}_t-D_t)R\hat{D}_t+D_tR(\hat{D}_t-D_t)]\hat{D}_t^{-1}R^{-1}\hat{D}_t^{-1}, \end{aligned}$$

it follows from Lemma B.2 and Hölder's inequality that for some $u^\star>0$,

$$\mathbb{E}[\sup_{\theta\in\Theta}|\mathrm{tr}\{X_tX_t'(H_t^{-1}-\hat{H}_t^{-1})\}|^{u^\star}]=O(t^k\phi^t). \tag{B.21}$$

Next let $h_{it}$ and $\hat{h}_{it}$ denote the $i$th element $(i=1,..,d)$ of $h_t(\theta)$ and $\hat{h}_t(\theta)$ respectively. Consider the second term in (B.20). From Ling and McAleer (2003, p.302),

$$\begin{aligned} |\log\{\det(H_t)\}-\log\{\det(\hat{H}_t)\}| &= |\log\{\det(D_t^2\hat{D}_t^{-2})\}| \\ &= |\log(\prod_{i=1}^d h_{it}/\hat{h}_{it})| = |\sum_{i=1}^d \log(h_{it}/\hat{h}_{it})|, \end{aligned}$$

where we have used that $h_{it}$ and $\hat{h}_{it}$ have a positive lower bound for each $i$ uniformly on $\Theta$. Since $\log(1+x)\leq x$ for $x>-1$, we have that

$$|\log\{\det(H_t)\}-\log\{\det(\hat{H}_t)\}| \leq \sum_{i=1}^d|\log\{1+(h_{it}-\hat{h}_{it})\hat{h}_{it}^{-1}\}| \leq \sum_{i=1}^d|(h_{it}-\hat{h}_{it})\hat{h}_{it}^{-1}|,$$

so, using Lemma B.2, for some $u^\star>0$,

$$\mathbb{E}[\sup_{\theta\in\Theta}\left|\log\{\det(H_t)\}-\log\{\det(\hat{H}_t)\}\right|^{u^\star}]=O(t^k\phi^t). \tag{B.22}$$

By combining (B.20), (B.21), (B.22), and Hölder's inequality, we conclude that point 2. holds.

Turning to point 3., expressions for $\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j$ for different choices of $i$ and $j$ are stated in Francq and Zakoïan (2012, pp.200-201) (note that in Francq and Zakoïan (2012) $\epsilon_t$ corresponds to $X_t$ here). By similar arguments as above, relying on Lemma B.2, we conclude that 3. holds.

In order to establish 4. it suffices to show that $\mathbb{E}[\sup_{\theta\in\Theta}|\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j|]<\infty$ for all

$i, j = 1, ..., s_0$. Again, by relying on expressions for $\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_i$ for different choices of $i$ and $j$ are stated in Francq and Zakoïan (2012, pp.200-201), it is seen that this moment restriction holds due to Lemma B.2 and Hölder's inequality. □

**Lemma B.5.** *Under the assumptions of Theorem 3.1, with $\hat{l}_t(\theta)$ and $l_t(\theta)$ given by (3.2) and (3.5), respectively,*

1. $\sup_{\theta\in\Theta} \|\frac{1}{T}\sum_{t=1}^T \partial l_t(\theta)/\partial\theta - \frac{1}{T}\sum_{t=1}^T \partial\hat{l}_t(\theta)/\partial\theta\| = o_p(T^{-1/2})$.
2. $\sup_{\theta\in\Theta} |\frac{1}{T}\sum_{t=1}^T l_t(\theta) - \frac{1}{T}\sum_{t=1}^T \hat{l}_t(\theta)| = o_p(T^{-1})$.
3. $\sup_{\theta\in\Theta} \|\frac{1}{T}\sum_{t=1}^T \partial^2 l_t(\theta)/\partial\theta\partial\theta' - \frac{1}{T}\sum_{t=1}^T \partial^2\hat{l}_t(\theta)/\partial\theta\partial\theta'\| = o(1) \quad a.s.$
4. $\sup_{\theta\in\Theta} \|\frac{1}{T}\sum_{t=1}^T \partial^2 l_t(\theta)/\partial\theta\partial\theta' - \mathbb{E}[\partial^2 l_t(\theta)/\partial\theta\partial\theta']\| = o(1) \quad a.s.$

*Proof.* In order to show 1., we use arguments similar to the ones given in Pedersen and Rahbek (2014, Proof of Lemma B.11), see also Hafner and Preminger (2009, Proof of Lemma 4). For any $\epsilon > 0$ and some $u > 0$, by the generalized Chebyshev inequality,

$$\mathbb{P}\left(\sqrt{T}\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^T\left[\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right]\right\| > \epsilon\right) \le \frac{T^{(u-2)/2}}{\epsilon^r}\sum_{t=1}^T \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\theta)}{\partial\theta} - \frac{\partial\hat{l}_t(\theta)}{\partial\theta}\right\|^u\right] = o(1),$$

choosing $u < 2$, where we have used Lemma B.4.1.
Using similar arguments and Lemma B.4.2, we conclude that point 2. holds.
Turning to point 3., for any $\epsilon > 0$ and some $\tilde{u} > 0$, by the generalized Chebyshev inequality,

$$\sum_{t=0}^\infty \mathbb{P}\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\hat{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| > \epsilon\right) \le \epsilon^{-\tilde{u}}\sum_{t=0}^\infty \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\hat{l}_t(\theta)}{\partial\theta\partial\theta'}\right\|^{\tilde{u}}\right] < \infty,$$

where we have used Lemma B.4.3. The Borel-Cantelli lemma then implies that almost surely

$$\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\hat{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| \to 0 \quad \text{as } t\to\infty,$$

and point 3. then follows by Cesàro's mean theorem.
The proof of 4. follows by Lemma B.4.4 and a uniform law of large numbers for ergodic processes, see e.g. Ranga Rao (1962). □

