Asymptotic Theory for Multivariate GARCH Processes

F. Comte* and O. Lieberman†

Revised, July 2001

Abstract

We provide in this paper asymptotic theory for the multivariate GARCH(p, q) process. Strong consistency of the quasi-maximum likelihood estimator (MLE) is established by appealing to conditions given in Jeantheau [19] in conjunction with a result given by Boussama [9] concerning the existence of a stationary and ergodic solution to the multivariate GARCH(p, q) process. We prove asymptotic normality of the quasi-MLE when the initial state is either stationary or fixed.

KEY WORDS: Asymptotic normality, BEKK, Consistency, GARCH, Martingale CLT.

*University of Paris 6, Laboratoire de Probabilités et Modèles Aléatoire, UMR 7599, FRANCE. Email: comte@ccr.jussieu.fr

†Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa, ISRAEL. Email: offerl@ie.technion.ac.il.

1. INTRODUCTION

There is now an insurmountable literature on Generalized Autoregressive Conditional Heteroscedasticity (GARCH). The model and its various subsidiaries have been one of the most successful econometric modelling schemes over the past two decades or so. For univariate GARCH, there is more or less coherent asymptotic theory for the maximum likelihood estimator (MLE), enabling practitioners to conduct statistical inference with a reasonable amount of confidence, given as usual, correct model specification and a large enough sample. The story is markedly different in the multivariate case. Here, since asymptotic theory is rare, practitioners often resort to asymptotic normality simply as a rule of thumb. See for instance, Bollerslev [5], pp. 306–307, or Comte and Lieberman [11].

Broadly speaking, most papers in the area concentrate on either the univariate or the multivariate case and on either the statistical or probabilistic properties of these processes. In the following we outline the ongoing research on asymptotic theory for GARCH. The univariate ARCH(p) model

$$x_t = \sqrt{h_t}\varepsilon_t, \;\; \varepsilon_t \sim \text{iid}(0,1), \;\; h_t = w + \sum_{i=1}^{p} \alpha_i x_{t-i}^2, \tag{1}$$

was originally presented by Engle [13]. It was generalized by Bollerslev [4] to GARCH(p, q), with

$$h_t = w + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}.$$

The model requires $w > 0$ and $\alpha_i \geq 0, \;\; \beta_i \geq 0, \;\; \forall i$. It has been shown by Bollerslev [4] to be second-order stationary if and only if

$$w > 0 \text{ and } \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

Weiss [30] established consistency and asymptotic normality of the MLE in a univariate linear dynamic model with ARCH(p) errors, a model slightly more general than (1). He proved asymptotic normality by appealing to conditions given by Basawa et al. [2]. These conditions appear to form the backbone of many related studies to follow. Nelson [26] gave necessary and sufficient conditions for strict stationarity and ergodicity of the univariate GARCH(1,1) model. His condition

$$E\{\log(\beta_1 + \alpha_1 \varepsilon_t^2)\} < 0 \tag{2}$$

does not exclude the case $\beta_1 + \alpha_1 = 1$ and hence, allows for the possibility of Integrated GARCH (IGARCH).

Lumsdaine [22] established consistency and asymptotic normality of the quasi-MLE in the GARCH(1,1) and IGARCH(1,1) models under the assumptions: (i) The true parameter $\theta_0 \in \text{int}(\Theta)$, $\Theta \subset \mathbb{R}^4$ is a compact parameter space, (ii) Nelson's [26] condition (2) and (iii): $\varepsilon_t \sim$ iid f_ε, with f_ε a symmetric unimodal density, bounded in the neighborhood of the origin, $\mathbb{E}(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = 1$ and $\mathbb{E}(\varepsilon_t^{32}) < \infty$. In addition, h_t is independent of $\{\varepsilon_t, \varepsilon_{t+1}, ...\}$. The main difference between Lumsdaine's [22] and Weiss' [30] conditions is that the former are imposed on the noise density whereas the latter are imposed on the process. In particular, Weiss assumed $E(x_t^4) < \infty$. Lee and Hansen [20] also considered the univariate GARCH(1,1) model with the possibility that the process is integrated or even mildly explosive. In contrast with Lumsdaine's [22] work, no assumption was made on the shape of the density. For nonintegrated GARCH, Lee and Hansen [20] gave a first proof of consistency of the quasi-MLE under the assumptions that ε_t is strictly stationary and ergodic with $E(|\varepsilon_t|^{2+\delta}|\mathcal{F}_{t-1}) \leq S_\delta < \infty$, where S_δ is a positive constant, $\delta > 0$, $\mathcal{F}_t = \sigma\{x_t, x_{t-1}, ...\}$ and $\alpha_1 + \beta_1 < 1$. Asymptotic normality for the IGARCH case was given under the additional assumption $\mathbb{E}(\varepsilon_t^4|\mathcal{F}_{t-1}) \leq K < \infty$. Ling and Li [21] established asymptotic theory for the estimators of the ARMA parameters in unstable ARMA processes with GARCH innovations. They derived the limiting distribution of the MLE in a unified manner for all types of roots of the ARMA part inside/outside the unit circle. The limiting distribution involves a sequence of independent bivariate Brownian motions with correlated components.

Parallel to the asymptotic theory of estimation, Bougerol and Picard [8] established strict stationarity and ergodicity of the univariate GARCH(p,q) model in terms of the top Lyapunov exponent

$$\lambda = \inf_{t \in \mathbb{N}} (t+1)^{-1} \mathbb{E}\{\log \|A(\varepsilon_0)A(\varepsilon_{-1}) \ \cdots \ A(\varepsilon_{-t})\|\} < 0,$$

where $A(\varepsilon_t)$ is a matrix composed of the coefficients of the process and the noise ε_t, the ε_t are iid and $\|.\|$ is the Euclidian norm. Their result is a generalization of Nelson's [26] result for the stable GARCH(1,1) case. Bougerol and Picard [8] proved their main theorem by writing the model as a first-order recursion with random coefficients. The intuition is given by Bougerol's ([7], Theorem 3.1) conditions under which the function of recursion is Lipschitz. A model $Y_{t+1} = F(Y_t, \eta_{t+1})$ with $\eta_t \sim \text{iid}(0,1)$ is said to satisfy a Lipschitz property if

$$\|F(x,\eta) - F(y,\eta)\| \leq \alpha(\eta)\|x - y\|$$

for a positive valued function α with $\mathbb{E}(\alpha(\eta_t)^m) < 1$ and $\mathbb{E}(\|F(0,\eta_t)\|^m) < \infty$ for some real number $m \geq 1$. In the GARCH(p,q) context the components of Y_t are the past and current values of x_t, and h_t. The Lipschitz idea works for univariate GARCH(p,q) models. Unfortunately, Bougerol and Picard's [8] approach does not extend in general to the multivariate case. Boussama [9] gave a counter-example to this extent. Recently, Hansen and Rahbek [16] used an operational drift criterion from Markov chain theory to obtain stationarity, ergodicity and existence of moments in a simple multivariate ARCH(1) process. The simple model discussed in their work retains the Lipschitz property used in Bougerol's [7] work. Boussama [9] proved the existence of a stationary and ergodic solution for multivariate GARCH(p,q) models by using Markov chain theory and algebraic topology.

Starting under Bougerol and Picard's [8] conditions, Elie and Jeantheau [12] established strong consistency of the quasi–MLE in the univariate GARCH(p,q) model and Boussama [10] proved asymptotic normality of the quasi–MLE in the same model under moment conditions of order 6 on the noise and under the minimal strict stationarity conditions that allow for IGARCH models.

As opposed to the univariate case, asymptotic theory of estimation for multivariate GARCH processes is far from being coherent. Bollerslev and Wooldridge [6] proposed the condition that the likelihood follows a uniform weak law of large numbers for consistency of the MLE. They also assumed asymptotic normality of the score but have not verified whether any of the conditions actually holds for specific multivariate GARCH models. Tuncer [29] established weak convergence of the MLE of a multivariate GARCH(1,1) BEKK[1] representation, a model proposed by Engle and Kroner [14]. Jeantheau [19] gave conditions for strong consistency of the MLE for multivariate GARCH and verified that the conditions hold for the multivariate model with constant correlation (e.g., Bollerslev [4]). Jeantheau's [19] work does not require conditions on the log-likelihood derivatives. His main condition is that the process admits a unique strictly stationary and ergodic solution.

In this paper we establish asymptotic theory for the multivariate GARCH(p, q) process. In Section 3 we prove strong consistency of the quasi-MLE by verifying conditions given by Jeantheau [19]. In Section 4 we establish asymptotic normality of the quasi-MLE when the initial state of the process is either in the stationary law or fixed. We assume existence of a density with support containing the origin for the rescaled innovation ε_t and the finiteness of moments of the process up to order 8. We emphasize that the tools adopted by Lumsdaine [22] and Lee and Hansen [20] in the univariate setting do not seem to be of much use in the multivariate framework. In addition, our model is non-Lipschitz and so our conditions are set on the process. Finally, our model excludes the IGARCH case. For the clarity of the exposition, we include only the chief results in the main body of the paper. All detailed proofs are placed in the Appendix.

2. NOTATION AND PRELIMINARIES

We consider the multivariate GARCH(p, q) model defined as follows. Let $(X_t)_{t\in\mathbf{Z}}$ be a sequence of random variables of $\mathbb{R}^d$ and let $\mathcal{F}_t$ be the σ -field generated by past X_t's, i.e., $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$. We assume that X_t is square integrable and such that

$$X_t = H_t^{1/2}\varepsilon_t \tag{3}$$

with

$$\varepsilon_t \sim \text{iid}(0, I_d) \tag{4}$$

where I_d is the $d \times d$ identity matrix. Without loss of generality, we choose $H_t^{1/2}$ to be symmetric. The process X_t is a martingale-difference

$$\mathbb{E}(X_t|\mathcal{F}_{t-1}) = 0 \ \text{ a.s.} \tag{5}$$

with a conditional covariance matrix

$$\mathbb{E}(X_t X_t'|\mathcal{F}_{t-1}) = H_t. \tag{6}$$

[1]The acronym BEKK stands for Baba, Engle, Kraft and Kroner who wrote an earlier version of the paper by Engle and Kroner [14].

