MANCHESTER 1824

The University of Manchester

Economics Discussion Paper Series EDP-0721

First order asymptotic theory for parametric misspecification tests of GARCH models

Andreea Halunga
Chris D. Orme

October 2007

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

First Order Asymptotic Theory for Parametric Misspecification Tests of GARCH Models*

Andreea G. Halunga,†‡
School of Business and Economics, University of Exeter,
Streatham Court, Rennes Drive, Exeter EX4 4PU, UK.

Chris D. Orme,
Economics, School of Social Science, University of Manchester,
Manchester M13 9PL, UK.

October 5, 2007

Abstract

This paper develops a framework for the construction and analysis of parametric misspecification tests for GARCH models, based on standard first order asymptotic theory. Here, the GARCH model is defined to be a regression model in which the errors, under the null, are conditionally heteroskedastic according to a GARCH process and it is this latter assumption which is the subject of misspecification tests. The principal finding is that estimation effects from the correct specification of the conditional mean (regression) function can be asymptotically non-negligible. This implies that certain procedures, such as the asymmetry tests of Engle and Ng (1993) and the non-linearity test of Lundbergh and Teräsvirta (2002), are asymptotically invalid. A second contribution is the proposed use of alternative tests for asymmetry and/or non-linearity which, it is conjectured, should enjoy improved power properties. A Monte Carlo study supports the principal theoretical findings and also suggests that the new tests have fairly good size and very good power properties, when compared with tests of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002).

JEL Classification: C12, C22

1 Introduction

A great deal of research has been undertaken on modelling volatility clustering in financial and economic time series, in which the GARCH model of Bollerslev (1986) represents a benchmark specification. The subsequent literature

*We are grateful for the insightful comments of three referees, Alastair Hall and Len Gill, which greatly improved the exposition of this paper. The standard disclaimer applies.

†The first author's research is part of her PhD thesis and was supported by a University Research Studentship and an Overseas Research Studentship; both of which are gratefully acknowledged.

‡Corresponding author. *Email*: a.g.halunga@exeter.ac.uk

has provided generalisations by, for example, allowing for possible asymmetric and/or non-linear behaviour. Prominent among these are: the EGARCH model of Nelson (1990); the GJR model of Glosten, Jagannathan and Runkle (1993); the TGARCH model of Zakoian (1994); and, the Smooth Transition GARCH (STGARCH) model of Hagerud (1997) and Gonzalez-Rivera (1998).

Notwithstanding these developments, the parametric GARCH model remains a popular choice among applied workers. Therefore, and as noted by Lundbergh and Teräsvirta (2002), it is important to perform misspecification tests to assess the adequacy of the parametric model being employed. In this paper, the GARCH model is defined to be a regression model in which the errors, under the null, are conditionally heteroskedastic according to a GARCH process and the parametric tests are ones which might be used to detect misspecification in the assumed GARCH process for the errors. Developing misspecification tests has not been a neglected area of research. Bollerslev (1986) suggested a natural score type test for testing a GARCH model against a higher order GARCH model. Asymmetry tests were proposed by Engle and Ng (1993), and these are now widely used in empirical finance. Li and Mak (1994) constructed a test for the adequacy of a GARCH (p, q) model with a null hypothesis that the squared standardised error process is serially uncorrelated. Lundbergh and Teräsvirta (2002) proposed tests of (i) no remaining ARCH in standardised errors; (ii) linearity; and, (iii) parameter constancy. All these procedures are important inferential tools for empirical researchers who are interested in obtaining accurate forecasts of financial volatility, in order to take the appropriate decisions on portfolio selection, asset management or pricing derivative assets.

However, in this paper it is argued that, on closer inspection, the standard first order theory employed to justify the asymptotic validity of such procedures has sometimes been misinterpreted. To establish this, a unifying framework for the construction and analysis of parametric misspecification tests in GARCH models, based on the conditional moment principle and first order asymptotic analysis, is developed. This provides a useful contribution in at least two respects.

Firstly, and most significantly, the theory predicts that the limit null distribution of the relevant test indicators must take account of asymptotically non-negligible estimation effects which arise due to the estimated conditional mean (regression) parameters in the null GARCH(p, q) model. (The importance of estimation effects was addressed by Durbin, 1970, when testing for serial correlation with lagged dependent variables.) This issue has been, apparently, overlooked in the GARCH testing literature because in the null GARCH(p, q) model, under conditional symmetry of the errors, the estimated conditional mean parameters are asymptotically orthogonal to estimated conditional heteroskedasticity parameters. In particular, and because of this orthogonality, it appears that the conditional mean estimation effects have been simply (but erroneously) assumed away, for example, by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002).[1]

The second contribution proposes "new" tests for asymmetry and/or nonlinearity. It is conjectured that these test procedures should have better power properties against the types of alternative models considered by both Engle and

[1] Although, the issue of estimation effects from the conditional heteroskedasticity parameters has been ackowledged; see, for example, Li and Mak (1994) and Lundbergh and Teräsvirta (2002).

Ng (1993) and Lundbergh and Teräsvirta (2002) in their Monte Carlo experiments, since their construction takes into account the recursive nature of the conditional heteroskedasticity (whereas the test procedures of Engle and Ng, 1993, and Lundbergh and Teräsvirta, 2002, do not). The results of a small Monte Carlo study reveal that the new tests do indeed have good size properties and very good power, when compared with the tests of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002).

This paper is organized as follows, with supporting Lemmas and Propositions, together with all proofs, relegated to Appendices. Section 2 describes the null GARCH model, and briefly discusses Quasi Maximum Likelihood (QML) estimation. Section 3 describes a framework for constructing a particular class of parametric misspecification tests. In Section 4 the tests proposed by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993) are reviewed and new asymptotically valid tests for asymmetry and non-linearity are introduced. Section 5 presents some Monte Carlo evidence in support of the theoretical findings and Section 6 concludes.

2 The Null GARCH(p, q) Model

The regression model for the variable of interest, y_t, is defined as

$$y_t = m\left(\mathbf{w}_t; \boldsymbol{\varphi}_0\right) + \varepsilon_{0t}, \qquad t = 1, ..., T \tag{1}$$

where $\mathbf{w}_t = (\mathbf{y}'_{t-1}, \mathbf{z}'_t)$, $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-l})' \in \Re^{l+1}$, $\mathbf{z}_t = (z_{t1}, ..., z_{tk})' \in \Re^k$ are exogenous variables, $\boldsymbol{\varphi}_0 = (\varphi_{01}, .., \varphi_{0r})'$ is the true parameter vector and the conditional mean (regression) function, $m\left(\mathbf{w}_t; \boldsymbol{\varphi}_0\right)$, is possibly non-linear.[2] The error $\{\varepsilon_{0t}, \mathcal{F}_t\}$, where $\mathcal{F}_{t-1} = \sigma\left(\left(y_{t-1}, \mathbf{z}'_t\right), \left(y_{t-2}, \mathbf{z}'_{t-1}\right), ...\right)$, is a martingale difference sequence given by

$$\varepsilon_{0t} = \xi_t h_{0t}^{1/2} \tag{2}$$

where the standardised error process, ξ_t, is an *i.i.d.* sequence with mean zero and variance one and define $k_c = E\left[\xi_t^4\right]$ and $v_c = E\left[\xi_t^3\right]$, both finite constants. The conditional variance is specified as

$$\begin{aligned} h_{0t} &= \boldsymbol{\eta}'_0 \mathbf{s}_{0,t-1} \\ &= \alpha_{00} + A_0(L)\varepsilon_{0t}^2 + B_0(L)h_{0t} \end{aligned} \tag{3}$$

where $\mathbf{s}_{0,t-1} = \left(1, \varepsilon_{0,t-1}^2, ..., \varepsilon_{0,t-q}^2, h_{0,t-1}, .., h_{0,t-p}\right)'$, $\boldsymbol{\eta}_0 = \left(\alpha_{00}, \alpha_{01}, ..., \alpha_{0q}, \beta_{01}, .., \beta_{0p}\right)'$, $A_0(L) = \alpha_{01}L + ... + \alpha_{0q}L^q$, $B_0(L) = \beta_{01}L + ... + \beta_{0p}L^p$.

The above process is defined for the true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}'_0, \boldsymbol{\eta}'_0)'$ and, correspondingly, the model for the unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\eta}')'$, is defined as

$$\begin{aligned} y_t &= m\left(\mathbf{w}_t; \boldsymbol{\varphi}\right) + \varepsilon_t, \qquad t = 1, ..., T \\ h_t &= \boldsymbol{\eta}' \mathbf{s}_{t-1} \\ &= \alpha_0 + A(L)\varepsilon_t^2 + B(L)h_t \\ &= a_t + B(L)h_t \end{aligned} \tag{4}$$

[2] For example, Lundbergh and Teräsvirta (1999) proposed the STAR-GARCH model and the statistical properties of this model were investigated by Chan and McAleer (2002).

where $a_t = \alpha_0 + A(L)\varepsilon_t^2 = \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2$. The following assumptions ensure the identifiability, stationarity and ergodicity of the above process.

Assumptions A

1. The parameter space, $\mathbf{\Theta}$, is compact and $\boldsymbol{\theta}_0$ lies in the interior of $\mathbf{\Theta}$.
2. The elements of $(y_t, \mathbf{z}_t')$ are strictly stationary and ergodic; and, $m(\mathbf{w}_t; \boldsymbol{\varphi})$ is continuous and $\mathcal{F}_{t-1}$ measurable for all $\boldsymbol{\varphi} \in \mathbf{\Theta}$.
3. (i) all the roots of $1 - A(z) - B(z) = 0$ lie outside the unit circle;
(ii) the parameter space is constrained such that $0 < \lambda \leq \min\{\eta_l\} \leq \max\{\eta_l\} < \Lambda$, $l = 1, ..., p+q+1$, where λ and Λ are independent of $\boldsymbol{\theta}$;
(iii) the polynomials $A(z)$ and $1 - B(z)$ are coprimes.

As in Ling and McAleer (2003), A3(i) is a stationarity assumption imposed over the whole parameter space. Notice that, with A3(ii), this implies that roots of $1 - B(z) = 0$ lie outside the unit circle. Thus, in addition to A3(ii), which restricts the parameter space so that zero values in $\boldsymbol{\eta}$ are ruled out, $\sum_{j=1}^{p} \beta_j < 1$. These restrictions are also imposed on $\mathbf{\Theta}$ by Berkes, Horváth and Kokoszka (2003) and are employed here because they afford uniform convergence of second derivatives of the log-likelihood over $\mathbf{\Theta}$, removing the need for third derivatives, thus greatly simplifying the algebra required to justify the substantive contribution.

