Bernoulli **10**(4), 2004, 605–637

Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes

CHRISTIAN FRANCQ[1] and JEAN-MICHEL ZAKOÏAN[2]

[1]*Université Lille 3, GREMARS, BP 149, 59653 Villeneuve d'Ascq Cedex, France, E-mail: francq@univ-lille3.fr*
[2]*Université Lille 3, GREMARS and CREST, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France, E-mail: zakoian@ensae.fr*

We prove the strong consistency and asymptotic normality of the quasi-maximum likelihood estimator of the parameters of pure generalized autoregressive conditional heteroscedastic (GARCH) processes, and of autoregressive moving-average models with noise sequence driven by a GARCH model. Results are obtained under mild conditions.

Keywords: ARMA; asymptotic normality; consistency; GARCH; heteroskedastic time series; maximum likelihood estimation

1. Introduction

Since the seminal papers by Engle (1982) and Bollerslev (1986), generalized autoregressive conditional heteroscedastic (GARCH) processes have received considerable attention in the literature devoted to the analysis of financial time series. These time series models capture several important features of financial series, such as leptokurticity and volatility clustering – see Mikosch (2001) for a recent paper on GARCH and stochastic volatility models.

The asymptotic properties of the quasi-maximum likelihood estimator (QMLE) were first established by Weiss (1986) for ARCH models, under fourth-order moment conditions on the ARCH process. Unfortunately, these conditions are typically violated when GARCH models are estimated on financial data. The problem of finding weak assumptions for the consistency and asymptotic normality of the QMLE in GARCH models has attracted much attention in the statistical literature. To our knowledge, the most significant contributions on the theoretical properties of the QMLE in GARCH models are those of Lee and Hansen (1994) and Lumsdaine (1996), both for the GARCH(1, 1) case, Straumann and Mikosch (2003) for a general heteroscedastic model including GARCH(1, 1), and Boussama (1998; 2000), Berkes and Horváth (2003a; 2003b) and Berkes *et al*. (2003) for general GARCH (p, q). The latter reference gives rigorous proofs of the strong consistency and asymptotic normality, under assumptions which we discuss in Section 2.

The first goal of the present paper is to establish, under weaker conditions than those in the existing literature, the convergence and asymptotic normality of the QMLE for the GARCH(p, q) process defined in (2.1) below. We will provide asymptotic results requiring strict stationarity but no moment assumption. An alternative method of proof allows us to weaken some of the technical assumptions used in the above references.

Our second goal is to extend these asymptotic results to ARMA-GARCH processes. In financial applications, it is common practice to fit return series by autoregressive moving-average (ARMA) models with GARCH innovations. It is therefore of interest to analyse the properties of QMLEs of ARMA-GARCH processes. As we will see, the extension leads to non-trivial problems. Recent works on the estimation of ARMA-GARCH processes are Ling and Li (1997; 1998) and Ling and McAleer (2003a). Comments on these papers are provided in Section 3. See also Ling and Li (2003) and Ling and McAleer (2003b) for related work.

The paper is organized as follows. Section 2 presents the assumptions for the GARCH model and states our results on this class. Section 3 is devoted to the ARMA-GARCH class. The proofs are postponed to Section 4.

The following notation will be used throughout. The norm of a matrix $A = (a_{ij})$ is defined by $\|A\| = \sum |a_{ij}|$. The spectral radius of a sqaure matrix A is denoted by $\rho(A)$. The Kronecker product is denoted by $\otimes$. The symbol $\Rightarrow$ denotes convergence in distribution.

2. The pure GARCH(p, q) case

Consider the GARCH(p, q) model

$$\begin{aligned} \epsilon_t &= \sqrt{h_t}\eta_t, \\ h_t &= \omega_0 + \sum_{i=1}^{q} \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}, \qquad \forall t \in \mathbb{Z}, \end{aligned} \tag{2.1}$$

where (η_t) is a sequence of independent and identically distributed (i.i.d.) random variables such that $E\eta_t^2 = 1$, $\omega_0 > 0$, $\alpha_{0i} \geqslant 0$ ($i = 1, \ldots, q$) and $\beta_{0j} \geqslant 0$ ($j = 1, \ldots, p$).

Bougerol and Picard (1992) showed that a unique non-anticipative strictly stationary solution (ϵ_t) to model (2.1) exists if and only if the sequence of matrices $\mathbf{A}_0 = (A_{0t})$, where

$$A_{0t} = \begin{pmatrix} \alpha_{01}\eta_t^2 & & \cdots & & \alpha_{0q}\eta_t^2 & \beta_{01}\eta_t^2 & & \cdots & & \beta_{0p}\eta_t^2 \\ 1 & 0 & \cdots & & 0 & 0 & & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 & 0 & & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 & 0 & & \cdots & 0 & 0 \\ \alpha_{01} & & \cdots & & \alpha_{0q} & \beta_{01} & & \cdots & & \beta_{0p} \\ 0 & & \cdots & & 0 & 1 & 0 & \cdots & & 0 \\ 0 & & \cdots & & 0 & 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 0 & 0 & 0 & & \cdots & 1 & 0 \end{pmatrix},$$

has a strictly negative top Lyapunov exponent, $\gamma(\mathbf{A}_0) < 0$, where

$$\gamma(\mathbf{A}_0) := \inf_{t \in \mathbb{N}^*} \frac{1}{t} \mathrm{E}(\log \|A_{0t} A_{0,t-1} \ldots A_{01}\|) = \lim_{t \to \infty} \frac{1}{t} \log \|A_{0t} A_{0,t-1} \ldots A_{01}\| \text{ a.s.} \tag{2.2}$$

The definition of $\gamma(\mathbf{A}_0)$ does not depend on the choice of a norm on the space of the $(p+q) \times (p+q)$ matrices. The second equality in (2.2) is a consequence of the subadditive ergodic theorem (Kingman, 1973). Note that the existence of $\gamma(\mathbf{A}_0)$ is guaranteed by the inequality $\mathrm{E}(\log^+ \|A_{01}\|) \leqslant \mathrm{E}\|A_{01}\| < \infty$.

Let $\underline{z}_t = (\epsilon_t^2, \ldots, \epsilon_{t-q+1}^2, h_t, \ldots, h_{t-p+1})^{\mathrm{T}} \in \mathbb{R}^{p+q}$ and $\underline{b}_t = (\omega, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{p+q}$. Then (2.1) is equivalently written as a vector stochastic recurrence equation

$$\underline{z}_t = b_t + A_{0t} \underline{z}_{t-1}, \tag{2.3}$$

and if $\gamma(A_0) < 0$, the unique strictly stationary solution to (2.3) is

$$\underline{z}_t = \underline{b}_t + \sum_{k=1}^{\infty} A_{0t} A_{0,t-1} \ldots A_{0,t-k+1} \underline{b}_{t-k}. \tag{2.4}$$

In view of (2.2), and using the Jensen inequality, it is clear that the conditions

$$\mathrm{E}(\log \|A_{0k} A_{0,k-1} \ldots A_{01}\|) < 0, \qquad \text{for some } k > 0,$$

and

$$\rho(\mathrm{E}A_{01}) < 1, \qquad \text{i.e. } \sum_{i=1}^{q} \alpha_{0i} + \sum_{i=1}^{p} \beta_{0i} < 1, \tag{2.5}$$

imply $\gamma(\mathbf{A}_0) < 0$. Note, however, that the sufficient condition (2.5) is much stronger than the strict stationarity condition $\gamma(\mathbf{A}_0) < 0$ ((2.5) implies $\mathrm{E}\epsilon_t^2 < \infty$).

Two well-known consequences of the strict stationarity condition are stated in the following proposition. We refer to Bougerol and Picard (1992) for the proof of the first result, and to Nelson (1990) and Berkes *et al*. (2003, Lemma 2.3) for the second.

Proposition 1. *If* $\gamma(\mathbf{A}_0) < 0$*, then the following equivalent conditions hold:*

(a) $\sum_{j=1}^{p} \beta_{0j} < 1$.
(b) *The roots of the polynomial* $1 - \beta_{01} z - \ldots - \beta_{0p} z^p$ *are outside the unit disc.*
(c) $\rho(B_0) < 1$*, where*

$$B_0 = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & & \beta_{0p} \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 \end{pmatrix}.$$

Moreover, if $\gamma(\mathbf{A}_0) < 0$*, then there exists* $s > 0$ *such that*

$$\mathrm{E}h_t^s < \infty \quad \textit{and} \quad \mathrm{E}\epsilon_t^{2s} < \infty.$$

We now turn to the QML estimation of model (2.1). The vector of parameters is

$$\theta = (\theta_1, \ldots, \theta_{p+q+1})^{\mathrm{T}} = (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$$

and belongs to a parameter space $\Theta \subset]0, +\infty[\times[0, \infty[^{p+q}$. The true parameter value is unknown and is denoted by $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})^{\mathrm{T}}$.

Let $(\epsilon_1, \ldots, \epsilon_n)$ be a realization of length n of the unique non-anticipative strictly stationary solution (ϵ_t) to model (2.1). Conditionally on initial values $\epsilon_0, \ldots, \epsilon_{1-q}$, $\tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-p}^2$, the Gaussian quasi-likelihood is given by

$$L_n(\theta) = L_n(\theta; \epsilon_1, \ldots, \epsilon_n) = \sum_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2}\right),$$

where the $\tilde{\sigma}_t^2$ are defined recursively, for $t \geqslant 1$, by

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2.$$

For instance, the initial values can be chosen as

$$\epsilon_0^2 = \ldots = \epsilon_{1-q}^2 = \tilde{\sigma}_0^2 = \ldots = \tilde{\sigma}_{1-p}^2 = \omega \tag{2.6}$$

or

$$\epsilon_0^2 = \ldots = \epsilon_{1-q}^2 = \tilde{\sigma}_0^2 = \ldots = \tilde{\sigma}_{1-p}^2 = \epsilon_1^2. \tag{2.7}$$

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

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\min_{\theta\in\Theta} \tilde{\mathbf{l}}_n(\theta). \tag{2.8}$$

where

$$\tilde{\mathbf{l}}_n(\theta) = n^{-1} \sum_{t=1}^{n} \tilde{\ell}_t \quad \text{and} \quad \tilde{\ell}_t = \tilde{\ell}_t(\theta) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log \tilde{\sigma}_t^2.$$

Remark 1.1. It will be shown that the choice of initial values does not matter for the asymptotic properties of the QMLE. However, it may be important from a practical point of view. Other ways of generating the sequence $\tilde{\sigma}_t^2$ have been considered in the literature; for instance by taking $\tilde{\sigma}_t^2 = c_0(\theta) + \sum_{i=1}^{t-1} c_i(\theta)\epsilon_{t-i}^2$, where the $c_i(\theta)$ are computed recursively (see Berkes *et al*. 2003). Note that, to compute $\tilde{\mathbf{l}}_n(\theta)$, their procedure requires a number of operations of order n^2. The number of operations required by our procedure is of order n.

Let $\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$. By convention, $\mathcal{A}_\theta(z) = 0$ if $q = 0$ and $\mathcal{B}_\theta(z) = 1$ if $p = 0$. To show the strong consistency, the following assumptions will be made.

(A1) $\theta_0 \in \Theta$ and Θ is compact.
(A2) $\gamma(\mathbf{A}_0) < 0$ and $\forall\theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

(A3) η_t^2 has a non-degenerate distribution with $E\eta_t^2 = 1$.
(A4) If $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common root, $\mathcal{A}_{\theta_0}(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.

It will be convenient to approximate the sequence $(\tilde{\ell}_t(\theta))$ by an ergodic stationary sequence. In the first part of Proposition 1, equivalence evidently holds for any $\theta \in \Theta$. Thus (A2) implies that the roots of $\mathcal{B}_\theta(z)$ are outside the unit disc. Therefore, denote by $(\sigma_t^2) = \{\sigma_t^2(\theta)\}$ the strictly stationary, ergodic and non-anticipative solution of

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \qquad \forall t. \tag{2.9}$$

Note that $\sigma_t^2(\theta_0) = h_t$. Let

$$\mathbf{l}_n(\theta) = \mathbf{l}_n(\theta; \epsilon_n, \epsilon_{n-1} \ldots,) = n^{-1} \sum_{t=1}^{n} \ell_t, \qquad \ell_t = \ell_t(\theta) = \frac{\epsilon_t^2}{\sigma_t^2} + \log \sigma_t^2.$$

We are now in a position to state our first result.

Theorem 2.1. *Let* $(\hat{\theta}_n)$ *be a sequence of QML estimators satisfying* (2.8)*, with the initial conditions* (2.6) *or* (2.7)*. Then, under* (A1)–(A4)*, almost surely* $\hat{\theta}_n \to \theta_0$*, as* $n \to \infty$.

Remark 2.1. Unlike Berkes *et al*. (2003), our assumptions on the i.i.d. process η_t do not impose the existence of $E(\eta_t^{2+\epsilon})$, for some $\epsilon > 0$, or any other technical assumption on η_t, such as those requiring that the cdf around zero is well behaved. In fact, their proof requires $E_{\theta_0} \sup_{\theta \in \Theta} |\ell_t(\theta)| < \infty$ (see their Lemma 5.1). Using an ergodic theorem for stationary processes (X_t) such that $EX_1 \in \mathbb{R} \cup \{+\infty\}$, our proof only requires $E_{\theta_0} |\ell_t(\theta_0)| < \infty$. Moreover, in Berkes *et al*. (2003), the parameter space is very constrained, ruling out zero coefficients in θ.

Remark 2.2. Straumann and Mikosch (2003) established asymptotic results for a general heteroscedastic time series model. When applied to the GARCH(1, 1) model, their consistency result coincides with ours. A slight difference is that they assume that the distribution of η_t is not concentrated at two points, whereas we assume that η_t is not concentrated at ± 1.

Remark 2.3. Lee and Hansen (1994) and Lumsdaine (1996) established asymptotic results for the GARCH(1, 1) model. In Lee and Hansen (1994) the η_t are required to form a strictly stationary martingale difference sequence. However, their QMLE is local, that is, $L_n(\theta)$ is maximized in a neighbourhood of θ_0. Moreover, the existence of $E(\eta_t^{2+\epsilon})$ is assumed for some $\epsilon > 0$.

Remark 2.4. The first part of the identifiability assumption (A4) concerning the common roots of the polynomials was also made by Berkes *et al*. (2003). It is worth noting that (A4) implies that θ_0 need not belong to the interior of Θ. This is essential, in particular, to handle situations of overidentification. For instance, our result shows that an ARCH(q) model ($\beta_{0j} = 0$, for all j) is consistently estimated when a GARCH(p, q) is fitted. More generally, one of the two orders p and q can be overidentified, but not both of them. Evidently, it is required that $\alpha_{0i} > 0$ for some i when $p > 0$. Without this assumption, the model solution would be an i.i.d. white noise, which could be represented as any GARCH(1, 0) process of the form $\sigma_t^2 = \sigma^2(1-\beta) + 0 \times \epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. Note also that the first part of (A4) is always satisfied when $p > 1$ and $q > 1$. If $q = 1$, the unique root of $\mathcal{A}_{\theta_0}(z)$ is 0 and $\mathcal{B}_{\theta_0}(0) \neq 0$. If $p = 1$ and $\beta_{01} \neq 0$, the unique root of $\mathcal{B}_{\theta_0}(z)$ is $1/\beta_{01} > 0$ (if $\beta_{01} = 0$, the polynomial does not have any zero), and, because the α_{0i} are positive, $\mathcal{A}_{\theta_0}(1/\beta_{01}) \neq 0$.

