(27 November 2006)

ESTIMATING AND TESTING GARCH PROCESSES WHEN THE PARAMETER IS ON A BOUNDARY

CHRISTIAN FRANCQ ,* *GREMARS Université Lille 3*

JEAN-MICHEL ZAKOIAN,** *GREMARS Université Lille 3 and CREST*

Abstract

In this paper we establish the asymptotic distribution of the quasi-maximum likelihood (QML) estimator 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 which, for important subclasses of the general GARCH, coincide with those made in the recent literature when the true parameter is in the interior of the parameter space. Furthermore, the QML estimator is shown to converge to its asymptotic distribution locally uniformly. Using these results, we consider the problem of testing that one or several GARCH coefficients are equal to zero. The null distribution and the local asymptotic powers of the Wald, score and quasi-likelihood ratio tests are derived. The one-sided nature of the problem is exploited and asymptotic optimality issues are addressed.

Keywords: Asymptotic efficiency of tests, Boundary, Chi-bar distribution, GARCH model, Quasi Maximum Likelihood Estimation, Local alternatives.

2000 Mathematics Subject Classification: Primary 62M10
Secondary 62F05; 62F12

* Postal address: GREMARS, UFR MSES, Université Lille 3, Domaine du Pont de bois, BP 149, 59653 Villeneuve d'Ascq Cedex, France
** Postal address: CREST, 3 Avenue P. Larousse, 92245 Malakoff Cedex,France

1. Introduction

Much attention has been given recently to the asymptotic properties of the quasi-maximum likelihood estimator (QMLE) in the context of GARCH(p, q) processes. Whereas ARCH (AutoRegressive Conditionally Heteroskedastic) models have been introduced by Engle in 1982, and generalized by Bollerslev in 1986, it took about twenty years to see the emergence 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 (1998), the monograph by Straumann (2005) and the papers by Berkes and Horváth (2003, 2004), Berkes, Horváth and Kokoszka (2003), Hall and Yao (2003) for GARCH models with heavy-tailed errors, and Francq and Zakoïan (2004) (hereafter FZ). It is benificial to use the QMLE in the GARCH framework because it is much less sensitive with respect to heavy tailed unconditional distributions than, for instance, the least-squares method. Other estimation procedures that are not demanding in terms of unconditional moments have recently been suggested by Liese (2004) and Ling (2005). See Giraitis, Leipus and Surgailis (2004) for a survey on GARCH modeling.

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 (1992) for weaker, but generally non explicit, conditions). A key regularity condition, imposed by the above cited papers, is that the true parameter must lie in the interior of the parameter space. This is essentially required for the asymptotic normality, not for the consistency of the QML estimator. For instance the asymptotic gaussian distribution of the QMLE does not obtain 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.

For hypothesis testing, it is crucial to be able to relax the assumption that the true parameter value is an inner point of the parameter space. One typical situation where the positivity condition is violated is of course the case of conditional homoskedasticity. The problem of conditional homoskedasticity testing is particularly important in the finance literature. The model then reduces, under the null, to an independent white noise which legitimates the use of the so-called Black-Scholes formula for option pricing. Under the alternative of conditional heteroskedasticity, option pricing or Value-at-Risk calculation demand much more sophisticated methods. More generally, testing that some coefficients are null is an important subject in the GARCH framework. The non gaussianity of the QMLE may obviously have consequences on the asymptotic distribution of the standard tests statistics. The usual asymptotic χ^2 distribution of the Wald and Likelihood ratio is no longer valid. This problem is well-known (see Weiss (1986)), and has been investigated by Demos and Sentana (1998) among others.

Our objective is to develop a complete asymptotic theory of estimation and testing in the context of GARCH processes, when the true parameter may be on the boundary of the parameter space. Fullfiling such an objective requires a series of steps:

(i) deriving the asymptotic distribution of the QML estimator under, if possible, the same mild conditions as those employed when the parameter is in the interior of the parameter space. The main difficulty is that the standard equivalence in probability between the rescaled centered estimator and an asymptotically gaussian vector (the normalized score multiplied by the inverse of the Hessian matrix), does not hold. The asymptotic distribution of the QMLE will be obtained by approximating the quasi-likelihood by a quadratic function, and will be shown to be given by the projection of a normal vector onto a convex cone.

(ii) deriving the asymptotic distributions of the commonly used tests, such as the Wald, Likelihood ratio and Rao-score tests, under the null of conditional homoskedasticity or, more generally, under the assumption that one or several GARCH coefficients are equal to zero. As is well-known, the asymptotic equivalence of the three tests does not hold when the parameter belongs to the boundary.

(iii) establishing the regularity of the QML estimator over the whole parameter space. This step consists in studying the change of the asymptotic distribution derived in step (i) under a small change (of size $n^{-1/2}$) of the true parameter value. The third lemma of Le Cam being difficult to apply in our framework, we give a direct proof.

(iv) finally, comparing the asymptotic powers of the tests statistics (in the local and Bahadur senses) and studying their optimality properties. Two important examples are considered. In the first the nullity of only one coefficient is assumed under the null. In the second, the null hypothesis of no conditional heteroskedasticity is tested. In the latter case, a test exploiting the one-sided nature of the alternative will be compared to the previous ones. We will show that this test is locally asymptotically most stringent somewhere most powerful, a concept which has been introduced by Akharif and Hallin (2003).

There are numerous antecedents in the literature to the results of the present paper. A systematic investigation of estimation with a parameter on a boundary, and a wide class of boundary hypothesis tests is in Andrews (1997, 1999, 2001). In particular, he considers the GARCH(1, q) model under assumptions we will further discuss. Klüppelberg et al. (2002), and May and Szimayer (2001) consider testing for conditional heteroskedasticity in the AR(1)-GARCH(1,1) framework. To our knowledge, asymptotic results for the general GARCH(p, q) when the parameter is on the boundary are not available in the literature. The use of a quadratic approximation to the objective function, and its optimization on a convex cone have been made by Chernoff (1954) and Andrews (2001) among many others (see the latter paper for a list of references). Testing problems in which, under the null hypothesis, the parameter is on the boundary of the maintained assumption have been considered e.g. by Chernoff (1954), Bartholomew (1959), Perlman (1969), Gouriéroux, Holly and Monfort (1982), Andrews (2001). Several papers consider one-sided alternatives. These include Wolak (1989), Rogers (1986), Silvapulle and Silvapulle (1995), King and Wu (1997); see the latter paper for further references. In particular, tests exploiting the one-sided nature of the ARCH alternative, against the null of no ARCH effect, have been proposed by Lee and King (1993), Hong (1997), Demos and Sentana (1998), Hong and Lee (2001), Andrews (2001), Dufour, Khalaf, Bernard and Genest (2004) among others.

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 moments assumptions on the observed process. Section 4 establishes, without additional assumptions, the regularity of the QLME. These results are applied in Section 5 to testing that some coefficients are equal to zero. We concentrate on the Wald, score and quasi-likelihood ratio tests. Conditions ensuring the asymptotic optimality of these tests are given. The one-sided conditional homoscedasticity test proposed by Lee and King (1993) is also investigated. This test is shown to be locally asymptotically most stringent somewhere most powerful under some regularity conditions. Section 6 concludes. 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. The notation $a \stackrel{c}{=} b$ will stand for $a = b + c$.

2. Assumptions and preliminary results

Let $(\epsilon_1, \dots, \epsilon_n)$ be a realization of length n of a nonanticipative strictly stationary solution (ϵ_t) to the GARCH(p,q) model:

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

where (η_t) is a sequence of iid random variables such that $E\eta_t^2 = 1$, $\omega_0 > 0$, $\alpha_{0i} \geq 0$ ($i = 1, \dots, q$), $\beta_{0j} \geq 0$ ($j = 1, \dots, p$). The vector of parameters is $\theta = (\theta_1, \dots, \theta_{p+q+1})' = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)'$ and it belongs to a parameter space $\Theta \subset (0, +\infty) \times [0, \infty)^{p+q}$. The true parameter value is denoted by $\theta_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})' \in \Theta$.

Bougerol and Picard (1992) showed that a unique nonanticipative strictly stationary solution (ϵ_t) to Model (2.1) 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

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

with I_k being the $k \times k$ identity matrix. The reader is referred to Bougerol and Picard (1992) for the definition and properties of the Lyapounov exponents.

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

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

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 $\overline{\omega} > \underline{\omega} > 0$. For instance one can take

$$\Theta = [\underline{\omega}, \overline{\omega}] \times [0, \overline{\alpha}_1] \times \cdots \times [0, \overline{\beta}_p] \tag{2.2}$$

where $\overline{\alpha}_1, \dots, \overline{\beta}_p > 0$.

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

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\min_{\theta\in\Theta} \tilde{\mathbf{l}}_n(\theta), \tag{2.3}$$

where

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

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 nonanticipative solution of

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

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

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

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 $\sum_{j=1}^{p} \beta_j < 1, \quad \forall \theta \in \Theta$.

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: $\kappa_\eta := E\eta_t^4 < \infty$.

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 for instance Berkes et al (2003). For detailed comments on these assumptions see FZ, in which the following result is established.

Theorem 2.1. *Let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (2.3). 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*

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

A crucial step in the proof of this theorem is to show that the L^2-norm of the vector $\sigma_t^{-2}(\theta_0)\partial\sigma_t^2(\theta_0)/\partial\theta$, and the L^1-norm of the matrix $\sigma_t^{-4}(\theta_0)\left(\partial\sigma_t^2(\theta_0)/\partial\theta\right)\left(\partial\sigma_t^2(\theta_0)/\partial\theta'\right)$ are finite. A bound for these norms was shown to be of the form Kc^{-1}, where K is a constant and $c > 0$ is the smallest component of θ_0. Obviously, the proof breaks down when one or several components of θ_0 are equal to zero. In the next section we will allow true parameter values belonging to $\partial\Theta := \{\theta_0 \in \Theta \;:\; \theta_{0i} = 0, \quad \text{for some } i > 0\}$. 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: $\omega_0 > \underline{\omega}$ and $\theta_0(\varepsilon) \in \overset{\circ}{\Theta}$ for some $\underline{\omega} > 0$ and $\varepsilon > 0$.

For instance, if Θ has the form (2.2), one can take $\omega_0 > \underline{\omega}$ 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 $\alpha_{0i} > 0$ $(i = 1, \dots, q)$, $\beta_{0j} > 0$ $(j = 1, \dots, p)$ 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 nonnegative 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 2.1, no moment assumption is required for the asymptotic distribution to hold when θ_0 is an interior point of Θ. Before deriving the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ when $\theta_0 \in \partial\Theta$, 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 $\quad \epsilon_t = \sigma_t\eta_t, \quad \sigma_t^2 = \omega_0 + \alpha_{01}\epsilon_{t-1}^2 + \alpha_{02}\epsilon_{t-2}^2$ 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

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

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

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

whence

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

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 2.1 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 A.1 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

$$\{j \mid \beta_{0,j} > 0\} \neq \varnothing \quad \text{and} \quad \prod_{i=1}^{j_0} \alpha_{0i} > 0 \quad \text{for} \quad j_0 = \min\{j \mid \beta_{0,j} > 0\}. \tag{3.1}$$

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. The proof is relegated to the end of the section.

Theorem 3.1. *Let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (2.3). Then if* **A2**–**A6** *and either* **A7** *or* **A8** *hold,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \lambda^{\Lambda} := \arg\inf_{\lambda\in\Lambda} \{\lambda - Z\}' J \{\lambda - Z\},$$

$$\text{with} \quad Z \sim \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right), \qquad \Lambda = \Lambda(\theta_0) = \Lambda_1 \times \cdots \times \Lambda_{p+q+1},$$

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 $\theta_{0i} = 0$.

Comments.

1. For $\theta_0 \in \overset{\circ}{\Theta}$, the result of this theorem reduces to that of Theorem 2.1. 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 3.1 has interest only when θ_0 belongs to $\partial\Theta$. In such a case the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is more complex than a Gaussian.

2. We insist on 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 those employed for the standard case where $\theta_0 \in \overset{\circ}{\Theta}$. For ARCH($q$) models, **A7** is required but **A4** vanishes.

3. Andrews (1997) considered the case of a GARCH(1,q) model. When $p = 1$ our model does not reduce to his specification because he allows for stochastic regressors in the conditional mean equation, whereas we assume that the GARCH process is observed. In Andrews it is assumed that (ϵ_t) is stationary and ergodic and that $Eh_t^s < \infty$ for some $s \in (0, 1)$. These properties are in fact consequences of **A2** (see FZ). Another difference is that our assumptions **A2-A4** imply the consistency of $\hat{\theta}$. In Andrews this property is proved under the existence of the second-order moment of ϵ_t. On the other hand Andrews assumes 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 by Andrews. In particular the ARCH(q) and the GARCH(1,1) models with coefficients equal to zero are not covered by his paper.

4. 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 A.2 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 section of subspaces allows to obtain this projection in a more explicit way (see e.g. Perlman (1969)). 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 $\mathcal{K} = \left\{K_1, \ldots, \; K_{2^{d_2}-1}\right\}$, 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

$$\begin{aligned} \lambda^{\Lambda} &= Z\mathbb{1}_{\Lambda}(Z) + \mathbb{1}_{\Lambda^c}(Z) \times \arg\min_{\lambda \in \mathcal{C}} \|\lambda - Z\|_J \\ &= Z\mathbb{1}_{\Lambda}(Z) + \sum_{i=1}^{2^{d_2}-1} P_i Z \mathbb{1}_{\mathcal{D}_i}(Z), \end{aligned} \tag{3.2}$$

where $\mathcal{C} = \{\lambda_{K_i} : K_i \in \mathcal{K} \text{ and } K\lambda_{K_i} \geq 0\}$, $\|\lambda_{K_i} - Z\|_J^2 = Z' M_i Z$, Λ and the $\mathcal{D}_i$ form a partition of $\mathbb{R}^d$. These formulas will be illustrated in Sections 5.1-5.2.

4. Regularity of $\hat{\theta}_n$ over the whole parameter space

We will show that the QMLE converges to its asymptotic distribution locally uniformly. More precisely, define a sequence of local parameters $\theta_n = \theta_0 + \tau/\sqrt{n}$, where $\tau = (\tau_0, \ldots, \tau_{p+q})' \in (0, +\infty)^{p+q+1}$ is such that $\theta_n \in \Theta$, at least for sufficiently large n. Write $\mathbf{A_0} = \mathbf{A}(\theta_0)$ and assume that **A2** holds. For n large enough, $\gamma\{\mathbf{A}(\theta_0 + \tau/\sqrt{n})\} < 0$ and we denote by $(\epsilon_{t,n})_{t\in\mathbb{Z}}$ the non anticipative and strictly stationary solution of

$$\begin{cases} \epsilon_{t,n} = \sqrt{h_{t,n}}\ \eta_t \\ h_{t,n} = \omega_0 + \frac{\tau_0}{\sqrt{n}} + \sum_{i=1}^q \left(\alpha_{0i} + \frac{\tau_i}{\sqrt{n}}\right)\epsilon_{t-i,n}^2 + \sum_{j=1}^p \left(\beta_{0j} + \frac{\tau_{q+j}}{\sqrt{n}}\right) h_{t-j,n}, \quad \forall t \in \mathbb{Z} \end{cases}$$

where (η_t) is iid $(0,1)$. Given the observations $\epsilon_{1,n}, \ldots, \epsilon_{n,n}$, the QMLE satisfies

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^n \tilde{\ell}_{t,n}, \quad \tilde{\ell}_{t,n} = \tilde{\ell}_{t,n}(\theta) = \tilde{\ell}_t(\theta; \epsilon_{n,n}, \ldots, \epsilon_{1,n}) = \frac{\epsilon_{t,n}^2}{\tilde{\sigma}_{t,n}^2} + \log\tilde{\sigma}_{t,n}^2, \tag{4.1}$$

where $\tilde{\sigma}_{t,n} = \tilde{\sigma}_{t,n}(\theta)$ is obtained by replacing ϵ_u by $\epsilon_{u,n}$, $1 \leq u < t$, in $\tilde{\sigma}_t$ but, for simplicity, with initial values independent of n. Similarly $\sigma_{t,n}^2(\theta)$ is defined by replacing ϵ_u by $\epsilon_{u,n}$, $u < t$, in $\sigma_t^2(\theta)$. Denote by $\mathbb{P}_{n,\tau}$ the distribution of $(\epsilon_{t,n})$.

Theorem 4.1. *Let $\theta_0 \in \Theta$ and let $\tau \in (0, +\infty)^{p+q+1}$. Let $(\hat{\theta}_n)$ be a sequence of QML estimators satisfying (4.1). Then, if* **A2-A4** *hold,* $\hat{\theta}_n \to \theta_0$, $\mathbb{P}_{n,\tau}-a.s.$ *as $n \to \infty$. Moreover, if the assumptions of Theorem 3.1 hold then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically distributed under $\mathbb{P}_{n,\tau}$ as*

$$\lambda^{\Lambda}(\tau) := \arg\inf_{\lambda\in\Lambda} \{\lambda - Z - \tau\}' J \{\lambda - Z - \tau\}, \quad \textit{with} \quad Z \sim \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right).$$

Given the limiting distribution of a statistic under $\mathbb{P}_0 = \mathbb{P}_{n,0}$, a usual method for establishing its limiting distribution under $\mathbb{P}_{n,\tau}$ is to show that $\mathbb{P}_{n,\tau}$ and $\mathbb{P}_0$ are contiguous and then to use Le Cam's third lemma (see e.g. van der Vaart p 90, 1998). Because the sequence $\{\sqrt{n}(\hat{\theta}_n - \theta_0)', \log L_n(\theta_0 + \tau/\sqrt{n}) - \log L_n(\theta_0)\}$ is not asymptotically Gaussian, Le Cam's third lemma seems difficult to apply. The same problem was encountered by Ling (2005). For this reason, the previous theorem is established directly.

