INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2008-04

Testing the Nullity of GARCH Coefficients : Correction of the Standard Tests and Relative Efficients Comparisons

C. FRANCQ[1]
J.-M. ZAKOÏAN[2]

[1] Université Lille III, GREMARS-EQUIPPE, BP 60149, 59653 Villeneuve d'Ascq Cédex, France.
Tél. : 33 (0) 3 20 41 64 87 (francq@univ-lille3.fr)
[2] CREST and GREMARS-EQUIPPE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cédex, France.
Tél. : 33 (0) 1 41 17 78 25 (zakoian@ensae.fr).

Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons

Christian Francq* and Jean-Michel Zakoïan†

Abstract

This article is concerned by testing the nullity of coefficients in GARCH models. The problem is non standard because the quasi-maximum likelihood estimator is subject to positivity constraints. The paper establishes the asymptotic null and local alternative distributions of Wald, score, and quasi-likelihood ratio tests. Efficiency comparisons under fixed alternatives are also considered. Two cases of special interest are: (i) tests of the null hypothesis of one coefficient equal to zero and (ii) tests of the null hypothesis of no conditional heteroscedasticity. The results are illustrated by means of simulation experiments. An empirical application to the Standard & Poor 500 and the CAC40 indexes is proposed.

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

Résumé

Cet article concerne les tests de nullité de coefficients dans les modèles GARCH models. Le problème est non standard car l'estimateur du quasi-maximum de vraisemblance est obtenu sous contraintes de positivité sur les coefficients. L'article établit les propriétés asymptotiques, sous l'hypothèse nulle et sous des alternatives locales, des tests de Wald, du score et du rapport de quasi-vraisemblance. Des comparaisons d'efficacité sous alternatives fixes sont également effectuées. Deux cas particuliers importants sont : (i) les tests de nullité d'un seul coefficient et (ii) les tests de l'hypothèse d'absence d'hétéroscédasticité conditionnelle. Les résultats sont illustrés par des expériences de simulations. Une application empirique aux indices Standard & Poor 500 et CAC40 est proposée.

Keywords : Efficacité asymptotique de tests, frontière, loi Chi-bar, modèle GARCH, estimation par quasi maximum de vraisemblance, alternatives locales.

*Université Lille III, GREMARS-EQUIPPE, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-mail: francq@univ-lille3.fr, tel: 33.3.20.41.64.87

†CREST and GREMARS-EQUIPPE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr, tel: 33.1.41.17.78.25

1 Introduction

Despite the development of stochastic volatility models, the class of generalized autoregressive conditionally heteroscedastic (GARCH) models introduced by Engle (1982) and generalized by Bollerslev (1986) remains very popular in finance. This is testified by the body of work using this class for financial applications such as Value At Risk, Option Pricing, and portfolio analysis. Contrary to a common opinion, a GARCH model is not a simple structure and before proceeding to its estimation, it is sensible to make sure that such a sophisticated model is justified. When a GARCH effect is present in the data, it is of interest to test if the orders of the fitted models can be reduced, by testing the nullity of the higher-lag ARCH or GARCH coefficient. In practice, testing the nullity of parameters in the GARCH framework is achieved by applying standard tests, such as the Wald test, the Rao-score (or Lagrange Multiplier) test and the Likelihood Ratio test. These standard tests are provided by most standard time series packages currently available for GARCH estimation (e.g. GAUSS, RATS, SAS, SPSS).

Unfortunately, as we will see, this common practice may be based on an invalid asymptotic theory. Tests in GARCH models have received much less attention than the theory of estimation. Despite its apparent simplicity, the problem of testing that some coefficients are equal to zero in a GARCH model is non trivial. The reason is that the coefficients and the QMLE are positively constrained. It follows that the standard distributions for some widely used tests are not asymptotically valid.

The primary objective of this paper is to derive asymptotically valid critical values for the Wald, Student, Rao-score and Quasi-likelihood ratio statistics. The tests presented here rest on the asymptotic properties of the Quasi Maximum-Likelihood Estimators (QMLE) when some GARCH coefficients are on the boundary of the parameter space. Our second goal is to compare the efficiencies of those tests under fixed and local alternatives. We will use the approximate Bahadur slope criterion and the Pitman analysis for power comparisons. Investigation of the asymptotic local powers requires an extension of the asymptotic properties of the QMLE to the case of local alternatives to a parameter at the boundary.

The most important cases for applications are: (i) tests of the null hypothesis of one coefficient equal to zero and (ii) tests of the null hypothesis of no conditional heteroscedasticity. In these two special cases, detailed asymptotic efficiency (local and non local) comparisons can be done. 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 procedures, namely the Wald, Rao-score and the Quasi-Likelihood Ratio (QLR) tests. For conditional homoscedasticity testing, we will also compare these three tests with the Lee and King (1993) test, which exploits the one-sided nature of the alternatives and enjoys optimality properties.

There exists a large amount of literature dealing with testing problems in which, under the null hypothesis, the parameter is at the boundary of the maintained assumption. Such problems 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. Tests of ARCH(1)-type effects in autoregressive processes, possibly with unit root, have been considered by Klüppelberg, Maller, van de Vyver and Wee (2002).

The article is organized as follows. Section 2 presents the estimation results, in particular when the true parameter value is on the boundary, and the main test statistics. Section 3 determines their asymptotic null distributions. Section 4 establishes the asymptotic distribution of the QMLE under sequences of local alternatives to the null parameter value. Section 5 uses these results to compare the local powers of the tests. Efficiency comparisons in the sense of Bahadur are also considered. Sections 6 and 7 apply these results to the two main examples: testing the nullity of one coefficient and testing the absence of ARCH effect. Some Monte Carlo results are reported. Section 8 is devoted to an application to the Standard & Poor (S&P) 500 index and to the French CAC40 index. Section 9 concludes. Proofs are relegated to an appendix.

If a matrix A is semi-positive definite, a semi-norm of a vector x of appropriate dimension is defined by $\|x\|_A = (x'Ax)^{1/2}$. The notation $a \overset{c}{=} b$ will stand for $a = b + c$. For a vector x, inequalities such as $x > 0$ or $x \geq 0$ have to be understood componentwise.

2 Model and test statistics

Assume that the observed time series $\epsilon_1, \dots, \epsilon_n$ is generated by 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 $\theta_0 := (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})$ is a parameter vector and the noise sequence (η_t) is iid with mean 0 and variance 1. Under the positivity constraints

$$\omega_0 > 0, \quad \alpha_{0i} \geq 0 \ \ (i = 1, \dots, q), \quad \beta_{0j} \geq 0 \ \ (j = 1, \dots, p),$$

Bougerol and Picard (1992) showed that a unique *nonanticipative strictly stationary* solution (ϵ_t) exists if and only if $\gamma(\mathbf{A_0}) < 0$ where, for any norm $\|\cdot\|$ on the space of the $(p+q) \times (p+q)$ matrices, $\gamma(\mathbf{A_0}) = \lim_{t\to\infty} \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\|$ a.s. and

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

with $\alpha_{01:q-1} = (\alpha_{01} \dots \alpha_{0q-1})$, $\beta_{01:q-1} = (\beta_{01} \dots \beta_{0p-1})$ and I_k being the $k \times k$ identity matrix. A nonanticipative solution (ϵ_t) of Model (2.1) is such that ϵ_t is a measurable function of the $\eta_{t-i}, i \geq 0$. Note that Nelson and Cao (1992) derived necessary and sufficient conditions for the positivity of the volatility process σ_t^2. However these conditions are not very explicit and thus seem difficult to use for statistical purposes.

The primary objective of this article is to develop a methodology for testing the nullity of a sub-vector of θ_0. More precisely, and without loss of generality we consider testing the nullity of the last d_2 coefficients of θ_0, split into two components as $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)})'$, where $\theta_0^{(i)} \in \mathbb{R}^{d_i}$, $d_1 + d_2 = p + q + 1 = d$. The null hypothesis is thus

$$H_0 : \quad \theta_0^{(2)} = 0_{d_2\times 1} \qquad \text{i.e. } K\theta_0 = 0_{d_2\times 1} \text{ with } \quad K = \begin{pmatrix} 0_{d_2\times d_1}, \ I_{d_2} \end{pmatrix}$$

and let

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

denote our maintained assumption. To proceed, we define the vector of parameters as $\theta = (\theta_1, \dots, \theta_{p+q+1})'$, with $\theta_1 = \omega$, and the parameter space Θ as any compact subset of $[0,\infty)^{p+q+1}$ that bounds the first component away from zero. For technical reasons we also assume that Θ contains some hypercube of the form $[\underline{\omega}, \overline{\omega}] \times [0,\varepsilon]^{p+q}$, for some $\varepsilon > 0$ and $\overline{\omega} > \underline{\omega} > 0$.

To define the QMLE, the initial values are, for simplicity, taken equal to zero, i.e. $\epsilon_0^2 = \dots = \epsilon_{1-q}^2 = \tilde{\sigma}_0^2 = \dots = \tilde{\sigma}_{1-p}^2 = 0$, and the variables $\tilde{\sigma}_t^2(\theta)$ are recursively defined, for $t \geq 1$, by

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

A QMLE of θ is defined as any measurable solution $\hat{\theta}_n$ of $\hat{\theta}_n = \arg\min_{\theta\in\Theta} \tilde{\mathbf{l}}_n(\theta)$, where $\tilde{\mathbf{l}}_n(\theta) = n^{-1}\sum_{t=1}^n \tilde{\ell}_t$, and $\tilde{\ell}_t = \tilde{\ell}_t(\theta) = \tilde{\ell}_t(\theta; \epsilon_n, \dots, \epsilon_1) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log \tilde{\sigma}_t^2$. 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 $\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$, for all t. Note that $\sigma_t^2(\theta_0) = h_t$. Let $\mathbf{l}_n(\theta) = n^{-1}\sum_{t=1}^n \ell_t$, and $\ell_t = \ell_t(\theta) = \frac{\epsilon_t^2}{\sigma_t^2} + \log \sigma_t^2$. Under appropriate conditions (see e.g. Francq and Zakoian, 2004), the information matrix $J = E_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial\theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial\theta'}\right)$ is well-defined and the QMLE is asymptotically normal:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\kappa_\eta - 1)J^{-1}), \qquad \kappa_\eta = E\eta_t^4. \tag{2.2}$$

The usual forms of the Wald, Rao-score and QLR statistics follow, and are given by

$$\begin{aligned}
\mathbf{W}_n &= \frac{n}{\hat{\kappa}_\eta - 1} \hat{\theta}_n^{(2)'} \left\{ K \hat{J}_n^{-1} K' \right\}^{-1} \hat{\theta}_n^{(2)}, \\
\mathbf{R}_n &= \frac{n}{\hat{\kappa}_{\eta|2} - 1} \frac{\partial \tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right)}{\partial\theta'} \hat{J}_{n|2}^{-1} \frac{\partial \tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right)}{\partial\theta}, \\
\mathbf{L}_n &= n\left\{ \tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right) - \tilde{\mathbf{l}}_n\left(\hat{\theta}_n\right) \right\},
\end{aligned}$$

where $\hat{\theta}_{n|2}$ denotes the *restricted* (by H_0) estimator of θ_0, $\hat{\kappa}_\eta$, $\hat{\kappa}_{\eta|2}$ denote consistent estimators of κ_η, and $\hat{J}_n, \hat{J}_{n|2}$ denote consistent estimators of the information matrix J. In general, $\hat{J}_n$ and $\hat{\kappa}_\eta$ are derived using the unconstrained estimator $\hat{\theta}_n$, whereas $\hat{J}_{n|2}$ and $\hat{\kappa}_{\eta|2}$ are computed using $\hat{\theta}_{n|2}$. For instance, one can take

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

and

$$\hat{\kappa}_\eta = \frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_t^4}{\tilde{\sigma}_t^4(\hat{\theta}_n)}, \quad \hat{\kappa}_{\eta|2} = \frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_t^4}{\tilde{\sigma}_t^4(\hat{\theta}_{n|2})},$$

because

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\hat{\theta}_n)} = \frac{1}{n}\sum_{t=1}^{n} \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\hat{\theta}_{n|2})} = 1, \quad a.s. \tag{2.3}$$

Note that the latter equalities imply that

$$\mathbf{L}_n = \frac{1}{n}\sum_{t=1}^{n} \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_{n|2})}{\tilde{\sigma}_t^2(\hat{\theta}_n)}, \quad a.s.$$

One rejects the null hypothesis for large values of $\mathbf{W}_n, \mathbf{R}_n, \mathbf{L}_n$. In the next section, we give the asymptotic distributions of these statistics under the null hypothesis.

3 Non standard asymptotic null distributions

Among the regularity assumptions required for (2.2) to hold, a crucial one is that $\theta_0 > 0$, componentwise. Indeed if, say, $\theta_{0i} = 0$, the variable $\sqrt{n}(\hat{\theta}_{ni} - \theta_{0i}) = \sqrt{n}\hat{\theta}_{ni}$ is nonnegative and thus cannot be asymptotically normal. Note that this problem cannot be solved by blowing up the parameter space Θ outside the positive quadrant, since the variable $\tilde{\sigma}_t^2(\theta)$ must be positive for the loglikelihood to be well-defined.

We now give the precise assumptions required to obtain the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ under H_0. 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 $p = 0$. Let $\theta_0(\varepsilon)$ be the vector obtained by replacing all zero coefficients of θ_0 by a number ε.

A1: $\theta_0(\varepsilon) \in \overset{\circ}{\Theta}$ for some $\varepsilon > 0$, 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 $E\eta_t^2 = 1$ and $\kappa_\eta = E\eta_t^4 < \infty$.

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 $\alpha_{0q} + \beta_{0p} \neq 0$.