Given Assumption A3(i)(ii) the process for h_t has the following representation

$$h_t^{\infty} = (1 - B(L))^{-1} a_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

where $(1 - B(L))^{-1} = \sum_{i=0}^{\infty} \psi_i L^i$, with $\psi_0 = 1$, $\psi_i > 0$ and satisfying $\psi_i = \sum_{j=1}^{p} \beta_j \psi_{i-j}$, with $\psi_s = 0$, $s < 0$, $0 < \sum_{i=0}^{\infty} \psi_i = \left(1 - \sum_{j=1}^{p} \beta_j\right)^{-1} < \infty$. The coefficients, ψ_i, decay exponentially fast, and there exist constants $K > 0$ and $0 < \rho < 1$, independent of $\boldsymbol{\theta}$, such that $\psi_i \leq K\rho^i$. Then, as in Ling and McAleer (2003), but under (1), additional Assumption A2 and $h_{0t} = h_t^{\infty}(\boldsymbol{\theta}_0)$, it can be shown that $\{\varepsilon_{0t}, h_{0t}\}$ is strictly stationary and ergodic.

Asymptotic theory for GARCH models has been considered by several authors. For example, Ling and McAleer (2003) required that $E\left(\varepsilon_{0t}^6\right) < \infty$ to ensure asymptotic normality of the QML estimator in the ARMA-GARCH model. Furthermore, Chan and McAleer (2002, 2003) argued that the results in Ling and McAleer (2003) also hold for a STAR-GARCH model. Berkes et al (2003) established the consistency and asymptotic normality of the QML estimator, under weaker moment assumptions, in the pure GARCH model and Francq and Zakoian (2004) established consistency and asymptotic normality of the QML estimator in both a pure GARCH and ARMA-GARCH model under further weakened conditions; for example, in the pure GARCH model the parameter space can contain zero elements (although the true parameter can not) and the only moment condition required for consistency is $E\left[\xi_t^2\right] = 1$ and for asymptotic normality, $E\left[\xi_t^4\right] < \infty$. All these authors assume, as here, that the ξ_t are *i.i.d.* Therefore, whilst it is possible that the assumptions employed in this paper could be weakened, it should be noted that the regression specification in (1) is more general than that employed in the literature referred to above and the corresponding assumptions employed are, nonetheless, sufficient and (importantly) permit a relatively straightforward justification of the required first order asymptotic theory, without obfuscating the principal issue that is addressed in the paper. In practice, and following Weiss (1986),[3] the existence of moments is assumed when required as follows, where $\|.\|$ denotes the Euclidean norm:

Assumptions B

1. $E\left|\varepsilon_{0t}\right|^{4(1+s)} < \infty$ for some $s > 0$, and all t.
2. $E\left|m(\mathbf{w}_t; \boldsymbol{\varphi}) - m(\mathbf{w}_t; \boldsymbol{\varphi}_0)\right|^2 > 0$, for all $\boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$
3. $m\left(\mathbf{w}_t; \boldsymbol{\varphi}\right)$ is at least twice continuously differentiable in $\boldsymbol{\varphi}$, with, for all t
   (i) $\sup_{\boldsymbol{\theta}} \left|m(\mathbf{w}_t; \boldsymbol{\varphi})\right|^{4(1+s)} < B(\mathbf{w}_t)$, with $E\left[B(\mathbf{w}_t)\right] < \infty$, for some $s > 0$;
   (ii) $E\left[\sup_{\boldsymbol{\theta}} \left\|\varepsilon_t^r \frac{\partial m(\mathbf{w}_{t-i}; \boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}\right\|^2\right] < \infty$, $r = 0, 2$, and all $i \geq 0$;
   (iii) $E\left[\sup_{\boldsymbol{\theta}} \left\|\varepsilon_t^r \frac{\partial^2 m(\mathbf{w}_{t-i}; \boldsymbol{\varphi})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'}\right\|^2\right] < \infty$, $r = 0, 1$ and all $i \geq 0$.

2.1 Estimation Framework

The (average) quasi log-likelihood, conditional on available pre-sample values $\tilde{\mathbf{y}} = (y_0, ..., y_{1-l})'$, is (ignoring constants)

$$L_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad l_t(\boldsymbol{\theta}) = -\frac{1}{2}\left[\ln(h_t) + \frac{\varepsilon_t^2}{h_t}\right], \tag{5}$$

although the ensuing asymptotic analysis does not restrict ξ_t to be normally distributed; see Bollerslev (1986). Note that (5) is not only conditional on available pre-sample values, $\tilde{\mathbf{y}}$, from which ε_t, $t = 1, ..., T$, can be constructed, but also on $\tilde{\boldsymbol{\varepsilon}}_0 = \left(\varepsilon_0^2, ..., \varepsilon_{1-q}^2, h_0, ..., h_{1-p}\right)'$, from which h_t can be constructed using (4). However, ε_t and the process h_t, $t \leq 0$, are unobserved. In order to simplify the algebra and asymptotic theory, it is assumed (in addition) that pre-sample observations $\mathbf{w}_0, ..., \mathbf{w}_{1-q}$ are also available (entailing $y_{-l}, ..., y_{1-(l+q)}$), so that ε_{1-k}, $k = 1, ..., q$, can be constructed, and that $h_t = 0$ for all $t \leq 0$.[4] The simplifications in the analysis derive from the fact that, now, h_t can be expressed as $h_t = \sum_{i=0}^{t-1} \psi_i a_{t-i} = \sum_{i=0}^{t-1} \left\{B^*(L)\right\}^i a_{t-i}$, $t = 1, ..., T$, where $B^*(L) = \beta_1 + \beta_2 L + ... + \beta_p L^{p-1}$. (In practice, and for all inferential procedures described in this paper, a constant value can be chosen for $\tilde{\boldsymbol{\varepsilon}}_0$, in order to generate h_t, $t = 1, ..., T$.) The unknown parameters can be estimated jointly by QML estimation of (5). Throughout, the estimated parameter vector will be denoted $\hat{\boldsymbol{\theta}}' = \left(\hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\eta}}'\right)$.

The unobserved log-likelihood function, conditioning on the infinite history of all past observations $\left(\mathbf{w}_0', \mathbf{w}_{-1}', \mathbf{w}_{-2}', ...\right)'$ is $L_T^\infty(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} l_t^\infty(\boldsymbol{\theta}) \neq$

[3] Weiss (1986) established the asymptotic theory for the ARCH model allowing for exogenous variables in the conditional mean.

[4] Note, that this is not the same start-up scheme employed by either Ling and McAleer (2003), who choose $\tilde{\boldsymbol{\varepsilon}}_0 = \mathbf{0}$, Berkes et al (2003), or Francq and Zakoian (2004).

$L_T(\boldsymbol{\theta})$, with $l_t^\infty(\boldsymbol{\theta}) = -\frac{1}{2}\left(\ln(h_t^\infty) + \frac{\varepsilon_t^2}{h_t^\infty}\right)$ and score vector contributions of $\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta}) = \frac{\partial l_t^\infty(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, where $\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta}) = \left(\mathbf{d}_{\boldsymbol{\varphi}t}^{\infty\prime}(\boldsymbol{\theta}), \mathbf{d}_{\boldsymbol{\eta}t}^{\infty\prime}(\boldsymbol{\theta})\right)'$ in an obvious manner. Assuming $L_T^\infty(\boldsymbol{\theta})$ and $L_T(\boldsymbol{\theta})$ are both twice continuously differentiable in $\boldsymbol{\theta}$, define $\mathbf{d}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta})' = T^{-1}\sum_{t=1}^T \mathbf{d}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}), \mathbf{P}_{\boldsymbol{\theta\theta}T}(\boldsymbol{\theta}) = -T^{-1}\sum_{t=1}^T \frac{\partial \mathbf{d}_{\boldsymbol{\theta}t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$, and, correspondingly, $\mathbf{D}_{\boldsymbol{\theta}T}^\infty(\boldsymbol{\theta})$ and $\mathbf{P}_{\boldsymbol{\theta\theta}T}^\infty(\boldsymbol{\theta})$ in an obvious manner for the unobserved $L_T^\infty(\boldsymbol{\theta})$. By introducing the unobserved log-likelihood, the methodology of Ling and McAleer (2003), Berkes et al (2003) and Francq and Zakoian (2004) is followed whereby it is established that $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta})$ has exactly the same first order asymptotic properties as $\hat{\boldsymbol{\theta}}^\infty = \arg\max_{\boldsymbol{\theta}} L_T^\infty(\boldsymbol{\theta})$, with the asymptotic properties of the latter being fairly easy to verify.

In order to develop these arguments, it will be useful to illustrate, and distinguish between, the various unobserved and observed quantities associated with $L_T^\infty(\boldsymbol{\theta})$ and $L_T(\boldsymbol{\theta})$, respectively, based on the assumed initial start up values embodied in $\tilde{\boldsymbol{\varepsilon}}_0$. Specifically, the unobserved scores are $\mathbf{D}_{\boldsymbol{\varphi}T}^\infty(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^T \left\{\frac{\varepsilon_t \mathbf{f}_t}{h_t^\infty} + \frac{1}{2}\left(\frac{\varepsilon_t^2}{h_t^\infty} - 1\right)\mathbf{c}_t^\infty\right\}$, $\mathbf{D}_{\boldsymbol{\eta}T}^\infty(\boldsymbol{\theta}) = T^{-1}\frac{1}{2}\sum_{t=1}^T \left(\frac{\varepsilon_t^2}{h_t^\infty} - 1\right)\mathbf{x}_t^\infty$, where $\mathbf{f}_t = \frac{\partial m(\mathbf{w}_t; \boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}}$, and by exploiting the recursions $\frac{\partial h_t}{\partial \boldsymbol{\varphi}} = -2\sum_{k=1}^q \alpha_k \varepsilon_{t-k}\mathbf{f}_{t-k} + B(L)\frac{\partial h_t}{\partial \boldsymbol{\varphi}}$ and $\frac{\partial h_t}{\partial \boldsymbol{\eta}} = \mathbf{s}_{t-1} + B(L)\frac{\partial h_t}{\partial \boldsymbol{\eta}}$,

$$\mathbf{c}_t^\infty = \frac{1}{h_t^\infty}\frac{\partial h_t^\infty}{\partial \boldsymbol{\varphi}} = -2\frac{1}{h_t^\infty}\sum_{k=1}^q \alpha_k \left\{\sum_{i=0}^\infty \psi_i \varepsilon_{t-k-i}\mathbf{f}_{t-k-i}\right\}, \tag{6}$$

and

$$\mathbf{x}_t^\infty = \frac{1}{h_t^\infty}\frac{\partial h_t^\infty}{\partial \boldsymbol{\eta}} = \frac{1}{h_t^\infty}\sum_{i=0}^\infty \psi_i \mathbf{s}_{t-1-i}^\infty, \tag{7}$$

where $\mathbf{s}_{t-1}^{\infty\prime} = \left(1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}^\infty, .., h_{t-p}^\infty\right)'$. Given $\tilde{\boldsymbol{\varepsilon}}_0$, the corresponding observed score $\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta})$, associated with $L_T(\boldsymbol{\theta})$, can be expressed analogously but with

