The Annals of Statistics
2009, Vol. 37, No. 5B, 2730–2759
DOI: 10.1214/08-AOS674

ASYMPTOTIC NORMALITY OF THE QUASI-MAXIMUM LIKELIHOOD ESTIMATOR FOR MULTIDIMENSIONAL CAUSAL PROCESSES

BY JEAN-MARC BARDET AND OLIVIER WINTENBERGER

University Paris 1

Strong consistency and asymptotic normality of the quasi-maximum likelihood estimator are given for a general class of multidimensional causal processes. For particular cases already studied in the literature [for instance univariate or multivariate ARCH(∞) processes], the assumptions required for establishing these results are often weaker than existing conditions. The QMLE asymptotic behavior is also given for numerous new examples of univariate or multivariate processes (for instance TARCH or NLARCH processes).

1. Introduction. Since the seminal paper of Engle [11], autoregressive conditional heteroscedasticity (ARCH) models have been favored by econometricians for modeling financial series in discrete time. There are several reasons to explain the success of these models. One of them is certainly that they may be applied to series of data with fat tails. For statistical inference in conditionally Gaussian ARCH models, maximum likelihood estimators (MLE) are rather simple to calculate and have nice asymptotic properties. But it is well known that the normality of the innovations is rejected in most applications dealing with fat tail data. However, Gaussian MLE remain the most simple estimators and may keep their nice asymptotic properties even in nonconditionally Gaussian cases. Estimators are then called (Gaussian) quasi-MLE, QMLE for short.

We give in this paper, for the first time, asymptotic properties, namely strong consistency and asymptotic normality (respectively, SC and AN for short), of the QMLE for many multivariate models. To establish results in a unified way, we consider almost everywhere (a.e.) solutions $X = (X_t, t \in \mathbb{Z})$ of equations of the type

$$X_t = M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \cdot \xi_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \qquad \forall t \in \mathbb{Z}. \tag{1.1}$$

Here, $\theta_0 \in \Theta \subset \mathbb{R}^d$ is the parameter of interest, $M_{\theta_0}(X_{t-1}, X_{t-2}, \dots)$ is a $(m \times p)$-random matrix having a.e. full rank m, $f_{\theta_0}(X_{t-1}, X_{t-2}, \dots)$ is a $\mathbb{R}^m$-random vector, the $\mathbb{R}^p$-random vectors $\xi_t = (\xi_t^{(k)})_{1 \le k \le p}$ are independent and identically distributed satisfying standard assumptions $\mathbb{E}[\xi_0^{(k)} \xi_0^{(k')}] = 0$ for $k \neq k'$ and $\mathbb{E}[{\xi_0^{(k)}}^2] =$

Received December 2007; revised November 2008.
AMS 2000 subject classifications. 62M10, 62F12.
Key words and phrases. Quasi-maximum likelihood estimator, strong consistency, asymptotic normality, multidimensional causal processes, multivariate ARMA–GARCH processes.

$\mathrm{Var}(\xi_0^{(k)}) = 1$. Many models admit such multidimensional autoregressive representation. ARCH, GARCH, VAR, multivariate AR–GARCH and TARCH all correspond to (1.1) associated with specific functions $\theta \mapsto f_\theta$ and $\theta \mapsto M_\theta$.

Only for defining properly the QMLE, we assume, in this paragraph, that the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is normally distributed. The conditional likelihood of X expresses as, up to an additional constant,

$$(1.2)\qquad \begin{aligned} &L_n(\theta) := -\frac{1}{2}\sum_{t=1}^{n} q_t(\theta) \qquad \text{for all } \theta \in \Theta \\ &\text{with } q_t(\theta) := [(X_t - f_\theta^t)'(H_\theta^t)^{-1}(X_t - f_\theta^t) + \log(\det(H_\theta^t))] \end{aligned}$$

and $f_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \ldots)$, $M_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \ldots)$ and $H_\theta^t := M_\theta^t M_\theta^{t\,\prime}$. The quasi-likelihood $\widehat{L}_n$ is obtained by plugging in L_n the approximations $\widehat{f_\theta^t} := f_\theta(X_{t-1}, \ldots, X_1, u)$, $\widehat{M_\theta^t} := M_\theta(X_{t-1}, \ldots, X_1, u)$ and $\widehat{H}_\theta^t := \widehat{M_\theta^t} \cdot (\widehat{M_\theta^t})'$, where $u = (u_n)_{n\in\mathbb{N}}$ is a finitely nonzero sequence[1] $(u_n)_{n\in\mathbb{N}}$,

$$(1.3)\qquad \begin{aligned} &\widehat{L}_n(\theta) := -\frac{1}{2}\sum_{t=1}^{n} \widehat{q}_t(\theta) \\ &\text{with } \widehat{q}_t(\theta) := [(X_t - \widehat{f_\theta^t})'(\widehat{H}_\theta^t)^{-1}(X_t - \widehat{f_\theta^t}) + \log(\det(\widehat{H}_\theta^t))]. \end{aligned}$$

The QMLE $\widehat{\theta}_n$ is the maximizer of the quasi-likelihood $\widehat{L}_n$; that is,

$$(1.4)\qquad \widehat{\theta}_n := \operatorname*{Arg\,max}_{\theta\in\Theta} \widehat{L}_n(\theta).$$

Remark that unobserved values $(X_t, t \le 0)$ have to be fixed a priori equal to $(u_n)_{n\in\mathbb{N}}$ in the quasi-likelihood $\widehat{L}_n$. In Section 3.1, we give sufficient conditions on the parameters set Θ in order to ensure that the choice of $(u_n)_{n\in\mathbb{N}}$ does not have any consequences on the asymptotic behavior of $\widehat{L}_n$. Even if $(u_n)_{n\in\mathbb{N}}$ may have some consequences in practical applications (see [14]), it plays no role in the asymptotic results of this paper, and we set $u_n = 0$ $\forall n \in \mathbb{N}$ for convenience.

From now on, we omit any Gaussian assumption on the distribution of ξ_t. Under some Lipschitz-type assumptions on M and f, we define parameters sets $\Theta(r) \subset \Theta$ if $\mathbb{E}(\|\xi_t\|^r) < \infty$ for $r = 2$ and $r = 4$. Thanks to a result of Doukhan and Wintenberger [9], we derive the existence of a solution X of the very general model (1.1) as $\theta_0 \in \Theta(r)$ (see Section 2). This solution X admits necessarily finite moments of corresponding orders [i.e., $\mathbb{E}(\|X_t\|^r) < \infty$]. Using these moments properties, we prove SC and AN of QMLE if, respectively, $\theta_0 \in \Theta(2)$ and $\theta_0 \in \overset{\circ}{\Theta} \cap \Theta(4)$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ and other assumptions (see Section 3 for more details). Then, SC and AN of QMLE are only given here if X has finite moments of order 2 and 4, which is less sharp than classical results

[1] This means that $u_n \neq 0$ only for finitely many $n \in \mathbb{N}$.

on univariate GARCH, ARMA–GARCH and univariate conditional heteroscedastic models (see Berkes, Horváth and Kokoszka [1], Francq and Zakoïan [14] and Straumann and Mikosch [25], resp.). But, compared with other existing results for multivariate models (see Jeantheau [17] and Boussama [4] for SC, Comte and Liberman [5] and Ling and McAleer [20] for AN), our assumptions are sharper and much simpler. Moreover, even for univariate models such that ARCH(∞) model, our results are competitive with those of Robinson and Zaffaroni [24].

Thus, in the sequel, we focus our presentation in models satisfying (1.1), except univariate conditional heteroscedastic models defined as in [25]. In this framework, most of existing results recalled in the previous paragraph are given under a condition of existence and finite moments on the process X without specifying assumptions on Θ and ξ that are related to this condition. The links between the conditions first on X and second on Θ and ξ are intricate in this framework and the classical approach of stochastic recurrence equation, introduced by Bougerol [3] and applied in Straumann and Mikosch [25] does not work. Moreover, another classical method using Markov chain representations and contraction assumptions, as in Duflo [10], does not work, too (see Boussama [4], page 131). For such models, the approach of Doukhan and Wintenberger [9] that we use here provides a nice and simple alternative.

We express, in Section 4, our conditions for classical models. For univariate ARCH(∞) processes, our conditions are different than those in Robinson and Zaffaroni [24] concerning SC and AN but may be more interesting in specific cases. For SC in multivariate GARCH models, we obtain similar conditions as in Jeantheau [17] and Boussama [4]. For AN multivariate ARCH(∞) and ARMA–GARCH models, our conditions are sharper than those in Comte and Lieberman [5] and Ling and McAleer [20], who derived the asymptotic normality for more specific models under moments of order 4, 6 or 8 on X. Reducing the order of finite moments of the processes X to 4 is consistent with financial data that usually exhibit fat tailed marginals. We also provide, for the first time, the SC and AN of the QMLE in TARCH, GLARCH and some multidimensional SV (stochastic volatility) models.

But, to begin with, Section 2 presents assumptions on the model (1.1).

2. Notation and assumptions.

Some standard notation is used:

- The symbol 0 denotes any null vector of any vector space;
- For $u \in \mathbb{R}^p$, $u = (u^{(i)})_{1 \le i \le p}$, and for A a $m \times p$-matrix, $A = (A_{ij})_{ij}$;
- For $u \in \mathbb{R}^p$, $\mathrm{Diag}(u)$ is the diagonal $p \times p$ matrix with $(\mathrm{Diag}(u))_{ij} = \delta_{ij} u_i$;
- If V is a vector space then $V^\infty = \{(x_n)_{n \in \mathbb{N}} \in V^{\mathbb{N}}, \exists N \in \mathbb{N}, x_k = 0 \text{ for all } k > N\}$;
- The symbol $\|\cdot\|$ denotes the usual Euclidean norm of a vector or a matrix (for A a $m \times p$-matrix, $\|A\| = \sup_{\|Y\| \le 1}\{\|AY\|, Y \in \mathbb{R}^p\}$);
- For the measurable vector- or matrix-valued function g defined on some set U, $\|g\|_U = \sup_{\theta \in U} \|g(\theta)\|$;

- From now on, Θ denotes a subset of $\mathbb{R}^d$, and if V is a Banach space, then $\mathcal{C}(\Theta, V)$ denotes the Banach space of V-valued continuous functions on Θ equipped with the uniform norm $\|\cdot\|_\Theta$ and $\mathbb{L}^r(\mathcal{C}(\Theta, V))$ $(r \geq 1)$ denotes the Banach space of random a.e. continuous functions f such that $\mathbb{E}[\|f\|_\Theta^r] < \infty$;
- For $\theta \in \Theta$, if $\Psi_\theta : (\mathbb{R}^m)^\infty \to V$ is a Borelian function on V a finite-dimensional vector space, then $\partial_\theta^k \Psi_\theta(x)$ denotes, respectively, for $k = 0, 1, 2$, when there exists $\Psi_\theta(x)$, $\frac{\partial \Psi_\theta(x)}{\partial\theta}$ and $\frac{\partial^2 \Psi_\theta(x)}{\partial\theta^2}$ $[x \in (\mathbb{R}^m)]$;
- If $h : \mathbb{R}^m \to V$ is a Borelian function on a vector space V equipped with the norm $\|\cdot\|$, then h is a $\operatorname{Lip} h$-Lipschitzian function if

$$\operatorname{Lip} h := \sup_{x, y \in \mathbb{R}^m, x \neq y} \frac{\|h(x) - h(y)\|}{\|x - y\|} < \infty.$$

2.1. *Solutions of* (1.1). In Proposition 1 below, we prove the existence of a stationary solution of order r to the general model (1.1) under some restrictions on the parameter θ_0. To settle these assumptions in a short way, let us introduce the generic symbol Ψ for any of the functions f, M or H and, for $k = 0, 1, 2$ and some compact subset Θ of $\mathbb{R}^d$, define

($\mathrm{A}_k(\Psi, \Theta)$) The function $\partial_\theta^k \Psi_\theta$ satisfies $\|\partial_\theta^k \Psi_\theta(0)\|_\Theta < \infty$, and there exists a sequence $(\alpha_j^{(k)}(\Psi, \Theta))_j$ of nonnegative numbers such that $\forall x, y \in (\mathbb{R}^m)^\infty$

$$\|\partial_\theta^k \Psi_\theta(x) - \partial_\theta^k \Psi_\theta(y)\|_\Theta \leq \sum_{j=1}^\infty \alpha_j^{(k)}(\Psi, \Theta) \|x_j - y_j\|$$

$$\text{with } \sum_{j=1}^\infty \alpha_j^{(k)}(\Psi, \Theta) < \infty,$$

if $\Psi = H$, x, y, x_j, y_j are, respectively, replaced by xx', yy', $x_j x_j'$ and $y_j y_j'$.

These Lipschitz-type inequalities on f and M are essential for establishing Proposition 1 and SC and AN of QMLE. Let us now provide some examples of models satisfying (1.1) and assumptions ($\mathrm{A}_0(\Psi, \Theta)$).