Assumption **A1** is intended to prevent θ_0 from reaching the upper bound of Θ. In **A2**, the strict stationarity condition is imposed only at the value θ_0. For all other parameter values, it is sufficient to make the given assumption on the β_i coefficients. Assumptions **A3** and **A4** are made for identifiability reasons. In some cases, no moment assumption on the observed process (ϵ_t) will be required. In other cases a moment condition is necessary. We therefore introduce two assumptions which will be made alternately.

A5: $E_{\theta_0}\epsilon_t^6 < \infty$.

A6: $\{j \mid \beta_{0,j} > 0\} \neq \varnothing$ and $\prod_{i=1}^{j_0} \alpha_{0i} > 0$ for $j_0 = \min\{j \mid \beta_{0,j} > 0\}$.

Note that **A6** does not cover the ARCH case, where all the β_{0i} coefficients are equal to zero. Let $\Lambda = \mathbb{R}^{d_1} \times [0, \infty)^{d_2}$. The proof of the following result is given in Francq and Zakoian (2007) (hereinafter FZ). We also display the asymptotic distribution of the score vector.

Theorem 3.1 *If H_0,* **A1**–**A4** *and either* **A5** *or* **A6** *hold,*

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_0) &\xrightarrow{d} \lambda^{\Lambda} := \arg\inf_{\lambda \in \Lambda} \{\lambda - Z\}' J \{\lambda - Z\}, \qquad Z \sim \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right), \\ \sqrt{n}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta} &\xrightarrow{d} \mathcal{N}\{0, (\kappa_\eta - 1)J\},\end{aligned}$$

where in the definition of J, derivatives with respect to the last d_2 components are replaced by right derivatives.

The asymptotic distribution of the QMLE is thus non standard when the true parameter has coefficients equal to zero, but it can be easily simulated. First note that λ^Λ can be interpreted as the projection of Z, for the metric defined by J, onto the convex set $\Lambda = \{\lambda \in \mathbb{R}^d \mid K\lambda \geq 0\}$. The faces of Λ are sections of the subspaces $\{\lambda \in \mathbb{R}^d \mid K_i\lambda = 0\}$, where the K_i are obtained by cancelling 0, 1 or several rows of K. Projecting Z onto those subspaces yields the vectors $\lambda_{K_i} = P_i Z$, where $P_i = I_d - J^{-1}K_i'\left(K_i J^{-1} K_i'\right)^{-1} K_i$. The solution is thus obtained as

$$\lambda^\Lambda = Z\mathbb{1}_\Lambda(Z) + \mathbb{1}_{\Lambda^c}(Z) \times \mathrm{argmin}_{\lambda \in \mathcal{C}} \|\lambda - Z\|_J = Z\mathbb{1}_\Lambda(Z) + \sum_{i=1}^{2^{d_2}-1} P_i Z \mathbb{1}_{\mathcal{D}_i}(Z), \quad (3.1)$$

where $\mathcal{C} = \{\lambda_{K_i} : i = 1, \ldots, 2^{d_2} - 1 \text{ and } K\lambda_{K_i} \geq 0\}$ is the set of admissible projections (those with nonnegative last d_2 components) and the $\mathcal{D}_i$ form a partition of $\mathbb{R}^d$. For instance, when all the coefficients α_{0i} are equal to zero in an ARCH(q) model ($d_1 = 1, d_2 = q, d = q + 1$), it can be seen that (3.1) reduces to

$$\lambda^\Lambda = \left(Z_1 + \omega \sum_{i=2}^{d} Z_i^-, Z_2^+, \cdots, Z_d^+\right)'. \quad (3.2)$$

We are now in position to derive the asymptotic distributions of the three test statistics introduced in Section 2. 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 3.2 *Under H_0 and the assumptions of Theorem 3.1 we have*

$$\mathbf{W}_n \xrightarrow{d} \mathbf{W} = \lambda^{\Lambda'}\Omega\lambda^{\Lambda}, \tag{3.3}$$

$$\mathbf{R}_n \xrightarrow{d} \chi^2_{d_2}, \tag{3.4}$$

$$\begin{aligned} \mathbf{L}_n \xrightarrow{d} \mathbf{L} &= -\frac{1}{2}(\lambda^\Lambda - Z)'J(\lambda^\Lambda - Z) + \frac{1}{2}Z'K'\left\{KJ^{-1}K'\right\}^{-1}KZ \\ &= -\frac{1}{2}\left\{\inf_{K\lambda\geq 0}\|Z-\lambda\|_J^2 - \inf_{K\lambda=0}\|Z-\lambda\|_J^2\right\}. \end{aligned} \tag{3.5}$$

An interesting point is that, contrary to the standard situation, the asymptotic distributions of those statistics are not the same. Only the score statistic has the standard $\chi^2_{d_2}$ distribution, which is a consequence of the gaussian asymptotic distribution of the score vector under H_0. This implies that the standard Rao score test remains valid whatever the position of θ_0, in the interior or on the boundary of Θ. On the contrary, valid tests based on the Wald and LR statistics require correction of the usual critical values. 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).

By Theorem 3.2, tests of asymptotic level α are defined by the critical regions

$$\{\mathbf{W}_n > \mathbf{w}_{1-\alpha}\}, \quad \{\mathbf{R}_n > \chi^2_{d_2,1-\alpha}\}, \quad \{\mathbf{L}_n > \mathbf{l}_{1-\alpha}\}$$

where $\mathbf{w}_{1-\alpha}$, $\chi^2_{d_2,1-\alpha}$ and $\mathbf{l}_{1-\alpha}$ are the $(1-\alpha)$-quantiles of the distributions of $\mathbf{W}, \chi^2_{d_2}, \mathbf{L}$ respectively. In the sequel the first test is referred to as the *modified* Wald test. The standard Wald test is defined by $\{\mathbf{W}_n > \chi^2_{d_2,1-\alpha}\}$ and its asymptotic level is not equal to α. Similar remarks apply to the QLR test.

4 Non regularity of the QMLE under local alternatives

For local power comparisons, the asymptotic distribution of the QMLE under sequences of local alternatives to the null parameter value θ_0 is required. Let $\theta_n = \theta_0 + \tau/\sqrt{n}$, where $\tau = (\tau_0, \dots, \tau_{p+q})' \in (0, +\infty)^{p+q+1}$ is such that $\theta_n \in \Theta$, at least for sufficiently large n.

We need to precisely define the data generating process.Write $\mathbf{A_0} = \mathbf{A}(\theta_0)$ and assume that **A2** holds. For n large enough, $\gamma\left\{\mathbf{A}(\theta_0 + \tau/\sqrt{n})\right\} < 0$ and we can define the nonanticipative and strictly stationary solution $(\epsilon_{t,n})_{t\in\mathbb{Z}}$ 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 $\epsilon_{1,n}, \dots, \epsilon_{n,n}$, the QMLE satisfies

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

where $\tilde{\sigma}_{t,n} = \tilde{\sigma}_{t,n}(\theta)$ is obtained by replacing ϵ_u by $\epsilon_{u,n}$, $1 \le u < t$, 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 QMLE 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 $\sqrt{n}(\hat{\theta}_n - \theta_n)$ is asymptotically distributed under $\mathbb{P}_{n,\tau}$ as $\lambda^\Lambda(\tau) - \tau$ where*

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

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 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 (2007). However the previous theorem can be established directly. For brevity we do not provide the proof but it is available from the authors.

When the true value θ_0 is not on the boundary, i.e. when H_0 does not hold, $\lambda^\Lambda(\tau) - \tau = Z$ is independent of τ. However, it is seen that under H_0, the QMLE *does not* converge to its asymptotic distribution locally uniformly since $\lambda^\Lambda(\tau) - \tau$ generally depends on τ. Thus, the QMLE is *regular* in the interior of Θ but not on the whole parameter space (see e.g. van der Vaart p 115, 1998).

5 Power comparisons

In this section, we consider two popular efficiency measures, in order to compare the asymptotic power functions of the tests. We start by Bahadur's (1960) approach in which the efficiency of a test is measured by the rate of convergence of its p-value under a fixed alternative $H_1 : \theta_0^{(2)} > 0$.

5.1 Bahadur slopes

Let

$$J(\theta) = E_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\right), \quad D(\theta) = E_{\theta_0}\left[\frac{1}{\sigma_t^2(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta^{(2)}}\left(1-\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right)\right].$$

Let $S_{\mathbf{W}}(t) = \mathbb{P}(\mathbf{W} > t)$, $S_{\mathbf{R}}(t) = \mathbb{P}(\mathbf{R} > t)$ where $\mathbf{R} \sim \chi^2_{d_2}$, and $S_{\mathbf{L}}(t) = \mathbb{P}(\mathbf{L} > t)$, be the asymptotic survival functions of the Wald, score and QLR statistics under the null hypothesis H_0.

Proposition 5.1 *Under the alternative $H_1 : \theta_0^{(2)} > 0$ and under* **A1-A4**, *the approximate Bahadur slope of the Wald test is*

$$\lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{W}}(\mathbf{W}_n) = \frac{1}{\kappa_\eta - 1}\theta_0^{(2)\prime}\left(KJ^{-1}K'\right)^{-1}\theta_0^{(2)}, \quad a.s. \tag{5.1}$$

Moreover, under regularity conditions discussed in the appendix, the approximate Bahadur slopes of the score and QLR tests are

$$\lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{R}}(\mathbf{R}_n) = \frac{1}{\kappa_{\eta|2} - 1}D'(\theta_{0|2})KJ_{0|2}^{-1}K'D(\theta_{0|2}), \tag{5.2}$$

$$\lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{L}}(\mathbf{L}_n) = E_{\theta_0}\left(\log\frac{\sigma_t^2(\theta_{0|2})}{\sigma_t^2(\theta_0)}\right), \tag{5.3}$$

where $\theta_{0|2}$ is the a.s. limit of $\hat{\theta}_{n|2}$, $J_{0|2} = J(\theta_{0|2})$ and $\kappa_{\eta|2}$ is the kurtosis coefficient of $\sigma_t^{-1}(\theta_{0|2})\epsilon_t$ under H_1. It follows that the Wald, score and QLR tests are consistent against H_1.

The term "approximate" Bahadur slopes serves to distinguish the limits in (5.1) and (5.2) from other quantities, called "exact" Bahadur slopes, which are defined by substituting the non-asymptotic survival functions for the asymptotic ones (e.g. $P(\mathbf{W}_n > t)$ for $S_{\mathbf{W}}(t)$) 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 and $J_{0|2}$ are not known in closed form. Numerical comparisons can be done however as will be seen later.

5.2 Pitman analysis

Whereas Bahadur's approach considers non-local alternatives and compares the rates at which the P-values of two tests converge to zero, the Pitman approach considers sequences of local alternatives, and compares the local asymptotic powers of the tests. We denote by $\chi^2_k(c)$ the noncentral chi-square distribution with noncentrality parameter c and k degrees of freedom. The asymptotic distributions of the 3 test statistics under the local alternatives are given in the following theorem.

Theorem 5.1 *Under the assumptions of Theorem 4.1, we have*

$$\mathbf{W}_n \xrightarrow{d} \mathbf{W}(\tau) = \lambda^{\Lambda}(\tau)'\Omega\lambda^{\Lambda}(\tau), \tag{5.4}$$

$$\mathbf{R}_n \xrightarrow{d} \chi^2_{d_2}\left\{\tau'\Omega\tau\right\}, \tag{5.5}$$

$$\begin{aligned}\mathbf{L}_n \xrightarrow{d} \mathbf{L}(\tau) &= -\frac{1}{2}\left\{\lambda^{\Lambda}_{\tau} - Z - \tau\right\}' J\left\{\lambda^{\Lambda}_{\tau} - Z - \tau\right\} + \frac{\kappa_\eta - 1}{2}(Z+\tau)'\Omega(Z+\tau) \\ &= -\frac{1}{2}\left\{\inf_{K\lambda\geq 0}\|Z+\tau-\lambda\|^2_J - \inf_{K\lambda=0}\|Z+\tau-\lambda\|^2_J\right\}.\end{aligned} \tag{5.6}$$

It is seen that the asymptotic distribution of the Rao statistic is very different from that of the two other statistics. The following proposition establishes that the asymptotic distributions of the Wald and the rescaled Quasi-Likelihood Ratio statistics are actually the same under the null or under the local alternatives.

Proposition 5.2 *Under the assumptions of Theorems 3.1 or 4.1,* $\mathbf{W}_n \overset{o_P(1)}{=} \frac{2}{\hat{\kappa}_\eta - 1}\mathbf{L}_n$.

Note that under non-local alternatives the Wald and rescaled Quasi-Likelihood Ratio tests might have different powers.

6 Testing the nullity of one coefficient

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

$$H_0 : \alpha_{0i} = 0 \quad (\text{or} \quad H_0 : \beta_{0j} = 0) \tag{6.1}$$

for some given $i \in \{1, \ldots, q\}$ (or $j \in \{1, \ldots, p\}$). 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 $d_2 = 1$. Let $\Phi(\cdot)$ denote the $\mathcal{N}(0,1)$ cumulative distribution function, $\tau^* = \tau_d/\sigma_d$ and $\sigma_d^2 = \text{Var}Z_d$. The critical regions of asymptotic level α and the local asymptotic powers are as follows.