Remark 2.5. Following the suggestion made by a referee, we have not imposed $\mathrm{E}\eta_t = 0$. The conditional variance of ϵ_t given $\{\epsilon_{t-i}, i > 0\}$ is only proportional to h_t in this case. The assumption that $\mathrm{E}\eta_t^2 = 1$ is made for identifiability reasons and is not restrictive provided $\mathrm{E}\eta_t^2 < \infty$. Berkes and Horváth (2003b) showed that this moment condition is necessary for the asymptotic normality of the Gaussian QMLE, and Berkes and Horváth (2003a) showed that the moment condition can be weakened when criteria different from the Gaussian quasi-likelihood are used.

To show the asymptotic normality, the following additional assumptions are made.

(A5) $\theta_0 \in \mathring{\Theta}$, where $\mathring{\Theta}$ denotes the interior of Θ.
(A6) $\kappa_\eta := \mathrm{E}\eta_t^4 < \infty$.

The second main result of this section is the following.

Theorem 2.2. *Under assumptions* (A1)–(A6)*,* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is asymptotically distributed as* $\mathcal{N}(0, (\kappa_\eta - 1)J^{-1})$*, where*

$$J := \mathrm{E}_{\theta_0}\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right) = \mathrm{E}_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta^{\mathrm{T}}}\right). \tag{2.10}$$

Remark 2.6. We show in Section 4 the existence and positive definiteness of J.

Remark 2.7. Assumption (A5) is clearly necessary to obtain asymptotic normality. For instance, when $\alpha_{01} = 0$, the distribution of $\sqrt{n}(\hat{\alpha}_1 - \alpha_{01})$ is concentrated on $[0, \infty[$, so the asymptotic distribution cannot be normal. Andrews (1999) studied such boundary problems in the GARCH(1, q) case.

Remark 2.8. As in Theorem 2.1, no technical assumption on the distribution of η_t is required, apart from the existence of a fourth-order moment. This assumption is clearly necessary for the existence of the variance of the score vector $\partial\ell_t(\theta_0)/\partial\theta$. Note also that this assumption does not imply the existence of a second-order moment for the observed process (ϵ_t). This is particularly interesting for financial applications, because such existence of the second-order moments is often found to be inappropriate.

Remark 2.9. In Berkes *et al*. (2003) it is assumed that $E|\eta_t|^{4+\delta} < \infty$ for some $\delta > 0$, and $t^{-\mu}P(\eta_t^2 \leqslant t) \to 0$ when $t \to 0$, for some $\mu > 0$. These assumptions are used to treat the right-hand terms of the inequality $\sum_{i=1}^{\infty} i^3 c_i \epsilon_{t-i}^2 (1 + \sum_{i=1}^{\infty} c_i \epsilon_{t-i}^2)^{-1} \leqslant M^3 + \sum_{M<i} i^3 c_i \epsilon_{t-i}^2$, for some absolutely summable positive sequence (c_i) and any $M \geqslant 1$ (see their Lemma 5.2). Instead, we use Proposition 1 and the inequality $\sum_{i=1}^{\infty} i^3 c_i \epsilon_{t-i}^2 (1 + \sum_{i=1}^{\infty} c_i \epsilon_{t-i}^2)^{-1} \leqslant \sum_{i=1}^{\infty} i^3 c_i^s \epsilon_{t-i}^{2s}$ for all $s \in]0, 1[$. The idea of exploiting the inequality $x/(1+x) \leqslant x^s$ for all $x > 0$ is due to Boussama (2000).

Remark 2.10. In Boussama (2000) it is assumed that $E\eta_t^6 < \infty$. The parameter space is supposed to be a hypercube of the form $[\underline{\omega}, \overline{\omega}] \times \prod_{i=1}^{q} [\underline{\alpha}_i, \overline{\alpha}_i] \times \prod_{j=1}^{p} [\delta\bar{\beta}_j, \bar{\beta}_j]$ with $\sum_{j=1}^{q} \bar{\beta}_j < 1$, which seems very restrictive. Moreover, it is not clear whether his results allow pure ARCH models to be treated, because an implicit assumption in his paper is that both α_{0q} and β_{0p} are non-zero.

3. Estimation of ARMA-GARCH models

In this section our aim is to extend the previous results to the case where the GARCH process is not directly observed. The process (ϵ_t), the solution to (2.1), is a martingale difference and can therefore be used as the innovation of an ARMA process. Even for financial series, it seems very restrictive to assume that the observed process is a pure GARCH. Allowing for an ARMA part considerably extends the range of applications, but it also entails serious technical difficulties.

The observations are now denoted $X_1, \ldots, X_n$ and are obtained from an ARMA(P, Q)-GARCH(p, q) process (X_t) satisfying

$$\begin{aligned} X_t - c_0 &= \sum_{i=1}^{P} a_{0i}(X_{t-i} - c_0) + e_t - \sum_{j=1}^{Q} b_{0j} e_{t-j}, \\ e_t &= \sqrt{h_t}\eta_t, \\ h_t &= \omega_0 + \sum_{i=1}^{q} \alpha_{0i} e_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}, \end{aligned} \tag{3.1}$$

where (η_t) and the coefficients ω_0, α_{0i} and β_{0j} are defined as in (2.1). The parameter vector is denoted $\varphi = (\vartheta^{\mathrm{T}}, \theta^{\mathrm{T}})^{\mathrm{T}} = (c, a_1, \ldots a_P, b_1, \ldots, b_Q, \theta^{\mathrm{T}})^{\mathrm{T}}$, the true value is $\varphi_0 = (\vartheta_0^{\mathrm{T}}, \theta_0^{\mathrm{T}})^{\mathrm{T}} = (c_0, a_{01}, \ldots a_{0P}, b_{01}, \ldots, b_{0Q}, \theta_0^{\mathrm{T}})^{\mathrm{T}}$, and the parameter space is $\Phi \subset \mathbb{R}^{P+Q+1} \times]0, +\infty[\times [0, \infty[^{p+q}$.

If $q \geqslant Q$, the initial values $X_0 \ldots, X_{1-(q-Q)-P}, \tilde{\epsilon}_{-q+Q}, \ldots, \tilde{\epsilon}_{1-q}, \tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-p}^2$ allow us to compute $\tilde{\epsilon}_t(\vartheta)$, for $t = -q + Q + 1, \ldots, n$, and $\tilde{\sigma}_t^2(\varphi)$, for $t = 1, \ldots, n$, from

$$\tilde{\epsilon}_t = \tilde{\epsilon}_t(\vartheta) = X_t - c - \sum_{i=1}^{P} \alpha_i (X_{t-i} - c) + \sum_{j=1}^{Q} b_j \tilde{\epsilon}_{t-j},$$

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\varphi) = \omega + \sum_{i=1}^{q} \alpha_i \tilde{\epsilon}_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2.$$

When $q < Q$, the required initial values are $X_0, \dots, X_{1-(q-Q)-P}$, $\tilde{\epsilon}_0, \dots, \tilde{\epsilon}_{1-Q}$, $\tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{1-p}^2$. For simplicity, these initial values will be taken to be fixed (neither random nor functions of the parameters).

A QMLE of φ is any measurable solution of

$$\hat{\varphi}_n = \arg\min_{\varphi \in \Phi} \tilde{\mathbf{l}}_n(\varphi), \tag{3.2}$$

where $\tilde{\mathbf{l}}_n(\varphi) = n^{-1} \sum_{t=1}^{n} \tilde{\ell}_t$ and $\tilde{\ell}_t = \tilde{\ell}_t(\varphi) = \tilde{\epsilon}_t^2(\vartheta)/\tilde{\sigma}_t^2(\varphi) + \log \tilde{\sigma}_t^2(\varphi)$.

Let $A_\vartheta(z) = 1 - \sum_{i=1}^{P} a_i z^i$ and $B_\vartheta(z) = 1 - \sum_{j=1}^{Q} b_j z^j$. We make standard assumptions on these autoregressive and moving-average polynomials, and we adapt assumption (A1) as follows:

(A7) $\varphi_0 \in \Phi$ and Φ is compact.
(A8) For all $\varphi \in \Phi$, $A_\vartheta(z) B_\vartheta(z) = 0$ implies $|z| > 1$.
(A9) $A_{\vartheta_0}(z)$ and $B_{\vartheta_0}(z)$ have no common root, $a_{0P} \neq 0$ or $b_{0Q} \neq 0$.

Under assumptions (A2) and (A8), (X_t) is assumed to be the unique non-anticipative strictly stationary solution to (3.1). Let $\epsilon_t = \epsilon_t(\vartheta) = A_\vartheta(L) B_\vartheta^{-1}(L)(X_t - c)$, where L denotes the lag operator, and let $\ell_t = \ell_t(\varphi) = \epsilon_t^2/\sigma_t^2 + \log \sigma_t^2$, where $\sigma_t^2 = \sigma_t^2(\varphi)$ is the strictly stationary, ergodic and non-anticipative solution of (2.9). Note that $e_t = \epsilon_t(\vartheta_0)$ and $h_t = \sigma_t^2(\varphi_0)$. The following result extends Theorem 2.1.

Theorem 3.1. *Let* $(\hat{\varphi}_n)$ *be a sequence of QMLEs satisfying* (3.2). *Assume that* $\mathrm{E}\eta_t = 0$. *Then, under* (A2)–(A4) *and* (A7)–(A9), $\hat{\varphi}_n \to \varphi_0$, *almost surely, as* $n \to \infty$.

Remark 3.1. Ling and Li (1997; 1998) announced theoretical results for the MLE and QMLE of unstable and fractionally integrated ARMA models with GARCH innovations. However, they only obtained results for local estimators, that is, for sequences of solutions to the likelihood equation.

Remark 3.2. Ling and MacAleer (2003a) considered QMLEs for vector ARMA-GARCH models. Their consistency result requires the existence of a second-order moment.

Remark 3.3. (A9) is an identifiability assumption. In the literature on ARMA estimation, the assumption that $a_{0P} \neq 0$ and $b_{0Q} \neq 0$ is often made. This excludes interesting situations where, for instance, and AR(1) model is fitted to a white noise.

Remark 3.4. As in the pure GARCH case, the process ϵ_t (and hence X_t) need not have finite variance. In the pure ARMA case, where $\epsilon_t = \eta_t$ has finite variance, our theorem reduces to a classical result on ARMA models based on i.i.d. innovations (see Brockwell and Davis, 1991, p. 384). For ARMA processes with i.i.d. infinite-variance innovations, the asymptotic distribution of the QMLE is not standard – see Mikosch *et al*. (1995) and Kokoszka and Taqqu (1996).

Remark 3.5. Apart from the condition $\mathrm{E}\eta_t = 0$, the assumptions required for the strong consistency are not strengthened when an ARMA part is added. One might wonder whether the normality Theorem 2.2 also extends without cost in terms of assumptions. Unfortunately, the answer is negative, as the following example reveals.

Consider the AR(1)-ARCH(1) model

$$X_t = a_{01}X_{t-1} + e_t, \qquad e_t = \sqrt{h_t}\eta_t, \qquad h_t = \omega_0 + \alpha_{01}e_{t-1}^2 \tag{3.3}$$

where $|a_{01}| < 1$, $\omega_0 > 0$, $\alpha_{01} \geqslant 0$, and (η_t) is an i.i.d. sequence such that, for some $a > 1$,

$$P(\eta_t = a) = P(\eta_t = -a) = \frac{1}{2a^2}, \qquad P(\eta_t = 0) = 1 - \frac{1}{a^2}.$$

It can easily be seen that for an ARCH(1) process, the strict stationarity condition is $\alpha_{01} < \exp\{-\mathrm{E}(\log \eta_t^2)\}$. For any α_{01}, the process (X_t) is therefore strictly stationary, since $\exp\{-\mathrm{E}(\log \eta_t^2)\} = +\infty$. However, X_t does not have a second-order moment, whence $\alpha_{01} \geqslant 1$. The first component of the (normalized) score vector is

$$\frac{\partial \ell_t(\theta_0)}{\partial a_1} = \left(1 - \frac{e_t^2}{h_t}\right)\left(\frac{1}{h_t}\frac{\partial \sigma_t^2(\theta_0)}{\partial a_1}\right) + \frac{2e_t}{h_t}\frac{\partial \epsilon_t(\theta_0)}{\partial a_1} = -2\alpha_{01}(1 - \eta_t^2)\left(\frac{e_{t-1}X_{t-2}}{h_t}\right) - 2\frac{\eta_t X_{t-1}}{\sqrt{h_t}}.$$

We have

$$\begin{aligned}
&\mathrm{E}\left\{\alpha_{01}(1 - \eta_t^2)\left(\frac{e_{t-1}X_{t-2}}{h_t}\right) + \frac{\eta_t X_{t-1}}{\sqrt{h_t}}\right\}^2 \\
&\qquad \geqslant \mathrm{E}\left[\left\{\alpha_{01}(1 - \eta_t^2)\left(\frac{e_{t-1}X_{t-2}}{h_t}\right) + \frac{\eta_t X_{t-1}}{\sqrt{h_t}}\right\}^2 \middle| \eta_{t-1} = 0\right] P(\eta_{t-1} = 0) \\
&\qquad = \frac{a_{01}^2}{\omega_0}\left(1 - \frac{1}{a^2}\right) E(X_{t-2}^2)
\end{aligned}$$

because $\eta_{t-1} = 0$ implies $e_{t-1} = 0$ and $X_{t-1} = a_{01}X_{t-2}$, and because η_t, η_{t-1} and X_{t-2} are independent. Therefore, if $\mathrm{E}X_t^2 = \infty$ and $a_{01} \neq 0$ the variance of the score vector is not defined. In Theorem 2.2, the asymptotic variance of the estimator of the pure GARCH parameter is proportional to the (finite) variance of the score vector (see Remark 2.8). This example shows that Theorem 2.2 does not extend to the ARMA-GARCH class. This is not very surprising since the asymptotic normality of the estimators of pure ARMA models with i.i.d. innovations (which belong to our general class) are obtained under second-order moment assumptions (see Brockwell and Davis 1991). For ARMA models with infinite-variance noise, the rate of convergence is faster than in the standard case and asymptotic stable laws are obtained (see Davis *et al*. 1992; Mikosch *et al*. 1995).

We establish asymptotic normality under a fourth-order moment assumption. Chen and An (1998) showed that there exists a non-anticipative and strictly stationary solution of (2.1) with finite fourth-order moment if and only if $\rho\{\mathrm{E}(A_{0t} \otimes A_{0t})\} > 1$. We assume that

(A10) $\rho\{\mathrm{E}(A_{0t} \otimes A_{0t})\} < 1$, and $\forall\theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

This assumption implies that $\kappa_\eta = \mathrm{E}(\eta_t^4) < \infty$. Also, (A2) becomes redundant. Analogously to the pure GARCH case, we assume that

(A11) $\varphi_0 \in \mathring{\Phi}$, where $\mathring{\Phi}$ denotes the interior of Φ.

For identifiability reasons we also make the following assumption, which is slightly stronger than the first part of (A3) when η_t has a non-symmetric distribution.

(A12) There exists no set Λ of cardinality 2 such that $P(\eta_t \in \Lambda) = 1$.