EXAMPLE 1. (i) In [9] multidimensional GLARCH(∞) [generalized linear ARCH(∞)] models were defined, such that

$$X_t = \operatorname{Diag}\left(B_0(\theta_0) + \sum_{k=1}^\infty B_k(\theta_0, X_{t-k})\right)\xi_t, \tag{2.1}$$

where, for $j \in \mathbb{N}$, $B_j(\theta, \cdot) : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^p$ are Lipschitzian functions. Then, ($\mathrm{A}_0(M, \Theta)$) is satisfied with $\alpha_j^{(0)}(M, \Theta) = \sup_{\theta \in \Theta} \operatorname{Lip}_x(B_j(\theta, x))$ when

$$\sum_{j=1}^\infty \sup_{\theta \in \Theta} \operatorname{Lip}_x(B_j(\theta, x)) < \infty. \tag{2.2}$$

(ii) Nonlinear ARCH(∞) [NLARCH(∞)] models refer to (1.1) satisfying

$$f_\theta \equiv 0 \quad \text{and} \quad H_\theta := M_\theta M_\theta' \text{ is a function of } xx' = (x_j x_j')_{j\geq 0}. \tag{2.3}$$

For NLARCH(∞) models, it is clear that assumption ($\mathrm{A}_0(H,\Theta)$) will be more straightforward than ($\mathrm{A}_0(M,\Theta)$). In particular, when H_θ is a Lipschitzian function of xx', then ($\mathrm{A}_0(H,\{\theta\})$) holds. The multivariate ARCH(∞) models are specific cases of NLARCH(∞) models defined by

$$X_t = \left(B_0(\theta_0) + \sum_{j=1}^{\infty} B_j(\theta_0) X_{t-j} X_{t-j}' B_j'(\theta_0)\right)^{1/2} \xi_t, \tag{2.4}$$

where $m = p$ and $A^{1/2}$ denotes the symmetric matrix such that $(A^{1/2})^2 = A$ for some symmetric positive matrix A. Here, $B_0(\theta)$ is assumed to be a symmetric matrix. Multivariate ARCH(∞) processes are processes with stochastic volatility that generalize multivariate GARCH(p,q) processes (see, e.g., [17] or [5]). Here, ($\mathrm{A}_0(H,\Theta)$) is satisfied with $\alpha_j^{(0)}(H,\Theta) = \|B_j(\theta)\|_\Theta^2$ when

$$\sum_{j=1}^{\infty} \|B_j(\theta)\|_\Theta^2 < \infty. \tag{2.5}$$

(iii) Third main examples are multidimensional extensions of complete models with stochastic volatility introduced in [18] as approximations of complete models with continuous time. The increments of the log of the price processes are, here, solutions of the recursive equation

$$X_t = \sigma(S_{t-1})\xi_t + \mu(S_{t-1}), \tag{2.6}$$

where σ and μ are, respectively, $m \times p$-matrices and $\mathbb{R}^p$-vector valued Lipschitz functions defined on $\mathbb{R}^p$ and the so-called offset functions S_t satisfy the equation

$$S_t = \big(I_p - B(\theta_0)\big) \sum_{i=1}^{\infty} B(\theta_0)^{i-1}(X_t + \cdots + X_{t-i+1}).$$

Here, $B(\theta)$ is a $p \times p$ matrix, and then ($\mathrm{A}_0(f,\Theta)$) and ($\mathrm{A}_0(M,\Theta)$) are satisfied as soon as $\|B(\theta)\| < 1$.

For ensuring a stationary r-order solution of (1.1), for $r \geq 1$, define the set

$$\Theta(r) := \Bigg\{\theta \in \mathbb{R}^d, (\mathrm{A}_0(f,\{\theta\})) \text{ and } (\mathrm{A}_0(M,\{\theta\})) \text{ hold},$$
$$\sum_{j=1}^{\infty} \alpha_j^{(0)}(f,\{\theta\}) + (\mathbb{E}\|\xi_0\|^r)^{1/r} \sum_{j=1}^{\infty} \alpha_j^{(0)}(M,\{\theta\}) < 1\Bigg\}.$$

It is clear that $\theta \in \Theta(r)$ only if $\mathbb{E}\|\xi_0\|^r < \infty$. Then, using results of Doukhan and Wintenberger [9], one obtains (all the proofs are given in Section 5) the following.

PROPOSITION 1. *If $\theta_0 \in \Theta(r)$ for some $r \geq 1$, then there exists a unique causal* [X_t *is independent of* $(\xi_i)_{i>t}$ *for* $t \in \mathbb{Z}$] *solution X to* (1.1), *which is stationary, ergodic and satisfies* $\mathbb{E}\|X_0\|^r < \infty$.

This result generalizes the one proved by Giraitis, Kokoszka and Leipus [15] for univariate ARCH(∞) models. It automatically yields weak dependence properties (see [9] for details). However, for specific processes, conditions on M can be advantageously replaced by conditions on H and the following.

COROLLARY 1. *For univariate* NLARCH(∞) *models satisfying* (2.3) *with* $m = p = 1$ *and* $f \equiv 0$, *the result of Proposition* 1 *holds, if* $\theta_0 \in \widetilde{\Theta}(r)$ *for* $r \geq 2$, *where*

$$\widetilde{\Theta}(r) := \left\{\theta \in \mathbb{R}^d, (A_0(H, \{\theta\})) \text{ holds}, (\mathbb{E}\|\xi_0\|^r)\left(\sum_{j=1}^{\infty} \alpha_j^{(0)}(H, \{\theta\})\right)^{r/2} < 1\right\}.$$

REMARK 1. Proposition 1 and Corollary 1 link the r-moment of innovations (ξ_t) to an r-moment of X_0. However, it is known that the consistency of QMLE can be obtained for $r = 2$, but requiring only that $\mathbb{E}(|\log(\det H_{\theta_0}^t)|) < \infty$ (see [21]). Then, we cannot consider FIGARCH models here, as they never have finite variance (see [24], page 1061). However, as far as we know, all existing proofs of consistency of QMLE in a multidimensional context imply a finite moment of order 2 for X (see [4, 5, 17] and [20]). This is due to the difficulty of proving the existence of a solution and its moments properties.

REMARK 2. The main example of process satisfying Corollary 1 is univariate ARCH(∞) models, the univariate version of (2.4), defined by Robinson [23] as the solution of the equation

$$X_t = \sigma_t \xi_t, \qquad \sigma_t^2 = b_0(\theta_0) + \sum_{j=1}^{\infty} b_j(\theta_0) X_{t-j}^2, \tag{2.7}$$

where, for all $\theta \in \mathbb{R}^d$, $(b_j(\theta))_{j\geq 1}$ are sequences of nonnegative real numbers. Here, $\alpha_j^{(0)}(M, \Theta) = \sqrt{\sup_{\theta\in\Theta} b_j(\theta)}$ and $\alpha_j^{(0)}(H, \Theta) = \sup_{\theta\in\Theta} b_j(\theta)$. Then, $(A_0(H, \{\theta_0\}))$ holds when $\sum_{j=1}^{\infty} b_j(\theta_0) < \infty$, and $\theta_0 \in \widetilde{\Theta}(r)$ when

$$(\mathbb{E}\|\xi_0\|^r)\left(\sum_{j=1}^{\infty} b_j(\theta_0)\right)^{r/2} < 1. \tag{2.8}$$

Working with $\widetilde{\Theta}(r)$ larger than $\Theta(r)$ gives weaker conditions in this context.

2.2. *Additional assumptions required for the convergence of QMLE.* Fix some compact subset Θ of $\mathbb{R}^d$. A first usual assumption for using QMLE is the following:

$$(\mathrm{D}(\Theta)) \qquad \inf_{\theta\in\Theta} \det(H_\theta(x)) \geq \underline{H} \qquad \text{for all } x \in (\mathbb{R}^m)^\infty \text{ and some } \underline{H} > 0.$$

The same primitive identifiability condition as in Jeantheau [17] will be required:

$$(\mathrm{Id}(\Theta)) \qquad \begin{aligned} &\forall\theta \in \Theta \text{ and some } t \in \mathbb{Z}, \\ &(f_\theta^t = f_{\theta_0}^t \text{ and } H_\theta^t = H_{\theta_0}^t \text{ a.s.}) \quad \Rightarrow \quad \theta = \theta_0. \end{aligned}$$

The following condition (Var) is needed for ensuring existence of the asymptotic variance in AN:

(Var) One of the families $(\partial f_{\theta_0}^t/\partial\theta_i)_{1\leq i\leq d}$ or $(\partial H_{\theta_0}^t/\partial\theta_i)_{1\leq i\leq d}$ is a.e. linearly independent, where

$$\frac{\partial f_\theta^t}{\partial\theta} := \frac{\partial f_\theta}{\partial\theta}(X_{t-1},\ldots) \quad \text{and} \quad \frac{\partial H_\theta^t}{\partial\theta} := \frac{\partial H_\theta}{\partial\theta}(X_{t-1},\ldots).$$

Such conditions (Id(Θ)) and (Var) are primitive. We give more explicit conditions for univariate and multivariate AR(∞)–ARCH(∞) models (see Section 4 for more details). We do not achieve explicit conditions when introducing nonlinearity in f, M or H.

3. Asymptotic behavior of the QMLE. The results of this section are more general if one replaces $\Theta(r)$ with $\widetilde{\Theta}(r)$ in the assumptions when dealing with univariate NLARCH(∞) models (see Remark 2 for more details).

3.1. *Asymptotic properties of the quasi-likelihood.* The QMLE is based on an approximation of $f_\theta^t = \mathbb{E}(X_t|X_{t-1}, X_{t-2},\ldots)$ and $H_\theta^t = \mathbb{E}((X_t - f_\theta^t)(X_t - f_\theta^t)'|X_{t-1}, X_{t-2},\ldots)$ by $\widehat{f}_\theta^t$ and $\widehat{H}_\theta^t$, which is defined as in the Introduction. The following lemma estimates the error of approximations. It is a crucial step in the proof of the QMLE consistency.

LEMMA 1. *Assume that $\theta_0 \in \Theta(r)$ for $r \geq 2$ and X is the stationary solution of* (1.1). *Let Θ be a compact set of $\mathbb{R}^d$*:

1. *If* ($\mathrm{A}_0(f,\Theta)$) *holds, then* $\forall\theta \in \Theta$, $f_\theta^t \in \mathbb{L}^r(\mathcal{C}(\Theta, \mathbb{R}^m))$ *and*

$$\mathbb{E}[\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta^r] \leq \mathbb{E}[\|X_0\|^r]\left(\left\|\sum_{j>t}\alpha_j(f)\right\|_\Theta\right)^r \qquad \text{for all } t \in \mathbb{N}^*; \tag{3.1}$$

2. *If* ($\mathrm{A}_0(M,\Theta)$) *holds, then* $\forall\theta \in \Theta$, $H_\theta^t \in \mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$, *and there exists $C > 0$ not depending on t such that*

$$\mathbb{E}[\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta^{r/2}] \leq C\left(\left\|\sum_{j>t}\alpha_j^{(0)}(M,\Theta)\right\|_\Theta\right)^{r/2} \qquad \text{for all } t \in \mathbb{N}^*; \tag{3.2}$$

3. *If* $(\mathrm{A}_0(H,\Theta))$ *holds, then* $\forall\theta\in\Theta$, $H_\theta^t\in\mathbb{L}^{r/2}(\mathcal{C}(\Theta,\mathcal{M}_m))$ *and*

$$\mathbb{E}[\|\widehat{H}_\theta^t-H_\theta^t\|_\Theta^{r/2}]\le\mathbb{E}[\|X_0\|^r]\Big(\Big\|\sum_{j>t}\alpha_j^{(0)}(H,\Theta)\Big\|_\Theta\Big)^{r/2}\qquad\textit{for all } t\in\mathbb{N}^*. \tag{3.3}$$

Moreover, under any of the two last conditions and with $(\mathrm{D}(\Theta))$, H_θ^t *is an invertible matrix and* $\|(\widehat{H}_\theta^t)^{-1}\|_\Theta\le\underline{H}^{-1/m}$.

3.2. *Strong consistency.* In the following theorem, we assume by convention that, if $(\mathrm{A}_0(M,\Theta))$ holds, then $\alpha_j^{(0)}(H,\Theta)=0$, and, if $(\mathrm{A}_0(H,\Theta))$ holds, then $\alpha_j^{(0)}(M,\Theta)=0$.

THEOREM 1. *Assume that* $\theta_0\in\Theta(2)\cap\Theta$ *and let* X *be the stationary solution of* (1.1). *If* $\theta_0\in\Theta$, *a compact set of* $\mathbb{R}^d$ *such that assumptions* $(\mathrm{D}(\Theta))$, $(\mathrm{Id}(\Theta))$, $(\mathrm{A}_0(f,\Theta))$ *and* $(\mathrm{A}_0(M,\Theta))$ [*or* $(\mathrm{A}_0(H,\Theta))$] *hold with*

$$\alpha_j^{(0)}(f,\Theta)+\alpha_j^{(0)}(M,\Theta)+\alpha_j^{(0)}(H,\Theta)=O(j^{-\ell})\qquad\textit{for some } \ell>3/2, \tag{3.4}$$

then the QMLE $\widehat{\theta}_n$ *defined by* (1.4) *is SC*; *that is,* $\widehat{\theta}_n\xrightarrow[n\to\infty]{\text{a.s.}}\theta_0$.

3.3. *Asymptotic normality.* We use the following convention: if $(\mathrm{A}_1(M,\Theta))$ holds, then $\alpha_j^{(1)}(H,\Theta)=0$, and, if $(\mathrm{A}_1(H,\Theta))$ holds, then $\alpha_j^{(1)}(M,\Theta)=0$.

THEOREM 2. *Assume that* $\theta_0\in\Theta(4)\cap\overset{\circ}{\Theta}$, *the interior of* Θ, *and let* X *be the stationary solution of* (1.1). *Under the assumptions of Theorem* 1 *and* (Var), *if, for* $i=1,2$, $(\mathrm{A}_i(f,\Theta))$ *and* $(\mathrm{A}_i(M,\Theta))$ [*or* $(\mathrm{A}_i(H,\Theta))$] *hold, with*

$$\alpha_j^{(1)}(f,\Theta)+\alpha_j^{(1)}(M,\Theta)+\alpha_j^{(1)}(H,\Theta)=O(j^{-\ell'})\qquad\textit{for some } \ell'>3/2, \tag{3.5}$$

then the QMLE $\widehat{\theta}_n$ *is SC and AN*; *that is,*

$$\sqrt{n}(\widehat{\theta}_n-\theta_0)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}_d(0,F(\theta_0)^{-1}G(\theta_0)F(\theta_0)^{-1}), \tag{3.6}$$

where $F(\theta_0)$ *and* $G(\theta_0)$ *are defined in* (5.10) *and* (5.14), *respectively.*

4. Examples. In this section, the previous asymptotic results are applied to several examples. As assumptions $(\mathrm{Id}(\Theta))$ and (Var) are primitive, we give more explicit sufficient conditions for ARCH(∞) models. Then, we will assume the following:

(Id$'$) ξ_0 is such that no quadratic form $Q\neq 0$, satisfies $Q(\xi_0)=\delta$ p.s. for some $\delta\in\mathbb{R}$.

This condition is the same as in [16], but we use it in the more general framework of multivariate ARCH(∞) models. In the univariate case, it is equivalent to the fact that the support of ξ_0 is not reduced to $\{-1, 1\}$. Whatever the dimensions are, if the support of ξ_0 has a nonempty interior, then (Id′) is automatically satisfied. If ξ_0 admits a density, identifiability holds under weaker conditions (see [12]). Things get easier then, as X_0 admits, also, a density (see Proposition 5.1 of [9]).