$$\begin{aligned}
\mathbf{c}_t &= -2\frac{1}{h_t}\sum_{k=1}^q \alpha_k \left\{\sum_{i=0}^{t-1} \psi_i \varepsilon_{t-i-k}\mathbf{f}_{t-i-k}\right\} = -2\frac{1}{h_t}\sum_{k=1}^q \alpha_k \left\{\sum_{i=0}^{t-1} \{B^*(L)\}^i \varepsilon_{t-i-k}\mathbf{f}_{t-i-k}\right\} \\
\mathbf{x}_t &= \frac{1}{h_t}\sum_{i=0}^{t-1} \psi_i \mathbf{s}_{t-1-i} = \frac{1}{h_t}\sum_{i=0}^{t-1} \{B^*(L)\}^i \mathbf{s}_{t-1-i}
\end{aligned}$$

replacing $\mathbf{c}_t^\infty$ and $\mathbf{x}_t^\infty$, respectively. For example in the GARCH$(1,1)$ case $h_t = \sum_{i=0}^{t-1}\beta_1^i\left\{\alpha_0 + \alpha_1\varepsilon_{t-i-1}^2\right\}$, $\mathbf{c}_t = -2h_t^{-1}\alpha_1\sum_{i=0}^{t-1}\beta_1^i\varepsilon_{t-1-i}\mathbf{f}_{t-1-i}$, whilst $\mathbf{x}_t = h_t^{-1}\sum_{i=0}^{t-1}\beta_1^i\mathbf{s}_{t-1-i}$, with $\mathbf{s}_{t-1}' = \left(1, \varepsilon_{t-1}^2, h_{t-1}\right)$ in this case. In practice, however, $\mathbf{c}_t$ and $\mathbf{x}_t$ can be constructed using the recursions for $\frac{\partial h_t}{\partial \boldsymbol{\varphi}}$ and $\frac{\partial h_t}{\partial \boldsymbol{\eta}}$, described above.

The consistency and asymptotic normality of the QMLE estimator $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta})$ is presented below, together with a consistent variance-covariance matrix estimator.

2.1.1 QML Estimation

The following Theorem establishes the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}$.

Theorem 1 *Given Assumptions A and B,* $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ *and*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{J}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}\mathbf{J}_{\boldsymbol{\theta\theta}}^{-1}\right)$$

where $\mathbf{J}_{\boldsymbol{\theta\theta}} = -E\left[\dfrac{\partial \mathbf{d}_{\boldsymbol{\theta}t}^{\infty}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}} = E\left[\mathbf{d}_{\boldsymbol{\theta}t}^{\infty}(\boldsymbol{\theta}_0)\mathbf{d}_{\boldsymbol{\theta}t}^{\infty}(\boldsymbol{\theta}_0)'\right]$ *are both finite and positive definite with*

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\varphi\varphi}} & \mathbf{J}_{\boldsymbol{\eta\varphi}}' \\ \mathbf{J}_{\boldsymbol{\eta\varphi}} & \mathbf{J}_{\boldsymbol{\eta\eta}} \end{bmatrix} = \tfrac{1}{2}E\begin{bmatrix} \mathbf{c}_t^{\infty}\mathbf{c}_t^{\infty\prime} & \mathbf{c}_t^{\infty}\mathbf{x}_t^{\infty\prime} \\ \mathbf{x}_t^{\infty}\mathbf{c}_t^{\infty\prime} & \mathbf{x}_t^{\infty}\mathbf{x}_t^{\infty\prime} \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + E\begin{bmatrix} \frac{1}{h_t^{\infty}}\mathbf{f}_t\mathbf{f}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and

$$\begin{aligned}
\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}} &= \begin{bmatrix} \boldsymbol{\Omega}_{\boldsymbol{\varphi\varphi}} & \boldsymbol{\Omega}_{\boldsymbol{\eta\varphi}}' \\ \boldsymbol{\Omega}_{\boldsymbol{\eta\varphi}} & \boldsymbol{\Omega}_{\boldsymbol{\eta\eta}} \end{bmatrix} \\
&= \tfrac{(k_c-1)}{4}E\begin{bmatrix} \mathbf{c}_t^{\infty}\mathbf{c}_t^{\infty\prime} & \mathbf{c}_t^{\infty}\mathbf{x}_t^{\infty\prime} \\ \mathbf{x}_t^{\infty}\mathbf{c}_t^{\infty\prime} & \mathbf{x}_t^{\infty}\mathbf{x}_t^{\infty\prime} \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\quad + \tfrac{v_c}{2}E\begin{bmatrix} \frac{1}{\sqrt{h_t^{\infty}}}\mathbf{f}_t\mathbf{c}_t^{\infty\prime} & \frac{1}{\sqrt{h_t^{\infty}}}\mathbf{f}_t\mathbf{x}_t^{\infty\prime} \\ \frac{1}{\sqrt{h_t^{\infty}}}\mathbf{x}_t^{\infty}\mathbf{f}_t' & \mathbf{0} \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + E\begin{bmatrix} \frac{1}{h_t^{\infty}}\mathbf{f}_t\mathbf{f}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.
\end{aligned}$$

Consistent standard errors follow from the next Lemma, in which $\hat{\mathbf{C}}$, $\hat{\mathbf{X}}$ and $\hat{\mathbf{F}}$ are matrices with rows $\hat{\mathbf{c}}_t'$, $\hat{\mathbf{x}}_t'$ and $\hat{\mathbf{f}}_t'$, respectively, and $\hat{\mathbf{H}} = diag\left(\hat{\boldsymbol{\eta}}'\hat{\mathbf{s}}_{t-1}\right)$, where "hats" denote evaluation at $\hat{\boldsymbol{\theta}}$.

Lemma 1 *Under Assumptions A and B,*

(a) $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\theta\theta}} - \boldsymbol{\Omega}_{\boldsymbol{\theta\theta}} = o_p(1)$, where

$$\begin{aligned}
\hat{\boldsymbol{\Omega}}_{\boldsymbol{\theta\theta}} &= \frac{\left(\hat{k}_c - 1\right)}{4}\frac{1}{T}\begin{bmatrix} \hat{\mathbf{C}}'\hat{\mathbf{C}} & \mathbf{C}'\hat{\mathbf{X}} \\ \hat{\mathbf{X}}'\hat{\mathbf{C}} & \hat{\mathbf{X}}'\hat{\mathbf{X}} \end{bmatrix} \\
&\quad + \frac{\hat{v}_c}{2}\frac{1}{T}\begin{bmatrix} \hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1/2}\hat{\mathbf{C}} & \hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1/2}\hat{\mathbf{X}} \\ \hat{\mathbf{X}}'\hat{\mathbf{H}}^{-1/2}\hat{\mathbf{F}} & \mathbf{0} \end{bmatrix} + \frac{1}{T}\begin{bmatrix} \hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},
\end{aligned}$$

where $\hat{k}_c - 1 = \frac{1}{T}\sum_{t=1}^{T}\left(\dfrac{\varepsilon_t^2}{h_t} - 1\right)^2_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\hat{v}_c = \frac{1}{T}\sum_{t=1}^{T}\left(\dfrac{\varepsilon_t}{\sqrt{h_t}}\right)^3_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

(b) $\hat{\mathbf{J}}_{\boldsymbol{\theta\theta}} - \mathbf{J}_{\boldsymbol{\theta\theta}} = o_p(1)$, where

$$\hat{\mathbf{J}}_{\boldsymbol{\theta\theta}} = \frac{1}{2}\frac{1}{T}\begin{bmatrix} \hat{\mathbf{C}}'\hat{\mathbf{C}} & \mathbf{C}'\hat{\mathbf{X}} \\ \hat{\mathbf{X}}'\hat{\mathbf{C}} & \hat{\mathbf{X}}'\hat{\mathbf{X}} \end{bmatrix} + \frac{1}{T}\begin{bmatrix} \hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Exploiting these results, and the method of proof, affords a framework in which to extend this asymptotic analysis to consideration of a specific class of misspecification tests.

3 A Class of Asymptotically Valid Test Procedures

In this section, first order asymptotic distribution results are developed for a class of parametric test statistics. The corresponding test procedures are derived from the conditional moment principle and are designed to detect misspecification in the null GARCH(p,q) error process, $h_t = \boldsymbol{\eta}'\mathbf{s}_{t-1}$, whilst assuming a correct regression function specification, $m(\mathbf{w}_t; \boldsymbol{\varphi})$.

If the GARCH model is correctly specified, then it follows from (2) that

$$E\left[\left(\xi_t^2 - 1\right)|\mathcal{F}_{t-1}\right] = 0.$$

Therefore, misspecification tests of GARCH models can be constructed as tests of the following moment conditions

$$E\left[\left(\xi_t^2 - 1\right)\mathbf{r}_t(\boldsymbol{\theta}_0)\right] = \mathbf{0} \tag{8}$$

where $\mathbf{r}_t(\boldsymbol{\theta}_0)$ is a $\mathcal{F}_{t-1}$ measurable function. The intuition, here, is that if the GARCH model is appropriate, then the squared standardised errors should be serially uncorrelated with any function of past information.[5]

Consistent with the notation introduced in the previous Section, let $\mathbf{d}_{\boldsymbol{\pi}t}(\boldsymbol{\theta}) = \left(\dfrac{\varepsilon_t^2}{h_t} - 1\right)\mathbf{r}_t(\boldsymbol{\theta})$, where the (test) variables in $\mathbf{r}_t$ will, in general, depend upon past history and, specifically, the process h_t. For example, $\mathbf{r}_t(\boldsymbol{\theta})$ could derive from a (quasi) score principle in which $\boldsymbol{\pi}$ denotes the unknown parameter vector in the alternative model, say h_t^a, and $H_0 : \boldsymbol{\pi} = \mathbf{0}$ is under test. In this case, and ignoring irrelevant factors of proportionality, $\mathbf{r}_t(\boldsymbol{\theta}) = \left[\frac{1}{h_t^a}\frac{\partial h_t^a}{\partial \boldsymbol{\pi}}\right]_{\boldsymbol{\pi}=\mathbf{0}}$; see Section 4.2. Therefore, as with $\mathbf{c}_t$ and $\mathbf{x}_t$, let $\mathbf{r}_t^{\infty}$ be the test variable constructed using h_t^{∞}.

To test the null of (8), the generic conditional moment test indicator is constructed as

$$\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\hat{\mathbf{r}}_t\right] = \frac{1}{T}\hat{\mathbf{R}}'\hat{\boldsymbol{\vartheta}} \tag{9}$$

where the matrix $\mathbf{R}$ has rows $\mathbf{r}_t' = \mathbf{r}_t(\boldsymbol{\theta})'$, $\hat{\boldsymbol{\vartheta}}$ is the vector with typical element $\left\{\dfrac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right\}$ and where "hats" denote that everything is evaluated at the consistent null parameter estimator, $\hat{\boldsymbol{\theta}}$. It should be noted that tests for non-linearity and/or asymmetry, discussed in Section 4 are special cases. Assessing the statistical significance of (9), which requires estimation only under the null GARCH model, provides the basis for a test procedure.