Theorem 3.2. *Assume that* $\mathrm{E}\eta_t = 0$. *Under assumptions* (A3)–(A4) *and* (A8)–(A12), $\sqrt{n}(\hat{\varphi}_n - \varphi_0)$ *is asymptotically distributed as* $\mathcal{N}(0, \Sigma)$, *where* $\Sigma = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$, with

$$\mathcal{I} = \mathrm{E}\varphi_0\left(\frac{\partial \ell_t(\varphi_0)}{\partial \varphi}\frac{\partial \ell_t(\varphi_0)}{\partial \varphi^{\mathrm{T}}}\right), \qquad \mathcal{J} = \mathrm{E}\varphi_0\left(\frac{\partial^2 \ell_t(\varphi_0)}{\partial \varphi \partial \varphi^{\mathrm{T}}}\right).$$

If, in addition, the distribution of η_t *is symmetric, we have*

$$\mathcal{I} = \begin{pmatrix} I_1 & 0 \\ 0 & I_2 \end{pmatrix}, \qquad \mathcal{J} = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix},$$

with

$$I_1 = (\kappa_\eta - 1)\mathrm{E}\varphi_0\left(\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \vartheta}\frac{\partial \sigma_t^2}{\partial \vartheta^{\mathrm{T}}}(\varphi_0)\right) + 4\mathrm{E}\varphi_0\left(\frac{1}{\sigma_t^2}\frac{\partial \epsilon_t}{\partial \vartheta}\frac{\partial \epsilon_t}{\partial \vartheta^{\mathrm{T}}}(\varphi_0)\right),$$

$$I_2 = (\kappa_\eta - 1)\mathrm{E}\varphi_0\left(\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta^{\mathrm{T}}}(\varphi_0)\right),$$

$$J_1 = \mathrm{E}\varphi_0\left(\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \vartheta}\frac{\partial \sigma_t^2}{\partial \vartheta^{\mathrm{T}}}(\varphi_0)\right) + \mathrm{E}\varphi_0\left(\frac{2}{\sigma_t^2}\frac{\partial \epsilon_t}{\partial \vartheta}\frac{\partial \epsilon_t}{\partial \vartheta^{\mathrm{T}}}(\vartheta_0)\right),$$

$$J_2 = \mathrm{E}\varphi_0\left(\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta^{\mathrm{T}}}(\varphi_0)\right).$$

Remark 3.6. When applied to the ARMA-GARCH case, the asymptotic normality result given by Ling and MacAleer (2003a) requires the existence of sixth-order moments. Moreover, the stationarity conditions are imposed over the whole parameter space.

Remark 3.7. In the proof of the theorem, we show the existence and positive definiteness of the matrices $\mathcal{I}$ and $\mathcal{J}$. Notice that when η_t has a symmetric distribution, Σ is block-diagonal, which is important in the testing of joint assumptions on ARMA and GARCH coefficients. In addition, the bottom right-hand block $J_2^{-1} I_2 J_2^{-1}$ of Σ depends on the GARCH coefficients only. In other words, the asymptotic accuracy of the GARCH estimators is not affected by the presence of an ARMA part.

Remark 3.8. It can easily be seen that assumption (A11) constrains only the GARCH coefficients. For any value of ϑ_0, the restriction of Φ to its first $P+Q+1$ components can be chosen sufficiently large that its interior contains ϑ_0 and assumption (A8) is not violated. Assumption (A11), however, requires the GARCH coefficients to be strictly positive.

4. Proofs

Let K and ρ be generic constants taking many different values $K > 0$ and $0 < \rho < 1$ throughout the proofs. For instance, we will allow ourselves to write, for $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$, $i_1 \geqslant 0$, $i_2 \geqslant 0$,

$$0 < K \sum_{i \geqslant i_1} \rho_1^i + K \sum_{i \geqslant i_2} i \rho_2^i \leqslant K \rho^{\min(i_1, i_2)}.$$

4.1. Proof of Theorem 2.1

Rewrite (2.9) in vector form as

$$\underline{\sigma}_t^2 = \underline{c}_t + B \underline{\sigma}_{t-1}^2, \tag{4.1}$$

where

$$\underline{\sigma}_t^2 = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \vdots \\ \sigma_{t-p+1}^2 \end{pmatrix}, \quad \underline{c}_t = \begin{pmatrix} \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{4.2}$$

We will establish the following intermediate results:

(i) $\lim_{n\to\infty} \sup_{\theta\in\Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| = 0$ a.s.
(ii) $(\exists t \in \mathbb{Z}$ such that $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$ P_{θ_0} a.s.$) \Rightarrow \theta = \theta_0$.
(iii) $\mathrm{E}_{\theta_0} |\ell_t(\theta_0)| < \infty$, and if $\theta \neq \theta_0$, $\mathrm{E}_{\theta_0} \ell_t(\theta) > \mathrm{E}_{\theta_0} \ell_t(\theta_0)$.
(iv) Any $\theta \neq \theta_0$ has a neighbourhood $V(\theta)$ such that $\liminf_{n\to\infty} \inf_{\theta^* \in V(\theta)} \tilde{\mathbf{l}}_n(\theta^*) > \mathrm{E}_{\theta_0} \ell_1(\theta_0)$ a.s.

To prove (i) first note that, using Proposition 1 and the compactness of Θ,

$$\sup_{\theta\in\Theta} \rho(B) < 1. \tag{4.3}$$

Hence, iterating (4.1), we obtain

$$\underline{\sigma}_t^2 = \underline{c}_t + B\underline{c}_{t-1} + B^2\underline{c}_{t-2} + \ldots + B^{t-1}\underline{c}_1 + B^t\underline{\sigma}_0^2 = \sum_{k=0}^{\infty} B^k \underline{c}_{t-k}. \tag{4.4}$$

Let $\underline{\tilde{\sigma}}_t^2$ be the vector obtained by replacing σ_{t-i}^2 by $\tilde{\sigma}_{t-i}^2$ in $\underline{\sigma}_{t-i}^2$. Let $\underline{\tilde{c}}$ be the vector obtained by replacing $\epsilon_0^2, \ldots, \epsilon_{1-q}^2$ by the initial values (2.6) or (2.7). We have

$$\underline{\tilde{\sigma}}_t^2 = \underline{c}_t + B\underline{c}_{t\text{-}1} + \ldots + B^{t-q-1}\underline{c}_{q+1} + B^{t-q}\underline{\tilde{c}}_q + \ldots + B^{t-1}\underline{\tilde{c}}_1 + B^t\underline{\tilde{\sigma}}_0^2. \tag{4.5}$$

In view of (4.3)–(4.5), almost surely,

$$\begin{aligned}\sup_{\theta\in\Theta}\|\sigma_t^2 - \underline{\tilde{\sigma}}_t^2\| &= \sup_{\theta\in\Theta}\left\|\left\{\sum_{k=1}^{q} B^{t-k}(\underline{c}_k - \underline{\tilde{c}}_k) + B^t(\underline{\sigma}_0^2 - \underline{\tilde{\sigma}}_0^2\right\}\right\| \\ &\leqslant K\rho^t, \qquad \forall t.\end{aligned} \tag{4.6}$$

Thus, using $\log x \leqslant x - 1$, almost surely,

$$\begin{aligned}\sup_{\theta\in\Theta}|\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| &\leqslant n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left\{\left|\frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\tilde{\sigma}_t^2\sigma_t^2}\right|\epsilon_t^2 + \left|\log\left(1 + \frac{\sigma_t^2 - \tilde{\sigma}_t^2}{\tilde{\sigma}_t^2}\right)\right|\right\} \\ &\leqslant \left\{\sup_{\theta\in\Theta}\frac{1}{\omega^2}\right\}Kn^{-1}\sum_{t=1}^{n}\rho^t\epsilon_t^2 + \left\{\sup_{\theta\in\Theta}\frac{1}{\omega}\right\}Kn^{-1}\sum_{t=1}^{n}\rho^t.\end{aligned} \tag{4.7}$$

To deduce (i) is suffices to use the Cesàro lemma and the fact $\rho^t\epsilon_t^2 \to 0$ almost surely. This convergence is obtained by the Borel–Cantelli lemma, the Markov inequality and the existence of a moment of order $s > 0$ for ϵ_t^2 (Proposition 1):

$$\sum_{t=1}^{\infty} P(\rho^t\epsilon_t^2 > \epsilon) \leqslant \sum_{t=1}^{\infty}\frac{\mathrm{E}(\rho^t\epsilon_t^2)^s}{\epsilon^s} < \infty.$$

We now will prove (ii). Suppose that $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$ P_{θ_0}-a.s. Note that $\mathcal{B}_\theta(B)$ is invertible under assumption (A2), by Proposition 1. If $\mathcal{A}_\theta(1) \neq 0$, it follows from (2.9) that

$$\frac{\mathcal{A}_\theta(B)}{\mathcal{B}_\theta(B)}\left\{\epsilon_t^2 + \frac{\omega}{\mathcal{A}_\theta(1)}\right\} = \frac{\mathcal{A}_{\theta_0}(B)}{\mathcal{B}_{\theta_0}(B)}\left\{\epsilon_t^2 + \frac{\omega_0}{\mathcal{A}_{\theta_0}(1)}\right\} \text{ a.s.} \qquad \forall t,$$

implying

$$\left\{\frac{\mathcal{A}_\theta(B)}{\mathcal{B}_\theta(B)} - \frac{\mathcal{A}_{\theta_0}(B)}{\mathcal{B}_{\theta_0}(B)}\right\}\epsilon_t^2 = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} - \frac{\omega}{\mathcal{B}_\theta(1)} \text{ a.s.} \qquad \forall t.$$

It can easily be seen that this equality also holds when $\mathcal{A}_\theta(1) = 0$. Hence, if $\mathcal{A}_\theta(B)/\mathcal{B}_\theta(B) \neq \mathcal{A}_{\theta_0}(B)/\mathcal{B}_{\theta_0}(B)$, there exists a constant linear combination of the ϵ_{t-j}^2, $j \geqslant 0$. This is impossible since, by assumption (A3),

$$\epsilon_t^2 - \mathrm{E}_{\theta_0}(\epsilon_t^2|\epsilon_{t-1}^2, \ldots) = \sigma_t^2(\theta_0)(\eta_t^2 - 1) \neq 0, \qquad \text{with positive probability.}$$

Therefore

$$\frac{\mathcal{A}_\theta(z)}{\mathcal{B}_\theta(z)} = \frac{\mathcal{A}_{\theta_0}(z)}{\mathcal{B}_{\theta_0}(z)}, \quad \forall |z| \leqslant 1, \quad \text{and} \quad \frac{\omega}{\mathcal{B}_\theta(1)} = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)}. \tag{4.8}$$

Under assumption (A4), this implies $\mathcal{A}_\theta(z) = \mathcal{A}_{\theta_0}(z)$, $\mathcal{B}_\theta(z) = \mathcal{B}_{\theta_0}(z)$ and $\omega = \omega_0$, which proves (ii).

Turning now to (iii), first note that $\mathrm{E}_{\theta_0}\mathbf{l}_n(\theta) = \mathrm{E}_{\theta_0}\ell_t(\theta)$ is well defined and belongs to $\mathbb{R} \cup \{+\infty\}$ because $\mathrm{E}_{\theta_0}\ell_t^-(\theta) \leqslant \mathrm{E}_{\theta_0} \log^- \sigma_t^2 \leqslant \max\{0, -\log\omega\} < \infty$. Remark that $\mathrm{E}_{\theta_0}\ell_t(\theta) = \infty$ when, for instance, $\theta = (\omega, 0, \ldots, 0)^{\mathrm{T}}$ and $\mathrm{E}_{\theta_0}\epsilon_t^2 = \infty$. However, we will show that $\mathrm{E}_{\theta_0}|\ell_t(\theta_0)| < \infty$. It remains to show that $\mathrm{E}_{\theta_0}\ell_t^+(\theta_0) < \infty$. By Proposition 1,

$$\mathrm{E}_{\theta_0} \log \sigma_t^2(\theta_0) = \mathrm{E}_{\theta_0} \frac{1}{s} \log\{\sigma_t^2(\theta_0)\}^s \leqslant \frac{1}{s} \log \mathrm{E}_{\theta_0}\{\sigma_t^2(\theta_0)\}^s < \infty.$$

Therefore

$$\mathrm{E}_{\theta_0}\ell_t(\theta_0) = \mathrm{E}_{\theta_0}\left\{\frac{\sigma_t^2(\theta_0)\eta_t^2}{\sigma_t^2(\theta_0)} + \log\sigma_t^2(\theta_0)\right\} = 1 + \mathrm{E}_{\theta_0}\log\sigma_t^2(\theta_0) < \infty.$$

Since $\log x \leqslant x - 1$ for all $x > 0$, and $\log x = x - 1$ if and only if $x = 1$, we have

$$\begin{aligned}\mathrm{E}_{\theta_0}\ell_t(\theta) - \mathrm{E}_{\theta_0}\ell_t(\theta_0) &= \mathrm{E}_{\theta_0}\log\frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} + \mathrm{E}_{\theta_0}\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} - 1 \\ &\geqslant \mathrm{E}_{\theta_0}\left\{\log\frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} + \log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right\} = 0 \end{aligned} \tag{4.9}$$

with equality if and only if $\sigma_t^2(\theta_0)/\sigma_t^2(\theta) = 1$ P_{θ_0}-a.s.

It remains to show (iv). For any $\theta \in \Theta$ and any positive integer k, let $V_k(\theta)$ be the open ball with centre θ and radius $1/k$. In view of (i),

$$\begin{aligned}\liminf_{n\to\infty} \inf_{\theta^*\in V_k(\theta)\cap\Theta} \tilde{\mathbf{l}}_n(\theta^*) &\geqslant \liminf_{n\to\infty} \inf_{\theta^*\in V_k(\theta)\cap\Theta} \mathbf{l}_n(\theta^*) - \limsup_{n\to\infty} \sup_{\theta\in\Theta}|\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| \\ &\geqslant \liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta} \ell_t(\theta^*).\end{aligned}$$

Now we use the following ergodic theorem: if (X_t) is a stationary and ergodic process such that $\mathrm{E}X_1 \in \mathbb{R}\cup\{+\infty\}$, then $n^{-1}\sum_{t=1}^n X_t$ converges almost surely to $\mathrm{E}X_1$ when $n \to \infty$ (see Billingsley 1995, pp. 284 and 495). Applying this theorem to $\{\inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_t(\theta^*)\}_t$ and using $\mathrm{E}_{\theta_0}\ell_t^-(\theta) < \infty$, we obtain

$$\liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta} \ell_t(\theta^*) = \mathrm{E}_{\theta_0} \inf_{\theta^*\in V_k(\theta)\cap\Theta} \ell_1(\theta^*).$$

By the Beppo-Levi theorem, when k increases to ∞, $\mathrm{E}_{\theta_0}\inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_1(\theta^*)$ increases to $\mathrm{E}_{\theta_0}\ell_1(\theta)$. In view of (4.9), (iv) is proved. By a standard compactness argument we complete the proof of Theorem 2.1.

4.2. Proof of Theorem 2.2

The proof rests classically on a Taylor series expansion of the score vector around θ_0. We have

$$\begin{aligned}0 &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\hat{\theta}_n)\\ &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)+\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_t(\theta_{ij}^*)\right)\sqrt{n}(\hat{\theta}_n-\theta_0)\end{aligned}$$

where the θ_{ij}^* are between $\hat{\theta}_n$ and θ_0. We will show that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)\Rightarrow\mathcal{N}(0,\,(\kappa_\eta-1)J) \tag{4.10}$$

and

$$n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_t(\theta_{ij}^*)\to J(i,\,j)\qquad\text{in probability.} \tag{4.11}$$

The theorem will straightforwardly follow. Again, we will split the proof into several intermediate results:

(i) $\mathrm{E}_{\theta_0}\|(\partial\ell_t(\theta_0)/\partial\theta)\ (\partial\ell_t(\theta_0)/\partial\theta^{\mathrm{T}})\|<\infty,\ \mathrm{E}_{\theta_0}\|\partial^2\ell_t(\theta_0)/\partial\theta\partial\theta^{\mathrm{T}}\|<\infty$.