5. Testing that some GARCH coefficients are equal to zero

In this section we consider the problem of testing that a subset of components of the parameter vector are equal to zero. A special attention will be given to testing conditional homoscedasticity. Given the variety of possible tests we decided to limit ourselves to the most widely used test procedures, namely those of Wald, Rao (or Lagrange Multiplier) and the quasi-likelihood ratio (QLR). For conditional homoscedasticity testing we will also compare these 3 tests with the Lee and King (1993) test, which exploits the one-sided nature of the alternatives and enjoys optimality properties.

Without loss of generality suppose we are interested in testing the nullity of the last d_2 components of θ. Partition θ as $\theta = (\theta_0^{(1)}, \theta_0^{(2)})'$ where $\theta^{(i)} \in \mathbb{R}^{d_i}$, $d_1 + d_2 = p + q + 1 = d$. The null hypothesis is

$$H_0 : \quad \theta_0^{(2)} = 0 \qquad \text{i.e. } K\theta_0 = 0_{d_2 \times 1} \text{ with } \quad K = \begin{pmatrix} 0, & I_{d_2} \end{pmatrix}.$$

The first alternatives we consider are local hypotheses of the form

$$H_n(\tau) : \quad \theta = \theta_0 + \frac{\tau}{\sqrt{n}}, \qquad \text{with} \qquad K\theta_0 = 0 \text{ and } \tau \in (0, +\infty)^{p+q+1}.$$

As a maintained assumption, we suppose that

$$\textbf{A9:} \qquad \theta_0^{(1)} > 0 \qquad \text{i.e. } \overline{K}\theta_0 > 0 \text{ with } \quad \overline{K} = \begin{pmatrix} I_{d_1}, & 0_{d_1 \times d_2} \end{pmatrix}.$$

The procedures we consider here are based on the statistics

$$\begin{aligned} \mathbf{W}_n &= \frac{n}{\hat{\kappa}_\eta - 1} \hat{\theta}_n^{(2)'} \left\{ K \hat{J}_n^{-1} K' \right\}^{-1} \hat{\theta}_n^{(2)} \quad \text{for the Wald test,} \qquad (5.1) \\ \mathbf{R}_n &= \frac{n}{\hat{\kappa}_\eta - 1} \frac{\partial \tilde{\mathbf{l}}_n \left(\hat{\theta}_{n|2} \right)}{\partial \theta'} \hat{J}_{n|2}^{-1} \frac{\partial \tilde{\mathbf{l}}_n \left(\hat{\theta}_{n|2} \right)}{\partial \theta} \quad \text{for the Rao score test,} \\ \mathbf{L}_n &= 2 \left\{ \log L_n(\hat{\theta}) - \log L_n(\hat{\theta}_{n|2}) \right\} \quad \text{for the QLR test,} \end{aligned}$$

where $\hat{\theta}_{n|2}$ denotes the estimation of θ_0 subject to the constraint $K\theta = 0$ implied by the null, $\hat{\kappa}_\eta$ denotes a consistent estimator of $\kappa_\eta = E\eta_t^4$, and $\hat{J}_n, \hat{J}_{n|2}$ denote consistent estimators of the information matrix J defined in Theorem 2.1. In general, $\hat{J}_n$ is derived using the unconstrained estimator $\hat{\theta}_n$, whereas $\hat{J}_{n|2}$ is computed using $\hat{\theta}_{n|2}$. For instance, one can take

$$\hat{J}_n = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}_n)} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}_n)}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}_n)}{\partial \theta'}, \quad \hat{J}_{n|2} = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}_{n|2})} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}_{n|2})}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}_{n|2})}{\partial \theta'}.$$

One rejects the null hypothesis for large values of these statistics. The asymptotic distributions of the 3 test statistics under both the null $H_0 = H_n(0)$ and the local alternatives $H_n(\tau)$ are given in the following theorem. Note that taking $\tau = 0$ in the definition of $\lambda^\Lambda(\tau)$ in Theorem 4.1, gives the variable λ^Λ of Theorem 3.1. Therefore we can set $\lambda^\Lambda(0) = \lambda^\Lambda$. We denote by $\chi^2_k(c)$ the noncentral chi-square distribution with noncentrality parameter c and k degrees of freedom. Let $\Omega = K'\left\{(\kappa_\eta - 1)KJ^{-1}K'\right\}^{-1}K$. Note that for any $z = (z^{(1)}, z^{(2)})' \in \mathbb{R}^d$ we have $z'\Omega z = \|z^{(2)}\|_{\{\mathrm{var}(Z^{(2)})\}^{-1}}$ where $Z = (Z^{(1)}, Z^{(2)})'$ is as in Theorem 3.1.

Theorem 5.1. *Under $H_n(\tau)$, with $\tau \geq 0$,* **A9** *and the assumptions of Theorem 3.1, we have*

$$\mathbf{W}_n \xrightarrow{\mathcal{L}} \mathbf{W}(\tau) = \lambda^\Lambda(\tau)'\Omega\lambda^\Lambda(\tau), \tag{5.2}$$

$$\mathbf{R}_n \xrightarrow{\mathcal{L}} \chi^2_{d_2}\left\{\tau'\Omega\tau\right\}, \tag{5.3}$$

$$\begin{aligned}\mathbf{L}_n \xrightarrow{\mathcal{L}} \mathbf{L}(\tau) &= -\frac{1}{2}\left\{\lambda^\Lambda_\tau - Z - \tau\right\}' J \left\{\lambda^\Lambda_\tau - Z - \tau\right\} + \frac{\kappa_\eta - 1}{2}(Z+\tau)'\Omega(Z+\tau) \\ &= -\frac{1}{2}\left\{\inf_{K\lambda \geq 0} \|Z + \tau - \lambda\|^2_J - \inf_{K\lambda = 0} \|Z + \tau - \lambda\|^2_J\right\}. \end{aligned} \tag{5.4}$$

where Z and $\lambda^\Lambda(\tau)$ are defined in Theorem 4.1 with $\Lambda = \mathbb{R}^{d_1} \times [0, \infty)^{d_2}$.

Comment. 5. Contrary to the classical situation, the asymptotic distributions of the 3 test statistics are not the same. Only the score statistic has the standard $\chi^2_{d_2}$ distribution under the null. This means that the standard Rao score test remains valid whatever the position of θ_0, in the interior or on the boundary of Θ. Valid tests based on the Wald and QLR statistics require correction of the usual critical values when θ_0 has zero components. This problem is well known in situations where the parameter is constrained both under the null and the alternatives (see Chernoff (1954) and the references in the introduction).

It is easily seen that, in Theorem 5.1, the asymptotic distribution of the Rao test is very different from that of the two other tests. The following proposition establishes that the asymptotic distributions of the latter tests are actually the same.

Proposition 5.1. *Under the assumptions of Theorem 5.1,*

$$\mathbf{W}_n \overset{o_P(1)}{=} \frac{2}{\hat{\kappa}_\eta - 1}\mathbf{L}_n.$$

Theorem 5.1 and Proposition 5.1 can be used to compare the local asymptotic behaviour of the 3 tests. The comparison of tests by means of their local asymptotic powers is generally referred to as the Pitman approach. Another popular approach is the non local approach of Bahadur in which the efficiency of a test is measured by the rate of convergence of its p-value under a fixed alternative. Let $S_{\mathbf{W}}(t) = P(\mathbf{W}(0) > t)$ and $S_{\mathbf{R}}(t) = P(\mathbf{R}(0) > t)$ be the asymptotic survival functions of the Wald and score statistics under the null hypothesis H_0.

Proposition 5.2. *Under the alternative $H_1 : \theta = \theta_1 > 0$, and under the assumptions of Theorem 3.1 with θ_0 replaced by θ_1, we have, almost surely*

$$\hat{\theta}_{n|2} \to \theta_{1|2} := \begin{pmatrix} \theta_1^{(1)} + J_{11}^{-1} J_{12} \theta_1^{(2)} \\ 0_{d_2 \times 1} \end{pmatrix},$$

the approximate Bahadur slope of the Wald test is

$$\lim_{n\to\infty} -\frac{2}{n} \log S_{\mathbf{W}}(\mathbf{W}_n) = \frac{1}{\kappa_\eta - 1} \theta_1^{(2)'} \left(K J_1^{-1} K'\right)^{-1} \theta_1^{(2)}, \tag{5.5}$$

and the approximate Bahadur slope of the score test is

$$\lim_{n\to\infty} -\frac{2}{n} \log S_{\mathbf{R}}(\mathbf{R}_n) = \frac{1}{\kappa_\eta - 1} \theta_1^{(2)'} \left(K J_1^{-1} K'\right)^{-1} K J_{1|2}^{-1} K' \left(K J_1^{-1} K'\right)^{-1} \theta_1^{(2)}, \tag{5.6}$$

where $J_1 := J(\theta_1) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}$ *and* $J_{1|2} := J(\theta_{1|2})$.

The term "approximate" Bahadur slopes serves to distinguish the limits in (5.5) and (5.6) from other quantities, called "exact" Bahadur slopes, which are defined by substituting the non-asymptotic survival functions $P(\mathbf{W}_n > t)$ and $P(\mathbf{R}_n > t)$ for $S_{\mathbf{W}}(t)$ and $S_{\mathbf{R}}(t)$ in the above definitions. We are unable to pursue the exact versions because we do not have large-deviation results for the statistics $\mathbf{W}_n$ and $\mathbf{R}_n$. For a discussion of approximate and exact slopes, see Bahadur (1967). In the Bahadur sense, a test is considered more efficient than another one when its slope is greater. This approach is sometimes criticized (see e.g. van der Vaart (1998)) and is not easy to use in our framework because the information matrices J_1 and $J_{1|2}$ are not known in closed form.

A more fruitful approach in the present context is that of Pitman which, based on Theorem 5.1, will be applied to two leading examples.

5.1. Testing that one GARCH coefficient is equal to zero

In this section we are interested in testing assumptions of the form

$$H_0 : \alpha_{0i} = 0 \quad (\text{or} \quad H_0 : \beta_{0j} = 0)$$

for some given $i \in \{1, \dots, q\}$ (or $j \in \{1, \dots, p\}$). This is for instance the case when a GARCH$(p-1, q)$ (or a GARCH$(p, q-1)$) is tested against a GARCH(p, q). The maintained assumption is that all other coefficients are positive, so that $d_2 = 1$. In view of Comment 4 we have $\Lambda = \mathbb{R}^{d_1} \times [0, \infty)$, $\quad K = (0, \dots, 0, 1)$, $\quad \mathcal{K} = \{K\}$, and $\lambda^\Lambda = Z\mathbb{1}_{Z_d \geq 0} + PZ\mathbb{1}_{Z_d < 0}$ with $Z = (Z_1, \dots, Z_d)'$, $\quad P = I_d - J^{-1}K'\left(KJ^{-1}K'\right)^{-1}K$. It follows that

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

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

In view of Proposition 5.1, it follows that

$$\mathbf{W}(0) = \frac{2}{\kappa_\eta - 1}\mathbf{L}(0) = \frac{\left\{\lambda_d^\Lambda\right\}^2}{\text{Var}Z_d} = U^2\mathbb{1}_{U\geq 0} \sim \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$$

where $U \sim \mathcal{N}(0,1)$ and δ_0 denotes the Dirac mass at 0. The distribution of $\mathbf{W}(0)$ is known as a $\overline{\chi}^2$ distribution (see Kudô, 1963). It follows that the tests defined by the critical regions $\{\mathbf{W}_n > \chi_1^2(1-2\underline{\alpha})\}$ and $\{\frac{2}{\hat{\kappa}_\eta - 1}\mathbf{L}_n > \chi_1^2(1-2\underline{\alpha})\}$ have asymptotic level $\underline{\alpha}$ (for $\underline{\alpha} \leq 1/2$). Note that the standard Wald test $\{\mathbf{W}_n > \chi_1^2(1-\underline{\alpha})\}$ has asymptotic level $\underline{\alpha}/2$. The standard QLR test $\{\mathbf{L}_n > \chi_1^2(1-\underline{\alpha})\}$ has the same asymptotic level $\underline{\alpha}/2$ when $\kappa = 3$.

Arguing as in the case $\tau = 0$, it can be shown that the last component of $\lambda^\Lambda(\tau)$ is $\lambda_d^\Lambda(\tau) = (Z_d + \tau_d)\,\mathbb{1}_{Z_d+\tau_d>0}$. We deduce that under the assumptions of Theorem 5.1

$$\mathbf{W}(\tau) = \frac{2}{\kappa_\eta - 1}\mathbf{L}(\tau) = \frac{\left\{\lambda_d^\Lambda(\tau)\right\}^2}{\text{Var}Z_d} \sim \left(U + \frac{\tau_d}{\sigma_d}\right)^2 \mathbb{1}_{\left\{U + \frac{\tau_d}{\sigma_d} > 0\right\}},$$

where $U \sim \mathcal{N}(0,1)$, and $\sigma_d^2 = \text{Var}Z_d$. Denote by $\Phi(\cdot)$ the $\mathcal{N}(0,1)$ cumulative distribution function. At $\tau^* = \tau_d/\sigma_d$, the local asymptotic power of the Wald test is

$$P\left(U + \tau^* > c_1\right) = 1 - \Phi(c_1 - \tau^*), \quad c_1 = \Phi^{-1}(1 - \underline{\alpha}). \tag{5.7}$$

The score test has the local asymptotic power

$$P\left\{(U + \tau^*)^2 > c_2^2\right) = 1 - \Phi(c_2 - \tau^*) + \Phi(-c_2 - \tau^*), \quad c_2 = \Phi^{-1}(1 - \underline{\alpha}/2). \tag{5.8}$$

We will show that the probability defined in (5.8) is less than that defined in (5.7).

Corollary 5.1. *Under the local alternatives $H_n(\tau)$, $\tau > 0$, and the assumptions of Theorem 5.1 with $d_2 = 1$, we have*

$$\lim_{n\to\infty} P\left\{\mathbf{W}_n > \chi_1^2(1 - 2\underline{\alpha})\right\} > \lim_{n\to\infty} P\left\{\mathbf{R}_n > \chi_1^2(1 - \underline{\alpha})\right\}.$$

Thus, for testing that one GARCH coefficient is equal to zero, the modified Wald test (as well as the QLR test when $\kappa_\eta = 3$) is locally asymptotically more powerful than the standard score test. This is illustrated in Figure 5.1.

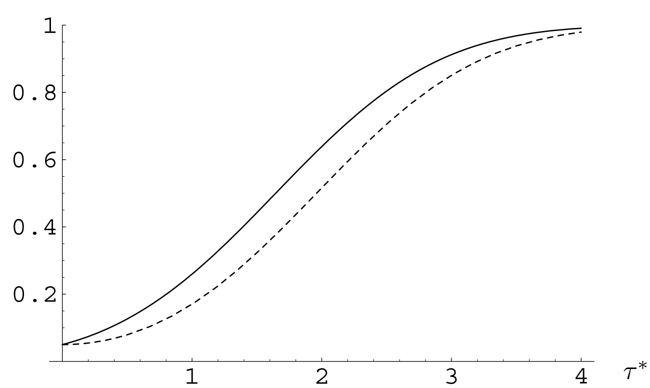

FIGURE 5.1: Local asymptotic power of the Wald test (full line) and of the score test (dashed line) for testing that one GARCH coefficient is equal to zero.

Now we will see that the modified Wald test enjoys optimality properties. Assume that η_t has a density f such that $\iota_f = \int \left\{1 + yf'(y)/f(y)\right\}^2 f(y)dy < \infty$. Note that ι_f is σ^2 times the Fisher information on the scale parameter $\sigma > 0$ in the density family

$\sigma^{-1}f(\cdot/\sigma)$. From Drost and Klaassen (1997), Drost, Klaassen and Werker (1997) and Ling and McAleer (2003) it is known that, under mild regularity conditions, GARCH processes are locally asymptotically normal (LAN) with information matrix

$$I_f = \frac{\iota_f}{4} E \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}(\theta_0) = \frac{\iota_f}{4} J.$$

In this framework the so-called local experiments $\{L_n(\theta_0 + \tau/\sqrt{n}), \tau \in \Lambda\}$ converge to the limiting gaussian experiment $\left\{\mathcal{N}\left(\tau, I_f^{-1}\right), \tau \in \Lambda\right\}$ (see van der Vaart (1998) for details about LAN properties and the notion of experiments). Testing $K\theta_0 = 0$ corresponds to testing $K\tau = 0$ in the limiting experiment. Suppose that X is $\mathcal{N}\left(\tau, I_f^{-1}\right)$ distributed. From the Neyman-Pearson lemma, the test rejecting for large values of KX is uniformly most powerful against the alternatives $K\tau > 0$. This optimal test has the power

$$\pi(\tau) = 1 - \Phi\left(c_{\underline{\alpha}} - \frac{K\tau}{\sqrt{KI_f^{-1}K'}}\right), \quad c_{\underline{\alpha}} = \Phi^{-1}(1 - \underline{\alpha}). \tag{5.9}$$

A test whose level and power jointly converge to $\underline{\alpha}$ and to the bound in (5.9), respectively, will be called asymptotically optimal.