Proposition 6.1 *(a) Under (6.1) and the assumptions of Theorem 3.1, tests of asymptotic level α (for $\alpha \le 1/2$) are defined by the critical regions*

$$\{\mathbf{W}_n > \chi^2_{1,1-2\alpha}\}, \quad \{\mathbf{R}_n > \chi^2_{1,1-\alpha}\}, \quad \{\frac{2}{\hat{\kappa}_\eta - 1}\mathbf{L}_n > \chi^2_{1,1-2\alpha}\}.$$

(b) Under the assumptions of Theorem 5.1, the local asymptotic power of the Wald and QLR tests is

$$\lim_{n\to\infty} \mathbb{P}_{n,\tau}\{\mathbf{W}_n > \chi^2_{1,1-2\alpha}\} = \lim_{n\to\infty} \mathbb{P}_{n,\tau}\{\frac{2}{\hat{\kappa}_\eta - 1}\mathbf{L}_n > \chi^2_{1,1-2\alpha}\} = 1 - \Phi(c_1 - \tau^*), \tag{6.2}$$

and that of the score test is

$$\lim_{n\to\infty} \mathbb{P}_{n,\tau}\{\mathbf{R}_n > \chi^2_{1,1-\alpha}\} = 1 - \Phi(c_2 - \tau^*) + \Phi(-c_2 - \tau^*), \tag{6.3}$$

where $c_1 = \Phi^{-1}(1-\alpha)$ and $c_2 = \Phi^{-1}(1-\alpha/2)$. (c) Moreover, for any $\tau > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{n,\tau}\left\{\mathbf{W}_n > \chi^2_{1,1-2\alpha}\right\} > \lim_{n\to\infty} \mathbb{P}_{n,\tau}\left\{\mathbf{R}_n > \chi^2_{1,1-\alpha}\right\}.$$

Proposition 6.1(c) shows that, for testing the nullity of one GARCH coefficient, the modified Wald test is locally asymptotically more powerful than the standard score test. This is illustrated in Figure 6.

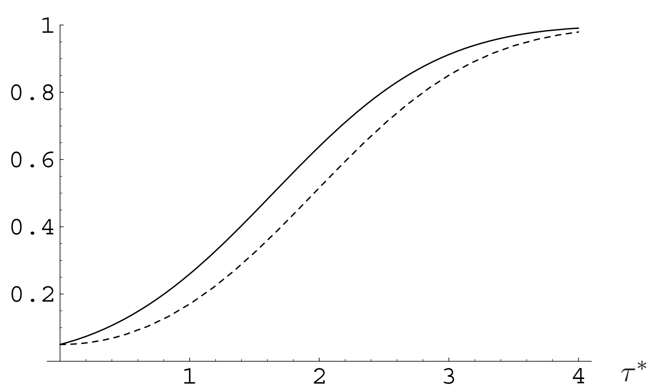

Figure 1: Local asymptotic power of the Wald test (full line) and of the score test (dashed line) for testing that one GARCH coefficient is equal to zero.

Now we will see that the modified Wald test enjoys optimality properties. Assume that η_t has a density f such that $\iota_f = \int \{1 + yf'(y)/f(y)\}^2 f(y)dy < \infty$. 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

$$I_f = \frac{\iota_f}{4} E \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}(\theta_0) = \frac{\iota_f}{4} J. \tag{6.4}$$

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

$$\pi(\tau) = 1 - \Phi\left(c_\alpha - \frac{K\tau}{\sqrt{KI_f^{-1}K'}}\right), \quad c_\alpha = \Phi^{-1}(1-\alpha). \tag{6.5}$$

A test whose level and power jointly converge to α and to the bound in (6.5), respectively, will be called asymptotically optimal.

Proposition 6.2 *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 and QLR tests are asymptotically optimal if and only if*

$$f(y) = \frac{a^a}{\Gamma(a)} \exp(-ay^2)|y|^{2a-1}, \quad a > 0, \quad \Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)dt. \tag{6.6}$$

The score test is never asymptotically optimal.

To conclude this section, it is important to note that the *standard* Wald test $\{\mathbf{W}_n > \chi^2_{1,1-\alpha}\}$ has asymptotic level $\alpha/2$. It is therefore too conservative and may lead to select too simple ARCH models. The *standard* QLR test $\{\mathbf{L}_n > \chi^2_{1,1-\alpha}\}$ has the same asymptotic level $\alpha/2$ when $\kappa = 3$. However, when the distribution of η_t is highly leptokurtic, which seems to be the case for many financial time series, Table 1 reveals that the *standard* QLR test can lead to overrejection of the null hypothesis.

7 Testing conditional homoscedaticity

In this section, we 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 no conditional heteroscedasticity

Table 1: Asymptotic levels of the standard Wald and QLR tests of nominal level 5%, for testing the nullity of one coefficient.

κ_η	2	3	4	5	6	7	8	9	10
Standard Wald	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Standard QLR	0.3	2.5	5.5	8.3	10.8	12.9	14.7	16.4	17.8

versus an ARCH(q) alternative. We therefore consider the hypothesis

$$H_0 : \alpha_{01} = \cdots = \alpha_{0q} = 0 \tag{7.1}$$

in the ARCH(q) model

$$\begin{cases} \epsilon_t & = \sigma_t \eta_t, \qquad \eta_t \quad \text{iid } (0,1) \\ \sigma_t^2 & = \omega_0 + \sum_{i=1}^q \alpha_{0i} \epsilon_{t-i}^2, \qquad \omega > 0, \quad \alpha_{0i} \geq 0. \end{cases} \tag{7.2}$$

7.1 Some simple test statistics

In his paper introducing ARCH, Engle (1982) noted that the score test is very simple to compute. Indeed, $\mathbf{R}_n = nR^2$, where R^2 is the determination coefficient of the regression of ϵ_t^2 on a constant and $\epsilon_{t-1}^2, \dots, \epsilon_{t-q}^2$. An asymptotically equivalent version is

$$\mathbf{R}_n^* = \frac{n}{(\hat{\kappa}_{\eta|2} - 1)^2} \sum_{i=1}^q \left\{ \frac{1}{n} \sum_{t=1}^n (1 - \frac{\epsilon_t^2}{\hat{\sigma}_\epsilon^2}) \frac{\epsilon_{t-i}^2}{\hat{\sigma}_\epsilon^2} \right\}^2 = n \sum_{i=1}^q \hat{\rho}_{\epsilon^2}^2(i), \tag{7.3}$$

where $\hat{\sigma}_\epsilon^2 = n^{-1} \sum_{t=1}^n \epsilon_t^2$, $\hat{\kappa}_{\eta|2} = (n\hat{\sigma}_\epsilon^4)^{-1} \sum_{t=1}^n \epsilon_t^4$ and $\hat{\rho}_{\epsilon^2}(i)$ is a standard estimator of the i-th autocorrelation of (ϵ_t^2). The score statistic thus has the interpretation of a portmanteau statistic for checking that (ϵ_t^2) is a white noise.

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

$$(\kappa_\eta - 1) J^{-1} = \begin{pmatrix} (\kappa_\eta + q - 1)\omega_0^2 & -\omega_0 & \cdots & -\omega_0 \\ -\omega_0 & & & \\ \vdots & & I_q & \\ -\omega_0 & & & \end{pmatrix}. \tag{7.4}$$

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 $\mathbf{W}_n^*$ is not the usual Wald statistic defined in (2.3), which uses the estimator $\hat{J}_n$ based on the unconstrained estimator $\hat{\theta}_n$. However, the asymptotic null and local alternative distributions of Wald statistics are not affected by the choice of a consistent estimator of J.

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

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

where $\hat{\sigma}_{LK}^2$ is an estimator of the variance of the numerator and $1_q = (1,\ldots,1)' \in \mathbb{R}^q$. In view of (A.3), (A.5), (A.6), (A.7) and (7.4) one can take

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

with $K = (0_{q\times 1}, I_q)$ and $\overline{K} = (1, 0_{1\times q})$. It follows that

$$\mathbf{LK}_n = \frac{1}{\sqrt{q}}\sum_{i=1}^{q}\sqrt{n}\hat{\rho}_{\epsilon^2}(i).$$

This form is not exactly the expression given in Lee and King (hereafter LK), but is asymptotically equivalent to it under the null (and under local alternatives). We will see that the **LK**-test enjoys some optimality properties.

7.2 Asymptotic null distributions

Using the results of Theorem 3.1, we now state the asymptotic distributions of the previous statistics under the null of independent observations. It was noted that in the ARCH case, **A6** could not be used and had to be replaced by the moment assumption **A5**. In the case of conditional homoscedasticity we do not need this assumption.

Proposition 7.1 *Under (7.1) and* **A3** *we have*

$$\mathbf{W}_n^* \xrightarrow{d} \frac{1}{2^q}\delta_0 + \sum_{i=1}^{q}\begin{pmatrix} q \\ i \end{pmatrix}\frac{1}{2^q}\chi_i^2, \qquad \mathbf{R}_n^* \xrightarrow{d} \chi_q^2, \qquad \mathbf{LK}_n \xrightarrow{d} \mathcal{N}(0,1). \tag{7.5}$$

Demos and Sentana (1998) obtained the same result for $\mathbf{W}_n^*$ 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.

7.3 Power comparisons under fixed alternatives

The next result allows to compare the efficiencies in the Bahadur sense of the "simple" tests for no conditional heteroscedasticity. Let ρ_{ϵ^2} denote the autocorrelation function of the process (ϵ_t^2), and let $\kappa_\epsilon = E_{\theta_0}(\epsilon_t^4)/\{E_{\theta_0}(\epsilon_t^2)\}^2$. The following gives the asymptotic relative efficiencies (ARE) of the simple conditional homoscedasticity tests in the presence of ARCH.

Proposition 7.2 *Let (ϵ_t) be a strictly stationary and nonanticipative solution of the ARCH(q) model (7.2) with $E(\epsilon_t^4) < \infty$ and $\sum_{i=1}^q \alpha_{0i} > 0$. Then,*

$$\begin{aligned}
\mathbf{ARE}(\mathbf{R}^*/\mathbf{LK}) &:= \lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{R}}(\mathbf{R}_n^*)\{\lim_{n\to\infty} -\frac{2}{n}\log\{1-\Phi(\mathbf{LK}_n)\}\}^{-1} \\
&= \frac{q\sum_{i=1}^q \rho_{\epsilon^2}^2(i)}{\{\sum_{i=1}^q \rho_{\epsilon^2}(i)\}^2} \geq 1, \\
\mathbf{ARE}(\mathbf{R}^*/\mathbf{W}^*) &:= \lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{R}}(\mathbf{R}_n^*)\{\lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{W}}(\mathbf{W}_n^*)\}^{-1} \\
&= \frac{\sum_{i=1}^q \rho_{\epsilon^2}^2(i)}{\sum_{i=1}^q \alpha_{0i}^2} \geq 1, \\
\mathbf{ARE}(\mathbf{R}/\mathbf{W}^*) &:= \lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{R}}(\mathbf{R}_n)\{\lim_{n\to\infty} -\frac{2}{n}\log S_{\mathbf{W}}(\mathbf{W}_n^*)\}^{-1} \\
&= \frac{\kappa_\epsilon - \kappa_\eta}{\kappa_\eta(\kappa_\epsilon - 1)\sum_{i=1}^q \alpha_{0i}^2} \geq 1,
\end{aligned}$$

with equalities when $q = 1$.

Because a test is consistent whenever its slope is positive, these conditional homoscedasticity tests are consistent under much more general assumptions than the ARCH(q) alternative.

Proposition 7.3 *Let (ϵ_t) be a strictly stationary and ergodic process. The tests based on $\mathbf{R_n^*}$, and $\mathbf{R_n}$, are consistent against alternatives of the form*

$$H_{a1} : E\epsilon_t^4 < \infty \quad \textit{and} \quad \sum_{i=1}^q \rho_{\epsilon^2}^2(i) > 0.$$

The test based on $\mathbf{LK_n}$, is consistent against alternatives of the form

$$H_{a2} : E\epsilon_t^4 < \infty \quad \textit{and} \quad \sum_{i=1}^{q} \rho_{\epsilon^2}(i) \neq 0.$$

The test based on $\mathbf{W_n}^*$*, is consistent against alternatives of the form*

$$H_{a3} : \epsilon_t = \sqrt{\omega_0 + \sum_{i=1}^{q} \alpha_{0i}\epsilon_{t-i}^2}\,\eta_t, \quad \textit{where } \omega_0 > 0, \alpha_{0i} \geq 0, \sum_{i=1}^{q} \alpha_{0i} > 0, \ (\eta_t) \textit{ is a}$$

sequence of (possibly non-iid) variables verifying **A3** *and* $E(\eta_t^2 \mid \epsilon_{t-1}, \ldots, \epsilon_{t-q}) = 1, a.s.$

Moreover $E|\epsilon_t|^{2s} < \infty$ *for some* $s > 0$.

It can be noted that the consistency of the test based on $\mathbf{W_n}^*$ is established under stronger dependence assumptions on the process (ϵ_t) than for the other tets. On the other hand, the fourth-order moment condition is avoided for the $\mathbf{W_n}^*$ test, which is interesting for financial series applications. It is also worth noting that the LK test may fail to detect particular forms of heteroscedasticity. For instance, let $\epsilon_t = a_t\sqrt{X_t}$ where (a_t) is an iid process such that $P(a_t = 1) = P(a_t = -1) = 0.5$, and (X_t) is the positive process defined by

$$X_t = 1 + 0.5X_{t-1} - 0.5X_{t-2} + u_t, \quad (u_t) \ iid \ \sim \mathcal{U}[-c, c], \quad (u_t) \ \perp \ (a_t),$$

for some sufficiently small c. Than it is easily seen that $\rho_{\epsilon^2}(1) = -\rho_{\epsilon^2}(2) = 1/3$. Thus, for $q = 2$, $\mathbf{LK_n}$ has a non-degenerate limiting distribution and the LK test will often fail to reject conditional homoscedaticity. The score tests are consistent for this alternative.