(ii) J is non-singular and $\mathrm{var}_{\theta_0}\{\partial\ell_t(\theta_0)/\partial\theta\}=\{\kappa_\eta-1\}J$.

(iii) There exists a neighbourhood $\mathcal{V}(\theta_0)$ of θ_0 such that, for all $i,\ j,\ k\in\{1,\,\dots,\,p+q+1\}$,

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

(iv) $\|n^{-1/2}\sum_{t=1}^{n}\{\partial\ell_t(\theta_0)/\partial\theta-\partial\tilde{\ell}_t(\theta_0)/\partial\theta\}\|\to 0$ and $\sup_{\theta\in\mathcal{V}(\theta_0)}\|n^{-1}\sum_{t=1}^{n}\{\partial^2\ell_t(\theta)/\partial\theta\partial\theta^{\mathrm{T}}-\partial^2\tilde{\ell}_t(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\}\|\to 0$ in probability when $n\to\infty$.

(v) $n^{-1/2}\sum_{t=1}^{n}\partial\ell_t(\theta_0)/\partial\theta\Rightarrow\mathcal{N}(0,\,(\kappa_\eta-1)J)$.

(vi) $n^{-1}\sum_{t=1}^{n}\partial^2\ell_t(\theta_{ij}^*)/\partial\theta_i\partial\theta_j\to J(i,\,j)$ a.s.

The derivatives of $\ell_t=\epsilon_t^2/\sigma_t^2+\log\sigma_t^2$ are given by

$$\frac{\partial\ell_t}{\partial\theta}=\left\{1-\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right\}, \tag{4.12}$$

$$\frac{\partial^2\ell_t}{\partial\theta\partial\theta^{\mathrm{T}}}=\left\{1-\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^{\mathrm{T}}}\right\}+\left\{2\frac{\epsilon_t^2}{\sigma_t^2}-1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta^{\mathrm{T}}}\right\}. \tag{4.13}$$

For $\theta=\theta_0$, $\epsilon_t^2/\sigma_t^2=\eta_t^2$ is independent of the terms involving σ_t^2 and its derivatives. To prove (i) it will therefore be sufficient to show that

$$E_{\theta_0}\left\|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0)\right\| < \infty, \qquad E_{\theta_0}\left\|\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^{\mathrm{T}}}(\theta_0)\right\| < \infty, \qquad E_{\theta_0}\left\|\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta^{\mathrm{T}}}(\theta_0)\right\| < \infty. \tag{4.14}$$

By (4.4) we have

$$\frac{\partial\underline{\sigma}_t^2}{\partial\omega} = \sum_{k=0}^{\infty} B^k \underline{1}, \qquad \frac{\partial\underline{\sigma}_t^2}{\partial\alpha_i} = \sum_{k=0}^{\infty} B^k \underline{\epsilon}_{t-k-i}^2, \tag{4.15}$$

$$\frac{\partial\underline{\sigma}_t^2}{\partial\beta_j} = \sum_{k=1}^{\infty}\left\{\sum_{i=1}^{k} B^{i-1}B^{(j)}B^{k-i}\right\}\underline{c}_{t-k}, \tag{4.16}$$

where $\underline{1} = (1, 0, \ldots, 0)^{\mathrm{T}}$, $\underline{\epsilon}_t^2 = (\epsilon_t^2, 0, \ldots, 0)^{\mathrm{T}}$, and $B^{(j)}$ is a $p \times p$ matrix with $(1, j)$th element 1 and all other elements 0. Notice that, by the positivity of the coefficients in (4.15)–(4.16), the derivatives of σ_t^2 are non-negative. From (4.15), it is clear that $\partial\sigma_t^2/\partial\omega$ is bounded. Since $\sigma_t^2 \geqslant \omega > 0$, this is also the case for $\{\partial\sigma_t^2/\partial\omega\}/\sigma_t^2$ which therefore possesses moments of any order. By (4.15) we have

$$\alpha_i\frac{\partial\underline{\sigma}_t^2}{\partial\alpha_i} = \sum_{k=0}^{\infty} B^k\alpha_i\underline{\epsilon}_{t-k-i}^2 \leqslant \sum_{k=0}^{\infty} B^k\underline{c}_{t-k} = \underline{\sigma}_t^2,$$

from which we deduce

$$\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\alpha_i} \leqslant \frac{1}{\alpha_i}. \tag{4.17}$$

Hence $\sigma_t^{-2}(\partial\sigma_t^2/\partial\alpha_i)$ has moments of all orders at $\theta = \theta_0$. In view of (4.16) and $\beta_j B^{(j)} \leqslant B$, we have

$$\beta_j\frac{\partial\underline{\sigma}_t^2}{\partial\beta_j} \leqslant \sum_{k=1}^{\infty}\left\{\sum_{i=1}^{k} B^{i-1}BB^{k-i}\right\}\underline{c}_{t=k} = \sum_{k=1}^{\infty} kB^k\underline{c}_{t-k}. \tag{4.18}$$

Using (4.3), we have $\|B^k\| \leqslant K\rho^k$ for all k. Invoking Proposition 1 and the elementary inequality $(a+b)^s \leqslant a^s + b^s$ for all $a, b \geqslant 0$, we deduce that $\underline{c}_t(1) = \omega + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2$ has a moment of order s, for some $s \in]0, 1[$. Using (4.18), the inequalities $\sigma_t^2 \geqslant \omega + B^k(1, 1)\underline{c}_{t-k}(1)$ and $x/(1+x) \leqslant x^s$ for all $x \geqslant 0$, we obtain

$$\begin{aligned} E_{\theta_0}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\beta_j} &\leqslant E_{\theta_0}\frac{1}{\beta_j}\sum_{k=1}^{\infty}\frac{kB^k(1, 1)\underline{c}_{t-k}(1)}{\omega + B^k(1, 1)\underline{c}_{t-k}(1)} \\ &\leqslant \frac{1}{\beta_j}\sum_{k=1}^{\infty} k\,E_{\theta_0}\left\{\frac{B^k(1, 1)\underline{c}_{t-k}(1)}{\omega}\right\}^s \\ &\leqslant \frac{K^s}{\omega^s\beta_j}E_{\theta_0}\{\underline{c}_{t-k}(1)\}^s\sum_{k=1}^{\infty} k\rho^{sk} \leqslant \frac{K}{\beta_j}. \end{aligned} \tag{4.19}$$

Under assumption (A5), we have $\beta_{0j} > 0$ for all j. This allows us to conclude that the first expectation in (4.14) exists.

Let us now turn to the higher derivatives of σ_t^2. It follows from (4.15) that

$$\frac{\partial^2 \underline{\sigma}_t^2}{\partial \omega^2} = \frac{\partial^2 \underline{\sigma}_t^2}{\partial \omega \partial \alpha_i} = 0, \qquad \frac{\partial^2 \underline{\sigma}_t^2}{\partial \omega \partial \beta_j} = \sum_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} \underline{1}. \tag{4.20}$$

Thus

$$\beta_i \frac{\partial^2 \underline{\sigma}_t^2}{\partial \omega \partial \beta_j} \leqslant \sum_{k=1}^{\infty} k B^k \underline{1},$$

whose components are finite, proving that $\partial^2 \sigma_t^2(\theta_0)/\partial\omega\partial\theta_i$ is bounded and admits moments of any order. The same conclusion holds for $\{\partial^2 \sigma_t^2(\theta_0)/\partial\omega\partial\theta_i\}/\sigma_t^2(\theta_0)$. By the second equality in (4.15) we find

$$\frac{\partial^2 \underline{\sigma}_t^2}{\partial \alpha_i \partial \alpha_j} = 0, \qquad \frac{\partial^2 \underline{\sigma}_t^2}{\partial \alpha_i \partial \beta_j} = \sum_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} \underline{\epsilon}_{t-k-i}^2, \tag{4.21}$$

and the arguments used to show (4.19) give

$$\mathrm{E}_{\theta_0} \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \alpha_i \partial \beta_j} \leqslant \frac{K^*}{\beta_j}.$$

This proves that $\{\partial^2 \sigma_t^2(\theta_0)/\partial\alpha_i\partial\theta\}/\sigma_t^2(\theta_0)$ is integrable. Differentiating (4.16) with respect to $\beta_{j'}$ gives

$$\begin{aligned}
\beta_j \beta_{j'} \frac{\partial^2 \underline{\sigma}_t^2}{\partial \beta_j \partial \beta_{j'}} &= \beta_j \beta_{j'} \sum_{k=2}^{\infty} \left[\sum_{i=2}^{k} \left\{ \left(\sum_{l=1}^{i-1} B^{l-1} B^{(j')} B^{i-1-l} \right) B^{(j)} B^{k-i} \right\} \right. \\
&\qquad \left. + \sum_{i=1}^{k-1} \left\{ B^{i-1} B^{(j)} \left(\sum_{l=1}^{k-i} B^{l-1} B^{(j')} B^{k-i-l} \right) \right\} \right] \underline{c}_{t-k} \\
&\leqslant \sum_{k=2}^{\infty} \left[\sum_{i=2}^{k} (i-1) B^k + \sum_{i=1}^{k-1} (k-i) B^k \right] \underline{c}_{t-k} = \sum_{k=2}^{\infty} k(k-1) B^k \underline{c}_{t-k}
\end{aligned} \tag{4.22}$$

since $\beta_j B^{(j)} \leqslant B$. Using the same arguments as for (4.19), we can conclude that

$$\mathrm{E}_{\theta_0} \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \beta_j \partial \beta_j'} \leqslant \frac{K^*}{\beta_j \beta_{j'}},$$

so the existence of the second expectation in (4.14) is proved. Now, since $\{\partial \sigma_t^2/\partial\omega\}/\sigma_t^2$ is bounded, and, by (4.17), the variables $\{\partial\sigma_t^2/\partial\alpha_i\}/\sigma_t^2$ are bounded at θ_0, it is clear that

$$\mathrm{E}_{\theta_0} \left\| \frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right\| < \infty,$$

for $i = 1, \ldots, q+1$. With the notation and arguments already used to prove (4.19), and the elementary inequality $x/(1+x) \leqslant x^{s/2}$ for all $x \geqslant 0$, the Minkowski inequality gives

$$\left\{ \mathrm{E}_{\theta_0} \left(\frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \beta_j} \right)^2 \right\}^{1/2} \leqslant \frac{1}{\beta_{0j}} \sum_{k=1}^{\infty} k \left\{ \mathrm{E}_{\theta_0} \left(\frac{B^k(1, 1)\underline{c}_{t-k}(1)}{\omega_0} \right)^s \right\}^{1/2} < \infty.$$

Finally, the Cauchy–Schwarz inequality allows us to conclude that the third expectation in (4.14) exists.

We now prove (ii). By (4.12) and (i), we have

$$\mathrm{E}_{\theta_0} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \right\} = \mathrm{E}_{\theta_0}(1 - \eta_t^2) \mathrm{E}_{\theta_0} \left\{ \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right\} = 0.$$

Now using (4.13) and (i), J exists and (2.10) holds. We also have

$$\begin{aligned} \mathrm{var}_{\theta_0} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \right\} &= \mathrm{E}_{\theta_0} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right\} \\ &= \mathrm{E}\{(1 - \eta_t^2)^2\} \mathrm{E}_{\theta_0} \left\{ \frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} \right\} \\ &= (\kappa_\eta - 1) J. \end{aligned} \tag{4.23}$$

Now suppose

$$\lambda^{\mathrm{T}} J \lambda = \mathrm{E} \left[\frac{1}{\sigma_t^4(\theta_0)} \left(\lambda^{\mathrm{T}} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right)^2 \right] = 0$$

for some vector $\lambda \in \mathbb{R}^{p+q+1}$. Then, almost surely, $\lambda^{\mathrm{T}}\{\partial \sigma_t^2(\theta_0)/\partial \theta\} = 0$. In view of (2.9) and the stationarity of $\{\partial \sigma_t^2(\theta_0)/\partial \theta\}_t$, we have

$$0 = \lambda^{\mathrm{T}} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} = \lambda^{\mathrm{T}} \begin{pmatrix} 1 \\ \epsilon_{t-1}^2 \\ \vdots \\ \epsilon_{t-q}^2 \\ \sigma_{t-1}^2(\theta_0) \\ \vdots \\ \sigma_{t-p}^2(\theta_0) \end{pmatrix} + \sum_{j=1}^{p} \beta_j \lambda^{\mathrm{T}} \frac{\partial \sigma_{t-j}^2(\theta_0)}{\partial \theta} = \lambda^{\mathrm{T}} \begin{pmatrix} 1 \\ \epsilon_{t-1}^2 \\ \vdots \\ \epsilon_{t-q}^2 \\ \sigma_{t-1}^2(\theta_0) \\ \vdots \\ \sigma_{t-p}^2(\theta_0) \end{pmatrix}.$$

Write $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_{q+p}).^{\mathrm{T}}$ It is clear that $\lambda_1 = 0$, otherwise ϵ_{t-1}^2 would be measurable with respect to the σ-field generated by $\{\eta_u, u < t-1\}$. For the same reason, it can be shown that $\lambda_2 = \ldots = \lambda_{2+i} = 0$ if $\lambda_{q+1} = \ldots = \lambda_{q+i} = 0$. Therefore $\lambda \neq 0$ entails a GARCH$(p-1, \ q-1)$ representation. This is impossible in view of (A4) using the arguments given to establish (4.8). Therefore $\lambda^{\mathrm{T}} J \lambda = 0$ implies $\lambda = 0$, which completes the proof of (ii).

