Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 1265–1284
stochastic processes and their applications
www.elsevier.com/locate/spa

Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero

Christian Francq[a], Jean-Michel Zakoian[b,*]

[a] *EQUIPPE-GREMARS, UFR MSES, Université Lille 3, Domaine du Pont de bois, BP 60149, 59653 Villeneuve d'Ascq, Cedex, France*

[b] *EQUIPPE-GREMARS and CREST, 15 Boulevard G. Péri, 92245 Malakoff, Cedex, France*

Received 13 February 2006; received in revised form 8 December 2006; accepted 11 January 2007
Available online 21 January 2007

Abstract

The asymptotic distribution of the quasi-maximum likelihood (QML) estimator is established for generalized autoregressive conditional heteroskedastic (GARCH) processes, when the true parameter may have zero coefficients. This asymptotic distribution is the projection of a normal vector distribution onto a convex cone. The results are derived under mild conditions. For an important subclass of models, no moment condition is imposed on the GARCH process. The main practical implication of these results concerns the estimation of overidentified GARCH models.

MSC: 62M10; 62F12

Keywords: Boundary of the parameter space; Conditional heteroskedasticity; GARCH model; Quasi-maximum likelihood estimation; Non-normal asymptotic distribution

1. Introduction

Much attention has been given recently to the asymptotic properties of the quasi-maximum likelihood estimator (QMLE) in the context of GARCH processes. Whereas ARCH (AutoRegressive Conditionally Heteroskedastic) models were introduced by Engle in 1982 [11], and generalized by Bollerslev in 1986 [7], it took about twenty years to see the emergence

* Corresponding address: CREST and Université Lille 3, Statistics, 15 Boulevard G. Péri, 92245 Malakoff, France. Tel.: +33 141177666; fax: +33 141177725.

E-mail address: zakoian@ensae.fr (J.-M. Zakoian).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2007.01.001

of consistency and asymptotic normality results for the general GARCH model under weak assumptions. Recent references dealing with the QML estimation of general GARCH(p, q) are the dissertation by Boussama [9], the monograph by Straumann [25] and the papers by Berkes and Horváth [4,3], Berkes, Horváth and Kokoszka [5], Hall and Yao [14] for GARCH models with heavy tailed errors, and Francq and Zakoian [12] (hereafter FZ). Using the QMLE in the GARCH framework is beneficial for it is much less sensitive with respect to heavy tailed unconditional distributions than, for instance, the least-squares method. Other estimation procedures which are not demanding in terms of unconditional moments have recently been suggested by Horváth and Liese [15] and Ling [18]. See Straumann and Mikosch [26], Ling and McAleer [20] for related work. See Ling and McAleer [19], Giraitis, Leipus and Surgailis [13] for recent surveys on theoretical results for GARCH models.

The GARCH estimation theory however suffers the major weakness of excluding the presence of zero coefficients in the true parameter value. Indeed, one important difference between GARCH and other popular time series models, such as ARMA models, is that the admissible parameter space needs to be inequality restricted. The data generation mechanism requires the conditional variance to be always strictly positive, which is generally obtained by imposing a strictly positive intercept and non-negative GARCH coefficients in the conditional variance equation (see however Nelson and Cao [23] for weaker, but generally non-explicit, conditions). A key regularity condition, imposed by the above cited papers to establish the asymptotic normality, is that the true parameter must lie in the interior of the parameter space. For instance, it is easily seen that the QMLE is not asymptotically Gaussian if, for instance, a GARCH(p, q) model is estimated when the underlying process is a GARCH$(p-1, q)$, or a GARCH$(p, q-1)$ process.

Our aim in this paper is to derive the asymptotic distribution of the QML estimator under, if possible, the same mild conditions as are employed when the parameter is in the interior of the parameter space. The quasi-likelihood will be approximated by a quadratic function, and the asymptotic distribution will be obtained as the projection of a normal vector onto a convex cone. A quadratic approximation to the objective function and its optimization on a convex cone have been used by Chernoff [10] and Andrews [1] among many others (see the latter paper for a list of references). To our knowledge, when the parameter is on the boundary, asymptotic results for the general GARCH(p, q), or even for the GARCH(1, 1), are not available in the literature. Partial results can be found in Andrews [1,2] and Jordan [16].

The article proceeds as follows. Section 2 describes the estimation problem of concern and recalls results available when θ_0 is not on the boundary. Section 3 establishes the asymptotic distribution of the QMLE when θ_0 is on the boundary. For a large class of GARCH models, the results are obtained without moment assumptions on the observed process. Section 4 shows how to practically compute the asymptotic distribution. Proofs are relegated to an Appendix.

For a matrix A of generic term $A(i, j)$ we use the norm $\|A\| = \sum |A(i, j)|$. The spectral radius of a square matrix A is denoted by $\rho(A)$. The symbols $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{P}$ denote the convergences in distribution and in probability.

2. Assumptions and preliminary results

Consider the GARCH(p, q) model:

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

where (η_t) is a sequence of iid random variables such that $E\eta_t^2 = 1$, $\omega_0 > 0$, $\alpha_{0i} \geq 0$ $(i = 1, \ldots, q)$, $\beta_{0j} \geq 0$ $(j = 1, \ldots, p)$. A strictly stationary solution (ϵ_t) is called non-anticipative if ϵ_t is a measurable function of the $\eta_{t-i}, i \geq 0$. Let $(\epsilon_1, \ldots, \epsilon_n)$ be a realization of length n of a non-anticipative strictly stationary solution (ϵ_t) to Model (1). The vector of parameters is $\theta = (\theta_1, \ldots, \theta_{p+q+1})' = (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$ and it belongs to a parameter space $\Theta \subset (0, +\infty) \times [0, \infty)^{p+q}$. The true parameter value is denoted by $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})' \in \Theta$.

Bougerol and Picard [8] showed that a unique non-anticipative strictly stationary solution (ϵ_t) to Model (1) exists if and only if the sequence of matrices $\mathbf{A_0} = (A_{0t})$ has a strictly negative top Lyapunov exponent, $\gamma(\mathbf{A_0}) < 0$, where

$$\gamma(\mathbf{A_0}) = \lim_{t\to\infty} \quad \text{a.s. } \frac{1}{t} \log \|A_{0t} A_{0t-1} \ldots A_{01}\|,$$

$\|\cdot\|$ denoting any norm on the space of the $(p+q) \times (p+q)$ matrices, and

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

with I_k being the $k \times k$ identity matrix.

Conditionally on initial values $\epsilon_0^2, \ldots, \epsilon_{1-q}^2, \tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-p}^2$, the Gaussian quasi-likelihood is given by

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

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

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

The parameter space Θ is a compact subset of $[0, \infty)^{p+q+1}$ that bounds the first component away from zero. We will also assume throughout that Θ contains some hypercube of the form $[\underline{\omega}, \overline{\omega}] \times [0, \varepsilon]^{p+q}$, for some $\varepsilon > 0$ and $\overline{\omega} > \underline{\omega} > 0$.

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

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\min_{\theta\in\Theta} \tilde{\mathbf{l}}_n(\theta), \tag{2}$$

where

$$\tilde{\mathbf{l}}_n(\theta) = n^{-1} \sum_{t=1}^{n} \tilde{\ell}_t, \quad \text{and} \quad \tilde{\ell}_t = \tilde{\ell}_t(\theta) = \tilde{\ell}_t(\theta; \epsilon_n, \ldots, \epsilon_1) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log \tilde{\sigma}_t^2.$$

Notice that $\tilde{\ell}_t$ may depend on the whole set of observations since it is customary to choose the empirical mean of the squared observations for the initial values. An ergodic and stationary approximation $(\ell_t(\theta))$ of the sequence $(\tilde{\ell}_t(\theta))$ is obtained as follows. Under the condition A2 below, denote by $\left(\sigma_t^2\right) = \left\{\sigma_t^2(\theta)\right\}$ the strictly stationary, ergodic and non-anticipative solution of

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

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

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

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 obtain the asymptotic properties of the QMLE in the classical case where θ_0 is not on the boundary, the following assumptions can be made.

A1: $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.
A2: $\gamma(\mathbf{A_0}) < 0$ and $\sum_{j=1}^{p} \beta_j < 1, \forall \theta \in \Theta$.
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 $\alpha_{0q} + \beta_{0p} \neq 0$.
A5: $\kappa_\eta := E\eta_t^4 < \infty$.

It is worth noting that, in A2, the strict stationarity condition is imposed on the true value only. For $\theta \neq \theta_0$ we only require the weaker condition that $\sum_{j=1}^{p} \beta_j < 1$ (see e.g. Kazakevičius and Leipus [17], Th. 2.2). One important consequence of $\gamma(\mathbf{A_0}) < 0$ is that $E\epsilon_t^{2s} < \infty$ for some $s \in (0, 1)$. For a proof of this statement see Nelson [22] and Berkes et al. [5, Lemma 2.3]. For detailed comments on these assumptions, and comparisons with similar conditions given in the aforementioned papers, see FZ, in which the following result is established.