4.1. *Univariate* ARCH(∞) *processes*. We use the definition and results of Remark 2. For $\theta_0 \in \widetilde{\Theta}(r)$, the existence of a stationary solution and of its rth order moments is already settled in Giraitis, Kokoszka and Leipus [15]. Here, we formulate a version of Theorems 1 and 2 adapted to this context.

PROPOSITION 2. *Let $\theta_0 \in \widetilde{\Theta}(2)$, let X be the stationary solution of* (2.7) *and let* (Id′) *hold. Assume that $\theta_0 \in \Theta$, a compact set of $\mathbb{R}^d$ such that $\forall j \in \mathbb{N}, \theta \in \Theta \mapsto b_j(\theta)$, is a nonnegative continuous injective function such that*

$$\inf_{\theta\in\Theta} b_0(\theta) > 0 \quad \textit{and} \quad \sup_{\theta\in\Theta} b_j(\theta) = O(j^{-\ell}) \qquad \textit{for some } \ell > 3/2: \tag{4.1}$$

SC. *Then, the QMLE $\widehat{\theta}_n$ is SC.*

AN. *Moreover, if $\theta_0 \in \widetilde{\Theta}(4) \cap \overset{\circ}{\Theta}$ and $\theta \in \Theta \mapsto b_j(\theta)$ is 2 times continuously differentiable for any $j \in \mathbb{N}$ with for $(k, k') \in \{1, \ldots, d\}^2$,*

$$\begin{aligned} &\sup_{\theta\in\Theta}\left|\frac{\partial b_j(\theta)}{\partial\theta_k}\right| = O(j^{-\ell'}) \qquad \textit{for some } \ell' > 3/2 \quad \textit{and} \\ &\sum_{j\geq 1}\sup_{\theta\in\Theta}\left|\frac{\partial^2 b_j(\theta)}{\partial\theta_k\,\partial\theta_{k'}}\right| < \infty, \end{aligned} \tag{4.2}$$

and, if there exists some injective function $k \in \mathbb{N} \mapsto j_k \in \mathbb{N}$ such that

$$\left(\frac{\partial b_{j_k}(\theta_0)}{\partial\theta_k}\right)_{1\leq k\leq d} \textit{ is linearly independent,} \tag{4.3}$$

then the QMLE $\widehat{\theta}_n$ is also AN.

The proof that assumptions (Id(Θ)) and (Var) hold under (Id′) and (4.3) is given in Section 5.5. A more explicit condition in Robinson and Zaffaroni [24], assumption A(r), F(1) and G, page 1053, ensures that both (Id(Θ)) and (Var). We do not consider this condition here, as it is more restrictive and useless in multivariate cases.

Let us compare the results of Proposition 2 with those of Theorems 1 and 2 in Robinson and Zaffaroni [24]. Those authors obtained SC of the QMLE under moments of order $r > 2$ (instead of $r = 2$ here) but only with $\mathbb{E}X_t^{2\rho} < \infty$, with $\rho < 1$ (instead of $\mathbb{E}X_t^2 < \infty$ here) and a decreasing rate $j^{-\ell}$ with $\ell > 1$ (instead of $\ell > 3/2$ here) for the sequence $(\sup_{\theta\in\Theta}|b_j(\theta)|)_{j\geq 1}$. It implies that a more general space of parameters Θ and that FIGARCH processes can be considered in [24]. However, note that no relation such as $\theta_0 \in \widetilde{\Theta}(2)$ [or $\widetilde{\Theta}(4)$] is specified in [24] with explicit sets $\widetilde{\Theta}(2)$ [or $\widetilde{\Theta}(4)$]. Such relation may derive from the sufficient existence conditions given in [7], but the link between these two complementary works has not been examined, as far as we know. Concerning identifiability condition, assumptions on the derivatives of $\theta \to b_j(\theta)$ are required in assumption F(1) and G of [24] as well as the existence of a probability density function in assumption A(r). There are no such conditions in Proposition 2, and assumption (Id′) seems to be more general and easier used than assumption A(r), F(1) and G of [24]. Moreover, concerning AN, their assumption F(l), which is obtained from a comparison of the derivatives of $\theta \to b_j(\theta)$ and $b_j(\theta)$, seems not unnatural; for instance, if $b_j(\theta) = C(\cos(\alpha j) - \alpha j)j^{-5}$ with $\theta = (C, \alpha)$ and $\alpha \le -1$, then our Proposition 2 shows AN, while assumption F(2) of [24] is not satisfied [in such a case, $|\frac{\partial^2 b_j}{\partial \alpha^2}(\theta)| = Cj^{-3}|\cos(\alpha j)|$ is not always smaller than $(b_j(\theta))^{1-\eta}$ for all $\eta > 0$ and $j \in \mathbb{N}$]. Finally, [24] requires conditions on the third derivatives, while no such assumption is supposed in our Proposition 2.

As a conclusion, it is not possible to compare Proposition 2 and Theorem 1 and 2 of [24], because the required conditions are different. But Proposition 2 is certainly simpler and more straightforward to use. However, the complexity of the conditions in [24] are certainly due to the efforts of these authors to deal with FIGARCH processes that we cannot consider here as all processes with infinite variance.

4.2. TARCH(∞) *models*. The process X is called threshold ARCH(∞) if it satisfies the equations

$$X_t = \sigma_t \xi_t,$$
$$\sigma_t = b_0(\theta_0) + \sum_{j=1}^{\infty} [b_j^+(\theta_0) \max(X_{t-j}, 0) - b_j^-(\theta_0) \min(X_{t-j}, 0)], \tag{4.4}$$

where the parameters $b_0(\theta)$, $b_j^+(\theta)$ and $b_j^-(\theta)$ are assumed to be nonnegative real numbers. This class of processes is a generalization of the class of TGARCH(p, q) processes (introduced by Rabemananjara and Zakoïan [22]) and AGARCH(p, q) processes (introduced by Ding, Granger and Engle [6]). Here,

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^d \,\Big|\, \sum_{j=1}^{\infty} \max(b_j^-(\theta), b_j^+(\theta)) \le (\mathbb{E}[|\xi_0|^r])^{-1/r} \right\},$$

since $\alpha_j^{(0)}(M, \{\theta\}) = \max(b_j^-(\theta), b_j^+(\theta))$. Consequently, we can settle, for the first time, the SC and AN of the QMLE for TARCH(∞) models as follows.

PROPOSITION 3. *Let $\theta_0 \in \Theta(2)$, let X be the stationary solution to* (4.4), *and assume that $\theta_0 \in \Theta$, a compact set of $\mathbb{R}^d$ such that* (Id(Θ)), *holds. Moreover, assume that* $\inf_{\theta\in\Theta} b_0(\theta) > 0$ *and*

$$\sup_{\theta\in\Theta} \max(b_j^-(\theta), b_j^+(\theta)) = O(j^{-\ell}) \qquad \textit{for some } \ell > 3/2:$$

SC. *Then, the QMLE is SC.*

AN. *Moreover, if $\theta_0 \in \overset{\circ}{\Theta} \cap \Theta(4)$, assume that $\theta \mapsto b_0(\theta)$, $\theta \mapsto b_j^+(\theta)$ and $\theta \mapsto b_j^-(\theta)$ are 2 times continuously differentiable on Θ, satisfying*

$$\sup_{\theta\in\Theta} \max\left(\left|\frac{\partial b_j^+(\theta)}{\partial\theta_k}\right|, \left|\frac{\partial b_j^+(\theta)}{\partial\theta_k}\right|\right) = O(j^{-\ell'}) \qquad \textit{for some } \ell' > 3/2$$

and

$$\sum_{j\ge 1} \sup_{\theta\in\Theta} \max\left(\left|\frac{\partial^2 b_j^+(\theta)}{\partial\theta_k\,\partial\theta_{k'}}\right|, \left|\frac{\partial^2 b_j^-(\theta)}{\partial\theta_k\,\partial\theta_{k'}}\right|\right) < \infty \qquad \textit{for all } (k,k') \in \{1,\dots,d\}^2.$$

If (Var) *holds, then the QMLE $\widehat{\theta}_n$ is also AN.*

As far as we know, more explicit conditions for (Id(Θ)) and (Var) in such nonlinear context do not exist. One possible way to obtain such conditions could be to work under (Id$'$), where the quadratic forms Q are replaced with functions $x \mapsto b^+ \max(x,0) - b^- \min(x,0)$ for some $b^+, b^- \in \mathbb{R}^+$.

4.3. *Multivariate* ARCH(∞) *processes.* Multivariate ARCH(∞) processes are already considered in Example 1(ii), and, since $\alpha_j^{(0)}(M,\{\theta\}) = \|B_j(\theta)\|$,

$$\Theta(r) = \left\{\theta \in \mathbb{R}^d \,\middle|\, \sum_{j=1}^{\infty} \|B_j(\theta)\| < (\mathbb{E}[\|\xi_0\|^r])^{-1/r}\right\}. \tag{4.5}$$

PROPOSITION 4. *Let $\theta_0 \in \Theta(2)$, X be the stationary solution of* (2.4) *and* (Id$'$) *hold. Assume that $B_0(\theta)$ is some $p \times p$ symmetric definite positive matrix, and $B_j(\theta)$ are, for all $j \ge 1$, some $p \times p$ null-matrices or matrices such that the symmetric matrix $(B_j(\theta) + B_j(\theta)')/2$ is positive definite. If $\theta_0 \in \Theta$, a compact set of $\mathbb{R}^d$ such that* $\inf_{\theta\in\Theta} \det B_0(\theta) > 0$ *and $\theta \in \Theta \mapsto B_j(\theta)$ is an injective function, for all $j \in \mathbb{N}$:*

SC. *Then, the QMLE is SC.*

AN. *Moreover, assume that $\theta_0 \in \Theta(4) \cap \overset{\circ}{\Theta}$ and $\forall j \in \mathbb{N}$, $\theta \in \Theta \mapsto B_j(\theta)$ are 2 times continuously differentiable, satisfying, for all $(k,k') \in \{1,\dots,d\}^2$,*

$$\left\|\frac{\partial B_j(\theta)}{\partial\theta_k}\right\|_{\Theta} = O(j^{-\ell'}) \qquad \textit{for some } \ell' > 3/2$$

and

$$\sum_{j\geq 1}\left\|\frac{\partial^2 B_j(\theta)}{\partial\theta_k\,\partial\theta_{k'}}\right\|_\Theta<\infty.$$

If there exists some subset $\mathcal{S}\subset\mathbb{N}$ *such that*

$$\left(\frac{\partial B_j(\theta_0)}{\partial\theta_k}\right)_{1\leq k\leq d,\, j\in\mathcal{S}} \text{ is linearly independent,} \tag{4.6}$$

then the QMLE $\widehat{\theta}_n$ *is also AN.*

The proof that assumptions (Id(Θ)) and (Var) hold under such conditions is given in Section 5.5. To the best of our knowledge, the asymptotic behavior of the QMLE for such models is studied here for the first time. It generalizes the work of Jeantheau [17] and Comte and Lieberman [5] on, respectively, VEC and BEKK multivariate GARCH(q,q') models, which both admit an ARCH(∞) representation. Conditions for SC are similar to those in [17] and in [5]. Existing results for AN in [5] were obtained under stronger assumptions than here.

4.4. *Multivariate* GLARCH(∞) *models.* Such models have been already considered in Example 1(i). Note that they are generalizations of multivariate LARCH(∞) models introduced by Doukhan, Teyssière and Winant [8]. For instance, consider the multidimensional extension of the TARCH models as

$$(B_j(\theta,x))_k=\sum_{i=1}^m B^+_{jki}(\theta)\max(x_{ji},0)+B^-_{jki}(\theta)\min(x_{ji},0),$$

where B^+_{jki} and B^-_{jki} are nonnegative real numbers. For GLARCH(∞) models,

$$\Theta(r)=\left\{\theta\in\mathbb{R}^d\;\Big|\;\sum_{j=1}^\infty \mathrm{Lip}_x(B_j(\theta,x))<(\mathbb{E}[\|\xi_0\|^r])^{-1/r}\right\}.$$

PROPOSITION 5. *Let* $\theta_0\in\Theta(2)$, X *be the stationary solution of* (2.1). *If* $\theta_0\in\Theta$, *a compact set of* $\mathbb{R}^d$ *such that* (Id(Θ)) *holds*, $\inf_{\theta\in\Theta}\|B_0(\theta)\|>0$ *and, for all* $j\in\mathbb{N}$, $\theta\in\Theta$, $B_j(\theta,\cdot)\in[0,\infty[^p$ *and*

$$\|\mathrm{Lip}_x(B_j(\theta,x))\|_\Theta=O(j^{-\ell})\qquad \text{for some } \ell>3/2:$$

SC. *Then, the QMLE is SC.*

AN. *Moreover, if* $\theta_0\in\overset{\circ}{\Theta}\cap\Theta(4)$, (Var) *holds, and* $\forall j\in\mathbb{N}$, $\theta\in\Theta\mapsto B_j(\theta,\cdot)$ *is* 2 *times continuously differentiable and satisfies*

$$\left\|\mathrm{Lip}_x\frac{\partial B_j(\theta,x)}{\partial\theta_k}\right\|_\Theta=O(j^{-\ell'})\qquad \text{with } \ell'>3/2$$

and

$$\sum_{j\geq 1}\left\|\mathrm{Lip}_x \frac{\partial^2 B_j(\theta,x)}{\partial\theta_k\,\partial\theta_{k'}}\right\|_\Theta < \infty \qquad \textit{for all } (k,k')\in\{1,\dots,d\}^2,$$

then the QMLE $\widehat{\theta}_n$ is also AN.

