QML estimation of a class of multivariate GARCH models without moment conditions on the observed process

Francq, Christian and Zakoian, Jean-Michel
Université Lille 3 GREMARS-EQUIPPE, CREST

February 2010

QML ESTIMATION OF A CLASS OF MULTIVARIATE GARCH MODELS WITHOUT MOMENT CONDITIONS ON THE OBSERVED PROCESS

By Christian Francq

University Lille 3, EQUIPPE-GREMARS

and

By Jean-Michel Zakoïan

CREST and University Lille 3

We establish the strong consistency and asymptotic normality of the quasi-maximum likelihood estimator of the parameters of a class of multivariate GARCH processes. The conditions are mild and coincide with the minimal ones in the univariate case. In particular, contrary to the current literature on the estimation of multivariate GARCH models, no moment assumption is made on the observed process. Instead, we require strict stationarity, for which a necessary and sufficient condition is established.

1. Introduction. Since the inception of the univariate ARCH and GARCH models by Engle (1982) and Bollerslev (1990), a wide variety of multivariate extensions have been proposed. Recent reviews on the rapidly changing literature on multivariate GARCH models are Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009).

Although the asymptotic theory for multivariate GARCH has been less investigated than for univariate models, several papers have established asymptotic results for different specifications. Jeantheau (1998) gave general conditions for the strong consistency of the QMLE for multivariate GARCH models. Comte and Lieberman (2003) showed the consistency and the asymptotic normality of the Quasi Maximum Likelihood Estimator (QMLE) for the BEKK formulation. Asymptotic results were established by Ling and McAleer (2003) for the CCC formulation of an ARMA-GARCH, by Hafner and Preminger (2009a) for the Vec model.

In all these references, moment assumptions are made on the observed process. Given that the existence of such moments is doubtful for many financial series, such conditions can be restrictive. To our knowledge, consistency and asymptotic normality results for multivariate GARCH without moments restriction have only been established by Hafner and Preminger (2009b), for a factor model of the form FF-GARCH. However, their model is a first-order model (it reduces to the standard GARCH(1,1) when the dimension is one). For univariate GARCH(p, q), it took almost twenty years to reach minimal assumptions for the strong consistency (SC) and the asymptotic normality (AN) of the QMLE. The most significant breakthrough in this direction was the paper by Berkes, Horváth and Kokoszka (2003), although slightly weaker conditions can be found in Francq and Zakoian (2004).

The main contribution of this article is to provide asymptotic results for the Constant Conditional Correlation (CCC) GARCH(p, q) under conditions which parallel those used in the univariate setting. The CCC-GARCH(p, q), introduced by Bollerslev (1990) and generalized by Jeantheau (1998), is undoubtedly one of the most popular multivariate GARCH models. The attractiveness of this class follows from its tractability: i) the number of unknown coefficients is less than in other specifications; ii) the conditions ensuring definite positiveness of the conditional variance are simple and explicit. Moreover, as we will see, the conditions ensuring the existence of strictly stationary solutions are explicit. Of course, more sophisticated classes of models can be seen as more realistic. This is in particular the case of the Dynamic Conditional Correlation (DCC) model introduced by Engle (2002), and studied by Engle and Sheppard (2001) and Nakatani and Teräsvirta (2009), among others. For such models, however, establishing a sound asymptotic theory of estimation seems a formidable task. We view the results of this paper as a first step in this direction.

An outline of the paper can be given as follows. In Section 2, we discuss the model assumptions and establish the strict stationarity condition. In Section 3 our main results concerning the asymptotic properties of the QMLE are stated. Proofs are relegated to Section 4,

2. Model and strict stationarity condition. Let (ϵ_t) denote a vector process with dimension $m \times 1$. The process (ϵ_t) is called a CCC-GARCH(p,q) if it verifies

$$\begin{cases} \epsilon_t &= H_t^{1/2}\eta_t, \\ H_t &= D_t R D_t, \quad D_t^2 = \mathrm{diag}(\underline{h}_t) \\ \underline{h}_t &= \underline{\omega} + \sum_{i=1}^{q} \mathbf{A}_i \underline{\epsilon}_{t-i} + \sum_{j=1}^{p} \mathbf{B}_j \underline{h}_{t-j}, \quad \underline{\epsilon}_t = \left(\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2\right)' \end{cases} \tag{2.1}$$

where R is a correlation matrix, $\underline{\omega}$ is a vector of size $m \times 1$ with strictly positive coefficients, the $\mathbf{A}_i$ and $\mathbf{B}_j$ are matrices of size $m \times m$ with positive coefficients, and (η_t) is an iid sequence of centered variables on $\mathbb{R}^m$ with identity covariance matrix.

The CCC model was introduced by Bollerslev (1990) in a simplest version, assuming that the matrices $\mathbf{A}_i$ and $\mathbf{B}_j$ are diagonal. By contrast, in (2.1) the conditional variance $h_{kk,t}$ of the k-th component of ϵ_t depends not only on its past values but also on the past values of the other components. For this reason, Model (2.1) is referred to as the *Extended* CCC model by He and Teräsvirta (2004).

In the latter reference, a sufficient condition for second-order and strict stationarity of a CCC-GARCH(1,1) is given. A sufficient condition for strict stationarity and the existence of fourth-order moments of the CCC-GARCH(p,q) is established in Aue, Hörmann, Horváth, and Reimherr (2009). Our first result provides a necessary and sufficient strict stationarity condition for the same model.

Write

$$\epsilon_t = D_t \tilde{\eta}_t, \quad \text{where} \quad \tilde{\eta}_t = R^{1/2}\eta_t \tag{2.2}$$

is a centered vector with covariance matrix R. Thus

$$\underline{\epsilon}_t = \Upsilon_t \underline{h}_t, \quad \text{where} \quad \Upsilon_t = \begin{pmatrix} \tilde{\eta}_{1t}^2 & 0 & \dots & 0 \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & & \tilde{\eta}_{mt}^2 \end{pmatrix}.$$

Let the $(p+q)m \times (p+q)m$ matrix

$$C_t = \begin{pmatrix} \Upsilon_t\mathbf{A}_1 & & \cdots & & \Upsilon_t\mathbf{A}_q & \Upsilon_t\mathbf{B}_1 & & \cdots & & \Upsilon_t\mathbf{B}_p \\ I_m & 0 & \cdots & & 0 & 0 & & \cdots & & 0 \\ 0 & I_m & \cdots & & 0 & 0 & & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \dots & I_m & 0 & 0 & & \dots & 0 & 0 \\ \mathbf{A}_1 & & \cdots & & \mathbf{A}_q & \mathbf{B}_1 & & \cdots & & \mathbf{B}_p \\ 0 & & \cdots & & 0 & I_m & 0 & \cdots & & 0 \\ 0 & & \cdots & & 0 & 0 & I_m & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \dots & 0 & 0 & 0 & & \dots & I_m & 0 \end{pmatrix} \quad (2.3)$$

We are now in a position to state the following result.

THEOREM 2.1. *A necessary and sufficient condition for the existence of a strictly stationary and non anticipative solution process to Model (2.1) is* $\gamma(\mathbf{C_0}) < 0$, *where* $\gamma(\mathbf{C_0})$ *is the top Lyapunov exponent of the sequence* $\mathbf{C_0} = \{C_t, t \in \mathbb{Z}\}$ *defined in (2.3). This stationary and non anticipative solution, when* $\gamma(\mathbf{C_0}) < 0$, *is unique and ergodic.*

The following result provides a necessary strict stationarity condition which is simple to check. Denote by $\det(A)$ or $|A|$ the determinant of a square matrix A.

COROLLARY 2.1. *Let the matrix polynomial defined by:* $\mathcal{B}(z) = I_m - z\mathbf{B}_1 - \ldots - z^p\mathbf{B}_p, \quad z \in \mathbb{C}$. *Let*

$$\mathbb{B} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & & \mathbf{B}_p \\ I_m & 0 & \cdots & & 0 \\ 0 & I_m & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & I_m & 0 \end{pmatrix}.$$

Then, if $\gamma(\mathbf{C_0}) < 0$ *the following equivalent properties hold:*

1. *The roots of* $\det \mathcal{B}(z)$ *are outside the unit disk,*
2. $\rho(\mathbb{B}) < 1$.

The following result will be extremely useful to prove the CAN of the QMLE under minimal conditions.

COROLLARY 2.2. *Suppose* $\gamma(\mathbf{C_0}) < 0$. *Let* ϵ_t *be the strictly stationary and non anticipative solution of Model (2.1). There exists* $s > 0$ *such that* $E\|\underline{h}_t\|^s < \infty$ *and* $E\|\epsilon_t\|^{2s} < \infty$.

3. QML estimation. The parameters consist of the coefficients of the matrices $\underline{\omega}, \mathbf{A}_i$ and $\mathbf{B}_j$, and the coefficients of the lower triangular part (excluding the diagonal) of the correlation matrix $R = (\rho_{ij})$. The number of unknown parameters is thus

$$s_0 = m + m^2(p+q) + \frac{m(m-1)}{2}.$$

The parameter vector is denoted

$$\theta = (\theta_1, \ldots, \theta_{s_0})' = (\underline{\omega}', \alpha_1', \ldots, \alpha_q', \beta_1', \ldots, \beta_p', \rho')' := (\underline{\omega}', \alpha', \beta', \rho')',$$

where $\rho' = (\rho_{21}, \ldots, \rho_{m1}, \rho_{32}, \ldots, \rho_{m2}, \ldots, \rho_{m,m-1})$, $\alpha_i = \operatorname{vec}(\mathbf{A}_i)$, $i = 1, \ldots, q$, and $\beta_j = \operatorname{vec}(\mathbf{B}_j)$, $j = 1, \ldots, p$. The parameter space is a sub-space Θ of

$$]0, +\infty[^m \times [0, \infty[^{m^2(p+q)} \times]-1, 1[^{m(m-1)/2}.$$

The true parameter valued is denoted

$$\theta_0 = (\underline{\omega}_0', \alpha_{01}', \dots, \alpha_{0q}', \beta_{01}', \dots, \beta_{0p}', \rho_0')' = (\underline{\omega}_0', \alpha_0', \beta_0', \rho_0')'.$$

Before detailing the estimation procedure and its properties, we discuss conditions to impose on the matrices $\mathbf{A}_i$ and $\mathbf{B}_j$ in order to ensure the uniqueness of the parameterization.