Corollary 5.2. *Assume that η_t has a density f such that ι_f exists. For testing that one GARCH coefficient is equal to zero, the modified Wald test is asymptotically optimal if and only if*

$$f(y) = \frac{a^a}{\Gamma(a)} \exp(-ay^2)|y|^{2a-1}, \quad a > 0, \quad \Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)dt. \tag{5.10}$$

The QLR test is asymptotically optimal in the gaussian case only, and the score is never asymptotically optimal.

5.2. Testing conditional homoscedaticity

Consider the case $d_1 = 1$, with $\theta^{(1)} = \omega$, $p = 0$ and $d_2 = q$. This case corresponds to the problem of testing the null hypothesis of conditional homoscedasticity versus an ARCH(q) alternative. We therefore consider the hypothesis

$$H_0 : \alpha_{01} = \cdots = \alpha_{0q} = 0$$

in the ARCH(q) model

$$\begin{cases} \epsilon_t & = \sigma_t \eta_t, \qquad \eta_t \quad \text{iid } (0,1) \\ \sigma_t^2 & = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2, \qquad \omega > 0, \quad \alpha_{0i} \geq 0. \end{cases}$$

In his paper introducing ARCH, Engle (1982) noted that the score test is very simple to compute. Indeed a standard interpretation of the score test shows that $\mathbf{R}_n$ is equal to n times the R^2 of the regression of ϵ_t^2 on a constant and $\epsilon_{t-1}^2, \ldots, \epsilon_{t-q}^2$. Lee (1991) demonstrated that the score test is unchanged when the alternative is GARCH(p, q).

Another very simple test is the Wald test. As remarked by Demos and Sentana (1998), at the point $\theta_0 = (\omega_0, 0, \ldots, 0)$, the information matrix $J = J(\theta_0)$ takes a simple form and we have

$$(\kappa_\eta - 1)J^{-1} = \begin{pmatrix} (\kappa_\eta + q - 1)\omega^2 & -\omega & \cdots & -\omega \\ -\omega & & & \\ \vdots & & I_q & \\ -\omega & & & \end{pmatrix}. \tag{5.11}$$

Because $(\kappa_\eta - 1)KJ^{-1}K' = I_q$, a simple version of the Wald statistic is $\mathbf{W}_n = n\sum_{i=1}^q \hat{\alpha}_i^2$. Note that this is not the version of the Wald statistic defined by (5.1) because we are not using here the estimator $\hat{J}_n$ based on the unconstrained estimator $\hat{\theta}_n$. The asymptotic distribution of the Wald statistic, under the null and local alternatives, is not affected by the choice of the consistent estimator of J. It is easy to see that $\lambda^\Lambda = \left(Z_1 + \omega \sum_{i=2}^d Z_i^-, Z_2^+, \cdots, Z_d^+\right)'$. The asymptotic distribution of $n\sum_{i=2}^d \hat{\alpha}_i^2$ is therefore that of $\sum_{i=2}^d \left(Z_i^+\right)^2$, where the Z_i are iid $\mathcal{N}(0,1)$. Thus, when an ARCH(q) is fitted to a centered iid sequence, the Wald statistic satisfies

$$\mathbf{W}_n = n\sum_{i=1}^q \hat{\alpha}_i^2 \xrightarrow{\mathcal{L}} \mathbf{W}(0) = \frac{1}{2^q}\delta_0 + \sum_{i=1}^q C_q^i \frac{1}{2^q}\chi_i^2. \tag{5.12}$$

Demos and Sentana (1998) obtained the same result by means of heuristic arguments and results established by Wolak (1989) in the iid case. They wrote on page 107 that their "analysis is based on the presumption that standard results one inequality testing can be extended" to the GARCH case. Our results allow to validate this presumption. Indeed almost all the assumptions of Theorem 5.1 are trivially satisfied when $d_2 = q$, $p = 0$ and H_0 holds true. More precisely, the convergence (5.12) holds under H_0, and the assumptions **A3** and $E\eta_t^6 < \infty$.

Lee and King (1993) proposed a test which exploits the one-sided nature of the ARCH alternative. Their test rejects conditional homoscedasticity for large values of

$$\mathbf{LK}_n = -\frac{\sqrt{n}1_q'\partial\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right)/\partial\theta^{(2)}}{\hat{\sigma}_{LK}} = \frac{n^{-1/2}\sum_{t=1}^{n}(\frac{\epsilon_t^2}{\hat{\sigma}_\epsilon^2}-1)\hat{\sigma}_\epsilon^{-2}\sum_{i=1}^{q}\epsilon_{t-i}^2}{\hat{\sigma}_{LK}},$$

where $\hat{\sigma}_\epsilon^2$ denotes the empirical variance of the observations, $\hat{\sigma}_{LK}^2$ is an estimator of the variance of the numerator and $1_q = (1,\dots,1)' \in \mathbb{R}^q$. The statistic $\mathbf{LK}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ under the null. In view of (A.55), (A.57), (A.58), (A.59) and (5.11) one can take

$$\begin{aligned}\hat{\sigma}_{LK}^2 &= (\hat{\kappa}_\eta - 1)1_q'\left\{K\hat{J}_{n|2}K' - (K\hat{J}_{n|2}\overline{K}')(\overline{K}\hat{J}_{n|2}\overline{K}')^{-1}(\overline{K}\hat{J}_{n|2}K')\right\}1_q \\ &= (\hat{\kappa}_\eta - 1)1_q'\left\{K\hat{J}_{n|2}^{-1}K'\right\}^{-1}1_q = q(\hat{\kappa}_\eta - 1)^2,\end{aligned}$$

with $K = (0_{q\times 1}, I_q)$ and $\overline{K} = (1, 0_{1\times q})$ (this form is not exactly the expression given in Lee and King (1993, 1994), but it is asymptotically equivalent). The **LK**-test enjoys some optimality properties. As already seen, under mild regularity conditions, in the limiting experiment our testing problem corresponds to testing $K\tau = 0$ with one observation $X = (X_1,\dots,X_{q+1})' \sim \mathcal{N}(\tau, I_f^{-1})$. Let $\overset{\bullet}{\tau}$ be a point of Λ whose last q components are equal to some $c > 0$, and let $\overset{\circ}{\tau} = \overset{\bullet}{\tau} - I_f^{-1}K'(KI_f^{-1}K')^{-1}K\overset{\bullet}{\tau}$, so that $K\overset{\circ}{\tau} = 0$. By the Neyman-Pearson lemma, the most powerful test for testing $\tau = \overset{\circ}{\tau}$ against $\tau = \overset{\bullet}{\tau}$ rejects for large values of

$$(X - \overset{\bullet}{\tau})'I_f(X - \overset{\bullet}{\tau}) - (X - \overset{\circ}{\tau})'I_f(X - \overset{\circ}{\tau}) = 2\overset{\bullet}{\tau}{}'K'(KI_f^{-1}K')^{-1}KI_f^{-1}I_fX + constant.$$

Using

$$KI_f^{-1}K' = 4\iota_f^{-1}(\kappa_\eta - 1)^{-1}I_q, \tag{5.13}$$

it is easy to see that this test rejects for large values of $\sum_{i=2}^{q+1}X_i$. This test is therefore uniformly most powerful to test $\tau_1 = \cdots = \tau_q = 0$ versus $\tau_1 = \cdots = \tau_q > 0$. Similarly it can be shown that the tests which are somewhere most powerful (SMP) in $\Lambda \setminus (0,\infty)\times\{0\}^q$ reject for large values of $\mathbf{d}'X$ with $\mathbf{d} \in [0,\infty)^{q+1}$ and $K\mathbf{d} \neq 0$. Such a test is uniformly most powerful for testing $\tau_1 = \cdots = \tau_q = 0$ versus $\tau = c\mathbf{d}$, $c > 0$. Of course, an optimal test in the "direction" $\mathbf{d}$ may have a very low power in other directions. The test rejecting for large values of $\sum_{i=2}^{q+1}X_i$ is however most stringent somewhere most powerful (MSSMP) (the reader is referred to Shi (1987), Shi and Kudô (1987)

The authors greatly thank Professor Shi for sending us these two papers and for his answers to our questions

and the references therein for the concept of MSSMP and SMP test). In view of (5.13), this MSSMP test has the power

$$\pi(\tau) = 1 - \Phi\left(c_{\underline{\alpha}} - \frac{\sum_{i=1}^{q}\tau_i}{\sqrt{4q\iota_f^{-1}(\kappa_\eta - 1)^{-1}}}\right), \quad c_{\underline{\alpha}} = \Phi^{-1}(1-\underline{\alpha}). \tag{5.14}$$

The following corollary gives the local asymptotic powers of the conditional homoscedasticity tests considered in this section, and shows that the Lee-King test is locally asymptotically MSSMP (Lee and King (1993) exhibit another optimality property for their test). The concept of locally asymptotically MSSMP test has been proposed by Akharif and Hallin (2003) in order to cope with one-sidedness in hypothesis testing.

Corollary 5.3. *Under the local alternatives $H_n(\tau)$, $\tau > 0$, and the assumptions of Theorem 5.1 with $p = 0$, $d_1 = 1$ and $d_2 = q$, the local asymptotic power of the modified Wald, score and Lee-King tests are given by*

$$\begin{aligned}
\lim_{n\to\infty} P\left\{\mathbf{W}_n > c_{\underline{\alpha}}^{\mathbf{W}}\right\} &= P\left\{\sum_{i=1}^{q}(U_i+\tau_i)^2\mathbb{1}_{\{U_i+\tau_i>0\}} > c_{\underline{\alpha}}^{\mathbf{W}}\right\} \\
\lim_{n\to\infty} P\left\{\mathbf{R}_n > c_{\underline{\alpha}}^{\mathbf{R}}\right\} &= P\left\{\chi_q^2\left(\sum_{i=1}^{q}\tau_i^2\right) > c_{\underline{\alpha}}^{\mathbf{R}}\right\} \\
\lim_{n\to\infty} P\left\{\mathbf{LK}_n > c_{\underline{\alpha}}\right\} &= 1 - \Phi\left(c_{\underline{\alpha}} - \frac{\sum_{i=1}^{q}\tau_i}{\sqrt{q}}\right),
\end{aligned} \tag{5.15}$$

where $U = (U_1, \dots, U_q)' \sim \mathcal{N}(0, I_q)$, the critical value $c_{\underline{\alpha}}^{\mathbf{W}}$ is the $(1-\underline{\alpha})$-quantile of the $\mathbf{W}$ distribution defined by (5.12) and $c_{\underline{\alpha}}^{\mathbf{R}}$ is the $(1-\underline{\alpha})$-quantile of the χ_q^2 distribution.

Under the assumptions of Corollary 5.2, the Lee-King test is asymptotically most stringent somewhere most powerful (in the sense that the right-hand side of (5.15) is equal to the upper bound $\pi(\tau)$ defined by (5.14)) if and only if the density f of η_t belongs to the class defined by (5.10).

It is well known that there exists no satisfactory notion of optimality for testing hypothesis on multidimensional parameters. The Lee-King test is asymptotically optimal in the direction $\alpha_1 = \cdots = \alpha_q$, but there is no objective reason to favour this direction. As shown in Figure 5.2, the local asymptotic power of Lee-King test may be lower than that of the Wald test, and even lower than that of the two-sided score test.

To assess the validity of these asymptotic developments in finite samples we simulated ARCH(q) models with $\eta_t \sim \mathcal{N}(0, 1)$. Table 1 displays the relative frequency of

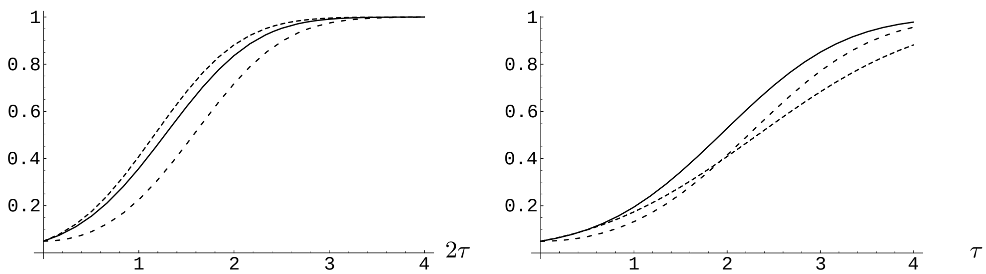

FIGURE 5.2: Local asymptotic power of the Wald (full line), score (dashed line) and Lee-King (dotted line) for testing conditional homoscedasticity with an ARCH(2) model where $\alpha_1 = \alpha_2 = \tau/\sqrt{n}$ (left figure) and $\alpha_1 = \tau/\sqrt{n},\ \alpha_2 = 0$ or $\alpha_1 = 0,\ \alpha_2 = \tau/\sqrt{n}$ (right figure).

rejection of the conditional homoscedasticity hypothesis for the Wald and score tests and for the QLR test without correction by the factor $2/(\hat{\kappa}_\eta - 1)$. In accordance with the asymptotic study, the one-sided Wald and QLR tests are more powerful that the standard two-sided score test. Similar experiments reveal that, as expected, the QLR test might not well control the error of first kind when η_t is not gaussian. Thus, in terms of power and robustness, the Wald test seems superior to the two other tests.

TABLE 1: Empirical power (in %) of the Wald, score and QLR tests for conditional homoscedasticity. The number of replications is $N = 5000$, the critical values are adjusted to obtain 5% relative rejection frequency when the observations are iid gaussian, the DGP is an ARCH(q) with ARCH coefficients $\alpha_i = n^{-1/2}$.

q	$n = 500$			$n = 1000$			$n = 10000$			$n = \infty$		
	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$
1	24.8	17.9	25.5	24.8	18.9	25.6	25.8	17.0	25.7	25.9	17.0	25.9
2	36.8	25.6	36.8	38.1	26.8	39.3	39.5	22.9	39.2	35.8	22.6	35.8
3	44.3	32.6	46.6	47.5	31.7	49.8	50.1	27.9	52.3	44.6	27.5	44.6

Table 2 compares the finite sample powers of the Wald and Lee-King one-sided tests. The conclusion drawn from the comparison of the local asymptotic powers (see Figure 5.2) remains valid for these simulation experiments. The Wald and Lee-King tests have similar powers in the case $q = 1$ (from Corollaries 5.2 and 5.3 these two tests are both locally asymptotically uniformly most powerful), the Lee-King test is slightly more powerful than the Wald test when the DGP is an ARCH(q) with $q > 1$ and $\alpha_1 = \cdots = \alpha_q > 0$ (see the fist part of Table 2), but the Lee-King test is much less powerful than the Wald for some other alternatives (as in the last part of the table).

TABLE 2: Empirical power (in %) of the Wald and Lee-King tests for conditional homoscedasticity. The number of replications is $N = 5000$, the critical values are adjusted to obtain 5% relative rejection frequency when the observations are iid gaussian, the DGP is an ARCH(q).

$\alpha_1 = \cdots = \alpha_q = 1.5n^{-1/2}$

q	$n = 500$		$n = 1000$		$n = 10000$		$n = \infty$	
	$\mathbf{W}_n$	$\mathbf{L}K_n$	$\mathbf{W}_n$	$\mathbf{L}K_n$	$\mathbf{W}_n$	$\mathbf{L}K_n$	$\mathbf{W}_n$	$\mathbf{L}K_n$
1	39.82	39.06	40.80	41.84	43.00	43.12	44.24	44.24
2	54.88	59.40	62.84	63.86	66.16	68.32	61.89	68.31
3	73.52	74.18	78.02	78.44	79.84	83.54	74.70	82.98

$\alpha_1 = \cdots = \alpha_{q-1} = 0,\ \alpha_q = q1.5n^{-1/2}$

q	$n = 500$		$n = 1000$		$n = 10000$		$n = \infty$	
	$\mathbf{W}_n$	$\mathbf{L}K_n$	$\mathbf{W}_n$	$\mathbf{L}K_n$	$\mathbf{W}_n$	$\mathbf{L}K_n$	$\mathbf{W}_n$	$\mathbf{L}K_n$
2	76.38	55.12	82.32	61.60	87.92	67.34	85.11	68.31
3	93.66	65.94	95.30	73.70	98.12	82.32	98.95	82.98

6. Concluding remarks

In this paper we derived the asymptotic distribution of the QMLE for general GARCH(p, q) models when components of the true parameter value are allowed to be zero. The asymptotic distribution is non standard, but is easily computable as the projection of a mutivariate gaussian distribution on a convex cone. The results were established under mild conditions. For important subclasses of the general GARCH(p, q), the conditions are not stronger than those recently obtained in the literature for parameters in the interior of the parameter space. For ARCH(q) models, the results require additional moment assumptions. A direct consequence of the non standard asymptotic distribution of the QMLE when the parameter is on the boundary concerns confidence intervals. Caution is needed when one coefficient is suspected to be zero. Investigations not reported here show that asymptotic conventional confidence intervals tend to be too large. The asymptotic behaviour of the QMLE was established for parameters on the boundary, and also for parameters approaching the boundary at the rate $\sqrt{n}$. One major application of these results concerns testing problems. We purposely limited ourselves to the most widely used tests (the Wald, the score and Quasi-Likelihood-ratio tests) and, for the sake of comparison, to the one-sided test of Lee and King (1993). From the derivation of the local asymptotic powers, several conclusions can be drawn: i) the Rao test remains valid for testing a value on the boundary, but looses its optimal properties; ii) the Wald and QLR tests need to be modified but remain equivalent under the null and local alternatives; iii) for testing the nullity of one coefficient the QML-based Wald test is optimal for a class of densities not restricted to the standard Gaussian; iv) for testing conditional homoscedasticity in the ARCH framework, the Wald test statistic reduces to the sum of the squared coefficients and is therefore very convenient; the one-sided Lee-King test is locally asymptotically most stringent somewhere most powerful, but may be less powerful than the 3 other tests for detecting alternatives that are not symmetric in the ARCH parameters.