Engle and Kroner's [14] BEKK representation is given by

$$H_t = C + \sum_{i=1}^{q} \left(\sum_{j=1}^{k} A_{ij} X_{t-i} X'_{t-i} A'_{ij} \right) + \sum_{i=1}^{p} \left(\sum_{j=1}^{k} B_{ij} H_{t-i} B'_{ij} \right), \tag{7}$$

where the matrices C, A_{ij}, for $i = 1, \dots, q, j = 1, \dots, k$ and B_{ij}, for $i = 1, \dots, p, j = 1, \dots, k$ satisfy the assumption:

$$C \text{ is positive definite, } A_{ij}, B_{ij} \text{ are real } d \times d \text{ matrices,} \tag{8}$$

and k is an integer less than $d(d+1)/2$. The main advantage of this model is that it guarantees positive definiteness of H_t. Denote by *vec* and *vech* the operator that stacks the columns of a matrix, and the vector-half operator, which stacks the lower triangular portion of a matrix, respectively. Then (7) can be rewritten as

$$vec(H_t) = vec(C) + \sum_{i=1}^{q} A_i^\star vec(X_{t-i} X'_{t-i}) + \sum_{i=1}^{p} B_i^\star vec(H_{t-i}) \tag{9}$$

with $A_i^\star = \sum_{j=1}^{k} A_{ij} \otimes A_{ij}$ for $i = 1, \dots, q$ and $B_i^\star = \sum_{j=1}^{k} B_{ij} \otimes B_{ij}$ for $i = 1, \dots, p$, and $\otimes$ denoting the Kronecker product.
Since the matrices involved in the representation are symmetric, we may also write

$$vech(H_t) = vech(C) + \sum_{i=1}^{q} \tilde{A}_i vech(X_{t-i} X'_{t-i}) + \sum_{i=1}^{p} \tilde{B}_i vech(H_{t-i}), \tag{10}$$

where L_d and K_d are matrices of dimension $d(d+1) \times d^2$ satisfying $\tilde{A}_i = L_d A_i^\star K'_d$ for $i = 1, \dots, q$ and $\tilde{B}_i = L_d B_i^\star K'_d$ for $i = 1, \dots, p$. Note that $\dim(vec(H_t)) = d^2$ and $\dim(vech(H_t)) = d(d+1)/2$. Without loss of generality, we can set $k = 1$. All proofs in the paper trivially extend to any arbitrary k. We denote by θ the parameter vector of the process, so that the matrices C, $\tilde{A}_i$ and $\tilde{B}_i$ are functions of θ: $C = C(\theta)$, $\tilde{A}_i = \tilde{A}_i(\theta)$ and $\tilde{B}_i = \tilde{B}_i(\theta)$. Note that in most applied work the entries in the matrices C, $\tilde{A}_i$ and $\tilde{B}_i$ are simply the components of θ.

The model is not assumed to be necessarily Gaussian, but we work with the Gaussian log-likelihood. So, the quasi-MLE $\hat{\theta}_n$ is defined as minimizing

$$L_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} \ell_t(\theta)$$

with

$$\ell_t(\theta) = \log[\det(H_{t,\theta})] + X'_t H_{t,\theta}^{-1} X_t$$

where $\det(A)$ denotes the determinant of the matrix A. Note that the likelihood depends on the observed X_t and also on H_t which needs to be calculated recursively. We consider two possibilities for the choice of the initial value of the process. The first option is to assume that the initial value of the H_t sequence is drawn from the stationary law. This approach is of little practical use but of important theoretical conveniency: indeed it allows to work first with a stationary process. For practical purposes it is easier to assume a fixed initial value. This leads to non–stationarity of H_t. We show in the paper that either option leads to the same asymptotic results. The key point is that non–stationary H_t's converge to stationarity with an exponential rate.

We further make use of the following notation. $\|\cdot\|$ is the Euclidian norm for both vectors and matrices, $\|A\|^2 = \text{Tr}(A'A) = \sum_{i,j} A_{i,j}^2$, $\rho(A)$ the spectral radius of A , i.e., the largest modulus of the eigenvalues of A. $\mathbf{N}(A)$ is the spectral norm of A, namely, the square root of $\rho(A'A)$. The following inequalities (see Magnus and Neudecker [23]) will be used extensively in our work.

$$|\text{Tr}(AB)| \leq \|A\| \, \|B\|, \;\; \mathbf{N}(AB) \leq \mathbf{N}(A) \;\, \mathbf{N}(B), \tag{11}$$

$$\|AB\| \leq \mathbf{N}(A) \;\, \|B\|, \;\; \|AB\| \leq \|A\|\mathbf{N}(B), \;\, \mathbf{N}(A+B) \leq \mathbf{N}(A) + \mathbf{N}(B). \tag{12}$$

If A is $d \times d$, then

$$\mathbf{N}(A) \leq \|A\| \leq \sqrt{d}\mathbf{N}(A). \tag{13}$$

3. STRONG CONSISTENCY

In this section we establish strong consistency of the quasi MLE by appealing to Jeantheau's [19] conditions. Let Θ be the parameter space and $\theta_0 \in \Theta \subset \mathbb{R}^r$ be the true parameter value. Jeantheau's [19] conditions for strong consistency of the quasi-MLE are:

A0 Θ is compact.

A1 $\forall \theta_0 \in \Theta$, the model admits a unique strictly stationary and ergodic solution, following a stationary law P_{θ_0}.

A2 There exists a deterministic constant c such that $\forall t, \forall \theta \in \Theta, \det(H_{t,\theta}) \geq c$.

A3 $\forall \theta_0 \in \Theta, \mathbb{E}_{\theta_0}(|\log(\det(H_{t,\theta_0}))|) < \infty$.

A4 The model is identifiable.

A5 $H_{t,\theta}$ is a continuous function of θ.

We now verify that the conditions hold for the model under consideration. First, **A0** is always assumed. For **A1**, we recall the following theorem from Boussama [9].

Theorem 1 *For the model given by* (3)–(4) *and* (7)*, assume that the* ε_t*'s admit a density absolutely continuous w.r.t. the Lebesgue measure, positive in a neighbourhood of the origin. Assume moreover that*

$$\rho(\sum_{i=1}^{q} \tilde{A}_i + \sum_{i=1}^{p} \tilde{B}_i) < 1,$$

and let Y *be defined by*

$$Y_t = (vech(H_{t+1})', vech(H_t)', \ldots, vech(H_{t-p+2})', X_t', X_{t-1}', \ldots, X_{t-q+1}')'. \tag{14}$$

Then the recurrence relations (3)–(4) and (7) for Y *have an almost surely unique strictly stationary causal solution which constitutes a positive Harris recurrent Markov chain which is geometrically ergodic and* β*-mixing.*

Boussama [9] proved the theorem on an application of theorems by Meyn and Tweedie [24] together with some results given in Mokkadem [25]. Both Boussama [9] and Mokkadem [25] make extensive use of algebraic topology.

To show **A2**, we note that for any positive definite matrix W and for any positive semidefinite matrix D, $\det(W + D) \geq \det(W) + \det(D)$. It follows from (7) that $\det(H_{t,\theta}) \geq \det(C(\theta))$. As Θ is compact, we may set $c := \inf_{\theta\in\Theta} det(C(\theta))$ as soon as $C(\theta)$ is a continuous function of θ and we assume that $c > 0$. For **A3**, let $x_i(\theta)$ be the (positive) eigenvalues of $H_{t,\theta}$ for a fixed t. Then

$$\log[\det(H_{t,\theta})] = \sum_{i=1}^{d} \log(x_i(\theta)) \leq \sum_{i=1}^{d} x_i(\theta) = \mathrm{Tr}(H_{t,\theta})$$

implying that

$$\mathbb{E}(\log(\det H_{t,\theta})) \leq \mathbb{E}(\mathrm{Tr}(H_{t,\theta})) = \sum_{i=1}^{d} \mathbb{E}([H_{t,\theta}]_{ii}).$$

By the square integrability of X_t, $\mathbb{E}(vech(H_{t,\theta})) < +\infty$. Thus, $\mathbb{E}(\log(\det H_{t,\theta}))^+ < \infty$ and $\mathbb{E}(\log(\det H_{t,\theta}))^- \leq \sup(-\log(c), 0) < \infty$. Therefore, $\mathbb{E}|\log(\det H_{t,\theta})| < +\infty$ and **A3** is fulfilled. For Assumption **A4**, we recall the following proposition by Engle and Kroner [14]. Defining two representations to be equivalent if each sequence $\{X_t\}$ generates the same sequence $\{H_t\}$ for both representations, they prove:

Proposition 1 *For the model*

$$H_t = C_0C_0' + \sum_{j=1}^{k} A_{1j}' X_{t-1}X_{t-1}' A_{1j} + \sum_{j=1}^{k} B_{1j}' H_{t-1} B_{1j},$$

suppose the diagonal elements of C_0 are restricted to be positive. Assume that A_{1k_s}, with $k_s = d(s-1)+1, \ldots, ds$ and $s = 1, \ldots, d$, is the matrix obtained by setting the first $s-1$ columns and the first $k_s - d(s-1) - 1$ rows to zero. Assume also that $[A_{1k_s}]_{dd} > 0$, $\forall k_s$ and that similar restrictions are set on the B_{1j} matrices. Then a fully general BEKK model is obtained which has no other equivalent representations in this class.

Similar conditions for identification can be set for higher order BEKK processes, see Engle and Kroner [14]. This definition is consistent with Jeantheau's ([19], p.72) definition of identifiability, namely, $\forall\theta \in \Theta, \forall\theta_0 \in \Theta$,

$$H_{t,\theta} = H_{t,\theta_0}, \mathbb{P}_{\theta_0} - a.s. \Rightarrow \theta = \theta_0.$$

Finally, **A5** is obviously fulfilled. We summarize our findings so far in Theorem 2.