To prove (iii) we differentiate (4.13), which gives

$$\frac{\partial^3 \ell_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\} \tag{4.24}$$
$$+\left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_j\partial\theta_k}\right\}$$
$$+\left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_k}\right\}$$
$$+\left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}\right\}$$
$$+\left\{2 - 6\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}\right\}.$$

We first prove that $\{1 - \epsilon_t^2/\sigma_t^2\}$ is integrable. This term is difficult to handle because ϵ_t^2/σ_t^2 is not uniformly integrable over Θ: in $\theta = (\omega, 0, \ldots, 0)^{\mathrm{T}}$, the ratio ϵ_t^2/σ_t^2 is not integrable when $\mathrm{E}\epsilon_t^2 = \infty$. However, we will show that $\{1 - \epsilon_t^2/\sigma_t^2\}$ is uniformly integrable in a neighbourhood of θ_0. Let Θ^* be a compact set containing θ_0 and included in $\overset{\circ}{\Theta}$. Denote by B_0 the matrix B (defined in (4.2)) evaluated at $\theta = \theta_0$. For all $\delta > 0$, there exists a neighbourhood $\mathcal{V}(\theta_0)$ of θ_0, with $\mathcal{V}(\theta_0) \subseteq \Theta^*$, such that $B_0 \leqslant (1+\delta)B$ for all $\theta \in \mathcal{V}(\theta_0)$. From (4.4), we obtain

$$\sigma_t^2 = \omega\sum_{k=0}^{\infty} B^k(1, 1) + \sum_{i=1}^{q}\alpha_i\left\{\sum_{k=0}^{\infty} B^k(1, 1)\epsilon_{t-k-i}^2\right\}.$$

Since $\mathcal{V}(\theta_0) \subset \Theta^*$, we have $\sup_{\theta\in\mathcal{V}(\theta_0)} 1/\alpha_i < \infty$. Using also $x/(1+x) \leqslant x^s$ for all $x \geqslant 0$ and any $s \in]0, 1[$, we obtain

$$\sup_{\theta\in\mathcal{V}(\theta_0)}\frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \leqslant \sup_{\theta\in\mathcal{V}(\theta_0)}\left\{\frac{\omega_0\sum_{k=0}^{\infty}B_0^k(1, 1)}{\omega} + \sum_{i=1}^{q}\alpha_{0i}\left(\sum_{k=0}^{\infty}\frac{B_0^k(1, 1)\epsilon_{t-k-i}^2}{\omega + \alpha_i B^k(1, 1)\epsilon_{t-k-i}^2}\right)\right\}$$
$$\leqslant K + \sum_{i=1}^{q}\sup_{\theta\in\mathcal{V}(\theta_0)}\left\{\frac{\alpha_{0i}}{\alpha_i}\sum_{k=0}^{\infty}\frac{B_0^k(1, 1)}{B^k(1, 1)}\left(\frac{\alpha_i B^k(1, 1)\epsilon_{t-k-i}^2}{\omega}\right)^s\right\}$$
$$\leqslant K + K\sum_{i=1}^{q}\sum_{k=0}^{\infty}(1+\delta)^k\rho^{ks}\epsilon_{t-k-i}^{2s}. \tag{4.25}$$

Choosing s such that $\mathrm{E}\epsilon_t^{2s} < \infty$ and, for instance, $\delta = (1-\rho^s)/(2\rho^s)$, we obtain

$$\mathrm{E}_{\theta_0}\sup_{\theta\in\mathcal{V}(\theta_0)}\frac{\epsilon_t^2}{\sigma_t^2} = \mathrm{E}_{\theta_0}\sup_{\theta\in\mathcal{V}(\theta_0)}\frac{\sigma_t^2(\theta_0)}{\sigma_t^2} < \infty$$

for some neighbourhood $\mathcal{V}(\theta_0)$. For the same choice of δ, with s such that $\mathrm{E}\epsilon_t^{4s} < \infty$, and using (4.25), we find

$$\left\| \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\epsilon_t^2}{\sigma_t^2} \right\|_2 = \kappa_\eta^{1/2} \left\| \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \right\|_2$$

$$\leqslant \kappa_\eta^{1/2} K + \kappa_\eta^{1/2} K q \sum_{k=0}^{\infty} (1+\delta)^k \rho^{ks} \|\epsilon_t^{2s}\|_2 < \infty. \tag{4.26}$$

Let us now turn to the second term in brackets in (4.24). By differentiating (4.20) and (4.21), arguments already used to show (4.17) yield

$$\sup_{\theta \in \Theta^*} \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} \leqslant K,$$

when at least one index i_1, i_2 or i_3 does not belong to $\{q+1,\ q+2, \ldots, q+1+p\}$. Following the lines of the proofs of (4.18) and (4.19), we obtain

$$\beta_i \beta_j \beta_k \frac{\partial^3 \sigma_t^2}{\partial \beta_i \partial \beta_j \partial \beta_k} \leqslant \sum_{k=3}^{\infty} k(k-1)(k-2) B^k(1,1) \underline{c}_{t-k}(1),$$

$$\sup_{\theta \in \Theta^*} \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial \beta_i \partial \beta_j \partial \beta_k} \leqslant K \left\{ \sup_{\theta \in \Theta^*} \frac{1}{\omega^s \beta_i \beta_j \beta_k} \right\} \sum_{k=3}^{\infty} k(k-1)(k-2) \rho^{ks} \left\{ \sup_{\theta \in \Theta^*} \underline{c}_{t-k}(1) \right\}^s$$

for all $s \in]0, 1[$. Since $\mathrm{E}_{\theta_0} \{\sup_{\theta \in \Theta^*} \underline{c}_{t-k}(1)\}^{2s} < \infty$ for some $s > 0$, we then have

$$\mathrm{E}_{\theta_0} \sup_{\theta \in \Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial \theta_i \partial \theta_j \partial \theta_k} \right|^2 < \infty. \tag{4.27}$$

More generally, it can similarly be shown that, for any d,

$$\mathrm{E}_{\theta_0} \sup_{\theta \in \Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial \theta_i \partial \theta_j \partial \theta_k} \right|^d < \infty. \tag{4.28}$$

By the Cauchy–Schwarz inequality, (4.26) and (4.27), we obtain

$$\mathrm{E}_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \left\{ 1 - \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial \theta_i \partial \theta_j \partial \theta_k} \right\} \right| < \infty.$$

To deal with the other terms of the sum in (4.24) we show that, similarly to (4.28),

$$\mathrm{E}_{\theta_0} \sup_{\theta \in \Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j} \right|^d < \infty, \qquad \mathrm{E}_{\theta_0} \sup_{\theta \in \Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_i} \right|^d < \infty, \tag{4.29}$$

for any integer d. This allows us to obtain, using the Hölder inequality,

$$\mathrm{E}_{\theta_0} \sup_{\theta\in\mathcal{V}(\theta_0)} \left| \left\{ 2 - 6\frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k} \right\} \right|$$

$$\leqslant \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \left| 2 - 6\frac{\epsilon_t^2}{\sigma_t^2} \right| \right\|_2 \left\| \max_i \left\| \sup_{\theta\in\Theta^*} \left| \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i} \right| \right\|_6 \right\|^3 < \infty.$$

The other terms of the sum in (4.24) can also be treated in this way, and we thus obtain (iii).

To show (iv), we use (4.5) to obtain the following results, analogous to (4.15)–(4.16):

$$\frac{\partial\underline{\tilde{\sigma}}_t^2}{\partial\omega} = \sum_{k=0}^{t-1-q} B^k \underline{1} + \sum_{k=1}^{q} B^{t-k} \frac{\partial\underline{\tilde{c}}_k}{\partial\omega} + B^t \frac{\partial\underline{\tilde{\sigma}}_0^2}{\partial\omega}, \tag{4.30}$$

$$\frac{\partial\underline{\tilde{\sigma}}_t^2}{\partial\alpha_i} = \sum_{k=0}^{t-1-q} B^k \underline{\epsilon}_{t-k-i}^2 + \sum_{k=1}^{q} B^{t-k} \frac{\partial\underline{\tilde{c}}_k}{\partial\alpha_i} + B^t \frac{\partial\underline{\tilde{\sigma}}_0^2}{\partial\alpha_i}, \tag{4.31}$$

$$\frac{\partial\underline{\tilde{\sigma}}_t^2}{\partial\beta_j} = \sum_{k=1}^{t-1-q} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} \underline{c}_{t-k} + \sum_{k=1}^{q} \left\{ \sum_{i=1}^{t-k} B^{i-1} B^{(j)} B^{t-k-i} \right\} \underline{\tilde{c}}_k, \tag{4.32}$$

where $\partial\underline{\tilde{\sigma}}_0^2/\partial\omega = (0, \ldots, 0)^{\mathrm{T}}$ when the initial conditions are given by (2.7), and $\partial\underline{\tilde{\sigma}}_0^2/\partial\omega = (1, \ldots, 1)^{\mathrm{T}}$ when the initial conditions are given in (2.6). The second-order derivatives have similar expressions. In view of (4.3), (4.15)–(4.16) and (4.30)–(4.32), we have, almost surely,

$$\sup_{\theta\in\Theta} \left\| \frac{\partial\sigma_t^2}{\partial\theta} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta} \right\| < K\rho^t, \quad \sup_{\theta\in\Theta} \left\| \frac{\partial^2\theta_t^2}{\partial\theta\partial\theta^{\mathrm{T}}} - \frac{\partial^2\tilde{\sigma}_t^2}{\partial\theta\partial\theta^{\mathrm{T}}} \right\| < K\rho^t, \qquad \forall t. \tag{4.33}$$

In view of (4.6), we have

$$\left| \frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2} \right| = \left| \frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2\tilde{\sigma}_t^2} \right| \leqslant \frac{K\rho^t}{\sigma_t^2}, \qquad \frac{\sigma_t^2}{\tilde{\sigma}_t^2} \leqslant 1 + K\rho^t. \tag{4.34}$$

Since

$$\frac{\partial\tilde{\ell}_t(\theta)}{\partial\theta} = \left\{ 1 - \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\sigma}_t^2}{\partial\theta} \right\} \quad \text{and} \quad \frac{\partial\ell_t(\theta)}{\partial\theta} = \left\{ 1 - \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta} \right\}$$

we have, using (4.34) and the first inequality in (4.33),

$$\begin{aligned} \left| \frac{\partial\ell_t(\theta_0)}{\partial\theta_i} - \frac{\partial\tilde{\ell}_t(\partial_0)}{\partial\theta_i} \right| &= \left| \left\{ \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i} \right\} + \left\{ 1 - \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2} \right\} \left\{ \frac{\partial\sigma_t^2}{\partial\theta_i} \right\} \right. \\ &\quad \left. + \left\{ 1 - \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\tilde{\sigma}_t^2} \right\} \left\{ \frac{\partial\sigma_t^2}{\partial\theta_i} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta_i} \right\} \right| (\theta_0) \\ &\leqslant K\rho^t(1+\eta_t^2) \left| 1 + \left\{ \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_i} \right\} \right|. \end{aligned}$$

It follows that

$$\left| n^{-1/2} \sum_{t=1}^{n} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta_i} - \frac{\partial \tilde{\ell}_t(\theta_0)}{\partial \theta_i} \right\} \right| \leqslant K^* n^{-1/2} \sum_{t=1}^{n} \rho^t (1+\eta_t^2) \left\{ 1 + \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \right\}. \quad (4.35)$$

The Markov inequality, (i), and the independence between η_t and $\sigma_t^2(\theta_0)$ entail that, for all $\epsilon > 0$,

$$P\left(n^{-1/2} \sum_{t=1}^{n} \rho^t (1+\eta_t^2) \left\{ 1 + \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right\} > \epsilon \right)$$

$$\leqslant \frac{2}{\epsilon} \left(1 + \mathrm{E}_{\theta_0} \left| \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right| \right) n^{-1/2} \sum_{t=1}^{n} \rho^t \to 0,$$

which, in view of (4.35), shows the first part of (iv).

We now turn to the second-order derivatives. By (4.13) and the inequalities in (4.33) and (4.34), we find

$$\begin{aligned}
\sup_{\theta \in \mathcal{V}(\theta_0)} \left| n^{-1} \sum_{t=1}^{n} \left\{ \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{\ell}_t(\theta)}{\partial \theta_i \partial \theta_j} \right\} \right| &\leqslant n^{-1} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \left\{ \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j} \right\} \right. \\
&\quad + \left\{ 1 - \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} \right\} \left\{ \left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2} \right) \frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j} + \frac{1}{\tilde{\sigma}_t^2} \left(\frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{\sigma}_t^2}{\partial \theta_i \partial \theta_j} \right) \right\} \\
&\quad + \left\{ 2 \frac{\epsilon_t^2}{\sigma_t^2} - 2 \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_i} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_j} \right\} \\
&\quad + \left\{ 2 \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right\} \left\{ \left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2} \right) \frac{\partial \sigma_t^2}{\partial \theta_i} + \frac{1}{\tilde{\sigma}_t^2} \left(\frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{\partial \tilde{\sigma}_t^2}{\partial \theta_i} \right) \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_j} \right\} \\
&\quad \left. + \left\{ 2 \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right\} \left\{ \frac{1}{\tilde{\sigma}_t^2} \frac{\partial \tilde{\sigma}_t^2}{\partial \theta_i} \right\} \left\{ \left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2} \right) \frac{\partial \sigma_t^2}{\partial \theta_j} + \frac{1}{\tilde{\sigma}_t^2} \left(\frac{\partial \sigma_t^2}{\partial \theta_j} - \frac{\partial \tilde{\sigma}_t^2}{\partial \theta_j} \right) \right\} \right| \\
&\leqslant K n^{-1} \sum_{t=1}^{n} \rho^t \Upsilon_t,
\end{aligned}$$

where

$$\Upsilon_t = \sup_{\theta \in \mathcal{V}(\theta_0)} \left\{ 1 + \frac{\epsilon_t^2}{\sigma_t^2} \right\} \left\{ 1 + \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \theta_i \partial \theta_j} + \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_i} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_j} \right\}.$$

From (4.26), (4.29) and the Hölder inequality, it follows that Υ_t is integrable for some neighbourhood $\mathcal{V}(\theta_0)$. Using the Markov inequality again, the second convergence of (iv) follows.

To prove (v) we will apply a central limit theorem for margingale differences. Let $\mathcal{F}_t$ be the σ-algebra generated by the random variables ϵ_{t-i}, $i \geqslant 0$. Notice that

$\mathrm{E}_{\theta_0}(\partial\ell_t(\theta_0)/\partial\theta|\mathcal{F}_{t-1})=0$ and that in view of (4.23), $\mathrm{var}_{\theta_0}(\partial\ell_t(\theta_0)/\partial\theta)$ exists. By assumptions (A3) and (A5), implying $0<\kappa_\eta-1<\infty$, and since J is non-singular, the matrix $\mathrm{var}_{\theta_0}(\partial\ell_t(\theta_0)/\partial\theta)$ is non-degenerate. Hence for any $\lambda\in\mathbb{R}^{p+q+1}$, the sequence $\{\lambda^{\mathrm{T}}(\partial/\partial\theta)\ell_t(\theta_0),\mathcal{F}_t\}_t$ is a square-integrable stationary margingale difference. The central limit theorem of Billingsley (1961) and the Wold–Cramér device allow us to derive the asymptotic normality result (v).

We now consider a Taylor series expansion of the second-order derivatives of the criterion about θ_0, in order to prove (vi). For all i and j, we have

$$n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta_{ij}^*)=n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta_0)+n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^{\mathrm{T}}}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\tilde{\theta}_{ij})\right\}(\theta_{ij}^*-\theta_0), \tag{4.36}$$

where $\tilde{\theta}_{ij}$ is between θ_{ij}^* and θ_0. The almost sure convergence of $\tilde{\theta}_{ij}$ to θ_0, the ergodic theorem and (iii) imply that, almost surely,

$$\begin{aligned}\limsup_{n\to\infty}\left\|n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^{\mathrm{T}}}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\tilde{\theta}_{ij})\right\}\right\| &\leqslant \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial}{\partial\theta^{\mathrm{T}}}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)\right\}\right\| \\ &= \mathrm{E}_{\theta_0}\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial}{\partial\theta^{\mathrm{T}}}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)\right\}\right\|<\infty.\end{aligned}$$

Therefore, since $\|\theta_{ij}^*-\theta_0\|\to 0$ a.s., the second term on the right-hand side of (4.36) converges to 0 with probability 1. The convergence in (vi) follows from an application of the ergodic theorem to the first term on the right-hand side of (4.36).

To complete the proof of Theorem 2.2 it suffices to apply the Slutsky lemma. In view of (iv), (v) and (vi), we obtain (4.10) and (4.11).

4.3. Proof of Theorem 3.1

Following the scheme of proof of Theorem 2.1 we make the following statements.

(i) $\lim_{n\to\infty}\sup_{\varphi\in\Phi}|\mathbf{l}_n(\varphi)-\tilde{\mathbf{l}}_n(\varphi)|=0$ a.s.
(ii) $(\exists t\in\mathbb{Z}$ such that $\epsilon_t(\vartheta)=\epsilon_t(\vartheta_0)$ and $\sigma_t^2(\varphi)=\sigma_t^2(\varphi_0)$ P_{φ_0}-a.s.$)\Rightarrow\varphi=\varphi_0$.
(iii) If $\varphi\neq\varphi_0$, then $\mathrm{E}_{\varphi_0}\ell_t(\varphi)>\mathrm{E}_{\varphi_0}\ell_t(\varphi_0)$.
(iv) Any $\varphi\neq\varphi_0$ has a neighbourhood $V(\varphi)$ such that

$$\liminf_{n\to\infty}\inf_{\varphi^*\in V(\varphi)}\tilde{\mathbf{l}}_n(\varphi^*)>\mathrm{E}_{\varphi_0}\ell_1(\varphi_0)\text{ a.s.}$$

We shall only prove (i)–(iii) since the proof of (iv) is similar to that given for Theorem 2.1.

