QML ESTIMATION OF A CLASS OF MULTIVARIATE ASYMMETRIC GARCH MODELS

By Christian Francq

CREST (CNRS) and University Lille 3 (EQUIPPE)

and

By Jean-Michel Zakoïan

CREST and University Lille 3 (EQUIPPE) *

We establish the strong consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE) of the parameters of a class of multivariate asymmetric GARCH processes, allowing for cross leverage effects. The conditions required to establish the asymptotic properties of the QMLE are mild and coincide with the minimal ones in the univariate case. In particular, no moment assumption is made on the observed process. Instead, we require strict stationarity, for which a necessary and sufficient condition is established. The asymptotic results are illustrated by Monte Carlo experiments and an application to a bivariate exchange rates series is proposed.

1. Introduction. Since the inception of the univariate ARCH and GARCH models by Engle (1982) and Bollerslev (1986), 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

*The authors sincerely thank the Editor Peter C.B. Phillips, the Co-Editor Pentti Saikkonen and two anonymous referees for their very helpful comments and suggestions.

consistency of the QMLE for multivariate GARCH models. Comte and Lieberman (2003) showed the consistency and the asymptotic normality (CAN) of the Quasi Maximum Likelihood Estimator (QMLE) for the BEKK formulation. Asymptotic results were established by Ling and McAleer (2003) for the Constant Conditional Correlation (CCC) formulation of an ARMA-GARCH. These volatility models reduce to the standard GARCH in the univariate case, and thus do not capture an important characteristic of financial series: the so-called leverage effect. Although other asymmetric effects may also be of interest, the leverage effect is by far the most documented one in the empirical literature. Typically, negative shocks tend to have more impact on volatility then positive shocks of the same magnitude.

Actually, while many asymmetric univariate GARCH models have been considered in the literature to capture the leverage effect, extensions to the multivariate setting have not been much developed. By extending the GJR model (Glosten, Jaganathan and Runkle, 1993), the CCC Asymmetric GARCH (AGARCH) model - recently studied by McAleer, Hoti and Chan (2009) - allows for asymmetric impacts of past returns on the current volatility. In this multivariate extension, positive and negative values of any component of a vector of assets are allowed to impact differently the volatilities (and co-volatilities) of all assets. Similarly, multivariate extensions of the Exponential GARCH (EGARCH) model introduced by Nelson (1991) have been proposed (see for instance Koutmos and Booth (1995)). Bardet and Wintenberger (2009) studied the asymptotic behaviour of the QMLE in a general class of multidimensional causal processes allowing for asymmetries. Another interesting extension is the Generalized Autoregressive Conditional Correlation (GARCC) model, recently proposed and analyzed by McAleer, Chan, Hoti and Liebermann (2009).

In this article, we study the estimation of the CCC-AGARCH, which includes the CCC-GARCH(p, q) introduced by Bollerslev (1990), and its generalization by Jeantheau (1998). The CCC-GARCH is undoubtedly one of the most popular class of multivariate GARCH models. The attractiveness of the CCC- and CCC-

AGARCH follows from their tractability: i) the number of unknown coefficients is less than in other specifications, and remains tractable in small dimension; ii) the coefficients are easy to interpret; iii) the conditions ensuring positive definiteness of the conditional variance are simple and explicit. Moreover, as we will see, the strict stationarity conditions are explicit too.

In all above-mentioned references on multivariate GARCH estimation, 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, CAN results for multivariate GARCH without moments restriction have only been established by Hafner and Preminger (2009), 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 CAN of the QMLE. The most significant breakthrough in this direction was the paper by Berkes, Horváth and Kokoszka (2003), although slightly weaker conditions have been obtained by Francq and Zakoian (2004). [1]

The main contribution of this article is to provide asymptotic results for the class of CCC-AGARCH models without moment assumptions on the observed process.

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. Section 4 presents a few Monte Carlo experiments and an application to a bivariate exchange rates series. Section 5 concludes. Proofs are relegated to an appendix.

[1] Extensions of these results to asymmetric univariate GARCH (see Pan, Wang and Tong (2008), Hamadeh and Zakoian (2011)) or to more general univariate GARCH (see Meitz and Saikkonen (2010)) exist. For some specifications, however, the asymptotic properties of the QMLE may be difficult to establish. In particular, Straumann and Mikosch (2006) established asymptotic results for the EGARCH(0,1) but suggest that the extension to the general EGARCH model may not be possible (see p. 2452).

2. Model and strict stationarity condition. The m-dimensional process $\{\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{mt})'\}$ is called a CCC-AGARCH(p,q) if it verifies

$$\begin{cases} \epsilon_t &= H_t^{1/2}\eta_t, \\ H_t &= D_t R_0 D_t, \quad D_t = \mathrm{diag}\left(\sqrt{h_{11,t}}, \dots, \sqrt{h_{mm,t}}\right), \\ \underline{h}_t &= \underline{\omega}_0 + \sum_{i=1}^{q} \mathbf{A}_{0i,+}\underline{\epsilon}_{t-i}^{+} + \mathbf{A}_{0i,-}\underline{\epsilon}_{t-i}^{-} + \sum_{j=1}^{p} \mathbf{B}_{0j}\underline{h}_{t-j}, \end{cases} \tag{2.1}$$

where $\underline{h}_t = (h_{11,t}, \dots, h_{mm,t})'$ and (with $x^+ = \max(x,0) = (-x)^-$)

$$\underline{\epsilon}_t^{+} = \left(\left\{\epsilon_{1t}^{+}\right\}^2, \cdots, \left\{\epsilon_{mt}^{+}\right\}^2\right)', \quad \underline{\epsilon}_t^{-} = \left(\left\{\epsilon_{1t}^{-}\right\}^2, \cdots, \left\{\epsilon_{mt}^{-}\right\}^2\right)',$$

R_0 is a correlation matrix, $\underline{\omega}_0$ is a vector of size $m \times 1$ with strictly positive coefficients, the $\mathbf{A}_{0i,+}$, $\mathbf{A}_{0i,-}$ and $\mathbf{B}_{0j}$ are matrices of size $m \times m$ with positive coefficients, and (η_t) is an iid sequence of variables on $\mathbb{R}^m$ with identity covariance matrix. Note that the standard assumption $E\eta_t = 0$ allows to interpret H_t as the volatility (*i.e.* the conditional variance) of ϵ_t, but this assumption is not required in the sequel.

A submodel of (2.1) of particular interest is the popular CCC-GARCH(p,q) model for which $\mathbf{A}_{0i,+} = \mathbf{A}_{0i,-}$, and thus

$$\underline{h}_t = \underline{\omega}_0 + \sum_{i=1}^{q} \mathbf{A}_{0i}\underline{\epsilon}_{t-i} + \sum_{j=1}^{p} \mathbf{B}_{0j}\underline{h}_{t-j}, \qquad \underline{\epsilon}_t = \left(\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2\right)'. \tag{2.2}$$

The main interest of the AGARCH version (2.1) is that it allows for the so-called leverage effect, or volatility asymmetry, observed for most financial series. Choosing $\mathbf{A}_{0i,-} > \mathbf{A}_{0i,+}$ element by element, the present volatility $\underline{h}_t$ is sensitive to the sign of the past returns, in the sense that the $\underline{h}_{kk,t}$'s are higher when $\epsilon_{j,t-i} = -c < 0$ than when $\epsilon_{j,t-i} = c > 0$ for some $i > 0$. In other words, the AGARCH allows for a higher increase of the volatility after a bad news (a negative return) than after a good news implying a return of the same magnitude.

The CCC-GARCH model was introduced by Bollerslev (1990) in a simplest version, assuming that the matrices $\mathbf{A}_i$ and $\mathbf{B}_{0j}$ are diagonal. By contrast, in (2.1)-(2.2) 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 (2.2) 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 and for its extension (2.1).