Theorem 1. *Let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (2). Then*

(i) *if* A2–A4 *hold, almost surely $\hat{\theta}_n \to \theta_0$, as $n \to \infty$,*
(ii) *if* A1–A5 *hold, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\kappa_\eta - 1)J^{-1})$, where*

$$J := E_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\right). \tag{3}$$

In the next section we will allow true parameter values belonging to $\partial\Theta := \{\theta_0 \in \Theta : \theta_{0i} = 0,$ for some $i > 0\}$. To prevent θ_0 from reaching the upper bound of Θ we define $\theta_0(\varepsilon)$ as the vector obtained by replacing all zero coefficients of θ_0 by ε and we make the following assumption.

A6: $\theta_0(\varepsilon) \in \overset{\circ}{\Theta}$ for some $\varepsilon > 0$.

For instance, if the parameter space is specified as $\Theta = [\underline{\omega}, \overline{\omega}] \times [0, \overline{\alpha}_1] \times \cdots \times [0, \overline{\alpha}_q] \times [0, \overline{\beta}_1] \times \cdots \times [0, \overline{\beta}_p]$, Assumption A6 is satisfied when $\overline{\omega} > \omega_0 > \underline{\omega} > 0$ and $0 \leq \theta_0 < \overline{\theta} := (\overline{\omega}, \overline{\alpha}_1, \ldots, \overline{\beta}_p)'$.

3. Asymptotic distribution of $\hat{\theta}_n$ when θ_0 is on the boundary

It is easy to understand why the positivity condition, namely $\alpha_{0i} > 0$ $(i = 1, \ldots, q)$, $\beta_{0j} > 0$ $(j = 1, \ldots, p)$, is crucial for the asymptotic normality of the QMLE $\hat{\theta}_n$. Obviously, a Gaussian asymptotic distribution for $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is precluded when the components $\hat{\theta}_{in}$ of $\hat{\theta}_n$ are constrained to be non-negative and $\theta_0 \in \partial\Theta$. If, for instance, $\theta_{0i} = 0$ then $\sqrt{n}(\hat{\theta}_{in} - \theta_{0i}) = \sqrt{n}\hat{\theta}_{in} \geq 0$ for all n and the asymptotic distribution of this variable cannot be a standard Gaussian.

By Theorem 1, no moment assumption is required for the asymptotic distribution to hold, and thus for the existence of J, when θ_0 is an interior point of Θ. When $\theta_0 \in \partial\Theta$, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ will to be seen to rely on the existence of J as well. Before deriving this asymptotic distribution, we give an example showing that the matrix J may not exist if $E_{\theta_0}\epsilon_t^4 = \infty$ and A1 is relaxed.

3.1. *Possible non-existence of J under A2–A5*

Consider the ARCH(2) model $\epsilon_t = \sigma_t\eta_t$, $\sigma_t^2 = \omega_0 + \alpha_{01}\epsilon_{t-1}^2 + \alpha_{02}\epsilon_{t-2}^2$ where $\omega_0 > 0$, $\alpha_{01} \geq 0$, $\alpha_{02} = 0$, and the distribution of the iid sequence (η_t) is defined, for $a > 1$, by

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

This ARCH(2) model is used to generate the quasi-likelihood function but ϵ_t is in fact an ARCH(1). The strict stationarity condition $\gamma(\mathbf{A_0}) < 0$ takes the form $\alpha_{01} < \exp\left\{-E(\log\eta_t^2)\right\}$ for an ARCH(1). The process (ϵ_t) is therefore strictly stationary for any value of α_{01} since $\exp\left\{-E(\log\eta_t^2)\right\} = +\infty$. However ϵ_t is not second-order stationary when $\alpha_{01} \geq 1$.

We have

$$\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\alpha_2}(\theta_0) = \frac{\epsilon_{t-2}^2}{\omega_0 + \alpha_{01}\epsilon_{t-1}^2},$$

whence

$$\begin{aligned} E_{\theta_0}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\alpha_2}(\theta_0)\right\}^2 &\geq E_{\theta_0}\left[\left\{\frac{\epsilon_{t-2}^2}{\omega_0 + \alpha_{01}\epsilon_{t-1}^2}\right\}^2 \middle| \eta_{t-1} = 0\right] P(\eta_{t-1} = 0) \\ &= \frac{1}{\omega_0^2}\left(1 - \frac{1}{a^2}\right) E_{\theta_0}(\epsilon_{t-2}^4) \end{aligned}$$

firstly because $\eta_{t-1} = 0$ entails $\epsilon_{t-1} = 0$ and secondly because η_{t-1} and ϵ_{t-2} are independent. It follows that J does not exist if $E_{\theta_0}\epsilon_t^4 = \infty$.

3.2. *Assumptions and main result*

It is then clear that the assumptions of Theorem 1 are not sufficient to ensure the existence of J when A1 is relaxed. In view of these remarks we introduce two alternative assumptions. The first one is a moment condition.

A7: $E_{\theta_0}\epsilon_t^6 < \infty$.

In many interesting cases, except the ARCH(q) models, no moment assumption on ϵ_t^2 will be required. Indeed, it will be sufficient to ensure the existence of moments for the score vector normalized by σ_t^2. Note that under the condition $\gamma(\mathbf{A_0}) < 0$, the strictly stationary solution $\sigma_t^2(\theta_0)$ has an expansion of the form: $\sigma_t^2(\theta_0) = c_0 + \sum_{j=1}^{\infty} b_{0j}\epsilon_{t-j}^2$ with $c_0 > 0$, $b_{0j} \geq 0$. Similar expansions hold for the derivatives (see the proof of Lemma 8 below). The control of moments of $\{\partial\sigma_t^2/\partial\theta\}/\sigma_t^2$ will rely on the fact that every term ϵ_{t-j}^2 appearing in the numerator of this ratio is also present in the denominator. We therefore consider the assumption

A8: $b_{0j} > 0$ for all $j \geq 1$, where $\sigma_t^2(\theta_0) = c_0 + \sum_{j=1}^{\infty} b_{0j}\epsilon_{t-j}^2$.

It should be noted that a simple sufficient condition for A8 is $\alpha_{01} > 0$ and $\beta_{01} > 0$ (because $b_{0j} \geq \alpha_{01}\beta_{01}^{j-1}$). A necessary condition is obviously that $\alpha_{01} > 0$ (because $b_{01} = \alpha_{01}$). More generally, a necessary and sufficient condition for A8 is

$$\{j \mid \beta_{0,j} > 0\} \neq \emptyset \quad \text{and} \quad \prod_{i=1}^{j_0} \alpha_{0i} > 0 \quad \text{for } j_0 = \min\{j \mid \beta_{0,j} > 0\}, \tag{4}$$

meaning that the ARCH coefficients α cannot cancel up to the order j_0 of the first GARCH coefficient β equal to zero. Assumption A8 does not apply to ARCH(q) models, which is not surprising in view of the example in Section 3.1. The main result of this section is the following.

Theorem 2. *Let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (2). Then if* A2–A6 *and either* A7 *or* A8 *hold,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \lambda^{\Lambda} := \arg\inf_{\lambda \in \Lambda} \{\lambda - Z\}' J \{\lambda - Z\},$$
$$\text{with } Z \sim \mathcal{N}(0, (\kappa_\eta - 1)J^{-1}), \qquad \Lambda = \Lambda(\theta_0) = \Lambda_1 \times \cdots \times \Lambda_{p+q+1},$$

where $\Lambda_1 = \mathbb{R}$, and, for $i = 2, \ldots, p+q+1$, $\Lambda_i = \mathbb{R}$ if $\theta_{0i} \neq 0$ and $\Lambda_i = [0, \infty)$ if $\theta_{0i} = 0$.

In the ARCH case, the result can be stated as follows.

Corollary 3. *Let $p = 0$ and let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (2). Then if $\gamma(\mathbf{A_0}) < 0$,* A3, *and* A5–A7 *hold,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \lambda^{\Lambda} := \arg\inf_{\lambda \in \Lambda} \{\lambda - Z\}' J \{\lambda - Z\},$$
$$\text{with } Z \sim \mathcal{N}(0, (\kappa_\eta - 1)J^{-1}), \qquad \Lambda = \Lambda(\theta_0) = \Lambda_1 \times \cdots \times \Lambda_{q+1},$$

where $\Lambda_1 = \mathbb{R}$, and, for $i = 2, \ldots, q+1$, $\Lambda_i = \mathbb{R}$ if $\theta_{0i} \neq 0$ and $\Lambda_i = [0, \infty)$ if $\theta_{0i} = 0$.

Comments.

1. For $\theta_0 \in \overset{\circ}{\Theta}$, the result of this theorem reduces to that of Theorem 1. Indeed, in this case $\Lambda = \mathbb{R}^{p+q+1}$ and $\lambda^{\Lambda} = Z \sim \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right)$. Hence, Theorem 2 has interest only when θ_0 belongs to $\partial\Theta$.
2. We stress the fact that the moment condition A3 is on the iid process, not on (ϵ_t). For values of θ_0 satisfying A8, the asymptotic distribution is derived under the same mild conditions as are employed for the standard case where $\theta_0 \in \overset{\circ}{\Theta}$.
3. Andrews [1] considered, as an example of a more general framework, the case of a GARCH$(1, q)$ model with $q > 1$, assuming that the parameters α_1 and β_1 are bounded away from zero. Hence the case when α_{01} or β_{01} are on the boundary is not covered. In particular, this precludes the ARCH(q) and the GARCH(1, 1) models with coefficients equal to zero. Jordan [16] allows a parameter belonging to the boundary of a non-compact set, but is restricted to an ARCH framework and requires the moment assumption $E\epsilon_t^8 < \infty$.