It is not being claimed that such procedures are consistent in the sense of rejecting against any departure from the null model when the null hypothesis is false. Given the framework set out in this paper, the general results of Godfrey

[5] Lundbergh and Teräsvirta (2002) employed a similar approach in order to test for no remaining ARCH effects, in a GARCH model, but with an implicit null of $E\left[\left(\xi_t^2 - 1\right)\mathbf{r}_t^{\mathcal{G}}\right] = \mathbf{0}$, where $\mathbf{r}_t^{\mathcal{G}} = \left(\xi_{t-1}^2, .., \xi_{t-m}^2\right)'$ is $\mathcal{G}_{t-1}^{t-m} = \sigma\left(\xi_{t-1}, .., \xi_{t-m}\right)$ measurable; see Section 3.1 of Lundbergh and Teräsvirta (2002). However, this could yield tests with lower power than one based on (8), since test variables of the form $\mathbf{r}_t^{\mathcal{G}}$ contain less information about $\mathcal{F}_{t-1}$ than the test variables $\mathbf{r}_t$.

and Orme (1996) could be employed to suggest alternatives against which tests based on (9), for a given choice of $\mathbf{r}_t(\boldsymbol{\theta})$, may be relatively insensitive. On the other hand, the conditional moment framework suggests that Newey's (1985) results can be exploited to determine the choice of $\mathbf{r}_t(\boldsymbol{\theta})$ which will provide optimal local power against particular forms of misspecification. Such issues are not the primary focus of the current paper, however.

The following Theorem provides sufficient conditions under which the familiar limit distribution for $\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})$ applies.

Theorem 2 *In addition to Assumptions A and B, if*
(i) $\sum_t E\sup_{\boldsymbol{\theta}} |\varepsilon_t|^l \|\mathbf{r}_t^{\infty} - \mathbf{r}_t\| = O(1)$, $l = 0, 2$
(ii) $E\sup_{\boldsymbol{\theta}} \|\mathbf{r}_t^{\infty}\|^2 < \infty$, *for all* t
(iii) $E\sup_{\boldsymbol{\theta}} \left\|\varepsilon_t^l \frac{\partial \mathbf{r}_t^{\infty}}{\partial \boldsymbol{\theta}}\right\| < \infty$, $l = 0, 2$, *for all* t
then

$$\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}'$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{\boldsymbol{\theta\theta}} & \boldsymbol{\Omega}'_{\boldsymbol{\pi\theta}} \\ \boldsymbol{\Omega}_{\boldsymbol{\pi\theta}} & \boldsymbol{\Omega}_{\boldsymbol{\pi\pi}} \end{bmatrix}, \quad \mathbf{A} = \left[-\mathbf{J}_{\boldsymbol{\pi\theta}}\mathbf{J}_{\boldsymbol{\theta\theta}}^{-1} : \quad \mathbf{I}_m\right]$$

and $\mathbf{I}_m$ *is the identity matrix of rank* $m = rank(\boldsymbol{\Omega}_{\boldsymbol{\pi\pi}})$,

$$\begin{aligned}
\boldsymbol{\Omega}_{\boldsymbol{\pi\varphi}} &= v_c E\left[\frac{1}{\sqrt{h_t^{\infty}}}\mathbf{r}_t^{\infty}\mathbf{f}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{(k_c-1)}{2}E\left[\mathbf{r}_t^{\infty}\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \\
\boldsymbol{\Omega}_{\boldsymbol{\pi\eta}} &= \frac{(k_c-1)}{2}E\left[\mathbf{r}_t^{\infty}\mathbf{x}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \\
\boldsymbol{\Omega}_{\boldsymbol{\pi\pi}} &= (k_c-1)E\left[\mathbf{r}_t^{\infty}\mathbf{r}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.
\end{aligned}$$

$\mathbf{J}_{\boldsymbol{\pi\theta}} = [\mathbf{J}_{\boldsymbol{\pi\varphi}} : \mathbf{J}_{\boldsymbol{\pi\eta}}]$ *with*

$$\mathbf{J}_{\boldsymbol{\pi\varphi}} = -E\left[\frac{\partial \mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] = E\left[\mathbf{r}_t^{\infty}\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{10}$$

$$\mathbf{J}_{\boldsymbol{\pi\eta}} = -E\left[\frac{\partial \mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\eta}'}\right] = E\left[\mathbf{r}_t^{\infty}\mathbf{x}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{11}$$

From the above result, the general form of the misspecification test statistic is the quadratic form

$$T\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\Sigma}}_T^{-1}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}}) \tag{12}$$

which has a χ_m^2 limiting distribution under the null, where $\hat{\boldsymbol{\Sigma}}_T$ is any consistent estimator for $\boldsymbol{\Sigma}$, i.e. $\hat{\boldsymbol{\Sigma}}_T = \boldsymbol{\Sigma} + o_p(1)$. Similar in spirit to Lemma 1, the following Lemma gives an expression for $\hat{\boldsymbol{\Sigma}}_T$.

Lemma 2 *Under Assumptions A and B, and those of Theorem 2,* $\hat{\mathbf{A}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{A}}' - \boldsymbol{\Sigma} = o_p(1)$ *where*

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} \hat{\boldsymbol{\Omega}}_{\boldsymbol{\theta\theta}} & \hat{\boldsymbol{\Omega}}'_{\boldsymbol{\pi\theta}} \\ \hat{\boldsymbol{\Omega}}_{\boldsymbol{\pi\theta}} & \hat{\boldsymbol{\Omega}}_{\boldsymbol{\pi\pi}} \end{bmatrix}, \quad \hat{\mathbf{A}} = \left[-\hat{\mathbf{J}}_{\boldsymbol{\pi\theta}}\hat{\mathbf{J}}_{\boldsymbol{\theta\theta}}^{-1} : \quad \mathbf{I}_m\right]$$

where $\hat{\mathbf{\Omega}}_{\boldsymbol{\theta\theta}}$ *and* $\hat{\mathbf{J}}_{\boldsymbol{\theta\theta}}$ *are given in Lemma 1 and*

$$\begin{aligned}
\hat{\mathbf{\Omega}}_{\boldsymbol{\pi\varphi}} &= \hat{v}_c \frac{\hat{\mathbf{R}}'\hat{\mathbf{H}}^{-1/2}\hat{\mathbf{F}}}{T} + \frac{\left(\hat{k}_c - 1\right)}{2} \frac{\hat{\mathbf{R}}'\hat{\mathbf{C}}}{T}, \\
\hat{\mathbf{\Omega}}_{\boldsymbol{\pi\eta}} &= \frac{\left(\hat{k}_c - 1\right)}{2} \frac{\hat{\mathbf{R}}'\hat{\mathbf{X}}}{T} \\
\hat{\mathbf{\Omega}}_{\boldsymbol{\pi\pi}} &= \left(\hat{k}_c - 1\right) \frac{\hat{\mathbf{R}}'\hat{\mathbf{R}}}{T}
\end{aligned}$$

and $\hat{\mathbf{J}}_{\boldsymbol{\pi\theta}} = \left[\hat{\mathbf{J}}_{\boldsymbol{\pi\varphi}} : \hat{\mathbf{J}}_{\boldsymbol{\pi\eta}}\right]$ *with*

$$\hat{\mathbf{J}}_{\boldsymbol{\pi\varphi}} = \frac{\hat{\mathbf{R}}'\hat{\mathbf{C}}}{T}, \quad \hat{\mathbf{J}}_{\boldsymbol{\pi\eta}} = \frac{\hat{\mathbf{R}}'\hat{\mathbf{X}}}{T}.$$

Observe that $\mathbf{\Sigma} = \mathbf{A\Omega A}'$ depends upon the "mode" of estimation only through $\mathbf{\Omega}$ and not $\mathbf{J}_{\boldsymbol{\pi\theta}}$, which is independent of the mode of estimation. In particular, and of relevance for later discussions, if $\mathbf{J}_{\boldsymbol{\pi\varphi}} = E\left[\mathbf{r}_t^\infty \mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$ then the limit distribution of $\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})$ is not influenced by the estimation of $\boldsymbol{\varphi}$. Indeed, it appears that this claim, $\mathbf{J}_{\boldsymbol{\pi\varphi}} = \mathbf{0}$, is always made when constructing parametric misspecification tests of GARCH, and ARCH, models under the assumption of conditional symmetry; see, for example Lundbergh and Teräsvirta (2002) and Engle and Ng (1993). Using the framework introduced here, it is argued in the next section that this is not the case, in general, and in particular it is not the case for the test procedures proposed by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993).

Section 4 describes how (9) accommodates existing misspecification tests and also provides alternative asymptotically valid test procedures. Before that, however, the important effects of (known) conditional symmetry, on the preceding results, are considered although normality of ξ_t is not necessarily assumed.

3.1 The Effects of Conditional Symmetry

Conditional symmetry implies that $E\left[\xi_t^3\right] = 0$, $E\left[\varepsilon_{0t}^3 | \mathcal{F}_{t-1}\right] = 0$ and thus $v_c = 0$. Although it can be tested, see for example Bai and Ng (1993), it is often assumed as in Lundbergh and Teräsvirta (2002) and Engle and Ng (1993), with the latter actually assuming normality of ξ_t. The effect of conditional symmetry simplifies the form $\mathbf{\Sigma}$ for the class of test indicators given by (9) as follows:

Lemma 3 *Under conditional symmetry*

(i) $\mathbf{J}_{\boldsymbol{\eta\varphi}} = E\left[-\frac{\partial \mathbf{d}_{\boldsymbol{\eta}t}^\infty(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] = \mathbf{0}$;

(ii) $\mathbf{\Omega}_{\boldsymbol{\eta\varphi}} = E\left[\mathbf{d}_{\boldsymbol{\eta}t}^\infty(\boldsymbol{\theta}_0)\mathbf{d}_{\boldsymbol{\varphi}t}^\infty(\boldsymbol{\theta}_0)'\right] = \mathbf{0}$;

(iii) $\mathbf{\Omega}_{\boldsymbol{\varphi\varphi}} = \frac{(k_c - 1)}{4} E\left[\mathbf{c}_t^\infty \mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + E\left[\frac{1}{h_t^\infty}\mathbf{f}_t\mathbf{f}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$;

(iv) $\mathbf{\Omega}_{\boldsymbol{\pi\varphi}} = \frac{(k_c - 1)}{2} E\left[\mathbf{r}_t^\infty \mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

This Lemma reveals that, under conditional symmetry, $\mathbf{\Sigma} = \mathbf{A\Omega A}'$ can be expressed as

$$\mathbf{\Sigma} = \mathbf{\Omega}_{\boldsymbol{\pi\pi}} - \mathbf{\Omega}_{\boldsymbol{\pi\eta}}\mathbf{\Omega}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{\Omega}_{\boldsymbol{\eta\pi}} - \mathbf{J}_{\boldsymbol{\pi\varphi}}\mathbf{J}_{\boldsymbol{\varphi\varphi}}^{-1}\left[(k_c - 1)\,\mathbf{J}_{\boldsymbol{\varphi\varphi}} - \mathbf{\Omega}_{\boldsymbol{\varphi\varphi}}\right]\mathbf{J}_{\boldsymbol{\varphi\varphi}}^{-1}\mathbf{J}_{\boldsymbol{\varphi\pi}}. \quad (13)$$