Versions of tests which are asymptotically equivalent under the null and local alternatives may have different slopes. The asymptotic efficiencies derived in Proposition 5.1 do not coincide with those just derived for the "simple" test statistics. However, they can be evaluated by simulation. It can be seen that

$$\theta_{0|2} = \begin{pmatrix} E_{\theta_0}(\epsilon_t^2) \\ 0_{q\times 1} \end{pmatrix}, \quad J = E_{\theta_0}(\sigma_t^{-4} Z_t Z_t'), \quad J_{0|2} = \{E_{\theta_0}(\epsilon_t^2)\}^{-2} E_{\theta_0}(Z_t Z_t'),$$

with $Z_t = (1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-q}^2)'$. The results displayed in Table 2 concern the ARCH(1), for α_1 ranging from 0 to 0.4, with gaussian conditional distributions. Note that when $q = 1$ the AREs computed in Proposition 7.2 are equal to 1. Moreover, the slope of the Rao statistic given by (5.2) coincides with those of the other versions of the score, and is equal to α_1^2. It is seen from Table 2 that

$$\mathbf{W} \prec \mathbf{L} \prec \mathbf{R} \sim \mathbf{R}^* \sim \mathbf{W}^* \sim \mathbf{LK}$$

where $\mathbf{S} \prec \mathbf{T}$ means that a test $\mathbf{S}$ is less efficient than $\mathbf{T}$, and $\mathbf{S} \sim \mathbf{T}$ means that the two tests have the same slope. Table 3 reports efficiency results for an ARCH(2) and shows,

Table 2: Asymptotic efficiencies of the score and QLR tests relative to the Wald test for testing conditional homoscedasticity in an ARCH(1). The number of replications of the ratio is $N = 10$, the expectations are evaluated by empirical means of size 10,000,000.

α_1	0.1	0.2	0.3	0.4	0.5
ARE($\mathbf{R}^*/\mathbf{W}$)	1.7	2.3	2.9	3.4	4.0
ARE($\mathbf{L}/\mathbf{W}$)	1.4	1.8	2.2	2.7	3.3

in particular, that the equivalence observed in the case $q = 1$ does not hold in general. Colors, from blue to red, indicate the rankings of those tests. To summarize, the tests can be ranked as follows

$$\mathbf{W} \prec \mathbf{L} \prec \mathbf{W}^* \prec \mathbf{R} \prec \mathbf{R}^*.$$

The LK cannot be ranked in general: it can have the lowest or the highest asymptotic efficiency depending on the parameter values.

7.4 Power comparisons under local alternatives

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

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

Since by (6.4) and (7.4),

$$KI_f^{-1}K' = 4\iota_f^{-1}(\kappa_\eta - 1)^{-1}I_q, \tag{7.6}$$

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

Table 3: Asymptotic efficiencies of conditional homoscedasticity tests relative to the Wald test, for an ARCH(2) alternative. The number of replications of the slopes is $N = 10$, the expectations are evaluated by empirical means of size 10,000,000. Missing values correspond to the non existence of the 4th-order moment or to $\alpha_{01} = \alpha_{02} = 0$.

ARE(L/W)	α_{02}					
α_{01}	0	0.1	0.2	0.3	0.4	0.5
0	-	**1.4**	**1.8**	**2.2**	**2.7**	**3.3**
0.1	**1.4**	**1.5**	**1.8**	**2.1**	**2.6**	**3.2**
0.2	**1.8**	**1.8**	**2.0**	**2.4**	**2.9**	-
0.3	**2.2**	**2.3**	**2.5**	**2.9**	-	-
0.4	**2.7**	**2.8**	**3.1**	-	-	-
0.5	**3.3**	**3.5**	-	-	-	-

ARE($\mathbf{R}^*$/W)	α_{02}					
α_{01}	0	0.1	0.2	0.3	0.4	0.5
0	-	**1.7**	**2.3**	**2.9**	**3.4**	**4.0**
0.1	**1.7**	**1.9**	**2.4**	**2.9**	**3.4**	**4.0**
0.2	**2.4**	**2.7**	**3.1**	**3.6**	**4.2**	-
0.3	**3.2**	**3.6**	**4.1**	**4.7**	-	-
0.4	**4.0**	**4.7**	**5.3**	-	-	-
0.5	**5.0**	**5.9**	-	-	-	-

ARE(R/W)	α_{02}					
α_{01}	0	0.1	0.2	0.3	0.4	0.5
0	-	**1.7**	**2.3**	**2.9**	**3.4**	**4.0**
0.1	**1.7**	**1.7**	**2.1**	**2.6**	**3.1**	**3.6**
0.2	**2.3**	**2.3**	**2.5**	**2.8**	**3.3**	-
0.3	**2.9**	**2.9**	**3.0**	**3.3**	-	-
0.4	**3.4**	**3.5**	**3.6**	-	-	-
0.5	**4.0**	**4.1**	-	-	-	-

ARE($\mathbf{W}^*$/W)	α_{02}					
α_{01}	0	0.1	0.2	0.3	0.4	0.5
0	-	**1.7**	**2.3**	**2.9**	**3.4**	**4.0**
0.1	**1.7**	**1.6**	**2.0**	**2.4**	**2.9**	**3.4**
0.2	**2.3**	**1.9**	**2.0**	**2.2**	**2.6**	-
0.3	**2.9**	**2.4**	**2.2**	**2.3**	-	-
0.4	**3.4**	**2.9**	**2.6**	-	-	-
0.5	**4.0**	**3.4**	-	-	-	-

ARE(LK/W)	α_{02}					
α_{01}	0	0.1	0.2	0.3	0.4	0.5
0	-	**0.8**	**1.1**	**1.4**	**1.7**	**2.0**
0.1	**1.0**	**1.9**	**2.2**	**2.5**	**2.9**	**3.3**
0.2	**1.7**	**2.6**	**3.1**	**3.6**	**4.1**	-
0.3	**2.4**	**3.4**	**4.1**	**4.7**	-	-
0.4	**3.4**	**4.5**	**5.3**	-	-	-
0.5	**4.5**	**5.7**	-	-	-	-

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)[1] and the references therein for the concept of MSSMP and SMP test). In view of (7.6), this MSSMP test has the power

$$\pi(\tau) = 1 - \Phi\left(c_\alpha - \frac{\sum_{i=1}^q \tau_i}{\sqrt{4q\iota_f^{-1}(\kappa_\eta - 1)^{-1}}}\right), \quad c_\alpha = \Phi^{-1}(1-\alpha). \tag{7.7}$$

The following corollary gives the local asymptotic powers of the conditional homoscedasticity tests considered in this section, and shows that the LK 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.

Proposition 7.4 *Under the local alternatives $H_n(\tau)$, $\tau > 0$, and the assumptions of Theorem 4.1 with $p = 0$, $d_1 = 1$ and $d_2 = q$, we have*

$$\lambda^\Lambda(\tau) = \left((Z_1 + \tau_1) + \omega \sum_{i=2}^d (Z_i + \tau_i)^-, (Z_2 + \tau_2)^+, \cdots, (Z_d + \tau_d)^+\right)', \tag{7.8}$$

where $Z \sim \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right)$ and $(\kappa_\eta - 1)J^{-1}$ is given in (7.4). Thus, the local asymptotic power of the modified Wald, score and LK tests are given by

$$\begin{aligned}
\lim_{n\to\infty} P\{\mathbf{W}_n > \mathbf{w}_{1-\alpha}\} &= P\left\{\sum_{i=1}^q (U_i + \tau_i)^2 \mathbb{1}_{\{U_i+\tau_i>0\}} > \mathbf{w}_{1-\alpha}\right\} \\
\lim_{n\to\infty} P\left\{\mathbf{R}_n > \chi^2_{q,1-\alpha}\right\} &= P\left\{\chi^2_q\left(\sum_{i=1}^q \tau_i^2\right) > \chi^2_{q,1-\alpha}\right\} \\
\lim_{n\to\infty} P\{\mathbf{LK}_n > c_\alpha\} &= 1 - \Phi\left(c_\alpha - \frac{\sum_{i=1}^q \tau_i}{\sqrt{q}}\right),
\end{aligned} \tag{7.9}$$

where $U = (U_1, \ldots, U_q)' \sim \mathcal{N}(0, I_q)$.

Under the assumptions of Proposition 6.2, the LK test is asymptotically MSSMP (in the sense that the right-hand side of (7.9) is equal to the upper bound $\pi(\tau)$ defined by (7.7)) if and only if the density f of η_t belongs to the class defined by (6.6).

It is well known that there exists no satisfactory notion of optimality for testing hypothesis on multidimensional parameters. The LK 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 7.4, the local asymptotic power of LK test may be lower than that of the Wald test, and even lower than that of the two-sided score test.

[1]The authors greatly thank Professor Shi for sending them these two papers.

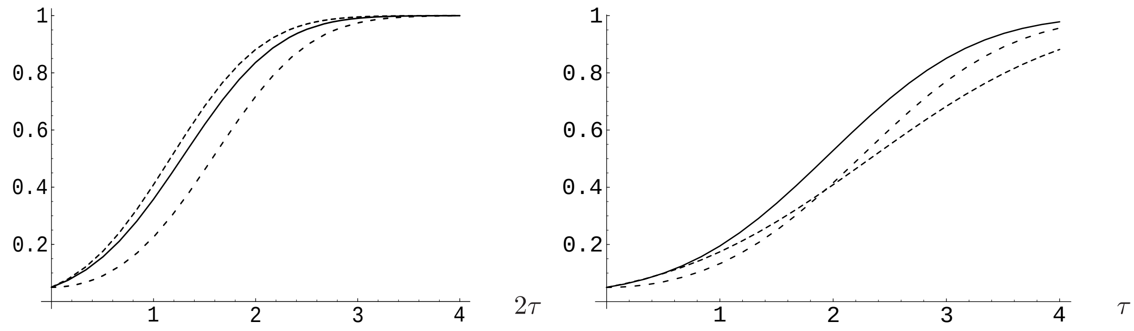

Figure 2: Local asymptotic power of the Wald (full line), score (dashed line) and LK (dotted line) for testing conditional homoscedasticity with an ARCH(2) model where $\alpha_1 = \alpha_2 = \tau/\sqrt{n}$ (left figure) and $\alpha_1 = \tau/\sqrt{n},\ \alpha_2 = 0$ or $\alpha_1 = 0,\ \alpha_2 = \tau/\sqrt{n}$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

7.5 Numerical experiments

In this section, we investigate the finite-sample properties of the tests for conditional homoscedasticity studied in this paper. First we generate $N = 5,000$ replications of samples of iid variables of size $n = 100, 500$ and $5,000$, for different distributions. The tests are designed for an ARCH(2) alternative. For $n = 5,000$ the relative rejection frequencies, presented in Table 4, are almost always within the 0.05 significant limits 4.38% and 5.62%. For smaller sample sizes, and non-gaussian distributions, the type I error is not perfectly well controlled by the Wald test. Deviations can also be noticed for the Rao and QLR tests for the uniform and exponential distributions but, even for $n = 100$, the sizes are never very far from the theoretical 5%.

We now turn to the power of those tests against local deviations from the null hypothesis. The results are presented in Table 5 and, for ease of reading, the highest rejection frequencies are written in bold for each experiment. In the upper part of the table, the DGP is an ARCH(q) with $q = 1, 2, 3$ and $\alpha_1 = \cdots = \alpha_q > 0$. The conclusion drawn from the comparison of the local asymptotic powers remains valid for these simulation experiments. The Rao test is clearly dominated by the three other ones, whatever the sample size. For $q = 1$, the local asymptotic powers of the Wald, QLR and LK tests are equal (by Propositions 6.2 and 7.4, these tests are locally asymptotically uniformly most powerful), and are very close for $n = 500$ and $n = 5,000$, with a slight advantage to the QLR test. This advantage can also be noticed for $q = 3$. For $q = 2$ the asymptotic superiority of the one-sided LK test is reflected in finite samples. However, when the alternative is not

Table 4: Empirical size (in %) of the Wald, score, QLR and LK tests for conditional homoscedasticity. The tests are based on an ARCH(2) model. 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 independent sequence, distributed as the $\mathcal{N}(0,1)$ ($\mathcal{N}$), the Student t with $\nu = 8$ degrees of freedom (St_8), the uniform ($\mathcal{U}$) on $(-1/2, 1/2)$, or the exponential distribution($\mathcal{E}$) of density $f(x) = e^{-x-1}1_{\{x>-1\}}$.

	$n = 100$				$n = 500$				$n = 5000$			
	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$
$\mathcal{N}$	4.38	4.40	4.44	4.96	5.36	4.30	5.22	4.86	4.70	4.86	4.90	4.84
St_8	6.46	4.56	5.46	4.76	5.98	4.92	4.76	4.86	5.96	4.76	4.70	4.98
$\mathcal{U}$	4.20	6.84	6.10	4.88	4.22	5.62	5.22	3.94	4.54	5.10	5.20	4.66
$\mathcal{E}$	6.30	4.94	4.98	4.68	6.66	6.22	3.98	5.32	6.74	4.92	4.42	5.48

symmetric in the ARCH coefficients, as it is the case in the lower part of Table 5, the LK test can be much less powerful than its competitors, both asymptotically and in finite samples. For this reason it cannot be recommended to practitioners.

8 Illustrative example

We now consider an application to the daily returns of the French CAC40 and the Standard & Poor's 500 indexes. The presence of GARCH in these series has been documented by many empirical studies. Our aim in this section is to compare the abilities of the various tests considered in this paper to detect the ARCH effect. As the sample size n increases, the p-values of the tests are expected to decrease. Assuming that the series is indeed a GARCH, the way those p values decrease to zero is an indication of the performances of the tests in finite sample.

The CAC data range from January 2, 2004 to December 29, 2006. The total length of the series is 771 but the sample size used for the tests ranges from $n = 400$ to $n = 600$. In the first experiment, the tests considered are the score and LK tests for conditional homoscedasticity, the ARCH order varying from $q = 1$ to $q = 9$. For each sample size n, a set of 201 p-values are computed based on the observations $X_{j+1}, \ldots, X_{j+n}$, for $j =$

Table 5: Empirical power (in %) of the Wald, score, QLR and LK 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 gaussian innovations.