Write

$$\epsilon_t = D_t \tilde{\eta}_t, \quad \text{where} \quad \tilde{\eta}_t = (\tilde{\eta}_{1t}, \ldots, \tilde{\eta}_{mt}) = R_0^{1/2} \eta_t \tag{2.3}$$

and

$$\underline{\epsilon}_t^+ = \Upsilon_t^+ \underline{h}_t, \quad \underline{\epsilon}_t^- = \Upsilon_t^- \underline{h}_t, \quad \text{where} \quad \Upsilon_t^\circ = \operatorname{diag}\left\{(\tilde{\eta}_{1t}^\circ)^2, \ldots, (\tilde{\eta}_{mt}^\circ)^2\right\}. \tag{2.4}$$

Introducing the $m \times pm$ matrix $\mathbf{B}_{01:p} = (\mathbf{B}_{01} \cdots \mathbf{B}_{0p})$, and similar other notations, let the $(p + 2q)m \times (p + 2q)m$ matrix

$$C_t = \begin{pmatrix} \Upsilon_t^+ \mathbf{A}_{01:q,+} & \Upsilon_t^+ \mathbf{A}_{01:q,-} & \Upsilon_t^+ \mathbf{B}_{01:p} \\ I_{(q-1)m} & \multicolumn{2}{c}{0_{(q-1)m \times (p+q+1)m}} \\ \Upsilon_t^- \mathbf{A}_{01:q,+} & \Upsilon_t^- \mathbf{A}_{01:q,-} & \Upsilon_t^- \mathbf{B}_{01:p} \\ 0_{(q-1)m \times qm} & I_{(q-1)m} & 0_{(q-1)m \times (p+1)m} \\ \mathbf{A}_{01:q,+} & \mathbf{A}_{01:q,-} & \mathbf{B}_{01:p} \\ 0_{(p-1)m \times 2qm} & I_{(p-1)m} & 0_{(p-1)m \times m} \end{pmatrix} \tag{2.5}$$

Let $\gamma(\mathbf{C_0})$ be the top Lyapunov exponent of the sequence $\mathbf{C_0} = \{C_t, t \in \mathbb{Z}\}$,

$$\gamma(\mathbf{C_0}) = \lim_{t \to \infty} \frac{1}{t} E\left(\log \|C_t C_{t-1} \ldots C_1\|\right) = \inf_{t \geq 1} \frac{1}{t} E(\log \|C_t C_{t-1} \ldots C_1\|).$$

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* $\gamma(\mathbf{C_0}) < 0$. *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, and by $\varrho(A)$ its spectral radius, that is, the greatest modulus of its eigenvalues.

COROLLARY 2.1. *Let the matrix polynomial defined by:* $\mathcal{B}(z) = I_m - z\mathbf{B}_{01} - \dots - z^p\mathbf{B}_{0p}, \quad z \in \mathbb{C}$. *Let*

$$\mathbb{B}_0 = \begin{pmatrix} & \mathbf{B}_{01:p} & \\ I_{(p-1)m} & & 0_{(p-1)m\times 1} \end{pmatrix}.$$

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. $\varrho(\mathbb{B}_0) < 1$.

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 vector $\underline{\omega}_0$, of the matrices $\mathbf{A}_{0i,+}$, $\mathbf{A}_{0i,-}$ and $\mathbf{B}_{0j}$, and of the coefficients of the lower triangular part (excluding the diagonal) of the correlation matrix $R_0 = (r_{0ij})$. The number of unknown parameters is thus

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

The parameter space is a sub-space Θ of

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

A generic element of Θ is denoted by

$$\begin{aligned}\theta = (\theta_1, \ldots, \theta_{s_0})' &= (\underline{\omega}_0', \alpha_{1,+}', \ldots, \alpha_{q,+}', \alpha_{1,-}', \ldots, \alpha_{q,-}', \beta_1', \ldots, \beta_p', r')' \\ &:= (\underline{\omega}_0', \alpha_+', \alpha_-', \beta', r')',\end{aligned}$$

where $r' = (r_{21}, \ldots, r_{m1}, r_{32}, \ldots, r_{m2}, \ldots, r_{m,m-1})$, $\alpha_{i,+} = \mathrm{vec}(\mathbf{A}_{i,+})$, $\alpha_{i,-} = \mathrm{vec}(\mathbf{A}_{i,-})$, $i = 1, \ldots, q$, and $\beta_j = \mathrm{vec}(\mathbf{B}_j)$, $j = 1, \ldots, p$. We denote by R the symmetric matrix with 1 on the diagonal and the r_{ij} in the lower triangular part. It is assumed that R is positive definite for any $\theta \in \Theta$. The true parameter value is denoted

$$\theta_0 = (\underline{\omega}_0', \alpha_{01,+}', \ldots, \alpha_{0q,-}', \beta_{01}', \ldots, \beta_{0p}', r_0')' = (\underline{\omega}_0', \alpha_{0,+}', \alpha_{0,-}', \beta_0', r_0')'.$$

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

3.1. *Identifiability Conditions.* Let $\mathcal{A}_\theta^+(z) = \sum_{i=1}^q \mathbf{A}_{i,+} z^i$, $\mathcal{A}_\theta^-(z) = \sum_{i=1}^q \mathbf{A}_{i,-} z^i$ and $\mathcal{B}_\theta(z) = I_m - \sum_{j=1}^p \mathbf{B}_j z^j$. By convention, $\mathcal{A}_\theta^\circ(z) = 0$ if $q = 0$ and $\mathcal{B}_\theta(z) = I_m$ if $p = 0$. If the roots of $\det(\mathcal{B}_{\theta_0}(z)) = 0$ are outside the unit disk, we have

$$\underline{h}_t = \mathcal{B}_{\theta_0}(1)^{-1}\underline{\omega}_0 + \mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}^+(B)\underline{\epsilon}_t^+ + \mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}^-(B)\underline{\epsilon}_t^- . \tag{3.1}$$

The parameter θ_0 is said to be identifiable if (3.1) does not hold true when θ_0 is replaced by $\theta \neq \theta_0$ belonging to Θ.

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 $\mathcal{A}_\theta^+(B)$, $\mathcal{A}_\theta^-(B)$ and $\mathcal{B}_\theta(B)$, the maximal degrees $q_i^+(\theta)$, $q_i^-(\theta)$ and $p_i(\theta)$, respectively. Suppose that these maximal values are imposed for these orders, that is

$$\forall \theta \in \Theta,\ \forall i = 1, \ldots, m, \quad q_i^+(\theta) \le q_i^+, \qquad q_i^-(\theta) \le q_i^- \quad \text{and} \quad p_i(\theta) \le p_i \tag{3.2}$$

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

PROPOSITION 3.1 (A simple identifiability condition). *If the matrix polynomials $\mathcal{A}_{\theta_0}^+(z)$, $\mathcal{A}_{\theta_0}^-(z)$ and $\mathcal{B}_{\theta_0}(z)$ are left-coprime, $\mathcal{A}_{\theta_0}^+(1)+\mathcal{A}_{\theta_0}^-(1)\neq 0$ and if the matrix*

$$M(\mathcal{A}_{\theta_0}^+, \mathcal{A}_{\theta_0}^-, \mathcal{B}_{\theta_0}) = \left[a_{q_1^+}^+(1)\cdots a_{q_m^+}^+(m)\;\; a_{q_1^-}^-(1)\cdots a_{q_m^-}^-(m)\;\; b_{p_1}(1)\cdots b_{p_m}(m)\right]$$

has full rank m, under the constraints (3.2) with $q_i^+ = q_i^+(\theta_0)$, $q_i^- = q_i^-(\theta_0)$ and $p_i = p_i(\theta_0)$ for any value of i, then

$$\begin{cases} \mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta^+(B) = \mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}^+(B) \\ \mathcal{B}_\theta(B)^{-1}\mathcal{A}_\theta^-(B) = \mathcal{B}_{\theta_0}(B)^{-1}\mathcal{A}_{\theta_0}^-(B) \end{cases} \Rightarrow (\mathcal{A}_\theta^+, \mathcal{A}_\theta^-, \mathcal{B}_\theta) = (\mathcal{A}_{\theta_0}^+, \mathcal{A}_{\theta_0}^-, \mathcal{B}_{\theta_0}).$$

Proof. Clearly the parameter θ_0 is identifiable if there exists no triplet $(\mathcal{A}_\theta^+, \mathcal{A}_\theta^-, \mathcal{B}_\theta) \neq (\mathcal{A}_{\theta_0}^+, \mathcal{A}_{\theta_0}^-, \mathcal{B}_{\theta_0})$, subject to the constraints (3.2) , such that This condition is equivalent to the existence of an operator $U(B)$ such that

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

From the proof of the theorem in Hannan (1969), the matrix $U(B)$ is unimodular because $\mathcal{A}_{\theta_0}^+(z)$, $\mathcal{A}_{\theta_0}^-(z)$ and $\mathcal{B}_{\theta_0}(z)$ are supposed to be left-coprime. Let $U(B) = U_0 + U_1B + \ldots + U_kB^k$. The i-th 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 the constraints (3.2) entail $U_j a_{q_i^-}^-(i) = 0$ and $U_j b_{p_i}(i) = 0$, for $j = 1, \ldots, k$ and $i = 1, \ldots m$. It follows that $U_j M(\mathcal{A}_{\theta_0}^+, \mathcal{A}_{\theta_0}^-, \mathcal{B}_{\theta_0}) = 0$, which implies $U_j = 0$ for $j = 1, \ldots, k$. Consequently $U(B) = U_0$ and, since $\mathcal{B}_\theta(0) = I_m$ for all θ, we have $U(B) = I_m$. □

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

$$M_1(\mathcal{A}_{\theta_0}^+, \mathcal{A}_{\theta_0}^-, \mathcal{B}_{\theta_0}) = [\mathbf{A}_{0q}^+ \; \mathbf{A}_{0q}^- \;\; \mathbf{B}_{0p}]$$

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

3.2. *Asymptotic Properties of the QML Estimator of the CCC-AGARCH.* 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

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

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

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\min_{\theta \in \Theta} \tilde{\mathbf{l}}_n(\theta). \tag{3.3}$$

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: For $i = 1, \dots, m$ the distribution of $\tilde{\eta}_{it}$ is not concentrated on 2 points and $P(\tilde{\eta}_{it} > 0) \in (0, 1)$.