Appendix A. Proofs and technical results

A.1. Asymptotic normality of the normalized score

Before proving Theorem 3.1 we will establish two lemmas. A first difficulty is 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. Therefore we cannot use a standard Taylor expansion of $\mathbf{l}_n(\theta) = n^{-1}\sum_{t=1}^n \ell_t(\theta)$ about θ_0. Instead, following Andrews (1999), we will use a Taylor expansion based on right derivatives. Indeed, it is clearly possible to define the right derivative of $\ell_t(\theta)$ at θ_0. For ease of notation, denote by $\partial\sigma_t^2(\theta_0)/\partial\theta := \left(\partial\sigma_t^2(\theta_0)/\partial\theta_i\right)_{i=1,\dots,p+q+1}$ and $\partial\ell_t(\theta_0)/\partial\theta := \left(\partial\ell_t(\theta_0)/\partial\theta_i\right)_{i=1,\dots,p+q+1}$ the vectors of partial derivatives of σ_t and ℓ_t at θ_0 with 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

$$\begin{aligned}
\frac{\partial\ell_t(\theta)}{\partial\theta} &= \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right\}, \\
\frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} &= \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right\} + \left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta'}\right\}. \qquad \text{(A.1)}
\end{aligned}$$

The first lemma allows to consider the L^1 norms of these derivatives at θ_0.

Lemma A.1. *Under the assumptions of Theorem 3.1,*

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

Proof. In this proof and the subsequent ones, K and ρ denote generic constants, whose values will be modified, such that $K>0$ and $0<\rho<1$. FZ introduce the following notations 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

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

and $B^{(j)}$ is a $p\times p$ matrix with $(1,j)$th element 1, and all other elements equal to zero. Elementary properties of these matrices are established in Lemma A.3 below. Similar formulas, given below, hold for the second derivatives. Since σ_t^{-2} is bounded by $1/\underline{\omega}$, the proof of Lemma A.1 is straightforward under Assumption **A7**.

Now assume that, **A8**, instead of **A7**, holds. From (A.3), $\partial\sigma_t^2(\theta_0)/\partial\omega$ is bounded since $\sum_{k=0}^{\infty}B^k$ is finite under **A2**. Since $\sigma_t^2(\theta_0)\geq\omega_0>0$, $\{\partial\sigma_t^2(\theta_0)/\partial\omega\}/\sigma_t^2(\theta_0)$ 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 $\ell\notin\{1,\dots,q\}$. By assumption **A8**, for all $k>0$ there exist an integer $i_k\in\{1,\dots,\min(q,k)\}$ such that

$$\sum_{\ell=1}^{k}\alpha_{0\ell}B_0^{k-\ell}(1,1)\geq\alpha_{0i_k}B_0^{k-i_k}(1,1)\geq\underline{\alpha}B_0^{k-i_k}(1,1)>0, \tag{A.6}$$

for some positive constant $\underline{\alpha}$ (one can take $\underline{\alpha}=\min\{\alpha_{0i}:\alpha_{0i}\neq0\}$). It follows that for any $s\in(0,1)$, in view of (A.2)-(A.3),

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

$$\leq\sum_{k=i}^{\infty}\frac{B_0^{k-i}(1,1)\epsilon_{t-k}^2}{\omega_0+\underline{\alpha}B_0^{k-i_k}(1,1)\epsilon_{t-k}^2}\leq\sum_{k=i}^{\infty}\frac{B_0^{k-i}(1,1)\epsilon_{t-k}^{2s}}{\omega_0^s\underline{\alpha}^{1-s}\{B_0^{k-i_k}(1,1)\}^{1-s}}, \tag{A.7}$$

where the last inequality follows from $ax/(b+cx) \leq ax^s/(b^s c^{1-s})$ for all $a, b, c, x \geq 0$. 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

$$B_0^{k-i}(1,1)/\{B_0^{k-i_k}(1,1)\}^{1-s} \leq K\rho^k \quad \text{for all } k. \tag{A.8}$$

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 $\{1,\dots,q\}$, 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 defined by (3.1) and let $r_i \in \{1,\dots,j_0\}$ 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.28), 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 (3.1). Thus one can take $i_k = r_i$ in (A.6), so that we have

$$B_0^{k-i}(1,1)/B_0^{k-i_k}(1,1) \leq 1/\beta_{0j_0}^{q_i}, \tag{A.9}$$

and thus (A.8) holds. Then (A.7) gives

$$E_{\theta_0}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\alpha_i}(\theta_0) \leq K\left\{\sum_{k=1}^{\infty}\rho^k\right\}E_{\theta_0}\epsilon_t^{2s}. \tag{A.10}$$

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

$$\frac{\partial\sigma_t^2(\theta_0)}{\partial\beta_j} = \omega_0\sum_{k=0}^{\infty}B_{k,j}(1,1) + \sum_{k=2}^{\infty}\sum_{\ell=1}^{k}\alpha_{0\ell}B_{k-\ell,j}(1,1)\epsilon_{t-k}^2 \tag{A.11}$$

where $B_{0,j} = 0$ and, for $k > 0$, the matrices $B_{k,j}$ defined in (A.5) are taken at θ_0. Using elementary properties of the matrix B (see Lemma A.3 below) we obtain, for any $0 \leq \ell \leq k-j$, by (A.24) and (A.25),

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

which, together with (A.26), entails that

$$\begin{aligned} B_{k-\ell,j}(1,1) &\leq (k-\ell-j)B_0^{k-\ell-j}(1,1) + B_0^{k-\ell-j}(1,1) \\ &\leq kB_0^{k-\ell-j}(1,1). \end{aligned} \tag{A.12}$$

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

$$B_{k-\ell,j} \leq \sum_{m=1}^{k-\ell} B_0^{m-1}B^{(j-k+\ell+m)}, \quad \text{and thus} \quad B_{k-\ell,j}(1,1) = 0. \tag{A.13}$$

Therefore, from (A.11) we deduce

$$\frac{\partial \sigma_t^2(\theta_0)}{\partial \beta_j} \quad \leq \quad \omega_0 \sum_{k=j}^{\infty} k B_0^{k-j}(1,1) + \sum_{k=j+1}^{\infty} \sum_{\ell=1}^{k-j} \alpha_{0\ell} k B_0^{k-\ell-j}(1,1)\epsilon_{t-k}^2.$$

Hence, proceeding as in (A.7) we get

$$\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta_j}(\theta_0) \leq K + \sum_{k=j+1}^{\infty} \sum_{\ell=1}^{k-j} \alpha_{0\ell} k \frac{B_0^{k-\ell-j}(1,1)\epsilon_{t-k}^{2s}}{\omega_0^s \underline{\alpha}^{1-s}\{B_0^{k-i_k}(1,1)\}^{1-s}}, \tag{A.14}$$

and thus

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

by arguments already used for (A.10). This allows to conclude that the first expectation in Lemma A.1 exists. Applying the Hölder inequality in (A.7) and (A.14) 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 A.1 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 $\partial^2 \sigma_t^2/\partial\omega\partial\beta_j \leq \sum_{k=j}^{\infty} k B_0^{k-j}(1,1) < \infty$, by (A.12) and (A.13) 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.11) with respect to $\beta_{j'}$ gives

$$\frac{\partial^2 \sigma_t^2}{\partial \beta_j \partial \beta_{j'}} \quad = \quad \omega_0 \sum_{k=0}^{\infty} B_{k,j,j'}(1,1) + \sum_{k=2}^{\infty} \sum_{\ell=1}^{k} \alpha_{0\ell} B_{k-\ell,j,j'}(1,1)\epsilon_{t-k}^2 \tag{A.16}$$

where

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

We first give a bound for the terms of the form $B^{(j)} B_{k,j'}$ involved in $B_{k,j,j'}^{(2)}$. First note that when $k \leq p$, 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

$$B^{(j)} B_{k,j'} \quad = \quad 0 \quad \text{for } k < j, \tag{A.17}$$

Using successively (A.25), (A.24) and (A.27), we obtain, for $j, j' = 1, \ldots, p$ and $k > 0$,

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

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.25) and (A.24), we obtain

$$B^{(j)} B_{k,j'} \leq B^{(j+j'-k)} + \sum_{n=j+1}^{k} B_0^{n-j} B^{(j'-k+n)} \quad \text{for } j \leq k < j + j', \tag{A.18}$$

$$\begin{aligned} B^{(j)} B_{k,j'} &= B^{(j')} B_0^{k-j} + \sum_{n=j+1}^{k-j'} B_0^{n-j} B^{(j')} B_0^{k-n} + \sum_{n=k-j'+1}^{k} B_0^{n-j} B^{(j')} B_0^{k-n} \\ &\leq (k - j' - j + 1) B_0^{k-j-j'+1} + \sum_{n=k-j'+1}^{k} B_0^{n-j} B^{(j'-k+n)}, \; k \geq j + j'. \end{aligned} \tag{A.19}$$

From (A.17) we obtain $B^{(2)}_{k,j,j'} := \sum_{m=1}^{k} B_0^{m-1} B^{(j)} B_{k-m,j'} = 0$ for $k \leq j$. Using the fact that the first column of $B^{(j)}$ is null for $j > 1$, (A.17) and (A.18) entail $B^{(2)}_{k,j,j'}(1,1) = 0$ for $j \leq k < j + j'$. With the same argument, (A.17), (A.18) and (A.19) show that for $k \geq j + j'$

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

Similarly we have $B^{(1)}_{k,j,j'}(1,1) \leq k^2 B_0^{k-j-j'}(1,1)$. Therefore, from (A.16) we deduce

$$\begin{aligned} \frac{\partial \sigma_t^2}{\partial \beta_j \partial \beta_j'}(\theta_0) \;\leq\; & 2\omega_0 \sum_{k=j+j'}^{\infty} k^2 B_0^{k-j-j'}(1,1) \\ & + 2 \sum_{k=j+j'+1}^{\infty} \sum_{\ell=1}^{k-j-j'} \alpha_{0\ell} k^2 B_0^{k-\ell-j-j'}(1,1) \epsilon_{t-k}^2. \end{aligned}$$

By the arguments used to show (A.15), 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 A.1. This completes the proof.

□

The following lemma establishes the asymptotic normality of the normalized score multiplied by the inverse of the Hessian matrix.

Lemma A.2. *Under the assumptions of Theorem 3.1,* $J_n := \frac{\partial^2 \mathbf{l}_n(\theta_0)}{\partial\theta\partial\theta'}$ *is an a.s. positive definite matrix for sufficiently large* n*, and*

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

Proof. In FZ, the proof of the asymptotic normality of Z_n relies on a set of six intermediate results, established under **A1-A5**. We start by examining the validity of these results when **A1** is replaced by **A6-A7**.

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

In view of Lemma A.1, the derivatives of σ_t^2 divided by σ_t^2 possess second-order moments. For $\theta = \theta_0$, the variable $\epsilon_t^2/\sigma_t^2 = \eta_t^2$ is independent of the terms involving σ_t^2 and its derivatives. The inequalities in *i)* follow straightforwardly, using the Hölder inequality.

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

The proof is the same as in the case where **A1** holds.

$$iii)\ \text{there exists a neighborhood } \mathcal{V}(\theta_0) \text{ of } \theta_0 \text{ such that,}$$
$$E_{\theta_0}\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left|\frac{\partial^3 \ell_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| < \infty \quad \forall i,j,k \in \{1,\dots,p+q+1\}.$$

The validity of this result is questionable without **A1**, because the third derivative of $\ell_t(\theta)$ involves terms such as

$$\left\{2 - 6\frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}\right\},$$

which might be non integrable. Fortunately, this result is not essential to our conclusion. Instead we will prove that for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that, almost surely

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

First assume that **A7** holds. The first result is a consequence of (A.1) and the moment assumption **A7**. Now assume that **A8**, instead of **A7**, holds. We will show that Lemma A.1 remains true in some neighborhood of θ_0. Let $j_0 = j_0(\theta_0)$ be the integer defined in (3.1). 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.6) 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

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

Similarly to (A.7) we then have

$$\sup_{\theta\in\mathcal{V}(\theta_0)} \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2(\theta_0)}{\partial\alpha_i} \quad\leq\quad K\sum_{k=i}^{\infty} \sup_{\theta\in\mathcal{V}(\theta_0)} \left\{\frac{B^{k-i}(1,1)}{B^{k-i_k}(1,1)}\right\}\rho^k\epsilon_{t-k}^{2s}, \tag{A.20}$$

using $\sup_{\theta\in\mathcal{V}(\theta_0)}\|B^k\| \leq K\rho^k$, 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

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

Indeed the last equality is straightforward when $B^{k-i}(1,1) = 0$ and follows from (A.9) when $B^{k-i}(1,1) \neq 0$. It follows that the sup inside the sum in (A.20) 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 $\ell_t(\theta)$. The first inequality in *iii)'* 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 the second result in *iii)'* is also proved.

$$iv)\quad \left\|n^{-1/2}\sum_{t=1}^{n}\left\{\frac{\partial\ell_t(\theta_0)}{\partial\theta} - \frac{\partial\tilde{\ell}_t(\theta_0)}{\partial\theta}\right\}\right\| \to 0 \quad\text{and} \tag{A.21}$$

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left\|n^{-1}\sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\tilde{\ell}_t(\theta)}{\partial\theta\partial\theta'}\right\}\right\| \overset{P}{\to} 0. \tag{A.22}$$

From FZ we have

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

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 (1990) and Berkes et al. (2003, Lemma 2.3)). Using the Hölder inequality, it follows that $E\Upsilon_t^s < \infty$. The Markov inequality and the elementary inequality $(a+b)^s \leq a^s + b^s$ for all $a,\, b \geq 0,\ s \in (0,1)$ entail

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

which shows the second convergence in *iv)*. The first convergence is shown by similar arguments. Note that *iv)* is shown without using the moment assumption **A7**.

$$v)\ n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ell_t(\theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (\kappa_\eta - 1)J\right) \quad \text{and} \quad Z_n \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right).$$

This result is a direct consequence of the central limit theorem for square-integrable stationary martingale differences, as was shown in the case where $\theta_0 \in \overset{\circ}{\Theta}$.

$$vi)\ n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta_{ij}^*) \to J(i,j)\ a.s., \text{ for all } \theta_{ij}^* \text{ between } \hat{\theta}_n \text{ and } \theta_0. \quad (A.23)$$

This result comes from *iii)'* and the strong consistency of $\hat{\theta}_n$.

□

In the following, we give elementary properties of the matrix B defined in (A.5).

Lemma A.3. *For* $j = 1, \ldots, p$

$$B^{(j)}B^k \leq B^{k-j+1}, \quad \text{for all } k \geq j-1, \quad (A.24)$$

$$B^{(j)}B^k = B^{(j-k)}, \quad \text{for all } 0 \leq k < j, \quad (A.25)$$

$$AB^{(1)} \leq A, \quad \text{and for } j \neq \ell_2,\ \{AB^{(j)}\}(\ell_1, \ell_2) = 0\ , \quad \forall A \geq 0, \quad (A.26)$$

$$B^{(1)}B^{(j)} = B^{(j)}, \quad \text{and} \quad B^{(i)}B^{(j)} = 0, \quad \text{for } i > 1, \quad (A.27)$$

$$B^k(1,1) \geq \beta_j B^{k-j}(1,1), \quad \text{for all } k \geq j. \quad (A.28)$$

Proof. First note that, when $j \leq p$, 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 zeroes. Thus $B^{(j)}B^j \leq B$. Multiplying the two sides of the previous inequality by B^{k-j}, whose elements are nonnegative, yields (A.24). For $k < j \leq p$, the j-th row of B^k is null, except one "1" in the j-th position, which shows (A.25). The j-th column of $AB^{(j)}$ is the first column of A, and the other elements of $AB^{(j)}$ are zeroes. Thus (A.26) and (A.27) are obvious. Inequality (A.28) comes from $B^k(1,1) = \sum_{j=1}^{p}\beta_j B^{k-1}(j,1) \geq \beta_j B^{k-1}(j,1) = \beta_j B^{k-j}(1,1)$.