Theorem 2 *For the GARCH(p,q) process defined by* (3)–(4) *and* (7) *and for $\hat{\theta}_n$ as defined above, assume that:*

1. *Θ is compact, C, $\tilde{A}_i$, $\tilde{B}_i$ are continuous functions of θ, and there exists $c > 0$ such that* $\inf_{\theta\in\Theta} \det\, C(\theta) \geq c > 0$,

2. *The model is identifiable,*

3. *The rescaled errors ε_t admit a density absolutely continuous w.r.t. the Lebesgue measure and positive in a neighbourhood of the origin,*

4. $\forall\theta \in \Theta$, $\rho(\sum_{i=1}^{q} \tilde{A}_i(\theta) + \sum_{i=1}^{p} \tilde{B}_i(\theta)) < 1$.

Then $\hat{\theta}_n$ is strongly consistent, that is, $\hat{\theta}_n \to_{n\to+\infty} \theta_0, \quad \mathbb{P}_{\theta_0} - \text{a.s.}$

The result was stated without proof by Boussama [9]. Theorem 2 is valid only under a random initial condition drawn in the stationary law (see Jeantheau [19]). An extension is required for the fixed initial condition case and this we shall provide in Section 4.2.

4. ASYMPTOTIC NORMALITY

4.1. The Initial State Is Stationary

In this section we establish asymptotic normality of the quasi-MLE. We first assume that the initial conditions for H_t are in the stationary law. In the next subsection we will deal with the fixed initial state case. Basawa, Feigin and Heyde [2] gave conditions for asymptotic normality of the MLE for general stochastic processes. These conditions were previously employed by, among others, Weiss ([30], p.130) and Lumsdaine ([22], p.594). The conditions are:

(i) $-\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{\mathbb{P}} C_1$ when $T \to +\infty$ for a nonrandom positive definite matrix C_1.

(ii) $\dfrac{1}{\sqrt{T}}\displaystyle\sum_{t=1}^{T} \frac{\partial \ell_t(\theta_0)}{\partial\theta} \xrightarrow{\mathcal{L}} \mathcal{N}(0, C_0)$ when $T \to +\infty$ for a nonrandom C_0.

(iii) For all i, j, k, $\mathbb{E}\left(\displaystyle\sup_{\|\theta-\theta_0\|\le\delta} \left|\frac{\partial^3 \ell_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right)$ is bounded for all $\delta > 0$.

Similar conditions are given by Amemiya ([1], Theorem 4.1.3). By Theorem 1 and under the assumptions of Theorem 2, $\partial^2\ell_t(\theta_0)/\partial\theta\partial\theta'$ is ergodic and so, condition (i) will be satisfied if C_1 is finite and positive definite. As

$$\frac{\partial \ell_t}{\partial\theta_i}(\theta) = \mathrm{Tr}\left(\frac{\partial H_{t,\theta}}{\partial\theta_i} H_{t,\theta}^{-1} - X_t X_t' H_{t,\theta}^{-1} \frac{\partial H_{t,\theta}}{\partial\theta_i} H_{t,\theta}^{-1}\right),$$

we find, using (6), that

$$\mathbb{E}_{\theta_0}\left[\frac{\partial \ell_t}{\partial\theta_i}(\theta_0)|\mathcal{F}_{t-1}\right] = 0 \quad a.s.$$

Thus, the score is a martingale difference. Moreover, it also follows from Theorem 1 and under the assumptions of Theorem 2 that $\partial\ell_t(\theta_0)/\partial\theta$ is a strictly stationary and ergodic process, because it is a measurable function of a strictly stationary and ergodic process. Thus, we may apply the CLT for martingales (e.g., Billingsley [3], p. 788) to obtain condition (ii) above, as long as $C_0 = \mathbb{E}_{\theta_0}\{(\partial\ell_t(\theta_0)/\partial\theta)(\partial\ell_t(\theta_0)/\partial\theta')\}$ is finite. Note that we only require the finiteness of the second moment of $\partial\ell_t(\theta_0)/\partial\theta$ for the application of Billingsley's [3] martingale CLT, whereas Lumsdaine [22] required the finiteness of the $2+\delta$ order moment of $\partial\ell_t(\theta_0)/\partial\theta$, for some δ. Lumsdaine [22] applied Theorem 6.3 of Serfling [28] which imposes this additional restriction but reduces the assumptions {strict stationarity, ergodicity} to {weak stationarity, eq'n (6.7) of Serfling }. See Serfling ([28] p. 1174) for discussion. This additional restriction effectively forced Lumsdaine ([22], p. 594) to show existence of the third order moment. Finally, we note that condition (iii) above follows from Basawa et al.'s [2] condition B7. In our case, we shall see in the proofs that condition (iii) holds with the supremum taken over all Θ.

To prove asymptotic normality of the quasi-MLE, it will suffice then to verify the following conditions:

B1 $C_1 = \mathbb{E}\left(\left(\dfrac{\partial^2\ell_t(\theta_0)}{\partial\theta_i\partial\theta_j}\right)_{1\le i,j\le r}\right)$ is finite and positive definite.

B2 $C_0 = \mathbb{E}\left(\dfrac{\partial\ell_t(\theta_0)}{\partial\theta}\dfrac{\partial\ell_t(\theta_0)}{\partial\theta'}\right)$ is finite.

B3 Condition (iii) above.

In addition, we require that the components of ε_t (for a fixed t) are independent. These requirements are fullfilled in the Gaussian case, but of course not in general. We prove **B1-B3** in the Appendix.

Theorem 3 *Under the Assumptions:*

(i) (1)–(4) of Theorem 2, and $C(\theta)$, $\tilde{A}_i(\theta)$, $\tilde{B}_i(\theta)$ *admit continuous derivatives up to order 3 on* Θ,

(ii) The components of ε_t *are independent,*

(iii) X_t *admits bounded moments of order 8,*

(iv) The initial value (in H*) is drawn for the stationary ergodic law,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}}_{n\to\infty} \mathcal{N}(0, C_1^{-1}C_0C_1^{-1}), \text{ under } I\!P_{\theta_0}.$$

Note that if moreover $\varepsilon_t \sim \mathcal{N}(0, I)$, then $C_0 = 2C_1$ and the asymptotic law reduces to $\mathcal{N}(0, 2C_1^{-1})$.

4.2. The Initial State Is Fixed

We considered above a random initial condition for the process, drawn from the stationary law. Here we assume that the initial value of the process is fixed. Let $\mathcal{H}_t = (vech(H_t)', \ldots, vech(H_{t-m+1})')'$ where $m = \max(p, q)$. Let $x = \mathcal{H}_0 \in$

$\mathbb{R}_+^{md(d+1)/2}$ be the initial state. Let $h_{t,x,\theta}$ be the values of h_t given the initial state. $\hat{\theta}_{x,n}$ denotes the quasi-MLE given the initial state. That is, the value that minimizes

$$L_n(x,\theta) = \frac{1}{2n}\sum_{t=1}^{n} \ell_t(x,\theta)$$

with

$$\ell_t(x,\theta) = \log[\det(H_{t,x,\theta})] + X_t' H_{t,x,\theta}^{-1} X_t,$$

with $H_{x,t,\theta}$ built of the $h_{x,t,\theta}$'s. We establish the following result:

Theorem 4 *Under the Assumptions of Theorem 3, with the exception that the initial condition* $x \in \mathbb{R}_+^{md(d+1)/2}$ *of the process* $\mathcal{H}_t$ *is fixed,* $\hat{\theta}_{x,n}$ *is strongly consistent and*

$$\sqrt{n}(\hat{\theta}_{x,n} - \theta_0) \xrightarrow{\mathcal{D}}_{n\to+\infty} \mathcal{N}(0, C_1^{-1} C_0 C_1^{-1}), \quad \text{under} \quad \mathbb{P}_{\theta_0}.$$

From p.19 and p.41 of Jeantheau [18] and from Theorem 2.2 of Elie and Jeantheau [12], strong consistency is obtained under the additional condition

$$\sup_{\theta\in\Theta} |\ell_t(\theta) - \ell_t(\theta, x)| \longrightarrow 0 \;\; a.s.$$

The condition is proved in Appendix **B**. The asymptotic normality is then a consequence of this result and Theorem 3.

5. REMARKS

For the univariate GARCH(1,1) model, Lumsdaine [22] established consistency and asymptotic normality of the quasi-MLE under strong assumptions on the shape of the normalized innovation density and boundedness of the conditional moment of order 32. Together with the contributions of Weiss [30], Nelson [26], Lee and Hansen [20], and others, asymptotic theory for the univariate GARCH(1,1) model is fairly well covered. The univariate GARCH(p, q) is treated in Boussama [9, 10].

In this paper, we established asymptotic theory for the multivariate GARCH(p, q) model. The tools usually used in the univariate case do not seem to be suitable for the multivariate model. We appealed to Jeantheau's [19] conditions in proving strong consistency of the MLE. Asymptotic normality of the MLE is proven then with the aid of Basawa et al.'s [2] conditions. The results of the paper enable practitioners to apply tools of statistical inference in a justified manner, whereas previously these tools were only used as a rule of thumb.

APPENDIX A: Proof of Theorem 3.