A4: If $p > 0$, $\mathcal{A}^{+}_{\theta_0}(1) + \mathcal{A}^{-}_{\theta_0}(1) \neq 0$, $\mathcal{A}^{+}_{\theta_0}(z)$, $\mathcal{A}^{-}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ are left coprime and $M_1(\mathcal{A}^{+}_{\theta_0}, \mathcal{A}^{-}_{\theta_0}, \mathcal{B}_{\theta_0})$ has full rank m.

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

If the space Θ is constrained by (3.2), that is if maximal orders are imposed for each component of $\underline{\epsilon}_t^{+}$, $\underline{\epsilon}_t^{-}$ and $\underline{h}_t$ in each equation, Assumption **A4** can be replaced by the more general condition:

A4': **A4** with $M_1(\mathcal{A}^+_{\theta_0}, \mathcal{A}^-_{\theta_0}, \mathcal{B}_{\theta_0})$ replaced $M(\mathcal{A}^+_{\theta_0}, \mathcal{A}^-_{\theta_0}, \mathcal{B}_{\theta_0})$.

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

THEOREM 3.1 (Strong consistency). *Let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (3.3). Then, under* **A1**-**A5** *(or* **A1**-**A4'**-**A5***),*

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

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

$$\underline{h}_t = \underline{\omega} + \sum_{i=1}^{q} \mathbf{A}_{i,+}\underline{\epsilon}^+_{t-i} + \mathbf{A}_{i,-}\underline{\epsilon}^-_{t-i} + \sum_{j=1}^{p} \mathbf{B}_j \underline{h}_{t-j}, \qquad \forall t. \tag{3.4}$$

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

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

Since there is no risk of confusion, $\underline{h}_t$, D_t and H_t now denote quantities evaluated at θ whereas these quantities were evaluated at θ_0 in (2.1).

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 positive semi-definite matrix, defined by*

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

REMARK 3.1 (On the assumptions in the univariate case). It is worth noting that the conditions ensuring the CAN are mild. When $m = 1$ and $\mathbf{A}_{i,+} = \mathbf{A}_{i,-}$ for all i, they reduce to those obtained by Francq and Zakoian (2004) for the standard GARCH case (except for **A3**, see Remark 3.4 below) and Hamadeh and Zakoian (2011) for the asymmetric GARCH models.

REMARK 3.2 (On the moment assumptions). No assumption is made concerning the existence of moments of the observed process. Corollary 2.2 however entails the existence of small-order moments, under the strict stationarity condition. The only moment assumption, **A7**, concerns the iid process and is required for the existence of the information matrix I. Moreover, it should be noted that for GARCH models the tail properties of the observed processes are very different from those of the errors: for instance the marginal distribution of a standard GARCH process with Gaussian noise has Pareto tails (see Basrak, Davis, and Mikosch, 2002).

REMARK 3.3 (Comparisons with assumptions used for other models). An advantage of the CCC model is that the strict stationarity assumption (first part of **A2**) and the identifiability conditions **A3-A4** are explicit. By comparison, for other models such assumptions have to be made in a non explicit way (for the BEKK model see Comte and Liebermann (2003), Assumptions **A1** and **A4**; for factor models see Hafner and Preminger (2009), Assumption **4.4**).

REMARK 3.4 (On the identifiability condition). The iid process $(\tilde{\eta}_t)$ is not supposed to be centered, but Assumption **A3** imposes that its components take values of both signs with a positive probability. Indeed, if for instance $\tilde{\eta}_{it} > 0$ with probability 1, then $\epsilon_{it} = h_{ii,t}^{1/2}\tilde{\eta}_{it} > 0$ and thus the coefficients of the ith row of any matrix $\mathbf{A}_{0j,-}$ cannot be identified. When no asymmetry is present in the model, as in the standard univariate GARCH, Assumption **A3** can be weakened.

4. Numerical illustrations.

4.1. *A Monte Carlo experiment.* To assess the performance of the QMLE in finite samples, we computed the estimator on 100 independent simulated trajecto-

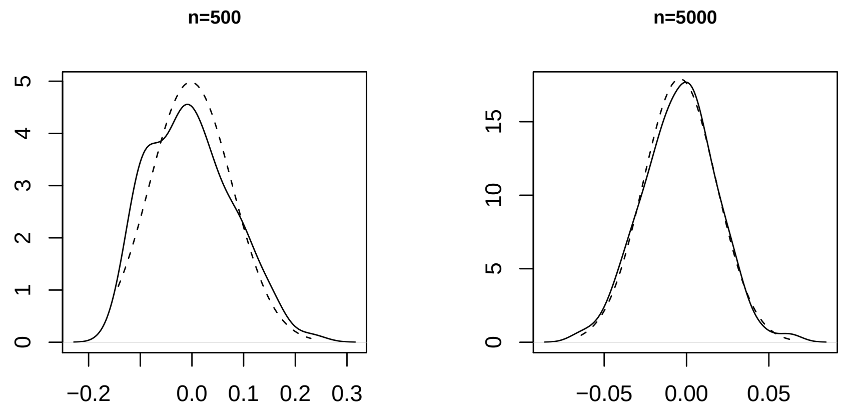

FIG 1. *Kernel density estimator (in full line) of the distribution of the QMLE errors for the estimation of* $\mathbf{A}_{01,+}(2,2)$*, and gaussian density (in dotted line) with the same mean and variance.*

ries of length $n = 500$ and $n = 5,000$ of a bivariate CCC-AGARCH of order $q = 1$ and $p = 0$, with $\eta_t \sim \mathcal{N}(0, I_2)$.

The implementation of the QMLE was made in R, with the numerical optimization procedure `nlminb()`. The code is available from the authors. The parameters of the simulated model are given in the second column of Table 1. With an appropriate choice of Θ, the assumptions of Theorems 3.1 and 3.2 are satisfied.

Table 1 summarizes the distribution of the QMLE over these simulations experiments. As expected, bias and RMSE decrease as the sample size increases. The empirical distribution of the QMLE is particularly well approximated by a Gaussian for the component $\mathbf{A}_{01,+}(2,2)$, at least for $n = 5,000$ (see Figure 1). For the other components of θ_0, the normal approximation is also satisfactory (see for instance the estimation of $R_0(1,2)$ in Figure 2). Results not reported here, however, show that for parameter values approaching the non stationarity region, the normal approximation deteriorates.

Table 1
Sampling distribution of the QMLE of θ_0 for the CCC-AGARCH(0,1) model (2.1).