**Lemma B.6.** *Let $Z_\beta$, $G_\delta$, and $J_{\delta\delta}$ be defined according to (A.1). Moreover, with $\Lambda = \Lambda_\beta\times\Lambda_\delta$ defined in Assumption 5, let $\lambda^\Lambda = (\lambda_\beta^{\Lambda\prime}, \lambda_\delta^{\Lambda\prime})'$satisfy $\|Z-\lambda^\Lambda\|_J^2 = \inf_{\lambda\in\Lambda}\|Z-\lambda\|_J^2$ and let $\lambda^{\Lambda_\beta}$ satisfy $\|Z_\beta - \lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} = \inf_{\lambda_\beta\in\Lambda_\beta}\|Z_\beta-\lambda_\beta\|^2_{(KJ^{-1}K')^{-1}}$. Under Assumptions 1-6,*

1. $Z'JZ = Z_\beta'(KJ^{-1}K')^{-1}Z_\beta + G_\delta' J_{\delta\delta}^{-1}G_\delta$,
2. $\|Z-\lambda^\Lambda\|_J^2 = \|Z_\beta-\lambda^{\Lambda_\beta}\|^2_{(KJ^{-1}K')^{-1}} = \|Z_\beta - \lambda_\beta^\Lambda\|^2_{(KJ^{-1}K')^{-1}}$,
3. $\lambda_\beta^\Lambda = \lambda^{\Lambda_\beta}$.

*Proof.* The proof follows the lines of Andrews (1999, Proof of Theorem 4). First, recall that for matrices $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{n\times m}$, and $D \in \mathbb{R}^{n\times n}$ satisfying that $E := \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, $D$, and $(A - BD^{-1}C)$ are nonsingular, then

$$E^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}. \tag{B.23}$$

Define the matrices

$$M := \begin{bmatrix} I_{s_1} \\ -J_{\delta\delta}^{-1}J_{\delta\beta} \end{bmatrix}, \quad P^{\perp} := MK, \quad P := I_{s_0} - P^{\perp}.$$

Observe that by orthogonality

$$(Px_1)'J(P^{\perp}x_2) = 0 \quad \forall x_1, x_2 \in \mathbb{R}^{s_0}. \tag{B.24}$$

By (B.23),

$$KJ^{-1}K' = (J_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta})^{-1}, \tag{B.25}$$

and, moreover,

$$M'JM = J_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta},$$

so

$$M'JM = (KJ^{-1}K')^{-1}. \tag{B.26}$$

Let $\bar{K} := (0_{s_2\times s_1}, I_{s_2})$. By definition $I_{s_0} = (K', \bar{K}')'$, so

$$PJ^{-1}G = \begin{bmatrix} KPJ^{-1}G \\ \bar{K}PJ^{-1}G \end{bmatrix}. \tag{B.27}$$

It holds that

$$\begin{aligned} KP &= K(I_{s_0} - MK) \\ &= K - KMK \\ &= K - [I_{s_1}, 0_{s_1\times s_2}] \begin{bmatrix} I_{s_1} \\ -J_{\delta\delta}^{-1}J_{\delta\beta} \end{bmatrix} K \\ &= 0_{s_1\times s_0}, \end{aligned} \tag{B.28}$$

so

$$KPJ^{-1}G = 0_{s_1\times 1}.$$

Furthermore, make the following partition $J^{-1} = \begin{bmatrix} J^{(1)} & J^{(2)} \\ J^{(3)} & J^{(4)} \end{bmatrix}$ according to (B.23) such that $J^{(1)} := (J_{\beta\beta} - J_{\beta\delta}J_{\delta\delta}^{-1}J_{\delta\beta})^{-1}$, $J^{(2)} := -J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}$, $J^{(3)} := -J_{\delta\delta}^{-1}J_{\delta\beta}J^{(1)}$, and $J^{(4)} :=$