□

A.2. Proof of Theorem 3.1

When $\theta_0 \in \overset{\circ}{\Theta}$, FZ showed that under **A1-A5**,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{o_P(1)}{=} Z_n := -J_n^{-1}\sqrt{n}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta}. \quad (A.29)$$

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 $\theta_0 \in \Theta$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{o_P(1)}{=} \lambda_n^\Lambda \tag{A.30}$$

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.30) reduces to (A.29) 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 $< x, y >_{J_n} = x' J_n y$. 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 an hypercube, increases to Λ. This projection can be written as $\sqrt{n}(\theta_{J_n}(Z_n) - \theta_0)$ with

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

The proof of Theorem 3.1 rests on a quadratic expansion about θ_0 of the quasi-likelihood function. Using a Taylor expansion for a function with right partial derivatives (see Andrews, Theorem 6, 1999) we get for all θ and θ_0 in Θ,

$$\begin{aligned}
\tilde{\mathbf{l}}_n(\theta) &= \tilde{\mathbf{l}}_n(\theta_0) + \frac{\partial\tilde{\mathbf{l}}_n(\theta_0)}{\partial\theta'}(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_0)}{\partial\theta\partial\theta'}(\theta - \theta_0) + R_n(\theta) && \text{(A.31)}\\
&= \tilde{\mathbf{l}}_n(\theta_0) - \frac{1}{2n}Z_n' J_n\sqrt{n}(\theta - \theta_0) - \frac{1}{2n}\sqrt{n}(\theta - \theta_0)' J_n Z_n \\
&\quad + \frac{1}{2}(\theta - \theta_0)' J_n(\theta - \theta_0) + R_n(\theta) + R_n^*(\theta) \\
&= \tilde{\mathbf{l}}_n(\theta_0) + \frac{1}{2n}\|Z_n - \sqrt{n}(\theta - \theta_0)\|_{J_n}^2 - \frac{1}{2n}Z_n' J_n Z_n + R_n(\theta) + R_n^*(\theta), && \text{(A.32)}
\end{aligned}$$

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 $\sqrt{n}(\theta_n - \theta_0) = O_P(1)$,
$R_n(\theta_n) = o_P(n^{-1})$, $R_n^*(\theta_n) = o_P(n^{-1})$,

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 A.2, the claim that $\|x\|_{J_n}$ is a norm, *a.s.* for n large, is justified. The triangle inequality gives

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

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 A.2. 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.31) satisfies

$$R_n(\theta) = \frac{1}{2}(\theta - \theta_0)' \left\{ \left[\frac{\partial^2 \tilde{\mathbf{l}}_n(\theta^*_{ij})}{\partial \theta \partial \theta'} \right] - \frac{\partial^2 \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta \partial \theta'} \right\} (\theta - \theta_0). \tag{A.33}$$

When $\theta = \hat{\theta}_n$, by (A.22) and (A.23), the term into accolades 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.31) is given by

$$R^*_n(\theta) = \left\{ \frac{\partial \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta} - \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta} \right\} (\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \left\{ \frac{\partial^2 \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta \partial \theta'} - J_n \right\} (\theta - \theta_0). \tag{A.34}$$

Therefore, in view of (A.21)-(A.22) 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

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

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

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

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

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

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

A straightforward extension of (A.23) allows to replace $\hat{\theta}_n$ by θ_n, with *a.s.* convergence replaced by convergence in probability. Therefore, $R_n(\theta_n) = o_P(\|\theta_n - \theta_0)\|^2) = o_P(n^{-1})$, by (A.33), which proves the first part of iii). The second equality similarly follows from (A.34) 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.32), 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

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

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 $< x, y >_{J_n}$, it is characterized by $\lambda_n^\Lambda \in \Lambda, \quad \left\langle Z_n - \lambda_n^\Lambda, \lambda_n^\Lambda - \lambda\right\rangle_{J_n} \geq 0, \quad \forall \lambda \in \Lambda$, see e.g. Zarantonello (1971), Lemma 1.1 pp. 239. Thus

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

Hence, by iv)

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

and v) is proved.

The continuous mapping theorem entails vi), because $(Z_n, J_n) \xrightarrow{\mathcal{L}} (Z, J)$ by Lemma A.2, $\lambda_n^\Lambda = f(Z_n, J_n)$ 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 3.1 readily follows from v) and vi).

A.3. Proof of Theorem 4.1.

Throughout, all expectations are taken with respect to the distribution of (η_t). Let $\ell_{t,n}(\theta) = \frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2(\theta)} + \log \sigma_{t,n}^2(\theta)$, so that the theoretical and empirical objective functions can still be denoted $\mathbf{l}_n(\theta) = n^{-1}\sum_{t=1}^n \ell_{t,n}(\theta)$, and $\tilde{\mathbf{l}}_n(\theta) = n^{-1}\sum_{t=1}^n \tilde{\ell}_{t,n}(\theta)$.

Denote by $A_{0t,n}$ the matrix obtained by substituting θ_n for θ_0 in the definition of A_{0t}. The following inequalities, which are straightforward consequences of $\tau > 0$, will be used throughout. For any $n \geq n_0$, we have $A_{0t,n_0} \geq A_{0t,n} \geq A_{0t}$ (componentwise), and thus, under **A2**, for $n \geq n_0$ and n_0 sufficiently large

$$\epsilon_{t,n_0}^2 \geq \epsilon_{t,n}^2 \geq \epsilon_t^2, \quad \text{and} \quad \sigma_{t,n_0}^2(\theta) \geq \sigma_{t,n}^2(\theta^*) \geq \sigma_t^2(\tilde{\theta}) \quad \text{for any } \theta \geq \theta^* \geq \tilde{\theta}. \quad \text{(A.35)}$$

A.3.1. *Consistency of* $\hat{\theta}_n$. Following the scheme of proof of Theorem 2.1 in FZ, we will establish the following intermediate results.

$$
\begin{aligned}
&i)\ \lim_{n\to\infty} \sup_{\theta\in\Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| = 0, \quad a.s.\\
&ii)\ \lim_{n\to\infty} \mathbf{l}_n(\theta_n) = E\ell_t(\theta_0), \quad a.s.\\
&iii)\ \text{for any } \theta \neq \theta_0 \text{ there exists a neighborhood } V(\theta) \text{ such that}\\
&\qquad \liminf_{n\to\infty} \inf_{\theta^*\in V(\theta)} \tilde{\mathbf{l}}_n(\theta^*) > E\ell_1(\theta_0), \quad a.s.
\end{aligned}
$$

First we show *i)*. Similar to (A.2) we have $\sigma_{t,n}^2(\theta) = \sum_{k=0}^{\infty} B^k(1,1)c_{t-k,n}$, where $c_{t,n} = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i,n}^2$. Let $\tilde{c}_{t,n}$ be obtained by replacing $\epsilon_{0,n}^2, \dots, \epsilon_{1-q,n}^2$ by their initial values in $c_{t,n}$. We have

$$
\tilde{\sigma}_{t,n}^2 = \sum_{k=0}^{t-(q+1)} B^k(1,1)c_{t-k,n} + \sum_{k=t-q}^{t-1} B^k(1,1)\tilde{c}_{t-k,n} + B^t(1,1)\tilde{\sigma}_0^2.
$$

Thus, almost surely,

$$
\begin{aligned}
\sup_{\theta\in\Theta} |\sigma_{t,n}^2 - \tilde{\sigma}_{t,n}^2| &= \sup_{\theta\in\Theta} \left| \sum_{k=1}^{q} B^{t-k}(1,1)\left(c_{k,n} - \tilde{c}_{k,n}\right) + B^t(1,1)\left(\sigma_{0,n}^2 - \tilde{\sigma}_0^2\right) \right| \\
&\leq \sup_{\theta\in\Theta} \left\{ \sum_{k=1}^{q} B^{t-k}(1,1)\left(c_{k,n_0} + \tilde{c}_{k,n_0}\right) + B^t(1,1)\left(\sigma_{0,n_0}^2 + \tilde{\sigma}_0^2\right) \right\} \\
&\leq K\rho^t, \quad \forall t. \qquad\qquad \text{(A.36)}
\end{aligned}
$$

Proceeding as in FZ (2004), we obtain, almost surely, for $n \geq n_0$,

$$
\sup_{\theta\in\Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_{n,\tau}(\theta)| \leq Kn^{-1}\sum_{t=1}^{n} \rho^t \epsilon_{t,n}^2 + Kn^{-1}\sum_{t=1}^{n}\rho^t \leq Kn^{-1}\sum_{t=1}^{n}\rho^t\epsilon_{t,n_0}^2 + Kn^{-1}.
$$

The a.s. convergence of $n^{-1}\sum_{t=1}^{n}\rho^t\epsilon_{t,n_0}^2$ to 0 follows by the arguments used in the aforementioned paper, provided n_0 is sufficiently large so that $\gamma(A_{0n_0}) < 0$. Hence *i)* is established.

Now we will prove *ii)*. We have

$$
\mathbf{l}_n(\theta_n) = \frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2} + \log\sigma_{t,n}^2 = \frac{1}{n}\sum_{t=1}^{n}\eta_t^2 + \frac{1}{n}\sum_{t=1}^{n}\log\sigma_{t,n}^2.
$$

In the right-hand side of the last equality, the first sample mean converges to 1, a.s., and the second one is between $\frac{1}{n}\sum_{t=1}^{n}\log\sigma_t^2$ and $\frac{1}{n}\sum_{t=1}^{n}\log\sigma_{t,n_0}^2$. By the ergodic theorem, these sample means a.s. converge to $E\log\sigma_t^2$ and $E\log\sigma_{t,n_0}^2$ respectively, when $n\to\infty$ (the existence of such expectations was shown in FZ (2004), Proof of Theorem 2.1, under the strict stationarity condition). The latter expectation decreases to the former one when n_0 tends to infinity, which establishes *ii)*.

It remains to show *iii)*. For any $\theta\in\Theta$ and any positive integer k, let $V_k(\theta)$ be the open ball with center θ and radius $1/k$. Proceeding as in FZ (2004), and in view of

(A.35), we find that

$$\begin{aligned}\liminf_{n\to\infty}\inf_{\theta^*\in V_k(\theta)\cap\Theta}\tilde{\mathbf{l}}_n(\theta^*) &\geq \liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta}\ell_{t,n}(\theta^*)\\ &= \liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta}\left(\log\sigma_{t,n}^2+\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}\right)(\theta^*)\\ &\geq \liminf_{n\to\infty} n^{-1}\sum_{t=1}^n \inf_{\theta^*\in V_k(\theta)\cap\Theta}\left(\log\sigma_{t}^2+\frac{\epsilon_{t}^2}{\sigma_{t,n_0}^2}\right)(\theta^*)\\ &= E\inf_{\theta^*\in V_k(\theta)\cap\Theta}\left(\log\sigma_{t}^2+\frac{\epsilon_{t}^2}{\sigma_{t,n_0}^2}\right)(\theta^*).\end{aligned}$$

The last equality follows from the ergodic theorem for stationary and ergodic processes (X_t) such that $E(X_t)$ exists in $\mathbb{R}\cup\{+\infty\}$ (see Billingsley (1995) p. 284 and 495). In the last equality, the infimum is larger than $\inf_{\theta^*\in\Theta}(\log\omega^*)$ which ensures the existence of its expectation. By the Beppo-Levi theorem, when k and n_0 increase to ∞, $E\inf_{\theta^*\in V_k(\theta)\cap\Theta}\left(\log\sigma_t^2+\frac{\epsilon_t^2}{\sigma_{t,n_0}^2}\right)(\theta^*)$ increases to $E\ell_1(\theta)$. In view of $E\ell_1(\theta)>E\ell_1(\theta_0)$, which was shown in FZ (2004), *iii)* is proved. This completes the proof of the strong consistency of $\hat\theta_n$.

A.3.2. *Asymptotic normality of the score at θ_n.* For the sake of brevity we will only establish the asymptotic distribution of $\hat\theta_n$ under the assumptions **A2**–**A6** and **A8**. The proof can be straightforwardly adapted when **A7**, instead of **A8**, holds. We will show that, when n tends to infinity

$$n^{-1/2}\sum_{t=1}^n\frac{\partial}{\partial\theta}\tilde\ell_{t,n}(\theta_n)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,(\kappa_\eta-1)J\right), \tag{A.37}$$

and

$$n^{-1}\sum_{t=1}^n\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde\ell_{t,n}(\theta_{ij}^*)\xrightarrow{P}J(i,j), \tag{A.38}$$

for any θ_{ij}^* between θ_n and $\hat\theta_n$. Let n_0 be a sufficiently large integer so that $\gamma(A_{0n_0})<0$ and $\theta_{n_0}\in\overset{\circ}{\Theta}$. We will show that

$$a)\ E\left\|\frac{\partial\ell_{t,n_0}(\theta_{n_0})}{\partial\theta}\frac{\partial\ell_{t,n_0}(\theta_{n_0})}{\partial\theta'}\right\|<\infty,\quad E\left\|\frac{\partial^2\ell_{t,n_0}(\theta_{n_0})}{\partial\theta\partial\theta'}\right\|<\infty,$$

$$b)\ n^{-1/2}\sum_{t=1}^n\frac{\partial}{\partial\theta}\ell_{t,n}(\theta_n)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,(\kappa_\eta-1)J\right), \tag{A.39}$$

$$c)\ E\sup_{n\geq n_0}\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left\|\frac{\partial^2\ell_{t,n}(\theta)}{\partial\theta\partial\theta'}\right\|<\infty,$$

$$d)\ \left\|n^{-1/2}\sum_{t=1}^n\left\{\frac{\partial\ell_{t,n}(\theta_n)}{\partial\theta}-\frac{\partial\tilde\ell_{t,n}(\theta_n)}{\partial\theta}\right\}\right\|\to0\ \text{ and} \tag{A.40}$$

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left\|n^{-1}\sum_{t=1}^n\left\{\frac{\partial^2\ell_{t,n}(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\tilde\ell_{t,n}(\theta)}{\partial\theta\partial\theta'}\right\}\right\|\xrightarrow{P}0, \tag{A.41}$$

$$e)\ n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_{t,n}(\theta_n)\to J(i,j)\ a.s, \tag{A.42}$$

$$f)\ \text{ for all } i,j,k\in\{1,\dots,p+q+1\},\quad E\sup_{n\geq n_0}\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left|\frac{\partial^3\ell_{t,n}(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|<\infty,$$

for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0. First notice that, formulas similar to (A.3), (A.4), and (A.5) hold with σ_t^2 (resp. ϵ_{t-k-i}^2) replaced by $\sigma_{t,n}^2$ (resp. $\epsilon_{t-k-i,n}^2$). When the derivatives are taken at θ_n, also the coefficients α_i and β_j have to be replaced by $\alpha_{i,n}$ and $\beta_{j,n}$. We begin to show that (A.37) and (A.38) follow from a)-f).

Proof of (A.37) and (A.38). The convergence (A.37) is a straightforward consequence of b) and the first part of d). To show (A.38) we start by using the second part of d) and the strong consistency, to prove that $\tilde{\ell}_{t,n}(\theta_{ij}^*)$ can be replaced by $\ell_{t,n}(\theta_{ij}^*)$. Then we make the Taylor expansion

$$n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_{t,n}(\theta_{ij}^*)=n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_{t,n}(\theta_n)+(\theta_{ij}^*-\theta_n)'n^{-1}\sum_{t=1}^{n}\frac{\partial^3}{\partial\theta\partial\theta_i\partial\theta_j}\ell_{t,n}(\theta_{ij}^{**}),$$

where θ_{ij}^{**} is between θ_{ij}^* and θ_n. To conclude, we use e), f) and again the strong consistency.

Proof of a)-f). Since θ_{n_0} belongs to the interior of Θ, a) is a consequence of the properties established in FZ (2004) (proof of Theorem 2.2). Turning to b), we first note that the central limit theorem used in the proof of Lemma A.2 does not apply here. Given that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ell_{t,n}(\theta_n)\quad=\quad n^{-1/2}\sum_{t=1}^{n}(1-\eta_t^2)\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\theta}:=n^{-1/2}\sum_{t=1}^{n}X_{t,n},$$

we will use the Lindeberg central limit theorem for triangular arrays of martingale differences. Indeed, recall that $\sigma_{t,n}^2$ and its derivatives are measurable with respect to the $\sigma-$field $\mathcal{F}_{t-1}$ generated by the variables $\eta_{t-i},\ i>0$. It follows that for any $n\geq 1$, $\{X_{t,n},\mathcal{F}_{t-1}\}_t$ is a strictly stationary martingale difference. Under the assumptions of the theorem, $(X_{t,n})$ is clearly square integrable for n large enough, because θ_n belongs to the interior of Θ (see FZ (2004)). Let $\lambda\in\mathbb{R}^{p+q+1}$, let $x_{t,n}=\lambda'X_{t,n}$ and let

$$s_{t,n}^2=E(x_{t,n}^2\mid\mathcal{F}_{t-1})=(\kappa_\eta-1)\lambda'\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'}\lambda.$$