3.1. *Identifiability Conditions.* Let $\mathcal{A}_\theta(z) = \sum_{i=1}^q \mathbf{A}_i z^i$ and $\mathcal{B}_\theta(z) = I_m - \sum_{j=1}^p \mathbf{B}_j z^j$. By convention, $\mathcal{A}_\theta(z) = 0$ if $q = 0$ and $\mathcal{B}_\theta(z) = I_m$ if $p = 0$.

If the roots of $\det(\mathcal{B}_\theta(z)) = 0$ are outside the unit disk, we deduce from $\mathcal{B}_\theta(B)\underline{h}_t = \underline{\omega} + \mathcal{A}_\theta(B)\underline{\epsilon}_t$ the representation

$$\underline{h}_t = \mathcal{B}_\theta(1)^{-1}\underline{\omega} + \mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta(B)\underline{\epsilon}_t. \quad (3.1)$$

In the vector case, assuming that the polynomials $\mathcal{A}_{\theta_0}$ and $\mathcal{B}_{\theta_0}$ have no common root does not suffice to ensure that there exists no other pair $(\mathcal{A}_\theta, \mathcal{B}_\theta)$, with the same degrees (p, q), such that

$$\mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta(B) = \mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}(B). \quad (3.2)$$

This condition is equivalent to the existence of an operator $U(B)$ such that

$$\mathcal{A}_\theta(B) = U(B)\mathcal{A}_{\theta_0}(B) \quad \text{and} \quad \mathcal{B}_\theta(B) = U(B)\mathcal{B}_{\theta_0}(B),$$

this common factor vanishing in $\mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta(B)$

The polynomial $U(B)$ is called *unimodular* if $\det\{U(B)\}$ is a non-zero constant. When the only common factors of the polynomials $P(B)$ and $Q(B)$ are unimodular, that is when

$$P(B) = U(B)P_1(B), \quad Q(B) = U(B)Q_1(B) \Longrightarrow \det\{U(B)\} = \text{ cst },$$

$P(B)$ and $Q(B)$ are called *left coprime.*

The following example shows that, in the vector case, assuming that $\mathcal{A}_{\theta_0}(B)$ and $\mathcal{B}_{\theta_0}(B)$ are left coprime is not sufficient to ensure that (3.2) has no solution $\theta \neq \theta_0$

(in the univariate case this is sufficient because the condition $\mathcal{B}_\theta(0) = \mathcal{B}_{\theta_0}(0) = 1$ imposes $U(B) = U(0) = 1$).

Example 3.1 (Non identifiable bivariate model). For $m = 2$, let

$$\mathcal{A}_{\theta_0}(B) = \begin{pmatrix} a_{11}(B) & a_{12}(B) \\ a_{21}(B) & a_{22}(B) \end{pmatrix}, \quad \mathcal{B}_{\theta_0}(B) = \begin{pmatrix} b_{11}(B) & b_{12}(B) \\ b_{21}(B) & b_{22}(B) \end{pmatrix},$$

$$U(B) = \begin{pmatrix} 1 & 0 \\ B & 1 \end{pmatrix}$$

with $\deg(a_{21}) = \deg(a_{22}) = q$, $\deg(a_{11}) < q$, $\deg(a_{12}) < q$ and $\deg(b_{21}) = \deg(b_{22}) = p$, $\deg(b_{11}) < p$, $\deg(b_{12}) < p$. The polynomial $\mathcal{A}(B) = U(B)\mathcal{A}_{\theta_0}(B)$ has the same degree q as $\mathcal{A}_{\theta_0}(B)$, and $\mathcal{B}(B) = U(B)\mathcal{B}_{\theta_0}(B)$ is a polynomial of the same degree p as $\mathcal{B}_{\theta_0}(B)$. On the other hand, $U(B)$ has a non-zero determinant which is independent of B, hence is it unimodular. Moreover $\mathcal{B}(0) = \mathcal{B}_{\theta_0}(0) = I_m$ and $\mathcal{A}(0) = \mathcal{A}_{\theta_0}(0) = 0$. It is thus possible to find θ such that $\mathcal{B}(B) = \mathcal{B}_\theta(B), \mathcal{A}(B) = \mathcal{A}_\theta(B)$ and $\underline{\omega} = U(1)\underline{\omega_0}$. The model is thus non identifiable, θ and θ_0 corresponding to the same representation (3.1).

Identifiability can be insured by several types of conditions (see for instance Reinsel, 1997, p. 37-40). To obtain a mild condition define, for any column i of the matrix operators $\mathcal{A}_\theta(B)$ and $\mathcal{B}_\theta(B)$, the maximal degrees $q_i(\theta)$ and $p_i(\theta)$, respectively. Suppose that these maximal values are imposed for these orders, that is

$$\forall \theta \in \Theta, \ \forall i = 1, \dots, m, \quad q_i(\theta) \le q_i \quad \text{and} \quad p_i(\theta) \le p_i \tag{3.3}$$

where $q_i \le q$ and $p_i \le p$ are fixed integers. Denote by $a_{q_i}(i)$ (resp. $b_{p_i}(i)$) the column vector of the coefficients of B^{q_i} (resp. B^{p_i}) in the i^{th} column of $\mathcal{A}_{\theta_0}(B)$ (resp. $\mathcal{B}_{\theta_0}(B)$).

Example 3.2 (Illustration of the notations on an example). For

$$\mathcal{A}_{\theta_0}(B) = \begin{pmatrix} 1 + a_{11}B^2 & a_{12}B \\ a_{21}B^2 + a_{21}^*B & 1 + a_{22}B \end{pmatrix}, \quad \mathcal{B}_{\theta_0}(B) = \begin{pmatrix} 1 + b_{11}B^4 & b_{12}B \\ b_{21}B^4 & 1 + b_{22}B \end{pmatrix},$$

with $a_{11}a_{21}a_{12}a_{22}b_{11}b_{21}b_{12}b_{22} \neq 0$, we have

$$q_1(\theta_0) = 2, \quad q_2(\theta_0) = 1, \quad p_1(\theta_0) = 4, \quad p_2(\theta_0) = 1$$

and

$$a_2(1) = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \quad a_1(2) = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}, \quad b_4(1) = \begin{pmatrix} b_{11} \\ b_{21} \end{pmatrix}, \quad b_1(2) = \begin{pmatrix} b_{12} \\ b_{22} \end{pmatrix}.$$

PROPOSITION 3.1 (A simple identifiability condition). *If the matrix*

$$M(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = [a_{q_1}(1) \cdots a_{q_m}(m) \;\; b_{p_1}(1) \cdots b_{p_m}(m)] \tag{3.4}$$

has full rank m, the parameters α_0 and β_0 are identified by the constraints (3.3) with $q_i = q_i(\theta_0)$ and $p_i = p_i(\theta_0)$ for any value of i.

Proof. Indeed, let $U(B) = U_0 + U_1B + \ldots + U_kB^k$. Since the term of highest degree (column by column) of $\mathcal{A}_{\theta_0}(B)$ is $[a_{q_1}(1)B^{q_1} \cdots a_{q_m}(m)B^{q_m}]$, the ith column of $\mathcal{A}_\theta(B) = U(B)\mathcal{A}_{\theta_0}(B)$ is a polynomial in B of degree less than q_i if and only if $U_j a_{q_i}(i) = 0$, for $j = 1, \ldots, k$. Similarly we must have $U_j b_{p_i}(i) = 0$, for $j = 1, \ldots, k$ and $i = 1, \ldots m$. It follows that $U_j M(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = 0$, which implies $U_j = 0$ for $j = 1, \ldots, k$ thanks to Condition (3.4). Consequently $U(B) = U_0$ and, since for all θ $\mathcal{B}_\theta(0) = I_m$, we have $U(B) = I_m$. □

EXAMPLE 3.3 (Illustration of the identifiability condition). In example 3.1,

$$M(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = [a_q(1)a_q(2)b_p(1)b_p(2)] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \times & \times & \times & \times \end{bmatrix}$$

is not a full-rank matrix. Hence, the identifiability condition of Proposition 3.1 is not satisfied. Indeed, the model is not identifiable.

A simpler, but more restrictive, condition is obtained by imposing that

$$M_1(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = [\mathbf{A}_q \;\; \mathbf{B}_p]$$

has full rank m. This entails uniqueness under the constraint that the degrees of $\mathcal{A}_\theta$ and $\mathcal{B}_\theta$ are less than p and q, respectively.

EXAMPLE 3.4 (Another illustration of the identifiability condition). Turning again to Example 3.2 with $a_{12}b_{21} = a_{22}b_{11}$ and, for instance, $a_{21} = 0$ and $a_{22} \neq 0$, observe that the matrix

$$M_1(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = \begin{bmatrix} 0 & a_{12} & b_{11} & 0 \\ 0 & a_{22} & b_{21} & 0 \end{bmatrix}$$

does not have full rank, but the matrix

$$M(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ 0 & a_{22} & b_{21} & b_{22} \end{bmatrix}$$

has full rank.