$J_{\delta\delta}^{-1} + J_{\delta\delta}^{-1} J_{\delta\beta} J^{(1)} J_{\beta\delta} J_{\delta\delta}^{-1}$, with $J_{\beta\beta}$, $J_{\beta\delta}$, $J_{\delta\beta}$, and $J_{\delta\delta}$ defined according to (A.1). Then

$$\begin{aligned}
J_{\delta\delta}\overline{K}PJ^{-1}G &= J_{\delta\delta}\bar{K}(I_{s_1+s_2} - MK)J^{-1}G \\
&= J_{\delta\delta}([0_{s_2\times s_1}, I_{s_2}] - \bar{K}M[I_{s_1}, 0_{s_1\times s_2}]) \begin{bmatrix} J^{(1)} & J^{(2)} \\ J^{(3)} & J^{(4)} \end{bmatrix} G \\
&= J_{\delta\delta}([J^{(3)}, J^{(4)}] - \bar{K}M[J^{(1)}, J^{(2)}])G \\
&= J_{\delta\delta}\left([J^{(3)}, J^{(4)}] - \bar{K}\begin{bmatrix} I_{s_1} \\ -J_{\delta\delta}^{-1}J_{\delta\beta} \end{bmatrix}[J^{(1)}, J^{(2)}]\right)G \\
&= J_{\delta\delta}\left([J^{(3)}, J^{(4)}] - [0_{s_2\times s_1}, I_{s_2}]\begin{bmatrix} [J^{(1)}, J^{(2)}] \\ -J_{\delta\delta}^{-1}J_{\delta\beta}[J^{(1)}, J^{(2)}] \end{bmatrix}\right)G \\
&= J_{\delta\delta}([J^{(3)}, J^{(4)}] + J_{\delta\delta}^{-1}J_{\delta\beta}[J^{(1)}, J^{(2)}])G \\
&= ([J_{\delta\delta}J^{(3)}, J_{\delta\delta}J^{(4)}] + J_{\delta\beta}[J^{(1)}, J^{(2)}])G
\end{aligned}$$

Hence,

$$\begin{aligned}
J_{\delta\delta}\bar{K}PJ^{-1}G &= ([-J_{\delta\beta}J^{(1)}, I_r + J_{\delta\beta}J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}] + [J_{\delta\beta}J^{(1)}, -J_{\delta\beta}J^{(1)}J_{\beta\delta}J_{\delta\delta}^{-1}])G \\
&= [0_{s_2\times s_1}, I_{s_2}]G \qquad \text{(B.29)} \\
&= G_\delta.
\end{aligned}$$

Combining (B.27), (B.32), and (B.29) yields

$$PJ^{-1}G = \begin{bmatrix} 0_{s_1\times 1} \\ J_{\delta\delta}^{-1}G_\delta \end{bmatrix}. \qquad \text{(B.30)}$$

Now (B.24) implies that

$$Z'JZ = (PZ)'\,J\,(PZ) + (P^{\perp}Z)'J(P^{\perp}Z).$$

This combined with (B.26), (B.30), and that $Z = -J^{-1}G$ (by definition) proves 1.
For $\lambda = (\lambda_\beta', \lambda_\delta') \in \Lambda_\beta \times \Lambda_\delta$ it holds that

$$P\lambda = \begin{bmatrix} 0_{s_1\times 1} \\ \lambda_\delta + J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta \end{bmatrix}. \qquad \text{(B.31)}$$

Using (B.24), (B.31), (B.30), and (B.26) gives

$$\begin{aligned}
\|Z-\lambda\|_J^2 &= \|P\,(Z-\lambda)\|_J^2 + \|P^{\perp}\,(Z-\lambda)\,\|_J^2 \\
&= \left\|\begin{bmatrix} 0_{s_1\times 1} \\ J_{\delta\delta}^{-1}G_\delta \end{bmatrix} - \begin{bmatrix} 0_{s_1\times 1} \\ \lambda_\delta + J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta \end{bmatrix}\right\|_J^2 + \|K\,(Z-\lambda)\|_{M'JM}^2 \\
&= \|J_{\delta\delta}^{-1}G_\delta - \lambda_\delta - J_{\delta\delta}^{-1}J_{\delta\beta}\lambda_\beta\|_{J_{\delta\delta}}^2 + \|Z_\beta - \lambda_\beta\|_{(KJ^{-1}K')^{-1}}^2. \qquad \text{(B.32)}
\end{aligned}$$

Since $\Lambda = \Lambda_\beta \times \Lambda_\delta$ and $\Lambda_\delta = \mathbb{R}^{s_2}$, for any $\lambda_\beta \in \Lambda_\beta$

$$\inf_{\lambda_\delta \in \Lambda_\delta} \|J_{\delta\delta}^{-1} G_\delta - \lambda_\delta - J_{\delta\delta}^{-1} J_{\delta\beta} \lambda_\beta\|_{J_{\delta\delta}}^2 = \inf_{\lambda_\delta \in \mathbb{R}^r} \|J_{\delta\delta}^{-1} G_\delta - \lambda_\delta - J_{\delta\delta}^{-1} J_{\delta\beta} \lambda_\beta\|_{J_{\delta\delta}}^2 = 0,$$

so

$$\inf_{\lambda \in \Lambda} \|Z - \lambda\|_J^2 = \inf_{\lambda_\beta \in \Lambda_\beta} \|Z_\beta - \lambda_\beta\|_{(KJ^{-1}K')^{-1}}^2,$$

which proves the first equality of 2. holds. Turning to the second equality of 2., notice that