The first term in the above expression, $\mathbf{\Sigma}^A = \mathbf{\Omega}_{\boldsymbol{\pi\pi}} - \mathbf{\Omega}_{\boldsymbol{\pi\eta}}\mathbf{\Omega}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{\Omega}_{\boldsymbol{\eta\pi}}$, is simply the variance in the conditional distribution of $\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta}_0)$ given $\mathbf{d}_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta}_0)$ and is the "appropriate" variance matrix only when the second expression is the zero matrix. Consequently, if it is "erroneously" assumed that $\mathbf{J}_{\boldsymbol{\varphi\pi}} = \mathbf{0}$ in constructing the test statistic the difference between the assumed, $\mathbf{\Sigma}^A$, and true, $\mathbf{\Sigma}$, variance matrices is $\mathbf{J}_{\boldsymbol{\pi\varphi}}\mathbf{J}_{\boldsymbol{\varphi\varphi}}^{-1}\left[(k_c-1)\,\mathbf{J}_{\boldsymbol{\varphi\varphi}} - \mathbf{\Omega}_{\boldsymbol{\varphi\varphi}}\right]\mathbf{J}_{\boldsymbol{\varphi\varphi}}^{-1}\mathbf{J}_{\boldsymbol{\varphi\pi}}$, where $(k_c-1)\,\mathbf{J}_{\boldsymbol{\varphi\varphi}} - \mathbf{\Omega}_{\boldsymbol{\varphi\varphi}} = \frac{(k_c-1)}{2}E\left[\mathbf{c}_t^{\infty}\mathbf{c}_t^{\infty\prime}\right] + (k_c-2)\,E\left[\frac{1}{h_t^{\infty}}\mathbf{f}_t\mathbf{f}_t^{\prime}\right]$. Under normality, $k_c - 1 = 2$, so that $\mathbf{\Sigma}^A - \mathbf{\Sigma}$ is positive definite, when $\mathbf{J}_{\boldsymbol{\varphi\pi}}$ has full column rank, and remains so under excess (normal) kurtosis.[6] Since both $\mathbf{\Sigma}^A$ and $\mathbf{\Sigma}$ can be consistently estimated (see below), the procedure which utilises the incorrect $\mathbf{\Sigma}^A$ will be asymptotically undersized (under normality) and increasingly so under heavier-tailed distributions.

3.1.1 Variance matrix estimators

Correspondingly, and given Lemmas 2 and 3, a consistent estimator for $\mathbf{\Sigma}$ can be obtained as

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_T &= \frac{1}{T}\left[\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{M}}_{\mathbf{X}}\hat{\mathbf{R}} - \frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{C}}\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\hat{\mathbf{C}}'\hat{\mathbf{R}}\right. \\
&\quad +\hat{\mathbf{R}}'\hat{\mathbf{C}}\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{4T}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right) \\
&\quad \left.\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\hat{\mathbf{C}}'\hat{\mathbf{R}}\right], \qquad (14)
\end{aligned}$$

where, as before, "hats" denote evaluation at $\hat{\boldsymbol{\theta}}$ and $\mathbf{M}_{\mathbf{X}} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$. Further modifications can be made according to whether $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ and/or $\mathbf{J}_{\boldsymbol{\pi\eta}}$ are null matrices. The former case, $\mathbf{J}_{\boldsymbol{\pi\varphi}} = \mathbf{0}$, yields

$$\hat{\mathbf{\Sigma}}_{1T} = \frac{1}{T}\left[\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{M}}_{\mathbf{X}}\hat{\mathbf{R}}\right] \qquad (15)$$

which is the form assumed by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993), and the test statistic (12) has the simple interpretation as T times the uncentred R^2 from regressing $\hat{\boldsymbol{\vartheta}}$ on $\left[\hat{\mathbf{R}}, \hat{\mathbf{X}}\right]$. The latter case, $\mathbf{J}_{\boldsymbol{\pi\eta}} = \mathbf{0}$, yields

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_{2T} &= \frac{1}{T}\left[\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{R}} - \frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{C}}\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\hat{\mathbf{C}}'\hat{\mathbf{R}}\right. \\
&\quad +\hat{\mathbf{R}}'\hat{\mathbf{C}}\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{4T}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right) \\
&\quad \left.\left(\hat{\mathbf{F}}'\hat{\mathbf{H}}^{-1}\hat{\mathbf{F}}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\hat{\mathbf{C}}'\hat{\mathbf{R}}\right]. \qquad (16)
\end{aligned}$$

[6] Under normality, $\mathbf{\Sigma}$ is conditional variance of $\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta}_0)$ given $\mathbf{d}_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta}_0)$ and $\mathbf{d}_{\boldsymbol{\varphi}t}^{\infty}(\boldsymbol{\theta}_0)$.

If both $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ and $\mathbf{J}_{\boldsymbol{\pi\eta}}$ are null matrices, we obtain

$$\hat{\mathbf{\Sigma}}_{3T} = \frac{1}{T}\left[\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{R}}\right], \tag{17}$$

and the test statistic (12) has the simple interpretation as the T times the uncentred R^2 from regressing $\hat{\boldsymbol{\vartheta}}$ on $\hat{\mathbf{R}}$.

3.1.2 Orthogonality

Importantly, Lemma 3 shows that $\hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{\eta}}$ are asymptotically orthogonal within a QML framework[7]. Thus, consistent estimation of $\boldsymbol{\eta}_0$ can be achieved by exploiting the QML approach, to obtain $\hat{\boldsymbol{\eta}}$, but utilizing *any* $\sqrt{T}$-consistent estimator, $\hat{\boldsymbol{\varphi}}$ (see Cox and Reid, 1987), without loss of asymptotic efficiency in estimating $\boldsymbol{\eta}_0$, although there will be a loss of efficiency in small samples; for example, $\hat{\boldsymbol{\varphi}}$ might be the least squares estimator. This might suggest that tests for the adequacy of h_t will not be influenced (asymptotically, at least) by the estimation of $\boldsymbol{\varphi}$. Whilst this intuition is correct, for example, when constructing tests for unconditional heteroskedasticity in the linear model, it is flawed when applied to certain misspecification tests for GARCH models (in particular, asymmetry and non-linearity tests). Formally, as the proof of Lemma 2 makes clear, what is required is that $\mathbf{J}_{\boldsymbol{\pi\varphi}} = \mathbf{0}$, and although this appears to have been taken for granted by many authors the following example illustrates, quite nicely, that it should not. The example employs an ARCH model which is technically not nested in the class of models characterised by Assumption A. However, assumptions such as those in Weiss (1986) could be exploited to get the same form of limit distribution as described in Theorem 2, with the obvious redefinitions of $\mathbf{x}_t$ and $\mathbf{c}_t$.

Example 1 *Suppose we have the following model*

$$\begin{aligned} y_t &= \varphi + \varepsilon_t, \\ h_t^a &= 1 + \alpha_1\varepsilon_{t-1}^2 + \pi\varepsilon_{t-1},\ 0 < \alpha_1 < 1,\ \pi > 0 \end{aligned}$$

and we want to the test the null hypothesis that $\pi = 0$, such that the null model for the conditional variance is

$$h_t = 1 + \alpha_1\varepsilon_{t-1}^2,\ 0 < \alpha_1 < 1$$

and the test indicator in (9) is $r_t = \frac{\varepsilon_{t-1}}{1+\alpha_1\varepsilon_{t-1}^2} = r_t^{\infty}$, so that var $(\varepsilon_t) = \dfrac{1}{1-\alpha_1}$ and $c_t = \frac{\partial h_t}{\partial \boldsymbol{\varphi}} = -2\frac{\alpha_1\varepsilon_{t-1}}{1+\alpha_1\varepsilon_{t-1}^2} = c_t^{\infty}$. We assume $E\left[\varepsilon_{0t}^3|\mathcal{F}_{t-1}\right] = 0$, such that Lemma 3 (i) implies that $\mathbf{J}_{\boldsymbol{\eta\varphi}} = 0$ and thus $\hat{\varphi}$ and $\hat{\eta} = \hat{\alpha}_1$ are asymptotically orthogonal. However, in this case, the scalar $J_{\pi\boldsymbol{\varphi}} = E\left[r_t c_t\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is given by

$$-2E\left\{E\left[\left.\frac{\alpha_1\varepsilon_{t-1}^2}{\left(1+\alpha_1\varepsilon_{t-1}^2\right)^2}\right|\mathcal{F}_{t-2}\right]\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -2E\left[\frac{\alpha_1\varepsilon_{t-1}^2}{\left(1+\alpha_1\varepsilon_{t-1}^2\right)^2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

[7] Although Bollerslev (1986, p.318) asserts this, he does not show it. Engle (1982) provides a useful theorem, exploiting symmetry, which enables this result to be established for the ARCH model. However, since he assumes normality, the importance of the conditional symmetry assumption of ξ_t is not stated explicitly.

Then, assuming $\Pr(\varepsilon_{t-1} \neq 0) > 0$, *it follows that* $\frac{\alpha_1 \varepsilon_{t-1}^2}{\left(1+\alpha_1 \varepsilon_{t-1}^2\right)^2} > 0$, *almost everywhere. Moreover,* $\Pr\left(\frac{\alpha_1 \varepsilon_{t-1}^2}{\left(1+\alpha_1 \varepsilon_{t-1}^2\right)^2} < 1\right) = 1$. *Therefore* $\Pr\left(0 < \frac{\alpha_1 \varepsilon_{t-1}^2}{\left(1+\alpha_1 \varepsilon_{t-1}^2\right)^2} < 1\right) = 1$, *so that* $J_{\pi\boldsymbol{\varphi}}$ *exists and is bounded between* -2 *and* 0.

This example is of relevance since it is a special case of the non-linearity test proposed by Lundbergh and Teräsvirta (2002). In the construction of that test statistic, QML is employed and it is explicitly "stated" that, because of symmetry, $E\left[T\mathbf{D}_{\boldsymbol{\pi}T}(\boldsymbol{\theta}_0)\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0)'\right] = \mathbf{0}$; Lundbergh and Teräsvirta (2002, p.433). From this they incorrectly assume that there are negligible estimation effects; i.e., that $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\varphi}} = E\left[\mathbf{r}_t^{\infty}\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$. However, the above simple example illustrates that this is not true. Generalising this example, Section 4 shows that for the tests proposed by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993), for the GARCH model $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\varphi}}$ is non-zero, rendering these test procedures asymptotically invalid even under conditional symmetry.