The proof of Theorem 3 requires **B1-B3**. The proofs make extensive use of the relations (7)–(9). First, we find a deterministic bound for the norms of H_t^{-1}, since it will be useful in several points. For a positive definite matrix C and a positive semidefinite matrix D, we have:

$$\begin{aligned} 0 \leq \mathrm{Tr}[(C+D)^{-2}] &= \|C^{-1/2}(I+C^{-1/2}DC^{-1/2})^{-1}C^{-1/2}\|^2 \\ &\leq \mathrm{Tr}(C^{-2}(I+C^{-1/2}DC^{-1/2})^{-2}) \\ &\leq \left(\mathrm{Tr}(C^{-4})Tr((I+C^{-1/2}DC^{-1/2})^{-4})\right)^{1/2}. \end{aligned}$$

As the eigenvalues of $I+C^{-1/2}DC^{-1/2}$ are all greater than unity, those of its inverse are necessarily in $(0,1]$ as well as those of any power of the inverse. This implies that

$$\mathrm{Tr}(I+C^{-1/2}DC^{-1/2})^{-4} < d\ .$$

Thus:

$$\mathbf{N}\left(H_{t,\theta}^{-1}\right)^2 \leq \|H_{t,\theta}^{-1}\|^2 \leq \sqrt{d}\|C^{-2}(\theta)\| \leq K^2. \tag{15}$$

The bound is uniform in t and also uniform on Θ using Assumption 1 of Theorem 2 which implies that all eigenvalues admit a uniform lower bound. Equation (15) implies that if X admits finite moments of order 8, i.e., if $\mathbb{E}(\|X\|^8) < +\infty$, then $\mathbb{E}\|\varepsilon\|^8 < +\infty$, because

$$\begin{aligned} \|\varepsilon\|^8 &= \mathrm{Tr}^4(\varepsilon'\varepsilon) = \mathrm{Tr}^4(X'H^{-1}X) \\ &= \mathrm{Tr}^4(XX'H^{-1}) \leq K^4\|X\|^8. \end{aligned}$$

Lemma 1 *Denote $\dot{H}_{t,i} = \partial H_t/\partial\theta_i$ and $\mu_{4,p} = \mathbb{E}(\varepsilon_{t,p}^4)$. Then*

$$\begin{aligned} \mathbb{E}_{t-1}\left[\frac{\partial \ell_t}{\partial\theta_i}(\theta_0)\frac{\partial \ell_t}{\partial\theta_j}(\theta_0)\right] &= \sum_{p=1}^{d}(\mu_{4,p}-3)[H_t^{-1/2}\ \dot{H}_{t,i}\ H_t^{-1/2}]_{pp}[H_t^{-1/2}\ \dot{H}_{t,j}\ H_t^{-1/2}]_{pp} \\ &\quad +2\mathrm{Tr}(\dot{H}_{t,i}\ H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1}) \end{aligned} \tag{16}$$

and

$$\mathbb{E}_{t-1}\left[\frac{\partial^2 \ell_t}{\partial\theta_i\partial\theta_j}(\theta_0)\right] = \mathrm{Tr}(\dot{H}_{t,i}\ H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1}). \tag{17}$$

Proof of Lemma 1. For simplicity, H_t denotes $H_{t,\theta}$ in the first two equalities and H_{t,θ_0} later on. First,

$$\frac{\partial \ell_t}{\partial\theta_i}(\theta) = \mathrm{Tr}\left[\dot{H}_{t,i}\ H_t^{-1} - X_tX_t'H_t^{-1}\ \dot{H}_{t,i}\ H_t^{-1}\right]$$

and

$$\begin{aligned} \frac{\partial^2 \ell_t}{\partial\theta_j\partial\theta_i}(\theta) &= \mathrm{Tr}\left[\ddot{H}_{t,i,j}\ H_t^{-1} - \dot{H}_{t,i}\ H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1} + X_tX_t'H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1}\ \dot{H}_{t,i}\ H_t^{-1}\right. \\ &\quad \left. -X_tX_t'H_t^{-1}\ \ddot{H}_{t,i,j}\ H_t^{-1} + X_tX_t'H_t^{-1}\ \dot{H}_{t,i}\ H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1}\right]. \end{aligned} \tag{18}$$

Using the fact that all terms in H_t and its derivatives are in $\mathcal{F}_{t-1}$, we obtain (17). Further,

$$\begin{aligned}
\mathbb{E}_{t-1}\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta_i}\frac{\partial \ell_t(\theta_0)}{\partial \theta_j}\right) &= \mathbb{E}_{t-1}\left(\mathrm{Tr}(X_tX_t'H_t^{-1}\ \dot{H}_{t,i}\ H_t^{-1})\mathrm{Tr}(X_tX_t'H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1})\right) \\
&\quad -\mathbb{E}_{t-1}\left(\mathrm{Tr}(X_tX_t'H_t^{-1}\ \dot{H}_{t,i}\ H_t^{-1})\mathrm{Tr}(\dot{H}_{t,j}\ H_t^{-1})\right) \\
&\quad -\mathbb{E}_{t-1}\left(\mathrm{Tr}(X_tX_t'H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1})\mathrm{Tr}(\dot{H}_{t,i}\ H_t^{-1})\right) \\
&\quad +\mathbb{E}_{t-1}\left(\mathrm{Tr}(\dot{H}_{t,i}\ H_t^{-1})\mathrm{Tr}(\dot{H}_{t,j}\ H_t^{-1})\right) \\
&= \mathbb{E}_{t-1}\left(\mathrm{Tr}(X_tX_t'H_t^{-1}\ \dot{H}_{t,i}\ H_t^{-1})\mathrm{Tr}(X_tX_t'H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1})\right) \\
&\quad -\mathrm{Tr}(\dot{H}_{t,i}\ H_t^{-1})\mathrm{Tr}(\dot{H}_{t,j}\ H_t^{-1}).
\end{aligned}$$

Let $H_t^{-1/2}$ be a symmetric root of H_t and $M_i = H_t^{-1/2}\ \dot{H}_{t,i}\ H_t^{-1/2}$. Then

$$\begin{aligned}
\mathbb{E}_{t-1}(\mathrm{Tr}(\varepsilon_t\varepsilon_t'M_i)\mathrm{Tr}(\varepsilon_t\varepsilon_t'M_j)) &= \mathbb{E}_{t-1}\left[\left(\sum_{r=1}^{d}\sum_{u=1}^{d}\varepsilon_{t,r}\varepsilon_{t,u}\,[M_i]_{k,p}\right)\left(\sum_{s=1}^{d}\sum_{v=1}^{d}\varepsilon_{t,s}\varepsilon_{t,v}\,[M_j]_{l,q}\right)\right] \\
&= \sum_{u=1}^{d}\sum_{r=1}^{d}\sum_{s=1}^{d}\sum_{v=1}^{d}[M_i]_{r,u}\,[M_j]_{v,s}\,\mathbb{E}\left(\varepsilon_{t,r}\varepsilon_{t,u}\varepsilon_{t,s}\varepsilon_{t,v}\right) \\
&= \sum_{r=1}^{d}[M_i]_{r,r}\,[M_j]_{r,r}\,(\mu_{4,r}-3) + \sum_{r=1}^{d}\sum_{s=1}^{d}[M_i]_{r,r}\,[M_j]_{s,s} \\
&\quad +\sum_{r=1}^{d}\sum_{u=1}^{d}[M_i]_{r,u}\,[M_j]_{r,u} + \sum_{r=1}^{d}\sum_{u=1}^{d}[M_i]_{u,r}\,[M_j]_{r,u} \\
&= \sum_{r=1}^{d}(\mu_{4,r}-3)\,[M_i]_{r,r}\,[M_j]_{r,r} + \mathrm{Tr}(M_i)\mathrm{Tr}(M_j) \\
&\quad +2\mathrm{Tr}(M_iM_j)\ ,
\end{aligned}$$

where we used that $\mathbb{E}_{t-1}(\varepsilon_{t,r}\varepsilon_{t,u}\varepsilon_{t,s}\varepsilon_{t,v}) = \mathbb{E}(\varepsilon_{t,r}\varepsilon_{t,u}\varepsilon_{t,s}\varepsilon_{t,v}) = \mu_{4,r}$ if $r = s = u = v$, 1 if $r = u, s = v$ with $r \neq s$, or $r = s, u = v$ with $r \neq u$ or $r = v, u = s$ with $r \neq u$ and 0 otherwise. The Lemma is proved on recalling that $\mathrm{Tr}(M_i)\mathrm{Tr}(M_j) = \mathrm{Tr}(\dot{H}_{t,i}\ H_t^{-1})\mathrm{Tr}(\dot{H}_{t,j}\ H_t^{-1})$. □

Proof of B1. From eq'n (14),

$$\mathbb{E}_{t-1}\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta_i \partial \theta_j}\right) = |Tr(\dot{H}_{t,i}\ H_t^{-1}\ \dot{H}_{t,j}\ H_t^{-1})| \leq \|H_t^{-1}\|^2\|\ \dot{H}_{t,i}\ \|\|\ \dot{H}_{t,j}\ \|$$

so that

$$\left|\mathbb{E}\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta_i \partial \theta_j}\right)\right| \leq K^2\mathbb{E}^{1/2}(\|\ \dot{H}_{t,i}\ \|^2)\mathbb{E}^{1/2}(\|\ \dot{H}_{t,j}\ \|^2).$$

To show that C_1 is finite, we require the following Lemma.

Lemma 2 *Assume that the true model for Y is such that X is strictly stationary and admits moments till order 4, and in particular that the initial condition in H is given (and depends only on θ_0) and is drawn in the stationary law. Then for all $1 \leq k, l \leq d$, $i = 1, \ldots, r$,*

$$I\!\!E \left\{ \sup_{\theta \in \Theta} \left[\frac{\partial H_t}{\partial \theta_i}(\theta) \right]^2_{k,l} \right\} < +\infty.$$

The uniformity requirement of Lemma 2 is only needed for **B3**.