$$\begin{aligned}
0 &\leq \|Z_\beta - \lambda_\beta^\Lambda\|_{(KJ^{-1}K')^{-1}}^2 - \|Z_\beta - \lambda^{\Lambda_\beta}\|_{(KJ^{-1}K')^{-1}}^2 \\
&\leq \|Z_\beta - \lambda_\beta^\Lambda\|_{(KJ^{-1}K')^{-1}}^2 + \|J_{\delta\delta}^{-1} G_\delta - \lambda_\delta^\Lambda - J_{\delta\delta}^{-1} J_{\delta\beta} \lambda_\beta^\Lambda\|_{J_{\delta\delta}}^2 - \|Z_\beta - \lambda^{\Lambda_\beta}\|_{(KJ^{-1}K')^{-1}}^2 \\
&= \|Z - \lambda^\Lambda\|_J^2 - \|Z_\beta - \lambda^{\Lambda_\beta}\|_{(KJ^{-1}K')^{-1}}^2 = 0,
\end{aligned}$$

where we have used (B.32) and the first equality of 2.

Point 3. follows from 2, and the fact that $\lambda^{\Lambda_\beta}$ is unique due to the convexity of $\Lambda_\beta$. $\square$

**Lemma B.7.** *Let $\{Y_t : t \in \mathbb{N}_0\}$, $Y_t = (X_t^{\odot 2\prime}, \sigma_t^{2\prime})'$, be the Markov chain generated by the ECCC-GARCH model (2.1)-(2.4) for $t \geq 1$, with fixed initial values $X_0 := x \in \mathbb{R}^d$ and $\sigma_0^2 := h \in (0,\infty)^d$, and with fixed $\theta = [\kappa_0', \text{vec}(A_0)', \text{vec}(B_0)', \text{vech}^0(R_0)']'$. Suppose that $\rho(B_0) < 1$ and that the diagonal elements of $A_0$ are strictly positive. Let $p \in \mathbb{N}$, and suppose that the distribution, $\Gamma$, of $\varepsilon_t := R_0^{1/2}\eta_t$ admits a probability density strictly positive on $\mathbb{R}^d$ with $\mathbb{E}[(\varepsilon_t^{\odot 2})^{\otimes p}] < \infty$, and $\rho(\mathbb{E}\{[A_0 \text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$. Then $\{Y_t : t \in \mathbb{N}_0\}$ is geometrically ergodic on $[0,\infty)^d \times (0,\infty)^d$, and the associated strictly stationary process $\{Y_t : t \in \mathbb{Z}\}$ is geometrically $\beta$-mixing with $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$.*

*Proof.* The proof is similar to Pedersen (2015, Proof of Lemma B.8). Consider the process $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ given by $\sigma_t^2 = \kappa_0 + [A_0 \,\text{diag}(\varepsilon_{t-1}^{\odot 2}) + B_0]\sigma_{t-1}^2$, with $\sigma_0^2 = h$. Relying on the theory of Boussama et al. (2011), it follows from Pedersen (2015, Proof of Lemma B.8) that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is a Markov chain which is aperiodic and $\psi$-irreducible on $(0,\infty)^d$, see Meyn and Tweedie (2009, Section 4.2). These properties of the Markov chain allow us, due to Tjøstheim (1990), to consider a $k$-step drift criterion for the Markov chain for some $k \in \mathbb{N}$. Specifically, with $\mathscr{B}((0,\infty)^d)$ the Borel $\sigma$-field of $(0,\infty)^d$, we want to show that there exists a small set $\mathcal{K} \in \mathscr{B}((0,\infty)^d)$, positive constants $a < 1$ and $b < \infty$, and a Lyapunov function $V_\sigma : (0,\infty)^d \to [1,\infty)$ such that for some fixed $k \in \mathbb{N}$,

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

With $\iota_{d^p}$ a $(d^p \times 1)$ vector of ones, consider the function $V_\sigma(h) := 1 + \iota_{d^p}' h^{\otimes p}$, and, for some constant $m$ sufficiently large, the set $\mathcal{K} := \{h \in (0,\infty)^d : \iota_{d^p}' h^{\otimes p} \leq m\}$.

For $t \in \mathbb{N}$, it holds that $(\sigma_{t+1}^2)^{\otimes p} = C_{t,p} + [A_0 \,\text{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}(\sigma_t^2)^{\otimes p}$, where for $p \geq 2$

$C_{t,p} := \{C_{t,p-1} \otimes \sigma_t^2 + [A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p-1} \otimes \kappa_0\}$ and $C_{t,1} := \kappa_0$. Recursions give that

$$(\sigma_{t+k}^2)^{\otimes p} = \sum_{i=0}^{k-1} \prod_{j=1}^{i} [A_0 \operatorname{diag}(\varepsilon_{t+k-j}^{\odot 2}) + B_0]^{\otimes p} C_{t+k-1-i,p} + \prod_{i=1}^{k} [A_0 \operatorname{diag}(\varepsilon_{t+k-i}^{\odot 2}) + B_0]^{\otimes p} (\sigma_t^2)^{\otimes p}.$$