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
			$n = 500$					
$\underline{\omega}_0$	1.00	0.01544	0.113	0.765	0.936	0.994	1.092	1.36
	1.00	-0.00532	0.125	0.606	0.912	0.979	1.065	1.45
$\mathrm{vec}\mathbf{A}_{01,+}$	0.25	-0.00250	0.071	0.100	0.199	0.246	0.291	0.48
	0.05	-0.00099	0.037	0.000	0.021	0.044	0.067	0.18
	0.05	-0.00023	0.044	0.000	0.015	0.045	0.067	0.18
	0.25	-0.00224	0.080	0.107	0.179	0.239	0.295	0.48
$\mathrm{vec}\mathbf{A}_{01,-}$	0.50	-0.01905	0.129	0.191	0.400	0.455	0.551	0.90
	0.50	0.00522	0.142	0.094	0.421	0.490	0.594	0.87
	0.50	0.00447	0.122	0.275	0.407	0.498	0.584	0.79
	0.50	0.00471	0.149	0.213	0.401	0.478	0.605	1.00
$R_0(1,2)$	0.50	-0.00667	0.044	0.375	0.460	0.492	0.528	0.58
			$n = 5,000$					
$\underline{\omega}_0$	1.00	-0.00513	0.039	0.877	0.975	0.994	1.015	1.095
	1.00	-0.00018	0.040	0.919	0.973	0.999	1.023	1.133
$\mathrm{vec}\mathbf{A}_{01,+}$	0.25	0.00173	0.024	0.182	0.235	0.252	0.267	0.321
	0.05	-0.00195	0.014	0.011	0.040	0.047	0.056	0.085
	0.05	0.00194	0.013	0.020	0.042	0.052	0.061	0.081
	0.25	-0.00409	0.023	0.186	0.231	0.247	0.259	0.313
$\mathrm{vec}\mathbf{A}_{01,-}$	0.50	-0.00132	0.040	0.404	0.472	0.498	0.523	0.597
	0.50	0.00079	0.040	0.398	0.477	0.496	0.524	0.592
	0.50	0.00434	0.042	0.399	0.475	0.501	0.532	0.609
	0.50	0.00654	0.044	0.413	0.475	0.508	0.536	0.619
$R_0(1,2)$	0.50	-0.00100	0.012	0.467	0.490	0.499	0.505	0.531

RMSE is the Root Mean Square Error, Q_i, $i = 1, 3$, denote the quartiles.

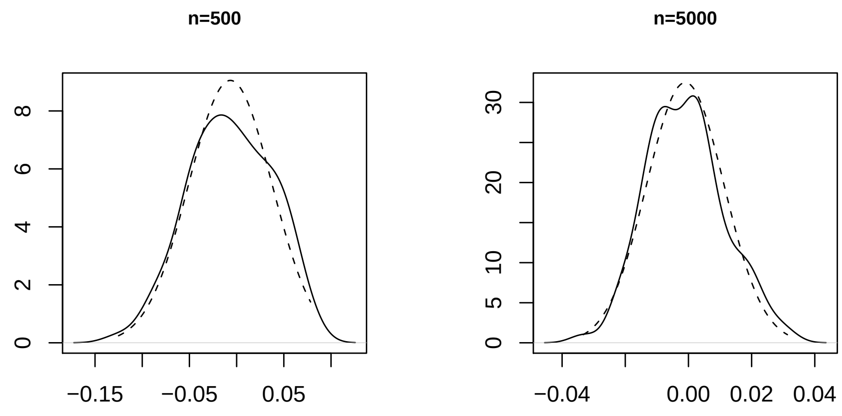

Fig 2. *Kernel density estimator (in full line) of the distribution of the QMLE errors for the estimation of $R_0(1,2)$, and gaussian density (in dotted line) with the same mean and variance.*

4.2. *An application to exchange rates.* We now consider the bivariate series of the returns of the daily exchange rates of the Dollar (USD) and the Yen (JPY) with respect to the Euro. The observations cover the period from January 5, 1999 to September 17, 2010, which corresponds to 2997 observations. The data were obtained from the web site of the National Bank of Belgium. Several CCC-AGARCH(1,1) models of the form (2.1) have been estimated by QML. The model fitted to the observations of the entire period, is given in the second column of Table 2. The following observations can be made: i) a strong leverage effect of the JPY with respect to its own past return; ii) almost no leverage effect in the volatility of the USD; iii) the diagonal form of the persistence matrix $\mathbf{B}_{01}$; iv) a strong positive correlation between the components of the iid noise.

Observe that the estimated parameter $\hat{\theta}_n$ is at the boundary of the parameter space Θ. Indeed, several components of $\hat{\theta}_n$ are equal to zero: the coefficients of USD_{t-1}^- and $h_{\text{USD},t-1}$ in $h_{\text{JPY},t}$, and the coefficient of $h_{\text{JPY},t-1}$ in $h_{\text{USD},t}$. Thus, one can suspect that Assumption **A6** is not satisfied, and thus that Theorem 3.2 does not apply. Indeed, the asymptotic distribution of $\hat{\theta}_n$ is not gaussian when the parameter θ_0 belongs to the boundary of the parameter space (see Andrews (2001) and Francq and Zakoïan (2009)). It is worth noting that the non-gaussianity of the asymptotic distribution does not only concern the components which reach the boundary (see Figure 1 in Francq and Zakoïan (2007)). For this reason, we do not provide standard errors for the estimates of this table.

To give an idea of the reliability of the parameter estimates obtained from the whole period of 2997 observations, the period has been divided into 3 subperiods (of 999 observations each), and the model has been reestimated. The models fitted from these three subperiods are also given in Table 2. The remarks concerning the diagonal form of the persistence matrix, the important leverage effect of JPY, and the instantaneous correlation of the noises remain valid over the different subperiods.

Table 2
Fitted CCC-AGARCH(1,1) model on the bivariate series of exchange rates of $(USD_t, JPY_t)'$ against the Euro.

	Full Period	First Period	Second Period	Third Period
$\underline{\omega}_0$	0.0125	0.2188	0.0643	0.0140
	0.0159	0.2238	0.0359	0.0149
$\mathrm{vec}\mathbf{A}_{01,+}$	0.0395	0.1000	0.0000	0.0869
	0.0359	0.0000	0.0275	0.1020
	0.0012	0.0147	0.0000	0.0421
	0.0434	0.0411	0.0621	0.0000
$\mathrm{vec}\mathbf{A}_{01,-}$	0.0215	0.0000	0.0000	0.0290
	0.0000	0.0000	0.0000	0.0032
	0.0099	0.0109	0.0000	0.0000
	0.2294	0.2439	0.1264	0.1367
$\mathrm{vec}\mathbf{B}_{01}$	0.9306	0.4842	0.7618	0.8768
	0.0000	0.0000	0.0000	0.0493
	0.0000	0.0000	0.0571	0.0005
	0.7514	0.5656	0.7654	0.8525
$R_0(1,2)$	0.5581	0.6490	0.4932	0.5573

The full period corresponds to observations from 1999-01-05 to 2010-09-16;
the first period runs from 1999-01-05 to 2002-11-26,
the second from 2002-11-27 to 2006-10-18,
and the third from 2006-10-19 to 2010-09-16.

5. Conclusion. In this paper, we investigated the asymptotic properties of the QMLE for the CCC-AGARCH(p, q) model. The conditions required for the CAN of the QMLE are natural extensions to those used to prove the same properties in the symmetric univariate setting. In particular, moment conditions are not needed. Instead, we required the strict stationarity, for which we established an explicit necessary and sufficient condition. These results also apply to the standard CCC-GARCH models.

Of course, more sophisticated classes of models can be seen as more realistic than the CCC-(A)GARCH. One could think of asymmetric extensions of the Dynamic Conditional Correlation (DCC) model introduced by Engle (2002), and studied by Engle and Sheppard (2001) and Nakatani and Teräsvirta (2009). Such extensions have been studied in Cappiello, Engle and Sheppard (2006), and employed to analyze the behavior of international equities and government bonds. See the recent textbook by Engle (2009) for a detailed analysis of DCC models. 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.

References

Andrews, D. W. K. (2001) Testing when a parameter is on a boundary of the maintained hypothesis. *Econometrica* 69, 683–734.

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.

Bardet, J-M. and O. Wintenberger (2009) Asymptotic normality of the Quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* 37, 2730–2759.

Basrak, B., Davis, R.A., and T. Mikosch (2002) Regular Variation of GARCH Processes. *Stochastic Processes and their Applications* 99, 95–115.

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.

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

Bougerol, P. and N. Picard (1992a) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.

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

Cappiello, L., Engle, R.F. and K. Sheppard (2006) Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 4, 537–572.

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. (2009) *Anticipating Correlations. A new paradigm for risk management.* Princeton University Press.

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.

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

Francq, C. and J-M. Zakoïan (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.

Glosten, L.R., Jaganathan, R. and D. Runkle (1993) On the relation between the expected values and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

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

Hamadeh, T and J-M. Zakoïan (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141, 488–507

Hannan, E. J. (1969) The Identification of Vector Mixed Autoregressive-Moving Average Systems. *Biometrika* 56, 223–225.

Harville, D. (1997) *Matrix algebra from a statistician's perspective.* New York: Springer.

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.

Koutmos, G. and G. G. Booth (1995) Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance* 14, 747–762.

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

Magnus, J.R. and H. Neudecker (1988) *Matrix differential calculus with applications in statistics and econometrics*. New York: John Wiley & Sons.

McAleer, M., Hoti, S. and F. Chan (2009) Structure and asymptotic theory for multivariate asymmetric conditional volatility. *Econometric Reviews* 28, 422–440.