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

q	$n = 500$				$n = 5000$				$n = \infty$			
	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$
1	40.0	30.0	**40.4**	39.1	42.4	32.2	**43.0**	42.3	**44.2**	32.3	**44.2**	**44.2**
2	58.6	45.1	59.1	**59.5**	63.6	46.5	63.7	**66.3**	61.9	46.0	61.9	**68.3**
3	73.4	57.0	**76.3**	74.1	81.1	57.8	**81.3**	81.1	74.7	57.2	74.7	**83.0**

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

q	$n = 500$				$n = 5000$				$n = \infty$			
	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$	$\mathbf{W}_n$	$\mathbf{R}_n$	$\mathbf{L}_n$	$\mathbf{LK}_n$
2	**79.4**	62.7	73.1	55.2	**85.9**	73.0	81.3	64.8	**85.1**	77.1	**85.1**	68.3
3	**93.7**	85.0	89.5	65.9	**97.4**	94.4	95.2	78.5	**99.0**	97.7	**99.0**	83.0

$0, \ldots, 201$. Figure 3 displays the averages of these p-values, for the score (left panel) and LK tests (right panel). Clearly, the tests based on $q = 1$ are dominated by the tests based on higher-order ARCH models. For $n = 600$ the average p-values are very small, except in the case $q = 1$. For the score test, the values of $q > 1$ lead to similar results, but this is less true for the LK test. Now for a given q, the LK test has better performances than the score test, in the sense that it is able to detect the ARCH effect more rapidly as n increases.

The S&P500 data range from January 2, 2003 to December 29, 2006. The total length of the series is 1007. The sample size used for the tests ranges from $n = 800$ to $n = 950$. Figure 4 plots the averages of the p-values of the Wald, score, QLR and LK conditional homoscedasticity tests, in the ARCH(2) model, for the CAC40 (left panel) and S&P500

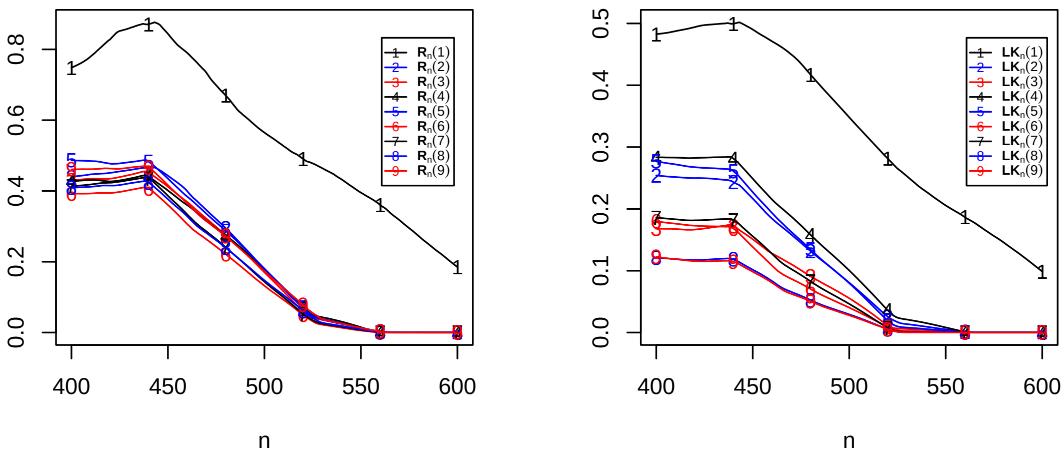

Figure 3: Average p-values of the score (left panel) and LK tests (right panel) for conditional homoscedasticity, in ARCH(q) models for $q = 1, \ldots, 9$, in function of the sample size n, for the CAC40 index.

(right panel). In both cases (i) the results for the QLR and Wald tests are similar, and (ii) the score test requires larger sample sizes to detect conditional heteroscedasticity. Looking at the results for the LK tests, the conclusions are opposite for the two series. For the CAC, this test does a better job than the three others, but for the S&P500, it is much less efficient than the Wald-QLR and is similar to the score.

9 Concluding remarks

The usual methodology for testing the nullity of coefficients in GARCH models is based on the standard Wald, score and QLR statistics. This article has shown that caution is needed in the use of such statistics, because the null hypothesis puts the parameter at the boundary of the parameter space. From the derivation of the asymptotic null and local alternative distributions of those statistics, four main conclusions can be drawn: i) the asymptotic sizes of the *standard* Wald and QLR tests can be very different from the *nominal* levels based on (invalid) χ^2 distributions; ii) the modified tests of this paper tackle the boundary problem; moreover, iii) the modified Wald and QLR tests remain equivalent under the null and local alternatives; iv) the usual Rao test remains valid for testing a value on the boundary, but looses its local optimality properties;

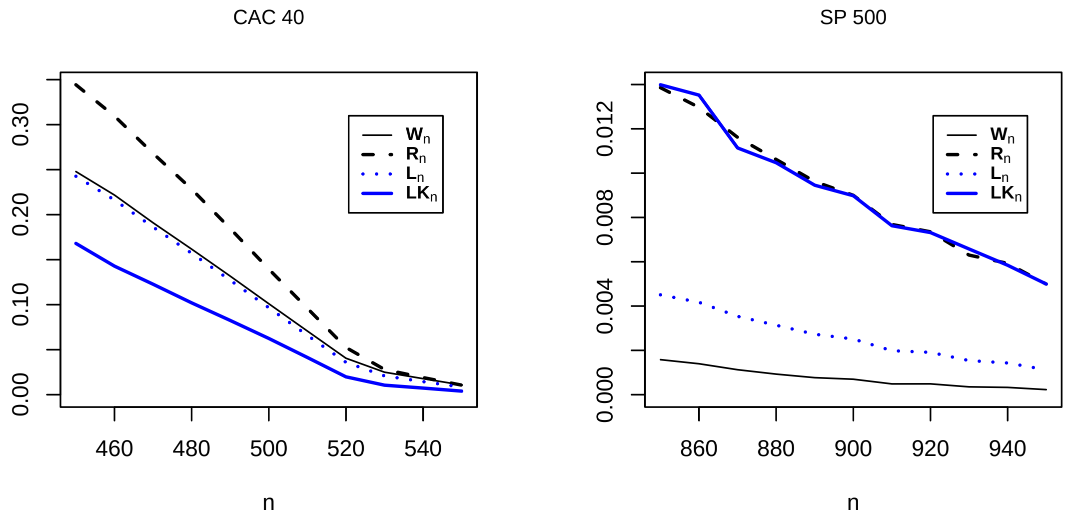

Figure 4: Average p-values of the Wald, score, QLR and LK conditional homoscedasticity tests for the CAC40 (left panel) and SP500 (right panel) .

For the two special cases considered in this paper, the approaches of Bahadur and Pitman allow efficiency comparisons, and shed light on the relative merits of the different tests. For the nullity of one coefficient, the modified Wald and QLR tests are locally asymptotically optimal, when the conditional density belongs to a class which is not restricted to the standard Gaussian. For the absence of conditional heteroscedasticity, several simple tests can be used, which have different powers under fixed alternatives. Efficiency comparisons for the ARCH(1) and ARCH(2) models suggest that the different versions of the score test are preferable to the other competitors in the Bahadur ARE sense. However, inverse conclusions are drawn when the local approach is adopted. Indeed, the score test appears to be locally dominated by the equivalent Wald and QLR tests. The one-sided version of the score test proposed by Lee and King enjoys optimality properties, but only for alternatives in certain directions. To conclude, a simple version of the Wald test, rejecting the null when the sum of the squared coefficients is large, can be recommended for testing for ARCH. From both local and non local points of view, our theoretical study and numerical experiments suggest that the behavior of this test is always close to the optimum.

Appendix: Proofs and technical results

A.1 Proof of Theorem 3.2

The convergence in distribution (3.3) is a direct application of the continuous mapping theorem, because $\sqrt{n}\hat{\theta}_n^{(2)\prime} = K\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} K\lambda^{\Lambda}$ under H_0 by Theorem 3.1.

We now turn to the proof of (3.4). 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

$$\frac{\partial\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right)}{\partial\theta} = K'\frac{\partial\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right)}{\partial\theta^{(2)}}. \tag{A.1}$$

A Taylor expansion yields

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

The last d_2 components of this vector relation give

$$\sqrt{n}\frac{\partial\tilde{\mathbf{l}}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta^{(2)}} + KJ\sqrt{n}\left(\hat{\theta}_{n|2} - \theta_0\right), \tag{A.3}$$

and the first d_1 components give

$$0 \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}} + \overline{K}J\overline{K}'\sqrt{n}\left(\hat{\theta}_{n|2}^{(1)} - \theta_0^{(1)}\right), \tag{A.4}$$

using

$$\left(\hat{\theta}_{n|2} - \theta_0\right) = \overline{K}'\left(\hat{\theta}_{n|2}^{(1)} - \theta_0^{(1)}\right). \tag{A.5}$$

In view of (A.4), we have

$$\sqrt{n}\left(\hat{\theta}_{n|2}^{(1)} - \theta_0^{(1)}\right) \overset{o_P(1)}{=} -\left(\overline{K}\hat{J}_{n|2}\overline{K}'\right)^{-1}\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}}. \tag{A.6}$$

Using (A.1), (A.3), (A.5) and (A.6) we obtain

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

Now recall that under H_0

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} := \sqrt{\frac{n}{\kappa_\eta - 1}}\begin{pmatrix} \frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta^{(1)}} \\ \frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta^{(2)}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{0, J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}\right\}. \tag{A.7}$$

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 $W_2 - J_{21}J_{11}^{-1}W_1 \sim \mathcal{N}\left(0, J_{22} - J_{21}J_{11}^{-1}J_{12}\right)$.

Turning to the proof of (3.5) and using (A.5) and (A.6), several Taylor expansions give

$$\begin{aligned} n\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right) &\overset{o_P(1)}{=} n\mathbf{l}_n(\theta_0) + n\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta'}\left(\hat{\theta}_{n|2} - \theta_0\right) + \frac{n}{2}\left(\hat{\theta}_{n|2} - \theta_0\right)' J \left(\hat{\theta}_{n|2} - \theta_0\right) \\ &\overset{o_P(1)}{=} n\mathbf{l}_n(\theta_0) - \frac{n}{2}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(1)'}}\left(\overline{K}J\overline{K}'\right)^{-1}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(1)}} \end{aligned} \tag{A.8}$$

and

$$n\mathbf{l}_n\left(\hat{\theta}_n\right) \overset{o_P(1)}{=} n\mathbf{l}_n(\theta_0) + n\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta'}\left(\hat{\theta}_n - \theta_0\right) + \frac{n}{2}\left(\hat{\theta}_n - \theta_0\right)' J \left(\hat{\theta}_n - \theta_0\right). \tag{A.9}$$

By subtraction,

$$\begin{aligned} \mathbf{L}_n \overset{o_P(1)}{=} & -n\left\{\frac{1}{2}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(1)'}}\left(\overline{K}J\overline{K}'\right)^{-1}\frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta^{(1)}}\right. \\ & \left. + \frac{\partial \mathbf{l}_n(\theta_0)}{\partial \theta'}\left(\hat{\theta}_n - \theta_0\right) + \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right)' J \left(\hat{\theta}_n - \theta_0\right).\right\} \end{aligned} \tag{A.10}$$

Under H_0, by 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

$$\begin{aligned} \mathbf{L} &= -\frac{1}{2}Z'JZ + \frac{1}{2}Z'(\kappa_\eta - 1)\Omega Z + Z'J'\lambda^\Lambda - \frac{1}{2}\lambda^{\Lambda'}J\lambda^\Lambda \\ &= -\frac{1}{2}(\lambda^\Lambda - Z)'J(\lambda^\Lambda - Z) + \frac{\kappa_\eta - 1}{2}Z'\Omega Z \end{aligned} \tag{A.11}$$

and the conclusion easily follows.

A.2 Proof of Proposition 5.1.

Under H_1 we have

$$\lim_{n\to\infty}\frac{\mathbf{W}_n}{n} = \frac{1}{\kappa_\eta - 1}\theta_0^{(2)'}\left(KJ^{-1}K'\right)^{-1}\theta_0^{(2)}.$$

Thus, (5.1) is obtained by showing that

$$\log S_{\mathbf{W}}(x) \quad \sim \quad \log P(\chi^2_{d_2} > x) \qquad x \to \infty, \tag{A.12}$$

and by noting that $\mathbf{W}_n \to \infty$ and $\log P(\chi^2_{d_2} > x) \sim -x/2$ as $x \to \infty$ (Bahadur, 1960).