Observe that

$$\mathbb{E}\left[V_\sigma(\sigma_k^2) | \sigma_0^2 = h\right] = \frac{1 + \iota_{d^p}' \tilde{C} + \iota_{d^p}' (\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\})^k h^{\otimes p}}{1 + \iota_{d^p}' h^{\otimes p}} V_\sigma(h),$$

where we have used that $\{\varepsilon_t\}$ is i.i.d. and where $\tilde{C}$ contains terms of $h$ of lower order than $p$. Since $\rho(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$ and choosing $k$ sufficiently large, there exists an $m$ large enough such that for $h \in \mathcal{K}^{\complement}$, $V_\sigma(h) \geq 1 + \iota_{d^p}' \tilde{C} + \iota_{d^p}' (\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\})^k h^{\otimes p}$. We conclude that suitable constants $a$ and $b$ exist. In line with Boussama et al. (2011, Section 4.6) it can be shown that $\mathcal{K}$ is small. It then holds that $\{\sigma_t^2 : t \in \mathbb{N}_0\}$ is $V_\sigma$-geometrically ergodic. From Meitz and Saikkonen (2008, Proposition 1 and the comments immediately after) we conclude that $\{Y_t : t \in \mathbb{N}_0\}$ is $V_Y$-geometrically ergodic, for some suitable function $V_Y : [0,\infty)^d \times (0,\infty)^d \to [1,\infty)$, and that the associated strictly stationary process $\{Y_t : t \in \mathbb{Z}\}$ is geometrically $\beta$-mixing. Moreover, $\mathbb{E}[\|(\sigma_t^2)^{\otimes p}\|] \leq \mathcal{C}\mathbb{E}[V_\sigma(\sigma_t^2)] < \infty$, and by using that $\mathbb{E}[(\varepsilon_t^{\odot 2})^{\otimes p}] < \infty$, we have that $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$. □

**Lemma B.8.** *Let $\{X_t : t \in \mathbb{Z}\}$, be a strictly stationary process generated by the ECCC-GARCH model (2.1)-(2.4) with fixed $\theta = [\kappa_0', \operatorname{vec}(A_0)', \operatorname{vec}(B_0)', \operatorname{vech}^0(R_0)']' \in \Theta$. For $p \in \mathbb{N}$ suppose that $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$. Then $\rho(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$, where $\varepsilon_t := R_0^{1/2} \eta_t$.*

*Proof.* The proof is similar to that of Ling and McAleer (2002, Proof of Theorem 2.1). Notice that $\mathbb{E}[(\varepsilon_t^{\odot 2})^{\otimes p}] < \infty$ is necessary for $\mathbb{E}[(X_t^{\odot 2})^{\otimes p}] < \infty$, and that $\mathbb{E}[(\sigma_t^2)^{\otimes p}] < \infty$. Similar to the proof of Lemma B.7, we obtain for $k \in \mathbb{N}$

$$(\sigma_t^2)^{\otimes p} = \sum_{i=0}^{k-1} \prod_{j=1}^{i} [A_0 \operatorname{diag}(\varepsilon_{t-j}^{\odot 2}) + B_0]^{\otimes p} C_{t-1-i,\otimes p} + \prod_{i=1}^{k} [A_0 \operatorname{diag}(\varepsilon_{t-i}^{\odot 2}) + B_0]^{\otimes p} (\sigma_{t-k}^2)^{\otimes p}.$$

Since $\prod_{i=1}^{k} [A_0 \operatorname{diag}(\varepsilon_{t-i}^{\odot 2}) + B_0]^{\otimes p} (\sigma_{t-k}^2)^{\otimes p} \geq 0 \ \forall k$ and $C_{t-1-i,\otimes p} \geq \kappa_0^{\otimes p}$, we obtain

$$\infty > \mathbb{E}[(\sigma_t^2)^{\otimes p}] \geq \sum_{i=0}^{\infty} \left(\mathbb{E}\left\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\right\}\right)^i \kappa_0^{\otimes p}. \tag{B.33}$$

Since $(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) \geq 0$ and $\kappa_0^{\otimes p} \in (0,\infty)^{d^p}$, we have, in light of (B.33), that $\sum_{i=0}^{\infty} (\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\})^i$ converges, which is necessary and sufficient for $\rho(\mathbb{E}\{[A_0 \operatorname{diag}(\varepsilon_t^{\odot 2}) + B_0]^{\otimes p}\}) < 1$. □

## C The $LM_{ECCC}$ statistic of Nakatani and Teräsvirta (2009)

**Proposition C.1.** *Let $\hat{J}_T(\theta)$ and $\hat{S}_T(\theta)$ be defined by (4.3), let $K_1$ be defined by (4.4), and let $\tilde{\theta}_T$ be the constrained estimator given in (4.2). Consider the test statistic given by*

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

*With the matrices $J$ and $\Sigma$ defined in (3.13) and $\lambda^{\Lambda_0}$ defined in (4.5), let $\mathcal{L}(G) = N(0, \Sigma)$ and consider the partitions of $J$, $G$, and $\lambda^{\Lambda_0}$, according to $\theta = (\beta_1', \beta_2', \delta')'$, given by*

$$J = \begin{pmatrix} J_{\beta_1\beta_1} & J_{\beta_1\beta_2} & J_{\beta_1\delta} \\ J_{\beta_2\beta_1} & J_{\beta_2\beta_2} & J_{\beta_2\delta} \\ J_{\delta\beta_1} & J_{\delta\beta_2} & J_{\delta\delta} \end{pmatrix}, \quad G = \begin{pmatrix} G_{\beta_1} \\ G_{\beta_2} \\ G_\delta \end{pmatrix}, \quad \text{and} \quad \lambda^{\Lambda_0} = \begin{pmatrix} \lambda^{\Lambda_0}_{\beta_1} \\ \lambda^{\Lambda_0}_{\beta_2} \\ \lambda^{\Lambda_0}_{\delta} \end{pmatrix}.$$