Proof of Lemma 2. Let $\mathcal{X}_t$ and $\mathcal{H}_t$ be defined by

$$\begin{aligned} \mathcal{X}_t &= (vech(X_t X_t')', \ldots, vech(X_{t-q+1} X'_{t-m+1}))', \\ \mathcal{H}_t &= (vech(H_t)', \ldots, vech(H_{t-m+1})')' \end{aligned} \tag{19}$$

where $m = \max(p, q)$ and let the vector $\mathcal{C}_1$ be defined by:

$$\mathcal{C}_1 = (vech(C)', 0, \ldots, 0)'$$

with size $md(d+1)/2$. Then $\mathcal{H}_{t+1} = \mathcal{C}_1 + B\mathcal{H}_t + A\mathcal{X}_t$, where A and B are defined by

$$A = \begin{pmatrix} \tilde{A}_1 & \ldots & \ldots & \ldots & \tilde{A}_m \\ I & 0 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & I & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \tilde{B}_1 & \ldots & \ldots & \ldots & \tilde{B}_m \\ I & 0 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & I & 0 \end{pmatrix}, \tag{20}$$

with convention $\tilde{A}_i = 0$ if $i > q$ and $\tilde{B}_i = 0$ if $i > p$. The model can be written as

$$\mathcal{H}_t(\theta) = \sum_{k=0}^{t-1} B^k(\theta)\mathcal{C}_1(\theta) + B^t(\theta)\mathcal{H}_0 + \sum_{k=0}^{t-1} B^k(\theta)A(\theta)L^k\mathcal{X}_{t-1}(\theta_0), \tag{21}$$

where L is the backshift operator $LX_t = X_{t-1}$.
Boussama [9] proved the following result:

Proposition 2

$$\rho\left(\sum_{i=1}^{q} \tilde{A}_i + \sum_{i=1}^{p} \tilde{B}_i\right) < 1 \Rightarrow \rho(\sum_{i=1}^{p} \tilde{B}_i) < 1 \quad \text{and} \quad \rho(\sum_{i=1}^{p} \tilde{B}_i) < 1 \Rightarrow \rho(B) < 1.$$

Thus Assumption 4 of Theorem 2 implies that $\forall \theta \in \Theta, \rho(B(\theta)) := \rho_1(\theta) < 1$. We shall denote by $\rho_0 = \sup_{\theta \in \Theta} \rho_1(\theta)$ and ρ_1 being continuous we know that $\rho_0 < 1$ (see Horn and Johnson [17], for the continuity of the eigenvalues). This allows to prove the following Lemma:

Lemma 3 *There exists a constant Ψ independent of θ such that $N(B^k) \leq \Psi k^{d_0} \rho_0^k$ for all $k \geq 1$, where $d_0 = md(d+1)/2$.*

Since under our assumptions,

$$\frac{\partial}{\partial\theta_i}\mathcal{H}_0 = \frac{\partial}{\partial\theta_i}\mathcal{X}_t = 0,$$

because $\mathcal{H}_0$ is fixed and $\mathcal{X}$ depends on θ_0 but is not a function of θ, we have,

$$\frac{\partial\mathcal{H}_t}{\partial\theta_i} = \frac{\partial}{\partial\theta_i}\left(\sum_{k=0}^{t-1} B^k\mathcal{C}_1\right) + \frac{\partial}{\partial\theta_i}(B^t)\mathcal{H}_0 + \frac{\partial}{\partial\theta_i}\left(\sum_{k=0}^{t-1} B^k L^k A\right)\mathcal{X}_{t-1}. \qquad (22)$$

As

$$\frac{\partial B^k}{\partial\theta_i} = \sum_{j=0}^{k-1} B^j \frac{\partial B}{\partial\theta_i} B^{k-1-j},$$

we get

$$\left\| B^j \frac{\partial B}{\partial\theta_i} B^{k-1-j} \right\| \leq \mathbf{N}(B^j)\left\|\frac{\partial B}{\partial\theta_i}\right\| \mathbf{N}(B^{k-1-j}), \quad j = 0, \dots, k-1,$$

so that for $j = 0, \dots, k-1$, using Lemma 3,

$$\|B^j \frac{\partial B}{\partial\theta_i} B^{k-1-j}\| \leq \Psi^2 k^{d_0} \rho_0^{k-1} \left\|\frac{\partial B}{\partial\theta_i}\right\|.$$

First, we note that the norms of the derivatives $\mathbf{N}(\partial B/\partial\theta_i)$, $\mathbf{N}(\partial A/\partial\theta_i)$ and $\|\partial\mathcal{C}_1/\partial\theta_i\|$ are uniformly bounded on Θ because these derivatives are all continuous functions of the parameters and θ belongs to a compact set. Then we denote by $S_i = \sup_{\theta\in\Theta}\mathbf{N}(\partial A/\partial\theta_i)$, $S_i' = \sup_{\theta\in\Theta}\mathbf{N}(\partial B/\partial\theta_i)$ and $S"_i = \sup_{\theta\in\Theta}\|\partial\mathcal{C}_1/\partial\theta_i\|$. Moreover let $S_0 = \sup_{\theta\in\Theta}\mathbf{N}(A)$. The derivative given by (22) involves three terms to bound:

$$\begin{aligned}
\left\|\frac{\partial}{\partial\theta_i}\left(\sum_{k=0}^{t-1} B^k\mathcal{C}_1\right)\right\| &= \left\|\sum_{k=1}^{t-1}\frac{\partial B^k}{\partial\theta_i}\mathcal{C}_1 + \sum_{k=0}^{t-1} B^k\frac{\partial\mathcal{C}_1}{\partial\theta_i}\right\| \\
&\leq \Psi^2\left\|\frac{\partial B}{\partial\theta_i}\right\|\sum_{k=1}^{t-1} k^{d_0}\rho_0^{k-1} + \Psi\left\|\frac{\partial\mathcal{C}_1}{\partial\theta_i}\right\|\sum_{k=0}^{t-1} k^{d_0}\rho_0^k \\
&\leq \frac{\Psi(d_0-1)!}{\rho_0(1-\rho_0)^{d_0}}\left(\Psi\left\|\frac{\partial B}{\partial\theta_i}\right\| + \left\|\frac{\partial\mathcal{C}_1}{\partial\theta_i}\right\|\right)
\end{aligned}$$

using $\sum_{k=1}^{t-1} k^{d_0}\rho_0^k \leq \sum_{k=1}^{\infty} k^{d_0}\rho_0^k = (d_0-1)!/(1-\rho_0)^{d_0}$. Of course, we implicitely assume that $\rho_0 \neq 0$, but if it is, the terms are straightforwardly bounded because B is then nilpotent and all sums are finite. Then we have:

$$\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_i}\left(\sum_{k=0}^{t-1} B^k\mathcal{C}_1\right)\right\| \leq \frac{\Psi(d_0-1)!\sqrt{d_0}}{\rho_0(1-\rho_0)^{d_0}}\left(\Psi S_i' + S"_i\right).$$

In the same way,

$$\left\|\frac{\partial}{\partial\theta_i}(B^t)\mathcal{H}_0\right\| \leq \frac{(d_0-1)!\Psi}{\rho_0(1-\rho_0)^{d_0}}\left\|\frac{\partial B}{\partial\theta_i}\right\| \|\mathcal{H}_0\|.$$

Finally,

$$\begin{aligned}
\left\| \frac{\partial}{\partial \theta_i} \left(\sum_{k=0}^{t-1} B^k L^k A \right) \mathcal{X}_{t-1} \right\| &\leq \left\| \left[\sum_{k=0}^{t-1} \left(\frac{\partial}{\partial \theta_i} B^k L^k A \right) \right] \mathcal{X}_{t-1} \right\| \\
&\quad + \left\| \left[\sum_{k=0}^{t-1} B^k L^k \left(\frac{\partial}{\partial \theta_i} A \right) \right] \mathcal{X}_{t-1} \right\| \\
&= \|T_1\| + \|T_2\|.
\end{aligned}$$

Then for the second term, we write

$$\|T_2\| \leq \Psi^2 \sum_{k=0}^{t-1} k^{d_0} \rho_0^k \mathbf{N} \left(\frac{\partial A}{\partial \theta_i} \right) \|\mathcal{X}_{t-k}\|,$$

and

$$\begin{aligned}
\mathbb{E} \left(\sup_{\theta \in \Theta} \|T_2\|^2 \right) &\leq \Psi^2 S_i^2 \sum_{k,k'=0}^{t-1} k^{d_0} (k')^{d_0} \rho_0^{k+k'} \mathbb{E} \left(\|\mathcal{X}_{t-k}\| \|\mathcal{X}_{t-k'}\| \right) \\
&\leq \Psi^2 S_i^2 \left(\sum_{k=0}^{t-1} k^{d_0} \rho_0^k \mathbb{E}^{1/2} \left(\|\mathcal{X}_{t-k}\|^2 \right) \right)^2 \\
&\leq \Psi^2 S_i^2 \frac{[(d_0 - 1)!]^2}{(1 - \rho_0)^{2d_0}} \mathbb{E}(\|\mathcal{X}_0\|^2).
\end{aligned}$$

For the first term we have

$$\begin{aligned}
\|T_1\| &\leq \Psi^2 \left(\sum_{k=1}^{t-1} k^{d_0+1} \rho_0^{k-1} \mathbf{N}(A) \|\mathcal{X}_{t-k-1}\| \right) \mathbf{N} \left(\frac{\partial B}{\partial \theta_i} \right) \\
&\leq \Psi^2 \left(\sum_{k=1}^{t-1} k^{d_0+1} \rho_1^{k-1} \|\mathcal{X}_{t-k-1}\| \right) \mathbf{N}(A) \mathbf{N} \left(\frac{\partial B}{\partial \theta_i} \right),
\end{aligned}$$

So,

$$\begin{aligned}
\mathbb{E} \left(\sup_{\theta \in \Theta} \|T_1\|^2 \right) &\leq \Psi^2 S_0^2 (S_i')^2 \mathbb{E} \left(\sum_{k=1}^{t-1} k^{d_0+1} \rho_0^{k-1} \|\mathcal{X}_{t-k-1}\| \right)^2 \\
&\leq \Psi^2 S_0^2 (S_i')^2 \mathbb{E} \left(\sum_{k,k'=1}^{t-1} (k^{d_0+1} \rho_0^{k-1})((k')^{d_0+1} \rho_0^{k'-1}) \|\mathcal{X}_{t-k-1}\| \|\mathcal{X}_{t-k'-1}\| \right) \\
&\leq \frac{(d_0! \Psi S_0 S_i')^2}{[\rho_0 (1 - \rho_0)^{d_0}]^2} \mathbb{E}(\|\mathcal{X}_0\|^2).
\end{aligned}$$