Using the Wold-Cramer device it will be sufficient to show that

$$\frac{1}{n}\sum_{t=1}^{n}s_{t,n}^2\xrightarrow{P}(\kappa_\eta-1)\lambda'J\lambda,\quad\text{ and} \tag{A.43}$$

$$\frac{1}{n}\sum_{t=1}^{n}E(x_{t,n}^2\mathbb{1}_{|x_{t,n}|\geq n^{1/2}\varepsilon})\to 0,\quad\text{when } n\to\infty, \tag{A.44}$$

for any $\varepsilon>0$. First consider the derivative of $\sigma_{t,n}^2$ with respect to β_j. In view of (A.4), this derivative evaluated at θ_n is given by

$$\frac{\partial\sigma_{t,n}^2}{\partial\beta_j}=\sum_{k=1}^{\infty}B_{k,j;n}(1,1)\left(\omega_n+\sum_{i=1}^{q}\alpha_{i,n}\epsilon_{t-k-i,n}^2\right),$$

where $B_{k,j;n}$ is the matrix obtained from $B_{k,j}$ by replacing the coefficients β_i by $\beta_{i,n}$. Similarly we have, by (A.3), $\sigma^2_{t,n} = \sum_{k=0}^{\infty} B_n^k(1,1)\left(\omega_n + \sum_{i=1}^q \alpha_{i,n}\epsilon^2_{t-k-i,n}\right)$. Denote by ${}_j\sigma^2_{t,n}$ (resp. ${}^j\sigma^2_{t,n}$) the variable obtained by replacing $\epsilon^2_{t-j,n}$ by ϵ^2_{t-j,n_0} (resp. ϵ^2_{t-j}) in the expansion of $\sigma^2_{t,n}$. Denote by ${}_j\sigma^2_t$ (resp. ${}^j\sigma^2_t$) the variable obtained by replacing the variables $\epsilon^2_{t-i,n}$ by ϵ^2_{t-i} (resp. by ϵ^2_{t-i,n_0}, and ϵ^2_{t-j,n_0} by ϵ^2_{t-j}) in ${}_j\sigma^2_{t,n}$, and the coefficients of θ_n by those of θ_0 (resp. θ_{n_0}). To make it clear, let us consider the example of a GARCH(1,1): we have $\sigma^2_{t,n} = \frac{\omega_n}{1-\beta_n} + \alpha_n \sum_{i\geq 1}\beta_n^{i-1}\epsilon^2_{t-i,n}$, ${}_j\sigma^2_{t,n} = \frac{\omega_n}{1-\beta_n} + \alpha_n\beta_n^{j-1}\epsilon^2_{t-j,n_0} + \alpha_n\sum_{i\geq 1, i\neq j}\beta_n^{i-1}\epsilon^2_{t-i,n}$ and ${}^j\sigma^2_{t,n} = \frac{\omega_n}{1-\beta_n} + \alpha_n\beta_n^{j-1}\epsilon^2_{t-j} + \alpha_n\sum_{i\geq 1, i\neq j}\beta_n^{i-1}\epsilon^2_{t-i,n}$, whereas ${}_j\sigma^2_t = \frac{\omega_0}{1-\beta_0} + \alpha_0\beta_0^{j-1}\epsilon^2_{t-j,n_0} + \alpha_0\sum_{i\geq 1, i\neq j}\beta_0^{i-1}\epsilon^2_{t-i}$ and ${}^j\sigma^2_t = \frac{\omega_{n_0}}{1-\beta_{n_0}} + \alpha_{n_0}\beta_{n_0}^{j-1}\epsilon^2_{t-j} + \alpha_{n_0}\sum_{i\geq 1, i\neq j}\beta_{n_0}^{i-1}\epsilon^2_{t-i,n_0}$. Notice that for any constants $a > 0$ and $b > 0$, the function $x \to x/(a+bx)$ is increasing over the positive line. Considering $\sigma^2_{t,n}$ as a function of ϵ^2_{t-j}, for $j > 0$, it follows that, using (A.35),

$$\frac{\epsilon^2_{t-j}}{{}^j\sigma^2_{t,n}} \;\leq\; \frac{\epsilon^2_{t-j,n}}{\sigma^2_{t,n}} \;\leq\; \frac{\epsilon^2_{t-j,n_0}}{{}_j\sigma^2_{t,n}}.$$

We also have, from (A.5)

$$B_{k,j;n} \;=\; \sum_{m=1}^{k} B_n^{m-1}B^{(j)}B_n^{k-m} \leq \sum_{m=1}^{k} B_{n_0}^{m-1}B^{(j)}B_{n_0}^{k-m} = B_{k,j;n_0}.$$

In view of the last inequalities, and (A.35), we have for $j = 1,\dots,p$,

$$\begin{aligned}
\frac{1}{\sigma^2_{t,n}}\frac{\partial\sigma^2_{t,n}}{\partial\beta_j} \;&\leq\; \sum_{k=1}^{\infty} B_{k,j;n}(1,1)\left(\omega_n + \sum_{i=1}^{q}\alpha_{i,n}\frac{\epsilon^2_{t-k-i,n_0}}{{}_{k+i}\sigma^2_{t,n}}\right)\\
&\leq\; \sum_{k=1}^{\infty} B_{k,j;n_0}(1,1)\left(\omega_{n_0} + \sum_{i=1}^{q}\alpha_{i,n_0}\frac{\epsilon^2_{t-k-i,n_0}}{{}_{k+i}\sigma^2_{t}}\right). \qquad \text{(A.45)}
\end{aligned}$$

The last inequality uses the fact that the components of θ_n are decreasing functions of n, and that all the quantities involved, in particular $B_{k,j;n}(1,1)$, are nonnegative. Similarly we have,

$$\frac{1}{\sigma^2_{t,n}}\frac{\partial\sigma^2_{t,n}}{\partial\beta_j} \;\geq\; \sum_{k=1}^{\infty} B_{k,j}(1,1)\left(\omega_0 + \sum_{i=1}^{q}\alpha_{0i}\frac{\epsilon^2_{t-k-i}}{{}^{k+i}\sigma^2_{t}}\right).$$

Similar lower and upper bounds hold for $\sigma^{-2}_{t,n}\frac{\partial\sigma^2_{t,n}}{\partial\alpha_i}$, $i = 1,\dots,q$ and $\sigma^{-2}_{t,n}\frac{\partial\sigma^2_{t,n}}{\partial\omega}$. It follows that

$$Y_t^{(1)}(n_0) \leq \frac{1}{\sigma^2_{t,n}}\frac{\partial\sigma^2_{t,n}}{\partial\theta} \leq Y_t^{(2)}(n_0) \qquad \text{(A.46)}$$

for some $(\mathbb{R}^+)^{p+q+1}$-valued, strictly stationary, processes $(Y_t^{(1)}(n_0))$ and $(Y_t^{(2)}(n_0))$. Because the vectors involved in the last inequality have positive components, it follows that

$$Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)' \leq \frac{1}{\sigma^4_{t,n}}\frac{\partial\sigma^2_{t,n}}{\partial\theta}\frac{\partial\sigma^2_{t,n}}{\partial\theta'} \leq Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)', \qquad \text{(A.47)}$$

componentwise. Note that the lower and upper bounds obtained for the matrix inside the inequalities are independent of n, whenever $n \geq n_0$. The ergodic theorem applies to $n^{-1}\sum_{t=1}^{n} Y_t^{(i)}(n_0)Y_t^{(i)}(n_0)'$ $(i=1,2)$ provided the expectation of $Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)'$ is finite. This can be shown by exactly the same techniques as those employed to establish Lemma A.1. More precisely, if **A8** holds true, proceeding as in the calculations leading to (A.14), we obtain an upper bound for the right-hand side of (A.45) as

$$\begin{aligned} Y_{q+1+j,t}^{(2)}(n_0) &\leq \omega_{n_0}\sum_{k=j}^{\infty} kB_{n_0}^{k-j}(1,1) + \sum_{k=j+1}^{\infty}\sum_{\ell=1}^{k-j}\alpha_{n_0\ell}kB_{n_0}^{k-\ell-j}(1,1)\frac{\epsilon_{t-k,n_0}^2}{{}_k\sigma_t^2} \\ &\leq K_{n_0} + \sum_{k=j+1}^{\infty}\sum_{\ell=1}^{k-j}\alpha_{n_0\ell}k\frac{B_{n_0}^{k-\ell-j}(1,1)\epsilon_{t-k,n_0}^{2s}}{\omega_0^s\underline{\alpha}^{1-s}\{B_0^{k-i_k}(1,1)\}^{1-s}}, \end{aligned}$$

where $Y_t^{(2)}(n_0) = (Y_{it}^{(2)}(n_0))_{1\leq i\leq p+q+1}$, for some positive constant $\underline{\alpha}$ and for any $s \in (0,1)$. It turns out that $Y_{q+1+j,t}^{(2)}(n_0)$ admit moments at any order. The same conclusion holds for the other components of $Y_t^{(2)}(n_0)$. It follows that $n^{-1}\sum_{t=1}^{n} Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)'$ $\xrightarrow{P} EY_t^{(2)}(n_0)Y_t^{(2)}(n_0)'$. By the Lebesgue theorem, this expectation converges to J when $n_0 \to \infty$. Similarly $n^{-1}\sum_{t=1}^{n} Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)' \xrightarrow{P} J$ when n and n_0 tend to infinity. In view of (A.47) we can conclude that

$$n^{-1}\sum_{t=1}^{n}\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} \to J \quad \text{in probability when } n \text{ tends to infinity,}$$

from which (A.43) straightforwardly follows. To prove (A.44) we first remark that the expectations in the right-hand side are independent of t, by strict stationarity of $(x_{t,n})$. In addition, the previous arguments show that $x_{t,n}$ admits moments at any order, which are bounded when n increases. By the Schwarz and Markov inequalities the convergence in (A.44) follows and the proof of b) is complete.

Now we prove c). The second derivative of $\ell_{t,n}(\theta)$ is given by

$$\frac{\partial^2\ell_{t,n}}{\partial\theta\partial\theta'} = \left\{1-\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}\right\}\frac{1}{\sigma_{t,n}^2}\frac{\partial^2\sigma_{t,n}^2}{\partial\theta\partial\theta'} + \left\{2\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}-1\right\}\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'}. \tag{A.48}$$

First we will show that a formula similar to (A.46) holds in some neighborhood $\mathcal{V}(\theta_0)$ of θ_0. Let n_0 be large enough so that $\theta_{n_0} \in \mathcal{V}(\theta_0)$. Let ${}_j\underline{\sigma}_t^2$ be obtained by replacing in ${}_j\sigma_t^2$, componentwise, θ_0 by the infimum of θ over $\mathcal{V}(\theta_0)\cap\Theta$. Then, in view of (A.45)

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\beta_j}(\theta) \leq \sum_{k=1}^{\infty}\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} B_{k,j}(1,1)\left(\omega+\sum_{i=1}^{q}\alpha_i\frac{\epsilon_{t-k-i,n_0}^2}{{}_{k+i}\underline{\sigma}_t^2}\right).$$

Note that, under **A8**, for $\mathcal{V}(\theta_0)$ sufficiently small, ϵ_{t-k-i,n_0}^2 appears in the expansion of ${}_{k+i}\underline{\sigma}_t^2$, by continuity arguments. Note also that the derivatives are nonnegative as can be seen from (A.3)-(A.4). Therefore, exactly the same arguments as those used to show b) apply, to establish that,

$$0 \leq \sup_{n\geq n_0}\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\theta}(\theta) \leq Y_t^{(3)}(n_0), \tag{A.49}$$

for some vector $Y_t^{(3)}(n_0)$ admitting moments at any order. Similar arguments show that for $i, j = 1, \ldots, p$,

$$0 \leq \sup_{n \geq n_0} \sup_{\theta \in \mathcal{V}(\theta_0) \cap \Theta} \frac{1}{\sigma_{t,n}^2} \frac{\partial^2 \sigma_{t,n}^2}{\partial \theta_i \partial \theta_j}(\theta) \ \leq \ Y_{i,j,t}^{(4)}(n_0), \tag{A.50}$$

for some variables $Y_{i,j,t}^{(4)}(n_0)$ admitting moments at any order.

To handle terms of (A.48) involving

$$\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2(\theta)} = \eta_t^2 \frac{\sigma_{t,n}^2(\theta_n)}{\sigma_{t,n}^2(\theta)},$$

we will use the expansion $\sigma_{t,n}^2(\theta) = c + \sum_{j=1}^{\infty} b_j \epsilon_{t-j,n}^2$ where $b_j = \sum_{\ell=1}^{j} \alpha_j B^{j-\ell}(1,1)$. Note that $b_j > 0$ over $\mathcal{V}(\theta_0) \cap \Theta$. Let $\delta > 0$. Using again the elementary inequality $ax/(b+cx) \leq ax^s/(b^s c^{1-s})$ for all $a, b, c, x \geq 0$ and any $s \in (0,1)$, we obtain, for $\mathcal{V}(\theta_0)$ sufficiently small

$$\frac{\sigma_{t,n}^2(\theta_n)}{\sigma_{t,n}^2(\theta)} \leq K + K \sum_{j=1}^{\infty} \frac{b_{j,n_0}}{b_j} b_j^s \epsilon_{t-j,n_0}^s \leq K + K \sum_{j=1}^{\infty} (1+\delta)^j \rho^{js} \epsilon_{t-j,n_0}^s, \tag{A.51}$$

uniformly in $\theta \in \mathcal{V}(\theta_0) \cap \Theta$, for some $\rho < 1$. The last inequality uses the fact that for n_0 sufficiently large, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that $B_{n_0} \leq (1+\delta)B$ for all $\theta \in \mathcal{V}(\theta_0) \cap \Theta$. Choosing s such that $E\epsilon_{t,n_0}^{2s} < \infty$ and, for instance, $\delta = (1-\rho^s)/(2\rho^s)$ we obtain

$$E \sup_{n \geq n_0} \sup_{\theta \in \mathcal{V}(\theta_0) \cap \Theta} \frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2(\theta)} = E \sup_{n \geq n_0} \sup_{\theta \in \mathcal{V}(\theta_0) \cap \Theta} \frac{\sigma_{t,n}^2(\theta_n)}{\sigma_{t,n}^2(\theta)} < \infty.$$

For the same choice of δ, with s such that $E\epsilon_t^{4s} < \infty$, and using (A.51), we find

$$\begin{aligned} \left\| \sup_{n \geq n_0} \sup_{\theta \in \mathcal{V}(\theta_0) \cap \Theta} \frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2(\theta)} \right\|_2 &= \kappa_\eta^{1/2} \left\| \sup_{n \geq n_0} \sup_{\theta \in \mathcal{V}(\theta_0) \cap \Theta} \frac{\sigma_{t,n}^2(\theta_n)}{\sigma_{t,n}^2(\theta)} \right\|_2 \\ \leq \quad K + K \sum_{j=1}^{\infty} (1+\delta)^j \rho^{js} \left\| \epsilon_{t,n}^{2s} \right\|_2 &< \infty. \end{aligned}$$

Using (A.48), (A.49), (A.50), (A.52) and the Schwarz inequality, it is straightforward to conclude that c) holds.

To prove d) first note that, analogue to (A.36), we have almost surely

$$\sup_{\theta \in \Theta} \left| \frac{\partial \sigma_{t,n}^2}{\partial \theta} - \frac{\partial \tilde{\sigma}_{t,n}^2}{\partial \theta} \right| \quad \leq \quad K\rho^t, \quad \sup_{\theta \in \Theta} \left| \frac{\partial^2 \sigma_{t,n}^2}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\sigma}_{t,n}^2}{\partial \theta \partial \theta'} \right| \leq K\rho^t, \quad \forall t$$

where K does not depend on n. It follows that

$$\begin{aligned} \left| n^{-1/2} \sum_{t=1}^{n} \left\{ \frac{\partial \ell_{t,n}(\theta_n)}{\partial \theta_i} - \frac{\partial \tilde{\ell}_{t,n}(\theta_n)}{\partial \theta_i} \right\} \right| \quad &\leq \quad K^* n^{-1/2} \sum_{t=1}^{n} \rho^t (1+\eta_t^2) \left\{ 1 + \frac{1}{\sigma_{t,n}^2} \frac{\partial \sigma_{t,n}^2}{\partial \theta_i} \right\}, \\ &\leq \quad K^* n^{-1/2} \sum_{t=1}^{n} \rho^t (1+\eta_t^2) \left\{ 1 + Y_{it}^{(2)}(n_0) \right\}, \end{aligned}$$

where $Y_{it}^{(2)}(n_0)$ is the i-th component of $Y_t^{(2)}(n_0)$ introduced in (A.46). The Markov inequality and the independence between η_t and $Y_t^{(2)}(n_0)$ allow to show the first convergence in d). By similarity with the proof of Theorem 3.1, we find that the supremum in d) is bounded by $Kn^{-1}\sum_{t=1}^n \rho^t \Upsilon_{t,n}$, where

$$\Upsilon_{t,n} = \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\{1+\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}\right\}\left\{1+\frac{1}{\sigma_{t,n}^2}\frac{\partial^2\sigma_{t,n}^2}{\partial\theta_i\partial\theta_j}+\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\theta_i}\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\theta_j}\right\}.$$

We have

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left\{1+\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}\right\} \le K(1+\epsilon_{t,n}^2) \le K(1+\epsilon_{t,n_0}^2),$$

where the right-hand side admits a moment of order $3s$. In view of the results established in the proof of c), it follows that $E\Upsilon_{t,n}^s < K$. The rest of the proof is identical to that of d) in the proof Theorem 3.1.

Now we show e). First consider the second group of terms in the second derivative of $\ell_{t,n}$, displayed in (A.48), at the value θ_n. In view of (A.47), we have

$$\begin{aligned} n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} \quad &\le \quad n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\mathbb{1}_{2\eta_t^2\ge 1}Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)' \\ &\quad +n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\mathbb{1}_{2\eta_t^2<1}Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)'. \end{aligned}$$

The ergodic theorem applies to the sums of the right hand side and yields, a.s.

$$\begin{aligned} \limsup_{n\to\infty} n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} \quad &\le \quad E\{(2\eta_t^2-1)\mathbb{1}_{2\eta_t^2\ge 1}\}E\{Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)'\} \\ &\quad +E\{(2\eta_t^2-1)\mathbb{1}_{2\eta_t^2<1}\}E\{Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)'\} \end{aligned}$$

from the independence between η_t and the variables $Y_t^{(i)}(n_0)$. We have already seen that $E\{Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)'\} \to J$, for $i = 1, 2$, as $n_0 \to \infty$. It follows that

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} \quad \le \quad E\{(2\eta_t^2-1)(\mathbb{1}_{2\eta_t^2\ge 1}+\mathbb{1}_{2\eta_t^2<1})\}J = J.$$

Similarly we have

$$\begin{aligned} \liminf_{n\to\infty} n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} \quad &\ge \quad E\{(2\eta_t^2-1)\mathbb{1}_{2\eta_t^2\ge 1}\}E\{Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)'\} \\ &\quad +E\{(2\eta_t^2-1)\mathbb{1}_{2\eta_t^2<1}\}E\{Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)'\}, \end{aligned}$$

which converges to J as $n_0 \to \infty$. Thus we have proved that, a.s.