McAleer, M., Chan, F., Hoti, S. and O. Liebermann (2009) Generalized autoregressive conditional correlation. *Econometric Theory* 28, 422–440.

Meitz, M. and P. Saikkonen (2010) Parameter estimation in nonlinear AR-GARCH models. Forthcoming *Econometric Theory.*

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

Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.

Pan, J., Wang, H., and H. Tong (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics*, 142, 352-378.

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.

Straumann, D. and T. Mikosch (2006) Quasi-maximum-likelihood estimation in heteroscedastic time series: a stochastic recurrence equation approach. *The Annals of Statistics* 34, 2449–2495.

APPENDIX: PROOFS

A.1. Proof of Theorem 2.1. In view of (2.4) and (2.1), we have

$$\underline{z}_t = \underline{b}_t + C_t \underline{z}_{t-1}, \tag{A.1}$$

where, with obvious notations, $\underline{z}_t = \left(\underline{\epsilon}^{+'}_{t:(t-q+1)}, \underline{\epsilon}^{-'}_{t:(t-q+1)}, \underline{h}'_{t:(t-p+1)}\right)'$ and $\underline{b}_t = \left(\underline{\omega}'_0 \Upsilon^+_t, 0'_{m(q-1)}, \underline{\omega}'_0 \Upsilon^-_t, 0'_{(q-1)m}, \underline{\omega}'_0, 0'_{(p-1)m}\right)'$.

Bougerol and Picard (1992a, 1992b) showed that $\gamma(\mathbf{C_0}) < 0$ is the necessary and sufficient condition for the existence of a stationary solution to an *irreducible* equation of the form (A.1), and showed that the condition is necessary and sufficient in the univariate GARCH(p, q) case. When $\gamma(\mathbf{C_0}) < 0$, Cauchy's root test shows that the series

$$\underline{\tilde{z}}_t = \underline{b}_t + \sum_{n=0}^{\infty} C_t C_{t-1} \cdots C_{t-n} \underline{b}_{t-n-1} \tag{A.2}$$

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

To prove uniqueness, let $(\underline{z}_t)$ denote a positive and strictly stationary solution of (A.1). For all $N \geq 0$,

$$\underline{z}_t = \underline{\tilde{z}}_t(N) + C_t \ldots C_{t-N} \underline{z}_{t-N-1}, \quad \underline{\tilde{z}}_t(N) = \underline{b}_t + \sum_{n=0}^{N} C_t C_{t-1} \cdots C_{t-n} \underline{b}_{t-n-1}$$

Then

$$\|\underline{z}_t - \tilde{\underline{z}}_t\| \leq \|\tilde{\underline{z}}_t(N) - \tilde{\underline{z}}_t\| + \|C_t \dots C_{t-N}\| \|\underline{z}_{t-N-1}\|.$$

The first term in the right-hand side tends to 0 a.s. when $N \to \infty$. In addition, because the series defining $\tilde{\underline{z}}_t$ converges a.s., we have $\|C_t \dots C_{t-N}\| \to 0$ with probability 1 when $n \to \infty$. Moreover the distribution of $\|\underline{z}_{t-N-1}\|$ is independent of N by stationarity. It follows that $\|C_t \dots C_{t-N}\| \|\underline{z}_{t-N-1}\| \to 0$ in probability as $N \to \infty$. We have shown that $\underline{z}_t - \tilde{\underline{z}}_t \to 0$ in probability when $N \to \infty$. This quantity being independent of N we have, necessarily, $\tilde{\underline{z}}_t = \underline{z}_t$ for any t, a.s.

Now consider the proof of the necessary part. From Lemma 3.4 in Bougerol and Picard (1992a), it is sufficient to prove that $\lim_{t\to\infty} \|C_0 \cdots C_{-t}\| = 0$. It thus suffices to show that, for $1 \leq i \leq p + 2q$

$$\lim_{t\to\infty} C_0 \cdots C_{-t}\underline{e}_i = 0, \quad \text{a.s.} \tag{A.3}$$

where $\underline{e}_i = e_i \otimes I_m$ and e_i is the ith element of the canonical base of $\mathbb{R}^{p+2q}$. If a strictly stationary solution $\underline{z}_t$ exists then, componentwise,

$$\underline{z}_t \geq \underline{b}_t + \sum_{n=0}^{k} C_t C_{t-1} \cdots C_{t-n} \underline{b}_{t-n-1} \qquad \forall k,$$

which implies that $C_0 \cdots C_{-k}\underline{b}_{-k-1} \to 0$ a.s. as $k \to \infty$. It follows that, using the relation $\underline{b}_{-k-1} = \underline{e}_1 \Upsilon^+_{-k-1}\underline{\omega}_0 + \underline{e}_{q+1}\Upsilon^-_{-k-1}\underline{\omega}_0 + \underline{e}_{2q+1}\underline{\omega}_0$, we have

$$\lim_{k\to\infty} C_0 \cdots C_{-k}\underline{e}_{2q+1}\underline{\omega}_0 = 0, \quad \text{a.s.}$$

Since the components of $\underline{\omega}_0$ are strictly positive, (A.3) thus holds for $i = 2q + 1$. We have

$$C_{-k}\underline{e}_{2q+i} = \Upsilon^+_{-k}\mathbf{B}_{0i}\underline{e}_1 + \Upsilon^-_{-k}\mathbf{B}_{0i}\underline{e}_{q+1} + \mathbf{B}_{0i}\underline{e}_{2q+1} + \underline{e}_{2q+i+1}, \quad i = 1, \dots, p \tag{A.4}$$

with the convention $\underline{e}_{p+2q+1} = 0$. Using this relation for $i = 1$ we obtain

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

where the inequalities are taken componentwise. Therefore, (A.3) holds true for $i = 2q + 2$, and by induction, for $i = 2q + j$, $j = 1, \dots, p$. Now $\underline{\omega}_0 > 0$ and $\lim_{k\to\infty} C_0 \cdots C_{-k+1}\underline{b}_{-k} = 0$ imply

$$C_0 \cdots C_{-k}\underline{e}_q = C_0 \cdots C_{-k+1} \begin{pmatrix} \Upsilon^+_{-k}\mathbf{A}_{0q,+} \\ 0_{(q-1)m} \\ \Upsilon^-_{-k}\mathbf{A}_{0q,+} \\ 0_{(q-1)m} \\ \mathbf{A}_{0q,+} \\ 0_{(p-1)m} \end{pmatrix} \to 0 \quad \text{a.s.} \quad \text{as } k \to \infty.$$

Therefore (A.3) holds for $i = q$. By the same argument, (A.3) holds for $i = 2q$. Since $C_{-k}\underline{e}_{i-1} \geq \underline{e}_i$ and $C_{-k}\underline{e}_{i+q-1} \geq \underline{e}_{i+q}$ for $i = 2, \dots, q$, we conclude for the other values of i using an ascendent recursion. □

A.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 $\mathbf{A}_{0i,+}$ and $\mathbf{A}_{0i,-}$ 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 $\mathbb{B}_0$, that is with the spectral radius $\varrho(\mathbb{B}_0)$. It follows that $\gamma(\mathbf{C_0}) \geq \log \varrho(\mathbb{B}_0)$. Hence $\gamma(\mathbf{C_0}) < 0$ entails that all the eigenvalues of $\mathbb{B}_0$ are outside the unit disk. Finally, the equivalence between the two properties follows from

$$\begin{aligned} \det(\mathbb{B}_0 - \lambda I_{mp}) &= (-1)^{mp} \det\left\{\lambda^p I_m - \lambda^{p-1}\mathbf{B}_{01} - \cdots - \lambda\mathbf{B}_{0p-1} - \mathbf{B}_{0p}\right\} \\ &= (-\lambda)^{mp} \det \mathcal{B}\left(\frac{1}{\lambda}\right), \qquad \lambda \neq 0. \end{aligned}$$

□

A.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 (A.2) 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 $\|\underline{h}_t\| \leq \|\underline{\tilde{z}}_t\|$. □

A.4. Proof of the Consistency and the Asymptotic Normality of the QMLE. We shall use the multiplicative norm defined by:

$$\|A\| := \sup_{\|x\|\leq 1} \|Ax\| = \varrho^{1/2}(A'A), \tag{A.5}$$

where A is a $d_1 \times d_2$ matrix, $\|x\| = \sqrt{x'x}$ for any vector $x \in \mathbb{R}^{d_2}$. This norm verifies, for any $d_2 \times d_1$ matrix B,

$$\|A\|^2 \leq \sum_{i,j} a_{i,j}^2 = \mathrm{Tr}(A'A) \leq d_2\|A\|^2, \quad |A'A| \leq \|A\|^{2d_2}, \tag{A.6}$$

$$|\mathrm{Tr}(AB)| \leq \left(\sum_{i,j} a_{i,j}^2\right)^{1/2} \left(\sum_{i,j} b_{i,j}^2\right)^{1/2} \leq \{d_2 d_1\}^{1/2}\|A\|\|B\|. \tag{A.7}$$