*Under Assumptions 1-7 and $\mathcal{H}_0$,*

$$LM_{ECCC} \overset{w}{\to} \frac{1}{2} \|\zeta\|^2_{(K_1 J^{-1} K_1')}, \tag{C.1}$$

*where $\zeta := G_{\beta_1} - J_{\beta_1\delta} J_{\delta\delta}^{-1} G_\delta + (J_{\beta_1\beta_2} - J_{\beta_1\delta} J_{\delta\delta}^{-1} J_{\delta\beta_2}) \lambda^{\Lambda_0}_{\beta_2}$.*
*Suppose in addition that $\tilde{s}_1 = s_1$ and that $\Sigma$ is positive definite. Then*

$$LM_{ECCC} \overset{w}{\to} \sum_{i=1}^{m} \xi_i \chi^2_{m_i}, \tag{C.2}$$

*where $\xi_i$, $i = 1, ..., m$, are the $m$ distinct eigenvalues of $(1/2)\Omega^{1/2}(K_1 J^{-1} K_1')\Omega^{1/2}$, with $\Omega^{1/2}$ the positive definite matrix square root of the $(s_1 \times s_1)$ matrix $\Omega$, given by*

$$\Omega = \Sigma_{\beta\beta} - J_{\beta\delta} J_{\delta\delta}^{-1} \Sigma_{\delta\beta} - \Sigma_{\beta\delta} J_{\delta\delta}^{-1} J_{\delta\beta} + J_{\delta\beta} J_{\delta\delta}^{-1} \Sigma_{\beta\beta} J_{\delta\delta}^{-1} J_{\delta\beta},$$

*and $\chi^2_{m_i}$, $i = 1, ..., m$ are mutually independent, and $m_i$ is the multiplicity of $\xi_i$.*
*Finally, suppose furthermore that $\mathcal{L}(\eta_t) = N(0, I_d)$, then*

$$LM_{ECCC} \overset{w}{\to} \chi^2_{s_1}. \tag{C.3}$$

*Proof.* Similar to the derivations in proof of Theorem 4.1 we obtain from a Taylor-type expansion,

$$\begin{aligned} \sqrt{T} K_1 \hat{S}_T(\tilde{\theta}_T) &= \sqrt{T} \frac{\hat{L}_T(\theta_0)}{\partial \beta_1} + \frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \beta_1 \partial \theta'} \sqrt{T}(\tilde{\theta}_T - \theta_0) \\ &= \sqrt{T} \frac{\hat{L}_T(\theta_0)}{\partial \beta_1} + \frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \beta_1 \partial \beta_2'} \sqrt{T}(\tilde{\beta}_{2,T} - \beta_{2,0}) + \frac{\partial^2 \hat{L}_T(\theta^\star)}{\partial \beta_1 \partial \delta'} \sqrt{T}(\tilde{\delta}_T - \delta_0) \end{aligned}$$

where $\theta^\star$ is between $\tilde{\theta}_T$ and $\theta_0$, and where the second equality follows from the fact that $\tilde{\beta}_{1,T} - \beta_{1,0} = 0_{\tilde{s}_1 \times 1}$. Since $\delta_0$ does not attain the bounds of $\Theta$, we have by a Taylor-type expansion that

$$
\begin{aligned}
0_{s_2\times 1} &= \frac{\hat{L}_T(\theta_0)}{\partial\delta} + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\theta'}(\tilde{\theta}_T-\theta_0) \\
&= \frac{\hat{L}_T(\theta_0)}{\partial\delta} + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\beta_2'}(\tilde{\beta}_{2,T}-\beta_{2,0}) + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}(\tilde{\delta}_T-\delta_0), \qquad \text{(C.4)}
\end{aligned}
$$

where $\theta^{\star\star}$ is between $\tilde{\theta}_T$ and $\theta_0$. Hence

$$
(\tilde{\delta}_T-\delta_0) = -\left(\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}\right)^{-1}\left[\frac{\hat{L}_T(\theta_0)}{\partial\delta} + \frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\beta_2'}(\tilde{\beta}_{2,T}-\beta_{2,0})\right], \qquad \text{(C.5)}
$$

and substituting (C.5) into (C.4) and rearranging yield

$$
\begin{aligned}
\sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{\hat{L}_T(\theta_0)}{\partial\beta_1} - \frac{\partial^2\hat{L}_T(\theta^{\star})}{\partial\beta_1\partial\delta'}\left(\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}\right)^{-1}\sqrt{T}\frac{\hat{L}_T(\theta_0)}{\partial\delta} \\
&\quad + \left[\frac{\partial^2\hat{L}_T(\theta^{\star})}{\partial\beta_1\partial\beta_2'} - \frac{\partial^2\hat{L}_T(\theta^{\star})}{\partial\beta_1\partial\delta'}\left(\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\delta'}\right)^{-1}\frac{\partial^2\hat{L}_T(\theta^{\star\star})}{\partial\delta\partial\beta_2'}\right]\sqrt{T}(\tilde{\beta}_{2,T}-\beta_{2,0}).
\end{aligned}
$$