4. Practical derivation of the limiting distribution

The vector λ^{Λ} appears to be the orthogonal projection of Z onto Λ, where orthogonality is defined in the metric associated with the covariance structure J (see the proof of Lemma 13 below), namely $x \perp y$ iff $x'Jy = 0$. It is uniquely determined because Λ is convex. Moreover, the fact that Λ is a convex cone whose faces are sections of subspaces allows one to obtain this projection in a more explicit way (see e.g. Perlman [24]). Suppose, without loss of generality, that the first d_1 components of θ_0 are positive, and that the last d_2 components are null, with $d_1 + d_2 = p + q + 1$. We have $\Lambda = \mathbb{R}^{d_1} \times [0, \infty)^{d_2} = \{\lambda \in \mathbb{R}^{d_1+d_2} \mid K\lambda \geq 0\}$, where $K = (0_{d_2 \times d_1}, I_{d_2})$. Let $\mathcal{K} = \{K_1, \ldots, K_{2^{d_2}-1}\}$, where the K_i are matrices obtained by cancelling 0, 1 or several (up to $d_2 - 1$) rows of K. Let $M_i = K_i' \left(K_i J^{-1} K_i'\right)^{-1} K_i$, let $P_i = I_{d_1+d_2} - J^{-1} M_i$ and denote by $\lambda_{K_i} = P_i Z$ the projection of Z onto the linear subspace of $\mathbb{R}^{d_1+d_2}$ spanned by one of the $2^{d_2} - 1$ faces of Λ (including the "face" $\mathbb{R}^{d_1} \times \{0\}^{d_2}$), defined by $K_i\lambda = 0$. Then we have, with $\mathcal{C} = \{\lambda_{K_i} : K_i \in \mathcal{K} \text{ and } K\lambda_{K_i} \geq 0\}$,

$$\begin{aligned} \lambda^{\Lambda} &= Z\mathbb{1}_{\Lambda}(Z) + \mathbb{1}_{\Lambda^c}(Z) \times \arg\min_{\lambda \in \mathcal{C}} \|\lambda - Z\|_J \\ &= Z\mathbb{1}_{\Lambda}(Z) + \sum_{i=1}^{2^{d_2}-1} P_i Z \mathbb{1}_{\mathcal{D}_i}(Z), \end{aligned} \tag{5}$$

for some partition $(\mathcal{D}_i)$ of $\mathbb{R}^d - \Lambda$.

For the sake of illustration we consider the following examples.

Example 4 (*One Zero Coefficient*). Suppose that only one component of θ_0 is zero, the other components being positive. Thus $d_2 = 1$, $\Lambda = \mathbb{R}^{d_1} \times [0, \infty)$, $K = (0, \ldots, 0, 1)$, $\mathcal{K} = \{K\}$, and, letting $Z = (Z_1, \ldots, Z_d)'$,

$$\lambda^{\Lambda} = Z\mathbb{1}_{Z_d \geq 0} + PZ\mathbb{1}_{Z_d < 0}, \quad P = I_d - J^{-1}K'(KJ^{-1}K')^{-1}K.$$

It follows that

$$\lambda^{\Lambda} = Z - Z_d^{-}\mathbf{c} \tag{6}$$

where $Z_d^{-} = Z_d \mathbb{1}_{Z_d < 0}$, and $\mathbf{c} = E(Z_d Z)/\mathrm{Var}(Z_d)$ is the last column of J^{-1} divided by the (d, d)-element of this matrix. Note that the last component of λ^{Λ} is $Z_d^{+} := Z_d \mathbb{1}_{Z_d > 0}$. Letting $\lambda^{\Lambda} = (\lambda_1^{\Lambda}, \ldots, \lambda_d^{\Lambda})'$, it is also seen that $\lambda_i^{\Lambda} = Z_i$ if and only if $\mathrm{Cov}(Z_i, Z_d) = 0$.

Example 5 (*Noise Estimated as an ARCH(1)*). To be more specific, consider Example 4 with d_1 $(=d_2) = 1$. Then $\theta_0 = (\omega_0, 0)'$ and

$$J = E_{\theta_0} \frac{1}{\sigma_t^4} \begin{pmatrix} 1 & \epsilon_{t-1}^2 \\ \epsilon_{t-1}^2 & \epsilon_{t-1}^4 \end{pmatrix} = \frac{1}{\omega_0^2} \begin{pmatrix} 1 & \omega_0 \\ \omega_0 & \omega_0^2 \kappa_\eta \end{pmatrix}, \qquad J^{-1} = \frac{1}{\kappa_\eta - 1} \begin{pmatrix} \omega_0^2 \kappa_\eta & -\omega_0 \\ -\omega_0 & 1 \end{pmatrix}.$$

Thus

$$\lambda^{\Lambda} = Z - Z_2^{-} \begin{pmatrix} -\omega_0 \\ 1 \end{pmatrix} = \begin{pmatrix} Z_1 + \omega_0 Z_2^{-} \\ Z_2^{+} \end{pmatrix}.$$

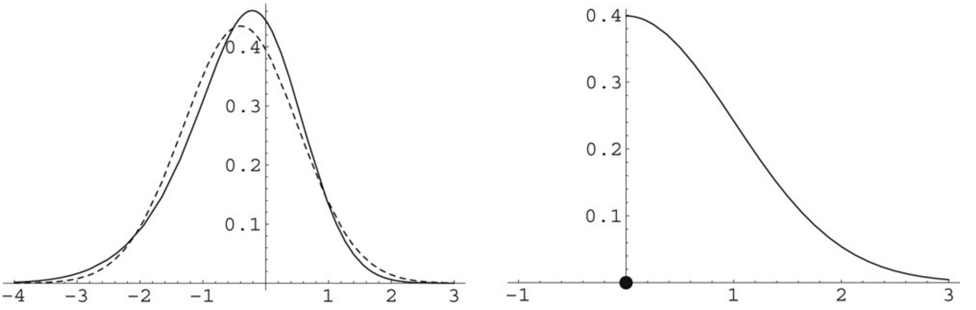

Fig. 1. Asymptotic marginal distributions of the QMLE, for the ARCH(1) model $\epsilon_t = \sqrt{\omega + \alpha\epsilon_{t-1}^2}\eta_t$ with $\omega_0 = 1$, $\alpha_0 = 0$ and $\kappa_\eta = 1.5$. The left-hand graph displays, as a full line, the density of λ_1^Λ, and as a dotted line the density of the normal distribution with the same first and second moments. The right-hand graph displays the distribution of λ_2^Λ, with the Dirac mass $1/2$ at 0 and the density $\mathcal{N}(0, 1)$ on $(0, +\infty)$.

Straightforward computations show that $E\lambda_1^\Lambda = -\omega_0 E\lambda_2^\Lambda = -\omega_0(2\pi)^{-1/2}$ and

$$\mathrm{Var}(\lambda^\Lambda) = \frac{1}{2}\left(1 - \frac{1}{\pi}\right)\begin{pmatrix} \frac{1+(1-2\kappa_\eta)\pi}{1-\pi}\omega_0^2 & -\omega_0 \\ -\omega_0 & 1 \end{pmatrix}.$$

It can be seen that $\mathrm{Var}(\lambda_1^\Lambda) = \omega_0^2\kappa_\eta - \frac{1}{2}(1 + \frac{1}{\pi})\omega_0^2 < \mathrm{Var}(Z_1) = \omega_0^2\kappa_\eta$ (see Example 6 below for a generalization). Fig. 1 displays the density of λ_1^Λ and that of the continuous part of λ_2^Λ. It is interesting to note that even the first component of λ^Λ is not Gaussian. Elementary computations show that its skewness coefficient is given by

$$\frac{E\left(\lambda_1^\Lambda - E\lambda_1^\Lambda\right)^3}{\left(\mathrm{Var}\,\lambda_1^\Lambda\right)^{3/2}} = \frac{-\left(\frac{1}{2} - \frac{1}{\pi}\right)\frac{1}{\sqrt{2\pi}}}{\left\{\kappa_\eta - \frac{1}{2}\left(1 + \frac{1}{\pi}\right)\right\}^{3/2}}.$$

Note that this skewness is always negative but vanishes when κ_η goes to infinity.