A.4.1. *Proof of Theorem 3.1.* Rewrite (3.4) in matrix form as

$$\mathbf{H}_t = \underline{c}_t + \mathbb{B}\mathbf{H}_{t-1} \tag{A.8}$$

where $\mathbb{B}$ is defined as in Corollary 2.1, replacing θ_0 by θ, and

$$\mathbf{H}_t = \begin{pmatrix} \underline{h}_t \\ \vdots \\ \underline{h}_{t-p+1} \end{pmatrix}, \quad \underline{c}_t = \begin{pmatrix} \underline{\omega} + \sum_{i=1}^{q} \mathbf{A}_{i,+}\underline{\epsilon}_{t-i}^{+} + \mathbf{A}_{i,-}\underline{\epsilon}_{t-i}^{-} \\ 0_{(p-1)m} \end{pmatrix}. \tag{A.9}$$

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 $\underline{h}_t(\theta) = \underline{h}_t(\theta_0)$ a.s. and $R = R_0) \implies \theta = \theta_0$,

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

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

In the sequel, $K > 0$ and $\rho \in (0, 1)$ denote generic constants whose exact values are unimportant and may vary in the proofs.

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

$$\sup_{\theta\in\Theta} \varrho(\mathbb{B}) < 1. \tag{A.10}$$

Using iteratively Equation (A.8), we deduce that, almost surely

$$\sup_{\theta\in\Theta} \|\mathbf{H}_t - \tilde{\mathbf{H}}_t\| \leq K\rho^t, \quad \forall t, \tag{A.11}$$

where $\tilde{\mathbf{H}}_t$ denotes the vector obtained by replacing the variables $\underline{h}_{t-i}$ by $\tilde{\underline{h}}_{t-i}$ in $\mathbf{H}_t$. Observe that K is a random variable which depends on the past values $\{\epsilon_t, t \leq 0\}$. From (A.11) we deduce that, almost surely,

$$\sup_{\theta\in\Theta} \|H_t - \tilde{H}_t\| \quad \leq \quad K\rho^t, \quad \forall t. \tag{A.12}$$

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

$$\sup_{\theta\in\Theta} \|\tilde{H}_t^{-1}\| \leq \sup_{\theta\in\Theta} \|\tilde{D}_t^{-1}\|^2 \|R^{-1}\| \leq \sup_{\theta\in\Theta} \{\min_i \underline{\omega}(i)\}^{-1} \|R^{-1}\| \leq K, \tag{A.13}$$

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

$$\sup_{\theta\in\Theta} \|H_t^{-1}\| \leq K. \tag{A.14}$$

Now

$$\begin{aligned}\sup_{\theta\in\Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| \quad \leq \quad & n^{-1} \sum_{t=1}^{n} \sup_{\theta\in\Theta} \left| \epsilon_t'(H_t^{-1} - \tilde{H}_t^{-1})\epsilon_t \right| \\ & + n^{-1} \sum_{t=1}^{n} \sup_{\theta\in\Theta} \left| \log|H_t| - \log|\tilde{H}_t| \right|.\end{aligned} \tag{A.15}$$

The first sum can be written as

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

as $n\to\infty$, using (A.7), (A.12), (A.13), (A.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 (A.6), the triangle inequality and, for $x\geq -1$, $\log(1+x)\leq x$, we have

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

and, by symmetry,

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

Using again (A.12), (A.13) and (A.14) we deduce that, in (A.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

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

Then, it readily follows that $r=r_0$ and, using the invertibility of the polynomial $\mathcal{B}_\theta(B)$ under Assumption **A2**, by (3.1)

$$\mathcal{B}_\theta^{-1}(1)\underline{\omega}-\mathcal{B}_{\theta_0}^{-1}(1)\underline{\omega}_0 = \mathcal{P}^+(B)\underline{\epsilon}_t^+ +\mathcal{P}^-(B)\underline{\epsilon}_t^- \quad \text{a.s.}\quad \forall t,$$

where

$$\mathcal{P}^{\circ}(B) = \mathcal{B}_{\theta_0}^{-1}(1)\mathcal{A}_{\theta_0}^{\circ}(B) - \mathcal{B}_{\theta}^{-1}(B)\mathcal{A}_{\theta}^{\circ}(B) := \sum_{i=0}^{\infty} P_i^{\circ} B^i. \tag{A.16}$$

With these notations and (2.4), there exists a random vector Z_{t-2} belonging to the σ-field generated by $\{\eta_{t-2}, \eta_{t-3}, \dots\}$ such that

$$P_1^+ \Upsilon_{t-1}^+ \underline{h}_{t-1} + P_1^- \Upsilon_{t-1}^- \underline{h}_{t-1} = Z_{t-2}, \quad \text{a.s.}$$

Since $P_1^+ \Upsilon_{t-1}^+ + P_1^- \Upsilon_{t-1}^-$ is independent from $(Z_{t-2}, \underline{h}_{t-1})$ and since $\underline{h}_{t-1} > 0$, $P_1^+ \Upsilon_{t-1}^+ + P_1^- \Upsilon_{t-1}^- = C$ for some constant matrix C. This matrix equality for the element (i,j) writes $P_1^+(i,j)\left(\tilde{\eta}_{j,t-1}^+\right)^2 + P_1^-(i,j)\left(\tilde{\eta}_{j,t-1}^-\right)^2 = C(i,j)$. If $P_1^+(i,j)P_1^-(i,j) \neq 0$, then $\tilde{\eta}_{j,t-1}$ takes at most two different values, which is in contradiction with **A3**. If $P_1^+(i,j) \neq 0$ and $P_1^-(i,j) = 0$, then $P_1^+(i,j)\left(\tilde{\eta}_{j,t-1}^+\right)^2 = C(i,j)$ which entails $C(i,j) = 0$, since $P(\tilde{\eta}_{j,t-1}^- > 0) \neq 0$, and then $\tilde{\eta}_{j,t-1}^- = 0$ a.s., which is also in contradiction with **A3**. We thus have $P_1^+ = P_1^- = 0$. Continuing in this way, we show that $\mathcal{P}^+(B) = \mathcal{P}^-(B) = 0$. Therefore, in view of (A.16), Proposition 3.1 shows that **A4** (or **A4'**) entails $\theta = \theta_0$. 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

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

At θ_0, Jensen's inequality, the second inequality in (A.6) and Corollary 2.2 entail

$$\begin{aligned} & E_{\theta_0} \log |H_t(\theta_0)| = E_{\theta_0} \frac{m}{s} \log |H_t(\theta_0)|^{s/m} \\ \leq\ & \frac{m}{s} \log E_{\theta_0} \|H_t^{1/2}(\theta_0)\|^{2s} \leq K + \frac{m}{s} \log E_{\theta_0} \|D_t(\theta_0)\|^{2s} \\ =\ & K + \frac{m}{s} \log E_{\theta_0} (\max_i h_{ii,t}(\theta_0))^s \leq K + \frac{m}{s} \log E_{\theta_0} \|\underline{h}_t(\theta_0)\|^s < \infty. \end{aligned}$$

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

$$E_{\theta_0}\ell_t(\theta) - E_{\theta_0}\ell_t(\theta_0) = E_{\theta_0}\left\{\sum_{i=1}^m (\lambda_{it} - 1 - \log\lambda_{it})\right\} \geq 0,$$

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, $\lambda_{it} = 1$ a.s. , that is if $H_t(\theta) = H_t(\theta_0)$, a.s. . This equality is equivalent to

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

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

Proof of iv). The last part of the proof of the consistency uses standard arguments, such as the compactness of Θ and the ergodicity of $(\ell_t(\theta))$. Therefore is it omitted. Theorem 3.1 is thus established, by the arguments used for the proof of Theorems 2.1 in Francq and Zakoian (2004). □

A.4.2. *Proof of Theorem 3.2.* By a standard Taylor expansion we have

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -\left(\frac{1}{n}\sum_{t=1}^n \frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_t(\theta_{ij}^*)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)\right), \tag{A.17}$$

where the θ_{ij}^*'s are between $\hat{\theta}_n$ and θ_0. It will be sufficient to show

$$n^{-1/2}\sum_{t=1}^n \frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I), \tag{A.18}$$

the invertibility of J, and

$$n^{-1}\sum_{t=1}^n \frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_t(\theta_{ij}^*) \to J(i,j) \text{ in probability.} \tag{A.19}$$