4 Testing for Non-linearity and Asymmetry

In this section, we illustrate the utility of the general framework described in Section 3 in two ways. Firstly, in Section 4.1, the general asymptotic analysis is applied to the Lundbergh and Teräsvirta (2002) test for non-linearity and the Engle and Ng (1993) negative size bias test for asymmetry. It is shown that both are asymptotically invalid procedures, even if the conditional distribution of ξ_t is symmetric. Secondly, the framework of Section 3 justifies two alternative, and asymptotically valid, tests for non-linearity and asymmetry in the conditional variance h_t. All the ensuing analysis is undertaken under the assumption of conditional symmetry of the errors so that Lemma 3 applies.

4.1 An Analysis of Existing Tests

4.1.1 Lundbergh and Teräsvirta Test

In order to test against non-linearity in the GARCH specification, Lundbergh and Teräsvirta (Theorem 4.1, 2002) proposed the following statistic

$$T_{LT} = T \times \frac{\hat{\boldsymbol{\vartheta}}'\hat{\mathbf{G}}(\hat{\mathbf{G}}'\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\hat{\boldsymbol{\vartheta}}}{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}} \tag{18}$$

where $\hat{\mathbf{G}}$ is a matrix with rows $\hat{\mathbf{g}}_t' = \left(\hat{\mathbf{x}}_t', \hat{\mathbf{v}}_{t-1}'\right)$ and $\hat{\mathbf{v}}_{t-1} = \left(\hat{\mathbf{v}}_{1,t-1}', \hat{\mathbf{v}}_{3,t-1}', ..., \hat{\mathbf{v}}_{n+2,t-1}'\right)'$, $\hat{\mathbf{v}}_{s,t-1} = (\hat{\varepsilon}_{t-1}^s, \hat{\varepsilon}_{t-2}^s, ..., \hat{\varepsilon}_{t-q}^s)'$. This can be interpreted as T times the uncentred R^2 following a regression of $\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)$ on $\hat{\mathbf{g}}_t'$, and is assumed to be asymptotically distributed as a $\chi^2_{(n+1)q}$ random variable under the null. In terms of the general framework of Section 3, the test indicator is of the form (9), with test variables $\hat{\mathbf{r}}_t = \hat{\mathbf{v}}_{t-1}$.[8] Lundbergh and Teräsvirta (2002) also defined an alternative regression based procedure, following Wooldridge (1991), which they

[8]Lundbergh and Teräsvirta (2002) obtain this statistic from a quasi-score principle but, given the alternative entertained, the test variables should have been $\hat{\mathbf{r}}_t = \hat{h}_t^{-1}\hat{\mathbf{v}}_{t-1}$.

suggested is robust to non-normality. However, the modification employed is actually designed to make the statistic robust to heterokurticity (as Wooldridge, 1991, p.29, makes clear), not non-normality. But heterokurticity is ruled out, anyway, by the assumptions made on ξ_t and so this alternative form is not considered further.

To focus discussion, consider a null GARCH $(1,1)$ model with $n = 1$, so that $\hat{\mathbf{r}}_t = \left(\hat{\varepsilon}_{t-1}, \hat{\varepsilon}^3_{t-1}\right)'$. The following Lemma generalises the example of the previous section and establishes that $\mathbf{J}_{\boldsymbol{\pi\varphi}} \neq \mathbf{0}$ whilst $\mathbf{J}_{\boldsymbol{\pi\eta}} = \mathbf{0}$. The former result implies that the test procedure proposed by Lundbergh and Teräsvirta (2002) is asymptotically invalid[9].

Lemma 4 *Assuming the GARCH(1,1) model under the null hypothesis and the test variables considered by Lundbergh and Teräsvirta (2002) of* $\hat{\mathbf{r}}_t = \left(\hat{\varepsilon}_{t-1}, \hat{\varepsilon}^3_{t-1}\right)'$, *(10) becomes*

$$\begin{aligned}\mathbf{J}_{\boldsymbol{\pi\varphi}} &= -2\alpha_{01} E\left[\frac{1}{h_t^{\infty}}\left(\begin{array}{c}\varepsilon_{t-1}\\ \varepsilon^3_{t-1}\end{array}\right)\sum_{i=0}^{\infty}\beta_1^i \varepsilon_{t-1-i}\mathbf{f}'_{t-1-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\ &\neq \mathbf{0}.\end{aligned}$$

whilst (11),

$$\mathbf{J}_{\boldsymbol{\pi\eta}} = E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon^3_{t-1}\end{pmatrix}\sum_{i=0}^{\infty}\beta_1^i \mathbf{s}^{\infty\prime}_{t-1-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

The implication of this is that, rather than employing the variance estimator $\hat{\mathbf{\Sigma}}_{1T}$, given in equation (15), Lundbergh and Teräsvirta (2002) should have employed version $\hat{\mathbf{\Sigma}}_{2T}$, given in equation (16), or an asymptotically equivalent version thereof.

4.1.2 Engle and Ng Test

Amongst the most popular asymmetry tests are those proposed by Engle and Ng (1993). In order to confirm the asymmetric behaviour of financial series, they constructed a number of score type tests. For purposes of exposition, consider the negative size bias test which examines the significance of (9), employing the test variable $\hat{r}_t = \hat{I}_{t-1}\hat{\varepsilon}_{t-1}$ where the indicator function I_{t-1} takes the value 1 if $\varepsilon_{t-1} \leq 0$ and 0 otherwise.

Specifically, the test statistic proposed by Engle and Ng (1993) is constructed as follows

$$T_{EN} = T \times \frac{\hat{\boldsymbol{\vartheta}}'\hat{\mathbf{G}}\left(\hat{\mathbf{G}}'\hat{\mathbf{G}}\right)^{-1}\hat{\mathbf{G}}'\hat{\boldsymbol{\vartheta}}}{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}, \tag{19}$$

[9] However, it can be shown that the test for remaining ARCH effects, also proposed by Lundbergh and Teräsvirta (2002), is asymptotically valid. The intuition for this is that because the alternative, being GARCH$(p, q+m)$, is of the same form as the null specification, asymptotically orthogonality of the regression parameter estimators and those of the GARCH process ensure that inferences concerning the latter are unaffected (asymptotoically) by the former. The same intuition also applies to the parameter constancy test, Lundbergh and Teräsvirta (2002), in which the alternative can be written as $h_t^a = \boldsymbol{\gamma}'_t\mathbf{s}_{t-1}$, $\boldsymbol{\gamma}_t = \boldsymbol{\eta} + \sum_{i=1}^{n} t^i\boldsymbol{\pi}_i$, which is still linear in the variables of $\mathbf{s}_{t-1}$.

where, here, $\hat{\mathbf{G}}$ has rows $\mathbf{g}_t' = \left(\hat{\mathbf{x}}_t', \ \hat{I}_{t-1}\hat{\varepsilon}_{t-1}\right)$, and T_{EN} is assumed to be asymptotically χ_1^2 under the null. This can be computed as T times the uncentred R^2 following a regression of $\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)$ on $\hat{\mathbf{g}}_t'$. The tests presented in their paper are derived assuming a conditional normal distribution for ξ_t, although asymptotically valid procedures can be derived assuming just conditional symmetry, as is the case here.

This case is not consistent with assumption that $\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})$ is continuously differentiable, as required for the analysis of Section 3. A direct mean value expansion of $\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})$ is not applicable, since it entails terms like $\partial \mathbf{r}_t/\partial \boldsymbol{\theta}'$ and this issue was not discussed by Engle and Ng (1993). Therefore, in general (and to deal with such a possibility) it will be assumed that

$$\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\mathbf{r}_t(\boldsymbol{\theta}_0)\right] + o_p(1). \tag{20}$$

(Note that employing (20) does not alter the generic expressions for $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ and $\mathbf{J}_{\boldsymbol{\pi\eta}}$ given by (10) and (11), respectively.) This assumption is innocuous when $\mathbf{r}_t\left(\boldsymbol{\theta}\right)$ is continuously differentiable in $\boldsymbol{\theta}$, since then $\mathbf{r}_t(\hat{\boldsymbol{\theta}}) = \mathbf{r}_t(\boldsymbol{\theta}_0) + \left(\partial \mathbf{r}_t\left(\bar{\boldsymbol{\theta}}\right)/\partial\boldsymbol{\theta}'\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, and $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is $O_p(T^{-1/2})$. When $\mathbf{r}_t$ is not continuously differentiable, (20) will have to be verified on a case by case basis, and the following result verifies this for the negative size bias test procedure:

Proposition 1 *For the negative size bias test of Engle and Ng (1993), in which $r_t = I_{t-1}\varepsilon_{t-1}$ is not continuously differentiable in $\boldsymbol{\theta}$, the equality in equation* (20) *holds.*

Again it is found that the Engle and Ng (1993) tests are asymptotically invalid, in general, under the null hypothesis since it is assumed that $\mathbf{J}_{\pi\boldsymbol{\varphi}} = \mathbf{0}$, contrary to the following Lemma.

Lemma 5 *Assuming the GARCH(1,1) model under the null hypothesis and test variable $\hat{r}_t = \hat{I}_{t-1}\hat{\varepsilon}_{t-1}$, (10)*

$$\mathbf{J}_{\pi\boldsymbol{\varphi}} = -2\alpha_{01}E\left[\frac{1}{h_t^\infty}I_{t-1}\varepsilon_{t-1}\sum_{i=0}^{\infty}\beta_1^i\varepsilon_{t-1-i}\mathbf{f}_{t-1-i}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \neq \mathbf{0},$$

in general.

4.2 Alternative Tests

The previous sub-section detailed the asymptotic invalidity of tests proposed by both Lundbergh and Teräsvirta (2002) and Engle and Ng (1993). Of course, asymptotically valid test procedures can be obtained using the framework of Section 3, together with the test variables employed by these authors.

However, these test variables are derived from a particular alternative specification for the conditional heteroskedasticity. Specifically, the alternative model employed by Lundbergh and Teräsvirta (2002, p.422) is

$$\varepsilon_t = \varsigma_t\sqrt{h_t + g_t}$$

where ς_t are *i.i.d.* (zero mean and unit variance) random variables, whereas that proposed by Engle and Ng (1993, p.1758) is of the form

$$\varepsilon_t = \varsigma_t \sqrt{h_t \exp(g_t)}$$

in which $h_t = \boldsymbol{\eta}' \mathbf{s}_{t-1}$, and $g_t = g(\boldsymbol{\pi}; \mathbf{v}_{t-1})$ characterises the misspecification where $\mathbf{v}_{t-1}$ is the vector of omitted variables. In particular, the non-linearity test of Lundbergh and Teräsvirta (2002) is constructed from the following alternative

$$h_t^a = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + g(\boldsymbol{\pi}; \mathbf{v}_{t-1}) + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{21}$$

whilst that of Engle and Ng (1993) is

$$\ln(h_t^a) = \ln\left(\alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}\right) + g(\boldsymbol{\pi}; \mathbf{v}_{t-1}), \tag{22}$$

respectively. Within the QML approach, which uses (5), the tests actually constructed by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993) can be interpreted as score tests of h_t against the alternatives of (21) and (22); i.e., tests of $H_0 : \boldsymbol{\pi} = \mathbf{0}$. Whilst this yields asymptotically valid (quasi-score) test procedures using the framework of Section 3, the alternative models proposed in the literature, and considered by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993) in their Monte Carlo studies, are not of the form used to construct these quasi-score test statistics. In those studies, the power of the test is evaluated against alternative models for the conditional heteroskedasticity (specifically GJR-GARCH and EGARCH models) which are "recursive" in nature, a characteristic which is not apparent in (21) or (22), where h_{t-i}, $i = 1, .., p$, appears on the right hand side and not the lagged values of h_t^a. For example, the GJR-GARCH(1,1) model can be expressed in the following form

$$h_t^a = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 I_{t-1} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^a$$

indicating that the conditional heteroskedasticity is "recursive" in nature, due to the inclusion of h_{t-1}^a on the right hand side. As a consequence, the non-linearity/asymmetry tests, which neglect this recursive behaviour under the alternative, may well lack power against these specifications. Similar remarks apply for the parameter constancy test constructed by Lundbergh and Teräsvirta (2002).