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^n(2\eta_t^2-1)\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} = J.$$

The first group of terms in the right-hand side of (A.48) can be treated analogously, using lower and upper bounds for $\sigma_{t,n}^{-2}\frac{\partial^2\sigma_{t,n}^2}{\partial\theta\partial\theta'}$. Therefore we have a.s.

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}(1-\eta_t^2)\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\theta\partial\theta'}=0.$$

The convergence in e) follows.

Finally, f) is proved in the same manner as c). Indeed, it can be seen from FZ that the third derivative of $\ell_{t,n}$ involves products of terms already encountered, plus a term involving the third derivative of $\sigma_{t,n}^2$ divided by $\sigma_{t,n}^2$. This term can be bounded independently of n, as in (A.49) and (A.50), which allows to conclude.

The next step is to prove the analogue of i)-vi) in the proof of Theorem 3.1.

A.3.3. *Asymptotic distribution of* $\hat\theta_n$. We start by introducing some notations. Let, for n sufficiently large

$$J_{n,\tau}=\frac{\partial^2\mathbf{l}_n(\theta_n)}{\partial\theta\partial\theta'},\qquad Z_{n,\tau}=-J_{n,\tau}^{-1}\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_n)}{\partial\theta},$$

where the non singularity of $J_{n,\tau}$ follows from (A.38), and let

$$\theta_{J_{n,\tau}}(Z_{n,\tau})=\arg\inf_{\theta\in\Theta}\|Z_{n,\tau}-\sqrt{n}(\theta-\theta_n)\|_{J_{n,\tau}},\qquad \lambda_{n,\tau}^{\Lambda}=\arg\inf_{\lambda\in\Lambda}\|Z_{n,\tau}+\tau-\lambda\|_{J_{n,\tau}}.$$

Similarly to (A.31), we have the following quadratic expansion of the quasi-likelihood function around θ_n

$$\tilde{\mathbf{l}}_n(\theta)\;=\;\tilde{\mathbf{l}}_n(\theta_n)+\frac{1}{2n}\|Z_{n,\tau}-\sqrt{n}(\theta-\theta_n)\|_{J_{n,\tau}}^2-\frac{1}{2n}Z_{n,\tau}'J_{n,\tau}Z_{n,\tau}+R_n(\theta),\quad\text{(A.52)}$$

where $R_n(\theta)$ is a remainder term. We will prove

i) $\sqrt{n}(\theta_{J_{n,\tau}}(Z_{n,\tau})-\theta_n)=O_P(1)$,

ii) $\sqrt{n}(\hat\theta_n-\theta_n)=O_P(1)$,

iii) for any sequence (θ_n^*) such that $\sqrt{n}(\theta_n^*-\theta_0)=O_P(1)$, $R_n(\theta_n^*)=o_P(n^{-1})$,

iv) $\|Z_{n,\tau}-\sqrt{n}(\hat\theta_n-\theta_n)\|_{J_{n,\tau}}^2\overset{o_P(1)}{=}\|Z_{n,\tau}+\tau-\lambda_{n,\tau}^{\Lambda}\|_{J_{n,\tau}}^2$,

v) $\sqrt{n}(\hat\theta_n-\theta_0)\overset{o_P(1)}{=}\lambda_{n,\tau}^{\Lambda}$,

vi) $\lambda_{n,\tau}^{\Lambda}\overset{\mathcal{L}}{\to}\lambda^{\Lambda}(\tau)$.

It suffices to adapt the arguments given in the proof of Theorem 3.1. For brevity we will only mention the points that need to be adapted.

In the proof of *i)* the same arguments apply, noting that $\|Z_{n,\tau}\|_{J_{n,\tau}}=O_P(1)$ because $J_{n,\tau}\overset{P}{\to}J$ by (A.42), and $\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_n)}{\partial\theta}=O_P(1)$ by (A.39).

The remainder term in (A.52) satisfies

$$\begin{aligned}R_n(\theta)\;=\;&\left\{n^{1/2}\left(\frac{\partial\tilde{\mathbf{l}}_n(\theta_n)}{\partial\theta}-\frac{\partial\mathbf{l}_n(\theta_n)}{\partial\theta}\right)\right\}n^{-1/2}(\theta-\theta_n)+\\&+\frac{1}{2}(\theta-\theta_n)'\left\{\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_n)}{\partial\theta\partial\theta'}-J_{n,\tau}+\left[\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_{ij}^*)}{\partial\theta\partial\theta'}\right]-\frac{\partial^2\tilde{\mathbf{l}}_n(\theta_n)}{\partial\theta\partial\theta'}\right\}(\theta-\theta_n),\end{aligned}$$

for some θ_{ij}^* between θ and θ_n. By (A.38) and the second part of (A.40), the last two terms into accolades tends to zero in probability as n tends to infinity. The first term into accolades converges to zero in probability by the first part of (A.40). To establish *ii)*, it is then straightforward to adjust the arguments given in the proof of Theorem 3.1. The same remark applies to the proof of *iii)*, and, noting that $\sqrt{n}(\theta_{J_{n,\tau}}(Z_{n,\tau}) - \theta_n) = \lambda_{n,\tau}^\Lambda$ for n sufficiently large, to that of *iv)*.

The vector $\lambda_{n,\tau}^\Lambda$ being the projection of $Z_{n,\tau} + \tau$ on the convex set Λ for the scalar product $< x, y >_{J_{n,\tau}}$, we have $\left\langle Z_{n,\tau} + \tau - \lambda_{n,\tau}^\Lambda, \lambda_{n,\tau}^\Lambda - \lambda \right\rangle_{J_{n,\tau}} \geq 0, \quad \forall \lambda \in \Lambda$. Thus, since $\sqrt{n}(\hat{\theta}_n - \theta_0) \in \Lambda$,

$$\begin{aligned} \left\| \sqrt{n}(\hat{\theta}_n - \theta_n) - Z_{n,\tau} \right\|_{J_{n,\tau}}^2 &= \left\| \sqrt{n}(\hat{\theta}_n - \theta_0) - (Z_{n,\tau} + \tau) \right\|_{J_{n,\tau}}^2 \\ &\geq \left\| \sqrt{n}(\hat{\theta}_n - \theta_0) - \lambda_n^\Lambda \right\|_{J_n}^2 + \left\| \lambda_n^\Lambda - (Z_{n,\tau} + \tau) \right\|_{J_{n,\tau}}^2 . \end{aligned}$$

Hence, *v)* follows from *iv)* and

$$\left\| \sqrt{n}(\hat{\theta}_n - \theta_0) - \lambda_n^\Lambda \right\|_{J_{n,\tau}}^2 \leq \| Z_{n,\tau} - \sqrt{n}(\hat{\theta}_n - \theta_n) \|_{J_{n,\tau}}^2 - \| Z_{n,\tau} + \tau - \lambda_n^\Lambda \|_{J_{n,\tau}}^2 = o_P(1).$$

Finally, *vi)* is proved by arguments already given.

A.4. Proof of Theorem 5.1

When $\tau = 0$, the convergence in distribution (5.2) is a direct application of the continuous mapping theorem, because $\sqrt{n}\hat{\theta}_n^{(2)'} = K\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} K\lambda^\Lambda$ under H_0 by Theorem 3.1. When $\tau > 0$ the same argument applies, based on Theorem 4.1.

We now turn to the proof of (5.3). Since $\hat{\theta}_{n|2}^{(1)}$ is a consistent estimator of $\theta_0^{(1)} > 0$, we have $\hat{\theta}_{n|2}^{(1)} > 0$ for n large enough. Therefore $\partial \tilde{\mathbf{l}}_n \left(\hat{\theta}_{n|2} \right) / \partial \theta_i = 0$ for $i = 1, \dots, d_1$, or equivalently

$$\frac{\partial \tilde{\mathbf{l}}_n \left(\hat{\theta}_{n|2} \right)}{\partial \theta} = K' \frac{\partial \tilde{\mathbf{l}}_n \left(\hat{\theta}_{n|2} \right)}{\partial \theta^{(2)}}. \tag{A.53}$$

A Taylor expansion and already given arguments yield

$$\sqrt{n} \frac{\partial \tilde{\mathbf{l}}_n (\hat{\theta}_{n|2})}{\partial \theta} \overset{o_P(1)}{=} \sqrt{n} \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta} + J\sqrt{n} \left(\hat{\theta}_{n|2} - \theta_0 \right). \tag{A.54}$$

The last d_2 components of this vector relation give

$$\sqrt{n} \frac{\partial \tilde{\mathbf{l}}_n (\hat{\theta}_{n|2})}{\partial \theta^{(2)}} \overset{o_P(1)}{=} \sqrt{n} \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(2)}} + KJ\sqrt{n} \left(\hat{\theta}_{n|2} - \theta_0 \right), \tag{A.55}$$

and the first d_1 components give

$$0 \overset{o_P(1)}{=} \sqrt{n} \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(1)}} + \overline{K} J \overline{K}' \sqrt{n} \left(\hat{\theta}_{n|2}^{(1)} - \theta_0^{(1)} \right), \tag{A.56}$$

using

$$\left(\hat{\theta}_{n|2} - \theta_0 \right) = \overline{K}' \left(\hat{\theta}_{n|2}^{(1)} - \theta_0^{(1)} \right). \tag{A.57}$$

In view of (A.56), we have

$$\sqrt{n}\left(\hat{\theta}^{(1)}_{n|2}-\theta^{(1)}_0\right)\overset{o_P(1)}{=}-\left(\overline{K}\hat{J}_{n|2}\overline{K}'\right)^{-1}\sqrt{n}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}}. \tag{A.58}$$

Using (A.53), (A.55), (A.57) and (A.58) we obtain

$$\begin{aligned}
\mathbf{R}_n &= \frac{n}{\hat{\kappa}_\eta-1}\frac{\partial \mathbf{l}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)\prime}}K\hat{J}^{-1}_{n|2}K'\frac{\partial \mathbf{l}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)}}\\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial \mathbf{l}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)}}\right\|^2_{KJ^{-1}K'}\\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(2)}}+KJ\overline{K}'\left(\hat{\theta}^{(1)}_{n|2}-\theta^{(1)}_0\right)\right\|^2_{KJ^{-1}K'}\\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(2)}}-KJ\overline{K}'\left(\overline{K}J\overline{K}'\right)^{-1}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}}\right\|^2_{KJ^{-1}K'}.
\end{aligned}$$

Now recall that under H_0

$$\begin{pmatrix} W_1\\ W_2\end{pmatrix} := \sqrt{\frac{n}{\kappa_\eta-1}}\begin{pmatrix}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}}\\ \frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(2)}}\end{pmatrix}\overset{\mathcal{L}}{\to}\mathcal{N}\left\{0, J=\begin{pmatrix}J_{11} & J_{12}\\ J_{21} & J_{22}\end{pmatrix}\right\}. \tag{A.59}$$

Using $KJ^{-1}K'=\left(J_{22}-J_{21}J_{11}^{-1}J_{12}\right)^{-1}$ it follows that the asymptotic distribution of $\mathbf{R}_n$ under H_0 is that of

$$\left(W_2-J_{21}J_{11}^{-1}W_1\right)'\left(J_{22}-J_{21}J_{11}^{-1}J_{12}\right)^{-1}\left(W_2-J_{21}J_{11}^{-1}W_1\right),$$

which follows the $\chi^2_{d_2}$ distribution since $W_2-J_{21}J_{11}^{-1}W_1\sim\mathcal{N}\left(0,J_{22}-J_{21}J_{11}^{-1}J_{12}\right)$.

Similarly, under $H_n(\tau)$, i.e. when $\theta_n=\theta_0+n^{-1/2}\tau$ obtains, we have

$$\sqrt{\kappa_\eta-1}\begin{pmatrix} W_1\\ W_2\end{pmatrix}\overset{o_P(1)}{=}\sqrt{n}\begin{pmatrix}\frac{\partial \mathbf{l}_n(\theta_n)}{\partial\theta^{(1)}}\\ \frac{\partial \mathbf{l}_n(\theta_n)}{\partial\theta^{(2)}}\end{pmatrix}+J\sqrt{n}(\theta_0-\theta_n)\overset{\mathcal{L}}{\to}\mathcal{N}\left\{-J\tau,(\kappa_\eta-1)J\right\}.$$

We then have

$$\left(W_2-J_{21}J_{11}^{-1}W_1\right)\sim\mathcal{N}\left\{-\left(J_{22}-J_{21}J_{11}^{-1}J_{12}\right)\frac{\tau^{(2)}}{\sqrt{\kappa_\eta-1}}, J_{22}-J_{21}J_{11}^{-1}J_{12}\right\},$$

and (5.3) follows.

To show (5.4), first note that $\mathbf{L}_n = n\left\{\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right)-\tilde{\mathbf{l}}_n\left(\hat{\theta}_n\right)\right\}$. Using (A.57) and (A.58), several Taylor expansions give

$$\begin{aligned}
n\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right) &\overset{o_P(1)}{=} n\mathbf{l}_n(\theta_0)+n\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta'}\left(\hat{\theta}_{n|2}-\theta_0\right)+\frac{n}{2}\left(\hat{\theta}_{n|2}-\theta_0\right)'J\left(\hat{\theta}_{n|2}-\theta_0\right)\\
&\overset{o_P(1)}{=} n\mathbf{l}_n(\theta_0)-\frac{n}{2}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(1)\prime}}\left(\overline{K}J\overline{K}'\right)^{-1}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}}
\end{aligned}$$

and

$$n\mathbf{l}_n\left(\hat{\theta}_n\right) \overset{o_P(1)}{=} n\mathbf{l}_n\left(\theta_0\right) + n\frac{\partial \mathbf{l}_n\left(\theta_0\right)}{\partial \theta'}\left(\hat{\theta}_n - \theta_0\right) + \frac{n}{2}\left(\hat{\theta}_n - \theta_0\right)' J\left(\hat{\theta}_n - \theta_0\right).$$

By subtraction,

$$\mathbf{L}_n \overset{o_P(1)}{=} -n\left\{\frac{1}{2}\frac{\partial \mathbf{l}_n\left(\theta_0\right)}{\partial \theta^{(1)'}}\left(\overline{K}J\overline{K}'\right)^{-1}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(1)}}\right. \tag{A.60}$$

$$\left. + \frac{\partial \mathbf{l}_n\left(\theta_0\right)}{\partial \theta'}\left(\hat{\theta}_n - \theta_0\right) + \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right)' J\left(\hat{\theta}_n - \theta_0\right).\right\} \tag{A.61}$$

Under H_0, showing

$$\sqrt{n}\begin{pmatrix} \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta} \\ \hat{\theta}_n - \theta_0 \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} -JZ \\ \lambda^\Lambda \end{pmatrix}$$

it can be seen that the asymptotic distribution of $\mathbf{L}_n$ is the law of

$$\mathbf{L} = -\frac{1}{2}Z'J'\overline{K}'J_{11}^{-1}\overline{K}JZ + Z'J'\lambda^\Lambda - \frac{1}{2}\lambda^{\Lambda'}J\lambda^\Lambda.$$

Now, because

$$J'\overline{K}'J_{11}^{-1}\overline{K}J = J - (\kappa_\eta - 1)\Omega \quad \text{with} \quad (\kappa_\eta - 1)\Omega = \begin{pmatrix} 0 & 0 \\ 0 & J_{22} - J_{21}J_{11}^{-1}J_{12} \end{pmatrix}$$

we obtain

$$\begin{aligned} \mathbf{L} &= -\frac{1}{2}Z'JZ + \frac{1}{2}Z'(\kappa_\eta - 1)\Omega Z + Z'J'\lambda^\Lambda - \frac{1}{2}\lambda^{\Lambda'}J\lambda^\Lambda \\ &= -\frac{1}{2}(\lambda^\Lambda - Z)'J(\lambda^\Lambda - Z) + \frac{\kappa_\eta - 1}{2}Z'\Omega Z \end{aligned} \tag{A.62}$$

and, in view of Comment 4 (see Theorem (3.1)) the conclusion easily follows when $\tau = 0$.

Now under $H_n(\tau)$, we obtain (5.4) by the same arguments and by showing that

$$\sqrt{n}\begin{pmatrix} \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta} \\ \hat{\theta}_n - \theta_0 \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} -J(Z+\tau) \\ \lambda^\Lambda(\tau) \end{pmatrix}.$$

A.5. Proof of Proposition 5.1.