As $\mathbb{E}(\|\mathcal{X}_0\|^2) < \infty$, the proof is completed. □

Proof of Lemma 3. If we can find a complex unitary matrix P which satisfies $P^* = P^{-1}$, P^* being the conjugate transpose of P, such that $B = P^* D P$, D diagonal, then it immediately follows that

$$\mathbf{N}(B^k) = \mathbf{N}(P^* D^k P) = \rho^{1/2}(P^* (DD^*)^k P) = \rho^{1/2}((DD^*)^k) = \rho_1^k(B)$$

so that the result holds with $\Psi = 1$. In general though, B cannot be diagonalized. In this case, we can still find a complex unitary matrix P and a complex lower triangular matrix T such that $B = P^*TP$. The diagonal terms of T are the eigenvalues of B. See Magnus and Neudecker ([23], Theorem 12). Write $T = D + L$ where D is diagonal and L is lower triangular with null diagonal. It follows that L is nilpotent with $L^{d_0} = 0$. We have $\mathbf{N}(B^k) = \mathbf{N}(T^k)$ as above and

$$\begin{aligned}
\mathbf{N}(T^k) &\leq \mathbf{N}(D^k) + \sum_{j=1}^{d_0} \binom{k}{j} \mathbf{N}(L)^j \mathbf{N}(D)^{k-j} \\
&\leq \rho_1^k + \sum_{j=1}^{d_0} \binom{k}{j} \mathbf{N}(L)^j \rho_1^{k-j},
\end{aligned}$$

for any $k \geq d_0$. Let $k = d_0 + n$. Then,

$$\mathbf{N}(T^k) \leq \rho_1^n \left(\rho_1^{d_0} + \sum_{j=1}^{d_0} \binom{k}{j} N(L)^j \rho_1^{d_0-j} \right),$$

and

$$\binom{k}{j} = \binom{d_0}{j} \prod_{p=1}^{n} \left(\frac{1}{1 - \frac{j}{d_0+p}} \right).$$

But for any real x, $0 \leq x \leq 1/d_0$, $-\log(1-x) \leq (d_0/(1+d_0))x$, so that for any $j \geq 1$,

$$\begin{aligned}
\log \left[\prod_{p=1}^{n} \left(\frac{1}{1 - \frac{j}{d_0+p}} \right) \right] &= -\sum_{p=1}^{n} \log \left(1 - \frac{j}{d_0+p} \right) \leq \frac{d_0}{1+d_0} \sum_{p=1}^{n} \frac{j}{d_0+p} \\
&\leq \frac{j}{1+d_0} \sum_{p=1}^{n} \frac{1}{1+p/d_0} \leq \frac{jd_0}{1+d_0} \log \left(1 + \frac{n}{d_0} \right) \\
&\leq d_0 \log \left(1 + \frac{n}{d_0} \right)
\end{aligned}$$

as $j \leq d_0$. Thus,

$$\binom{k}{j} \leq \binom{d_0}{j} \left(1 + \frac{n}{d_0} \right)^{d_0}$$

and

$$\mathbf{N}(T^{d_0+n}) \leq \rho_1^n \left(1 + \frac{n}{d_0} \right)^{d_0} (\mathbf{N}(L) + \rho_1)^{d_0}. \tag{23}$$

It is thus clear from (23) that

$$N(T^k) \leq \left(\frac{1 + N(L)/\rho_0}{d_0} \right)^{d_0} k^{d_0} \rho_0^k.$$

Since

$$\mathbf{N}(L) = \mathbf{N}(T - D) \leq \mathbf{N}(T) + \mathbf{N}(D) \leq \rho_0 + \mathbf{N}(B),$$

we find the result with $\Psi = [(2 + \sup_{\theta\in\Theta} \mathbf{N}(B(\theta))/\rho_0)/d_0]^{d_0}$, where the supremum is finite since B is a continuous function of θ and Θ is compact. □

Next, we prove that C_1 is positive definite. Let $A_i = H^{-1/2} \dot{H}_i H^{-1/2}$ where the t index is omitted for brevity and $H^{-1/2}$ is a symmetric root of H^{-1}. Using that for any matrices A, B, $\mathrm{Tr}(AB) = vec(A')'\, vec(B)$, we find that the matrix $\mathbb{E}_{t-1}(\partial^2\ell_t(\theta_0)/\partial\theta^2)$ is equal to $2WW'$ where $W' = (vec(A_1)|vec(A_2)|\ldots|vec(A_r))$. Then using that $vec(ABC) = (C' \otimes A)vec(B)$ (see Magnus and Neudecker [23], Theorem 2 p. 30), we have $vec(A_i) = (H^{-1/2} \otimes H^{-1/2})vec(\dot{H}_i)$. Using now that $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, we have

$$\begin{aligned}
\mathbb{E}_{t-1}\left(\partial^2\ell_t(\theta_0)/\partial\theta^2\right) &= 2WW' = 2\left(vec(A_i)'vec(A_j)\right)_{1\le i,j\le r} \\
&= 2\left(vec(\dot{H}_i)'(H^{-1/2}\otimes H^{-1/2})(H^{-1/2}\otimes H^{-1/2})vec(\dot{H}_j)\right)_{1\le i,j\le r} \\
&= 2\left(vec(\dot{H}_i)'(H^{-1}\otimes H^{-1})vec(\dot{H}_j)\right)_{1\le i,j\le r} \\
&= 2P'(H^{-1}\otimes H^{-1})P
\end{aligned}$$

where $P = (vec(\dot{H}_1)|vec(\dot{H}_2)|\ldots|vec(\dot{H}_r))$. We know that $H^{-1}\otimes H^{-1}$ is positive definite. Further, the eigenvalues of $A\otimes B$ are $\lambda_i\mu_j$ if λ_i are the eigenvalues of A and μ_j are those of B (see Magnus and Neudecker [23], Theorem 1, p. 28). This implies that the eigenvalues of $H^{-1}\otimes H^{-1}$ are positive, H^{-1} being positive definite. It follows that C_1 is at least positive semi-definite. Now, assume that C_1 is not full rank. Then there exists a vector x, independent of t, such that

$$x'\mathbb{E}\left(P_t'(H_t^{-1}\otimes H_t^{-1})P_t\right)x = 0.$$

But $x'\mathbb{E}(P_t'(H_t^{-1}\otimes H_t^{-1})P_t)x = \mathbb{E}((P_tx)'(H_t^{-1}\otimes H_t^{-1})P_tx) = 0$. As the term under the expectation is nonnegative, it is necessarily zero, and $H_t^{-1}\otimes H_t^{-1}$ being positive definite we deduce that $P_tx = 0, \mathbb{P}_{\theta_0} - a.s., \forall t\in\mathbb{N}$. This implies that there exists a vector y such that

$$y'\frac{\partial h_t}{\partial\theta} = \sum_{i=1}^{r} y_i'\frac{\partial h_t}{\partial\theta_i} = 0, \quad \mathbb{P}_{\theta_0}\text{-a.s.}, \quad \forall t\in\mathbb{N},$$

using the notations of Lemma 2 and denoting by $\partial h_t/\partial\theta$ the vector

$$(\partial h_t'/\partial\theta_1, \partial h_t'/\partial\theta_2, \ldots, \partial h_t'/\partial\theta_r)'.$$

Then differentiating relation (10)

$$y'\left(\frac{\partial vech(C)}{\partial\theta} + \sum_{i=1}^{m}\frac{\partial\tilde{A}_i}{\partial\theta}\eta_{t-i} + \sum_{i=1}^{m}\frac{\partial\tilde{B}_i}{\partial\theta}h_{t-i}\right) = 0, \mathbb{P}_{\theta_0}\text{-a.s.}$$

or

$$\left(y'\frac{\partial vech(C)}{\partial\theta}\right) + \sum_{i=1}^{m}\left(y'\frac{\partial\tilde{A}_i}{\partial\theta}\right)\eta_{t-i} + \sum_{i=1}^{m}\left(y'\frac{\partial\tilde{B}_i}{\partial\theta}\right)h_{t-i} = 0, \mathbb{P}_{\theta_0}\text{-a.s.}$$

This would allow one to find another representation of h_t and of the model and imply a contradiction of the identifiability conditions which ensure (see Engle and Kroner [14]) that the representation is unique.

We have thus shown that C_1 is finite and positive definite.□

Proof of B2. The expectation of the second term on the rhs of (16) equals $2C_1$ which is finite by **B1**. As for the first term,

$$\begin{aligned}
& \left| \mathbb{E} \sum_{p=1}^{d} (\mu_{4,p} - 3)[H_t^{-1/2} \dot{H}_{t,i} H_t^{-1/2}]_{pp} [H_t^{-1/2} \dot{H}_{t,j} H_t^{-1/2}]_{pp} \right| \\
\leq & \max_{1 \leq p \leq d} |\mu_{4,p} - 3| \mathbb{E} \left[\sum_{p=1}^{d} [H_t^{-1/2} \dot{H}_{t,i} H_t^{-1/2}]_{pp}^2 \sum_{p=1}^{d} [H_t^{-1/2} \dot{H}_{t,j} H_t^{-1/2}]_{pp}^2 \right]^{1/2} \\
\leq & \max_{1 \leq p \leq d} |\mu_{4,p} - 3| \left\{ \mathbb{E} \left(\sum_{p=1}^{d} [H_t^{-1/2} \dot{H}_{t,i} H_t^{-1/2}]_{pp}^2 \right) \mathbb{E} \left(\sum_{p=1}^{d} [H_t^{-1/2} \dot{H}_{t,j} H_t^{-1/2}]_{pp}^2 \right) \right\}^{1/2} \\
\leq & \max_{1 \leq p \leq d} |\mu_{4,p} - 3| \left\{ \mathbb{E} \left(\mathrm{Tr}(\dot{H}_{t,i} H_t^{-1} \dot{H}_{t,i} H_t^{-1}) \right) \mathbb{E} \left(\mathrm{Tr}(\dot{H}_{t,j} H_t^{-1} \dot{H}_{t,j} H_t^{-1}) \right) \right\}^{1/2}
\end{aligned}$$

using that for $M = (m_{i,j})_{1 \leq i,j \leq d}$, $\sum_{i=1}^{d} m_{i,i}^2 \leq \sum_{i=1}^{d} \sum_{k=1}^{d} m_{i,k}^2 = \mathrm{Tr}(MM')$ and that $\mathrm{Tr}(H_t^{-1/2} M) = \mathrm{Tr}(M H_t^{-1/2})$. So the finiteness of C_0 is ensured by **B1**. □