The proof is divided into several steps, which are analogous to steps (i)-(vi) in the proof of Theorem 2.2 in Francq and Zakoian (2004).

a) First derivative of the criterion. We shall use standard results on the differentiation of matrices which can be found in Magnus and Neudecker (1988) or

Harville (1997), in particular the chain rule for differentiation of composed functions: $\partial f(A)/\partial x = \mathrm{Tr}\left\{(\partial f(A)/\partial A')(\partial A/\partial x)\right\}$. We obtain

$$\begin{aligned}\frac{\partial \ell_t(\theta)}{\partial \theta_i} &= \mathrm{Tr}\left(\epsilon_t\epsilon_t'\frac{\partial D_t^{-1}R^{-1}D_t^{-1}}{\partial \theta_i}\right)+2\frac{\partial \log|\det D_t|}{\partial \theta_i}\\ &= -\mathrm{Tr}\left\{\left(\epsilon_t\epsilon_t'D_t^{-1}R^{-1}+R^{-1}D_t^{-1}\epsilon_t\epsilon_t'\right)D_t^{-1}\frac{\partial D_t}{\partial \theta_i}D_t^{-1}\right\}\\ &\quad +2\mathrm{Tr}\left(D_t^{-1}\frac{\partial D_t}{\partial \theta_i}\right), \qquad (A.20)\end{aligned}$$

for $i = 1, \dots, s_1 = m + (p + 2q)m^2$, and using $\partial \mathrm{Tr}(CA^{-1}B)/\partial A' = -A^{-1}BCA^{-1}$,

$$\frac{\partial \ell_t(\theta)}{\partial \theta_i} = -\mathrm{Tr}\left(R^{-1}D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}R^{-1}\frac{\partial R}{\partial \theta_i}\right)+\mathrm{Tr}\left(R^{-1}\frac{\partial R}{\partial \theta_i}\right), \qquad (A.21)$$

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

$$\begin{aligned}\frac{\partial \ell_t(\theta_0)}{\partial \theta_i} &= \mathrm{Tr}\left\{\left(I_m - R_0^{-1}\tilde{\eta}_t\tilde{\eta}_t'\right)D_{0t}^{(i)}D_{0t}^{-1}\right. \qquad (A.22)\\ &\quad \left. + \left(I_m - \tilde{\eta}_t\tilde{\eta}_t'R_0^{-1}\right)D_{0t}^{-1}D_{0t}^{(i)}\right\},\end{aligned}$$

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

$$\frac{\partial \ell_t(\theta_0)}{\partial \theta_i} = \mathrm{Tr}\left\{\left(I_m - R_0^{-1}\tilde{\eta}_t\tilde{\eta}_t'\right)R_0^{-1}R_0^{(i)}\right\}, \qquad (A.23)$$

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

b) Existence of moments at any order for the score. In view of (A.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 (A.8) and (A.10),

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

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

$$\theta_i \frac{\partial \mathbf{H}_t}{\partial \theta_i} \leq \mathbf{H}_t, \qquad i = m + 1, \dots, s_2.$$

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 $\underline{h}_t(\theta_0)$. Using the inequality $x/(1+x) \leq x^s$ for all $x \geq 0$ and $s \in [0, 1]$, the inequalities

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

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

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

we obtain

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

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 extend to higher-order derivatives, 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 $r_0 \geq 0$, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\underline{h}_t(i_1)} \frac{\partial \underline{h}_t(i_1)}{\partial\theta_i}(\theta) \right|^{r_0} < \infty, \tag{A.24}$$

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\underline{h}_t(i_1)} \frac{\partial^2 \underline{h}_t(i_1)}{\partial\theta_i\partial\theta_j}(\theta) \right|^{r_0} < \infty, \tag{A.25}$$

and

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \frac{1}{\underline{h}_t(i_1)} \frac{\partial^3 \underline{h}_t(i_1)}{\partial\theta_i\partial\theta_j\partial\theta_k}(\theta) \right|^{r_0} < \infty. \tag{A.26}$$

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 $\{\eta_u, u \leq t\}$. From (A.22) and (A.23) we have $E\left\{\partial\ell_t(\theta_0)/\partial\theta \mid \mathcal{F}_{t-1}\right\} = 0$. Property (A.24) ensures the existence of matrix I in Theorem 3.2. It follows that $\forall\lambda \in \mathbb{R}^{p+2q+1}$, 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

$$n^{-1/2}\sum_{t=1}^{n} \frac{\partial}{\partial\theta}\ell_t(\theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I). \tag{A.27}$$

d) Convergence to J. Starting from (A.20), applying several times the chain rule and using $\partial \mathrm{Tr}(CAB)/\partial A' = BC$ 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

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

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 (A.24) and (A.25) ensure the existence of the matrix $J=E\partial^2\ell_t(\theta_0)/\partial\theta\partial\theta'$. Note that with our parameterization, $\partial^2R/\partial\theta_i\partial\theta_j=0$.

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. For instance, let us consider the derivative of the last term in $\mathrm{Tr}(c_1)$. We have, for $i,j,k=1,\dots,s_1$,

$$\frac{\partial}{\partial\theta_k}\mathrm{Tr}\left(D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}R^{-1}D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j}\right) = -\mathrm{Tr}\,(c_{11}+c_{12}+c_{13}+c_{14})\,,$$

where

$$\begin{aligned}c_{11} &= D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}R^{-1}D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j}D_t^{-1}\frac{\partial D_t}{\partial\theta_k},\\ c_{12} &= D_t^{-1}R^{-1}D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j}D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}\frac{\partial D_t}{\partial\theta_k},\\ c_{13} &= D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j}D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}R^{-1}D_t^{-1}\frac{\partial D_t}{\partial\theta_k},\\ c_{14} &= -D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}R^{-1}D_t^{-1}\frac{\partial^3 D_t}{\partial\theta_i\partial\theta_j\partial\theta_k}.\end{aligned}$$

Using (A.7) and (2.3), we have

$$\begin{aligned}E\sup_{\theta\in\mathcal{V}(\theta_0)}|\mathrm{Tr}(c_{11})| &\le KE\|\tilde\eta_t\tilde\eta_t'\|E\sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1}D_{0t}\|^2\|R^{-1}\|\left\|D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j}\right\|\left\|D_t^{-1}\frac{\partial D_t}{\partial\theta_k}\right\|\\ &\le KE\sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1}D_{0t}\|^2\left\|D_t^{-1}\frac{\partial^2 D_t}{\partial\theta_i\partial\theta_j}\right\|\left\|D_t^{-1}\frac{\partial D_t}{\partial\theta_k}\right\|. \qquad \text{(A.28)}\end{aligned}$$

To prove that the right-hand side is finite for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0, we will first show that, for any $r_0 \geq 0$,

$$E \sup_{\theta \in \mathcal{V}(\theta_0)} \|D_t^{-1} D_{0t}\|^{r_0} < \infty. \tag{A.29}$$

In view of (A.8) the i_1^{th} component of $\underline{h}_t(\theta_0)$ is equal to

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

where c_0 is a strictly positive constant. For a sufficiently small neighborhood $\mathcal{V}(\theta_0)$ of θ_0, we have

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

for all $i_1, j_1, j_2 \in \{1, \dots, m\}$, any integer k, and any $\delta > 0$. Moreover, the coefficients $\mathbf{A}_{0i}^{+}(j_1, j_2)$ and $\mathbf{A}_{0i}^{-}(j_1, j_2)$ are bounded below by a constant $c > 0$ uniformly on $\theta \in \mathcal{V}(\theta_0)$. We thus have

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

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

$$E \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\underline{h}_{0t}(i_1)}{\underline{h}_t(i_1)} \right|^{r_0} < \infty. \tag{A.30}$$

Thus (A.29) follows. Now, because $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| D_t^{-1} (\partial D_t / \partial \theta_k) \right\|$ and $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| D_t^{-1} (\partial^2 D_t / \partial \theta_i \partial \theta_j) \right\|$ admit moments at any order by (A.24) and

(A.25), the right-hand side of (A.28) is finite. The other terms c_{1i}, for $i = 2, \ldots, 4$, as well as all the derivatives of $\mathrm{Tr}(c_i)$, for $i = 1, \ldots, 6$, are sums of products of matrices similar to those involved in c_{11}. The integrable matrix $\tilde{\eta}_t \tilde{\eta}_t'$ 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 $D_t^{-1} D_{0t}$, $D_t^{-1}\partial D_t/\partial\theta_i$, $D_t^{-1}\partial^2 D_t/\partial\theta_i\partial\theta_j$ and $D_t^{-1}\partial^3 D_t/\partial\theta_i\partial\theta_j\partial\theta_k$, whose norms admit moments at any orders in the neighborhood of θ_0. Finally, $E\sup_{\theta\in\mathcal{V}(\theta_0)} |\partial\mathrm{Tr}(c_i)/\partial\theta_k| < \infty$ for $i = 1, \ldots, 6$ and $k = 1, \ldots, s_0$.