From Lemma B.5, using that $\sqrt{T}(\tilde{\theta}_T-\theta_0)$ and $\sqrt{T}S_T(\theta_0)$ are $O_p(1)$,

$$
\begin{aligned}
\sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{L_T(\theta_0)}{\partial\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}\sqrt{T}\frac{L_T(\theta_0)}{\partial\delta} \\
&\quad + \left[J_{\beta_1\beta_2} - J_{\beta_1\delta}J_{\delta\delta}^{-1}J_{\delta\beta_2}\right]\sqrt{T}(\tilde{\beta}_{2,T}-\beta_{2,0}) + o_p(1) \\
&\overset{w}{\to} G_{\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}G_\delta + (J_{\beta_1\beta_2} - J_{\beta_1\delta}J_{\delta\delta}^{-1}J_{\delta\beta_2})\lambda_{\beta_2}^{\Lambda_0}, \qquad \text{(C.6)}
\end{aligned}
$$

where we have used that the terms converge jointly due to point 1. of the proof of Theorem 3.1, and that $\sqrt{T}(\tilde{\theta}_T-\theta_0) \overset{w}{\to} \lambda^{\Lambda_0}$. Moreover, Lemma B.5 and the consistency of $\tilde{\theta}_T$ imply that

$$
K_1\hat{J}_T(\tilde{\theta}_T)^{-1}K_1' = K_1J^{-1}K_1' + o_p(1). \qquad \text{(C.7)}
$$

Hence by combining (C.6) and (C.7) and by applying the continuous mapping theorem, we have shown that (C.1) holds.
Next, for the case $\tilde{s}_1 = s_1$, we have that $\beta_2$ vanishes such that

$$
\begin{aligned}
\sqrt{T}K_1\hat{S}_T(\tilde{\theta}_T) &= \sqrt{T}\frac{L_T(\theta_0)}{\partial\beta_1} - J_{\beta_1\delta}J_{\delta\delta}^{-1}\sqrt{T}\frac{L_T(\theta_0)}{\partial\delta} + o_p(1) \\
&= (I_{s_1}, -J_{\beta_1\delta}J_{\delta\delta}^{-1})\sqrt{T}\frac{L_T(\theta_0)}{\partial\theta} + o_p(1) \\
&\overset{w}{\to} (I_{s_1}, -J_{\beta_1\delta}J_{\delta\delta}^{-1})G, \qquad \text{(C.8)}
\end{aligned}
$$

where we have used arguments similar to the ones given above. It holds that

$$
\mathcal{L}[(I_{s_1}, -J_{\beta_1\delta}J_{\delta\delta}^{-1})G] = N(0,\Omega). \qquad \text{(C.9)}
$$

Combining (C.7)-(C.9) and using White (1996, Theorem 8.6), we conclude that (C.2) holds. In the case where $\tilde{s}_1 = s_1$ and $\mathcal{L}(\eta_t) = N(0, I_d)$, the information equality implies that $2\Sigma = J$ and it is straightforward, using (B.23) and the continuous mapping theorem, to establish that (C.3) holds. $\square$

# D Additional details about the simulations

This section contains some additional details about the simulations reported in Section 5.

- The simulations are carried out in OxMetrics 7.0.
- All replications are based on a burn-in period of 1,000 observations, and all simulations are based on the same seed value.
- The computation of the QMLE $\hat{\theta}_T$ and the constrained QMLE $\tilde{\theta}_T$ is based on maximization of the log-likelihood function according to the `MaxSQP` function.
  For the computation of $\hat{\theta}_T$ we use the starting values:

$$\kappa = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.10 & 0.05 \\ 0.05 & 0.11 \end{bmatrix}, \quad B = \begin{bmatrix} 0.85 & 0.05 \\ 0.05 & 0.80 \end{bmatrix}, \quad \rho = 0.5.$$

  For the computation of $\tilde{\theta}_T$ we use the starting values:

$$\kappa = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.10 & \\ & 0.11 \end{bmatrix}, \quad B = \begin{bmatrix} 0.85 & \\ & 0.80 \end{bmatrix}, \quad \rho = 0.5.$$

- For the computation of the log-likelihood function, we use as initial value $\hat{h}_0 = T^{-1} \sum_{t=1}^{T} X_t^{\odot 2}$.
- The following constraints are imposed on the parameters for the optimization: With $\kappa = (\kappa_1, \kappa_2)'$, $\kappa_1, \kappa_2 \geq 0.000001$, $\rho \in [-0.99999, 0.99999]$, and all elements of the matrices $A$ and $B$ are nonnegative.
- All derivatives of the log-likelihood function are obtained by numerical techniques.
- If a replication yields an estimate $\hat{J}_T(\hat{\theta}_T)$ or $\hat{J}_T(\tilde{\theta}_T)$ that is found to be (numerically) singular, this replication is discarded from the calculations. The singularity of the matrices was mainly an issue for the replications with $T = 1{,}000$ observations.