3.2. *Asymptotic Properties of the QML Estimator of the CCC-GARCH.* Let $(\epsilon_1, \dots, \epsilon_n)$ be an observation of length n of the unique non anticipative and strictly stationary solution (ϵ_t) of Model (2.1). Conditionally to nonnegative initial values $\epsilon_0, \dots, \epsilon_{1-q}, \underline{\tilde{h}}_0, \dots, \underline{\tilde{h}}_{1-p}$, the Gaussian quasi-likelihood writes

$$L_n(\theta) = L_n(\theta; \epsilon_1, \dots, \epsilon_n) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{m/2}|\tilde{H}_t|^{1/2}} \exp\left(-\frac{1}{2}\epsilon_t' \tilde{H}_t^{-1} \epsilon_t\right),$$

where the $\tilde{H}_t$ are recursively defined, for $t \geq 1$, by

$$\begin{cases} \tilde{H}_t &= \tilde{D}_t R \tilde{D}_t, \quad \tilde{D}_t = \{\mathrm{diag}(\underline{\tilde{h}}_t)\}^{1/2} \\ \underline{\tilde{h}}_t &= \underline{\tilde{h}}_t(\theta) = \underline{\omega} + \sum_{i=1}^{q} \mathbf{A}_i \underline{\epsilon}_{t-i} + \sum_{j=1}^{p} \mathbf{B}_j \underline{\tilde{h}}_{t-j} \end{cases}$$

A QML estimator of θ is defined as any measurable solution $\hat{\theta}_n$ of

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\min_{\theta \in \Theta} \tilde{\mathbf{l}}_n(\theta). \tag{3.5}$$

where

$$\tilde{\mathbf{l}}_n(\theta) = n^{-1} \sum_{t=1}^{n} \tilde{\ell}_t, \qquad \text{et} \quad \tilde{\ell}_t = \tilde{\ell}_t(\theta) = \epsilon_t' \tilde{H}_t^{-1} \epsilon_t + \log|\tilde{H}_t|.$$

The following assumptions will be used to establish the strong consistency of the QML estimator.

A1: $\theta_0 \in \Theta$ and Θ is compact.

A2: $\gamma(\mathbf{C_0}) < 0$ and $\forall \theta \in \Theta, \quad |\mathcal{B}_\theta(z)| = 0 \Rightarrow |z| > 1.$

A3: The components of η_t are independent and their squares have non degenerate distributions.

A4: If $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ are left coprime and $M_1(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0})$ has full rank m.

A5: R is a positive-definite correlation matrix for all $\theta \in \Theta$.

If the space Θ is constrained by (3.3), that is if maximal orders are imposed for each component of $\underline{\epsilon}_t$ and $\underline{h}_t$ in each equation, Assumption **A4** can be replaced by the more general condition:

A4': If $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ are left coprime and $M(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0})$ has full rank m.

It will be useful to approximate the sequence $(\tilde{\ell}_t(\theta))$ by an ergodic and stationary sequence. Assumption **A2** implies that there exists a strictly stationary, non anticipative and ergodic solution $(\underline{h}_t)_t = \{\underline{h}_t(\theta)\}_t$ of

$$\underline{h}_t = \underline{\omega} + \sum_{i=1}^{q} \mathbf{A}_i \underline{\epsilon}_{t-i} + \sum_{j=1}^{p} \mathbf{B}_j \underline{h}_{t-j}, \qquad \forall t. \tag{3.6}$$

Now, letting $D_t = \{\mathrm{diag}(\underline{h}_t)\}^{1/2}$ and $H_t = D_t R D_t$, we define

$$\mathbf{l}_n(\theta) = \mathbf{l}_n(\theta; \epsilon_n, \epsilon_{n-1} \ldots,) = n^{-1} \sum_{t=1}^{n} \ell_t, \quad \ell_t = \ell_t(\theta) = \epsilon_t' H_t^{-1} \epsilon_t + \log |H_t|.$$

We are now in a position to state the following consistency theorem.

THEOREM 3.1 (Strong consistency). *Let $(\hat{\theta}_n)$ a sequence of QML estimators satisfying (3.5). Then, under* **A1**-**A5** *(or* **A1**-**A4'**-**A5***),*

$$\hat{\theta}_n \to \theta_0, \quad \textit{almost surely when } n \to \infty.$$

To establish the asymptotic normality we require the following additional assumptions.

A6: $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ is the interior of Θ.

A7: $E\|\eta_t \eta_t'\|^2 < \infty$.

THEOREM 3.2 (Asymptotic normality). *Under the assumptions of Theorem 3.1 and* **A6-A7** $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *converges in distribution to* $\mathcal{N}(0, J^{-1}IJ^{-1})$, *where* J *is a positive-definite matrix and* I *is a semi positive-definite matrix, defined by*

$$I = E\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta}\frac{\partial \ell_t(\theta_0)}{\partial \theta'}\right), \qquad J = E\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'}\right).$$

It is worth noting that the conditions ensuring the CAN are mild. When $m = 1$, they reduce to the minimal ones in the univariate setting. In particular, no assumption is made concerning the existence of moments of the observed process.

4. Proofs.

4.1. *Proof of Theorem 2.1.* The proof is similar to that given by Bougerol and Picard (1992) for univariate GARCH(p, q) models. The variables η_t admitting a variance, the condition $E \log^+ \|C_t\| < \infty$ is satisfied.

It follows that when $\gamma(\mathbf{C_0}) < 0$ the series

$$\underline{\tilde{z}}_t = \underline{b}_t + \sum_{n=0}^{\infty} C_t C_{t-1} \dots C_{t-n} \underline{b}_{t-n-1} \tag{4.1}$$

converges almost surely for all t. A strictly stationary solution to model (2.1) is obtained as $\epsilon_t = \{\mathrm{diag}(\underline{\tilde{z}}_{q+1,t})\}^{1/2} R^{1/2} \eta_t$ where $\underline{\tilde{z}}_{q+1,t}$ denotes the $(q+1)$th sub-vector of size m of $\underline{\tilde{z}}_t$. This solution is thus non anticipative and ergodic. The proof of the uniqueness is exactly the same as in the univariate case.

The proof of the necessary part can also be easily adapted. From Bougerol and Picard (1992) Lemma 3.4, it is sufficient to prove that $\lim_{t\to\infty} \|C_0 \dots C_{-t}\| = 0$. It suffices to show that, for $1 \le i \le p + q$

$$\lim_{t\to\infty} C_0 \dots C_{-t} \underline{e}_i = 0, \quad \text{a.s.} \tag{4.2}$$

where $\underline{e}_i = e_i \otimes I_m$ and e_i is the ith element of the canonical base of $\mathbb{R}^{p+q}$, since any vector x of $\mathbb{R}^{m(p+q)}$ can be decomposed, in a unique way, as $x = \sum_{i=1}^{p+q} \underline{e}_i x_i$ where $x_i \in \mathbb{R}^m$. As in the univariate case, the existence of a strictly stationary solution implies that $C_0 \dots C_{-k}\underline{b}_{-k-1}$ tends to 0, almost surely, as $k \to \infty$. It follows that, using the relation $\underline{b}_{-k-1} = \underline{e}_1 \Upsilon_{-k-1}\underline{\omega} + \underline{e}_{q+1}\underline{\omega}$, we have

$$\lim_{k\to\infty} C_0 \dots C_{-k}\underline{e}_1 \Upsilon_{-k-1}\underline{\omega} = 0, \quad \lim_{k\to\infty} C_0 \dots C_{-k}\underline{e}_{q+1}\underline{\omega} = 0, \quad \text{a.s.} \tag{4.3}$$

Since the components of $\underline{\omega}$ are strictly positive, (4.2) thus holds for $i = q+1$. Using

$$C_{-k}\underline{e}_{q+i} = \Upsilon_{-k}\mathbf{B}_i\underline{e}_1 + \mathbf{B}_i\underline{e}_{q+1} + \underline{e}_{q+i+1}, \quad i = 1, \dots, p \tag{4.4}$$

with by convention $\underline{e}_{p+q+1} = 0$, for $i = 1$ we obtain

$$0 = \lim_{t\to\infty} C_0 \dots C_{-k}\underline{e}_{q+1} \geq \lim_{k\to\infty} C_0 \dots C_{-k+1}\underline{e}_{q+2} \geq 0,$$

where the inequalities are taken componentwise. Therefore, (4.2) holds true for $i = q+2$, and by induction, for $i = q+j,\ j = 1, \dots, p$ in view of (4.4). Moreover, since $C_{-k}\underline{e}_q = \Upsilon_{-k}\mathbf{A}_q\underline{e}_1 + \mathbf{A}_q\underline{e}_{q+1}$, (4.2) holds for $i = q$. We conclude for the other values of i using an ascendent recursion, as in the univariate case. □

4.2. *Proof of Corollary 2.1.* Because all the entries of the matrices C_t are positive, it is clear that $\gamma(\mathbf{C_0})$ is larger than the top Lyapunov exponent of the sequence (C_t^*) obtained by replacing the matrices $\mathbf{A}_i$ by 0 in C_t. It is easily seen that the top Lyapunov coefficient of (C_t^*) coincides with that of the constant sequence equal to $\mathbb{B}$, that is with $\rho(\mathbb{B})$. It follows that $\gamma(\mathbf{C_0}) \geq \log \rho(\mathbb{B})$. Hence $\gamma(\mathbf{C_0}) < 0$ entails that all the eigenvalues of $\mathbb{B}$ are outside the unit disk. Finally, the equivalence between the two properties follows from

$$\begin{aligned} \det(\mathbb{B} - \lambda I_{mp}) &= (-1)^{mp} \det\left\{\lambda^p I_m - \lambda^{p-1}\mathbf{B}_1 - \dots - \lambda\mathbf{B}_{p-1} - \mathbf{B}_p\right\} \\ &= (-\lambda)^{mp} \det \mathcal{B}(\frac{1}{\lambda}), \qquad \lambda \neq 0. \end{aligned}$$

□

4.3. *Proof of Corollary 2.2.* It follows from the proof of Lemma 2.3 in Berkes, Horváth and Kokoszka (2003), that the strictly stationary solution defined by (4.1) satisfies $E\|\underline{\tilde{z}}_t\|^s < \infty$ for some $s > 0$. The conclusion follows from: $\|\underline{\epsilon}_t\| \leq \|\underline{\tilde{z}}_t\|$ and $\|\underline{h}_t\| \leq \|\underline{\tilde{z}}_t\|$. □

4.4. *Proof of the Consistency and the Asymptotic Normality of the QML.* The proof follows the lines of that of Theorems 2.1 and 2.2 in Francq and Zakoian (2004) for the univariate case.