For convenience we recall some notations: $K = (0_{d_2 \times d_1}, I_{d_2})$, K_i are matrices obtained by cancelling 0, 1 or several (up to $d_2 - 1$) rows of K, $M_i = K_i'\left(K_iJ^{-1}K_i'\right)^{-1}K_i$, $P_i = I_{d_1+d_2} - J^{-1}M_i$. We have

$$\begin{aligned} \mathbf{W}_n &= \frac{n}{\hat{\kappa}_\eta - 1}(\hat{\theta}_n^{(2)} - \theta_0^{(2)})'\left\{K\hat{J}^{-1}K'\right\}^{-1}(\hat{\theta}_n^{(2)} - \theta_0^{(2)}) \\ &\overset{o_P(1)}{=} \frac{n}{\kappa_\eta - 1}(\hat{\theta}_n - \theta_0)'K'\left\{KJ^{-1}K'\right\}^{-1}K(\hat{\theta}_n - \theta_0) \\ &= \|\sqrt{n}(\hat{\theta}_n - \theta_0)\|_\Omega^2 \\ &\overset{o_P(1)}{=} \|\lambda_{n,\tau}^\Lambda\|_\Omega^2, \end{aligned} \tag{A.63}$$

where the last equality, up to an $o_P(1)$ term, follows from *v)* in the proof of Theorem 4.1. Now, similarly to (3.2), we have

$$\lambda_{n,\tau}^{\Lambda} \overset{o_P(1)}{=} \tilde{Z}_{n,\tau}\mathbb{1}_{\Lambda}(\tilde{Z}_{n,\tau}) + \sum_{i=1}^{2^{d_2}-1} P_i\tilde{Z}_{n,\tau}\mathbb{1}_{\mathcal{D}_i}(\tilde{Z}_{n,\tau}), \tag{A.64}$$

where $\tilde{Z}_{n,\tau} = Z_{n,\tau} + \tau$. This equality holds only up to $o_P(1)$ terms because it uses $\lambda_{n,\tau}^{\Lambda} = \arg\inf_{\lambda\in\Lambda} \|\tilde{Z}_{n,\tau} - \lambda\|_{J_{n,\tau}} \overset{o_P(1)}{=} \arg\inf_{\lambda\in\Lambda} \|\tilde{Z}_{n,\tau} - \lambda\|_J$. In view of (A.63), it follows that

$$\mathbf{W}_n \overset{o_P(1)}{=} \|\tilde{Z}_{n,\tau}\|_{\Omega}^2\mathbb{1}_{\Lambda}(\tilde{Z}_{n,\tau}) + \sum_{i=1}^{2^{d_2}-1} \|P_i\tilde{Z}_{n,\tau}\|_{\Omega}^2\mathbb{1}_{\mathcal{D}_i}(\tilde{Z}_{n,\tau}).$$

Turning to $\mathbf{L}_n$, using (A.60), $\kappa_\eta = 3$ and the step *xi)* already mentionned, we obtain, similarly to (A.62)

$$\begin{aligned}\mathbf{L}_n &\overset{o_P(1)}{=} -\frac{1}{2}Z_n'JZ_n + Z_n'\Omega Z_n + Z_n'J'\lambda_{n,\tau}^{\Lambda} - \frac{1}{2}\lambda_{n,\tau}^{\Lambda'}J\lambda_{n,\tau}^{\Lambda}\\ &= -\frac{1}{2}(\lambda_{n,\tau}^{\Lambda} - Z_n)'J(\lambda_{n,\tau}^{\Lambda} - Z_n) + Z_n'\Omega Z_n\end{aligned}$$

where $Z_n = -J_n^{-1}\sqrt{n}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial\theta}$. A Taylor expansion shows that $Z_n \overset{o_P(1)}{=} Z_{n,\tau} + \tau = \tilde{Z}_{n,\tau}$, from which we deduce

$$\mathbf{L}_n \overset{o_P(1)}{=} -\frac{1}{2}\|\lambda_{n,\tau}^{\Lambda} - \tilde{Z}_{n,\tau}\|_J^2 + \|\tilde{Z}_{n,\tau}\|_{\Omega}^2.$$

By (A.64) we have

$$\frac{1}{2}\|\tilde{Z}_{n,\tau} - \lambda_{n,\tau}^{\Lambda}\|_J^2 = \frac{1}{2}\sum_{i=1}^{2^{d_2}-1} \|(I_d - P_i)\tilde{Z}_{n,\tau}\|_J^2\mathbb{1}_{\mathcal{D}_i}(\tilde{Z}_{n,\tau}) = \sum_{i=1}^{2^{d_2}-1} \|\tilde{Z}_{n,\tau}\|_{\Omega_i}^2\mathbb{1}_{\mathcal{D}_i}(\tilde{Z}_{n,\tau}),$$

where $\Omega_i = (\kappa_\eta - 1)^{-1}(I_d - P_i)'J(I - P_i) = K_i'\left((\kappa_\eta - 1)K_iJ^{-1}K_i'\right)^{-1}K_i$. Moreover

$$\|\tilde{Z}_{n,\tau}\|_{\Omega}^2 = \|\tilde{Z}_{n,\tau}\|_{\Omega}^2\mathbb{1}_{\Lambda}(\tilde{Z}_{n,\tau}) + \sum_{i=1}^{2^{d_2}-1} \|\tilde{Z}_{n,\tau}\|_{\Omega}^2\mathbb{1}_{\mathcal{D}_i}(\tilde{Z}_{n,\tau}).$$

It follows that

$$\begin{aligned}\mathbf{L}_n - \mathbf{W}_n &\overset{o_P(1)}{=} \sum_{i=1}^{2^{d_2}-1}\left(\|\tilde{Z}_{n,\tau}\|_{\Omega}^2 - \|\tilde{Z}_{n,\tau}\|_{\Omega_i}^2 - \|P_i\tilde{Z}_{n,\tau}\|_{\Omega}^2\right)\mathbb{1}_{\mathcal{D}_i}(\tilde{Z}_{n,\tau})\\ &= \sum_{i=1}^{2^{d_2}-1} \|\tilde{Z}_{n,\tau}\|_{\Omega-\Omega_i-P_i'\Omega P_i}^2\mathbb{1}_{\mathcal{D}_i}(Z) = 0\end{aligned}$$

because $A - A_i - P_i'A = 0$. This equality is obtained by noting that K_i is of the form $K_i = B_iK$ for some matrix B_i (recall that K_i is deduced from K by cancellation of rows). Hence $P_i'\Omega P_i = P_i'(\Omega - M_i) = P_i'\Omega$ and

$$(I-P_i)'\Omega = K_i'\left(K_iJ^{-1}K_i'\right)^{-1}K_iJ^{-1}K'\left(KJ^{-1}K'\right)^{-1}K = K_i'\left(K_iJ^{-1}K_i'\right)^{-1}B_iK = \Omega_i.$$

A.6. Proof of Proposition 5.2.

Clearly, the constrained estimator $\hat{\theta}_{n|2}$ does not converges to θ_1 under H_1. Its almost sure limit is obtained along the same lines as for the unconstrained estimator. For the sake of brevity we will only show that the limit criterion $\ell_\infty(\theta) := E_{\theta_1}\{\ell_t(\theta)\}$ is uniquely minimized at the value $\theta_{1|2}$ under the constraint $\theta^{(2)} = 0$. A Taylor expansion about θ_1 gives $\ell_\infty(\theta) = \ell_\infty(\theta_1) + (\theta-\theta_1)'J_1(\theta-\theta_1) + o(\|\theta-\theta_1\|^2)$, because the first derivative of $\ell_\infty(\theta)$ cancels at the value θ_1. Thus the optimization problem reduces to minimizing $\|\theta-\theta_1\|^2_{J_1}$ under the constraint $K\theta = 0$. The solution is $(I_d - J_1^{-1}K'\left(KJ_1^{-1}K'\right)^{-1}K)\theta_1 = \theta_{1|2}$. It follows from the a.s. convergence of $\hat{\theta}_{n|2}$ to $\theta_{1|2}$ that the matrix $\hat{J} = J(\hat{\theta}_{n|2})$ is a consistent estimator of $J_{1|2}$.

We now adapt several intermediate results given in the proof of Theorem 5.1. It is easy to check that (A.54) still holds under H_1 when θ_0 is replaced by θ_1. Thus we have

$$\sqrt{n}\frac{\partial\tilde{\mathbf{l}}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta^{(2)}} + KJ_1\sqrt{n}\left(\hat{\theta}_{n|2}-\theta_1\right). \tag{A.65}$$

The analogue of (A.57) is

$$\left(\hat{\theta}_{n|2}-\theta_1\right) = \overline{K}'\left(\hat{\theta}^{(1)}_{n|2}-\theta^{(1)}_1\right) - K'\theta^{(2)}_1. \tag{A.66}$$

Therefore (A.56) becomes

$$0 \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta^{(1)}} + \overline{K}J_1\overline{K}'\sqrt{n}\left(\hat{\theta}^{(1)}_{n|2}-\theta^{(1)}_1\right) - \sqrt{n}\,\overline{K}J_1K'\theta^{(2)}_1,$$

which gives

$$\sqrt{n}\left(\hat{\theta}^{(1)}_{n|2}-\theta^{(1)}_1\right) \overset{o_P(1)}{=} -\left(\overline{K}J_1\overline{K}'\right)^{-1}\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta^{(1)}} + \sqrt{n}\left(\overline{K}J_1\overline{K}'\right)^{-1}\overline{K}J_1K'\theta^{(2)}_1. \tag{A.67}$$

Using (A.53), (A.65), (A.66) and (A.67) we obtain

$$\begin{aligned}
\mathbf{R}_n &\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial\mathbf{l}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)}}\right\|^2_{KJ_{1|2}^{-1}K'} \\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta^{(2)}} + KJ_1\overline{K}'\left(\hat{\theta}^{(1)}_{n|2}-\theta^{(1)}_1\right) - KJ_1K'\theta^{(2)}_1\right\|^2_{KJ_{1|2}^{-1}K'} \\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta^{(2)}} - KJ_1\overline{K}'\left(\overline{K}J_1\overline{K}'\right)^{-1}\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta^{(1)}}\right. \\
&\qquad\left. + KJ_1\overline{K}'\left(\overline{K}J_1\overline{K}'\right)^{-1}\overline{K}J_1K'\theta^{(2)}_1 - KJ_1K'\theta^{(2)}_1\right\|^2_{KJ_{1|2}^{-1}K'} \\
&\overset{o_P(n)}{=} \frac{n}{\kappa_\eta-1}\left\|KJ_1\overline{K}'\left(\overline{K}J_1\overline{K}'\right)^{-1}\overline{K}J_1K'\theta^{(2)}_1 - KJ_1K'\theta^{(2)}_1\right\|^2_{KJ_{1|2}^{-1}K'}.
\end{aligned}$$

The last equality follows from the fact that $\frac{\partial\mathbf{l}_n(\theta_1)}{\partial\theta} = O_P(1/\sqrt{n})$. Using $KJ_1\overline{K}' = J_{21}$, $\overline{K}J_1\overline{K}' = J_{11}$, $KJ_1K' = J_{22}$, and $J_{22} - J_{21}J_{11}^{-1}J_{12} = \left(KJ_1^{-1}K'\right)^{-1}$, we have under H_1

$$\lim_{n\to\infty}\frac{\mathbf{R}_n}{n} = \frac{1}{\kappa_\eta-1}\theta^{(2)'}_1\left(KJ_1^{-1}K'\right)^{-1}KJ_{1|2}^{-1}K'\left(KJ_1^{-1}K'\right)^{-1}\theta^{(2)}_1.$$

To show (5.6) it suffices to note that

$$\log S_{\mathbf{R}}(\mathbf{R}_n) = \log P(\chi^2_{d_2} > \mathbf{R}_n) \sim -\mathbf{R}_n/2$$

because $\mathbf{R}_n \to \infty$ and $\log P(\chi^2_{d_2} > x) \sim -x/2$ as $x \to \infty$.

The behaviour of the Wald statistic under H_1 is clearly given by

$$\lim_{n\to\infty} \frac{\mathbf{W}_n}{n} = \frac{1}{\kappa_\eta - 1} \theta_1^{(2)'} \left(K J_1^{-1} K' \right)^{-1} \theta_1^{(2)},$$

and (5.5) is obtained by showing that $\log S_{\mathbf{W}}(\mathbf{W}_n) \sim -\mathbf{W}_n/2$.

A.7. Proof of Corollary 5.1.

Note that (5.7) is the power of the test of critical region $\{X > c_1\}$ for testing the null hypothesis $H_0 : EX = 0$ versus the alternative $H_1 : EX = \tau^* > 0$, when the unique observation X follows a gaussian distribution with unknown mean EX and variance 1. The power (5.8) is that of the two-sided test $\{|X| > c_2\}$. The two tests $\{X > c_1\}$ and $\{|X| > c_2\}$ have the same level, but it is well-known that the first test is uniformly most powerful under one-sided alternatives of the form H_1.

A.8. Proof of Corollary 5.2.

In view of (5.7) and (5.9), the Wald test is asymptotically optimal if and only if $(\kappa_\eta - 1)KJ^{-1}K' = KI_f^{-1}K'$, which is equivalent to $(\kappa_\eta - 1) = 4/\iota_f$. We have

$$\begin{aligned} \int (y^2-1)\left(1 + \frac{f'(y)}{f(y)} y\right) f(y)dy &= E\eta_t^2 - 1 + \int y^3 f'(y)dy - \int y f'(y)dy \\ &= \lim_{a,b\to\infty} [y^3 f(y)]_a^{-b} - \int 3y^2 f(y)dy + 1 = -2. \end{aligned}$$

Thus, the Cauchy-Schwarz inequality yields

$$4 \quad \le \quad \int (y^2-1)^2 f(y)dy \int \left(1 + \frac{f'(y)}{f(y)} y\right)^2 f(y)dy = (E\eta_t^4 - 1)\iota_f$$

with equality iff there exists $a \neq 0$ such that $1 + \eta_t f'(\eta_t)/f(\eta_t) = -2a\left(\eta_t^2 - 1\right)$ a.s. The latter equality holds iff $f'(y)/f(y) = -2ay + (2a-1)/y$ almost everywhere. The solution of this differential equation, under the constraint $f \geq 0$ and $\int f(y)dy = 1$, is given by (5.10). Note that when f is defined by (5.10), we have $\kappa_\eta = \int y^4 f(y)dy = a(a+1)/a^2 = 3$ iff $a = 1/2$ which corresponds to the case $\eta_t \sim \mathcal{N}(0,1)$.

References

Akharif, A. and M. Hallin (2003) Efficient detection of random coefficients in autoregressive models. *The Annals of Statistics* 31, 675–704.

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

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

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

Bahadur, R. R. (1967) Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics* 38, 303–324.

Batholomew, D. J. (1959) A test of homogeneity of ordered alternatives, I. *Biometrika* 46, 36–48.

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

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

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

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

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

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

Boussama, F. (1998) *Ergodicité, mélange et estimation dans les modèles GARCH.* Thèse de l'Université Paris-7.

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

Demos, A. and E. Sentana (1998) Testing for GARCH effects: A one-sided approach. *Journal of Econometrics* 86, 97–127.

Drost, F. C. and C. A. J. Klaassen (1997) Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193–221.

Drost, F. C., Klaassen, C. A. J. and B. J. M. Werker (1997) Adaptive estimation in time-series models. *Annals of Statistics* 25, 786–817.

Dufour, J.-M., Khalaf, L., Bernard, J.-T. and Genest, I. (2004) Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. *Journal of Econometrics* 122, 317–347.

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

Francq, C., and J. M. Zakoïan (2004) Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli* 10, 605–637.

Giraitis L., Leipus R. and D. Surgailis (2004) Recent advances in ARCH modelling. In *Long-Memory in Economics. Eds. A. Kirman and G. Teyssiere.* Springer, Berlin, forthcoming.

Gouriéroux, C., Holly A., and A. Monfort (1982) Likelihood Ratio tests, Wald tests, and Kuhn-Ticker Test in Linear Models with inequality constraints on the regression parameters. *Econometrica* 50, 63–80.

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

Hong, Y. (1997) One-sided ARCH testing in time series models. *Journal of Time Series Analysis* 18, 253–277.

Hong, Y. and J. Lee (2001) One-sided testing for ARCH effects using wavelets. *Econometric Theory* 17, 1051–1081.

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

King, M. L. and P. X. Wu (1997) Locally optimal one-sided tests for multiparameter hypotheses. *Econometric Reviews* 16, 131–156.

Klüppelberg, C., Maller R. A., Van de Vyver, M. and D. Wee (2002) Testing for reduction to random walk in autoregressive conditional heteroskedasticity models. *The Econometrics Journal* 5, 387–416.

Kudô, A. (1963) A multivariate analogue of the one-sided test. *Biometrika* 50, 403–418.

Lee, J. H. H. (1991) A Lagrange multiplier test for GARCH models. *Economics Letters* 37, 265–271.

Lee, J. H. H. and M. L. King (1993) A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics* 11, 17–27.

Ling, S. (2005) Self-weighted LSE and MLE for ARMA-GARCH models. *Unpublished document* University of Science and Technology, Hong Kong.

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

May, A. and A. Szimayer (2001) Testing for conditional heteroskedasticity in financial time-series. *Unpublished Document,* Darmstadt University of Technology and University of Bonn.

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

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

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

Rogers, A. J. (1986) Modified Lagrange multiplyer tests for problems with one-sided alternatives. *Journal of Econometrics* 31, 341–361.

Shi, N. Z. (1987) Testing a normal mean vector against the alternative determined by a convex cone. *Memoirs of the Faculty of Science, Kyushu University, Ser. A* 41, 133–145.

Shi, N. Z. and A. Kudô (1987) The most stringent somewhere most powerful one sided test of the multivariate normal mean. *Memoirs of the Faculty of Science, Kyushu University, Ser. A* 41, 37–44.

Silvapulle, M. J. and P. Silvapulle (1995) A score test against one-sided alternatives. *Journal of the American Statistical Association* 90, 342–349.

Straumann, D. (2005) *Estimation in conditionally heteroscedastic time series models.* Lecture Notes in Statistics, Springer Berlin Heidelberg. *Technical Report* University of Copenhagen.

van der Vaart, A. W. (1998) *Asymptotic statistics.* Cambridge University Press, Cambridge.

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

Wolak, F. A. (1989) Local and global testing of linear and non linear inequality constraints in non linear econometric models. *Econometric Theory* 5, 1–35.

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