Example 6 (*Asymptotic Mean Squared Error Comparisons*). The results of Lovell and Prescott [21] for the Gaussian multiple regression model show that the Mean Squared Error (MSE) of the least-squares estimator subject to a positivity constraint is smaller than the conventional (unconstrained) MSE. As noted by Andrews [2], this property is probably true for non-standard models with a parameter on the boundary of a convex set, but this property remains a conjecture. We will verify this conjecture on Example 4. In view of (6), we have

$$\begin{aligned}\mathrm{MSE}(\lambda^\Lambda) &:= \mathrm{Var}(\lambda^\Lambda) + E\lambda^\Lambda E\lambda^{\Lambda\prime} \\ &= \mathrm{Var}(Z) + \mathrm{Var}(Z_d^-)\mathbf{c}\mathbf{c}' - 2E(Z_d^- Z)\mathbf{c}' + (EZ_d^-)^2\mathbf{c}\mathbf{c}'.\end{aligned}$$

Since $EZ_d^- Z = EZ_d^+ Z = EZ_d Z/2 = \mathrm{Var}(Z_d)\mathbf{c}/2$, $EZ_d^- = -\sqrt{\mathrm{Var}(Z_d)/2\pi}$ and $E(Z_d^-)^2 = \mathrm{Var}(Z_d)/2$, we have

$$\mathrm{MSE}(\lambda^\Lambda) - \mathrm{MSE}(Z) = \{E(Z_d^-)^2 - \mathrm{Var}(Z_d)\}\mathbf{c}\mathbf{c}' = -\frac{\mathrm{Var}(Z_d)}{2}\mathbf{c}\mathbf{c}' \tag{7}$$

showing that the difference $\mathrm{MSE}(\lambda^\Lambda) - \mathrm{MSE}(Z)$ is a semi-negative definite matrix. Thus, when one GARCH coefficient is null, the conventional asymptotic standard errors for the QML estimators of the other coefficients are too large. This may have consequences in practical applications.

Acknowledgements

The authors are grateful to Professor B.P. Pötscher who gave us helpful comments on the preliminary version of this paper which was presented at the ESEM-EEA 2006 congress. The authors are also very grateful to the Editor and the referees for their comments and constructive suggestions.

Appendix A. Proofs and technical results

A.1. Asymptotic normality of the normalized score

Before proving Theorem 2 we will establish several lemmas. Note that when $\theta_0 \in \partial\Theta$, the function $\sigma_t^2(\theta)$ may be negative in a neighborhood of θ_0 and $\ell_t(\theta)$ may be non-defined in this neighborhood. Instead of a standard Taylor expansion, we will use a one-sided expansion, based on right derivatives of $\mathbf{l}_n(\theta) = n^{-1}\sum_{t=1}^n \ell_t(\theta)$ about θ_0. For ease of notation, denote by $\partial\sigma_t^2(\theta_0)/\partial\theta := \left(\partial\sigma_t^2(\theta_0)/\partial\theta_i\right)_{i=1,\ldots,p+q+1}$ and $\partial\ell_t(\theta_0)/\partial\theta := (\partial\ell_t(\theta_0)/\partial\theta_i)_{i=1,\ldots,p+q+1}$ the vectors of partial derivatives of σ_t and ℓ_t at θ_0 with the i-th derivative replaced by the right derivative when $\theta_{0i} = 0$. We use the same convention for the derivatives of $\mathbf{l}_n$, $\tilde{\ell}_t$ and $\tilde{\mathbf{l}}_n$ at θ_0, and for the second partial derivatives. Under this convention, the derivatives of $\ell_t(\theta) = \epsilon_t^2/\sigma_t^2 + \log\sigma_t^2$ are given by

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

$$\frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} = \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right\} + \left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta'}\right\}. \tag{A.1}$$

FZ [12] introduce the following notation for σ_t^2 and its derivatives:

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

$$\frac{\partial\sigma_t^2}{\partial\omega} = \sum_{k=0}^{\infty} B^k(1,1), \qquad \frac{\partial\sigma_t^2}{\partial\alpha_i} = \sum_{k=0}^{\infty} B^k(1,1)\epsilon_{t-k-i}^2, \tag{A.3}$$

$$\frac{\partial\sigma_t^2}{\partial\beta_j} = \sum_{k=1}^{\infty} B_{k,j}(1,1)\left(\omega + \sum_{i=1}^{q}\alpha_i\epsilon_{t-k-i}^2\right) \tag{A.4}$$

where

$$B_{k,j} = \frac{\partial B^k}{\partial\beta_j} = \sum_{m=1}^{k} B^{m-1}B^{(j)}B^{k-m}, \quad B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{A.5}$$

and $B^{(j)}$ is a $p \times p$ matrix with $(1, j)$th element 1, and all other elements equal to zero. Similar formulas, given below, hold for the second derivatives. Elementary properties of the matrix B are established in the following lemma.

Lemma 7. *For $j = 1, \ldots, p$*

$$B^{(j)} B^k \le B^{k-j+1}, \quad \text{for all } k \ge j-1, \tag{A.6}$$

$$B^{(j)} B^k = B^{(j-k)}, \quad \text{for all } 0 \le k < j, \tag{A.7}$$

$$AB^{(1)} \le A, \quad \text{and} \quad \text{for } j \ne \ell_2, \quad \{AB^{(j)}\}(\ell_1, \ell_2) = 0, \quad \forall A \ge 0, \tag{A.8}$$

$$B^{(1)} B^{(j)} = B^{(j)}, \quad \text{and} \quad B^{(i)} B^{(j)} = 0, \quad \text{for } i > 1, \tag{A.9}$$

$$B^k(1,1) \ge \beta_j B^{k-j}(1,1), \quad \text{for all } k \ge j. \tag{A.10}$$

Proof. First note that, when $j \le p$, the j-th row of B^j is the first row of B. The first row of $B^{(j)}A$ is the j-th row of A, and the other elements of $B^{(j)}A$ are zeros. Thus $B^{(j)}B^j \le B$. Multiplying the two sides of the previous inequality by B^{k-j}, whose elements are non-negative, yields (A.6). For $k < j \le p$, the j-th row of B^k is null, except one "1" in the j-th position, which shows (A.7). The j-th column of $AB^{(j)}$ is the first column of A, and the other elements of $AB^{(j)}$ are zeros. Thus (A.8) and (A.9) are obvious. Inequality (A.10) comes from $B^k(1,1) = \sum_{j=1}^{p} \beta_j B^{k-1}(j,1) \ge \beta_j B^{k-1}(j,1) = \beta_j B^{k-j}(1,1)$. □

The second lemma allows to consider the L^1 norms of the derivatives of σ_t^2 at θ_0.

Lemma 8. *Under the assumptions of Theorem 2,*

$$E_{\theta_0} \left\| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta}(\theta_0) \right\| < \infty, \qquad E_{\theta_0} \left\| \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}(\theta_0) \right\| < \infty,$$

$$E_{\theta_0} \left\| \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'}(\theta_0) \right\| < \infty.$$

Proof. In this proof and the subsequent ones, K and ρ denote generic constants, whose values might change from line to line but always satisfy $K > 0$ and $0 < \rho < 1$. Since σ_t^{-2} is bounded by $1/\underline{\omega}$, the proof of Lemma 8 is straightforward under Assumption A7.

Now assume that, A8, instead of A7, holds. By (A.3), $\partial \sigma_t^2(\theta_0)/\partial \omega$ is bounded since $\sum_{k=0}^{\infty} B^k$ is finite under A2. Since $\sigma_t^2(\theta_0) \ge \omega_0 > 0$, $\{\partial \sigma_t^2(\theta_0)/\partial \omega\}/\sigma_t^2(\theta_0)$ therefore possesses moments of any order. Consider the derivatives with respect to α_i. Let B_0 be the matrix B for $\theta = \theta_0$. We have, in view of (A.2),

$$\sigma_t^2(\theta_0) = \omega_0 \sum_{k=0}^{\infty} B_0^k(1,1) + \sum_{k=1}^{\infty} \sum_{\ell=1}^{k} \alpha_{0\ell} B_0^{k-\ell}(1,1) \epsilon_{t-k}^2$$

with by convention $\alpha_{0\ell} = 0$ when $\ell \notin \{1, \ldots, q\}$. By assumption A8, for all $k > 0$ there exists an integer $i_k \in \{1, \ldots, \min(q, k)\}$ such that

$$\sum_{\ell=1}^{k} \alpha_{0\ell} B_0^{k-\ell}(1,1) \ge \alpha_{0 i_k} B_0^{k-i_k}(1,1) \ge \underline{\alpha} B_0^{k-i_k}(1,1) > 0, \tag{A.11}$$

for some positive constant $\underline{\alpha}$ (one can take $\underline{\alpha} = \min\{\alpha_{0i} : \alpha_{0i} \neq 0\}$). It follows that for any $s \in (0, 1)$, in view of (A.2) and (A.3),

$$\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \alpha_i}(\theta_0) = \frac{\sum_{k=0}^{\infty} B_0^k(1,1)\epsilon_{t-i-k}^2}{\sum_{k=0}^{\infty} B_0^k(1,1)\left(\omega_0 + \sum_{j=1}^{q}\alpha_{0j}\epsilon_{t-j-k}^2\right)}$$

$$\leq \sum_{k=i}^{\infty} \frac{B_0^{k-i}(1,1)\epsilon_{t-k}^2}{\omega_0 + \underline{\alpha} B_0^{k-i_k}(1,1)\epsilon_{t-k}^2} \leq \sum_{k=i}^{\infty} \frac{B_0^{k-i}(1,1)\epsilon_{t-k}^{2s}}{\omega_0^s \underline{\alpha}^{1-s}\{B_0^{k-i_k}(1,1)\}^{1-s}}, \tag{A.12}$$

where the last inequality follows from $ax/(b+cx) \leq ax^s/(b^s c^{1-s})$ for all $a, x \geq 0$, $b, c > 0$ and $s \in (0, 1)$. The latter inequality comes from the elementary inequality $x/(1+x) \leq x^s$ for all $x \geq 0$ and all $s \in (0, 1)$. Now, for any fixed $s \in (0, 1)$, we will show that

$$B_0^{k-i}(1,1)/\{B_0^{k-i_k}(1,1)\}^{1-s} \leq K\rho^k \quad \text{for all } k. \tag{A.13}$$