We shall use the multiplicative norm defined by:

$$\|A\| := \sup_{\|x\|\leq 1} \|Ax\| = \rho^{1/2}(A'A), \tag{4.5}$$

where A is a $d_1 \times d_2$ matrix, $\|x\|$ is the euclidian norm of vector $x \in \mathbb{R}^{d_2}$, and $\rho(\cdot)$ denotes the spectral radius. This norm verifies, for any $d_2 \times d_1$ matrix B,

$$\|A\|^2 \leq \sum_{i,j} a_{i,j}^2 = \mathrm{Tr}(A'A) \leq d_2\|A\|^2, \quad |A'A| \leq \|A\|^{2d_2}, \tag{4.6}$$

$$|\mathrm{Tr}(AB)| \leq \left(\sum_{i,j} a_{i,j}^2\right)^{1/2} \left(\sum_{i,j} b_{i,j}^2\right)^{1/2} \leq \{d_2 d_1\}^{1/2}\|A\|\|B\|. \tag{4.7}$$

4.4.1. *Proof of Theorem 3.1.* Rewrite (3.6) in matrix form as

$$\mathbf{H}_t = \underline{c}_t + \mathbb{B}\mathbf{H}_{t-1} \tag{4.8}$$

where $\mathbb{B}$ is defined in Corollary 2.1 and

$$\mathbf{H}_t = \begin{pmatrix} \underline{h}_t \\ \underline{h}_{t-1} \\ \vdots \\ \underline{h}_{t-p+1} \end{pmatrix}, \quad \underline{c}_t = \begin{pmatrix} \underline{\omega} + \sum_{i=1}^{q} \mathbf{A}_i \underline{\epsilon}_{t-i} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{4.9}$$

We will establish the following intermediate results.

i) $\lim_{n\to\infty} \sup_{\theta\in\Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| = 0$, a.s.

ii) $(\exists t \in \mathbb{Z}$ such that $\underline{h}_t(\theta) = \underline{h}_t(\theta_0)$ P_{θ_0} a.s. and $R(\theta) = R(\theta_0))$ $\Longrightarrow \theta = \theta_0$,

iii) $E_{\theta_0}|\ell_t(\theta_0)| < \infty$, and if $\theta \neq \theta_0$, $E_{\theta_0}\ell_t(\theta) > E_{\theta_0}\ell_t(\theta_0)$,

iv) for any $\theta \neq \theta_0$ there exists a neighborhood $V(\theta)$ such that

$$\liminf_{n\to\infty} \inf_{\theta^*\in V(\theta)} \tilde{\mathbf{l}}_n(\theta^*) > E_{\theta_0}\ell_1(\theta_0), \quad \text{a.s.}$$

Proof of i). In view of Assumption **A2** and Corollary 2.1, we have $\rho(\mathbb{B}) < 1$. By the compactness of Θ we even have

$$\sup_{\theta\in\Theta} \rho(\mathbb{B}) < 1. \tag{4.10}$$

Using iteratively Equation (4.8), we deduce that, almost surely

$$\sup_{\theta\in\Theta} \|\mathbf{H}_t - \tilde{\mathbf{H}}_t\| \leq K\rho^t, \quad \forall t, \tag{4.11}$$

where $\tilde{\mathbf{H}}_t$ denotes the vector obtained by replacing the variables $\underline{h}_{t-i}$ by $\underline{\tilde{h}}_{t-i}$ in $\mathbf{H}_t$. Observe that K is a random variable which depends on the past values $\{\epsilon_t, t \leq 0\}$. Since K does not depend on n, it can be considered as a constant, such as ρ. From (4.11) we deduce that, almost surely,

$$\sup_{\theta\in\Theta} \|H_t - \tilde{H}_t\| \quad \leq \quad K\rho^t, \quad \forall t. \tag{4.12}$$

Noting that $\|R^{-1}\|$ is the inverse of the eigenvalue of smaller module of R, and that $\|\tilde{D}_t^{-1}\| = \{\min_i(h_{ii,t})\}^{-1}$, we have

$$\sup_{\theta\in\Theta} \|\tilde{H}_t^{-1}\| \leq \sup_{\theta\in\Theta} \|\tilde{D}_t^{-1}\|^2\|R^{-1}\| \leq \sup_{\theta\in\Theta}\{\min_i \underline{\omega}(i)\}^{-2}\|R^{-1}\| \leq K, \tag{4.13}$$

using **A5**, the compactness of Θ and the strict positivity of the components of $\underline{\omega}$. Similarly we have

$$\sup_{\theta\in\Theta} \|H_t^{-1}\| \leq K. \tag{4.14}$$

Now

$$\begin{aligned} \sup_{\theta\in\Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| \quad \leq \quad & n^{-1}\sum_{t=1}^{n} \sup_{\theta\in\Theta} \left|\epsilon_t'(H_t^{-1} - \tilde{H}_t^{-1})\epsilon_t\right| \\ & + n^{-1}\sum_{t=1}^{n} \sup_{\theta\in\Theta} \left|\log|H_t| - \log|\tilde{H}_t|\right|. \end{aligned} \tag{4.15}$$

The first sum can be written as

$$\begin{aligned}
& n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|\epsilon_t'\tilde{H}_t^{-1}(H_t-\tilde{H}_t)H_t^{-1}\epsilon_t\right| \\
= \;& n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|\text{ Tr }\{\epsilon_t'\tilde{H}_t^{-1}(H_t-\tilde{H}_t)H_t^{-1}\epsilon_t\}\right| \\
= \;& n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|\text{ Tr }\{\tilde{H}_t^{-1}(H_t-\tilde{H}_t)H_t^{-1}\epsilon_t\epsilon_t'\}\right| \\
\leq \;& Kn^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\|\tilde{H}_t^{-1}\|\|H_t-\tilde{H}_t\|\|H_t^{-1}\|\|\epsilon_t\epsilon_t'\| \\
\leq \;& Kn^{-1}\sum_{t=1}^{n}\rho^t\|\epsilon_t\epsilon_t'\|\to 0
\end{aligned}$$

as $n\to\infty$, using (4.7), (4.12), (4.13), (4.14), the Cesàro lemma and the fact that $\rho^t\|\epsilon_t\epsilon_t'\|=\rho^t\epsilon_t'\epsilon_t\to 0$ a.s. The latter statement can be shown by using the Borel-Cantelli lemma, the Markov inequality and by applying Corollary 2.2:

$$\sum_{t=1}^{\infty}\mathbb{P}(\rho^t\epsilon_t'\epsilon_t>\varepsilon)\leq\sum_{t=1}^{\infty}\frac{\rho^{st}E(\epsilon_t'\epsilon_t)^s}{\varepsilon^s}=\sum_{t=1}^{\infty}\frac{\rho^{st}E\|\epsilon_t\|^{2s}}{\varepsilon^s}<\infty.$$

Now, using (4.6), the triangle inequality and, for $x\geq -1$, $\log(1+x)\leq x$, we have

$$\begin{aligned}
\log|H_t|-\log|\tilde{H}_t| &= \log|I_m+(H_t-\tilde{H}_t)\tilde{H}_t^{-1}| \\
&\leq m\log\|I_m+(H_t-\tilde{H}_t)\tilde{H}_t^{-1}\| \\
&\leq m\log(\|I_m\|+\|(H_t-\tilde{H}_t)\tilde{H}_t^{-1}\|) \\
&\leq m\log(1+\|(H_t-\tilde{H}_t)\tilde{H}_t^{-1}\|) \\
&\leq m\|H_t-\tilde{H}_t\|\|\tilde{H}_t^{-1}\|,
\end{aligned}$$

and, by symmetry,

$$\log|\tilde{H}_t|-\log|H_t|\leq m\|H_t-\tilde{H}_t\|\|H_t^{-1}\|.$$

Using again (4.12), (4.13) and (4.14) we deduce that, in (4.15), the second sum tends to 0. We thus have shown i).

Proof of ii). Suppose that for some $\theta\neq\theta_0$, the following holds

$$\underline{h}_t(\theta)=\underline{h}_t(\theta_0),\quad P_{\theta_0}\text{-a.s.}\quad\text{and}\quad R(\theta)=R(\theta_0).$$

Then, it readily follows that $\rho = \rho_0$ and, using the invertibility of the polynomial $\mathcal{B}_\theta(B)$ under Assumption **A2**, by (3.1)

$$\mathcal{B}_\theta(1)^{-1}\underline{\omega} + \mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta(B)\underline{\epsilon}_t = \mathcal{B}_{\theta_0}(1)^{-1}\underline{\omega}_0 + \mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}(B)\underline{\epsilon}_t$$

that is

$$\begin{aligned}\mathcal{B}_\theta(1)^{-1}\underline{\omega} - \mathcal{B}_{\theta_0}(1)^{-1}\underline{\omega}_0 &= \{\mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}(B) - \mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta(B)\}\underline{\epsilon}_t \\ &:= \mathcal{P}(B)\underline{\epsilon}_t \quad \text{a.s.} \quad \forall t.\end{aligned}$$

Write $\mathcal{P}(B) = \sum_{i=0}^{\infty}\mathcal{P}_i B^i$. Noting that $\mathcal{P}_0 = \mathcal{P}(0) = 0$ and isolating the terms functions of the components of η_{t-1}, we obtain

$$\mathcal{P}_1(h_{11,t-1}\eta^2_{1,t-1}, \dots, h_{mm,t-1}\eta^2_{m,t-1})' = Z_{t-2}, \quad \text{a.s.}$$

where Z_{t-2} belongs to the σ-field generated by $\{\eta_{t-2}, \eta_{t-3}, \dots\}$.Since η_{t-1} is independent from this σ-field, the latter equality contradicts **A3** unless if, for $i, j = 1, \dots, m$, $p_{ij}h_{jj,t} = 0$, a.s., where the p_{ij} are the entries of $\mathcal{P}_1$. Because $h_{jj,t} > 0$ for all j, we thus have $\mathcal{P}_1 = 0$. Similarly, we show that $\mathcal{P}(B) = 0$ by successively considering the past values of η_{t-1}. Therefore, in view of **A4** (or **A4'**), we have $\alpha = \alpha_0$ and $\beta = \beta_0$ (see Section 3.1). It readily follows that $\underline{\omega} = \underline{\omega}_0$. Hence $\theta = \theta_0$. We thus have established *ii)*.