The behaviour of the two other statistics is more intricate because the constrained estimator $\hat{\theta}_{n|2}$ does not converges to θ_0 under H_1. Under general conditions, see White (1982), the QMLE $\hat{\theta}_{n|2}$ in the misspecified (by H_0) model converges to

$$\theta_{0|2} = \arg \min_{\theta \in \Theta:\, \theta^{(2)}=0} E_{\theta_0}\{\ell_t(\theta)\},$$

provided that this minimum exists and is unique. For the existence, moments of order 4 are required. For the uniqueness, a necessary condition is the local identifiability of $\theta_{0|2}$ (see White, 1982). This is achieved in our model because it can be shown that, for any $\theta \in \Theta$

$$J(\theta) = E_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'}\right) \quad \text{is a positive definite matrix.} \tag{A.13}$$

Let $J^*_{0|2} = J^*(\theta_{0|2})$ where

$$J^*(\theta) = E_{\theta_0}\left(\frac{\partial^2 \ell_t}{\partial \theta \partial \theta'}(\theta)\right).$$

The existence of $J^*(\theta)$ is ensured when $E\epsilon_t^6 < \infty$. Note that $J^*(\theta_0) = J(\theta_0)$ but $J^*(\theta_{0|2}) \neq J(\theta_{0|2})$. It follows from the a.s. convergence of $\hat{\theta}_{n|2}$ to $\theta_{0|2}$ that, similar to (A.2)-(A.3),

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

and then, assuming that $\overline{K} J^*_{0|2} \overline{K}'$ is non-singular,

$$\begin{aligned} \sqrt{n}\left(\hat{\theta}^{(1)}_{n|2} - \theta^{(1)}_{0|2}\right) &\overset{o_P(1)}{=} -(\overline{K} J^*_{0|2} \overline{K}')^{-1} \sqrt{n}\frac{\partial \mathbf{l}_n(\theta_{0|2})}{\partial \theta^{(1)}} \\ &= -(\overline{K} J^*_{0|2} \overline{K}')^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{\sigma_t^2(\theta_{0|2})} \frac{\partial \sigma_t^2(\theta_{0|2})}{\partial \theta^{(1)}} \left(1 - \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_{0|2})}\eta_t^2\right). \end{aligned}$$

Note that the summand is centered because $\theta_{0|2}$ minimizes the limit criterion $E_{\theta_0}\{\ell_t(\theta)\}$. However it is not a martingale difference. To apply a central limit theorem, one can rely on the strong mixing properties of GARCH processes. Such properties require additional assumptions on the density of η_t (see e.g. Carrasco and Chen (2002), Francq and Zakoian (2006)) and are beyond the scope of this paper. Applying this central limit theorem we have under H_1,

$$\sqrt{n}\left(\hat{\theta}_{n|2} - \theta_{0|2}\right) = O_P(1). \tag{A.14}$$

Therefore

$$\sqrt{n}\frac{\partial \tilde{\mathbf{l}}_n(\hat{\theta}_{n|2})}{\partial \theta^{(2)}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial \mathbf{l}_n(\theta_{0|2})}{\partial \theta^{(2)}} + K J^*_{0|2} \sqrt{n}\left(\hat{\theta}_{n|2} - \theta_{0|2}\right) \overset{O_P(\sqrt{n})}{=} \sqrt{n}\frac{\partial \mathbf{l}_n(\theta_{0|2})}{\partial \theta^{(2)}}.$$

It follows that, using the convergence of $\hat{J}_{n|2}$ to $J_{0|2}$, and of $\hat{\kappa}_{\eta|2}$ to $\kappa_{\eta|2}$,

$$\frac{\mathbf{R}_n}{n} \overset{o_P(1)}{=} \frac{1}{\kappa_{\eta|2}-1}\left\|\frac{\partial\tilde{\mathbf{l}}_n(\hat{\theta}_{n|2})}{\partial\theta^{(2)}}\right\|^2_{KJ_{0|2}^{-1}K'} \overset{o_P(1)}{=} \frac{1}{\kappa_{\eta|2}-1}\left\|\frac{\partial\mathbf{l}_n(\theta_{0|2})}{\partial\theta^{(2)}}\right\|^2_{KJ_{0|2}^{-1}K'},$$

from which (5.2) can be deduced by application of the ergodic theorem and arguments already used to establish (5.1).

Now similar to (A.8) and (A.9) we have

$$n\tilde{\mathbf{l}}_n\left(\hat{\theta}_{n|2}\right) \overset{o_P(1)}{=} n\mathbf{l}_n\left(\theta_{0|2}\right)+n\frac{\partial\mathbf{l}_n\left(\theta_{0|2}\right)}{\partial\theta'}\left(\hat{\theta}_{n|2}-\theta_{0|2}\right)+\frac{n}{2}\left(\hat{\theta}_{n|2}-\theta_{0|2}\right)'J^*_{0|2}\left(\hat{\theta}_{n|2}-\theta_{0|2}\right),$$
$$n\mathbf{l}_n\left(\hat{\theta}_n\right) \overset{o_P(1)}{=} n\mathbf{l}_n\left(\theta_0\right)+n\frac{\partial\mathbf{l}_n\left(\theta_0\right)}{\partial\theta'}\left(\hat{\theta}_n-\theta_0\right)+\frac{n}{2}\left(\hat{\theta}_n-\theta_0\right)'J\left(\hat{\theta}_n-\theta_0\right).$$

It follows, using (A.14), that

$$\frac{\mathbf{L}_n}{n} \overset{o_P(1)}{=} \mathbf{l}_n\left(\theta_{0|2}\right)-\mathbf{l}_n\left(\theta_1\right) \overset{o_P(1)}{=} E_{\theta_0}\{\ell_t(\theta_{0|2})-\ell_t(\theta_1)\},$$

from which (5.3) can be deduced, using

$$E_{\theta_0}\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_{0|2})}\right)=1. \tag{A.15}$$

The consistency of the three test follows from the positivity of the Bahadur slopes. From (5.1) it is seen that, in view of the positive definiteness of J, the Wald test is consistent. In (5.2) the positivity of the right-hand side is ensured if $D(\theta_{0|2})$ is not equal to zero. The consistency of the QLR test follows from

$$-E_{\theta_0}\left(\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_{0|2})}\right)\geq -\log E_{\theta_0}\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_{0|2})}\right)=0,$$

by (A.15) and Jensen's inequality, with strict inequality when $\sigma_t^2(\theta_{0|2})\neq\sigma_t^2(\theta_0)$. The latter is a consequence of the identifiability assumptions **A3-A4**.

A.3 Proof of Theorem 5.1

By arguments used in the proof of Theorem 4.1, it can be shown that with probability 1 under $\mathbb{P}_{n,\tau}$

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

The convergence in distribution (5.4) is then obtained by the same arguments as in the proof of (3.3), using Theorem 4.1. With the notation introduced in (A.7) a Taylor expansion gives

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

For the convergence in distribution we use $\sqrt{n}\frac{\partial \mathbf{l}_n(\theta_n)}{\partial\theta} \xrightarrow{d} \mathcal{N}\left(0,(\kappa_\eta-1)J\right)$, which is established in the proof of Theorem 4.1, and we note that $\sqrt{n}(\theta_0-\theta_n)=-\tau$. We then have

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

and (5.5) follows by the arguments used to establish (3.4). Similarly, (5.6) follows from the arguments used to prove (3.5) and from

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

A.4 Proof of Proposition 5.2.

We start by introducing some notations. Let

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

where, for ease of notation, $\mathbf{l}_n$ is as in Section 2, but with variables indexed by $\{t,n\}$ instead of t. In the proof of Theorem 4.1 it is proved that

$$\sqrt{n}(\hat\theta_n-\theta_0)\overset{o_P(1)}{=}\lambda^\Lambda_{n,\tau}:=\arg\inf_{\lambda\in\Lambda}\|\tilde Z_{n,\tau}-\lambda\|_{J_{n,\tau}}\overset{o_P(1)}{=}\arg\inf_{\lambda\in\Lambda}\|\tilde Z_{n,\tau}-\lambda\|_J,$$

where $\tilde Z_{n,\tau}=Z_{n,\tau}+\tau$. We then have

$$\begin{aligned}\mathbf{W}_n &= \frac{n}{\hat\kappa_\eta-1}(\hat\theta_n^{(2)}-\theta_0^{(2)})'\left\{K\hat J^{-1}K'\right\}^{-1}(\hat\theta_n^{(2)}-\theta_0^{(2)})\\ &\overset{o_P(1)}{=}\frac{n}{\kappa_\eta-1}(\hat\theta_n-\theta_0)'K'\left\{KJ^{-1}K'\right\}^{-1}K(\hat\theta_n-\theta_0)\\ &= \|\sqrt{n}(\hat\theta_n-\theta_0)\|^2_\Omega\overset{o_P(1)}{=}\|\lambda^\Lambda_{n,\tau}\|^2_\Omega.\end{aligned}$$

Now, similarly to (3.1), we have

$$\lambda^\Lambda_{n,\tau}\overset{o_P(1)}{=}\tilde Z_{n,\tau}\mathbb{1}_\Lambda(\tilde Z_{n,\tau})+\sum_{i=1}^{2^{d_2}-1}P_i\tilde Z_{n,\tau}\mathbb{1}_{\mathcal{D}_i}(\tilde Z_{n,\tau}),\tag{A.16}$$

where $P_i=I_d-J^{-1}M_i$ and $M_i=K_i'\left(K_iJ^{-1}K_i'\right)^{-1}K_i$. It follows that

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

Let $Z_n=-J_n^{-1}\sqrt{n}\frac{\partial\mathbf{l}_n(\theta_0)}{\partial\theta}$. Turning to $\mathbf{L}_n$, using (A.10) we obtain, similarly to (A.11)

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

A Taylor expansion shows that $Z_n \overset{o_P(1)}{=} Z_{n,\tau} + \tau = \tilde{Z}_{n,\tau}$, from which we deduce

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

By (A.16) we have

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

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

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

It follows that

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

because $\Omega - \Omega_i - P_i'\Omega = 0$. 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.5 Proof of Proposition 6.1.

(a) We have $\Lambda = \mathbb{R}^{d_1} \times [0, \infty)$, $K = (0, \dots, 0, 1)$, $\mathcal{K} = \{K\}$, and $\lambda^{\Lambda} = Z\mathbb{1}_{Z_d \geq 0} + PZ\mathbb{1}_{Z_d < 0}$ with $Z = (Z_1, \dots, Z_d)'$, $P = I_d - J^{-1}K'\left(KJ^{-1}K'\right)^{-1}K$. It follows that

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

where $Z_d^- = Z_d\mathbb{1}_{Z_d<0}$, and $\mathbf{c} = E(Z_dZ)/\mathrm{Var}(Z_d)$ is the last column of J^{-1} divided by the (d, d)-element of this matrix. Note that the last component of $\lambda^{\Lambda} = (\lambda^{\Lambda}_1, \dots, \lambda^{\Lambda}_d)'$ is $\lambda^{\Lambda}_d = Z_d^+ := Z_d\mathbb{1}_{Z_d>0}$. It is also seen that $\lambda^{\Lambda}_i = Z_i$ if and only if $\mathrm{Cov}(Z_i, Z_d) = 0$.

In view of Proposition 5.2, it follows that

$$\mathbf{W}(0) = \frac{2}{\kappa_\eta - 1}\mathbf{L}(0) = \frac{\left\{\lambda^{\Lambda}_d\right\}^2}{\mathrm{Var}Z_d} = U^2\mathbb{1}_{U\geq 0} \sim \frac{1}{2}\delta_0 + \frac{1}{2}\chi^2_1$$

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).

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

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

where $U \sim \mathcal{N}(0,1)$. Equalities (6.2) and (6.3) follow.

(c) Note that (6.2) 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 (6.3) 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.6 Proof of Proposition 6.2.

In view of (6.2) and (6.5), 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

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

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 (6.6). Note that when f is defined by (6.6), 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 $\eta_t \sim \mathcal{N}(0,1)$.

A.7 Proof of Proposition 7.1.

Under (7.1), thorough inspection of the proof (given in FZ) shows that Theorem 3.1 holds without the moment assumption in **A5** (and without **A6** which does not make sense in the ARCH case). In particular we have, for some constant C,

$$\begin{aligned}E_{\theta_0} \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\| \frac{\partial^2 \ell_t(\theta)}{\partial\theta\partial\theta'} \right\| &= E_{\theta_0} \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\| \left\{ 2\frac{\epsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta'} \right\} \right\| \\ &\leq E_{\theta_0} \left\| C(1+\epsilon_t^2) \right\| E \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \left\| \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta} \right\} \left\{ \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta'} \right\} \right\| < \infty,\end{aligned}$$

where the first inequality follows from the independence between ϵ_t and σ_t^2 and its derivative under (7.1), and the second inequality follows from $E(\epsilon_t^4) = \omega^2 E(\eta_t^4) < \infty$.

In view of (3.2), 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 $\mathcal{N}(0,1)$. The asymptotic null distribution of $\mathbf{W}_n^*$ follows.

A.8 Proof of Proposition 7.2.

By arguments used in the proof of Proposition 5.1, $\log\{1-\Phi(\mathbf{LK}_n)\} \sim -\mathbf{LK}_n^2/2$. Moreover

$$\frac{\mathbf{LK}_n^2}{n} \;\to\; \frac{1}{q}\left(\sum_{i=1}^q \rho_{\epsilon^2}(i)\right)^2, \quad a.s. \tag{A.17}$$

Similarly, in view of (7.3)

$$\frac{\mathbf{R}_n^*}{n} \to \sum_{i=1}^{q} \rho_{\epsilon^2}^2(i), \quad a.s. \tag{A.18}$$

The expressions for the asymptotic efficiencies follow. Using $(q^{-1}\sum_{i=1}^q a_i)^2 \leq q^{-1}\sum_{i=1}^q a_i^2$, for any real numbers a_i, we then have $\mathbf{ARE}(\mathbf{R}^*/\mathbf{LK}) \geq 1$, with equality when $q = 1$. To show that $\mathbf{ARE}(\mathbf{R}^*/\mathbf{W}^*) \geq 1$ note that, because (ϵ_t^2) has an AR(q) representation under H_1, and because $\rho_{\epsilon^2}(i) \geq 0$, for $i = 1, \dots, q$,

$$\rho_{\epsilon^2}(i) = \alpha_1\rho_{\epsilon^2}(i-1) + \cdots + \alpha_{i-1}\rho_{\epsilon^2}(1) + \alpha_i + \alpha_{i+1}\rho_{\epsilon^2}(1) + \cdots + \alpha_q\rho_{\epsilon^2}(q-i) \geq \alpha_i,$$

with equality when $q = 1$. The conclusion directly follows.

Finally, introducing the linear innovation $\nu_t = (\eta_t^2 - 1)\sigma_t^2(\theta_0)$ of ϵ_t^2 under the alternative, we have

$$\frac{\mathbf{R}_n}{n} \to 1 - \frac{\mathrm{Var}(\nu_t)}{\mathrm{Var}(\epsilon_t^2)} = \frac{\kappa_\epsilon - \kappa_\eta}{\kappa_\eta(\kappa_\epsilon - 1)}, \quad a.s. \tag{A.19}$$

The desired inequality $\mathbf{ARE}(\mathbf{R}/\mathbf{W}^*) \geq 1$ is equivalent to $\kappa_\epsilon(\kappa_\eta^{-1} - \sum_{i=1}^q \alpha_i^2) \geq 1 - \sum_{i=1}^q \alpha_i^2$. On the other hand, straight computation of $E\sigma_t^4$ yields, using again $\rho_{\epsilon^2}(i) \geq 0$,

$$\kappa_\epsilon(\kappa_\eta^{-1} - \sum_{i=1}^q \alpha_i^2) = 1 - \left(\sum_{i=1}^q \alpha_i\right)^2 + 2\sum_{i<j}\alpha_i\alpha_j\frac{E(\epsilon_{t-i}^2\epsilon_{t-j}^2)}{(E\epsilon_t^2)^2} \geq 1 - \sum_{i=1}^q \alpha_i^2.$$

A.9 Proof of Proposition 7.3.