4.5. *Multivariate ARMA–GARCH models.* Here, M_θ is concentrated on its diagonal and f is not necessarily identically zero. If $f\equiv 0$, the model coincides with the VEC–GARCH model (see Jeantheau [17]). Multidimensional ARMA–GARCH processes were introduced by Ling and McAleer [20] as the solution of the system of equations

$$\begin{cases} \Phi_\theta(L)\cdot X_t = \Psi_\theta(L)\cdot\varepsilon_t, \\ \varepsilon_t = M_\theta(X_{t-1},X_{t-2},\dots)\xi_t, \end{cases} \tag{4.7}$$

with $\mathrm{Diag}(H_\theta^t) = C_0(\theta)+\sum_{i=1}^q C_i(\theta)\,\mathrm{Diag}(\varepsilon_{t-i}\varepsilon'_{t-i})+\sum_{i=1}^{q'} D_i(\theta)\,\mathrm{Diag}(H_\theta^{t-i})$. Here, $C_0(\theta)$, $C_i(\theta)$ and $D_j(\theta)$ are positive definite matrices, $\mathrm{Diag}(A)$ is the diagonal of a matrix A, $\Phi_\theta(L) := I_m - \Phi_1 L - \cdots - \Phi_s L^s$ and $\Psi_\theta(L) := I_m - \Psi_1 L - \cdots - \Psi_{s'} L^{s'}$ are polynomials in the lag operator L, Φ_i and Ψ_j are squared matrices. We define, for all $\theta\in\mathbb{R}^d$,

$$\Gamma_\theta(L) := I_m + \sum_{i=1}^\infty \Gamma_i(\theta)L^i = \Psi_\theta^{-1}(L)\Phi_\theta(L)$$

and

$$\sum_{i=1}^\infty B_i(\theta)Z^i := \left(1-\sum_{i=1}^{q'} D_i(\theta)Z^i\right)^{-1} \times \sum_{i=1}^q C_i(\theta)Z^i \qquad \text{for all } Z\in\mathbb{C}^m,$$

where the polynomials of the two right-hand side products are assumed to be coprime. Equation (4.7) satisfies (1.1) with $f_\theta(X_{t-1},X_{t-2},\dots) = \sum_{i=1}^\infty \Gamma_i(\theta)X_{t-i}$. Then,

$$\Theta(r) = \left\{\theta\in\mathbb{R}^d \;\Big|\; \sum_{i=1}^\infty \|\Gamma_i(\theta)\| + (\mathbb{E}[\|\xi_0\|^r])^{1/r}\sum_{j=1}^\infty \|B_j(\theta)\| < 1\right\}.$$

If $\theta_0\in\Theta(r)$, then the existence of a solution is ensured. This existence condition is more explicit than the one of Theorem 2.1 of Ling and McAleer [20]. Now, we give a version of Theorems 1 and 2 when

$$\theta = (\Phi_1,\dots,\Phi_s,\Psi_1,\dots,\Psi_{s'},C_0,C_1,\dots,C_q,D_1,\dots,D_{q'}).$$

PROPOSITION 6. *Let $\theta_0\in\Theta(2)$ and X be the stationary solution of (4.7). If $\theta_0\in\Theta$, a compact set of $\mathbb{R}^d$ such that* $\inf_{\theta\in\Theta}\det(C_0(\theta))>0$, *such that the formulation for the multivariate GARCH part of X is minimal and such that* $\det(\Phi_\theta(z)\Psi_\theta(z))\neq 0$ *for all $\|z\|\leq 1$, $\Phi_\theta(z)$ and $\Psi_\theta(z)$ are coprime on Θ:*

SC. *Then*, *the QMLE is SC*.

AN. *Moreover*, *if* $\theta_0 \in \overset{\circ}{\Theta} \cap \widetilde{\Theta}(4)$, *and* (Var) *holds*, *then the QMLE* $\widehat{\theta}_n$ *is also AN*.

We refer the reader to Proposition 3.4 of [20], which proves that assuming the minimal representation of the GARCH part of the processes is enough for ensuring that assumption (Id(Θ)) holds. Remark that Proposition 6 improves the results of Ling and McAleer [20], which also provide (weak) consistency and asymptotic normality of the QMLE under finite moments of higher order for X. Notice, also, that the QMLE consistency of VEC–GARCH models was already established by Jeantheau [17] under similar conditions.

4.6. *Complete models with stochastic volatility.* For such process solution of (2.6), the set $\Theta(r)$ is defined easily for any $r \geq 2$ by the equation

$$\Theta(r) = \{\theta \in \mathbb{R}^d \mid \mathrm{Lip}(\mu) + (\mathbb{E}\|\xi_0\|^r)^{1/r}\,\mathrm{Lip}(\sigma) < 1 - \|B_\theta\|\}.$$

Now, we are able to give asymptotic properties of QMLE for complete models with stochastic volatility.

PROPOSITION 7. *Let* $\theta_0 \in \Theta(2)$, X *be the stationary solution of* (2.6), *with* σ *verifying that* $\inf_{x\in\mathbb{R}^m} \det(\sigma(x)\sigma(x)') \geq \underline{H}$ *for some* $\underline{H} > 0$. *If* $\theta_0 \in \Theta$, *a compact set of* $\mathbb{R}^d$ *such that* (Id(Θ)), *holds*:

SC. *Then*, *the QMLE is SC*.

AN. *Moreover*, *if* $\theta_0 \in \overset{\circ}{\Theta} \cap \Theta(4)$, (Var) *holds*, *and* $\forall j \in \mathbb{N}$, $\theta \in \Theta \mapsto B_j(\theta, \cdot)$ *is* 2 *times continuously differentiable and satisfies*

$$\left\| \frac{\partial B_j(\theta, x)}{\partial \theta_k} \right\|_\Theta < 1$$

and

$$\left\| \frac{\partial^2 B_j(\theta, x)}{\partial \theta_k \, \partial \theta_{k'}} \right\|_\Theta < 1 \qquad \text{for all } (k, k') \in \{1, \dots, d\}^2,$$

then the QMLE $\widehat{\theta}_n$ *is also AN*.

To our knowledge, this is the first result of this type for complete models with stochastic volatility. This example enlightens the fact that QMLE procedure also provides satisfying estimators for some models with stochastic volatility.

5. Proofs. In this section, the proofs of the main results are collected in the order of appearance in the paper. First, we prove Proposition 1 and Corollary 1, then Lemma 1, which settles the asymptotic properties of the quasi-likelihood. With the help of this property, we prove the main theorems that state consistency and asymptotic normality of the QMLE.

5.1. *Proofs of Proposition* 1 *and Corollary* 1. We apply a result of Doukhan and Wintenberger [9] that provides conditions for the existence of a stationary solution of an equation of type

$$X_t = F(X_{t-1}, X_{t-2}, \dots; \xi_t) \qquad \text{a.e. for all } t \in \mathbb{Z}. \tag{5.1}$$

If $\mathbb{E}\|\xi_0\|^r < \infty$ and F satisfies, for $x = (x_i)_{i\ge 1}$, $y = (y_i)_{i\ge 1} \in (\mathbb{R}^m)^\infty$:

- $\mathbb{E}\|F(0; \xi_0)\|^r < \infty$;
- $(\mathbb{E}\|F(x; \xi_0) - F(y; \xi_0)\|^r)^{1/r} \le \sum_{j\ge 1} a_j \|x_j - y_j\|$ with $\sum_{j\ge 1} a_j < 1$.

The existence of a unique causal stationary solution X of (5.1), such that $\mathbb{E}[\|X_0\|^r] < \infty$, is proved in [9]. We identify F from (1.1) as follows:

$$F(X_{t-1}, X_{t-2}, \dots; \xi_t) = M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \cdot \xi_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \dots).$$

Obviously, $\mathbb{E}[\|F(0; \xi_0)\|^r] < \infty$ if $\mathbb{E}\|\xi_0\|^r < \infty$, and we have

$$\begin{aligned}
&\big(\mathbb{E}\|F(x; \xi_0) - F(y; \xi_0)\|^r\big)^{1/r} \\
&\qquad \le \big(\mathbb{E}\big\|\big(M_{\theta_0}(x) - M_{\theta_0}(y)\big) \cdot \xi_0\big\|^r\big)^{1/r} + \|f_{\theta_0}(x) - f_{\theta_0}(y)\| \\
&\qquad \le (\mathbb{E}\|\xi_0\|^r)^{1/r} \|M_{\theta_0}(x) - M_{\theta_0}(y)\| + \|f_{\theta_0}(x) - f_{\theta_0}(y)\|.
\end{aligned}$$

The condition of Proposition 1 then implies those of [9] on F. For univariate NLARCH models satisfying (2.3), as H_θ is a function of $xx' = x^2$, we have

$$\begin{aligned}
\big(\mathbb{E}[|M_{\theta_0}^2(x)\xi_0^2 - M_{\theta_0}^2(y)\xi_0^2|^{r/2}]\big)^{2/r} &= (\mathbb{E}[|\xi_0|^r])^{2/r} |H_{\theta_0}(x^2) - H_{\theta_0}(y^2)| \\
&\le (\mathbb{E}[|\xi_0|^r])^{2/r} \sum_{j=1}^{\infty} \alpha_j(H, \theta_0) |x_j^2 - y_j^2|.
\end{aligned}$$

The result of [9] yields the existence, in $\mathbb{L}^{r/2}$, of $(X_t^2)_{t\in\mathbb{Z}}$, satisfying the equation

$$X_t^2 = M_{\theta_0}^2(X_{t-1}, X_{t-2}, \dots)\xi_t^2 = H_{\theta_0}(X_{t-1}^2, X_{t-2}^2, \dots)\xi_t^2 \qquad \text{a.e.}$$

Moreover, by [9], there exists a measurable function φ such that $X_t = \varphi(\xi_t, \xi_{t-1}, \dots)$ for all $t \in \mathbb{Z}$. The ergodicity of X follows from the Proposition 4.3 in Krengel [19]; it states that, if $(E, \mathcal{E})$ and $(\widetilde{E}, \widetilde{\mathcal{E}})$ are measurable spaces, $(v_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of E-valued random elements and $\varphi : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \mapsto (\widetilde{E}, \widetilde{\mathcal{E}})$ is a measurable function, then the sequence $(\widetilde{v}_t)_{t\in\mathbb{Z}}$ defined by $\widetilde{v}_t = \varphi(v_t, v_{t-1}, \dots)$ is a stationary ergodic process.

5.2. *Proof of Lemma* 1. We treat the three assertions of the lemma one after the other.

1. Define $f_\theta^{t,p} = f_\theta(X_{t-1},\dots,X_{t-p},0,0,\dots)$ for all $t\in\mathbb{Z}$ and $p\in\mathbb{N}$. We have $f_\theta^{t,p}\in\mathbb{L}^r(\mathcal{C}(\Theta,\mathbb{R}^m))$ because $\theta_0\in\Theta(r)$ and, using Proposition 1, all the following quantities are finite:

$$\begin{aligned}(\mathbb{E}[\|f_\theta^{t,p}\|_\Theta^r])^{1/r} &\le (\mathbb{E}[\|f_\theta^{t,0}-f_\theta^{t,p}\|_\Theta^r])^{1/r} + (\mathbb{E}[\|f_\theta^{t,0}\|_\Theta^r])^{1/r}\\ &\le \Big(\sum_{j\ge1}\alpha_j^{(0)}(f,\Theta)\Big)(\mathbb{E}[\|X_0\|^r])^{1/r} + \|f_\theta(0)\|_\Theta.\end{aligned}$$

For $p<q$,

$$\begin{aligned}\mathbb{E}[\|f_\theta^{t,p}-f_\theta^{t,q}\|_\Theta^r] &\le \mathbb{E}\Big[\Big\|\sum_{p<j\le q}\alpha_j^{(0)}(f,\Theta)X_{t-j}\Big\|^r\Big]\\ &\le \mathbb{E}[\|X_0\|^r]\Big(\sum_{p<j\le q}\alpha_j^{(0)}(f,\Theta)\Big)^r.\end{aligned}$$

Since $\sum_{j\ge1}\alpha_j^{(0)}(f,\Theta)<\infty$, $(f_\theta^{t,p})_{p\ge0}$ satisfies the Cauchy criteria in $\mathbb{L}^r(\mathcal{C}(\Theta,\mathbb{R}^m))$, and it converges to $f_\theta^{t,\infty}$; that is, f_θ^t on $\sigma(X_{t_1},\dots,X_{t_n})$, for all $n\in\mathbb{N}^*$ and $t>t_1>\cdots>t_n$ [those σ-algebras generate $\sigma(X_{t-1},X_{t-2},\dots)$ and, therefore, $f_\theta^{t,\infty}=_{\text{a.s.}} f_\theta^t$].

2. Define $H_\theta^{t,p} = H_\theta(X_{t-1},\dots,X_{t-p},0,\dots)$ for all $p\in\mathbb{N}$ and $t\in\mathbb{N}$. From Proposition 1, $\theta_0\in\Theta(r)$ and common inequalities satisfied by matrix norms, $H_\theta^{t,p}\in\mathbb{L}^{r/2}(\mathcal{C}(\Theta,\mathcal{M}_m))$, since, denoting $M_\theta^{t,p}=M_\theta(X_{t-1},\dots,X_{t-p},0,\dots)$,

$$\|H_\theta^{t,p}\|_\Theta^{r/2} \le \|M_\theta^{t,p}\|_\Theta^r \le \Big(\|M_\theta(0)\|_\Theta + \sum_{j=1}^\infty \|X_{t-j}\|\alpha_j^{(0)}(M,\Theta)\Big)^r.$$

We conclude as above that $H_\theta^t\in\mathbb{L}^{r/2}(\mathcal{C}(\Theta,\mathcal{M}_m))$ by bounding, for $p<q$,

$$\|H_\theta^{t,p}-H_\theta^{t,q}\|_\Theta^{r/2} \le \|M_\theta^{t,p}-M_\theta^{t,q}\|_\Theta^{r/2}(\|M_\theta^{t,p}\|_\Theta^{r/2}+\|M_\theta^{t,q}\|_\Theta^{r/2}).$$

The Cauchy–Schwarz inequality implies that

$$\begin{aligned}\mathbb{E}[\|H_\theta^{t,p}-H_\theta^{t,q}\|_\Theta^{r/2}] &\le (\mathbb{E}[\|M_\theta^{t,p}-M_\theta^{t,q}\|_\Theta^r])^{1/2}\\ &\qquad\times\big[(\mathbb{E}[\|M_\theta^{t,p}\|_\Theta^r])^{1/2}+(\mathbb{E}[\|M_\theta^{t,q}\|_\Theta^r])^{1/2}\big]\\ &\le B\Big(\mathbb{E}\Big[\Big(\sum_{p<j\le q}\alpha_j^{(0)}(M,\Theta)\|X_{t-j}\|\Big)^r\Big]\Big)^{1/2}\\ &\le B(\mathbb{E}[\|X_0\|^r])^{1/2}\Big(\sum_{p<j\le q}\alpha_j^{(0)}(M,\Theta)\Big)^{r/2}\end{aligned}$$

for some constant $B>0$.