With this in mind, alternative tests for non-linearity and asymmetry are now constructed with the following alternative specification in mind

$$\begin{aligned} \varepsilon_t &= \varsigma_t (h_t^a)^{1/2} \\ h_t^a &= \boldsymbol{\eta}' \mathbf{s}_{t-1}^a + g_t = (a_t + g_t) + B(L) h_t^a \end{aligned} \tag{23}$$

where $\mathbf{s}_{t-1}^a = \left(1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}^a, .., h_{t-p}^a\right)'$ and $g_t = g(\mathbf{v}_{t-1}; \boldsymbol{\pi})$ is a non-linear and/or asymmetric function of ε_{t-j}, $j \geq 1$ with $\mathbf{v}_{t-1}$ being the vector of omitted variables. Thus the test indicator is of the form (9), with test variables constructed as $\hat{\mathbf{r}}_t = \left[\frac{1}{h_t^a} \frac{\partial h_t^a}{\partial \boldsymbol{\pi}}\right]_{\boldsymbol{\pi}=\mathbf{0}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

4.2.1 Testing for Non-Linearity

Following Lundbergh and Teräsvirta (2002), non-linearity is introduced in the intercept and the term containing the squared past errors via a smooth transition function $F_n(\varepsilon_{t-j};\gamma;\mathbf{c})$, $j=1,...,q$, i.e.,

$$g_t=\sum_{j=1}^{q}(\alpha_{0j}+\alpha_{1j})\,\varepsilon_{t-j}^{2}F_n(\varepsilon_{t-j};\gamma,\mathbf{c}).$$

with

$$F_n(\varepsilon_{t-j};\gamma,\mathbf{c})=\left(1+\exp\left(-\gamma\prod_{l=1}^{n}(\varepsilon_{t-j}-c_l)\right)\right)^{-1}-\frac{1}{2},\ \gamma>0,c_1\leq...\leq c_n. \tag{24}$$

For example, if the location parameter (threshold) of the transition function is zero, i.e., $c=0$, then the transition is made between the regime characterized by negative shocks to the one characterized by positive shocks. Under the null of $\gamma=0$, it follows that $F_n=0$, and taking a first-order Taylor expansion of F_n around $\gamma=0$, yields

$$g_t=\boldsymbol{\pi}'\mathbf{v}_{t-1} \tag{25}$$

where $\mathbf{v}_{t-1}=\left(\mathbf{v}_{1,t-1}',\mathbf{v}_{3,t-1}',...,\mathbf{v}_{n+2,t-1}'\right)$, with $\mathbf{v}_{s,t-1}=(\varepsilon_{t-1}^{s},\varepsilon_{t-2}^{s},...,\varepsilon_{t-q}^{s})'$, $s=1,3,..,n+2$.

Combining (23) and (25), a quasi-score test of $\boldsymbol{\pi}=\mathbf{0}$ can be based on assessing the significance of the test indicator (9) in which the test variables, given $\tilde{\boldsymbol{\varepsilon}}_0$, are constructed as

$$\begin{aligned}\hat{\mathbf{r}}_t &= \left[\frac{1}{h_t^a}\frac{\partial h_t^a}{\partial\boldsymbol{\pi}}\right]_{\boldsymbol{\pi}=\mathbf{0},\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\\ &= \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\psi}_i\hat{\mathbf{v}}_{t-1-i}'\\ &= \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\left\{\hat{B}^*(L)\right\}^{i}\hat{\mathbf{v}}_{t-1-i}'\end{aligned} \tag{26}$$

where $\hat{B}^*(L)=\hat{\beta}_1+...+\hat{\beta}_pL^{p-1}$ and, in practice, $\frac{\partial h_t^a}{\partial\boldsymbol{\pi}}$ can de derived from the recursion $\frac{\partial h_t^a}{\partial\boldsymbol{\pi}}=\mathbf{v}_{t-1}+B(L)\frac{\partial h_t^a}{\partial\boldsymbol{\pi}}$. For example, in the GARCH (1,1) model assuming $n=1$, the test variables take the form

$$\hat{\mathbf{r}}_t=\frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\beta}_1^{i}\begin{pmatrix}\hat{\varepsilon}_{t-1-i}\\ \hat{\varepsilon}_{t-1-i}^{3}\end{pmatrix},$$

compared with those employed by Lundbergh and Teräsvirta (2002) which are simply $\hat{\mathbf{r}}_t=\left(\hat{\varepsilon}_{t-1},\hat{\varepsilon}_{t-1}^{3}\right)'$, in this case.

The following Lemma, stated for the general GARCH (p,q) model, establishes that $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\varphi}}$ cannot be guaranteed to be zero even under conditional symmetry, although it turns out that $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\eta}}=\mathbf{0}$ (so that $\boldsymbol{\Omega}_{\boldsymbol{\pi}\boldsymbol{\eta}}=\mathbf{0}$, also).

Lemma 6 *Under the null GARCH(p, q) model and assuming for simplicity, but without loss of generality, test variables given by* $\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\psi}_i\hat{\mathbf{v}}'_{t-1-i}$, *with* $n = 1$, *so that* $\mathbf{v}_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}, \varepsilon^3_{t-1}, \varepsilon^3_{t-2}, ..., \varepsilon^3_{t-q})'$,

$$\mathbf{J}_{\boldsymbol{\pi\varphi}} = -2\sum_{k=1}^{q}\alpha_{0k}E\left[\frac{1}{(h_t^{\infty})^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\varepsilon_{t-k-j}\mathbf{v}_{t-1-i}\mathbf{f}'_{t-k-j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \neq \mathbf{0},$$

in general, but

$$\mathbf{J}_{\boldsymbol{\pi\eta}} = E\left[\frac{1}{(h_t^{\infty})^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\mathbf{v}_{t-1-i}\mathbf{s}^{\infty\prime}_{t-1-j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

Using these results and those of Section 3.1.1, an asymptotically valid nonlinearity test statistic can be constructed as

$$T_N = T\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\Sigma}}_{2T}^{-1}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}}) \tag{27}$$

which is asymptotically distributed as $\chi^2_{(n+1)q}$ under the null, for the general n case, where $\dim(\mathbf{v}_{t-1}) = (n+1)\,q$ and $\hat{\boldsymbol{\Sigma}}_{2T}$ is given by (16).

4.2.2 Testing for Asymmetry

The asymmetry test, of whether important negative shocks have more impact on volatility than important positive shocks, assesses if the variables $\mathbf{v}_{t-1} = (I_{t-1}\varepsilon_{t-1}, .., I_{t-q}\varepsilon_{t-q})'$ have been omitted from the null GARCH (p, q) model. Again a quasi-score test statistic is constructed from the "alternative" volatility model of (23), with asymmetry characterised by $g_t = \boldsymbol{\pi}'\mathbf{v}_{t-1}$. Within this framework and under the null of $\boldsymbol{\pi} = \mathbf{0}$, the test indicator in (9) employs test variables

$$\begin{aligned}\hat{\mathbf{r}}_t &= \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\psi}_i\hat{\mathbf{v}}_{t-1-i}\\ &= \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{B}^*(L)^i\,\hat{\mathbf{v}}_{t-1-i}.\end{aligned} \tag{28}$$

If the null model is the GARCH (1,1) specification, the test variable is

$$\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\beta}_1^i\hat{I}_{t-1-i}\hat{\varepsilon}_{t-1-i}.$$

(This test variable differs from the Engle and Ng test variable of $\hat{r}_t = \hat{I}_{t-1}\hat{\varepsilon}_{t-1}$, in this case.)

For this test indicator, neither $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ and $\mathbf{J}_{\boldsymbol{\pi\eta}}$ are null matrices, in general, as stated by the following Lemma:

Lemma 7 *Under the null GARCH(p, q) model, with test variables given by* $\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\psi}_i\hat{\mathbf{v}}_{t-1-i}$, $\mathbf{v}_{t-1} = (I_{t-1}\varepsilon_{t-1}, .., I_{t-q}\varepsilon_{t-q})'$, $\mathbf{J}_{\boldsymbol{\pi\varphi}} \neq \mathbf{0}$ *and* $\mathbf{J}_{\boldsymbol{\pi\eta}} \neq \mathbf{0}$, *in general.*

The discussion in Section 3.1 provides the following test statistic

$$T_A = T\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\Sigma}}_T^{-1}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}}), \tag{29}$$

where $\hat{\boldsymbol{\Sigma}}_T$ is given by (14), and T_A asymptotically distributed as χ_q^2.

As argued by Engle and Ng (1993), we can also test asymmetry for more extreme values of past errors. The asymptotic distribution of the test, in this case, is the same as the previous one except that the test indicator employs variables $\hat{\mathbf{r}}_t = \hat{h}_t^{-1}\sum_{i=0}^{t-1}\hat{B}(L)^i\left(\hat{I}_{t-1}\hat{\varepsilon}_{t-1}^2, .., \hat{I}_{t-q}\hat{\varepsilon}_{t-q}^2\right)'$.

5 Monte Carlo Study

In this section Monte Carlo evidence is presented on the finite sample size and power performance of the various asymmetry and non-linearity tests discussed in Section 4.

The Monte Carlo experiment for assessing the size properties of the tests is based on an AR(1)-GARCH$(1,1)$ data generation process. We consider the following sets of parameter values for the conditional mean:

Model (1): $y_t = \varepsilon_t$
Model (2)-(4): $y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t$ with $\varphi_0 = 1$ and $\varphi_1 \in \{0.1, 0.5, 0.9\}$.

where $\varepsilon_t = \sqrt{h_t}\xi_t$ with $\xi_t \sim N(0,1)$, $\xi_t \sim t(\eta)$ (standardised Student t-distribution with υ degrees of freedom) where $\upsilon \in \{7, 5, 3\}$. The inclusion of $t(3)$, for example, offers some evidence on the robustness of the procedures to violations of the moment assumptions employed. The conditional variance equation follows Engle and Ng (1993)

Model H (high persistence): $h_t = 0.01 + 0.09\varepsilon_{t-1}^2 + 0.9h_{t-1}$
Model M (medium persistence): $h_t = 0.05 + 0.05\varepsilon_{t-1}^2 + 0.9h_{t-1}$
Model L (low persistence): $h_t = 0.2 + 0.05\varepsilon_{t-1}^2 + 0.75h_{t-1}$

such that, without loss of generality, the unconditional variance of ε_t equals one.