Proof of B3. The third order log-likelihood derivative involves terms of the form $Tr(\dddot{H}_{t,i,j,k} H_t^{-1})$, $Tr(\ddot{H}_{t,i,j} H_t^{-1} \dot{H}_{t,k} H_t^{-1})$, $Tr(\dot{H}_{t,i} H_t^{-1} \dot{H}_{t,j} H_t^{-1} \dot{H}_{t,k} H_t^{-1})$, or the traces of the same matrices premultiplicated by $X_t X_t' H_t^{-1}$. Thus, for instance

$$\begin{aligned}
& \mathbb{E} \left(\sup_{\theta \in \Theta} \left| Tr \left(X_t X_t' H_t^{-1} \dot{H}_{t,i} H_t^{-1} \dot{H}_{t,j} H_t^{-1} \dot{H}_{t,k} H_t^{-1} \right) \right| \right) \\
\leq & K^4 \mathbb{E} \left(\mathbf{N} (X_t X_t') \sup_{\theta \in \Theta} \left(\| \dot{H}_{t,i} \| \| \dot{H}_{t,j} \| \| \dot{H}_{t,k} \| \right) \right) \\
\leq & K^4 \mathbb{E}^{1/4} (\|X\|^8) \mathbb{E}^{1/4} (\sup_{\theta \in \Theta} \| \dot{H}_{t,i} \|^4) \mathbb{E}^{1/4} (\sup_{\theta \in \Theta} \| \dot{H}_{t,j} \|^4) \mathbb{E}^{1/4} (\sup_{\theta \in \Theta} \| \dot{H}_{t,k} \|^4).
\end{aligned}$$

It is clear from Lemma 2 that if X admits moments of order 8, since

$$\mathbb{E}^{1/4} (\sup_{\theta \in \Theta} \| \dot{H}_{t,i} \|^4) \leq \mathbb{E}^{1/4} \left(\sup_{\theta \in \Theta} \left\| \frac{\partial \mathcal{H}_t}{\partial \theta_i} \right\|^4 \right),$$

where $\mathcal{H}$ is defined by (19), the terms above are bounded. Indeed, the result of Lemma 2 can also be extended to get

$$\mathbb{E} \left(\sup_{\theta \in \Theta} \left[\frac{\partial H_t}{\partial \theta_i} \right]_{k,l}^4 \right) < K_1, \ \mathbb{E} \left(\sup_{\theta \in \Theta} \left[\frac{\partial^2 H_t}{\partial \theta_i \partial \theta_j} \right]_{k,l}^2 \right) < K_2, \ \mathbb{E} \left(\sup_{\theta \in \Theta} \left[\frac{\partial^3 H_t}{\partial \theta_i \partial \theta_j \partial \theta_k} \right]_{k,l}^2 \right) < K_3$$

for $1 \leq k, l \leq d$, under our moment condition of order 8. □

APPENDIX B: Proof of Theorem 4.

First, $\hat{\theta}_{x,n}$ is strongly consistent (see Jeantheau [16] p. 19 and p. 41) if, for any $x \in \mathbb{R}_+^{md(d+1)/2}$,

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\ell_t(x,\theta)-\ell_t(\theta)\right]\right| \to 0 \text{ almost surely.} \tag{24}$$

To prove (24), it suffices to check that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}|\ell_t(\theta)-\ell_t(x,\theta)|\right]$$

is bounded by a summable sequence in t. Indeed then

$$\sum_{t=1}^{+\infty}\mathbb{P}\left(\sup_{\theta\in\Theta}|\ell_t(\theta)-\ell_t(x,\theta)|>\xi\right) < +\infty \text{ for all } \xi>0,$$

so that the Borel–Cantelli Lemma implies that $\sup_{\theta\in\Theta}|\ell_t(\theta)-\ell_t(x,\theta)|$ tends to zero almost surely. Cesaro's mean theorem implies then that (24) holds. Now

$$\begin{aligned}\ell_t(x,\theta)-\ell_t(\theta) &= \log\left(\det(I_r+H_{t,\theta}^{-1/2}(H_{t,x,\theta}-H_{t,\theta})H_{t,\theta}^{-1/2})\right)\\ &\quad +X_t'H_{t,\theta}^{-1}(H_{t,\theta}-H_{t,x,\theta})H_{t,x,\theta}^{-1}X_t.\end{aligned}$$

For the second term, we write

$$\begin{aligned}|X_t'H_{t,\theta}^{-1}(H_{t,\theta}-H_{t,x,\theta})H_{t,x,\theta}^{-1}X_t| &= |Tr(X_t'H_{t,\theta}^{-1}(H_{t,\theta}-H_{t,x,\theta})H_{t,x,\theta}^{-1}X_t)|\\ &= |Tr(H_{t,\theta}^{-1}(H_{t,\theta}-H_{t,x,\theta})H_{t,x,\theta}^{-1}X_tX_t')|\\ &\le \|X_tX_t'\|N(H_{t,\theta}^{-1})N(H_{t,x,\theta}^{-1})\|H_{t,\theta}-H_{t,x,\theta}\|\\ &\le K^2\|X_t\|^2\|H_{t,\theta}-H_{t,x,\theta}\|.\end{aligned}$$

Using (21), we have

$$\mathcal{H}_t-\mathcal{H}_{x,t}=B^t(\mathcal{H}_0-\mathcal{H}_{x,0})$$

where $\mathcal{H}_{x,0}$ is the initial condition for $\mathcal{H}$ associated with x. This implies that

$$\begin{aligned}\mathbf{N}\left(H_{t,x}-H_t\right) &\le \|H_{t,x}-H_t\| \le \|\mathcal{H}_t-\mathcal{H}_{x,t}\| \le \mathbf{N}\left(B^t\right)\|\mathcal{H}_0-\mathcal{H}_{x,0}\|\\ &\le \Psi t^{d_0}\rho_0^t\|\mathcal{H}_0-\mathcal{H}_{x,0}\| \end{aligned}\tag{25}$$

by using Lemma 3. Therefore

$$\left|X_t'(H_{t,x,\theta}^{-1}-H_{t,\theta}^{-1})X_t\right| \le K^2\Psi\|\mathcal{H}_0-\mathcal{H}_{x,0}\|\|X_t\|^2t^{d_0}\rho_0^t.$$

This implies

$$\mathbb{E}\left[\sup_{\theta}\left|X_t'(H_{t,x,\theta}^{-1}-H_{t,\theta}^{-1})X_t\right|\right]=O(t^{d_0}\rho_0^t) \tag{26}$$

if $\mathbb{E}(\|X_t\|^4) < +\infty$, and the bound in (26) is summable. For the first term, let $\lambda_i(t,x,\theta)$ be the eigenvalues of the symmetric matrix $H_{t,\theta}^{-1/2}(H_{t,x,\theta} - H_{t,\theta})H_{t,\theta}^{-1/2}$, $i = 1, \dots, d$. Then

$$\log\left\{\det\left[I_r + H_{t,\theta}^{-1/2}(H_{t,x,\theta} - H_{t,\theta})H_{t,\theta}^{-1/2}\right]\right\} = \sum_{i=1}^{d}\log(1 + \lambda_i(t,x,\theta)).$$

We have

$$|\lambda_i(t,x,\theta)| \leq \rho(H_{t,\theta}^{-1/2}(H_{t,x,\theta} - H_{t,\theta})H_{t,\theta}^{-1/2}) \leq N(H_{t,\theta}^{-1/2})^2 N(H_{t,x,\theta} - H_{t,\theta})$$

and therefore, using (25), there exists a constant κ independent of θ such that

$$|\lambda_i(t,x,\theta)| \leq \kappa t^{d_0}\rho_0^t.$$

Then there exist some fixed t_0 such that for $t \geq t_0$, $|\lambda_i(t,x,\theta)| \leq 1/2$, and since $|\log(1+u)| \leq 2|u|$ for $|u| \leq 1/2$, for $t \geq t_0$,

$$\begin{aligned}\left|\log\left(\det(I_r + H_{t,\theta}^{-1/2}(H_{t,x,\theta} - H_{t,\theta})H_{t,\theta}^{-1/2})\right)\right| &\leq \sum_{i=1}^{d}|\log(1+\lambda_i(t,x,\theta))| \\ &\leq 2\sum_{i=1}^{d}|\lambda_i(t,x,\theta)| \leq 2d\kappa t^{d_0}\rho_0^t\end{aligned}$$

This implies that

$$\sup_{\theta\in\Theta}\left|\log\left[\det(H_{t,\theta}^{-1}H_{t,x,\theta})\right]\right| = O(t^{d_0}\rho_0^t) \qquad (27)$$

and the bound in (27) is summable (over t). The compactness of Θ is implicitely used in the above considerations. Gathering the above summable bounds (26) and (27) proves (24) and gives the strong consistency of $\hat{\theta}_{n,x}$.

For the asymptotic normality, write

$$\nabla^2 L_n(x,\theta_{n,x}^*)\sqrt{n}(\hat{\theta}_{x,n} - \theta_0) = \sqrt{n}\left(\nabla L_n(\theta_0) - \nabla L_n(x,\theta_0)\right) + \nabla^2 L_n(\theta_n^*)\sqrt{n}(\hat{\theta}_n - \theta_0) \qquad (28)$$

where $|\theta_{n,x}^* - \theta_0| \leq |\hat{\theta}_{n,x} - \theta_0|$ and $|\theta_n^* - \theta_0| \leq |\hat{\theta}_n - \theta_0|$. In view of (28), the sufficient conditions for $\hat{\theta}_{x,n}$ to have the same asymptotic distribution as $\hat{\theta}_n$ are:

$$\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{n}\left(\nabla\ell_t(\theta_0) - \nabla\ell_t(x,\theta_0)\right)\right| \to 0 \text{ in probability,} \qquad (29)$$

and

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\nabla^2\ell_t(x,\theta) - \nabla^2\ell_t(\theta)\right)\right| \to 0 \text{ in probability.} \qquad (30)$$

Using Markov's Inequality which implies that $\mathbb{P}(|X| \geq a) \leq \mathbb{E}(|X|^p)/a^p$ for any $p \geq 0$, we know that (29) holds if for all i,

$$\sum_{t=1}^{\infty}\mathbb{E}\left\|\frac{\partial}{\partial\theta_i}\ell_t(\theta_0) - \frac{\partial}{\partial\theta_i}\ell_t(x,\theta_0)\right\| \text{ is bounded.} \qquad (31)$$

We omit the index θ_0 for simplicity.