By A2 and the compactness of Θ, we have $\sup_{\theta\in\Theta}\rho(B) < 1$. Thus $\|B_0^k\| \leq K\rho^k$ for all k, and since i_k belongs to the finite set $\{1, \ldots, q\}$, we have $\left\{B_0^{k-i_k}(1,1)\right\}^s \leq K\rho^k$, and it suffices to show that $B_0^{k-i}(1,1)/B_0^{k-i_k}(1,1)$ is bounded by a constant independent of k. It is sufficient to consider k such that $B_0^{k-i}(1,1) \neq 0$. Let j_0 be defined by (4) and let $r_i \in \{1, \ldots, j_0\}$ be such that $i - 1 \equiv r_i - 1 \pmod{j_0}$, that is $i = q_i j_0 + r_i$ with $q_i \geq 0$. In view of (A.10), we have $B_0^{k-r_i}(1,1) = B_0^{k-i+q_i j_0}(1,1) \geq \beta_{0j_0}^{q_i} B_0^{k-i}(1,1) > 0$. Moreover, $\alpha_{0r_i} \neq 0$ by (4). Thus one can take $i_k = r_i$ in (A.11), so that we have

$$B_0^{k-i}(1,1)/B_0^{k-i_k}(1,1) \leq 1/\beta_{0j_0}^{q_i}, \tag{A.14}$$

and thus (A.13) holds. Then (A.12) gives

$$E_{\theta_0}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \alpha_i}(\theta_0) \leq K\left\{\sum_{k=1}^{\infty}\rho^k\right\} E_{\theta_0}\epsilon_t^{2s}. \tag{A.15}$$

Since ϵ_t^2 has a moment of order s, for some $s \in (0, 1)$, the right-hand side in the last inequality is finite. Hence $\sigma_t^{-2}(\partial\sigma_t^2/\partial\alpha_i)$ has a moment of order 1 at $\theta = \theta_0$.

Let us turn to the derivatives with respect to β_j. In view of (A.4) we have

$$\frac{\partial \sigma_t^2(\theta_0)}{\partial \beta_j} = \omega_0 \sum_{k=0}^{\infty} B_{k,j}(1,1) + \sum_{k=2}^{\infty}\sum_{\ell=1}^{k} \alpha_{0\ell} B_{k-\ell,j}(1,1)\epsilon_{t-k}^2 \tag{A.16}$$

where $B_{0,j} = 0$ and, for $k > 0$, the matrices $B_{k,j}$ defined in (A.5) are taken at θ_0. We obtain, for any $0 \leq \ell \leq k - j$, by (A.6) and (A.7),

$$B_{k-\ell,j} \leq \sum_{m=1}^{k-\ell-j} B_0^{m-1}B_0^{k-\ell-m-j+1} + \sum_{m=k-\ell-j+1}^{k-\ell} B_0^{m-1}B^{(j-k+\ell+m)}$$

$$= (k-\ell-j)B_0^{k-\ell-j} + \sum_{m=k-\ell-j+1}^{k-\ell} B_0^{m-1}B^{(j-k+\ell+m)},$$

which, together with (A.8), entails that

$$B_{k-\ell,j}(1,1) \le (k-\ell-j)B_0^{k-\ell-j}(1,1) + B_0^{k-\ell-j}(1,1) \le kB_0^{k-\ell-j}(1,1). \tag{A.17}$$

For $k-j < \ell \le k$ we similarly obtain

$$B_{k-\ell,j} \le \sum_{m=1}^{k-\ell} B_0^{m-1}B^{(j-k+\ell+m)}, \quad \text{and thus } B_{k-\ell,j}(1,1) = 0. \tag{A.18}$$

Therefore, from (A.16) we deduce

$$\frac{\partial \sigma_t^2(\theta_0)}{\partial \beta_j} \le \omega_0 \sum_{k=j}^{\infty} kB_0^{k-j}(1,1) + \sum_{k=j+1}^{\infty}\sum_{\ell=1}^{k-j} \alpha_{0\ell} kB_0^{k-\ell-j}(1,1)\epsilon_{t-k}^2.$$

Hence, proceeding as in (A.12) we get

$$\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \beta_j}(\theta_0) \le K + \sum_{k=j+1}^{\infty}\sum_{\ell=1}^{k-j} \alpha_{0\ell} k \frac{B_0^{k-\ell-j}(1,1)\epsilon_{t-k}^{2s}}{\omega_0^s \underline{\alpha}^{1-s}\{B_0^{k-i_k}(1,1)\}^{1-s}}, \tag{A.19}$$

and thus

$$E_{\theta_0}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \beta_j}(\theta_0) \le K + K\left\{\sum_{k=1}^{\infty} k\rho^{ks}\right\} E_{\theta_0}\epsilon_t^{2s} < \infty, \tag{A.20}$$

by arguments already used for (A.15). This allows to conclude that the first expectation in Lemma 8 exists. Applying the Hölder inequality in (A.12) and (A.19) with s such that $E\epsilon_t^{4s} < \infty$, it can be shown that $\|\sigma_t^{-2}\partial\sigma_t^2(\theta_0)/\partial\theta\|_2 < \infty$. Thus the second expectation in Lemma 8 exists.

Let us now turn to the second-order derivatives of σ_t^2. It follows from (A.3) that

$$\frac{\partial^2 \sigma_t^2}{\partial \omega^2} = \frac{\partial^2 \sigma_t^2}{\partial \omega \partial \alpha_i} = 0 \quad \text{and} \quad \frac{\partial^2 \sigma_t^2}{\partial \omega \partial \beta_j} = \sum_{k=1}^{\infty} B_{k,j}(1,1).$$

Thus $\partial^2\sigma_t^2/\partial\omega\partial\beta_j \le \sum_{k=j}^{\infty} kB_0^{k-j}(1,1) < \infty$, by (A.17) and (A.18) with $\ell = 0$, which proves that $\partial^2\sigma_t^2(\theta_0)/\partial\omega\partial\theta_i$ is bounded and admits moments at any order. The same conclusion holds for $\left\{\partial^2\sigma_t^2(\theta_0)/\partial\omega\partial\theta_i\right\}/\sigma_t^2(\theta_0)$. By (A.3) and (A.4) we find

$$\frac{\partial^2 \sigma_t^2}{\partial \alpha_i \partial \alpha_j} = 0 \quad \text{and} \quad \frac{\partial^2 \sigma_t^2}{\partial \alpha_i \partial \beta_j} = \sum_{k=2}^{\infty} B_{k-i,j}(1,1)\epsilon_{t-k}^2,$$

and the arguments used for the first-order derivative with respect to β_j prove that $\left\{\partial^2\sigma_t^2(\theta_0)/\partial\alpha_i\partial\theta\right\}/\sigma_t^2(\theta_0)$ is integrable. Differentiating (A.16) with respect to $\beta_{j'}$ gives

$$\frac{\partial^2 \sigma_t^2}{\partial \beta_j \partial \beta_{j'}} = \omega_0 \sum_{k=0}^{\infty} B_{k,j,j'}(1,1) + \sum_{k=2}^{\infty}\sum_{\ell=1}^{k} \alpha_{0\ell} B_{k-\ell,j,j'}(1,1)\epsilon_{t-k}^2 \tag{A.21}$$

where

$$\begin{aligned}B_{k,j,j'} &= \frac{\partial B_{k,j}}{\partial \beta_{j'}} = \sum_{m=1}^{k} B_{m-1,j'} B^{(j)} B_0^{k-m} + \sum_{m=1}^{k} B_0^{m-1} B^{(j)} B_{k-m,j'} \\ &:= B^{(1)}_{k,j,j'} + B^{(2)}_{k,j,j'}.\end{aligned}$$