Proof of iii). We first show that $E_{\theta_0}\ell_t(\theta)$ is well defined in $\mathbb{R} \cup \{+\infty\}$ for all θ, and in $\mathbb{R}$ for $\theta = \theta_0$. We have

$$E_{\theta_0}\ell_t^-(\theta) \le E_{\theta_0}\log^-|H_t| \le \max\{0, -\log(|R|\min_i \underline{\omega}(i)^m)\} < \infty.$$

At θ_0, Jensen's inequality, the second inequality in (4.6) and Corollary 2.2 entail

$$\begin{aligned}
& E_{\theta_0}\log|H_t(\theta_0)| = E_{\theta_0}\frac{m}{s}\log|H_t(\theta_0)|^{s/m} \\
\le\ & \frac{m}{s}\log E_{\theta_0}\|H_t(\theta_0)\|^s \le \frac{m}{s}\log E_{\theta_0}\|R\|^s\|D_t(\theta_0)\|^{2s} \qquad (\mathbf{Pb}?) \\
\le\ & K + \frac{m}{s}\log E_{\theta_0}\|D_t(\theta_0)\|^{2s} = K + \frac{m}{s}\log E_{\theta_0}(\max_i h_{ii,t}(\theta_0))^s \\
\le\ & K + \frac{m}{s}\log E_{\theta_0}\left\{\sum_i h^2_{ii,t}(\theta_0)\right\}^{s/2} = K + \frac{m}{s}\log E_{\theta_0}\|\underline{h}_t(\theta_0)\|^s < \infty.
\end{aligned}$$

It follows that

$$\begin{aligned} E_{\theta_0}\ell_t(\theta_0) &= E_{\theta_0}\left\{\eta_t' H_t(\theta_0)^{1/2'} H_t(\theta_0)^{-1} H_t(\theta_0)^{1/2}\eta_t + \log|H_t(\theta_0)|\right\} \\ &= m + E_{\theta_0}\log|H_t(\theta_0)| < \infty. \end{aligned}$$

Because $E_{\theta_0}\ell_t^-(\theta_0) < \infty$, the existence of $E_{\theta_0}\ell_t(\theta_0)$ in $\mathbb{R}$ holds. It is thus not restrictive to study the minimum of $E_{\theta_0}\ell_t(\theta)$ for the values of θ such that $E_{\theta_0}|\ell_t(\theta)| < \infty$. Denoting by $\lambda_{i,t}$, the positive eigenvalues of $H_t(\theta_0)H_t^{-1}(\theta)$, we have

$$\begin{aligned} & E_{\theta_0}\ell_t(\theta) - E_{\theta_0}\ell_t(\theta_0) \\ = \; & E_{\theta_0}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + E_{\theta_0}\left\{\eta_t'[H_t^{1/2}(\theta_0)' H_t^{-1}(\theta) H_t^{1/2}(\theta_0) - I_m]\eta_t\right\} \\ = \; & E_{\theta_0}\log\{|H_t(\theta)H_t^{-1}(\theta_0)|\} \\ & + \mathrm{Tr}\,\left(E_{\theta_0}\left\{[H_t^{1/2}(\theta_0)' H_t^{-1}(\theta) H_t^{1/2}(\theta_0) - I_m]\right\} E(\eta_t\eta_t')\right) \\ = \; & E_{\theta_0}\log\{|H_t(\theta)H_t^{-1}(\theta_0)|\} + E_{\theta_0}\left(\mathrm{Tr}\left\{[H_t(\theta_0)H_t^{-1}(\theta) - I_m]\right\}\right) \\ = \; & E_{\theta_0}\left\{\sum_{i=1}^m (\lambda_{it} - 1 - \log\lambda_{it})\right\} \geq 0 \end{aligned}$$

because $\log x \leq x - 1$, $\forall x > 0$. Since $\log x = x - 1$ if and only if $x = 1$, the inequality is strict unless if, for all i, $\lambda_{it} = 1$ P_{θ_0}-a.s. , that is if $H_t(\theta) = H_t(\theta_0)$, P_{θ_0}-a.s. . This equality is equivalent to

$$\underline{h}_t(\theta) = \underline{h}_t(\theta_0), \quad P_{\theta_0}\text{-a.s.} \quad \text{and} \quad R(\theta) = R(\theta_0)$$

and thus to $\theta = \theta_0$, from ii).

Proof of iv). The last part of the proof of the consistency uses the compactness of Θ and the ergodicity of $(\ell_t(\theta))$, as in the univariate case. Therefore is it omitted.

Theorem 3.1 is thus established. □

4.4.2. *Proof of Theorem 3.2.* We start by stating a few elementary results on the differentiation of expressions involving matrices. If $f(A)$ is a real valued function of a matrix A whose entries a_{ij} are functions of some variable x, the chain rule for differentiation of composed functions entails

$$\frac{\partial f(A)}{\partial x} = \sum_{i,j} \frac{\partial f(A)}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial x} = \mathrm{Tr}\left\{ \frac{\partial f(A)}{\partial A'} \frac{\partial A}{\partial x} \right\}. \tag{4.16}$$

Moreover, for A invertible we have

$$\frac{\partial c'Ac}{\partial A'} = cc' \tag{4.17}$$

$$\frac{\partial \mathrm{Tr}(CA'BA')}{\partial A'} = C'AB' + B'AC' \tag{4.18}$$

$$\frac{\partial \log|\det(A)|}{\partial A'} = A^{-1} \tag{4.19}$$

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1}\frac{\partial A}{\partial x}A^{-1} \tag{4.20}$$

$$\frac{\partial \mathrm{Tr}(CA^{-1}B)}{\partial A'} = -A^{-1}BCA^{-1} \tag{4.21}$$

$$\frac{\partial \mathrm{Tr}(CAB)}{\partial A'} = BC \tag{4.22}$$

The proof is divided into several steps.

a) First derivative of the criterion. Applying (4.16) and (4.17), then (4.18), (4.19) and (4.20), we obtain

$$\begin{aligned}
\frac{\partial \ell_t(\theta)}{\partial \theta_i} &= \mathrm{Tr}\left(\epsilon_t \epsilon_t' \frac{\partial D_t^{-1} R^{-1} D_t^{-1}}{\partial \theta_i} \right) + 2\frac{\partial \log|\det D_t|}{\partial \theta_i} \\
&= -\mathrm{Tr}\left\{ \left(\epsilon_t\epsilon_t' D_t^{-1} R^{-1} + R^{-1} D_t^{-1} \epsilon_t \epsilon_t' \right) D_t^{-1} \frac{\partial D_t}{\partial \theta_i} D_t^{-1} \right\} \\
&\quad +2\mathrm{Tr}\left(D_t^{-1} \frac{\partial D_t}{\partial \theta_i} \right)
\end{aligned} \tag{4.23}$$

for $i = 1, \dots, s_1 = m + (p+q)m^2$, and using (4.21)

$$\frac{\partial \ell_t(\theta)}{\partial \theta_i} = -\mathrm{Tr}\left(R^{-1} D_t^{-1} \epsilon_t \epsilon_t' D_t^{-1} R^{-1} \frac{\partial R}{\partial \theta_i} \right) + \mathrm{Tr}\left(R^{-1} \frac{\partial R}{\partial \theta_i} \right) \tag{4.24}$$

for $i = s_1 + 1, \dots, s_0$. Letting $D_{0t} = D_t(\theta_0)$, $R_0 = R(\theta_0)$,

$$D_{0t}^{(i)} = \frac{\partial D_t}{\partial \theta_i}(\theta_0),\ R_0^{(i)} = \frac{\partial R}{\partial \theta_i}(\theta_0),\ D_{0t}^{(i,j)} = \frac{\partial^2 D_t}{\partial \theta_i \partial \theta_j}(\theta_0)\ \ R_0^{(i,j)} = \frac{\partial^2 R}{\partial \theta_i \partial \theta_j}(\theta_0),$$

and $\tilde{\eta}_t = R^{1/2}\eta_t$, the score vector writes

$$\begin{aligned}
\frac{\partial \ell_t(\theta_0)}{\partial \theta_i} &= \mathrm{Tr}\left\{ \left(I_m - R_0^{-1} \tilde{\eta}_t \tilde{\eta}_t' \right) D_{0t}^{(i)} D_{0t}^{-1} \right. \\
&\quad \left. + \left(I_m - \tilde{\eta}_t \tilde{\eta}_t' R_0^{-1} \right) D_{0t}^{-1} D_{0t}^{(i)} \right\},
\end{aligned} \tag{4.25}$$

for $i = 1, \dots, s_1$, and

$$\frac{\partial \ell_t(\theta_0)}{\partial \theta_i} = \mathrm{Tr}\left\{\left(I_m - R_0^{-1}\tilde{\eta}_t\tilde{\eta}_t'\right) R_0^{-1} R_0^{(i)}\right\}, \qquad (4.26)$$

for $i = s_1 + 1, \dots, s_0$.