Under H_{a1} (resp. H_{a2}) (A.18) (resp. (A.17)) follows from the ergodic theorem, which proves that $\mathbf{R}_n^*$ (resp. $\mathbf{LK}_n$)is consistent.

Similarly, the convergence in (A.19) holds, where $\nu_t = \epsilon_t^2 - EL(\epsilon_t^2 \mid \epsilon_{t-i}^2, i = 1, \dots, q)$. Under H_a, $\mathrm{Var}(\nu_t) < \mathrm{Var}(\epsilon_t)$, which proves that $\mathbf{R}_n$ is consistent.

To handle $\mathbf{W}_n^*$, we note that the proof of (i)-(iv) in Francq and Zakoian (2004), given for the case of an iid noise sequence, remains valid under H_{a3} with a slight adaptation concerning the identifiability step. Suppose that $\sigma_t^2(\theta_0) = \sigma_t^2(\theta)$ with $\theta \neq \theta_0$. By stationarity, it follows that ϵ_t^2 is a function of the $\epsilon_{t-i}^2, i = 1, \dots, q-1$. Thus

$$\epsilon_t^2 - E(\epsilon_t^2 \mid \epsilon_{t-i}^2, i = 1, \dots, q) = \sigma_t^2(\theta_0)(\eta_t^2 - 1) = 0.$$

It follows that $\eta_t^2 = 1, a.s.$ which is in contradiction with H_{a3}. Thus $\theta = \theta_0$ and the identifiability step is proved. The consistency of the QMLE follows and therefore

$$\frac{\mathbf{W}_n^*}{n} \to \sum_{i=1}^{q} \alpha_{0i}^2 > 0, \quad a.s.$$

which establish the consistency of the test. Note that Escanciano (2007) proves the consistency and asymptotic normality of the QMLE of GARCH models with a martingale difference noise sequence. For the consistency of the $\mathbf{W}_n^*$ test we only need the weaker assumption H_{a3}.

Complementary results

A Proof of (2.3)

We only prove the second equality, the first one being obtained by the same arguments. Recall that $\hat{\theta}_{n|2}$ minimizes

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

under the constraint $\theta^{(2)} = 0$. For any $c > 0$, there exists $\hat{\theta}^*_{n|2}$ such that $\tilde{\sigma}_t^2(\hat{\theta}^*_{n|2}) = c\tilde{\sigma}_t^2(\hat{\theta}_{n|2})$ for all $t \geq 0$. Note that $\hat{\theta}^*_{n|2} \neq \hat{\theta}_{n|2}$ iff $c \neq 1$. For instance, for the GARCH(1,2) constrained by $\theta^{(2)} = \beta_2 = 0$, if $\hat{\theta}_{n|2} = (\hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, 0)$ then $\hat{\theta}^*_{n|2} = (c\hat{\omega}, c\hat{\alpha}_1, \hat{\beta}_1, 0)$. Let $f(c) = \tilde{\mathbf{l}}_n(\hat{\theta}^*_{n|2})$. The minimum of f is obtained at only one point given by

$$c = n^{-1}\sum_{t=1}^{n} \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\hat{\theta}_{n|2})}.$$

Thus, for this value c, we have $\hat{\theta}^*_{n|2} = \hat{\theta}_{n|2}$. Hence $c = 1$ with probability 1, which is the announced result.

B Proof of (3.2)

To avoid unnecessary computations, we only prove this formula in the case $q = 2$. Let $\theta_0 = (\omega_0, 0, 0)$. We have $d_2 = 2$, $d_1 = 1$ and

$$\Lambda = \mathbb{R} \times (0,\infty)^2, \qquad K = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathcal{K} = \{K_1, K_2, K_3\},$$

where $K_1 = K$, $K_2 = (0,1,0)$ and $K_3 = (0,0,1)$. We have

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim \mathcal{N}\left\{0, \Sigma = (\kappa_\eta - 1)J^{-1} = \begin{pmatrix} (\kappa_\eta + 1)\omega_0^2 & -\omega_0 & -\omega_0 \\ -\omega_0 & 1 & 0 \\ -\omega_0 & 0 & 1 \end{pmatrix}\right\}.$$

Thus, using $K\Sigma K' = I_2$ and $K_i \Sigma K_i' = 1$ for $i = 2, 3$, we get

$$\begin{aligned} P_1 Z &= (Z_1 + \omega_0(Z_2 + Z_3), 0, 0)', \\ P_2 Z &= (Z_1 + \omega_0 Z_2, 0, Z_3)', \\ P_3 Z &= (Z_1 + \omega_0 Z_3, Z_2, 0)'. \end{aligned}$$

Let also $P_0 = I_3$. We have

$$\|P_i Z - Z\|_J^2 = (\kappa_\eta - 1)\begin{cases} 0 & \text{for} \quad i = 0 \\ Z_2^2 + Z_3^2 & \text{for} \quad i = 1 \\ Z_2^2 & \text{for} \quad i = 2 \\ Z_3^2 & \text{for} \quad i = 3 \end{cases}$$

This shows that

$$\lambda^\Lambda = \left(Z_1 + \omega_0 Z_2^- + \omega_0 Z_3^-, Z_2^+, Z_3^+\right)'.$$

C Proof of (A.12)

In view of (3.1),

$$\begin{aligned}\log S_{\mathbf{W}}(x) &= \log \mathbb{P}(\|\lambda^{\Lambda}\|_{\Omega}^2 > x) \\ &= \log \mathbb{P}(\|Z\|_{\Omega}^2 \mathbb{1}_{\Lambda}(Z) + \sum_{i=1}^{2^{d_2}-1} \|P_i Z\|_{\Omega}^2 \mathbb{1}_{\mathcal{D}_i}(Z) > x).\end{aligned}$$

Because $\|P_i Z\|_{\Omega}^2 \leq \|Z\|_{\Omega}^2$ we have

$$\log S_{\mathbf{W}}(x) \leq \log \mathbb{P}(\|Z\|_{\Omega}^2 > x) = \log \mathbb{P}(\|Z^{(2)}\|_{\{\mathrm{var}(Z^{(2)})\}^{-1}}^2 > x) = \log \mathbb{P}(\chi_{d_2}^2 > x).$$

Moreover, letting $U = \mathrm{var}^{-1/2}(Z^{(2)})Z^{(2)}$, which follows the $\mathcal{N}(0, I_{d_2})$ distribution, we have

$$\begin{aligned}\log S_{\mathbf{W}}(x) &\geq \log \mathbb{P}(\|Z\|_{\Omega}^2 \mathbb{1}_{\Lambda}(Z) > x) \\ &= \log \mathbb{P}(\|Z^{(2)}\|_{\{\mathrm{var}(Z^{(2)})\}^{-1}}^2 \mathbb{1}_{Z^{(2)} \geq 0} > x) \\ &= \log \mathbb{P}(\|U\|^2 \mathbb{1}_{U \in \mathcal{C}} > x) \\ &= \log\{\mathbb{P}(\|U\|^2 > x)\mathbb{P}(U \in \mathcal{C})\}\end{aligned}$$

for the cone $\mathcal{C} = \{u \in \mathbb{R}^{d_2} : \mathrm{var}^{1/2}(Z^{(2)})u > 0\}$. (A.12) follows.

D Proof of (A.13)

In Francq and Zakoian (2004), it is shown that $J(\theta_0)$ is positive definite. The same proof can be conducted for $\theta \neq \theta_0$.

E Proof of Theorem 4.1.

Throughout, all expectations are taken with respect to the distribution of (η_t). Let C and ρ be generic constants, whose values will be modified along the proofs, such that $C > 0$ and $0 < \rho < 1$.

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

$$\epsilon_{t,n_0}^2 \geq \epsilon_{t,n}^2 \geq \epsilon_t^2, \quad \text{and} \quad \sigma_{t,n_0}^2(\theta) \geq \sigma_{t,n}^2(\theta^*) \geq \sigma_t^2(\tilde{\theta}) \quad \text{for any} \quad \theta \geq \theta^* \geq \tilde{\theta}. \tag{E.1}$$

E.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) in FZ we have $\sigma^2_{t,n}(\theta)=\sum_{k=0}^{\infty}B^k(1,1)c_{t-k,n}$, where $c_{t,n}=\omega+\sum_{i=1}^{q}\alpha_i\epsilon^2_{t-i,n}$ and

$$
B=\begin{pmatrix}\beta_1 & \beta_2 & \cdots & \beta_p\\ 1 & 0 & \cdots & 0\\ \vdots & & & \\ 0 & \cdots & 1 & 0\end{pmatrix}.
$$

Let $\tilde{c}_{t,n}$ be obtained by replacing $\epsilon^2_{0,n},\dots,\epsilon^2_{1-q,n}$ by their initial values in $c_{t,n}$. We have

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

Thus, almost surely,

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

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

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

The a.s. convergence of $n^{-1}\sum_{t=1}^{n}\rho^t\epsilon^2_{t,n_0}$ 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

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

In the right-hand side of the last equality, the first sample mean converges to 1, a.s., and the second one is between $\frac{1}{n}\sum_{t=1}^{n}\log\sigma^2_t$ and $\frac{1}{n}\sum_{t=1}^{n}\log\sigma^2_{t,n_0}$. 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 (E.1), 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)[2] 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.

E.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

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_{t,n}(\theta_n)\xrightarrow{d}\mathcal{N}\left(0,(\kappa_\eta-1)J\right), \tag{E.3}$$

and

$$n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_{t,n}(\theta_{ij}^*)\xrightarrow{P}J(i,j), \tag{E.4}$$

[2] *Probability and Measure.* John Wiley, New York.

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

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

$$b)\ n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ell_{t,n}(\theta_n) \xrightarrow{d} \mathcal{N}\left(0, (\kappa_\eta - 1)J\right), \tag{E.5}$$

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

$$d)\ \left\|n^{-1/2}\sum_{t=1}^{n}\left\{\frac{\partial \ell_{t,n}(\theta_n)}{\partial\theta} - \frac{\partial \tilde{\ell}_{t,n}(\theta_n)}{\partial\theta}\right\}\right\| \to 0 \ \text{ and} \tag{E.6}$$

$$\sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta}\left\|n^{-1}\sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_{t,n}(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \tilde{\ell}_{t,n}(\theta)}{\partial\theta\partial\theta'}\right\}\right\| \xrightarrow{P} 0, \tag{E.7}$$

$$e)\ n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_{t,n}(\theta_n) \to J(i,j)\ a.s, \tag{E.8}$$

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

for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0. We begin to show that (E.3) and (E.4) follow from a)-f).

Proof of (E.3) and (E.4). The convergence (E.3) is a straightforward consequence of b) and the first part of d). To show (E.4) 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), in view of

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

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 η_{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

$$\frac{1}{n}\sum_{t=1}^{n} s_{t,n}^2 \xrightarrow{P} (\kappa_\eta - 1)\lambda' J\lambda, \quad \text{and} \tag{E.9}$$

$$\frac{1}{n}\sum_{t=1}^{n} E(x_{t,n}^2 \mathbb{1}_{|x_{t,n}|\geq n^{1/2}\varepsilon}) \to 0, \quad \text{when } n\to\infty, \tag{E.10}$$

for any $\varepsilon > 0$. First consider the derivative of $\sigma_{t,n}^2$ with respect to β_j. In view of (A.3)-(A.5) in FZ, we have

$$\sigma_{t,n}^2 = \sum_{k=0}^{\infty} B_n^k(1,1)\left(\omega_n + \sum_{i=1}^{q}\alpha_{i,n}\epsilon_{t-k-i,n}^2\right), \quad \frac{\partial\sigma_{t,n}^2}{\partial\beta_j} = \sum_{k=1}^{\infty} B_{k,j;n}(1,1)\left(\omega_n + \sum_{i=1}^{q}\alpha_{i,n}\epsilon_{t-k-i,n}^2\right),$$

where B_n (resp. $B_{k,j;n}$) is the matrix obtained from B (resp. $B_{k,j}$) by replacing the coefficients β_i by $\beta_{i,n}$. Denote by ${}_j\sigma_{t,n}^2$ (resp. ${}^j\sigma_{t,n}^2$) the variable obtained by replacing $\epsilon_{t-j,n}^2$ by ϵ_{t-j,n_0}^2 (resp. ϵ_{t-j}^2) in the expansion of $\sigma_{t,n}^2$. Denote by ${}_j\sigma_t^2$ (resp. ${}^j\sigma_t^2$) the variable obtained by replacing the variables $\epsilon_{t-i,n}^2$ by ϵ_{t-i}^2 (resp. by ϵ_{t-i,n_0}^2, and ϵ_{t-j,n_0}^2 by ϵ_{t-j}^2) in ${}_j\sigma_{t,n}^2$, 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 $\sigma_{t,n}^2 = \frac{\omega_n}{1-\beta_n} + \alpha_n\sum_{i\geq 1}\beta_n^{i-1}\epsilon_{t-i,n}^2$, ${}_j\sigma_{t,n}^2 = \frac{\omega_n}{1-\beta_n} + \alpha_n\beta_n^{j-1}\epsilon_{t-j,n_0}^2 + \alpha_n\sum_{i\geq 1, i\neq j}\beta_n^{i-1}\epsilon_{t-i,n}^2$ and ${}^j\sigma_{t,n}^2 = \frac{\omega_n}{1-\beta_n} + \alpha_n\beta_n^{j-1}\epsilon_{t-j}^2 + \alpha_n\sum_{i\geq 1,i\neq j}\beta_n^{i-1}\epsilon_{t-i,n}^2$, whereas ${}_j\sigma_t^2 = \frac{\omega_0}{1-\beta_0} + \alpha_0\beta_0^{j-1}\epsilon_{t-j,n_0}^2 + \alpha_0\sum_{i\geq 1,i\neq j}\beta_0^{i-1}\epsilon_{t-i}^2$ and ${}^j\sigma_t^2 = \frac{\omega_{n_0}}{1-\beta_{n_0}} + \alpha_{n_0}\beta_{n_0}^{j-1}\epsilon_{t-j}^2 + \alpha_{n_0}\sum_{i\geq 1,i\neq j}\beta_{n_0}^{i-1}\epsilon_{t-i,n_0}^2$. Notice that for any constants $a>0$ and $b>0$, the function $x\to x/(a+bx)$ is increasing over the positive line. Considering $\sigma_{t,n}^2$ as a function of ϵ_{t-j}^2, for $j>0$, it follows that, using (E.1),

$$\frac{\epsilon_{t-j}^2}{{}^j\sigma_{t,n}^2} \leq \frac{\epsilon_{t-j,n}^2}{\sigma_{t,n}^2} \leq \frac{\epsilon_{t-j,n_0}^2}{{}_j\sigma_{t,n}^2}.$$