We first give a bound for the terms of the form $B^{(j)} B_{k,j'}$ involved in $B^{(2)}_{k,j,j'}$. First note that when $k \le p$, only the first k rows of B_0^k contain terms depending on the β_j. Thus the last $p-k+1$ rows of $B_{k,j'}$ are equal to zero, and it follows that

$$B^{(j)} B_{k,j'} = 0 \quad \text{for } k < j. \tag{A.22}$$

Using successively (A.7), (A.6) and (A.9), we obtain, for $j, j' = 1, \dots, p$ and $k > 0$,

$$\begin{aligned}B^{(j)} B_{k,j'} &= \sum_{n=1}^{k} B^{(j)} B_0^{n-1} B^{(j')} B_0^{k-n} \\ &\le \sum_{n=1}^{j} B^{(j-n+1)} B^{(j')} B_0^{k-n} + \sum_{n=j+1}^{k} B_0^{n-j} B^{(j')} B_0^{k-n} \\ &= B^{(j')} B_0^{k-j} + \sum_{n=j+1}^{k} B_0^{n-j} B^{(j')} B_0^{k-n},\end{aligned}$$

where by convention $B_0^k = B^{(k+1)} = 0$ for $k < 0$ and $\sum_{n=k}^{k'} x_n = 0$ for $k > k'$. Using again (A.7) and (A.6), we obtain

$$B^{(j)} B_{k,j'} \le B^{(j+j'-k)} + \sum_{n=j+1}^{k} B_0^{n-j} B^{(j'-k+n)} \quad \text{for } j \le k < j + j', \tag{A.23}$$

$$\begin{aligned}B^{(j)} B_{k,j'} &= B^{(j')} B_0^{k-j} + \sum_{n=j+1}^{k-j'} B_0^{n-j} B^{(j')} B_0^{k-n} + \sum_{n=k-j'+1}^{k} B_0^{n-j} B^{(j')} B_0^{k-n} \\ &\le (k-j'-j+1) B_0^{k-j-j'+1} \\ &\quad + \sum_{n=k-j'+1}^{k} B_0^{n-j} B^{(j'-k+n)}, \quad k \ge j + j'.\end{aligned} \tag{A.24}$$

From (A.22) we obtain $B^{(2)}_{k,j,j'} := \sum_{m=1}^{k} B_0^{m-1} B^{(j)} B_{k-m,j'} = 0$ for $k \le j$. Using the fact that the first column of $B^{(j)}$ is null for $j > 1$, (A.22) and (A.23) entail $B^{(2)}_{k,j,j'}(1,1) = 0$ for $j \le k < j + j'$. With the same argument, (A.22)–(A.24) show that for $k \ge j + j'$

$$\begin{aligned}B^{(2)}_{k,j,j'}(1,1) &= \sum_{m=1}^{k-j-j'} B_0^{m-1} B^{(j)} B_{k-m,j'}(1,1) \\ &\le \sum_{m=1}^{k-j-j'} (k-m-j'-j+1) B_0^{k-j-j'}(1,1) \\ &\le \frac{(k-j-j')(k-j-j'+1)}{2} B_0^{k-j-j'}(1,1) \le k^2 B_0^{k-j-j'}(1,1).\end{aligned}$$

Similarly we have $B^{(1)}_{k,j,j'}(1,1) \le k^2 B_0^{k-j-j'}(1,1)$. Therefore, from (A.21) we deduce

$$\frac{\partial \sigma_t^2}{\partial \beta_j \partial \beta'_j}(\theta_0) \le 2\omega_0 \sum_{k=j+j'}^{\infty} k^2 B_0^{k-j-j'}(1,1) + 2 \sum_{k=j+j'+1}^{\infty} \sum_{\ell=1}^{k-j-j'} \alpha_{0\ell} k^2 B_0^{k-\ell-j-j'}(1,1)\epsilon_{t-k}^2.$$

By the arguments used to show (A.20), we conclude that

$$E_{\theta_0} \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \beta_j \partial \beta_{j'}}(\theta_0) < \infty,$$

which shows the existence of the last expectation in Lemma 8. □

The following two lemmas show the existence of the information matrix J defined in (3), under the assumptions of Theorem 2.

Lemma 9. *Under the assumptions of Theorem 2,*

$$E_{\theta_0} \left\| \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right\| < \infty \quad \textit{and} \quad E_{\theta_0} \left\| \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right\| < \infty.$$

Proof. In view of Lemma 8, the derivatives of σ_t^2 divided by σ_t^2 possess second-order moments. For $\theta = \theta_0$, the variable $\epsilon_t^2/\sigma_t^2 = \eta_t^2$ possesses a first-order moment and is independent of the terms involving σ_t^2 and its derivatives. The results then follow from (A.1), using the Hölder inequality. □

Lemma 10. *Under the assumptions of Theorem 2,*

$$J \textit{ is non-singular} \quad \textit{and} \quad \mathrm{Var}_{\theta_0}\left\{\frac{\partial \ell_t(\theta_0)}{\partial \theta}\right\} = \left\{\kappa_\eta - 1\right\} J.$$

Proof. The proof follows from Lemma 9 and the identifiability assumptions A3, A4 (see FZ, Proof of Theorem 2.2(ii)). □

The following lemma, together with Theorem 1(i), readily shows that J can be consistently estimated by $\hat{J} := \partial^2 \mathbf{l}_n(\hat{\theta}_n)/\partial\theta\partial\theta'$.

Lemma 11. *Under the assumptions of Theorem 2, for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that, almost surely,*

$$E_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)\cap\Theta} \left\| \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} \right\| < \infty \tag{A.25}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{V}(\theta_0)\cap\Theta} \left\| \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right\| \le \varepsilon. \tag{A.26}$$

Proof. When A7 is assumed, (A.25) is a consequence of (A.1). Now assume that A8, instead of A7, holds. We will show that Lemma 8 remains true in some neighborhood of θ_0. Let $j_0 = j_0(\theta_0)$

be the integer defined in (4). Let $\mathcal{V}(\theta_0)$ be a neighborhood of θ_0 such that

$$\inf_{\theta\in\mathcal{V}(\theta_0)} \prod_{i=1}^{j_0} \alpha_i > 0 \quad \text{and} \quad \inf_{\theta\in\mathcal{V}(\theta_0)} \beta_{j_0} > 0.$$

For the sequence $(i_k) = (i_k(\theta_0))$ satisfying (A.11) and some $\underline{\alpha} > 0$ (for instance one can take $\underline{\alpha} = \inf_{\theta\in\mathcal{V}(\theta_0)} \min_{\{i:1\le i\le j_0\}} \alpha_i$), we have

$$\inf_{\theta\in\mathcal{V}(\theta_0)} \alpha_{i_k} B^{k-i_k}(1,1) \ge \underline{\alpha} \inf_{\theta\in\mathcal{V}(\theta_0)} B^{k-i_k}(1,1) > 0.$$

Like for (A.12) we then have

$$\sup_{\theta\in\mathcal{V}(\theta_0)} \frac{1}{\sigma_t^2} \frac{\partial\sigma_t^2(\theta_0)}{\partial\alpha_i} \le K \sum_{k=i}^{\infty} \sup_{\theta\in\mathcal{V}(\theta_0)} \left\{ \frac{B^{k-i}(1,1)}{B^{k-i_k}(1,1)} \right\} \rho^k \epsilon_{t-k}^{2s}, \tag{A.27}$$

using $\sup_{\theta\in\mathcal{V}(\theta_0)} \|B^k\| \le K\rho^k$, which is a consequence of $\sup_{\theta\in\Theta} \rho(B) < 1$. Note that $B_0^{k-i_k}(1,1) \neq 0$ implies $B^{k-i_k}(1,1) \neq 0$ in $\mathcal{V}(\theta_0)$, but that $B_0^{k-i}(1,1) = 0$ does not imply $B^{k-i}(1,1) = 0$ in $\mathcal{V}(\theta_0)$. However, in any case we have

$$\frac{B^{k-i}(1,1)}{B^{k-i_k}(1,1)} \le \frac{1}{\beta_{j_0}^{q_i}}.$$

Indeed the last equality is straightforward when $B^{k-i}(1,1) = 0$ and follows from (A.14) when $B^{k-i}(1,1) \neq 0$. It follows that the sup inside the sum in (A.27) is bounded. Therefore

$$\left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{1}{\sigma_t^2} \frac{\partial\sigma_t^2(\theta_0)}{\partial\alpha_i} \right\|_3 < \infty.$$

Similar existence of moments can be shown for the other derivatives involved in the second derivative of $\ell_t(\theta)$. (A.25) follows.

Now, under A7 or A8, the ergodic theorem shows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\| \frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\ell_t(\theta_0)}{\partial\theta\partial\theta'} \right\| = E_{\theta_0} \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\| \frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\ell_t(\theta_0)}{\partial\theta\partial\theta'} \right\|.$$

This expectation decreases to 0 when the neighborhood $\mathcal{V}(\theta_0)$ decreases to the singleton $\{\theta_0\}$. Thus (A.26) is also proved. □

The following lemma shows that the initial values are asymptotically negligible.