b) Existence of moments at any order for the score. In view of (4.7) and the Cauchy-Schwarz inequality, we obtain

$$E\left|\frac{\partial \ell_t(\theta_0)}{\partial \theta_i}\frac{\partial \ell_t(\theta_0)}{\partial \theta_j}\right| \leq K\left\{E\left\|D_{0t}^{-1}D_{0t}^{(i)}\right\|^2 E\left\|D_{0t}^{-1}D_{0t}^{(j)}\right\|^2\right\}^{1/2},$$

for $i, j = 1, \dots, s_1$,

$$E\left|\frac{\partial \ell_t(\theta_0)}{\partial \theta_i}\frac{\partial \ell_t(\theta_0)}{\partial \theta_j}\right| < KE\left\|D_{0t}^{-1}D_{0t}^{(i)}\right\|,$$

for $i = 1, \dots, s_1$ and $j = s_1 + 1, \dots, s_0$, and $E\left|\frac{\partial \ell_t(\theta_0)}{\partial \theta_i}\frac{\partial \ell_t(\theta_0)}{\partial \theta_j}\right| < K$ for $i, j = s_1 + 1, \dots, s_0$. Note also that

$$D_{0t}^{(i)} = \frac{1}{2}D_{0t}^{-1}\mathrm{diag}\left\{\frac{\partial \underline{h}_t}{\partial \theta_i}(\theta_0)\right\}.$$

To show that the score admits a second-order moment, it is thus sufficient to prove that

$$E\left|\frac{1}{\underline{h}_t(i_1)}\frac{\partial \underline{h}_t(i_1)}{\partial \theta_i}(\theta_0)\right|^{r_0} < \infty$$

for all $i_1 = 1, \dots, m$, all $i = 1, \dots, s_1$ and $r_0 = 2$. By (4.8) and (4.10),

$$\sup_{\theta\in\Theta}\left\|\frac{\partial \mathbf{H}_t}{\partial \theta_i}\right\| < \infty, \qquad i = 1, \dots, m$$

and, setting $s_2 = m + qm^2$,

$$\left\|\frac{\partial \mathbf{H}_t}{\partial \theta_i}\right\| \leq \epsilon_{t-j(i)}^2 \inf_{m<k\leq s_2}\theta_k, \qquad i = m+1, \dots, s_2, \qquad (???)$$

where $j(i) \in \{1, \dots, q\}$. On the other hand we have

$$\frac{\partial \mathbf{H}_t}{\partial \theta_i} = \sum_{k=1}^{\infty}\left\{\sum_{j=1}^{k}\mathbb{B}^{j-1}\mathbb{B}^{(i)}\mathbb{B}^{k-j}\right\}\underline{c}_{t-k}, \qquad i = s_2 + 1, \dots, s_1,$$

where $\mathbb{B}^{(i)} = \partial\mathbb{B}/\partial\theta_i$ is a matrix whose entries are all 0, apart from a 1 located at the same place as θ_i in $\mathbb{B}$. By abuse of notation, we denote by $\mathbf{H}_t(i_1)$ and $\underline{h}_{0t}(i_1)$

the i_1^{th} components of $\mathbf{H}_t$ and $\underline{h}_t(\theta_0)$. With arguments similar to those used in the univariate case, that is the inequality $x/(1+x) \le x^s$ for all $x \ge 0$ and $s \in [0,1]$, and the inequalities

$$\theta_i \frac{\partial \mathbf{H}_t}{\partial \theta_i} \le \sum_{k=1}^{\infty} k \mathbb{B}^k \underline{c}_{t-k}, \quad \theta_i \frac{\partial \mathbf{H}_t(i_1)}{\partial \theta_i} \le \sum_{k=1}^{\infty} k \sum_{j_1=1}^{m} \mathbb{B}^k(i_1, j_1) \underline{c}_{t-k}(j_1)$$

and, setting $\omega = \inf_{1 \le i \le m} \underline{\omega}(i)$,

$$\mathbf{H}_t(i_1) \ge \omega + \sum_{j_1=1}^{m} \mathbb{B}^k(i_1, j_1) \underline{c}_{t-k}(j_1), \qquad \forall k,$$

we obtain

$$\frac{\theta_i}{\mathbf{H}_t(i_1)} \frac{\partial \mathbf{H}_t(i_1)}{\partial \theta_i} \le \sum_{j_1=1}^{m} \sum_{k=1}^{\infty} k \left\{ \frac{\mathbb{B}^k(i_1, j_1) \underline{c}_{t-k}(j_1)}{\omega} \right\}^{\frac{s}{r_0}} \le K \sum_{j_1=1}^{m} \sum_{k=1}^{\infty} k \rho_{j_1}^k \underline{c}_{t-k}^{s/r_0}(j_1),$$

where the constants ρ_{j_1} (which also depend of i_1, s and r_0) belong to the interval $[0,1)$. Noting that these inequalities are uniform on a neighborhood of $\theta_0 \in \overset{\circ}{\Theta}$, that they can be extended to higher-order derivatives, as in the univariate case, and that Corollary 2.2 implies $\|\underline{c}_t\|_s < \infty$, we can show a stronger result than the one announced: for all $i_1 = 1, \dots, m$, all $i, j, k = 1, \dots, s_1$ and all $r_0 \ge 0$, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that

$$E \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{1}{\underline{h}_t(i_1)} \frac{\partial \underline{h}_t(i_1)}{\partial \theta_i}(\theta) \right|^{r_0} < \infty, \tag{4.27}$$

$$E \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{1}{\underline{h}_t(i_1)} \frac{\partial^2 \underline{h}_t(i_1)}{\partial \theta_i \partial \theta_j}(\theta) \right|^{r_0} < \infty \tag{4.28}$$

and

$$E \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{1}{\underline{h}_t(i_1)} \frac{\partial^3 \underline{h}_t(i_1)}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta) \right|^{r_0} < \infty. \tag{4.29}$$

c) Asymptotic normality of the score vector. Clearly, $\{\partial \ell_t(\theta_0)/\partial \theta\}_t$ is stationary and $\partial \ell_t(\theta_0)/\partial \theta$ is measurable with respect to the σ-field $\mathcal{F}_t$ generated by $\{\eta_u, u \le t\}$. From (4.25) and (4.26) we have $E\{\partial \ell_t(\theta_0)/\partial \theta \mid \mathcal{F}_{t-1}\} = 0$. The property b), and in particular (4.27), ensures the existence of the matrix I in Theorem 3.2. It follows that $\forall \lambda \in \mathbb{R}^{p+q+1}$, the sequence $\left\{\lambda' \frac{\partial}{\partial \theta} \ell_t(\theta_0), \mathcal{F}_t\right\}_t$ is an ergodic, stationary and square integrable martingale difference. The central limit theorem of

Billingsley (1961) entails

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ell_t(\theta_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,I).$$

d) Higher-order derivatives of the criterion. Starting from a) and applying several times (4.16) and (4.21), as well as (4.22), we obtain

$$\frac{\partial\ell_t^2(\theta)}{\partial\theta_i\partial\theta_j} = \mathrm{Tr}\left(c_1+c_2+c_3\right), \quad i,j=1,\dots,s_1,$$

where

$$\begin{aligned}
c_1 &= D_t^{-1}R^{-1}D_t^{-1}\frac{\partial D_t}{\partial\theta_i}D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}\frac{\partial D_t}{\partial\theta_j}+D_t^{-1}\frac{\partial D_t}{\partial\theta_i}D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}D_t^{-1}\frac{\partial D_t}{\partial\theta_j}\\
&\quad +D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}D_t^{-1}\frac{\partial D_t}{\partial\theta_i}D_t^{-1}\frac{\partial D_t}{\partial\theta_j}-D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j},\\
c_2 &= -2D_t^{-1}\frac{\partial D_t}{\partial\theta_i}D_t^{-1}\frac{\partial D_t}{\partial\theta_j}+2D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j},
\end{aligned}$$

and c_3 is obtained by permuting $\epsilon_t\epsilon_t'$ and R^{-1} in c_1. We also obtain

$$\begin{aligned}
\frac{\partial\ell_t^2(\theta)}{\partial\theta_i\partial\theta_j} &= \mathrm{Tr}\left(c_4+c_5\right), \quad i=1,\dots,s_1, \quad j=s_1+1,\dots,s_0\\
\frac{\partial\ell_t^2(\theta)}{\partial\theta_i\partial\theta_j} &= \mathrm{Tr}\left(c_6\right), \quad i,j=s_1+1,\dots,s_0
\end{aligned}$$

where

$$\begin{aligned}
c_4 &= R^{-1}D_t^{-1}\frac{\partial D_t}{\partial\theta_i}D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}\frac{\partial R}{\partial\theta_j}\\
c_6 &= R^{-1}D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}\frac{\partial R}{\partial\theta_i}R^{-1}\frac{\partial R}{\partial\theta_j}+R^{-1}\frac{\partial R}{\partial\theta_i}R^{-1}D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}\frac{\partial R}{\partial\theta_j}\\
&\quad -R^{-1}D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}R^{-1}\frac{\partial^2 R}{\partial\theta_i\partial\theta_j}-R^{-1}\frac{\partial^2 R}{\partial\theta_i\partial\theta_j}-R^{-1}\frac{\partial R}{\partial\theta_i}R^{-1}\frac{\partial R}{\partial\theta_j},
\end{aligned}$$

and c_5 is obtained by permuting $\epsilon_t\epsilon_t'$ and $\partial D_t/\partial\theta_i$ in c_4. Results (4.27) and (4.28) ensure the existence of the matrix $J=E\partial^2\ell_t(\theta_0)/\partial\theta\partial\theta'$, which is invertible, as will be shown in e) below. Note that with our parameterization, $\partial^2 R/\partial\theta_i\partial\theta_j=0$.