$$\begin{aligned}\left|\frac{\partial}{\partial\theta_i}\ell_t(\theta_0) - \frac{\partial}{\partial\theta_i}\ell_t(x,\theta_0)\right| &= \Big|X_t'(H_t^{-1}\,\dot{H}_{t,i}\,H_t^{-1} - H_{x,t}^{-1}\,\dot{H}_{t,x,i}\,H_{x,t}^{-1})X_t \\ &\quad -Tr(H_t^{-1}\,\dot{H}_{t,i}) - Tr(H_{x,t}^{-1}\,\dot{H}_{t,x,i})\Big| \ . \qquad (32)\end{aligned}$$

There are thus two types of terms to study. First,

$$\begin{aligned}& |Tr(H_t^{-1}\,\dot{H}_{t,i}) - Tr(H_{t,x}^{-1}\,\dot{H}_{t,x,i})| \\ \leq\ & |Tr(H_t^{-1}(\dot{H}_{t,i} - \dot{H}_{t,x,i}))| + |Tr((H_t^{-1} - H_{t,x}^{-1})\,\dot{H}_{t,x,i})| \\ \leq\ & \|H_t^{-1}\|\|\,\dot{H}_{t,i} - \dot{H}_{t,x,i})\| + \|H_t^{-1} - H_{t,x}^{-1}\|\|\,\dot{H}_{t,x,i}\,\| \\ \leq\ & \sqrt{d}K\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i} - \frac{\partial\mathcal{H}_{t,x}}{\partial\theta_i}\right\| + \mathbf{N}\left(H_t^{-1}\right)\mathbf{N}\left(H_{t,x}^{-1}\right)\|H_t - H_{t,x}\|\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i}\right\| \\ \leq\ & \sqrt{d}K\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i} - \frac{\partial\mathcal{H}_{t,x}}{\partial\theta_i}\right\| + K^2\|H_t - H_{t,x}\|\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i}\right\| \ . \qquad (33)\end{aligned}$$

By Lemma 2, $\mathbb{E}(\|\frac{\partial\mathcal{H}_t}{\partial\theta_i}\|^2) < \infty$. Moreover, we saw above that the series with general term $\mathbb{E}^{1/2}\left(\|H_t - H_{t,x}\|^2\right)$ is summable. For the other term in (33), it is easy to see that

$$\frac{\partial\mathcal{H}_t}{\partial\theta_i} - \frac{\partial\mathcal{H}_{t,x}}{\partial\theta_i} = \frac{\partial B^t}{\partial\theta_i}(\mathcal{H}_0 - \mathcal{H}_{0,x})$$

which implies, by using Lemma 3

$$\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i} - \frac{\partial\mathcal{H}_{t,x}}{\partial\theta_i}\right\| \leq \Psi^2 t^{d_0+1}\rho_0^{t-1}\mathbf{N}\left(\frac{\partial B}{\partial\theta}\right)\|\mathcal{H}_0 - \mathcal{H}_{0,x}\|$$

which is a convergent series in t. For the second term of (32), we write

$$\begin{aligned}& X_t'(H_t^{-1}\,\dot{H}_{t,i}\,H_t^{-1} - H_{t,x}^{-1}\,\dot{H}_{t,x,i}\,H_{t,x}^{-1})X_t \\ =\ & X_t'(H_t^{-1} - H_{t,x}^{-1})\,\dot{H}_{t,i}\,H_t^{-1}X_t + X_t'H_{t,x}^{-1}(\dot{H}_{t,i} - \dot{H}_{t,x,i})H_t^{-1}X_t \\ & +X_t'H_{t,x}^{-1}\,\dot{H}_{t,x,i}\,(H_t^{-1} - H_{t,x}^{-1})X_t'\end{aligned}$$

and we bound each of the three terms in the same way. For instance

$$\begin{aligned}& \mathbb{E}\left|X_t'\left(H_t^{-1} - H_{t,x}^{-1}\right)\dot{H}_{t,i}\,H_t^{-1}X_t\right| \\ =\ & \mathbb{E}\left|X_t'H_t^{-1}\left(H_t - H_{t,x}\right)H_{t,x}^{-1}\,\dot{H}_{t,i}\,H_t^{-1}X_t\right| \\ =\ & \mathbb{E}\left|Tr\left(H_{t,x}^{-1}\,\dot{H}_{t,i}\,H_t^{-1}X_tX_t'H_t^{-1}\left(H_t - H_{t,x}\right)\right)\right| \\ \leq\ & \mathbb{E}\left(\|\,\dot{H}_{t,i}\,\|\mathbf{N}\left(H_t^{-1}\right)\left\|H_{t,x}^{-1}X_tX_t'H_t^{-1}\right\|\|H_t - H_{t,x}\|\right) \\ \leq\ & \mathbb{E}\left(\|\,\dot{H}_{t,i}\,\|\mathbf{N}\left(H_t^{-1}\right)\mathbf{N}\left(H_{t,x}^{-1}H_t^{-1}\right)\mathbf{N}\left(X_tX_t'\right)\|H_t - H_{t,x}\|\right) \\ \leq\ & K^3\mathbb{E}^{1/4}(\mathbf{N}\left(X_tX_t'\right)^4)\mathbb{E}^{1/4}(\|\,\dot{H}_{t,i}\,\|^4)\mathbb{E}^{1/2}(\|H_t - H_{t,x}\|^2) \\ \leq\ & K^3\mathbb{E}^{1/4}(\|X_t\|^4)\mathbb{E}^{1/4}\left(\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i}\right\|^4\right)\mathbb{E}^{1/2}(2\|\mathcal{H}_t - \mathcal{H}_{t,x}\|^2) \\ \leq\ & K^3\Psi t^{d_0}\rho_0^t\mathbb{E}^{1/4}(\|X_t\|^4)\mathbb{E}^{1/4}\left(\left\|\frac{\partial\mathcal{H}_t}{\partial\theta_i}\right\|^4\right)\mathbb{E}^{1/2}(2\|\mathcal{H}_0 - \mathcal{H}_{0,x}\|^2) \ .\end{aligned}$$

Thus again, the general term which appears here is summable (as of order constant $\times t^{d_0}\rho_0^t$). The two other terms can be treated in the same way. Note that our bounds are uniform on Θ.
The same method is suitable for dealing with (30), i.e., when looking for a uniform bound on

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(x,\theta)\right|\right)$$

with the second order derivatives given by (18). □

References

[1] T. Amemiya, "Advanced Econometrics", Blackwell, 1985.

[2] I.V. Basawa, P.D. Feigin and C.C. Heyde, Asymptotic properties of maximum likelihood estimators for stochastic processes, *Sankhya* **38** (1976), Series A, 259–270.

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

[4] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *J. Economet.* **31** (1986), 307–327.

[5] T. Bollerslev, Modelling the coherence in short run nominal exchange rates: a multivariate generalized ARCH model, *in* ARCH Selected Readings, edited by R.F. Engle, 300–313, Oxford University Press, 1995.

[6] T. Bollerslev and J.M. Wooldridge, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Economet. Rev.* **11** (1992), 143–17.

[7] P. Bougerol, Kalman filtering with random coefficients and contractions, *SIAM J. Control and Optim.* **31** (1993), 942–959.

[8] P. Bougerol and N. Picard, Stationarity of GARCH processes and some non-negative time series, *J. Economet.* **52** (1992), 115–127.

[9] F. Boussama, Ergodicity, Mixing and Estimation in GARCH Models, Ph.D. dissertation, University Paris 7, 1998.

[10] F. Boussama, Asymptotic normality for the quasi-maximum likelihood estimator of a GARCH model, *C. R. Acad. Sci. Paris* Série I 331 (2000), 81-84.

[11] F. Comte and O. Lieberman, Second-order noncausality in multivariate GARCH processes, *J. Time Ser. Anal.* **21** (2000), 5, 535–557.

[12] L. Elie and T. Jeantheau, Estimation in conditionally heteroskedastic models, *C.R. Acad. Sci. Paris*, Série I 320 (1995), 1255-1258.

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

[14] R.F. Engle, and K.F. Kroner, Multivariate simultaneous generalized ARCH, *Economet. Theory* **11** (1995), 122–150.

[15] E.J. Hannan and M. Deistler, The statistical theory of linear systems, Wiley, 1988.

[16] E. Hansen and A. Rahbek, Stationarity and asymptotics of multivariate ARCH time series with an application to robustness of cointegration analysis, Preprint, University of Copenhagen, December 1998.

[17] R.A. Horn and C.R. Johnson, "Matrix Analysis", *Cambridge University Press*: Cambridge, 1985.

[18] T. Jeantheau, Modèles autorégressifs à erreurs conditionnellement heteroscedastiques, PhD Thesis, University Paris 7, (1993).

[19] T. Jeantheau, Strong consistency of estimators for multivariate ARCH models, *Economet. Theory* **14** (1998), 70–86.

[20] S.W. Lee and B.E. Hansen, Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Economet. Theory* **10** (1994), 29–52.

[21] S. Ling and W.K. Li, Limiting distributions of maximum likelihood estimators for unstable autoregressive moving–average time series with general autoregressive heteroscedastic errors, *Ann. Statist.* **26** (1998), 84–125.

[22] R.L. Lumsdaine, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, *Economet. Theory* **64** (1996), 575–596.

[23] J.R. Magnus and H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, New York, 1988.

[24] S.P. Meyn and R.L. Tweedie, Markov chains and stochastic stability, Springer-Verlag , London, 1993.

[25] A. Mokkadem, Propiétés de mélange des processus autorégressifs polynomiaux, *Ann. Inst. Henri Poincaré* , Probabilités et Statistiques **26** (1990), 219-260.

[26] D.B. Nelson, Stationarity and persistence in the GARCH(1,1) model, *Economet. Theory* **6** (1990), 318–334.

[27] G.C. Reinsel, Elements of Multivariate time Series Analysis, Springer-Verlag, 1993.

[28] R.J. Serfling, Contributions to central limit theory for dependent variables, *Ann. Math. Stat.* **39** (1968), 4, 1158–1175.

[29] R. Tuncer, Convergence in probability of the maximum likelihood estimators of a multivariate ARMA model with GARCH(1,1) errors, Preprint CREST 9441, Paris, 1994.

[30] A.A. Weiss, Asymptotic theory for ARCH models: estimation and testing, *Economet. Theory* **2** (1986), 107–131.