We first prove (i). Equations (4.1)–(4.4) remain valid under the convention that $\epsilon_t=\epsilon_t(\vartheta)$. Equation (4.5) has to be replaced by

$$\underline{\tilde{\sigma}}_t^2=\underline{\tilde{c}}_t+B\underline{\tilde{c}}_{t-1}+\ldots+B^{t-1}\underline{\tilde{c}}_1+B^t\underline{\tilde{\sigma}}_0^2, \tag{4.37}$$

where $\tilde{\underline{c}}_t = (\omega + \sum_{i=1}^{q} \alpha_i \tilde{\epsilon}_{t-i}^2, 0, \ldots, 0)^{\mathrm{T}}$, the variables with 'tilde' being initialized as indicated in Section 3. By assumptions (A7)–(A8) we have,

$$\text{for all } k \geqslant 1 \text{ and } 1 \leqslant i \leqslant q, \qquad \sup_{\varphi \in \Phi} |\epsilon_{k-i} - \tilde{\epsilon}_{k-i}| \leqslant K\rho^k \text{ a.s.} \tag{4.38}$$

It follows that, almost surely,

$$\|\underline{c}_k - \tilde{\underline{c}}_k\| \leqslant \sum_{i=1}^{q} |\alpha_i| |\tilde{\epsilon}_{k-i}^2 - \epsilon_{k-i}^2| \leqslant K\rho^k \left(\sum_{i=1}^{q} |\epsilon_{k-i}| + 1 \right)$$

and thus, by (4.4), (4.37) and Proposition 1,

$$\begin{aligned} \|\underline{\sigma}_t^2 - \tilde{\underline{\sigma}}_t^2\| &= \left\| \sum_{k=1}^{t} B^{t-k} (\underline{c}_k - \tilde{\underline{c}}_k) + B^t (\underline{\sigma}_0^2 - \tilde{\underline{\sigma}}_0^2) \right\| \\ &\leqslant K \sum_{k=1}^{t} \rho^{t-k} \rho^k \left(\sum_{i=1}^{q} |\epsilon_{k-i}| + 1 \right) + K\rho^t \\ &\leqslant K\rho^t \sum_{k=1-q}^{t-1} (|\epsilon_k| + 1). \end{aligned} \tag{4.39}$$

Therefore we have

$$\begin{aligned} \sup_{\varphi \in \Phi} |\mathbf{l}_n(\varphi) - \tilde{\mathbf{l}}_n(\varphi)| &\leqslant n^{-1} \sum_{t=1}^{n} \sup_{\varphi \in \Phi} \left\{ \left| \frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\tilde{\sigma}_t^2 \sigma_t^2} \right| \epsilon_t^2 + \left| \log\left(1 + \frac{\sigma_t^2 - \tilde{\sigma}_t^2}{\tilde{\sigma}_t^2} \right) \right| + \frac{|\epsilon_t^2 - \tilde{\epsilon}_t^2|}{\tilde{\sigma}_t^2} \right\} \\ &\leqslant \left\{ \sup_{\varphi \in \Phi} \max\left(\frac{1}{\omega^2}, \frac{1}{\omega} \right) \right\} K n^{-1} \sum_{t=1}^{n} \rho^t (\epsilon_t^2 + 1) \sum_{k=-q}^{t} (|\epsilon_k| + 1). \end{aligned}$$

The latter equation is analogous to (4.7), $\epsilon_t^2 + 1$ being replaced by $\xi_t = (\epsilon_t^2 + 1) \sum_{k=-q}^{t} (|\epsilon_k| + 1)$. It therefore suffices to show that $\sum_{t=1}^{\infty} \mathrm{E}(\rho^t \xi_t)^r < \infty$, for some $r > 0$. For all positive random variables X and Y, and all $r \in]0, 1]$, we have $\mathrm{E}(X + Y)^r \leqslant \mathrm{E}(X)^r + \mathrm{E}(Y)^r$. Hence, in view of Proposition 1,

$$\begin{aligned} \mathrm{E}(\rho^t \xi_t)^{s/2} &\leqslant \rho^{ts/2} \sum_{k=-q}^{t} \mathrm{E}(\epsilon_t^2 |\epsilon_k| + \epsilon_t^2 + |\epsilon_k| + 1)^{s/2} \\ &\leqslant \rho^{ts/2} \sum_{k=-q}^{t} [\{\mathrm{E}(\epsilon_t^{2s}) \mathrm{E}|\epsilon_k|^s\}^{1/2} + \mathrm{E}|\epsilon_t|^s + \mathrm{E}|\epsilon_k|^{s/2} + 1] \\ &= O(t\rho^{ts/2}), \end{aligned}$$

which completes the proof of (i).

If $\vartheta \neq \vartheta_0$, the first equality in (ii) and assumptions (A8)–(A9) imply the existence of a constant linear combination of the X_{t-j}, $j \geqslant 0$. Note that the linear innovation of (X_t) is

$X_t - \mathrm{E}(X_t|\mathcal{F}_{t-1}) = \eta_t\sigma_t(\varphi_0) \neq 0$ with positive probability, because $\sigma_t^2(\varphi_0) \geqslant \omega_0 > 0$ and $\mathrm{E}(\eta_t^2) = 1$. This implies $\vartheta = \vartheta_0$, and finally $\theta = \theta_0$ by arguments given in the proof of Theorem 2.1. Thus (ii) holds.

We now have

$$\mathrm{E}\varphi_0\ell_t(\varphi) - \mathrm{E}\varphi_0\ell_t(\varphi_0) = \mathrm{E}\varphi_0\left\{\log\frac{\sigma_t^2(\varphi)}{\sigma_t^2(\varphi_0)} + \frac{\sigma_t^2(\varphi_0)}{\sigma_t^2(\varphi)} - 1\right\} + \mathrm{E}\varphi_0\frac{\{\epsilon_t(\vartheta) - \epsilon_t(\vartheta_0)\}^2}{\sigma_t^2(\varphi)}$$
$$+ \mathrm{E}\varphi_0\frac{2\eta_t\sigma_t(\varphi_0)\{\epsilon_t(\vartheta) - \epsilon_t(\vartheta_0)\}}{\sigma_t^2(\varphi)}.$$

In this equality, the last expectation is null and the first expectation is positive by arguments already given. Hence $\mathrm{E}\varphi_0\ell_t(\varphi) - \mathrm{E}\varphi_0\ell_t(\varphi_0) \geqslant 0$, with equality if and only if $\epsilon_t(\vartheta) = \epsilon_t(\vartheta_0)$ and $\sigma_t^2(\varphi_0)/\sigma_t^2(\varphi) = 1$ P_{θ_0}-a.s. which, by (ii), implies $\varphi = \varphi_0$, thus establishing (iii).

4.4. Proof of Theorem 3.2

The proof utilizes the following lemma.

Lemma 4.1. *If the distribution of η_t is symmetric, then,*

$$\forall j, \qquad \mathrm{E}\varphi_0\{g(\epsilon_t^2, \epsilon_{t-1}^2, \ldots)\epsilon_{t-j}f(\epsilon_{t-j-1}, \epsilon_{t-j-2}, \ldots)(\varphi_0)\} = 0,$$

for all functions f and g such that the expectation exists.

Proof. The result is obvious when $j < 0$. For $j \geqslant 0$, note that $e_t = \epsilon_t^2(\vartheta_0)$ is a measurable function of $\eta_t^2, \ldots, \eta_{t-j+1}^2, e_{t-j}^2, e_{t-j-1}^2, \ldots$. Consequently, $\mathrm{E}\{g(e_t^2, e_{t-1}^2, \ldots)|e_{t-j}, e_{t-j-1}, \ldots\}$ is an even function of each conditioning variable, and can therefore be denoted by $h(e_{t-j}^2, e_{t-j-1}^2, \ldots)$ for some measurable function h. Therefore the expectation in the lemma reads

$$\mathrm{E}[\mathrm{E}\{h(\eta_{t-j}^2\sigma_{t-j}^2, e_{t-j-1}^2, \ldots)\eta_{t-j}\sigma_{t-j}f(e_{t-j-1}, e_{t-j-2}, \ldots)|\eta_{t-j-1}, \eta_{t-j-2}, \ldots\}]$$
$$= \mathrm{E}\left[\int h(x^2\sigma_{t-j}^2, e_{t-j-1}^2, \ldots)x\sigma_{t-j}f(e_{t-j-1}, e_{t-j-2}, \ldots)\mathrm{d}P_\eta(x)\right] = 0.$$

□

We now turn to the proof of Theorem 3.2. Following the scheme of proof of Theorem 2.2, it will be sufficient to establish the following properties:

(i) $\mathrm{E}\varphi_0\|(\partial\ell_t(\varphi_0)/\partial\varphi)(\partial\ell_t(\varphi_0)/\partial\varphi^{\mathrm{T}})\| < \infty$, $\mathrm{E}\varphi_0\|\partial^2\ell_t(\varphi_0)/\partial\varphi\partial\varphi^{\mathrm{T}}\| < \infty$.
(ii) $\mathcal{I}$ and $\mathcal{J}$ are not singular and, when η_t is symmetrically distributed, they have the block-diagonal forms stated in the theorem.
(iii) $\|n^{-1/2}\sum_{t=1}^n\{\partial\ell_t(\varphi_0)/\partial\varphi - \partial\tilde{\ell}_t(\varphi_0)/\partial\varphi\}\| \to 0$ and $\sup_{\varphi\in\mathcal{V}(\varphi_0)}\|n^{-1}\sum_{t=1}^n\{\partial^2\ell_t(\varphi)/\partial\varphi\partial\varphi^{\mathrm{T}} - \partial^2\tilde{\ell}_t(\varphi)/\partial\varphi\partial\varphi^{\mathrm{T}}\}\| \to 0$ in probability when $n \to \infty$.

(iv) $n^{-1/2}\sum_{t=1}^{n}\partial\ell_t(\varphi_0)/\partial\varphi \Rightarrow \mathcal{N}(0, \mathcal{I})$.
(v) $n^{-1}\sum_{t=1}^{n}\partial^2\ell_t(\varphi^*)/\partial\varphi_i\partial\varphi_j \to \mathcal{J}(i, j)$ a.s., for any φ^* between $\hat{\varphi}_n$ and φ_0.

We begin with the expressions for the derivatives of ℓ_t. Formulae (4.12)–(4.13) providing the derivatives with respect to θ remain valid, with $\epsilon_t^2 = \epsilon_t^2(\vartheta)$. The other derivatives are given by

$$\frac{\partial\ell_t(\varphi)}{\partial\vartheta} = \left(1 - \frac{\epsilon_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta} + \frac{2\epsilon_t}{\sigma_t^2}\frac{\partial\epsilon_t}{\partial\vartheta}, \tag{4.40}$$

$$\begin{aligned}\frac{\partial^2\ell_t(\varphi)}{\partial\vartheta\partial\vartheta^{\mathrm{T}}} = {}&\left(1 - \frac{\epsilon_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\vartheta\partial\vartheta^{\mathrm{T}}} + \left(2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta^{\mathrm{T}}}\\ &+ \frac{2}{\sigma_t^2}\frac{\partial\epsilon_t}{\partial\vartheta}\frac{\partial\epsilon_t}{\partial\vartheta^{\mathrm{T}}} + \frac{2\epsilon_t}{\sigma_t^2}\frac{\partial^2\epsilon_t}{\partial\vartheta\partial\vartheta^{\mathrm{T}}} - \frac{2\epsilon_t}{\sigma_t^2}\left(\frac{\partial\epsilon_t}{\partial\vartheta}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta^{\mathrm{T}}} + \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta}\frac{\partial\epsilon_t}{\partial\vartheta^{\mathrm{T}}}\right),\end{aligned} \tag{4.41}$$

$$\frac{\partial^2\ell_t(\varphi)}{\partial\vartheta\partial\theta^{\mathrm{T}}} = \left(1 - \frac{\epsilon_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\vartheta\partial\theta^{\mathrm{T}}} + \left(2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta^{\mathrm{T}}} - \frac{2\epsilon_t}{\sigma_t^2}\frac{\partial\epsilon_t}{\partial\vartheta}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta^{\mathrm{T}}}. \tag{4.42}$$

Next we turn to the derivatives of ϵ_t. Let

$$v_t(\vartheta) = -A_\vartheta^{-1}(B)\epsilon_t(\vartheta), \qquad u_t(\vartheta) = B_\vartheta^{-1}(B)\epsilon_t(\vartheta), \tag{4.43}$$

let $H_{k,l}(t)$ denote the $k \times l$ Hankel matrix with generic term ϵ_{t-i-j}, and let $0_{k,l}$ denote the $k \times l$ null matrix. Then we have

$$\frac{\partial\epsilon_t}{\partial\vartheta} = (-A_\vartheta(1)B_\vartheta^{-1}(1), v_{t-1}(\vartheta), \ldots, v_{t-P}(\vartheta), u_{t-1}(\vartheta), \ldots, u_{t-Q}(\vartheta))^{\mathrm{T}}$$

and

$$\frac{\partial^2\epsilon_t}{\partial\vartheta\partial\vartheta^{\mathrm{T}}} = \begin{pmatrix} \multicolumn{2}{c}{\multirow{2}{*}{$0_{P+1,P+1}$}} & 0_{1,Q} \\ & & -A_\vartheta^{-1}(B)B_\vartheta^{-1}(B)H_{P,Q}(t) \\ 0_{Q,1} & -A_\vartheta^{-1}(B)B_\vartheta^{-1}(B)H_{Q,P}(t) & 2B_\vartheta^{-2}(B)H_{Q,Q}(t) \end{pmatrix}. \tag{4.44}$$

In view of (4.4), we also have

$$\frac{\partial\sigma_t^2}{\partial\vartheta_j} = \sum_{k=0}^{\infty} B^k(1, 1)\sum_{i=1}^{q} 2\alpha_i\epsilon_{t-k-i}\frac{\partial\epsilon_{t-k-i}}{\partial\vartheta_j}, \tag{4.45}$$

where ϑ_j denotes the jth component of ϑ, and

$$\frac{\partial^2\sigma_t^2}{\partial\vartheta_j\partial\vartheta_l} = \sum_{k=0}^{\infty} B^k(1, 1)\sum_{i=1}^{q} 2\alpha_i\left(\frac{\partial\epsilon_{t-k-i}}{\partial\vartheta_j}\frac{\partial\epsilon_{t-k-i}}{\partial\vartheta_l} + \epsilon_{t-k-i}\frac{\partial^2\epsilon_{t-k-i}}{\partial\vartheta_j\partial\vartheta_l}\right). \tag{4.46}$$