Continuing the differentiations, it is seen that $\partial\ell_t^3(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$ is also the trace of a sum of products of matrices similar to the c_i's. The integrable matrix $\epsilon_t\epsilon_t'$ appears at most one time in each of these products. The other terms are, on the one hand, the bounded matrices R^{-1}, $\partial R/\partial\theta_i$ and D_t^{-1} and, on the other hand, the matrices $D_t^{-1}\partial D_t/\partial\theta_i$, $D_t^{-1}\partial^2 D_t/\partial\theta_i\partial\theta_j$ and $D_t^{-1}\partial^3 D_t/\partial\theta_i\partial\theta_j\partial\theta_k$. From (4.27)-(4.29), the norms of these last 3 matrices admit moments at any orders in the neighborhood of θ_0. This shows that

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

e) Invertibility of the matrix J**.** The expression for J obtained in d), as a function of the partial derivatives of D_t and R, is not convenient to show its invertibility. Instead, we follow the approach of Comte and Lieberman (2003) p.77-78. We start by writing J as a function of H_t and of its derivatives. Starting from

$$\ell_t(\theta)=\epsilon_t' H_t^{-1}\epsilon_t+\log|H_t|,$$

the differentiation formulas (4.16), (4.19) and (4.21) give

$$\frac{\partial\ell_t}{\partial\theta_i}=\mathrm{Tr}\left\{\left(H_t^{-1}-H_t^{-1}\epsilon_t\epsilon_t' H_t^{-1}\right)\frac{\partial H_t}{\partial\theta_i}\right\},$$

and then, using (4.20) and (4.22),

$$\begin{aligned}\frac{\partial^2\ell_t}{\partial\theta_i\partial\theta_j} &= \mathrm{Tr}\left(H_t^{-1}\frac{\partial^2 H_t}{\partial\theta_i\partial\theta_j}\right)-\mathrm{Tr}\left(H_t^{-1}\frac{\partial H_t}{\partial\theta_j}H_t^{-1}\frac{\partial H_t}{\partial\theta_i}\right)\\ &\quad \mathrm{Tr}\left(H_t^{-1}\epsilon_t\epsilon_t' H_t^{-1}\frac{\partial H_t}{\partial\theta_i}H_t^{-1}\frac{\partial H_t}{\partial\theta_j}\right)+\mathrm{Tr}\left(H_t^{-1}\frac{\partial H_t}{\partial\theta_i}H_t^{-1}\epsilon_t\epsilon_t' H_t^{-1}\frac{\partial H_t}{\partial\theta_j}\right)\\ &\quad -\mathrm{Tr}\left(H_t^{-1}\epsilon_t\epsilon_t' H_t^{-1}\frac{\partial^2 H_t}{\partial\theta_i\partial\theta_j}\right).\end{aligned}$$

Using the relation $\mathrm{Tr}(A'B)=(\mathrm{vec}A)'\mathrm{vec}B$, we deduce

$$E\left(\frac{\partial^2\ell_t(\theta_0)}{\partial\theta_i\partial\theta_j}\mid\mathcal{F}_{t-1}\right)=\mathrm{Tr}\left(H_{0t}^{-1}H_{0t}^{(i)}H_{0t}^{-1}H_{0t}^{(j)}\right)=\mathbf{h}_i'\mathbf{h}_j,$$

where, in view of $\mathrm{vec}(ABC)=(C'\otimes A)\mathrm{vec}B$,

$$\mathbf{h}_i=\mathrm{vec}\left(H_{0t}^{-1/2}H_{0t}^{(i)}H_{0t}^{-1/2}\right)=\left(H_{0t}^{-1/2}\otimes H_{0t}^{-1/2}\right)\mathbf{d}_i,\quad \mathbf{d}_i=\mathrm{vec}\left(H_{0t}^{(i)}\right).$$

Introducing the matrices $m^2\times s_0$

$$\mathbf{h}=(\mathbf{h}_1\mid\cdots\mid\mathbf{h}_{s_0})\quad\text{and}\quad\mathbf{d}=(\mathbf{d}_1\mid\cdots\mid\mathbf{d}_{s_0}),$$

we have $\mathbf{h} = \mathbf{Hd}$ with $\mathbf{H} = H_{0t}^{-1/2} \otimes H_{0t}^{-1/2}$. Now suppose that $J = E\mathbf{h}'\mathbf{h}$ is singular. Then, there exists a non-zero vector $\mathbf{c} \in \mathbb{R}^{s_0}$, such that $\mathbf{c}'J\mathbf{c} = E\mathbf{c}'\mathbf{h}'\mathbf{hc} = 0$. Since $\mathbf{c}'\mathbf{h}'\mathbf{hc} \geq 0$ almost surely, it means that

$$\mathbf{c}'\mathbf{h}'\mathbf{hc} = \mathbf{c}'\mathbf{d}'\mathbf{H}^2\mathbf{dc} = 0 \quad a.s. \tag{4.30}$$

Because $\mathbf{H}^2$ is a positive-definite matrix, with probability 1, this entails $\mathbf{dc} = 0_{m^2}$ with probability 1. Decompose $\mathbf{c}$ under form $\mathbf{c} = (\mathbf{c}_1', \mathbf{c}_2')'$ with $\mathbf{c_1} \in \mathbb{R}^{s_1}$ and $\mathbf{c_2} \in \mathbb{R}^{s_3}$, where $s_3 = s_0 - s_1 = m(m-1)/2$. The rows $1, m+1, \dots, m^2$ of the equations

$$\mathbf{dc} = \sum_{i=1}^{s_0} c_i \frac{\partial}{\partial \theta_i} \text{vec} H_{0t} = \sum_{i=1}^{s_0} c_i \frac{\partial}{\partial \theta_i} \left(D_{0t} \otimes D_{0t} \right) \text{vec} R_0 = 0_{m^2}, \quad a.s. \tag{4.31}$$

give

$$\sum_{i=1}^{s_1} c_i \frac{\partial}{\partial \theta_i} \underline{h}_t(\theta_0) = 0_m, \quad a.s. \tag{4.32}$$

Differentiating Equation (3.6) yields

$$\sum_{i=1}^{s_1} c_i \frac{\partial}{\partial \theta_i} \underline{h}_t \quad = \quad \underline{\omega}^* + \sum_{j=1}^{q} \mathbf{A}_j^* \underline{\epsilon}_{t-j} + \sum_{j=1}^{p} \mathbf{B}_j^* \underline{h}_{t-j} + \sum_{j=1}^{p} \mathbf{B}_j \sum_{i=1}^{s_1} c_i \frac{\partial}{\partial \theta_i} \underline{h}_{t-j}$$

where

$$\underline{\omega}^* = \sum_{i=1}^{s_1} c_i \frac{\partial}{\partial \theta_i} \underline{\omega}, \quad \mathbf{A}_j^* = \sum_{i=1}^{s_1} c_i \frac{\partial}{\partial \theta_i} \mathbf{A}_j, \quad \mathbf{B}_j^* = \sum_{i=1}^{s_1} c_i \frac{\partial}{\partial \theta_i} \mathbf{B}_j.$$

Because (4.32) is satisfied for all t, we have

$$\underline{\omega}_0^* + \sum_{j=1}^{q} \mathbf{A}_{0j}^* \underline{\epsilon}_{t-j} + \sum_{j=1}^{p} \mathbf{B}_{0j}^* \underline{h}_{t-j}(\theta_0) = 0,$$

where quantities evaluated at $\theta = \theta_0$ are indexed by 0. This entails

$$\underline{h}_t(\theta_0) = \underline{\omega}_0 - \underline{\omega}_0^* + \sum_{j=1}^{q} \left(\mathbf{A}_{0j} - \mathbf{A}_{0j}^* \right) \underline{\epsilon}_{t-j} + \sum_{j=1}^{p} \left(\mathbf{B}_{0j} - \mathbf{B}_{0j}^* \right) \underline{h}_{t-j}(\theta_0),$$

and finally, introducing a vector θ_1 whose first s_1 components are $\text{vec}\left(\underline{\omega}_0 - \underline{\omega}_0^* \mid \mathbf{A}_{01} - \mathbf{A}_{01}^* \mid \cdots \mid \mathbf{B}_{0p} - \mathbf{B}_{0p}^* \right)$,

$$\underline{h}_t(\theta_0) = \underline{h}_t(\theta_1)$$

by choosing $\mathbf{c_1}$ small enough so that $\theta_1 \in \Theta$. If $\mathbf{c_1}$ is not equal to zero then $\theta_1 \neq \theta_0$. This is in contradiction with the identifiability of the parameter, hence $\mathbf{c_1} = 0$. Equations (4.31) thus become

$$(D_{0t} \otimes D_{0t}) \sum_{i=s_1+1}^{s_0} c_i \frac{\partial}{\partial \theta_i} \mathrm{vec} R_0 = 0_{m^2}, \quad a.s.$$

Therefore,

$$\sum_{i=s_1+1}^{s_0} c_i \frac{\partial}{\partial \theta_i} \mathrm{vec} R_0 = 0_{m^2}.$$

Because the vectors $\partial \mathrm{vec} R / \partial \theta_i$, $i = s_1 + 1, \dots, s_0$, are linearly independent, the vector $\mathbf{c_2} = (c_{s_1+1}, \dots, c_{s_0})'$ is nul, and thus $\mathbf{c} = 0$. This is in contradiction with (4.30), and shows that the assumption that J is singular is absurd.