Lemma 12. *Under the assumptions of Theorem 2,*

$$\left\| n^{-1/2} \sum_{t=1}^{n} \left\{ \frac{\partial\ell_t(\theta_0)}{\partial\theta} - \frac{\partial\tilde{\ell}_t(\theta_0)}{\partial\theta} \right\} \right\| \to 0 \tag{A.28}$$

and

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\| n^{-1} \sum_{t=1}^{n} \left\{ \frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\tilde{\ell}_t(\theta)}{\partial\theta\partial\theta'} \right\} \right\| \xrightarrow{P} 0. \tag{A.29}$$

Proof. From FZ, p. 625, we have

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left|n^{-1}\sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\theta)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\tilde{\ell}_t(\theta)}{\partial\theta_i\partial\theta_j}\right\}\right|\le Kn^{-1}\sum_{t=1}^{n}\rho^t\Upsilon_t,$$

where $K>0$, $\rho\in(0,1)$, and

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

It is known that, under the strict stationarity assumption A2, ϵ_t admits a moment of order $6s$ for some $s>0$ (see Nelson [22] and Berkes et al. [5, Lemma 2.3]). Using the Hölder inequality, it follows that $E\Upsilon_t^s<\infty$. The Markov inequality and the elementary inequality $(a+b)^s\le a^s+b^s$ for all $a,b\ge 0$, $s\in(0,1)$ entail

$$\forall\varepsilon>0,\quad P\left(Kn^{-1}\sum_{t=1}^{n}\rho^t\Upsilon_t>\varepsilon\right)\le KE(\Upsilon_t^s)\varepsilon^{-s}n^{-s}\sum_{t=1}^{n}\rho^{st}\to 0,\quad \text{as } n\to\infty,$$

which shows (A.29). The convergence (A.28) is shown by similar arguments. □

The following lemma establishes the asymptotic normality of the normalized score.

Lemma 13. *Under the assumptions of Theorem 2, $J_n := \frac{\partial^2\mathbf{l}_n(\theta_0)}{\partial\theta\partial\theta'}$ is an a.s. positive definite matrix for sufficiently large n, and*

$$Z_n := -J_n^{-1}\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta}\xrightarrow{\mathcal{L}} Z,\quad \textit{with } Z\sim\mathcal{N}\{0,(\kappa_\eta-1)J^{-1}\}.$$

Proof. The central limit theorem of Billingsley [6] for square-integrable stationary martingale differences entails that

$$\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta}=n^{-1/2}\sum_{t=1}^{n}(1-\eta_t^2)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0)\xrightarrow{\mathcal{L}}\mathcal{N}\{0,(\kappa_\eta-1)J\}.$$

The ergodic theorem and Lemma 10 show that $J_n\to J$ almost surely as $n\to\infty$. The conclusion follows from the Slutsky lemma. □

A.2. Proof of Theorem 2

The notation $a_n \overset{o_P(1)}{=} b_n$ stands for sequences (a_n) and (b_n) such that a_n-b_n converges to zero in probability. When $\theta_0\in\overset{\circ}{\Theta}$, FZ [12] (proof of Theorem 2.2) showed that under A1–A5,

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\overset{o_P(1)}{=}Z_n:=-J_n^{-1}\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta}. \tag{A.30}$$

This relation cannot hold when $\theta_0\in\partial\Theta$ because then, at least one component of the left-hand side vector is a positive random variable. Instead we will establish that, for all $\theta_0\in\Theta$,

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\overset{o_P(1)}{=}\lambda_n^{\Lambda} \tag{A.31}$$

where $\lambda_n^{\Lambda}=\arg\inf_{\lambda\in\Lambda}\{\lambda-Z_n\}'J_n\{\lambda-Z_n\}$. When $\theta_0\in\overset{\circ}{\Theta}$ we have $\lambda_n^{\Lambda}=Z_n$ because $\Lambda=\mathbb{R}^{p+q+1}$, so (A.31) reduces to (A.30) in this case. In the general case, λ_n^{Λ} can be interpreted as the orthogonal projection of Z_n on Λ for the inner product $\langle x, y\rangle_{J_n} = x' J_n y$. It will be convenient to approximate this projection by that of Z_n on the space $\sqrt{n}(\Theta - \theta_0)$ which, by the assumption that Θ contains a hypercube, increases to Λ. This projection can be written as $\sqrt{n}(\theta_{J_n}(Z_n) - \theta_0)$ with

$$\theta_{J_n}(Z_n) = \arg\inf_{\theta\in\Theta} \|Z_n - \sqrt{n}(\theta - \theta_0)\|_{J_n}, \quad \text{whereas } \lambda_n^{\Lambda} = \arg\inf_{\lambda\in\Lambda} \|Z_n - \lambda\|_{J_n}.$$

The proof of Theorem 2 rests on a quadratic expansion about θ_0 of the quasi-likelihood function. Using a Taylor expansion for a function with right partial derivatives we get, for all θ and θ_0 in Θ,

$$\begin{aligned}\tilde{\mathbf{l}}_n(\theta) &= \tilde{\mathbf{l}}_n(\theta_0) + \frac{\partial\tilde{\mathbf{l}}_n(\theta_0)}{\partial\theta'}(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)'\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_0)}{\partial\theta\partial\theta'}(\theta-\theta_0) + R_n(\theta) \qquad (A.32)\\ &= \tilde{\mathbf{l}}_n(\theta_0) - \frac{1}{2n}Z_n' J_n\sqrt{n}(\theta-\theta_0) - \frac{1}{2n}\sqrt{n}(\theta-\theta_0)' J_n Z_n\\ &\quad + \frac{1}{2}(\theta-\theta_0)' J_n(\theta-\theta_0) + R_n(\theta) + R_n^*(\theta)\\ &= \tilde{\mathbf{l}}_n(\theta_0) + \frac{1}{2n}\|Z_n - \sqrt{n}(\theta-\theta_0)\|_{J_n}^2 - \frac{1}{2n}Z_n' J_n Z_n + R_n(\theta) + R_n^*(\theta), \qquad (A.33)\end{aligned}$$

where $R_n(\theta)$ and $R_n(\theta)^*$ are remainder terms which will be discussed below. We will establish the following intermediate results. For all $\theta_0 \in \Theta$,

(i) $\sqrt{n}(\theta_{J_n}(Z_n) - \theta_0) = O_P(1)$,
(ii) $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$,
(iii) for any sequence (θ_n) such that $\sqrt{n}(\theta_n - \theta_0) = O_P(1)$,

$$R_n(\theta_n) = o_P(n^{-1}), \qquad R_n^*(\theta_n) = o_P(n^{-1}),$$

(iv) $\|Z_n - \sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n}^2 \overset{o_P(1)}{=} \|Z_n - \lambda_n^{\Lambda}\|_{J_n}^2$,
(v) $\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{o_P(1)}{=} \lambda_n^{\Lambda}$,
(vi) $\lambda_n^{\Lambda} \xrightarrow{\mathcal{L}} \lambda^{\Lambda}$.

To prove (i) we first remark that, in view of Lemma 10, the claim that $\|x\|_{J_n}$ is a norm, a.s. for n large, is justified. The triangle inequality gives

$$\begin{aligned}\|\sqrt{n}(\theta_{J_n}(Z_n) - \theta_0)\|_{J_n} &\le \|Z_n - \sqrt{n}(\theta_{J_n}(Z_n) - \theta_0)\|_{J_n} + \|Z_n\|_{J_n}\\ &\le \|Z_n\|_{J_n} + \|Z_n\|_{J_n} = O_P(1),\end{aligned}$$

where the second inequality holds because $\theta_0 \in \Theta$ and $\theta_{J_n}(Z_n)$ minimizes $\|Z_n - \sqrt{n}(\theta - \theta_0)\|_{J_n}$ over Θ, and the equality follows from Lemma 13. Thus (i) is proved.

By the Taylor expansion

$$\tilde{\mathbf{l}}_n(\theta) = \tilde{\mathbf{l}}_n(\theta_0) + \frac{\partial\tilde{\mathbf{l}}_n(\theta_0)}{\partial\theta'}(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)'\left[\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_{ij}^*)}{\partial\theta\partial\theta'}\right](\theta-\theta_0),$$

where the θ_{ij}^* lie between θ and θ_0, the first remainder term in (A.32) satisfies

$$R_n(\theta) = \frac{1}{2}(\theta-\theta_0)'\left\{\left[\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_{ij}^*)}{\partial\theta\partial\theta'}\right] - \frac{\partial^2\tilde{\mathbf{l}}_n(\theta_0)}{\partial\theta\partial\theta'}\right\}(\theta-\theta_0). \qquad (A.34)$$

When $\theta = \hat{\theta}_n$, by Theorem 1(i), (A.29) and (A.26), the difference of second-order derivatives tends to zero in probability as n tends to infinity. Hence

$$R_n(\hat{\theta}_n) = o_P(\|\hat{\theta}_n - \theta_0\|^2) = o_P(\|\hat{\theta}_n - \theta_0\|^2_{J_n}).$$

The second remainder term in (A.33) is given by

$$R_n^*(\theta) = \left\{\frac{\partial \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta} - \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta}\right\}(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'\left\{\frac{\partial^2 \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta \partial \theta'} - J_n\right\}(\theta - \theta_0). \quad \text{(A.35)}$$

Therefore, in view of (A.28) and (A.29) we have

$$R_n^*(\hat{\theta}_n) = o_P(n^{-1/2}\|\hat{\theta}_n - \theta_0\|_{J_n}) + o_P(\|\hat{\theta}_n - \theta_0\|^2_{J_n}).$$