We now prove (i). First, note that the existence of the expectations in (4.14) still holds. In view of (4.40)–(4.42), from the independence among $(\epsilon_t/\sigma_t)(\varphi_0) = \eta_t$ and $\sigma_t^2(\varphi_0)$, the derivatives of the latter, and the derivatives of $\epsilon_t(\vartheta_0)$, together with $\mathrm{E}(\eta_t^4) < \infty$ and $\sigma_t^2(\varphi_0) > \omega_0 > 0$, it is plain that proving

$$\mathrm{E}\varphi_0\left\|\frac{\partial\epsilon_t}{\partial\vartheta}\frac{\partial\epsilon_t}{\partial\vartheta^{\mathrm{T}}}(\theta_0)\right\| < \infty, \qquad \mathrm{E}\varphi_0\left\|\frac{\partial^2\epsilon_t}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}(\theta_0)\right\| < \infty, \tag{4.47}$$

$$\mathrm{E}\varphi_0\left\|\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\vartheta}\frac{\partial\sigma_t^2}{\partial\vartheta^{\mathrm{T}}}(\varphi_0)\right\| < \infty, \qquad \mathrm{E}\varphi_0\left\|\frac{\partial^2\sigma_t^2}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}(\varphi_0)\right\| < \infty, \qquad \mathrm{E}\varphi_0\left\|\frac{\partial^2\sigma_t^2}{\partial\theta\partial\vartheta^{\mathrm{T}}}(\varphi_0)\right\| < \infty \tag{4.48}$$

will suffice to establish (i), as well as the existence of $\mathcal{I}$ and $\mathcal{J}$. From the expressions for the derivatives of ϵ_t, (4.43)–(4.44), and using $\mathrm{E}\epsilon_t^2(\vartheta_0) < \infty$, (4.47) holds. Now the Cauchy–Schwarz inequality implies

$$\left|\sum_{i=1}^{q}\alpha_{0i}\epsilon_{t-k-i}(\vartheta_0)\frac{\partial\epsilon_{t-k-i}(\vartheta_0)}{\partial\vartheta_j}\right| \leqslant \left\{\sum_{i=1}^{q}\alpha_{0i}\epsilon_{t-k-i}^2(\vartheta_0)\right\}^{1/2}\left\{\sum_{i=1}^{q}\alpha_{0i}\left(\frac{\partial\epsilon_{t-k-i}(\vartheta_0)}{\partial\vartheta_j}\right)^2\right\}^{1/2}.$$

and thus, by (4.45) and the positivity of ω_0,

$$\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0) \leqslant 2\sum_{k=0}^{\infty}B_0^k(1,\,1)\left\{\omega_0 + \sum_{i=1}^{q}\alpha_{0i}\epsilon_{t-k-i}^2(\vartheta_0)\right\}^{1/2}\left\{\sum_{i=1}^{q}\alpha_{0i}\left(\frac{\partial\epsilon_{t-k-i}(\vartheta_0)}{\partial\vartheta_j}\right)^2\right\}^{1/2}.$$

Therefore, it follows from the triangular inequality and the elementary inequalities $(\sum|x_i|)^{1/2} \leqslant \sum|x_i|^{1/2}$ and $x/(1+x^2) \leqslant 1$, that

$$\begin{aligned}
\left\|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0)\right\|_2 &\leqslant \left\|2\sum_{k=0}^{\infty}\frac{B^{k/2}(1,\,1)\underline{c}_{t-k}^{1/2}(1)B^{k/2}(1,\,1)\sum_{i=1}^{q}\alpha_i^{1/2}|\partial\epsilon_{t-k-i}/\partial\vartheta_j|}{\omega + B^k(1,\,1)\underline{c}_{t-k}(1)}(\varphi_0)\right\|_2 \\
&\leqslant \left\|\frac{2}{\sqrt{\omega}}\sum_{k=0}^{\infty}B^{k/2}(1,\,1)\frac{B^{k/2}(1,\,1)\underline{c}_{t-k}^{1/2}(1)/\sqrt{\omega}}{1+(B^{k/2}(1,\,1)\underline{c}_{t-k}^{1/2}(1)/\sqrt{\omega})^2}\sum_{i=1}^{q}\alpha_i^{1/2}\left|\frac{\partial\epsilon_{t-k-i}}{\partial\vartheta j}\right|(\varphi_0)\right\|_2 \\
&\leqslant \frac{K}{\sqrt{\omega_0}}\sum_{k=0}^{\infty}\rho^{k/2}\sum_{i=1}^{q}\alpha_{0i}^{1/2}\left\|\frac{\partial\epsilon_{t-k-i}(\vartheta_0)}{\partial\vartheta_j}\right\|_2 < \infty.
\end{aligned} \tag{4.49}$$

The first inequality in (4.48) follows.

The existence of the second expectation in (4.48) is a straightforward consequence of (4.46), the Cauchy–Schwarz inequality, and the fact that ϵ_t and its derivatives are square integrable.

To deal with the second-order cross derivatives of σ_t^2, first note that $(\partial^2\sigma_t^2/\partial\vartheta\partial\omega)(\varphi_0) = 0$ by (4.15). Next, in view of (4.45),

$$\mathrm{E}\varphi_0\left|\frac{\partial^2\sigma_t^2}{\partial\vartheta_j\partial\alpha_i}(\varphi_0)\right| = \mathrm{E}\varphi_0\left|2\sum_{k=0}^{\infty}B^k(1,\,1)\epsilon_{t-k-i}\frac{\partial\epsilon_{t-k-i}}{\partial\vartheta_j}(\varphi_0)\right| < \infty \tag{4.50}$$

and, arguing as in the proof of (4.18),

$$\beta_{0\ell}\underline{\mathrm{E}}\varphi_0\left|\frac{\partial^2\sigma_t^2}{\partial\vartheta_j\partial\beta_l}(\varphi_0)\right| \leqslant \mathrm{E}\varphi_0\left|\sum_{k=1}^{\infty} kB^k(1,\,1)\sum_{i=1}^{q}2\alpha_{0i}\epsilon_{t-k-i}\frac{\partial\epsilon_{t-k-i}}{\partial\vartheta_j}(\varphi_0)\right| < \infty, \tag{4.51}$$

from which the existence of the third expectation in (4.48) follows. Hence the proofs of (i) is complete.

Assume that $\mathcal{I}$ is singular. Then there exists a vector $\lambda \neq 0$ such that $\lambda^{\mathrm{T}}\partial\ell_t(\varphi_0)/\partial\varphi^{\mathrm{T}} = 0$ a.s. From (4.12) and (4.40), we deduce that

$$(1-\eta_t^2)\frac{1}{\sigma_t^2(\varphi_0)}\lambda^{\mathrm{T}}\frac{\partial\sigma_t^2(\varphi_0)}{\partial\varphi} + \frac{2\eta_t}{\sigma_t(\varphi_0)}\lambda^{\mathrm{T}}\frac{\partial\epsilon_t(\vartheta_0)}{\partial\varphi} = 0 \text{ a.s.} \tag{4.52}$$

Taking (a representation of) the variance of the left-hand side conditional on the σ-field generated by $\{\eta_u,\, u < t\}$, we obtain almost surely, at $\varphi = \varphi_0$,

$$\begin{aligned} 0 &= (\kappa_\eta - 1)\left(\frac{1}{\sigma_t^2}\lambda^{\mathrm{T}}\frac{\partial\sigma_t^2}{\partial\varphi}\right)^2 - 2\nu_\eta\frac{1}{\sigma_t^2}\lambda^{\mathrm{T}}\frac{\partial\sigma_t^2}{\partial\varphi}\frac{2}{\sigma_t}\lambda^{\mathrm{T}}\frac{\partial\epsilon_t}{\partial\varphi} + \left(\frac{2}{\sigma_t}\lambda^{\mathrm{T}}\frac{\partial\epsilon_t}{\partial\varphi}\right)^2 \\ &:= (\kappa_\eta - 1)a_t^2 - 2\nu_\eta a_t b_t + b_t^2, \end{aligned}$$

where $\nu_\eta = \mathrm{E}(\eta_t^3)$. Since the right-hand side can be written as $(\kappa_\eta - 1 - \nu_\eta^2)a_t^2 + (b_t - \nu_\eta a_t)^2$, we find that $\kappa_\eta - 1 - \nu_\eta^2$ must be negative, and $b_t = a_t\{\nu_\eta \pm (\nu_\eta^2 + 1 - \kappa_\eta)^{1/2}\}$ a.s. By stationarity, we have either $b_t = a_t\{\nu_\eta - (\nu_\eta^2 + 1 - \kappa_\eta)^{1/2}\}$ a.s. for all t, or $b_t = a_t\{\nu_\eta + (\nu_\eta^2 + 1 - \kappa_\eta)^{1/2}\}$ a.s. for all t. Considering for instance the latter case, and turning to (4.52), we find that $a_t[1 - \eta_t^2 + \{\nu_\eta + (\nu_\eta^2 + 1 - \kappa_\eta)^{1/2}\}\eta_t] = 0$ a.s. But the term in brackets cannot cancel almost surely because, otherwise, η_t would take at most two different values, thus contradicting assumption (A12). Hence $a_t = 0$ a.s. and thus $b_t = 0$ a.s. We have proved that, almost surely,

$$\lambda^{\mathrm{T}}\frac{\partial\epsilon_t(\varphi_0)}{\partial\varphi} = \lambda_1^{\mathrm{T}}\frac{\partial\epsilon_t(\vartheta_0)}{\partial\vartheta} = 0 \quad\text{and}\quad \lambda^{\mathrm{T}}\frac{\partial\sigma_t^2(\varphi_0)}{\partial\varphi} = 0, \tag{4.53}$$

where λ_1 denotes the vector of the first $P+Q+1$ components of λ. By stationarity of $(\partial\epsilon_t/\partial\varphi)_t$, the first equality implies

$$0 = \lambda_1^{\mathrm{T}}\begin{pmatrix} -A_{\vartheta_0}(1) \\ c_0 - X_{t-1} \\ \vdots \\ c_0 - X_{t-P} \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-Q} \end{pmatrix} + \sum_{j=1}^{Q} b_{0j}\lambda_1^{\mathrm{T}}\frac{\partial\epsilon_{t-j}(\vartheta_0)}{\partial\vartheta} = \lambda_1^{\mathrm{T}}\begin{pmatrix} -A_{\vartheta_0}(1) \\ c_0 - X_{t-1} \\ \vdots \\ c_0 - X_{t-P} \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-Q} \end{pmatrix}.$$

In view of the minimality assumption (A9) on the ARMA representation, we concluded that $\lambda_1 = 0$. The second equality in (4.53) then reads $\lambda_2^{\mathrm{T}}\partial\sigma_t^2(\varphi_0)/\partial\theta = 0$, with obvious notation.

We have shown in the proof of Theorem 2.2 that this entails $\lambda_2 = 0$. Hence we have found a contradiction, which proves that $\mathcal{I}$ is non-singular.

From (4.13) and (4.41)–(4.42) we have

$$\mathcal{J} = \mathrm{E}\varphi_0\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\varphi}\frac{\partial\sigma_t^2}{\partial\varphi^{\mathrm{T}}}(\varphi_0)\right) + 2\mathrm{E}_{\varphi_0}\left(\frac{1}{\sigma_t^2}\frac{\partial\epsilon_t}{\partial\varphi}\frac{\partial\epsilon_t}{\partial\varphi^{\mathrm{T}}}(\varphi_0)\right).$$

We have just proved that the first matrix on the right-hand side is positive definite, whereas the second one is positive semi-definite. Hence the non-singularity of $\mathcal{J}$.

We now prove that $\mathcal{I}$ and $\mathcal{J}$ have the form given by the theorem when the distribution of η_t is symmetric. The expression for I_1 is a trivial consequence of (4.40) and $\mathrm{cov}(1-\eta_t^2, \eta_t) = 0$. Similarly, the form of I_2 follows directly from (4.12). We now turn to the non-diagonal blocks. From (4.12) and (4.40) we obtain

$$\mathrm{E}\varphi_0\left\{\frac{\partial\ell_t}{\partial\theta_i}\frac{\partial\ell_t}{\partial\vartheta_j}(\varphi_0)\right\} = \mathrm{E}(1-\eta_t^2)^2\mathrm{E}\varphi_0\left\{\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta_i}\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0)\right\}.$$

In view of (4.15), (4.16), (4.45) and Lemma 4.1, we have

$$\mathrm{E}\varphi_0\left\{\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\omega}\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0)\right\} = \sum_{k_1,k_2=0}^{\infty} B_0^{k_1}(1,1)B_0^{k_2}(1,1)\sum_{i=1}^{q}2\alpha_{0_i}\mathrm{E}_{\varphi_0}\left\{\sigma_t^{-4}\epsilon_{t-k_2-i}\frac{\partial\epsilon_{t-k_2-i}}{\partial\vartheta_j}(\varphi_0)\right\} = 0,$$

$$\mathrm{E}\varphi_0\left\{\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\alpha_{i_0}}\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0)\right\}$$

$$= \sum_{k_1,k_2=0}^{\infty} B_0^{k_1}(1,1)B_0^{k_2}(1,1)\sum_{i=1}^{q}2\alpha_{0_i}\mathrm{E}_{\varphi_0}\left\{\sigma_t^{-4}\epsilon_{t-k_1-i_0}^2\epsilon_{t-k_2-i}\frac{\partial\epsilon_{t-k_2-i}}{\partial\vartheta_j}(\varphi_0)\right\} = 0$$

and

$$\mathrm{E}_{\varphi 0}\left\{\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\beta_{j_0}}\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0)\right\} = \sum_{k_1,k_2=0}^{\infty}\left\{\sum_{\ell=1}^{k_1}B_0^{\ell-1}B_0^{(j_0)}B_0^{k_1-\ell}\right\}(1,1)B_0^{k_2}(1,1)\sum_{i=1}^{q}2\alpha_{0i}$$

$$\times\,\mathrm{E}_{\varphi_0}\left\{\sigma_t^{-4}\left(\omega_0+\sum_{i'=1}^{q}\alpha_{0i'}\epsilon_{t-k_1-i}^2\right)\epsilon_{t-k_2-i}\frac{\partial\epsilon_{t-k_2-i}}{\partial\vartheta_j}(\varphi_0)\right\} = 0.$$

Therefore

$$\forall i,j, \qquad \mathrm{E}_{\varphi_0}\left\{\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta_i}\frac{\partial\sigma_t^2}{\partial\vartheta_j}(\varphi_0)\right\} = 0, \tag{4.54}$$

and $\mathcal{I}$ is a block-diagonal matrix.

It is now straightforward to show that $\mathcal{J}$ has the form given by the theorem. The expressions for J_1 and J_2 are trivial consequences of (4.13) and (4.41). The block-diagonality results from (4.42) and (4.54). Hence (ii) is proved.