3. First, notice that $\|X_0X_0'\| \le \|X_0\|^2$. Next, as in the previous proofs, $(H_\theta^{t,p})_{p\in\mathbb{N}^*}$ converges to H_θ^t in $\mathbb{L}^{r/2}(\mathcal{C}(\Theta,\mathcal{M}_m))$. Thus, there exists a subsequence $(p_k)_{k\in\mathbb{N}}$ such that $\|H_\theta^{t,p_k} - H_\theta^t\|_\Theta \xrightarrow[k\to\infty]{\text{a.s.}} 0$. Thanks to the continuity of the determinant, $(\det H_\theta^{t,p_k})_{k\in\mathbb{N}}$ also converges a.s. to $\det H_\theta^t$. Then, $\det H_\theta^t \ge \underline{H}$, H_θ^t is an invertible matrix, and, in view of elementary relations between matrix norm and determinant, $\|(\widehat{H}_\theta^t)^{-1}\|_\Theta \le \underline{H}^{-1/m}$.

5.3. *Proof of Theorem* 1. The proof of the theorem is divided into two parts. In (i), a uniform (in θ) law of large numbers on $(\widehat{q}_t)_{t\in\mathbb{N}^*}$ [defined in (1.3)] is established. In (ii), it is proved that $L(\theta) := -\mathbb{E}(q_t(\theta))/2$ has a unique maximum in θ_0. Those two conditions lead to the consistency of $\widehat{\theta}_n$.

(i) Using Proposition 1, with $q_t = G(X_t, X_{t-1},\ldots)$, one deduces that $(q_t)_{t\in\mathbb{Z}}$ [defined in (1.2)] is a stationary ergodic sequence. From Straumann and Mikosch [25], we know that, if $(v_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of random elements with values in $\mathbb{C}(\Theta,\mathbb{R}^m)$, then the uniform (in $\theta\in\Theta$) law of large numbers is implied by $\mathbb{E}\|v_0\|_\Theta < \infty$. As a consequence, $(q_t)_{t\in\mathbb{Z}}$ satisfies a uniform (in $\theta\in\Theta$) strong law of large numbers as soon as $\mathbb{E}[\sup_\theta |q_t(\theta)|] < \infty$. But, from the inequality $\log(x) \le x-1$, for all $x \in]0,\infty[$ and Lemma 1, for all $t\in\mathbb{Z}$,

$$
\begin{aligned}
&|q_t(\theta)| \le \frac{\|X_t - f_t(\theta)\|^2}{(\underline{H})^{1/m}} + m\left|\frac{1}{m}\log \underline{H} + \frac{\|H_\theta^t\|}{\underline{M}^{1/m}} - 1\right| \qquad \text{for all } \theta\in\Theta \\
&\Longrightarrow \quad \sup_{\theta\in\Theta} |q_t(\theta)| \le \frac{\|X_t - f_t(\theta)\|_\Theta^2}{(\underline{H})^{1/m}} + |\log \underline{H}| + m\times \frac{\|H_\theta^t\|_\Theta}{\underline{H}^{1/m}}.
\end{aligned}
\tag{5.2}
$$

But, $\forall t\in\mathbb{Z}$, $\mathbb{E}\|X_t\|^r < \infty$ (see Proposition 1) and $\mathbb{E}[\|f_\theta^t\|_\Theta^r] + \mathbb{E}[\|H_\theta^t\|_\Theta^{r/2}] < \infty$ (see Lemma 1). As a consequence, the right-hand side of (5.2) has a finite first moment and, therefore,

$$\mathbb{E}\Big[\sup_{\theta\in\Theta} |q_t(\theta)|\Big] < \infty.$$

The uniform strong law of large numbers for $(q_t(\theta))$ follows; hence,

$$\left\|\frac{L_n(\theta)}{n} - L(\theta)\right\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0 \qquad \text{with } L(\theta) := -\frac{1}{2}\mathbb{E}[q_0(\theta)]. \tag{5.3}$$

Now, one shows that $\frac{1}{n}\|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0$. Indeed, for all $\theta\in\Theta$ and $t\in\mathbb{N}^*$,

$$
\begin{aligned}
&|\widehat{q}_t(\theta) - q_t(\theta)| \\
&\quad = (\log\det \widehat{H}_\theta^t - \log\det H_\theta^t) + (X_t - \widehat{f}_\theta^t)'(\widehat{H}_\theta^t)^{-1}(X_t - \widehat{f}_\theta^t) \\
&\qquad - (X_t - f_\theta^t)'(H_\theta^t)^{-1}(X_t - f_\theta^t)
\end{aligned}
$$

$$\le |C|^{-1}|\det(\widehat{H}_\theta^t) - \det(H_\theta^t)| + (X_t - \widehat{f}_\theta^t)'[(\widehat{H}_\theta^t)^{-1} - (H_\theta^t)^{-1}](X_t - \widehat{f}_\theta^t) \tag{5.4}$$

$$
\begin{aligned}
&\quad + (2X_t - \widehat{f}_\theta^t - f_\theta^t)'(H_\theta^t)^{-1}(f_\theta^t - \widehat{f}_\theta^t) \\
&\le \underline{H}^{-1}\|\det(\widehat{H}_\theta^t) - \det(H_\theta^t)\|_\Theta + 2(\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta)\|(\widehat{H}_\theta^t)^{-1} - (H_\theta^t)^{-1}\|_\Theta \\
&\quad + (2\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta)\|(H_\theta^t)^{-1}\|_\Theta\|f_\theta^t - \widehat{f}_\theta^t\|_\Theta
\end{aligned}
$$

by the mean value theorem, with $C \in [\det(H_\theta^t), \det(\widehat{H}_\theta^t)]$ and, therefore, $|C| > \underline{H}$. On the one hand,

$$\|(\widehat{H}_\theta^t)^{-1} - (H_\theta^t)^{-1}\|_\Theta \le \|(\widehat{H}_\theta^t)^{-1}\|_\Theta\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta \cdot \|(H_\theta^t)^{-1}\|_\Theta.$$

On the other hand, for an invertible matrix $A \in \mathcal{M}_m(\mathbb{R})$, and $H \in \mathcal{M}_m(\mathbb{R})$,

$$\det(A + H) = \det(A) + \det(A) \cdot \mathrm{Tr}((A^{-1})'H) + o(\|H\|),$$

where $|\mathrm{Tr}((A^{-1})'H)| \le \|A^{-1}\| \cdot \|H\|$. Using the relation $\|(H_\theta^t)^{-1}\|_\Theta \ge \underline{H}^{-m}$ for all $t \in \mathbb{Z}$, there exists $C > 0$ not depending on t, such that inequality (5.4) becomes

$$\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)| \le C(\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta) \times (\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta + \|f_\theta^t - \widehat{f}_\theta^t\|_\Theta).$$

From the Hölder and Minkowski inequalities and by virtue of $3/2 = 1 + 1/2$,

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|^{2/3}\Big] &\le C(\mathbb{E}[\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta]^2)^{1/3} \\
&\quad \times (\mathbb{E}[\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta] + \mathbb{E}[\|f_\theta^t - \widehat{f}_\theta^t\|_\Theta])^{2/3} \qquad (5.5)\\
&\le C'\Big(\sum_{j\ge t}[\alpha_j^{(0)}(f,\Theta) + \alpha_j^{(0)}(M,\Theta)]\Big)^{2/3}
\end{aligned}
$$

with $C' > 0$ not depending on θ and t. Now, consider, for $n \in \mathbb{N}^*$,

$$S_n := \sum_{t=1}^n \frac{1}{t} \sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|.$$

Applying the Kronecker lemma (see Feller [13], page 238), if $\lim_{n\to\infty} S_n < \infty$ a.s., then $\frac{1}{n} \cdot \|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0$. Following Feller's arguments, it remains to show that, for all $\varepsilon > 0$,

$$\mathbb{P}(\forall n \in \mathbb{N}, \exists m > n \text{ such that } |S_m - S_n| > \varepsilon) := \mathbb{P}(A) = 0.$$

Let $\varepsilon > 0$, and denote

$$A_{m,n} := \{|S_m - S_n| > \varepsilon\}$$

for $m > n$. Notice that $A = \bigcap_{n\in\mathbb{N}} \bigcup_{m>n} A_{m,n}$. For $n \in \mathbb{N}^*$, the sequence of sets $(A_{m,n})_{m>n}$ is obviously increasing, and, if $A_n := \bigcup_{m>n} A_{m,n}$, then $\lim_{m\to\infty} \mathbb{P}(A_{m,n}) = \mathbb{P}(A_n)$. Observe that $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of sets and, thus,

$$\lim_{n\to\infty}\lim_{m\to\infty} \mathbb{P}(A_{m,n}) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

It remains to bound $\mathbb{P}(A_{m,n})$. From the Bienaymé–Chebyshev inequality,

$$\begin{aligned}\mathbb{P}(A_{m,n}) &= \mathbb{P}\Biggl(\sum_{t=n+1}^{m} \frac{1}{t} \sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)| > \varepsilon\Biggr)\\ &\le \frac{1}{\varepsilon^{2/3}}\mathbb{E}\Biggl[\Biggl(\sum_{t=n+1}^{m} \frac{1}{t} \sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|\Biggr)^{2/3}\Biggr]\\ &\le \frac{1}{\varepsilon^{2/3}} \sum_{t=n+1}^{m} \frac{1}{t^{2/3}}\mathbb{E}\Bigl[\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|^{2/3}\Bigr].\end{aligned}$$

Using (5.5) and condition (3.4), since $\ell > 3/2$, there exists $C > 0$ such that

$$\Biggl(\sum_{j=t}^{\infty} \alpha_j^{(0)}(f,\Theta) + \alpha_j^{(0)}(M,\Theta) + \alpha_j^{(0)}(H,\Theta)\Biggr)^{2/3} \le \frac{C}{t^{2(\ell-1)/3}}.$$

Thus, $t^{-2/3}\mathbb{E}[\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|^{2/3}] \le C(t^{-2\ell/3})$ for some $C > 0$, and

$$\sum_{t=1}^{\infty} \frac{1}{t^{2/3}}\mathbb{E}\Bigl[\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|^{2/3}\Bigr] < \infty \qquad \text{as } \ell > 3/2.$$

Thus, $\lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}(A_{m,n}) \underset{n\to\infty}{\longrightarrow} 0$ and $\frac{1}{n} \cdot \|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0$.

(ii) See Proposition 2.1 of Jeantheau [17].

5.4. *Proof of Theorem* 2. Let V be a Banach space (thereafter $V = \mathbb{R}^m$ or $V = \mathcal{M}_m$) and $\mathcal{D}^{(2)}\mathcal{C}(\Theta, V)$ denote the Banach space of V-valued 2 times continuously differentiable functions on Θ equipped with the uniform norm

$$\|g\|_{(2),\Theta} = \|g\|_\Theta + \left\|\frac{\partial g}{\partial\theta}\right\|_\Theta + \left\|\frac{\partial^2 g}{\partial\theta\,\partial\theta'}\right\|_\Theta.$$

We start by proving the following preliminary lemma.

LEMMA 2. *Let* $\theta_0 \in \Theta(r)$ $(r \ge 2)$, *and assume that* $\theta_0 \in \Theta$, *a compact set of* $\mathbb{R}^d$ *such that*, *for* $i = 1, 2$, $(\mathrm{A}_i(f,\Theta))$ *and* $(\mathrm{A}_i(M,\Theta))$ [*or* $(\mathrm{A}_i(H,\Theta))$] *hold. Then*,

$$f_\theta^t \in \mathbb{L}^r\bigl(\mathcal{D}^{(2)}\mathcal{C}(\Theta, \mathbb{R}^m)\bigr) \quad \textit{and} \quad H_\theta^t \in \mathbb{L}^{r/2}\bigl(\mathcal{D}^{(2)}\mathcal{C}(\Theta, \mathcal{M}_m)\bigr).$$

In view of the results of Lemmas 1 and 2, the functions $\partial L_n(\theta)/\partial\theta$ and $\partial^2 L_n(\theta)/\partial\theta^2$ are measurable and a.s. finite for all $\theta \in \Theta$. Their asymptotic properties are described in the next two lemmas.

LEMMA 3. *Let $\theta_0 \in \Theta(r)$ $(r \geq 4)$ and assume that $\theta_0 \in \Theta$, a compact set of $\mathbb{R}^d$ such that, for $i = 1, 2$, $(\mathrm{A}_i(f, \Theta))$ and $(\mathrm{A}_i(M, \Theta))$ [or $(\mathrm{A}_i(H, \Theta))$] hold. Then,*

$$n^{-1/2} \frac{\partial L_n(\theta_0)}{\partial \theta} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0, G(\theta_0)), \tag{5.6}$$

where $G(\theta_0) = (G(\theta_0))_{1 \leq i, j \leq d}$ is finite and its expression is given in (5.14).