We then have

$$\begin{aligned}
\tilde{\mathbf{l}}_n(\hat{\theta}_n) - \tilde{\mathbf{l}}_n(\theta_0) &= \frac{1}{2n}\|Z_n - \sqrt{n}(\hat{\theta}_n - \theta_0)\|^2_{J_n} - \frac{1}{2n}\|Z_n\|^2_{J_n} + R_n(\hat{\theta}_n) + R_n^*(\hat{\theta}_n) \\
&= \frac{1}{2n}\{\|Z_n - \sqrt{n}(\hat{\theta}_n - \theta_0)\|^2_{J_n} - \|Z_n\|^2_{J_n} \\
&\quad + o_P(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n}) + o_P(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^2_{J_n})\} \le 0,
\end{aligned}$$

because $\hat{\theta}_n$ minimizes $\tilde{\mathbf{l}}_n(\cdot)$ over Θ. It follows that

$$\begin{aligned}
\|Z_n - \sqrt{n}(\hat{\theta}_n - \theta_0)\|^2_{J_n} &\le \|Z_n\|^2_{J_n} + o_P(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n}) + o_P(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^2_{J_n}) \\
&\le \{\|Z_n\|_{J_n} + o_P(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n})\}^2,
\end{aligned}$$

where the last inequality holds because $\|Z_n\|_{J_n} = O_P(1)$. By the triangle inequality we deduce that

$$\begin{aligned}
\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n} &\le \|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n\|_{J_n} + \|Z_n\|_{J_n} \\
&\le 2\|Z_n\|_{J_n} + o_P(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n}).
\end{aligned}$$

Thus $\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_{J_n}\{1 + o_P(1)\} \le 2\|Z_n\|_{J_n} = O_P(1)$, and (ii) readily follows.

In view of (A.34), (A.29) and (A.26), we have $R_n(\theta_n) = o_P(\|\theta_n - \theta_0\|^2) = o_P(n^{-1})$, which proves the first part of (iii). The second equality similarly follows from (A.35) and $R_n^*(\theta_n) = o_P(n^{-1/2}\|\theta_n - \theta_0\|) + o_P(\|\theta_n - \theta_0\|^2) = o_P(n^{-1})$.

By (A.33), by the fact that $\hat{\theta}_n$ minimizes $\tilde{\mathbf{l}}_n(\cdot)$ and that $\theta_{J_n}(Z_n)$ minimizes $\|Z_n - \sqrt{n}(\theta - \theta_0)\|_{J_n}$, and by (i)–(iii) we have

$$\begin{aligned}
0 &\le \|Z_n - \sqrt{n}(\hat{\theta}_n - \theta_0)\|^2_{J_n} - \|Z_n - \sqrt{n}(\theta_{J_n}(Z_n) - \theta_0)\|^2_{J_n} \\
&= 2n\{\tilde{\mathbf{l}}_n(\hat{\theta}_n) - \tilde{\mathbf{l}}_n(\theta_{J_n}(Z_n))\} - 2n\{(R_n + R_n^*)(\hat{\theta}_n) - (R_n + R_n^*)(\theta_{J_n}(Z_n))\} \\
&\le -2n\{(R_n + R_n^*)(\hat{\theta}_n) - (R_n + R_n^*)(\theta_{J_n}(Z_n))\} = o_P(1).
\end{aligned}$$

Now since $\sqrt{n}(\theta_{J_n}(Z_n) - \theta_0) = \lambda_n^{\Lambda}$ for n sufficiently large, (iv) holds.

The vector λ_n^{Λ} being the projection of Z_n on the convex set Λ for the scalar product $\langle x, y\rangle_{J_n}$, it is characterized by $\lambda_n^{\Lambda} \in \Lambda$, $\langle Z_n - \lambda_n^{\Lambda}, \lambda_n^{\Lambda} - \lambda\rangle_{J_n} \ge 0, \forall \lambda \in \Lambda$; see e.g. Zarantonello [27], Lemma 1.1, p. 239. Thus

$$\begin{aligned}\|\sqrt{n}(\hat{\theta}_n-\theta_0)-Z_n\|_{J_n}^2 &= \|\sqrt{n}(\hat{\theta}_n-\theta_0)-\lambda_n^{\Lambda}\|_{J_n}^2+\|\lambda_n^{\Lambda}-Z_n\|_{J_n}^2\\ &\quad +2\langle\sqrt{n}(\hat{\theta}_n-\theta_0)-\lambda_n^{\Lambda},\lambda_n^{\Lambda}-Z_n\rangle_{J_n}\\ &\geq \|\sqrt{n}(\hat{\theta}_n-\theta_0)-\lambda_n^{\Lambda}\|_{J_n}^2+\|\lambda_n^{\Lambda}-Z_n\|_{J_n}^2.\end{aligned}$$

Hence, by (iv)

$$\|\sqrt{n}(\hat{\theta}_n-\theta_0)-\lambda_n^{\Lambda}\|_{J_n}^2 \leq \|Z_n-\sqrt{n}(\hat{\theta}_n-\theta_0)\|_{J_n}^2-\|Z_n-\lambda_n^{\Lambda}\|_{J_n}^2=o_P(1),$$

and (v) is proved.

The continuous mapping theorem entails (vi), because $(Z_n, J_n) \xrightarrow{\mathcal{L}} (Z, J)$ by Lemma 13, $\lambda_n^{\Lambda} = f(Z_n, J_n)$ and $\lambda^{\Lambda} = f(Z, J)$ where f is a continuous function, except on the set of the points (Z_n, J_n) such that J_n is singular, which is a set of $P_{(Z,J)}$-probability zero. The proof of Theorem 2 follows from (v) and (vi).

References

[1] D.W.K. Andrews, Estimation when a parameter is on a boundary of the parameter space: Part II, Yale University, 1997 (Unpublished manuscript).

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

[3] I. Berkes, L. Horváth, The efficiency of the estimators of the parameters in GARCH processes, Annals of Statistics 32 (2004) 633–655.

[4] I. Berkes, L. Horváth, The rate of consistency of the quasi-maximum likelihood estimator, Statistics and Probability Letters 61 (2003) 133–143.

[5] I. Berkes, L. Horváth, P.S. Kokoszka, GARCH processes: Structure and estimation, Bernoulli 9 (2003) 201–227.

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

[7] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (1986) 307–327.

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

[9] F. Boussama, Ergodicité, mèlange et estimation dans les modèles GARCH, Thèse de l'Université Paris-7, 1998.

[10] H. Chernoff, On the distribution of the likelihood ratio, Annals of Mathematical Statistics 54 (1954) 573–578.

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

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

[13] L. Giraitis, R. Leipus, D. Surgailis, Recent advances in ARCH modelling, in: A. Kirman, G. Teyssiere (Eds.), Long-Memory in Economics, Springer, Berlin, 2007.

[14] P. Hall, Q. Yao, Inference in ARCH and GARCH models with heavy-tailed errors, Econometrica 71 (2003) 285–317.

[15] L. Horváth, F. Liese, L_p-estimators in ARCH-models, Journal of Statistical Planning and Inference 119 (2004) 277–309.

[16] H. Jordan, Asymptotic Properties of ARCH(p) Quasi Maximum Likelihood Estimators under Weak Conditions, Ph.D. Thesis, University of Vienna, 2003.

[17] V. Kazakevičius, R. Leipus, On stationarity in the ARCH(∞) model, Econometric Theory 18 (2002) 1–16.

[18] S. Ling, Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models, Journal of Econometrics, 2006 (in press).

[19] S. Ling, M. McAleer, A survey of recent theoretical results for time series models with GARCH errors, Journal of Economic Survey 16 (2002) 245–269.

[20] S. Ling, M. McAleer, Adaptive estimation in nonstationary ARMA models with GARCH noises, Annals of Statistics 31 (2003) 642–674.

[21] M.C. Lovell, E. Prescott, Multiple regression with inequality constraints: Pretesting bias, hypothesis testing and efficiency, Journal of the American Statistical Association 65 (1970) 913–925.

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

[23] D.B. Nelson, C.Q. Cao, Inequality constraints in the univariate GARCH model, Journal of Business and Economic Statistics 10 (1992) 229–235.

[24] M.D. Perlman, One-sided testing problems in multivariate analysis, The Annals of mathematical Statistics 40 (1969) 549–567. Corrections in The Annals of mathematical Statistics (1971) 42, 1777.

[25] D. Straumann, Estimation in Conditionally Heteroscedastic Time Series Models, in: Lecture Notes in Statistics, Springer, Berlin, Heidelberg, 2005.

[26] D. Straumann, T. Mikosch, Quasi-maximum-likelihood estimation in heteroscedastic time series: A stochastic recurrence equations approach, Annals of Statistics 34 (2006) 2449–2495.

[27] E.H. Zarantonello, Projections on convex sets in Hilbert spaces and spectral theory, in: E.H. Zarantonello (Ed.), Contributions to Nonlinear Functional Analysis, Acad. Press, New York, 1971.