f) Forgetting of the initial values. First remark that (4.11) and the arguments used to show (4.13) and (4.14) entail

$$\sup_{\theta \in \Theta} \|D_t - \tilde{D}_t\| \leq K\rho^t, \quad \sup_{\theta \in \Theta} \|\tilde{D}_t^{-1}\| \leq K, \quad \sup_{\theta \in \Theta} \|D_t^{-1}\| \leq K, \tag{4.33}$$

and thus

$$\begin{aligned} \sup_{\theta \in \Theta} \|D_t^{1/2} - \tilde{D}_t^{1/2}\| &\leq K\rho^t, \quad \sup_{\theta \in \Theta} \|\tilde{D}_t^{-1/2}\| \leq K, \quad \sup_{\theta \in \Theta} \|D_t^{-1/2}\| \leq K, \\ \sup_{\theta \in \Theta} \|D_t^{1/2} \tilde{D}_t^{-1/2}\| &\leq K(1 + \rho^t) \quad \sup_{\theta \in \Theta} \|\tilde{D}_t^{1/2} D_t^{-1/2}\| \leq K(1 + \rho^t). \end{aligned} \tag{4.34}$$

From (4.8), we have

$$\mathbf{H}_t = \sum_{k=0}^{t-r-1} \mathbb{B}^k \underline{c}_{t-k} + \mathbb{B}^{t-r} \mathbf{H}_r, \qquad \tilde{\mathbf{H}}_t = \sum_{k=0}^{t-r-1} \mathbb{B}^k \underline{\tilde{c}}_{t-k} + \mathbb{B}^{t-r} \tilde{\mathbf{H}}_r$$

where $r = \max\{p, q\}$ and the tilde means that initial values are taken into account. Since $\underline{\tilde{c}}_t = \underline{c}_t$ for all $t > r$, we have $\mathbf{H}_t - \tilde{\mathbf{H}}_t = \mathbb{B}^{t-r} \left(\mathbf{H}_r - \tilde{\mathbf{H}}_r \right)$ and

$$\frac{\partial}{\partial \theta_i} \left(\mathbf{H}_t - \tilde{\mathbf{H}}_t \right) = \mathbb{B}^{t-r} \frac{\partial}{\partial \theta_i} \left(\mathbf{H}_r - \tilde{\mathbf{H}}_r \right) + \sum_{j=1}^{t-r} \mathbb{B}^{j-1} \mathbb{B}^{(i)} \mathbb{B}^{t-r-j} \left(\mathbf{H}_r - \tilde{\mathbf{H}}_r \right).$$

Thus (4.10) entails

$$\sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta_i} \left(D_t - \tilde{D}_t \right) \right\| \leq K\rho^t. \tag{4.35}$$

Because

$$D_t^{-1} - \tilde{D}_t^{-1} = D_t^{-1}\left(\tilde{D}_t - D_t\right)\tilde{D}_t^{-1},$$

we thus have (4.33), implying

$$\sup_{\theta\in\Theta}\left\|\left(D_t^{-1} - \tilde{D}_t^{-1}\right)\right\| \leq K\rho^t, \qquad \sup_{\theta\in\Theta}\left\|\left(D_t^{-1/2} - \tilde{D}_t^{-1/2}\right)\right\| \leq K\rho^t. \tag{4.36}$$

Denoting by $\underline{h}_{0t}(i_1)$ the i_1^{th} component of $\underline{h}_t(\theta_0)$,

$$\underline{h}_{0t}(i_1) \quad = \quad c_0 + \sum_{k=0}^{\infty}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\sum_{i=1}^{q}\mathbf{A}_{0i}(j_1,j_2)\mathbb{B}_0^k(i_1,j_1)\epsilon_{j_2,t-k-i}^2$$

where c_0 is a strictly positive constant and, with the usual convention, the index 0 corresponding to quantities evaluated at $\theta = \theta_0$. For a sufficiently small neighborhood $\mathcal{V}(\theta_0)$ of θ_0, we have

$$\sup_{\theta\in\mathcal{V}(\theta_0)}\frac{\mathbf{A}_{0i}(j_1,j_2)}{\mathbf{A}_i(j_1,j_2)} < K, \qquad \sup_{\theta\in\mathcal{V}(\theta_0)}\frac{\mathbb{B}_0^k(i_1,j_1)}{\mathbb{B}^k(i_1,j_1)} < (1+\delta)$$

for all $i_1, j_1, j_2 \in \{1,\ldots,m\}$ and all $\delta > 0$. Moreover, in $\underline{h}_t(i_1)$, the coefficient of $\mathbb{B}^k(i_1,j_1)\epsilon_{j_2,t-k-i}^2$ is bounded below by a constant $c > 0$ uniformly on $\theta \in \mathcal{V}(\theta_0)$. We thus have

$$\begin{aligned}\frac{\underline{h}_{0t}(i_1)}{\underline{h}_t(i_1)} \quad &\leq \quad K + K\sum_{k=0}^{\infty}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\sum_{i=1}^{q}\frac{(1+\delta)^k\mathbb{B}^k(i_1,j_1)\epsilon_{j_2,t-k-i}^2}{\omega + c\mathbb{B}^k(i_1,j_1)\epsilon_{j_2,t-k-i}^2}\\ &\leq \quad K + K\sum_{j_2=1}^{m}\sum_{i=1}^{q}\sum_{k=0}^{\infty}(1+\delta)^k\rho^{ks}\epsilon_{j_2,t-k-i}^{2s},\end{aligned}$$

for some $\rho \in [0,1)$, all $\delta > 0$ and all $s \in [0,1]$. Corollary 2.2 then implies that, for all $r_0 \geq 0$,

$$E\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\frac{\underline{h}_{0t}(i_1)}{\underline{h}_t(i_1)}\right|^{r_0} < \infty.$$

From this we deduce

$$E\sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1/2}\epsilon_t\|^2 \quad = \quad E\sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1/2}D_{0t}^{1/2}\tilde{\eta}_t\|^2 < \infty, \tag{4.37}$$

$$\sup_{\theta\in\mathcal{V}(\theta_0)}\|\tilde{D}_t^{-1/2}\epsilon_t\| \quad \leq \quad (1+K\rho^t)\sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1/2}\epsilon_t\|. \tag{4.38}$$

The last inequality follows from (4.33) because

$$\tilde{D}_t^{-1/2}\epsilon_t = \tilde{D}_t^{-1/2}\left(\tilde{D}_t^{1/2} - D_t^{1/2}\right) D_t^{-1/2}\epsilon_t - D_t^{-1/2}\epsilon_t.$$

By (4.23) and (4.24)

$$\frac{\partial \ell_t(\theta)}{\partial \theta_i} - \frac{\partial \tilde{\ell}_t(\theta)}{\partial \theta_i} = \mathrm{Tr}(c_1 + c_2 + c_3)$$

where

$$\begin{aligned} c_1 &= -D_t^{-1/2}\epsilon_t\epsilon_t'\tilde{D}_t^{-1}R^{-1}\left(D_t^{-1} - \tilde{D}_t^{-1}\right) D_t^{1/2} D_t^{-1/2}\frac{\partial D_t}{\partial \theta_i} D_t^{-1/2}, \\ c_2 &= -D_t^{-1/2}\epsilon_t\epsilon_t'\tilde{D}_t^{-1}R^{-1}\tilde{D}_t^{-1}\left(\frac{\partial D_t}{\partial \theta_i} - \frac{\partial \tilde{D}_t}{\partial \theta_i}\right) D_t^{-1/2} \end{aligned}$$

and c_3 contains terms which can be handled as c_1 and c_2. Using (4.33)–(4.38), the Cauchy-Schwarz inequality, and

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left\| D_t^{-1/2}\frac{\partial D_t}{\partial \theta_i} D_t^{-1/2} \right\|^2 < \infty,$$

which follows from (4.27), we obtain

$$\sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{\partial \ell_t(\theta)}{\partial \theta_i} - \frac{\partial \tilde{\ell}_t(\theta)}{\partial \theta_i} \right| \leq K\rho^t \mathrm{u}_t,$$

where u_t is an integrable variable. From the Markov inequality, $n^{-1/2}\sum_{t=1}^n \rho^t \mathrm{u}_t = o_P(1)$, which implies

$$\left\| n^{-1/2}\sum_{t=1}^n \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{\ell}_t(\theta_0)}{\partial \theta} \right\} \right\| = o_P(1).$$

We have in fact shown that this convergence is uniform on a neighborhood of θ_0, but this is not directly useful for what follows. By exactly the same arguments,

$$\sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{\ell}_t(\theta)}{\partial \theta_i \partial \theta_j} \right| \leq K\rho^t \mathrm{u}_t^*,$$

where u_t^* is an integrable random variable, which entails

$$n^{-1/2}\sum_{t=1}^n \sup_{\theta\in\mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial \tilde{\ell}_t^2(\theta)}{\partial \theta \partial \theta'} \right\| = O_P(n^{-1}) = o_P(1).$$

It now suffices to observe that the analogous of the steps a)-f) in Section have been verified, which allows to conclude.

References

Aue, A., Hörmann, S., Horváth, L. and M. Reimherr (2009) Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* 37, 4046–4087.

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

Berkes, I., Horváth, L. and P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

Billingsley, P. (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–327.

Bougerol, P. and N. Picard (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

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

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.

Engle, R.F. (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.

Engle, R.F. and K. Sheppard (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, University of California San Diego, Discussion paper.

Francq, C. and J-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Hafner, C. and A. Preminger (2009a) On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* 100, 2044–2054.

Hafner, C. and A. Preminger (2009b) 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.

Ling, S. and M. McAleer (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

Nakatani, T. and T. Teräsvirta (2009) Testing for volatility interactions in the Constant Conditional Correlation GARCH model. *Econometrics Journal* 12, 147–163.

Reinsel, G.C. (1997) *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag.

Silvennoinen, A. and T. Teräsvirta (2009) Multivariate GARCH models. *Handbook of Financial Time Series* T.G. Andersen, R.A. Davis, J-P. Kreiss and T. Mikosch, eds. New York: Springer.

BP 60149
59653 Villeneuve d'Ascq cedex
France
E-mail: christian.francq@univ-lille3.fr

15 Boulevard Gabriel Peri
92245 Malakoff Cedex
France
E-mail: zakoian@ensae.fr