LEMMA 4. *Under the assumptions of Lemma* 3,

$$\left\| \frac{1}{n} \frac{\partial^2 L_n(\theta)}{\partial \theta\, \partial \theta'} - \frac{\partial^2 L(\theta)}{\partial \theta\, \partial \theta'} \right\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0 \qquad \textit{with } \frac{\partial^2 L(\theta)}{\partial \theta\, \partial \theta'} := -\frac{1}{2} \mathbb{E}\left[\frac{\partial^2 q_0}{\partial \theta\, \partial \theta'}(\theta) \right]. \tag{5.7}$$

We postponed the proofs of Lemmas 1–4 to the end of the section and continue with the proof of Theorem 2. From Theorem 1, we have

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{\text{a.s.}} \theta_0. \tag{5.8}$$

Since $\theta_0 \in \overset{\circ}{\Theta}$, a Taylor expansion of $\partial L_n(\theta_0)/\partial \theta_i \in \mathbb{R}$ implies

$$\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta_i} = \frac{\partial L_n(\theta_0)}{\partial \theta_i} + \frac{\partial^2 L_n(\overline{\theta}_{n,i})}{\partial \theta\, \partial \theta_i} (\widehat{\theta}_n - \theta_0) \tag{5.9}$$

for n sufficiently large such that the $\overline{\theta}_{n,i} \in \Theta$, which are between $\widehat{\theta}_n$ and θ_0, for all $1 \leq i \leq d$. Using (5.7) and (5.8), we conclude, with the uniform convergence theorem, that

$$F_n := -2 \left(\frac{1}{n} \frac{\partial^2 L_n(\overline{\theta}_{n,i})}{\partial \theta\, \partial \theta_i} \right)_{1 \leq i \leq d} \xrightarrow[n\to\infty]{\text{a.s.}} F(\theta_0).$$

One obtains $(F(\theta_0))_{ij} = \mathbb{E}[\partial^2 q_0(\theta_0)/\partial \theta_i\, \partial \theta_j]$ for $1 \leq i, j \leq d$. With similar arguments as for (5.13), since $X_t - f^t_{\theta_0} = M_{\theta_0} \xi_t$, with ξ_t independent of $(X_{t-1}, X_{t-2}, \ldots)$,

$$\begin{aligned} &\mathbb{E}\left[(X_t - f^t_{\theta_0})' \frac{\partial^2 (H^t_\theta)^{-1}}{\partial \theta_i\, \partial \theta_j} (X_t - f^t_{\theta_0}) \right] \\ &\quad = 2\mathbb{E}\left[\operatorname{Tr}\left((H^t_{\theta_0})^{-2} \frac{\partial H^t_{\theta_0}}{\partial \theta_j} \frac{\partial H^t_{\theta_0}}{\partial \theta_i} \right) - \operatorname{Tr}\left((H^t_{\theta_0})^{-1} \frac{\partial^2 H^t_{\theta_0}}{\partial \theta_j\, \partial \theta_i} \right) \right]. \end{aligned}$$

From (5.15), we then derive the explicit expression

$$(F(\theta_0))_{ij} = \mathbb{E}\left[2 \left(\frac{\partial f^t_{\theta_0}}{\partial \theta_j} \right)' (H^t_{\theta_0})^{-1} \frac{\partial f^t_{\theta_0}}{\partial \theta_i} + \operatorname{Tr}\left((H^t_{\theta_0})^{-2} \frac{\partial H^t_{\theta_0}}{\partial \theta_j} \frac{\partial H^t_{\theta_0}}{\partial \theta_i} \right) \right]. \tag{5.10}$$

Under assumption (Var), $F(\theta_0)$ is a positive definite $d \times d$ matrix. Indeed, for all $Y = (y_1, \dots, y_d) \in \mathbb{R}^d$,

$$Y' F(\theta_0) Y = \mathbb{E}\Big[2\Big(\sum_{1\le i\le d} y_i \frac{\partial f^t_{\theta_0}}{\partial \theta_i}\Big)' (H^t_{\theta_0})^{-1} \Big(\sum_{1\le i\le d} y_i \frac{\partial f^t_{\theta_0}}{\partial \theta_i}\Big) \\ + \operatorname{Tr}\Big((H^t_{\theta_0})^{-2}\Big(\sum_{1\le i\le d} y_i \frac{\partial H^t_{\theta_0}}{\partial \theta_i}\Big)^2\Big)\Big].$$

These two terms are nonnegative and at least one of them is positive under assumption (Var). Then, $F(\theta_0)$ is an invertible matrix, and there exists n large enough such that F_n is an invertible matrix. Moreover, (5.9) implies,

$$n(\widehat{\theta}_n - \theta_0) = -2F_n^{-1}\Big(\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta} - \frac{\partial L_n(\theta_0)}{\partial \theta}\Big).$$

Therefore, if $\frac{1}{\sqrt{n}}\|\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta}\| \xrightarrow[n\to\infty]{\mathcal{P}} 0$, using Lemma 3, one obtains Theorem 2. Since $\frac{\partial \widehat{L}_n(\widehat{\theta}_n)}{\partial \theta} = 0$ ($\widehat{\theta}_n$ is a local extremum for $\widehat{L}_n$),

$$\mathbb{E}\Big[\frac{1}{\sqrt{n}}\Big\|\frac{\partial L_n}{\partial \theta} - \frac{\partial \widehat{L}_n}{\partial \theta}\Big\|_\Theta\Big] \xrightarrow[n\to\infty]{} 0. \tag{5.11}$$

Using the relation (5.12), the following inequality

$$|a_1b_1c_1 - a_2b_2c_2| \le |a_1 - a_2||b_2||c_2| + |a_1||b_1 - b_2||c_2| + |a_1||b_1||c_1 - c_2|$$

and the bounds $\|(\widehat{H}^t_\theta)^{-1}\|_\Theta \le \underline{H}^{-1/m}$, $\|(H^t_\theta)^{-1}\|_\Theta \le \underline{H}^{-1/m}$, one obtains:

$$\begin{aligned}
&\Big\|\frac{\partial q_t(\theta)}{\partial \theta_i} - \frac{\partial \widehat{q}_t(\theta)}{\partial \theta_i}\Big\|_\Theta \\
&\quad\le \frac{2}{\underline{H}^{1/m}}\Big[\Big\|\frac{\partial \widehat{f}^t_\theta}{\partial \theta_i} - \frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta + \Big\|\frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta \|\widehat{f}^t_\theta - f^t_\theta\|_\Theta\Big] \\
&\qquad + 2\Big\|\frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta \|(H^t_\theta)^{-1} - (\widehat{H}^t_\theta)^{-1}\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta \\
&\qquad + \|\widehat{f}^t_\theta - f^t_\theta\|_\Theta \Big\|\frac{\partial (\widehat{H}^t_\theta)^{-1}}{\partial \theta_i}\Big\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta \\
&\qquad + \|X - f^t_\theta\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta \Big\|\frac{\partial (H^t_\theta)^{-1}}{\partial \theta_i} - \frac{\partial (\widehat{H}^t_\theta)^{-1}}{\partial \theta_i}\Big\|_\Theta \\
&\qquad + \|(\widehat{H}^t_\theta)^{-1}\|_\Theta \Big\|\frac{\partial H^t_\theta}{\partial \theta_i} - \frac{\partial \widehat{H}^t_\theta}{\partial \theta_i}\Big\|_\Theta \\
&\qquad + \|(H^t_\theta)^{-1} - (\widehat{H}^t_\theta)\|_\Theta \Big\|\frac{\partial (H^t_\theta)^{-1}}{\partial \theta_i}\Big\|_\Theta.
\end{aligned}$$

Since, for $i=1,2$, $(\mathrm{A}_i(f,\Theta))$ and $(\mathrm{A}_i(M,\Theta))$ [or $(\mathrm{A}_i(H,\Theta))$] hold, there exists $C>0$ such that

$$\mathbb{E}\|f_\theta^t-\widehat{f}_\theta^t\|_\Theta^r\le C\biggl(\sum_{j\ge t}\alpha_j^{(0)}(f,\Theta)\biggr)^r$$

and

$$\mathbb{E}\biggl\|\frac{\partial f_\theta^t}{\partial\theta_i}-\frac{\partial\widehat{f}_\theta^t}{\partial\theta_i}\biggr\|_\Theta^r\le C\biggl(\sum_{j\ge t}\alpha_j^{(1)}(f,\Theta)\biggr)^r.$$

The differences $\mathbb{E}\|H_\theta^t-\widehat{H}_\theta^t\|_\Theta^{r/2}\le C(\sum_{j\ge t}\alpha_j^{(0)}(M,\Theta))^{r/2}$ can also be bounded in the following way:

$$\mathbb{E}\biggl\|\frac{\partial H_\theta^t}{\partial\theta_i}-\frac{\partial\widehat{H}_\theta^t}{\partial\theta_i}\biggr\|_\Theta^{r/2}\le C\biggl(\biggl(\sum_{j\ge t}\alpha_j^{(0)}(M,\Theta)\biggr)^{r/2}+\biggl(\sum_{j\ge t}\alpha_j^{(1)}(M)\biggr)^{r/2}\biggr),$$

$$\mathbb{E}\biggl\|\frac{\partial(H_\theta^t)^{-1}}{\partial\theta_i}-\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta_i}\biggr\|_\Theta^{r/2}\le C\biggl(\biggl(\sum_{j\ge t}\alpha_j^{(0)}(M,\Theta)\biggr)^{r/2}+\biggl(\sum_{j\ge t}\alpha_j^{(1)}(M)\biggr)^{r/2}\biggr).$$

Finally, using Hölder inequalities, there exists another constant $C\ge 0$ satisfying

$$\begin{aligned}\mathbb{E}\biggl\|&\frac{\partial q_t(\theta)}{\partial\theta_i}-\frac{\partial\widehat{q}_t(\theta)}{\partial\theta_i}\biggr\|_\Theta\\&\le C\sum_{j\ge t}\bigl(\alpha_j^{(0)}(f,\Theta)+\alpha_j^{(0)}(M,\Theta)+\alpha_j^{(0)}(H,\Theta)\\&\qquad\qquad+\alpha_j^{(1)}(f,\Theta)+\alpha_j^{(1)}(M,\Theta)+\alpha_j^{(1)}(H,\Theta)\bigr).\end{aligned}$$

Under (3.5), $\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathbb{E}\|\frac{\partial q_t(\theta)}{\partial\theta_i}-\frac{\partial\widehat{q}_t(\theta)}{\partial\theta_i}\|_\Theta\xrightarrow[n\to\infty]{}0$, and Theorem 2 follows.

PROOF OF LEMMA 2. Here, we focus on the case of H_θ under $(\mathrm{A}_i(f,\Theta))$ and $(\mathrm{A}_i(M,\Theta))$, $i=1,2$. The other cases are similar and simpler.

With the same method and notation as in the proof of Lemma 1, the result holds as soon as the function $\theta\in\Theta\to H_\theta^{t,p}$ is proved to satisfy a Cauchy criterion in $\mathbb{L}^{r/2}(\mathcal{D}^{(2)}\mathcal{C}(\Theta,\mathcal{M}_m))$. Using the proof of Lemma 1, we already have $\mathbb{E}\|H_\theta^{t,p}\|_\Theta^{r/2}<\infty$. It remains to bound the quantities

$$\mathbb{E}\biggl\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}\biggr\|_\Theta^{r/2}\quad\text{and}\quad\mathbb{E}\biggl\|\frac{\partial^2 H_\theta^{t,p}}{\partial\theta_i\,\partial\theta_j}\biggr\|_\Theta^{r/2}\qquad\forall i,j\in\{1,\dots,d\},\forall p\in\mathbb{N}^*.$$

Using assumption ($\mathrm{A}_1(M, \Theta)$),

$$\begin{aligned}\left\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}\right\|_\Theta &\le 2\|M_\theta^{t,p}\|_\Theta\left\|\frac{\partial M_\theta^{t,p}}{\partial\theta_i}\right\|_\Theta\\ &\le \left(\|M_\theta(0)\|_\Theta+\sum_{j=1}^\infty \alpha_j^{(0)}(M,\Theta)\|X_{t-j}\|\right)\\ &\qquad\times\left(\left\|\frac{\partial M_\theta(0)}{\partial\theta_i}\right\|_\Theta+\sum_{j=1}^\infty \alpha_j^{(1)}(M,\Theta)\|X_{t-j}\|\right).\end{aligned}$$

Using $\mathbb{E}[\|X_0\|^r]<\infty$ and the Hölder and Minkowski inequalities,

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}\right\|_\Theta^{r/2}\right] &\le C\left(\|M_\theta(0)\|_\Theta^r+\mathbb{E}[\|X_0\|^r]\left(\sum_{j=1}^\infty \alpha_j^{(0)}(M,\Theta)\right)^r\right)^{1/2}\\ &\qquad\times\left(\left\|\frac{\partial M_\theta(0)}{\partial\theta_i}\right\|_\Theta^r+\mathbb{E}[\|X_0\|^r]\left(\sum_{j=1}^\infty \alpha_j^{(1)}(M,\Theta)\right)^r\right)^{1/2}.\end{aligned}$$

In the same way, there exists another constant $C>0$ such that

$$\begin{aligned}\mathbb{E}\left\|\frac{\partial^2 H_\theta^{t,p}}{\partial\theta_i\,\partial\theta_j}\right\|_\Theta^{r/2} &\le C\left[\left(\left(\sum_{j=1}^\infty \alpha_j^{(1)}(M,\Theta)\right)^r\left(\sum_{j=1}^\infty \alpha_j^{(1)}(M,\Theta)\right)^r\right)^{1/2}\right.\\ &\qquad\left.+\left(\left(\sum_{j=1}^\infty \alpha_j^{(0)}(M,\Theta)\right)^r\left(\sum_{j=1}^\infty \alpha_j^{(2)}(M,\Theta)\right)^r\right)^{1/2}\right].\end{aligned}$$

From $\sum_j \alpha_j^{(0)}(M,\Theta)<\infty$, $\sum_j \alpha_j^{(1)}(M,\Theta)<\infty$ and $\sum_j \alpha_j^{(2)}(M,\Theta)<\infty$, we deduce that $\mathbb{E}[\|H_\theta^{t,p}\|_{(2),\Theta}^{r/2}]<\infty$ for all $p\in\mathbb{N}^*$. In the same way as in the proof of Lemma 1, we can also prove that the sequence $(H_\theta^{t,p})_{p\in\mathbb{N}^*}$ satisfies the Cauchy criterion in the Banach space $\mathbb{L}^{r/2}(\mathcal{D}^{(2)}\mathcal{C}(\Theta,\mathcal{M}_m))$. For the first derivatives, the result easily follows from the inequality

$$\begin{aligned}\left\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}-\frac{\partial H_\theta^{t,q}}{\partial\theta_i}\right\|_\Theta &\le 2\|M_\theta^{t,p}-M_\theta^{t,q}\|_\Theta\left\|\frac{\partial M_\theta^{t,p}}{\partial\theta_i}\right\|_\Theta\\ &\qquad+2\|M_\theta^{t,q}\|_\Theta\left\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}-\frac{\partial H_\theta^{t,q}}{\partial\theta_i}\right\|_\Theta.\end{aligned}$$

For the second derivatives, a similar argument finishes the proof. □

PROOF OF LEMMA 3. Simple calculations give the relations

$$\frac{\partial(H_\theta^t)^{-1}}{\partial\theta_k}=-(H_\theta^t)^{-1}\frac{\partial H_\theta^t}{\partial\theta_k}(H_\theta^t)^{-1}\quad\text{and}\quad\frac{\partial\ln\det(H_\theta^t)}{\partial\theta_k}=\mathrm{Tr}\left((H_\theta^t)^{-1}\frac{\partial H_\theta^t}{\partial\theta_k}\right).$$