To show (iii), we differentiate (4.37). Recalling that the initial values are fixed, we obtain

$$\frac{\partial \underline{\tilde{\sigma}}_t^2}{\partial \omega} = \sum_{k=0}^{t-1} B^k \underline{1}, \qquad \frac{\partial \underline{\tilde{\sigma}}_t^2}{\partial \alpha_i} = \sum_{k=0}^{t-1} B^k \underline{\tilde{\epsilon}}_{t-k-i}^2, \qquad \frac{\partial \underline{\tilde{\sigma}}_t^2}{\partial \beta_j} = \sum_{k=1}^{t-1} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} \underline{\tilde{c}}_{t-k},$$

in the notation introduced in (4.15)–(4.16) and (4.30)–(4.31). Analogously to (4.45),

$$\frac{\partial \tilde{\sigma}_t^2}{\partial \vartheta_j} = \sum_{k=0}^{t-1} B^k(1, 1) \sum_{i=1}^{q} 2\alpha_i \tilde{\epsilon}_{t-k-i} \frac{\partial \tilde{\epsilon}_{t-k-i}}{\partial \vartheta_j}.$$

By a trivial extension (4.38) we have

$$\sup_{\varphi \in \Phi} \max \left\{ |\epsilon_k - \tilde{\epsilon}_k|, \left| \frac{\partial \epsilon_k}{\partial \vartheta_j} - \frac{\partial \tilde{\epsilon}_k}{\partial \vartheta_j} \right| \right\} \leqslant K\rho^k \text{ a.s.} \tag{4.55}$$

Hence

$$\begin{aligned}
\left| \frac{\partial \sigma_t^2}{\partial \vartheta_j} - \frac{\partial \tilde{\sigma}_t^2}{\partial \vartheta_j} \right| &\leqslant \left| \sum_{k=t}^{\infty} B^k(1, 1) \sum_{i=1}^{q} 2\alpha_i \epsilon_{t-k-i} \frac{\partial \epsilon_{t-k-i}}{\partial \vartheta_j} \right| \\
&\quad + \sum_{k=0}^{t-1} B^k(1, 1) \sum_{i=1}^{q} 2\alpha_i \left| (\epsilon_{t-k-i} - \tilde{\epsilon}_{t-k-i}) \frac{\partial \epsilon_{t-k-i}}{\partial \vartheta_j} + \tilde{\epsilon}_{t-k-i} \left(\frac{\partial \epsilon_{t-k-i}}{\partial \vartheta_j} - \frac{\partial \tilde{\epsilon}_{t-k-i}}{\partial \vartheta_j} \right) \right| \\
&\leqslant K\rho^t \sum_{k=0}^{\infty} \rho^k \left| \epsilon_{-k-1} \frac{\partial \epsilon_{-k-1}}{\partial \vartheta_j} \right| + K \sum_{k=0}^{t-1} \rho^k \sum_{i=1}^{q} \rho^{t-k-i} \left\{ \left| \frac{\partial \epsilon_{t-k-i}}{\partial \vartheta_j} \right| + |\tilde{\epsilon}_{t-k-i}| \right\} \\
&\leqslant K\rho^t \sum_{k=0}^{\infty} \rho^k \left| \epsilon_{-k-1} \frac{\partial \epsilon_{-k-1}}{\partial \vartheta_j} \right| + K\rho^{t/2} \sum_{k=1}^{t-1+q} \rho^{k/2} \rho^{(t-k)/2} \left\{ \left| \frac{\partial \epsilon_{t-k}}{\partial \vartheta_j} \right| + |\tilde{\epsilon}_{t-k}| \right\} \\
&\leqslant K\rho^{t/2} \sum_{k=0}^{\infty} \rho^{k/2} \left\{ \left| \epsilon_{-k-1} \frac{\partial \epsilon_{-k-1}}{\partial \vartheta_j} \right| + \left| \frac{\partial \epsilon_{k+1-q}}{\partial \vartheta_j} \right| + |\tilde{\epsilon}_{k+1-q}| \right\}.
\end{aligned}$$

It is clear that the last sum converges almost surely because its expectation is finite. Thus we have proved that

$$\sup_{\varphi \in \Phi} \left| \frac{\partial \sigma_t^2}{\partial \vartheta_j} - \frac{\partial \tilde{\sigma}_t^2}{\partial \vartheta_j} \right| \leqslant K\rho^t \text{ a.s.}$$

The other derivatives of σ_t^2 are handled in the same way, and we obtain

$$\sup_{\varphi \in \Phi} \left\| \frac{\partial \sigma_t^2}{\partial \varphi} - \frac{\partial \tilde{\sigma}_t^2}{\partial \varphi} \right\| \leqslant K\rho^t \text{ a.s.}$$

In view of (4.39) we have

$$\left| \frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2} \right| = \left| \frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2 \tilde{\sigma}_t^2} \right| \leqslant \frac{K}{\sigma_t^2} \rho^t S_{t-1}, \qquad \frac{\sigma_t^2}{\tilde{\sigma}_t^2} \leqslant 1 + K\rho^t S_{t-1},$$

where $S_{t-1} = \sum_{k=1-q}^{t-1} (|\epsilon_k| + 1)$. It is also straightforward to verify that, for $\varphi = \varphi_0$,

$$|\tilde{\epsilon}_t^2 - \epsilon_t^2| \leqslant K\rho^t(1+\sigma_t\eta_t), \qquad \left|1 - \frac{\tilde{\epsilon}_t^2}{\tilde{\sigma}_t^2}\right| \leqslant 1 + \eta_t^2 + K\rho^t(1 + |\eta_t|S_{t-1} + \eta_t^2 S_{t-1}).$$

Hence, using (4.55),

$$\begin{aligned}\left|\frac{\partial \ell_t(\varphi_0)}{\partial \varphi_i} - \frac{\partial \tilde{\ell}_t(\varphi_0)}{\partial \varphi_i}\right| = \Bigg|&\{\tilde{\epsilon}_t^2 - \epsilon_t^2\}\left\{\frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi_i}\right\} + \epsilon_t^2\left\{\frac{1}{\tilde{\sigma}_t^2} - \frac{1}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi_i}\right\} \\ &+ \left\{1 - \frac{\tilde{\epsilon}_t^2}{\tilde{\sigma}_t^2}\right\}\left\{\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial \sigma_t^2}{\partial \varphi_i}\right\} + \left\{1 - \frac{\tilde{\epsilon}_t^2}{\tilde{\sigma}_t^2}\right\}\left\{\frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial \sigma_t^2}{\partial \varphi_i} - \frac{\partial \tilde{\sigma}_t^2}{\partial \varphi_i}\right\} \\ &+ 2\{\epsilon_t - \tilde{\epsilon}_t\}\left\{\frac{1}{\sigma_t^2}\right\}\left\{\frac{\partial \epsilon_t}{\partial \varphi_i}\right\} + 2\tilde{\epsilon}_t\left\{\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial \epsilon_t}{\partial \varphi_i}\right\} \\ &2\tilde{\epsilon}_t\left\{\frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial \epsilon_t}{\partial \varphi_i} - \frac{\partial \tilde{\epsilon}_t}{\partial \varphi_i}\right\}\Bigg|(\varphi_0) \\ \leqslant K\rho^t\{1 + S_{t-1}^2(|\eta_t| + \eta_t^2)\}&\left\{1 + \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi_i} + \left|\frac{\partial \epsilon_t}{\partial \varphi_i}\right|\right\}(\varphi_0).\end{aligned}$$

We have, invoking the independence between η_t and S_{t-1}, (4.14), (4.49), the Cauchy–Schwarz inequality, and $\mathrm{E}(\epsilon_t^4) < \infty$,

$$\begin{aligned}&P\left(\left|n^{-1/2}\sum_{t=1}^n\left\{\frac{\partial \ell_t(\varphi_0)}{\partial \varphi_i} - \frac{\partial \tilde{\ell}_t(\varphi_0)}{\partial \varphi_i}\right\}\right| > \epsilon\right) \\ &\leqslant \frac{K}{\epsilon}\left\|1 + \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi_i}(\varphi_0) + \left|\frac{\partial \epsilon_t}{\partial \varphi_i}\right|(\varphi_0)\right\|_2 n^{-1/2}\sum_{t=1}^n \rho^t\{1 + (\|\eta_t^2\|_2 + \|\eta_t\|_2)\|S_{t-1}^2\|_2\} \\ &\leqslant \frac{K}{\epsilon}\left\|1 + \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^{2*}}{\partial \varphi_i}(\varphi_0) + \left|\frac{\partial \epsilon_t}{\partial \varphi_i}\right|(\varphi_0)\right\|_2 n^{-1/2}\sum_{t=1}^n \rho^t t^2 \to 0,\end{aligned}$$

which shows the first convergence in (iii). The same arguments can be used to establish the second convergence in (iii).

The proof of (iv) readily follows from the arguments given to establish (v) in the proof of Theorem 2.2.

Now let us consider (v). For pure GARCH models, we used a Taylor expansion of the second-order derivatives of the criterion, and we showed its uniform integrability in a neighbourhood of θ_0. Without additional moment assumptions, the method fails in the ARMA-GARCH framework because variables of the form $\sigma_t^{-2}(\partial\sigma_t^2/\partial\vartheta)$ do not always admit moments of any order (see (3.3)).

First, notice that, since we have shown that $\mathcal{J}$ exists, the ergodic theorem entails

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n\frac{\partial^2 \ell_t(\varphi_0)}{\partial\varphi\partial\varphi^{\mathrm{T}}} = \mathcal{J} \text{ a.s.}$$

In view of the already established consistency of $\hat{\phi}_n$, it will therefore be sufficient to show that, for any $\epsilon > 0$, there exists a neighbourhood $\mathcal{V}(\varphi_0)$ of φ_0 such that, almost surely

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left\| \frac{\partial^2 \ell_t(\varphi)}{\partial\varphi\partial\varphi^{\mathrm{T}}} - \frac{\partial^2 \ell_t(\varphi_0)}{\partial\varphi\partial\varphi^{\mathrm{T}}} \right\| \leqslant \epsilon. \tag{4.56}$$

We begin by proving that there exists $\mathcal{V}(\varphi_0)$ such that

$$\mathrm{E} \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left\| \frac{\partial^2 \ell_t(\varphi)}{\partial\varphi\partial\varphi^{\mathrm{T}}} \right\| < \infty. \tag{4.57}$$

In view of the Hölder inequality, (4.13), (4.41) and (4.42), it suffices to prove that, for any neighbourhood $\mathcal{V}(\varphi_0) \subset \Phi$ whose elements have their components α_i and β_j bounded away from zero, the norms

$$\left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \epsilon_t^2 \right\|_2, \qquad \left\| \sup_{\varphi\mathcal{V}(\varphi_0)} \left| \frac{\partial \epsilon_t}{\partial \vartheta} \right| \right\|_4, \qquad \left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left| \frac{\partial^2 \epsilon_t}{\partial\vartheta\partial\vartheta^{\mathrm{T}}} \right| \right\|_4, \tag{4.58}$$

$$\left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \frac{1}{\sigma_t^2} \right\|_\infty, \qquad \left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta} \right| \right\|_4, \qquad \left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \right| \right\|_4, \tag{4.59}$$

$$\left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left| \frac{\partial^2 \sigma_t^2}{\partial\vartheta\partial\vartheta^{\mathrm{T}}} \right| \right\|_2, \qquad \left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left| \frac{\partial^2 \sigma_t^2}{\partial\theta\partial\vartheta^{\mathrm{T}}} \right| \right\|_2, \qquad \left\| \sup_{\varphi\in\mathcal{V}(\varphi_0)} \left| \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial\theta\partial\theta^{\mathrm{T}}} \right| \right\|_2 \tag{4.60}$$

are finite. The norms in (4.58) are clearly finite, in view of $\epsilon_t(\vartheta) = A_\vartheta(B)B_\vartheta^{-1}(B)A_{\vartheta_0}^{-1}(B)B_{\vartheta_0}(B)\epsilon_t(\vartheta_0)$, similar expressions for the derivatives, and $\|\epsilon_t(\vartheta_0)\|_4 < \infty$. The finiteness of the first norm in (4.59) is an obvious consequence of $\sigma_t^2 \geqslant \inf_{\phi\in\Phi}\omega$, the last term being positive by the compactness of Φ.

It is clear that inequality (4.49) can be extended to obtain

$$\sup_{\varphi\in\Phi} \left\| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \vartheta_j}(\varphi) \right\|_4 \leqslant K \sum_{k=0}^{\infty} \rho^{k/2} \sup_{\varphi\in\Phi} \left\| \frac{\partial \epsilon_{t-k}}{\partial \vartheta_j} \right\|_4 < \infty.$$

On the other hand, because (4.15)–(4.18) remain valid when ϵ_t is replaced by $\epsilon_t(\vartheta)$, it can be seen that

$$\sup_{\varphi\in\mathcal{V}(\varphi_0)} \left\| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta}(\varphi) \right\|_d < \infty$$

for any $d > 0$ and any neighbourhood $\mathcal{V}(\varphi_0)$ whose elements have their components α_i and β_j bounded away from zero. Hence the finiteness of the norms in (4.59).

The existence of the first norm in (4.60) is a consequence of (4.46) and (4.58). To deal with the second norm we use (4.50), (4.51), (4.58), and the fact that $\sup_{\varphi\in\mathcal{V}(\varphi_0)}\beta_j^{-1} < \infty$. Finally, we show that the third norm is finite by (4.21), (4.22) and arguments already used.

Hence (4.57) is proved. Now the ergodic theorem shows that the left-hand term in (4.56) converges almost surely to

$$\mathrm{E} \sup_{\varphi \in \mathcal{V}(\varphi_0)} \left\| \frac{\partial^2 \ell_t(\varphi)}{\partial \varphi \partial \varphi^{\mathrm{T}}} - \frac{\partial^2 \ell_t(\varphi_0)}{\partial \varphi \partial \varphi^{\mathrm{T}}} \right\|.$$

This expectation decreases to 0 when the neighbourhood $\mathcal{V}(\varphi_0)$ decreases to the singleton $\{\varphi_0\}$. Thus (4.56) holds and (v) is proved, which completes the proof of Theorem 3.2.

Acknowledgement

We are very grateful to the anonymous referees for their comments.

References

Andrews, D.W.K. (1999) Estimation when a parameter is on a boundary. *Econometrica*, **67**, 1341–1384.

Berkes, I. and Horváth, L. (2003a) The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* To appear.

Berkes, I. and Horváth, L. (2003b) The rate of consistency of the quasi-maximum likelihood estimator. *Statist. Probab. Lett.*, **61**, 133-143.

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

Billingsley, P. (1961) The Lindeberg–Levy theorem for martingales. *Proc. Amer. Math. Soc.*, **12**, 788–792.

Billingsley, P. (1995) *Probability and Measure*. New York: Wiley.

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

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

Boussama, F. (1998) Ergodicité, mélange et estimation dans les modèles GARCH. Doctoral thesis, Université Paris 7.

Boussama, F. (2000) Normalité asymptotique de l'estimateur du pseudo-maximum de vraisemblance d'un modèle GARCH. *C. R. Acad, Sci. Paris*, **331**, 81–84.

Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods*. New York: Springer-Verlag.

Chen, M. and An, H.Z. (1998) A note on the stationarity and the existence of moments of the GARCH model. *Statist. Sinica*, **8**, 505–510.

Davis, R.A., Knight, K. and Liu, J. (1992) *M*-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.*, **40**, 145–180.

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

Kingman, J.F.C. (1973) Subadditive ergodic theory. *Ann. Probab.*, **6**, 883–909.

Kokoszka, P.S. and Taqqu, M.S. (1996) Parameter estimation for infinite variance fractional ARIMA. *Ann. Statist.*, **24**, 1880–1913.

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

Ling, S. and Li, W.K. (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, **92**, 1184–1194.

Ling, S. and Li, W.K. (1998) Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Ann Statist.*, **26**, 84–125.

Ling, S. and Li, W.K. (2003) Asymptotic inference for unit root processes with GARCH(1, 1) errors. *Econometric Theory*, **19**, 541–564.

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

Ling, S. and McAleer, M. (2003b) On adaptative estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.*, **31**, 642–674.

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

Mikosch, T. (2001) Modeling financial time series. Notes for lecture on 'Modeling heavy tails and dependence in financial data' presented to the meeting on 'New Directions in Time Series Analysis', Centre International de Rencontres Mathématiques, Luminy, France.

Mikosch, T., Gadrich, T., Klüppelberg, C. and Adler, R.J. (1995) Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.*, **23**, 305–326.

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

Straumann, D. and Mikosch, T. (2003) Quasi-maximum-likelihood estimation in heteroscedastic time series: a stochastic recurrence equation approach. Technical report, University of Copenhagen.

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

Received January 2003 and revised September 2003