We also have, from (A.5) in FZ,

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

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

$$\begin{aligned}\frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\beta_j} &\leq \sum_{k=1}^{\infty} B_{k,j;n}(1,1)\left(\omega_n + \sum_{i=1}^{q}\alpha_{i,n}\frac{\epsilon_{t-k-i,n_0}^2}{{}_{k+i}\sigma_{t,n}^2}\right)\\ &\leq \sum_{k=1}^{\infty} B_{k,j;n_0}(1,1)\left(\omega_{n_0} + \sum_{i=1}^{q}\alpha_{i,n_0}\frac{\epsilon_{t-k-i,n_0}^2}{{}_{k+i}\sigma_t^2}\right).\end{aligned} \tag{E.11}$$

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,

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

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

$$Y_t^{(1)}(n_0) \leq \frac{1}{\sigma_{t,n}^2}\frac{\partial\sigma_{t,n}^2}{\partial\theta} \leq Y_t^{(2)}(n_0) \tag{E.12}$$

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

$$Y_t^{(1)}(n_0)Y_t^{(1)}(n_0)' \leq \frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'} \leq Y_t^{(2)}(n_0)Y_t^{(2)}(n_0)', \tag{E.13}$$

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 8 in FZ. More precisely, if **A8** holds true, proceeding as in the calculations leading to (A.16) in FZ, we obtain an upper bound for the right-hand side of (E.11) as

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

where $Y_t^{(2)}(n_0) = (Y_{it}^{(2)}(n_0))_{1\leq i\leq p+q+1}$, for some positive constant α 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 (E.13) 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 (E.9) straightforwardly follows. To prove (E.10) 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 (E.10) follows and the proof of b) is complete.

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

$$\frac{\partial^2\ell_{t,n}}{\partial\theta\partial\theta'} = \left\{1-\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}\right\}\frac{1}{\sigma_{t,n}^2}\frac{\partial^2\sigma_{t,n}^2}{\partial\theta\partial\theta'} + \left\{2\frac{\epsilon_{t,n}^2}{\sigma_{t,n}^2}-1\right\}\frac{1}{\sigma_{t,n}^4}\frac{\partial\sigma_{t,n}^2}{\partial\theta}\frac{\partial\sigma_{t,n}^2}{\partial\theta'}. \tag{E.14}$$

First we will show that a formula similar to (E.12) 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 (E.11)

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

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

$$0 \leq \sup_{n\geq n_0} \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \frac{1}{\sigma^2_{t,n}} \frac{\partial \sigma^2_{t,n}}{\partial\theta}(\theta) \ \leq \ Y^{(3)}_t(n_0), \tag{E.15}$$

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

$$0 \leq \sup_{n\geq n_0} \sup_{\theta\in\mathcal{V}(\theta_0)\cap\Theta} \frac{1}{\sigma^2_{t,n}} \frac{\partial^2 \sigma^2_{t,n}}{\partial\theta_i\partial\theta_j}(\theta) \ \leq \ Y^{(4)}_{i,j,t}(n_0), \tag{E.16}$$

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

To handle terms of (E.14) involving

$$\frac{\epsilon^2_{t,n}}{\sigma^2_{t,n}(\theta)} = \eta^2_t \frac{\sigma^2_{t,n}(\theta_n)}{\sigma^2_{t,n}(\theta)},$$

we will use the expansion $\sigma^2_{t,n}(\theta) = c + \sum_{j=1}^{\infty} b_j\epsilon^2_{t-j,n}$ 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 $ax/(b+cx) \leq ax^s/(b^sc^{1-s})$ for all $a,b,c,x \geq 0$ and any $s \in (0,1)$, we obtain, for $\mathcal{V}(\theta_0)$ sufficiently small

$$\frac{\sigma^2_{t,n}(\theta_n)}{\sigma^2_{t,n}(\theta)} \leq C + C\sum_{j=1}^{\infty} \frac{b_{j,n_0}}{b_j} b^s_j \epsilon^s_{t-j,n_0} \leq C + C\sum_{j=1}^{\infty}(1+\delta)^j\rho^{js}\epsilon^s_{t-j,n_0}, \tag{E.17}$$

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 $E\epsilon^{2s}_{t,n_0} < \infty$ and, for instance, $\delta = (1-\rho^s)/(2\rho^s)$ we obtain

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

For the same choice of δ, with s such that $E\epsilon^{4s}_t < \infty$, and using (E.17), we find

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

Using (E.14), (E.15), (E.16), (E.18) and the Schwarz inequality, it is straightforward to conclude that c) holds.

To prove d) first note that, analogue to (E.2), we have almost surely

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

where C does not depend on n. It follows that

$$\left| n^{-1/2} \sum_{t=1}^{n} \left\{ \frac{\partial \ell_{t,n}(\theta_n)}{\partial \theta_i} - \frac{\partial \tilde{\ell}_{t,n}(\theta_n)}{\partial \theta_i} \right\} \right| \leq C^* n^{-1/2} \sum_{t=1}^{n} \rho^t (1+\eta_t^2) \left\{ 1 + \frac{1}{\sigma_{t,n}^2} \frac{\partial \sigma_{t,n}^2}{\partial \theta_i} \right\},$$

$$\leq C^* n^{-1/2} \sum_{t=1}^{n} \rho^t (1+\eta_t^2) \left\{ 1 + Y_{it}^{(2)}(n_0) \right\},$$

where $Y_{it}^{(2)}(n_0)$ is the i-th component of $Y_t^{(2)}(n_0)$ introduced in (E.12). 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 $Cn^{-1} \sum_{t=1}^{n} \rho^t \Upsilon_{t,n}$, 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

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

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 $E\Upsilon_{t,n}^s < C$. 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 (E.14), at the value θ_n. In view of (E.13), we have

$$n^{-1} \sum_{t=1}^{n} (2\eta_t^2 - 1) \frac{1}{\sigma_{t,n}^4} \frac{\partial \sigma_{t,n}^2}{\partial \theta} \frac{\partial \sigma_{t,n}^2}{\partial \theta'} \leq n^{-1} \sum_{t=1}^{n} (2\eta_t^2 - 1) \mathbb{1}_{2\eta_t^2 \geq 1} Y_t^{(2)}(n_0) Y_t^{(2)}(n_0)'$$

$$+ n^{-1} \sum_{t=1}^{n} (2\eta_t^2 - 1) \mathbb{1}_{2\eta_t^2 < 1} Y_t^{(1)}(n_0) Y_t^{(1)}(n_0)'.$$

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

$$\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} (2\eta_t^2 - 1) \frac{1}{\sigma_{t,n}^4} \frac{\partial \sigma_{t,n}^2}{\partial \theta} \frac{\partial \sigma_{t,n}^2}{\partial \theta'} \leq E\{(2\eta_t^2 - 1) \mathbb{1}_{2\eta_t^2 \geq 1}\} E\{Y_t^{(2)}(n_0) Y_t^{(2)}(n_0)'\}$$

$$+ E\{(2\eta_t^2 - 1) \mathbb{1}_{2\eta_t^2 < 1}\} E\{Y_t^{(1)}(n_0) Y_t^{(1)}(n_0)'\}$$

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

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

Similarly we have

$$\liminf_{n\to\infty} n^{-1} \sum_{t=1}^{n} (2\eta_t^2 - 1) \frac{1}{\sigma_{t,n}^4} \frac{\partial \sigma_{t,n}^2}{\partial \theta} \frac{\partial \sigma_{t,n}^2}{\partial \theta'} \geq E\{(2\eta_t^2 - 1) \mathbb{1}_{2\eta_t^2 \geq 1}\} E\{Y_t^{(1)}(n_0) Y_t^{(1)}(n_0)'\}$$

$$+ E\{(2\eta_t^2 - 1) \mathbb{1}_{2\eta_t^2 < 1}\} E\{Y_t^{(2)}(n_0) Y_t^{(2)}(n_0)'\},$$

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 (E.14) 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 (E.15) and (E.16), which allows to conclude.

E.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 (E.4), 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.33) in FZ, we have the following quadratic expansion of the quasi-likelihood function around θ_n

$$\tilde{\mathbf{l}}_n(\theta) \;=\; \tilde{\mathbf{l}}_n(\theta_n) + \frac{1}{2n} \|Z_{n,\tau} - \sqrt{n}(\theta - \theta_n)\|_{J_{n,\tau}}^2 - \frac{1}{2n} Z_{n,\tau}' J_{n,\tau} Z_{n,\tau} + R_n(\theta), \qquad \text{(E.18)}$$

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

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

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

(vi) $\lambda_{n,\tau}^{\Lambda} \xrightarrow{d} \lambda^{\Lambda}(\tau)$.

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 $J_{n,\tau} \xrightarrow{P} J$ by (E.8), and $\sqrt{n}\frac{\partial \mathbf{l}_n(\theta_n)}{\partial\theta} = O_P(1)$ by (E.5).

The remainder term in (E.18) satisfies

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

for some θ^*_{ij} between θ and θ_n. By (E.4) and the second part of (E.6), 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 (E.6). 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 $\sqrt{n}(\theta_{J_{n,\tau}}(Z_{n,\tau})-\theta_n) = \lambda^{\Lambda}_{n,\tau}$ for n sufficiently large, to that of (iv).

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

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

Hence, (v) follows from (iv) and

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

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

F Proof of (A.15)

The proof is similar to that of (2.3). Note that $\theta_{0|2}$ minimizes

$$E_{\theta_0}\left(\frac{\epsilon_t^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta)\right) = E_{\theta_0}\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta)\right)$$

under the constraint $\theta^{(2)}=0$. For any $c>0$, there exists $\theta^*_{0|2}$ such that $\sigma_t^2(\theta^*_{0|2}) = c\sigma_t^2(\theta_{0|2})$ for all $t\geq 0$. Note that $\theta^*_{0|2}\neq\theta_{0|2}$ iff $c\neq 1$. Let

$$f(c) = E_{\theta_0}\left(\frac{\sigma_t^2(\theta_0)}{c\sigma_t^2(\theta_{0|2})} + \log c\sigma_t^2(\theta_{0|2})\right).$$

The minimum of f is obtained at a unique point, given by

$$c = E_{\theta_0}\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_{0|2})}\right).$$

Thus, for this value c, we have $\theta^*_{0|2} = \theta_{0|2}$. Hence $c=1$, which is the announced result.

G Proof that $\rho_{\epsilon^2}(h) > 0$ (used in the proof of Proposition 7.2)

It suffices to show that we have a MA(∞) of the form

$$\epsilon_t^2 = c + \nu_t + \sum_{\ell=1}^{\infty} \phi_\ell \nu_{t-\ell}, \quad \text{with} \quad \phi_\ell \geq 0 \quad \forall \ell.$$

Indeed, $\nu_t := \epsilon_t^2 - \sigma_t^2 = (\eta_t^2 - 1)\sigma_t^2$ being a white noise, we have

$$\gamma_{\epsilon^2}(h) = E\nu_1^2 \sum_{\ell=0}^{\infty} \phi_\ell \phi_{\ell+|h|}, \qquad \text{with the notation } \phi_0 = 1.$$

Denoting by B the backshift operator, and introducing the notation $\alpha(z) = \sum_{i=1}^{q} \alpha_i z^i$, $\beta(z) = \sum_{j=1}^{p} \beta_j z^j$ and $\phi(z) = \sum_{\ell=1}^{\infty} \phi_\ell z^\ell$, we obtain

$$\epsilon_t^2 = \{1 - (\alpha+\beta)(1)\}^{-1}\omega + \{1 - (\alpha+\beta)(B)\}^{-1}(1 - \beta(B))\nu_t = c + \phi(B)\nu_t.$$

Since $1 - \beta(B) = 1 - (\alpha+\beta)(B) + \alpha(B)$, we obtain ϕ_ℓ as the coefficient of z^ℓ in the division of $\alpha(z)$ by $1 - (\alpha+\beta)(z)$ according to the increasing powers of z. By recurrence on ℓ, it is easy to see that these coefficients are positive because the polynomials $\alpha(z)$ and $(\alpha+\beta)(z)$ have positive coefficients.

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

Bahadur, R. R. (1960) Stochastic comparison of tests. *The Annals of Mathematical Statistics* 31, 276–295.

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

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

Carrasco, M. and X. Chen (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory*, **18**, 17-39.

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.

Francq, C., and J-M. Zakoïan (2006) Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory*, 22, 815-834.

Francq, C., and J. M. Zakoïan (2007) Quasi-Maximum Likelihood Estimation in GARCH Processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*, 117, 1265-1284.

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.

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.

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. 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. (2007) Self-weighted LSE and MLE for ARMA-GARCH models. *Journal of Econometrics* 140, 849–873.

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

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.

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

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–26.

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.