From Lemma 2, $\partial f_\theta^t/\partial\theta$, $\partial H_\theta^t/\partial\theta$ and $(\widehat{H}_\theta^t)^{-1}$ are a.s. finite. Then, $\partial L_n(\theta)/\partial\theta$ is an a.s. finite measurable function satisfying, for all $1 \le i \le d$, $\partial L_n(\theta)/\partial\theta_i = -\frac{1}{2}\sum_{t=1}^n \partial q_t(\theta)/\partial\theta_i$ with

$$\begin{aligned}\frac{\partial q_t(\theta)}{\partial\theta_k} &= -2\Big(\frac{\partial f_\theta^t}{\partial\theta_k}\Big)'(H_\theta^t)^{-1}(X_t - f_\theta^t)\\ &\quad + (X_t - f_\theta^t)'\frac{\partial (H_\theta^t)^{-1}}{\partial\theta_k}(X_t - f_\theta^t) + \operatorname{Tr}\Big((H_\theta^t)^{-1}\frac{\partial H_\theta^t}{\partial\theta_k}\Big).\end{aligned}\tag{5.12}$$

Denoting $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$, let us prove that $(\frac{\partial q_t(\theta_0)}{\partial\theta}, \mathcal{F}_t)_{t\in\mathbb{Z}}$ is a $\mathbb{R}^m$-valued martingale difference process. Indeed, for all $t \in \mathbb{Z}$,

$$\mathbb{E}\big((X_t - f_{\theta_0}^t)|\mathcal{F}_t\big) = 0 \quad \text{and} \quad \mathbb{E}\big((X_t - f_{\theta_0}^t)(X_t - f_{\theta_0}^t)'|\mathcal{F}_t\big) = H_{\theta_0}^t.$$

As a consequence,

$$\begin{aligned}\mathbb{E}\Big(\frac{\partial q_t(\theta_0)}{\partial\theta_k}\Big|\mathcal{F}_t\Big) &= \mathbb{E}\Big((X_t - f_{\theta_0}^t)'\frac{\partial (H_{\theta_0}^t)^{-1}}{\partial\theta_k}(X_t - f_{\theta_0}^t)\Big|\mathcal{F}_t\Big)\\ &\quad + \operatorname{Tr}\Big((H_{\theta_0}^t)^{-1}\frac{\partial H_{\theta_0}^t}{\partial\theta_k}\Big).\end{aligned}$$

We conclude by noticing that the first term of the sum is equal to

$$\mathbb{E}\Big(\operatorname{Tr}\Big(\frac{\partial (H_{\theta_0}^t)^{-1}}{\partial\theta_k}(X_t - f_{\theta_0}^t)(X_t - f_{\theta_0}^t)'\Big)\Big|\mathcal{F}_t\Big) = \operatorname{Tr}\Big(\frac{\partial (H_{\theta_0}^t)^{-1}}{\partial\theta_k}H_{\theta_0}^t\Big).$$

In order to apply the central limit theorem for martingale-differences (see [2]), we have to prove $\mathbb{E}[\|\frac{\partial q_t(\theta_0)}{\partial\theta}\|^2] < \infty$. Using the relation $X_t - f_{\theta_0}^t = M_{\theta_0}^t\xi_t$ for all $t \in \mathbb{Z}$, then

$$\begin{aligned}\frac{\partial q_t(\theta_0)}{\partial\theta_k} &= -2\Big(\frac{\partial f_{\theta_0}^t}{\partial\theta_k}\Big)'(H_{\theta_0}^t)^{-1}M_{\theta_0}^t\xi_t - \xi_t' M_{\theta_0}^{t\prime}(H_{\theta_0}^t)^{-1\prime}\frac{\partial H_{\theta_0}^t}{\partial\theta_k}(H_{\theta_0}^t)^{-1}M_{\theta_0}^t\xi_t\\ &\quad + \operatorname{Tr}\Big((H_{\theta_0}^t)^{-1}\frac{\partial H_{\theta_0}^t}{\partial\theta_k}\Big).\end{aligned}$$

With $\operatorname{Tr}(ABC) = \operatorname{Tr}(CAB) = \operatorname{Tr}(ACB)$ for symmetric matrices A, B, C,

$$\begin{aligned}&\mathbb{E}\Big[\Big(\xi_t' M_{\theta_0}^{t\prime}(H_{\theta_0}^t)^{-1\prime}\frac{\partial H_{\theta_0}^t}{\partial\theta_k}(H_{\theta_0}^t)^{-1}M_{\theta_0}^t\xi_t\Big)^2\Big]\\ &\qquad = \mathbb{E}\Big[(\xi_t'\xi_t)^2\operatorname{Tr}\Big(M_{\theta_0}^{t\prime}(H_{\theta_0}^t)^{-1}\frac{\partial H_{\theta_0}^t}{\partial\theta_k}(H_{\theta_0}^t)^{-1}\frac{\partial H_{\theta_0}^t}{\partial\theta_k}(H_{\theta_0}^t)^{-1}M_{\theta_0}^t\Big)\Big]\\ &\qquad = \mathbb{E}\Big[(\xi_t'\xi_t)^2\operatorname{Tr}\Big((H_{\theta_0}^t)^{-2}\Big(\frac{\partial H_{\theta_0}^t}{\partial\theta_k}\Big)^2\Big)\Big].\end{aligned}\tag{5.13}$$

Using this relation, the bound $\|(H^t_{\theta_0})^{-1}\|_\Theta \le \underline{H}^{-1/m}$ and the independence of ξ_t and $\mathcal{F}_t$, there exists $C > 0$ such that

$$\mathbb{E}\biggl[\biggl(\frac{\partial q_t(\theta_0)}{\partial \theta_k}\biggr)^2\biggr] \le C\biggl(\mathbb{E}\biggl[\biggl\|\frac{\partial f^t_{\theta_0}}{\partial \theta_k}\biggr\|^2 \|M^t_{\theta_0}\|^2\biggr] \\ \times \mathbb{E}[\|\xi_t\|^2] + \mathbb{E}\biggl[\biggl\|\frac{\partial H^t_{\theta_0}}{\partial \theta_k}\biggr\|^2\biggr] \\ + \mathbb{E}[\|\xi_t'\xi_t\|^2] \times \mathbb{E}\biggl[\biggl\|\frac{\partial H^t_{\theta_0}}{\partial \theta_k}\biggr\|^2\biggr]\biggr).$$

Therefore, since $r \ge 4$, the moment conditions for the CLT are fulfilled

$$\mathbb{E}\biggl[\biggl\|\frac{\partial q_t(\theta_0)}{\partial \theta}\biggr\|^2\biggr] = \sum_{k=1}^d \mathbb{E}\biggl[\frac{\partial q_t(\theta_0)}{\partial \theta_k}\biggr]^2 < \infty.$$

We compute the asymptotic covariance matrix of $\frac{\partial q_t(\theta_0)}{\partial \theta}$. Thus,

$$\begin{aligned}(G(\theta_0))_{ij} &= \mathbb{E}\biggl[\frac{\partial q_t(\theta_0)}{\partial \theta_i}\frac{\partial q_t(\theta_0)}{\partial \theta_j}\biggr] \\ &= \mathbb{E}\biggl[4\biggl(\frac{\partial f^t_{\theta_0}}{\partial \theta_i}\biggr)'(H^t_{\theta_0})^{-1}\biggl(\frac{\partial f^t_{\theta_0}}{\partial \theta_j}\biggr) \\ &\qquad - \mathrm{Tr}\biggl((H^t_{\theta_0})^{-1}\frac{\partial H^t_{\theta_0}}{\partial \theta_i}\biggr)\mathrm{Tr}\biggl((H^t_{\theta_0})^{-1}\frac{\partial H^t_{\theta_0}}{\partial \theta_j}\biggr) \\ &\qquad + p\bigl(m_4 + (p-1)\bigr)\mathrm{Tr}\biggl((H^t_{\theta_0})^{-2}\frac{\partial H^t_{\theta_0}}{\partial \theta_i}\frac{\partial H^t_{\theta_0}}{\partial \theta_j}\biggr)\biggr].\end{aligned} \tag{5.14}$$

To simplify the expression, we assume here that ξ_t and $-\xi_t$ have the same distribution, in order that $\mathbb{E}[\xi_t\xi_t' A\xi_t] = 0$ for A, a matrix. $\square$

PROOF OF THE LEMMA 4. From the proof of Proposition 1 and from the result of Lemma 2, the second derivative process $(\partial^2 q_t(\theta)/\partial\theta^2)_{t\in\mathbb{Z}}$ is stationary ergodic (it is a measurable function of $X_t, X_{t-1}, \ldots$). Therefore, it satisfies a Uniform Law of Large Numbers (ULLN) if its first uniform moment is bounded.

From (5.12), the second partial derivatives of $q_t(\theta)$ are

$$\begin{aligned}\frac{\partial^2 q_t(\theta)}{\partial \theta_i\,\partial \theta_j} &= -2\biggl(\frac{\partial^2 f^t_\theta}{\partial \theta_i\,\partial \theta_j}\biggr)'(H^t_\theta)^{-1}(X_t - f^t_\theta) \\ &\qquad + (X_t - f^t_\theta)'\frac{\partial^2 (H^t_\theta)^{-1}}{\partial \theta_i\,\partial \theta_j}(X_t - f^t_\theta) \\ &\qquad - 2\biggl(\biggl(\frac{\partial f^t_\theta}{\partial \theta_i}\biggr)'\frac{\partial (H^t_\theta)^{-1}}{\partial \theta_j} + \biggl(\frac{\partial f^t_\theta}{\partial \theta_j}\biggr)'\frac{\partial (H^t_\theta)^{-1}}{\partial \theta_i}\biggr)(X_t - f^t_\theta)\end{aligned} \tag{5.15}$$

$$+ 2\left(\frac{\partial f_\theta^t}{\partial\theta_i}\right)'(H_\theta^t)^{-1}\left(\frac{\partial f_\theta^t}{\partial\theta_i}\right)$$
$$+ \mathrm{Tr}\left(\left(\frac{\partial(H_\theta^t)^{-1}}{\partial\theta_j}\right)\left(\frac{\partial H_\theta^t}{\partial\theta_i}\right)\right) + \mathrm{Tr}\left((H_\theta^t)^{-1}\left(\frac{\partial^2 H_\theta^t}{\partial\theta_i\,\partial\theta_j}\right)\right).$$

Therefore, using the bound $\|(H_\theta^t)^{-1}\|_\Theta \le \underline{M}^{-1/m}$ of Lemma 1 and usual relations between norms and traces of matrix, there exists $C > 0$ such that

$$\begin{aligned}\left\|\frac{\partial^2 q_t(\theta)}{\partial\theta_i\,\partial\theta_j}\right\|_\Theta \le C\Bigg[\Bigg(&\left\|\frac{\partial^2 f_\theta^t}{\partial\theta_i\,\partial\theta_j}\right\|_\Theta + \left\|\frac{\partial H_\theta^t}{\partial\theta_j}\right\|_\Theta\left\|\frac{\partial f_\theta^t}{\partial\theta_i}\right\|_\Theta \\ &+ \left\|\frac{\partial H_\theta^t}{\partial\theta_i}\right\|_\Theta\left\|\frac{\partial f_\theta^t}{\partial\theta_j}\right\|\Bigg)\|X_t - f_\theta^t\|_\Theta \\ &+ \left\|\frac{\partial^2 H_\theta^t}{\partial\theta_i\,\partial\theta_j}\right\|_\Theta\|X_t - f_\theta^t\|_\Theta^2 + \left\|\frac{\partial f_\theta^t}{\partial\theta_i}\right\|_\Theta\left\|\frac{\partial f_\theta^t}{\partial\theta_j}\right\|_\Theta \\ &+ \left\|\frac{\partial H_\theta^t}{\partial\theta_i}\right\|_\Theta\left\|\frac{\partial H_\theta^t}{\partial\theta_j}\right\|_\Theta\Bigg].\end{aligned}$$

We conclude that $\mathbb{E}\|\frac{\partial^2 q_t(\theta)}{\partial\theta_i\,\partial\theta_j}\|_\Theta^{r/4} < \infty$ ($r \ge 4$), since, for $t \in \mathbb{Z}$, $1 \le i, j \le d$,

$$\mathbb{E}[\|X_t\|^r] < +\infty, \qquad \mathbb{E}[\|f_\theta^t\|_\Theta^r] < +\infty,$$
$$\mathbb{E}\left[\left\|\frac{\partial f_\theta^t}{\partial\theta_i}\right\|_\Theta^r\right] < +\infty, \qquad \mathbb{E}\left[\left\|\frac{\partial^2 f_\theta^t}{\partial\theta_i\,\partial\theta_j}\right\|_\Theta^r\right] < +\infty;$$
$$\mathbb{E}[\|H_\theta^t\|_\Theta^{r/2}] < +\infty, \qquad \mathbb{E}\left[\left\|\frac{\partial H_\theta^t}{\partial\theta_i}\right\|_\Theta^{r/2}\right] < +\infty, \qquad \mathbb{E}\left[\left\|\frac{\partial^2 H_\theta^t}{\partial\theta_i\,\partial\theta_j}\right\|_\Theta^{r/2}\right] < \infty.$$

As a consequence, the ULLN holds for $\partial^2 q_t(\theta)/\partial\theta^2$. $\square$

5.5. *Proof of sufficient conditions for* (Id(Θ)) *and* (Var) *for ARCH processes.* First, we express a key lemma derived from Lemma 3.1 in [16].