## References

Andrews, D. W. K. (1998): "Estimation when a parameter is on a boundary: Part II," Yale University, unpublished manscript.

——— (1999): "Estimation when a parameter is on a boundary," *Econometrica*, 67, 1341–1383.

——— (2001): "Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 69, 683–734.

Bali, T. G. and A. Hovakimian (2009): "Volatility spreads and expected stock returns," *Management Science*, 55, 1797–1812.

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model," *The Review of Economics and Statistics*, 72, 498–505.

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

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

Brown, B. M. (1971): "Martingale Central Limit Theorems," *The Annals of Mathematical Statistics*, 42, 59–66.

Chernoff, H. (1954): "On the distribution of the likelihood ratio," *The Annals of Mathematical Statistics*, 25, 573–578.

Comte, F. and O. Lieberman (2000): "Second-order noncausality in multivariate GARCH processes," *Journal of Time Series Analysis*, 21, 535–557.

Conrad, C. and M. Karanasos (2010): "Negative volatility spillovers in the unrestricted ECCC-GARCH model," *Econometric Theory*, 26, 838–862.

Conrad, C. and E. Weber (2013): "Measuring persistence in volatility spillovers," Discussion Paper 543, Department of Economics, University of Heidelberg.

Conrad, J., M. N. Gultekin, and G. Kaul (1991): "Asymmetric predictability of conditional variances," *The Review of Financial Studies*, 4, 597–622.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Francq, C. and J.-M. Zakoïan (2007): "Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero," *Stochastic Processes and their Applications*, 117, 1265–1284.

——— (2009): "Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons," *Journal of the American Statistical Association*, 104, 313–324.

——— (2010): *GARCH Models: Structure, Statistical Inference and Financial Applications*, Chichester: Wiley.

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

Granger, C. W. J. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 37, 424–438.

——— (1980): "Testing for causality: A personal viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.

Granger, C. W. J., R. P. Robins, and R. F. Engle (1986): *Wholesale and retail prices: Bivariate time-series modeling with forecastable error variances*, Cambridge, MA: The MIT Press, 1–16.

Hafner, C. M. and H. Herwartz (2008): "Testing for causality in variance using multivariate GARCH models," *Annales d'Économie et de Statistique*, 215–241.

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

He, C. and T. Teräsvirta (2004): "An extended constant conditional correlation GARCH model and its fourth-moment structure," *Econometric Theory*, 20, 904–926.

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

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

Ketz, P. (2014): "Testing near or at the boundary of the parameter space," Unpublished manuscript, Department of Economics, Brown University.

Ling, S. and M. McAleer (2002): "Necessary and sufficient moment conditions for the GARCH$(r, s)$ and Asymmetric Power GARCH$(r, s)$ models," *Econometric Theory*, 18, 722–729.

——— (2003): "Asymptotic theory for a vector ARMA-GARCH model," *Econometric Theory*, 19, 280–310.

——— (2010): "A general asymptotic theory for time-series models," *Statistica Neerlandica*, 64, 97–111.

Meitz, M. and P. Saikkonen (2008): "Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models," *Econometric Theory*, 24, 1291–1320.

Meyn, S. and R. L. Tweedie (2009): *Markov Chains and Stochastic Stability*, Cambridge: Cambridge University Press, 2 ed.

Nakatani, T. and T. Teräsvirta (2008): "Positivity constraints on the conditional variances in the family of conditional correlation GARCH models," *Finance Research Letters*, 5, 88–95.

——— (2009): "Testing for volatility interactions in the constant conditional correlation GARCH model," *The Econometrics Journal*, 12, 147–163.

Pedersen, R. S. (2015): "Targeting estimation of CCC-GARCH models with infinite fourth moments," Forthcoming in *Econometric Theory*, https://sites.google.com/site/rspecon/research.

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

Perlman, M. D. (1969): "One-sided testing problems in multivariate analysis," *The Annals of Mathematical Statistics*, 40, 549–567, Corrections in *The Annals of Mathematical Statistics* (1971), 42, 1777.

Ranga Rao, R. (1962): "Relations between weak and uniform convergence of measures with applications," *The Annals of Mathematical Statistics*, 33, 659–680.

Ross, S. A. (1989): "Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy," *The Journal of Finance*, 44, 1–17.

Tjøstheim, D. (1990): "Non-linear time series and Markov chains," *Advances in Applied Probability*, 22, 587–611.

White, H. (1996): *Estimation, Inference and Specification Analysis*, Econometric Society Monographs, New York, NY: Cambridge University Press.

Woźniak, T. (2015): "Testing causality between two vectors in multivariate GARCH models," *International Journal of Forecasting*, 31, 876–894.

Zarantonello, E. H. (1971): "Projections on convex sets in Hilbert spaces and spectral theory," in *Contributions to Nonlinear Functional Analysis*, ed. by E. H. Zarantonello, New York, NY: Academic Press, 237–341.