This shows that

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

It follows, by the arguments used for the proof of Theorems 2.2 in Francq and Zakoian (2004), that

$$n^{-1} \sum_{t=1}^{n} \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta_{ij}^*) \to J(i,j) \text{ in probability.} \tag{A.31}$$

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

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

elementary results already used 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,

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

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

$$\mathbf{c}'\mathbf{h}'\mathbf{hc} = \mathbf{c}'\mathbf{d}'\mathbf{H}^2\mathbf{dc} = 0 \quad a.s. \tag{A.32}$$

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 $1, m+1, \ldots, m^2$ of the equations

$$\mathbf{dc} = \sum_{i=1}^{s_0} c_i \frac{\partial}{\partial\theta_i}\mathrm{vec}H_{0t} = \sum_{i=1}^{s_0} c_i \frac{\partial}{\partial\theta_i}\left(D_{0t} \otimes D_{0t}\right)\mathrm{vec}R_0 = 0_{m^2}, \quad a.s. \tag{A.33}$$

give

$$\sum_{i=1}^{s_1} c_i \frac{\partial}{\partial\theta_i}\underline{h}_t(\theta_0) = 0_m, \quad a.s. \tag{A.34}$$

Differentiating Equation (3.4) yields

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

where

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

Because (A.34) is satisfied for all t, we have

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

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

$$\mathrm{vec}\left(\underline{\omega}_0 - \underline{\omega}_0^* \mid \mathbf{A}_{01}^{+} - \mathbf{A}_{01}^{+*} \mid \cdots \mid \mathbf{A}_{01}^{-} - \mathbf{A}_{01}^{-*} \mid \cdots \mid \mathbf{B}_{0p} - \mathbf{B}_{0p}^{*}\right),$$

we obtain $\underline{h}_t(\theta_0) = \underline{h}_t(\theta_1)$ 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 (A.33) 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 $\partial \mathrm{vec} R / \partial \theta_i$, $i = s_1 + 1, \ldots, s_0$, are linearly independent, the vector $\mathbf{c_2} = (c_{s_1+1}, \ldots, c_{s_0})'$ is null, and thus $\mathbf{c} = 0$. This is in contradiction with (A.32), and shows that the assumption that J is singular is absurd.

f) Forgetting of the initial values. To deduce (A.18)-(A.19) from (A.27)-(A.31), it remains to show that the effect of the initial values on the derivatives of the criterion vanish asymptotically. More precisely, we will prove that

$$\left\| n^{-1/2} \sum_{t=1}^{n} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{\ell}_t(\theta_0)}{\partial \theta} \right\} \right\| = o_P(1) \tag{A.35}$$

and, for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0,

$$n^{-1} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\ell}_t(\theta)}{\partial \theta \partial \theta'} \right\| = o_P(1). \tag{A.36}$$

First remark that (A.11) and the arguments used to show (A.13) and (A.14) entail

$$\sup_{\theta\in\Theta}\|D_t-\tilde{D}_t\|\leq K\rho^t,\quad \sup_{\theta\in\Theta}\|\tilde{D}_t^{-1}\|\leq K,\quad \sup_{\theta\in\Theta}\|D_t^{-1}\|\leq K. \tag{A.37}$$

Because

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

(A.37) entails

$$\sup_{\theta\in\Theta}\left\|D_t\left(D_t^{-1}-\tilde{D}_t^{-1}\right)\right\|\leq K\rho^t. \tag{A.38}$$

From (A.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. Let $i\in\{1,\dots,s_1\}$. 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 (A.10) entails $\sup_{\theta\in\Theta}\left\|\partial\left(\mathbf{H}_t-\tilde{\mathbf{H}}_t\right)/\partial\theta_i\right\|\leq K\rho^t$, or equivalently,

$$\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_i}\left(\underline{h}_t-\underline{\tilde{h}}_t\right)\right\|\leq K\rho^t. \tag{A.39}$$

In view of (A.24), we also have

$$E\sup_{\theta\in\Theta}\left\|D_t^{-2}\mathrm{diag}\left\{\frac{\partial\underline{h}_t}{\partial\theta_i}(\theta)\right\}\right\|^2<\infty,\quad E\sup_{\theta\in\Theta}\left\|D_t^{-1}\frac{\partial D_t}{\partial\theta_i}(\theta)\right\|^2<\infty. \tag{A.40}$$

Because

$$\frac{\partial}{\partial\theta_i}D_t=\frac{1}{2}D_t^{-1}\mathrm{diag}\left\{\frac{\partial\underline{h}_t}{\partial\theta_i}(\theta)\right\},\quad \frac{\partial}{\partial\theta_i}\tilde{D}_t=\frac{1}{2}\tilde{D}_t^{-1}\mathrm{diag}\left\{\frac{\partial\underline{\tilde{h}}_t}{\partial\theta_i}(\theta)\right\}$$

we have

$$\begin{aligned}D_t^{-1}\frac{\partial}{\partial\theta_i}\left(D_t-\tilde{D}_t\right) &= \frac{1}{2}D_t\left(D_t^{-1}-\tilde{D}_t^{-1}\right)D_t^{-2}\mathrm{diag}\left\{\frac{\partial\underline{h}_t}{\partial\theta_i}(\theta)\right\}\\ &\quad+\frac{1}{2}\tilde{D}_t^{-1}D_t^{-1}\mathrm{diag}\left\{\frac{\partial\underline{\tilde{h}}_t}{\partial\theta_i}(\theta)-\frac{\partial\underline{\tilde{h}}_t}{\partial\theta_i}(\theta)\right\}.\end{aligned}$$

Using (A.37)-(A.40) we then obtain

$$\sup_{\theta\in\Theta}\left\|D_t^{-1}\frac{\partial}{\partial\theta_i}\left(D_t-\tilde{D}_t\right)\right\|\leq K\rho^t\mathrm{u}_t, \tag{A.41}$$

where u_t is a squared integrable variable. From (A.30) and (A.38) we deduce

$$\sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1}\epsilon_t\| = \sup_{\theta\in\mathcal{V}(\theta_0)}\|D_t^{-1}D_{0t}\tilde{\eta}_t\|\leq \mathrm{v}_t\|\tilde{\eta}_t\|, \tag{A.42}$$

$$\sup_{\theta\in\mathcal{V}(\theta_0)}\|\tilde{D}_t^{-1}\epsilon_t\| = \sup_{\theta\in\mathcal{V}(\theta_0)}\|\tilde{D}_t^{-1}D_t\|\|D_t^{-1}\epsilon_t\|\leq (1+K\rho^t)\mathrm{v}_t\|\tilde{\eta}_t\|, \tag{A.43}$$

where the random variable v_t possesses fourth-order moments. By (A.20)

$$\frac{\partial\ell_t(\theta)}{\partial\theta_i}-\frac{\partial\tilde{\ell}_t(\theta)}{\partial\theta_i}=\mathrm{Tr}(d_1+d_2+d_3)$$

where

$$d_1 = -D_t^{-1}\epsilon_t\epsilon_t'D_t^{-1}D_t\left(D_t^{-1}-\tilde{D}_t^{-1}\right)R^{-1}D_t^{-1}\frac{\partial D_t}{\partial\theta_i},$$

$$d_2 = -D_t^{-1}\epsilon_t\epsilon_t'\tilde{D}_t^{-1}R^{-1}D_t^{-1}\left(\frac{\partial D_t}{\partial\theta_i}-\frac{\partial\tilde{D}_t}{\partial\theta_i}\right)$$

and d_3 is a sum of terms which can be handled as d_1 and d_2. Using (A.37)–(A.43) and the Cauchy-Schwarz inequality, we obtain

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

where w_t is an integrable variable. From the Markov inequality, $n^{-1/2}\sum_{t=1}^n\rho^t\mathrm{w}_t = o_P(1)$, which implies (A.35). By exactly the same arguments,

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

where w_t^* is an integrable random variable, which entails (A.36) and completes the proof.

□

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