LEMMA 5. *Assume that $(\xi_t)_t$ is a sequence of i.i.d. r.v. satisfying* (Id′). *Then, for all $t \in \mathbb{Z}$, if Q_t and $\delta\ \xi_t$ are $\sigma((X_{t-k})_{k\in\mathbb{N}^*})$-measurable, respectively, quadratic form and real variable, and if $(\xi_t)_t$ is independent of $\sigma((X_{t-k})_{k\in\mathbb{N}^*})$, then*

$$Q_t(\xi_t) = \delta_t, \qquad a.s. \quad \Rightarrow \quad Q_t = 0 \quad and \quad \delta_t = 0, \qquad a.s.$$

PROOF. The proof is similar to that in [16]. As ξ_t is independent of $(X_{t-1}, X_{t-2}, \ldots)$, we apply the Fubini theorem thusly:

$$\mathbb{P}(Q_t(\xi_t) = \delta_t) = \int \mathbb{P}(Q_t^{(w)}(\xi_t) = \delta_t^{(w)})\,d\mu(w),$$

denoting by μ the distribution of $(X_{t-1}, X_{t-2}, \dots)$. As this quantity is equal to 1, and since $\mathbb{P}(Q_t^{(w)}(\xi_t) = \delta_t^{(w)}) = \mathbb{P}(Q_t^{(w)}(\xi_0) = \delta_t^{(w)})$, we derive that $\mathbb{P}(Q_t^{(w)} = \delta_t^{(w)}) = 1$; thus, $Q_t^{(w)} = 0 = \delta_t^{(w)}$ using (Id′) μ-almost everywhere. This ends the proof. $\square$

LEMMA 6. *Under assumptions of the part* SC *of Proposition* 4, *then* (Id(Θ)) *holds*.

The basic idea is a recursive use of Lemma 5. Assume that, for some $t \in \mathbb{Z}$ and some $\theta \in \Theta$, we have $H_\theta^t = H_{\theta_0}^t$ a.s. Identifying $H_\theta^t = B_0(\theta) + \sum_{j=1}^\infty B_j(\theta) X_{t-j} X'_{t-j} B'_j(\theta)$, we express the equation $H_\theta^t = H_{\theta_0}^t$ a.s. for the first element, indexed by $(1, 1)$, thusly:

$$
\begin{aligned}
&B_0^{(1,1)}(\theta_0) + \sum_{j=1}^\infty B_j^{(1.)}(\theta_0) X_{t-j} X'_{t-j} \bigl(B_j^{(1.)}(\theta_0)\bigr)' \\
&\quad = B_0^{(1,1)}(\theta) + \sum_{j=1}^\infty B_j^{(1.)}(\theta) X_{t-j} X'_{t-j} \bigl(B_j^{(1.)}(\theta)\bigr)',
\end{aligned}
\tag{5.16}
$$

where $A^{(i.)}$ is the ith row of A and $A^{(i,k)}$ the element indexed by (i, k) of A. For some $\sigma(X_{t-2}, X_{t-3}, \dots)$-measurable random variable δ_1, we have the equation

$$
\begin{aligned}
Q_{t-1}(\xi_{t-1}) = {} & B_1^{(1.)}(\theta_0) M_{\theta_0}^{t-1} \xi_{t-1} \xi'_{t-1} (M_{\theta_0}^{t-1})' \bigl(B_1^{(1.)}(\theta_0)\bigr)' \\
& - B_1^{(1.)}(\theta) M_{\theta_0}^{t-1} \xi_{t-1} \xi'_{t-1} (M_{\theta_0}^{t-1})' \bigl(B_1^{(1.)}(\theta)\bigr)' = \delta_{t-1}.
\end{aligned}
$$

Applying Lemma 5, we get that $Q_{t-1} = 0$ and $\delta_{t-1} = 0$ a.s. Using the relation $\delta_{t-1} = 0$ a.s. and the expression of δ_{t-1}, we prove recursively on j that $Q_{t-j} = 0$ and $\delta_{t-j} = 0$ for any $j \ge 1$. As $Q_{t-j} j(\xi_{t-j}) = 0$, we get that

$$
B_j^{(1.)}(\theta_0) X_{t-j} X'_{t-j} \bigl(B_j^{(1.)}(\theta_0)\bigr)' = B_j^{(1.)}(\theta) X_{t-j} X'_{t-j} \bigl(B_j^{(1.)}(\theta)\bigr)'
$$

and we use these relations for all $j \ge 1$ in (5.16) to obtain that $B_0^{(1,1)}(\theta_0) = B_0^{(1,1)}(\theta)$. We also use that $Q_{t-j}(e_k) = 0$ p.s. for e_k the kth element of the canonical basis of $\mathbb{R}^p$ to get the equation

$$
\bigl(B_1^{(1.)}(\theta_0) M_{\theta_0}^{t-1\,(.k)}\bigr)^2 = \bigl(B_1^{(1.)}(\theta) M_{\theta_0}^{t-1\,(.k)}\bigr)^2.
$$

The same arguments applied to the second element, indexed by $(1, 2)$, implies that $B_0^{(1,2)}(\theta_0) = B_0^{(1,2)}(\theta)$ and the equation

$$
B_j^{(1.)}(\theta_0) M_{\theta_0}^{t-1\,(.k)} B_j^{(2.)}(\theta_0) M_{\theta_0}^{t-1\,(.k)} = B_j^{(1.)}(\theta) M_{\theta_0}^{t-1\,(.k)} B_j^{(2.)}(\theta) M_{\theta_0}^{t-1\,(.k)}.
$$

Hence, we have $B_0^{(i,i')}(\theta_0) = B_0^{(i,i')}(\theta)$ and $a_{i,k,j}(\theta_0) a_{i',k,j}(\theta_0) = a_{i,k,j}(\theta) \times a_{i',k,j}(\theta)$, for all $1 \le i, i', k \le m$ and $j \ge 1$, where $a_{i,k,j}(\theta) = B_j^{(i.)}(\theta) M_{\theta_0}^{t-1\,(.k)}$.

It directly follows that $B_0(\theta_0)=B_0(\theta)$. If $B_j(\theta_0)\neq 0$, then $B_j(\theta_0)$ is definite positive by assumption and, for any $1\leq k\leq m$, there exists some $1\leq i\leq m$ such that $a_{i,k,j}(\theta_0)\neq 0$. It leads to the existence of some $\epsilon_k\in\{-1,1\}$ satisfying

$$\big(B_j(\theta_0)-\epsilon_k B_j(\theta)\big)M_{\theta_0}^{t-1^{(.k)}}=0.$$

If $\epsilon_k=-1$, then $B_j(\theta_0)+B_j(\theta)$ is invertible as any definite positive matrix. We obtain a contradiction with $M_{\theta_0}^{t-1^{(.k)}}\neq 0$ and, thus, $\epsilon_k=1$. In other words, $(B_j(\theta_0)-B_j(\theta))M_{\theta_0}^{t-1^{(.k)}}=0$, and we conclude that $B_j(\theta_0)=B_j(\theta)$, for all $j>0$, as the family $\{M_{\theta_0}^{t-1^{(.k)}},1\leq k\leq p\}$ generates the whole space $\mathbb{R}^m$. Now, if $B_j(\theta_0)=0$, using the relation $a_{i,k,j}(\theta_0)^2=a_{i,k,j}(\theta)^2$, we derive that $B_j(\theta)=0$ as the family $\{M_{\theta_0}^{t-1^{(.k)}},1\leq k\leq p\}$ generates the whole space $\mathbb{R}^m$. Finally, in all cases, we conclude that $B_j(\theta_0)=B_j(\theta)$ for all $j\in\mathbb{N}$, and, if $\theta\in\Theta\mapsto B_j(\theta)$ is injective, then (Id(Θ)) is implied.

LEMMA 7. *Under assumptions of the part* AN *of Proposition* 4, *then* (Var) *holds.*

Simple calculations provide

$$\frac{\partial H_{\theta_0}^t}{\partial\theta_k}=\frac{\partial B_0(\theta_0)}{\partial\theta_k}+\sum_{j\geq 1}\bigg(\frac{\partial B_j(\theta_0)}{\partial\theta_k}X_{t-j}X'_{t-j}B_j(\theta_0)' + B_j(\theta_0)X_{t-j}X'_{t-j}\frac{\partial B_j(\theta_0)'}{\partial\theta_k}\bigg). \tag{5.17}$$

Let us now fix some d-uplet $(y_1,\dots,y_d)$ such that $\sum_{i=1}^d y_i\partial H_{\theta_0}^t/\partial\theta_i=0$. Using the expression (5.17) and the same arguments than in the previous proof of Lemma 6, we get that $\sum_{i=1}^d y_i\partial B_0(\theta_0)/\partial\theta_i=0$, and, for all $j\geq 1$ and $1\leq k\leq m$,

$$\sum_{i=1}^d y_i\bigg(\frac{\partial B_j(\theta_0)}{\partial\theta_i}M_{\theta_0}^{t-1^{(.k)}}\big(M_{\theta_0}^{t-1^{(.k)}}\big)'B_j(\theta_0)' + B_j(\theta_0)M_{\theta_0}^{t-1^{(.k)}}\big(M_{\theta_0}^{t-1^{(.k)}}\big)'\frac{\partial B_j(\theta_0)'}{\partial\theta_i}\bigg)=0. \tag{5.18}$$

Remark that, if $B_j(\theta_0)=0$, then $\frac{\partial B_j(\theta_0)}{\partial\theta_i}=0$ and, thus, necessarily $B_j(\theta_0)\neq 0$ for all $j\in\mathcal{S}$. Now, since the statement that if $B_j(\theta_0)\neq 0$, then $B_j(\theta_0)$ is invertible, we have that, for any $1\leq k\leq m$, there exists some $1\leq\ell\leq m$ such that $B_j^{(\ell.)}(\theta)M_{\theta_0}^{t-1^{(.k)}}\neq 0$. Then, as

$$\sum_{i=1}^d y_i\frac{\partial B_j(\theta_0)}{\partial\theta_i}^{(\ell.)}M_{\theta_0}^{t-1^{(.k)}}B_j(\theta_0)^{(\ell.)}M_{\theta_0}^{t-1^{(.k)}}=0,$$

we obtain that $\sum_{i=1}^{d} y_i \partial B_j(\theta_0)/\partial\theta_i{}^{(\ell.)} M_{\theta_0}^{t-1}{}^{(.k)} = 0$. Now, using this relation and (5.18) on the (ℓ, ℓ')th element for any ℓ', we easily obtain that $\sum_{i=1}^{d} y_i \frac{\partial B_j(\theta_0)}{\partial\theta_i}^{(\ell'.)} \times M_{\theta_0}^{t-1}{}^{(.k)} = 0$. Considering these results, we get the relation $\sum_{i=1}^{d} y_i \frac{\partial B_j(\theta_0)}{\partial\theta_i} \times M_{\theta_0}^{t-1}{}^{(.k)} = 0$ and, as the family $\{M_{\theta_0}^{t-1}{}^{(.k)}, 1 \le k \le p\}$ generates the whole space $\mathbb{R}^m$, we obtain that $\sum_{i=1}^{d} y_i \frac{\partial B_j(\theta_0)}{\partial\theta_i} = 0$ for all $j \in \mathcal{S}$. We conclude, necessarily, that $y_i = 0$ for all $1 \le i \le d$ and (Var) automatically follows.

Acknowledgments. We thank Thomas Mikosch, who helped us in the presentation of the drafts. We also thank the referee and the Associate Editor for comments that have been very useful in improving the content of this article.

REFERENCES

[1] BERKES, I., HORVÁTH, L. and KOKOSZKA, P. (2003). GARCH processes: Structure and estimation. *Bernoulli* **9** 201–227. MR1997027

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[3] BOUGEROL, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.* **31** 942–959. MR1227540

[4] BOUSSAMA, F. (1998). Ergodicity, mixing and estimation in GARCH models. Ph.D. thesis, Univ. de Paris 07, Paris.

[5] COMTE, F. and LIEBERMAN, O. (2003). Asymptotic theory for multivariate GARCH processes. *J. Multivariate Anal.* **84** 61–84. MR1965823

[6] DING, Z., GRANGER, C. W. and ENGLE, R. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance* **1** 83–106.

[7] DOUC, R., ROUEFF, F. and SOULIER, P. (2008). On the existence of some ARCH(∞) processes. *Stochastic Process. Appl.* **118** 755–761. MR2411519

[8] DOUKHAN, P., TEYSSIÈRE, G. and WINANT, P. (2006). A LARCH(∞) vector valued process. In *Dependence in Probability and Statistics* (P. Bertail, P. Doukhan and P. Soulier, eds.) 245–258. Springer, New York. MR2283258

[9] DOUKHAN, P. and WINTENBERGER, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* **118** 1997–2013. MR2462284

[10] DUFLO, M. (1997). *Random Iterative Models*. *Applications of Mathematics (New York)* **34**. Springer, Berlin. MR1485774

[11] ENGLE, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121

[12] ENGLE, R. and KRONER, K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11** 122–150. MR1325104

[13] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York. MR0210154

[14] FRANCQ, C. and ZAKOÏAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. *Bernoulli* **10** 605–637. MR2076065

[15] GIRAITIS, L., KOKOSZKA, P. and LEIPUS, R. (2000). Stationary arch models: Dependence structure and central limit theorem. *Econometric Theory* **16** 3–22. MR1749017

[16] JEANTHEAU, T. (1993). Modèles autorégressifs à erreur conditionnellement hétéroscédastique. Ph.D. thesis, Univ. Paris VII.

[17] JEANTHEAU, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory* **14** 70–86. MR1613694

[18] JEANTHEAU, T. (2004). A link between complete models with stochastic volatility and arch models. *Finance Stoch.* **8** 111–132. MR2022981

[19] KRENGEL, U. (1985). *Ergodic Theorems* **6**. De Gruyter, Berlin. MR0797411

[20] LING, S. and MCALEER, M. (2003). Asymptotic theory for a vector ARMA–GARCH model. *Econometric Theory* **19** 280–310. MR1966031

[21] PFANZAGL, J. (1969). On the mesurability and consistency of minimum contrast estimates. *Metrika* **14** 249–334.

[22] RABEMANANJARA, R. and ZAKOÏAN, J. (1993). Threshold ARCH models and asymmetries in volatility. *J. Appl. Econometrics* **8** 31–49.

[23] ROBINSON, P. (1991). Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression. *J. Econometrics* **47** 67–84. MR1087207

[24] ROBINSON, P. and ZAFFARONI, P. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist.* **34** 1049–1074. MR2278351

[25] STRAUMANN, D. and MIKOSCH, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* **34** 2449–2495. MR2291507

SAMOS (CENTRE D'ECONOMIE
DE LA SORBONNE)
UNIVERSITE PARIS 1
90 RUE DE TOLBIAC
75634 PARIS CEDEX 13
FRANCE
E-MAIL: Jean-Marc.Bardet@univ-paris1.fr
Olivier.Wintenberger@univ-paris1.fr
URL: http://samos.univ-paris1.fr/-Jean-Marc-Bardet
http://wintenberger.fr