Combining the conditional mean and variance specifications yields twelve models to consider. For this purpose, a series of 1200 data realizations were generated using the random generator number in GAUSS 5.0, with the first 200 observations being discarded, in order to avoid initialization effects, yielding a sample size of 1000 observations. Each model is replicated and estimated 1000 times by QML. The test statistics considered were T_A of (29) with $\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\beta}_1^i\hat{I}_{t-1-i}\hat{\varepsilon}_{t-1-i}$; T_N of (27) with $\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=0}^{t-1}\hat{\beta}_1^i\hat{\varepsilon}_{t-1-i}^3$; the Engle and Ng statistic, T_{EN}, of (19); and, the Lundbergh and Teräsvirta statistic, T_{LT}, of (18) with $\mathbf{v}_{t-1} = \hat{\varepsilon}_{t-1}^3$.

Table 1 reports the actual rejection frequencies when the null is true for the tests described above. The results are reported for a nominal size of 5% and the correct model for the mean is estimated. When $\xi_t \sim N(0,1)$ and there are no estimation effects (i.e., $y_t = \varepsilon_t$), the empirical sizes for T_A and T_{EN} are close to the nominal size of 5%, with the exception of low persistence volatility, when the size of T_A is 6%. When there are estimation effects from the conditional mean generated as an AR process, T_{EN} tends to be slightly undersized for medium and low persistence volatility model, whereas T_A is slightly oversized for the low volatility models.

[INSERT TABLE 1 ABOUT HERE]

The empirical size of the non-linearity test, T_N, is close to the nominal size, except for the low volatility persistence, whereas T_{LT} is undersized in all the experiments, especially for a high persistence volatility model and Student-t errors. When the conditional mean is generated as an AR process, the empirical size of T_N is close to the nominal size, whereas that of T_{LT} is lower than the nominal size of 5% for all volatility models examined and significantly so under Student-t errors. By ignoring asymptotically non-negligible estimation effects, the theoretical arguments of Section 3.1 imply that the procedures based on T_{EN} or T_{LT} will be asymptotically undersized, and increasingly so under excess-kurtosis; the Monte Carlo evidence supports this, although T_{EN} is "relatively" more robust that T_{LT}.

The results of the Monte Carlo study for assessing the power of the tests are reported in Table 2, where the nominal size is again 5%. The reported rejection frequencies are size-adjusted in the sense that they are constructing using empirical critical values obtained under the null experiments. The alternative models used are the GJR(1,1) model, with the parameter values considered by Lundbergh and Teräsvirta (2002) in their simulations; the logistic STGARCH (1,1) model, in which the transition between negative to positive shocks is made smooth by using the logistic function; the EGARCH (1,1) model with parameter values considered by Engle and Ng (1993); and, the TGARCH (1,1) model. In the last case, the parameter values used are estimates obtained by Zakoian (1994) for the CAC 40 daily stock index. Note that in these experiments, for the non-linearity tests, the "omitted variable" is $v_{t-1} = \varepsilon_{t-1}^3$ when the data is generated from the GJR and STGARCH models, but $v_{t-1} = \varepsilon_{t-1}$ for the EGARCH and TGARCH models. The models for the conditional mean equation are M1, M2 and M4 and we consider $\xi_t \sim N(0,1)$ and $\xi_t \sim t(7)$.

When the true data generating process is a GJR(1,1) model, the asymmetry test, T_A, performs remarkably well compared with the test proposed by Engle and Ng (1993), T_{EN}. This is true, as well, when the distribution of ξ_t is non-normal. Similarly, for the model with larger asymmetry, and under normality, the simulated power for the non-linearity test T_N is 89.2%, whereas that of the test proposed by Lundbergh and Teräsvirta (2002), T_{LT}, is 16.5%, when there are no estimation effects from the conditional mean. This implies that T_{LT} is relatively insensitive to this alternative model[10]. Similar conclusions can be drawn for the model with smaller asymmetry.

For smooth transitions between negative to positive shocks, i.e. the true data process is generated by STGARCH (1,1) model, the differences between the powers of T_A and T_{EN}, and T_N and T_{LT}, respectively, are quite large. When estimation effects from the conditional mean are present, say M2, and the model with larger asymmetry is examined, the power of T_N is 97% whereas that of T_{LT} is 45.9%. Similarly, the asymmetry test T_A attains a simulated power of 95.8%, whereas the actual rejection frequency of T_{EN} is 64.7%. For the non-normal distribution, the differences are also significant.

[10] If the omitted variable $\mathbf{v}_{t-1} = \left(\varepsilon_{t-1}, \varepsilon_{t-1}^3\right)'$ is considered as in the Monte Carlo study of Lundbergh and Teräsvirta (2002) and for a nominal size of 10%, then the size adjusted power are: for T_N is 95.3% and T_{LT} is 65.1%.

[INSERT TABLE 2 ABOUT HERE]

For the other data generating processes, i.e. the EGARCH (1,1) and TGARCH (1,1) models, the results are similar. The simulated power of the tests T_A and T_N is much higher than the power of the tests proposed by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002).

Overall, the Monte Carlo simulations confirm the theoretical derivations undertaken in the previous sections. The "new" tests, namely T_A and T_N, have fairly good size properties and very good power when compared with T_{EN} and T_{LT}. Moreover, the simulations reveal that these tests can be employed as general misspecification tests of asymmetry and non-linearity since they have power against the asymmetry and/or non-linear models proposed in the literature.

6 Conclusion

This paper has provided some unifying results for parametric misspecification testing in regression models with GARCH errors, which have practical implications for empirical research. Firstly, a general analytical approach has been provided for the construction of asymptotically valid test statistics that can accommodate, for example, misspecification tests for the STAR-GARCH model, something which has not been considered in the literature to date. The principal theoretical finding from this analysis is that even under conditional symmetry, implying that the estimated conditional mean (regression) and variance parameters are asymptotically orthogonal, estimation effects from the conditional mean (regression) parameters cannot be treated as asymptotically negligible. Exploiting this, it is established that the non-linearity and asymmetry tests proposed by Lundbergh and Teräsvirta (2002) and Engle and Ng (1993), respectively, are not asymptotically valid (since they ignore asymptotically non-negligible estimation effects) and, more generally, all test procedures which erroneously neglect such estimation effects will be asymptotically undersized when the error distribution is fat-tailed. Secondly, new tests have been introduced for non-linearity and asymmetry which, it is conjectured, should have better power properties, than some existing tests, against many popular alternatives to the GARCH(p,q) model.

The principal theoretical findings are supported by Monte Carlo results which also suggests that the new tests are quite powerful against various non-linear models proposed in the literature, suggesting that they can be useful as general misspecification tests against non-linearity and/or asymmetry in GARCH models.

7 Appendix A

We shall exploit the results contained in the following preliminary three Propositions.

Proposition 2 *(i) For any vector* $\mathbf{c} \in \Re^r$, $\frac{\partial m(\mathbf{w}_t;\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}'}\mathbf{c} = 0$, *almost surely (a.s.), only if* $\mathbf{c} = \mathbf{0}$. *(ii) For any vector* $\mathbf{b} \in \Re^{p+q+1}$, $\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}'}\mathbf{b} = 0$, *a.s., only if* $\mathbf{b} = \mathbf{0}$.

Proof. (i) follows immediately from the identification Assumption B2, which implies that $|m(\mathbf{w}_t;\boldsymbol{\varphi}) - m(\mathbf{w}_t;\boldsymbol{\varphi}_0)|^2 > 0$, a.s. for all $\boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$. Then, by Assumption B3 and a mean value expansion, $|m_t - m_{0t}|^2 = (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \frac{\partial m(\mathbf{w}_t;\overline{\boldsymbol{\varphi}})}{\partial \boldsymbol{\varphi}} \frac{\partial m(\mathbf{w}_t;\overline{\boldsymbol{\varphi}})}{\partial \boldsymbol{\varphi}'} (\boldsymbol{\varphi} - \boldsymbol{\varphi}_0) > 0$, a.s. for all $\boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$ and some mean value $\overline{\boldsymbol{\varphi}}$. Correspondingly, the identification condition, A3(iii) establishes (ii); see, for example, Ling and McAleer (2003), or Berkes et al (2003). ■

Proposition 3 *Under Assumptions B1 and B3(i),* $E\sup_{\boldsymbol{\theta}\in\Theta} |\varepsilon_t|^{4(1+s)} < \infty$, *for some* $s > 0$, *uniformly in* t.

Proof. *Let* $m_t \equiv m(\mathbf{w}_t; \boldsymbol{\varphi})$ *and* $m_{0t} \equiv m(\mathbf{w}_t; \boldsymbol{\varphi}_0)$, *so that* $\varepsilon_t \equiv \varepsilon_t(\boldsymbol{\theta}) = \varepsilon_{0t} - (m_t - m_{0t})$. *By Assumption B3(i) and the* c_r*-inequality, for some constant* $C > 0$ *and* $r \geq 0$

$$E \sup_{\boldsymbol{\theta}} |\varepsilon_t|^r \leq C \left(E |\varepsilon_{0t}|^r + 2^r E |B(\mathbf{w}_t)|^r \right) < \infty. \tag{A.1}$$

■

Definition 3 *In the following exposition* C, K *and* ρ *denote generic constants, independent of* $\boldsymbol{\theta}$, *whose values might change from line to line but which always satisfy* $C > 0$, $K > 0$ *and* $0 < \rho < 1$.

Remark 4 *(i) By Assumption A3(ii), for all* $r > 0$, $E\sup_{\boldsymbol{\theta}} |a_t|^r < \infty$, *provided* $E\sup_{\boldsymbol{\theta}} \left|\varepsilon_t^{2r}\right| < \infty$, *uniformly in* t.
(ii) The following inequalities will be useful: (a) $h_t^\infty = \sum_{i=0}^\infty \psi_i a_{t-i} \geq \lambda > 0$; *(b)* $h_t^\infty \geq \alpha_0 + \psi_i a_{t-i}$, $i \geq 1$; *or, (c)* $h_t^\infty \geq \alpha_0 + \psi_i \alpha_m \varepsilon_{t-i-m}^2$, $i \geq 0$, $m = 1, ..., q$. *Given the construction of initial values,* $\tilde{\boldsymbol{\varepsilon}}_0$, *we can also write (a)* $h_t = \sum_{i=0}^{t-1} \psi_i a_{t-i} \geq \lambda > 0$; *(b)* $h_t \geq \alpha_0 + \psi_i a_{t-i}$, $i = 1, ..., t-1$; *or, (c)* $h_t \geq \alpha_0 + \psi_i \alpha_m \varepsilon_{t-i-m}^2$, $i = 0, ..., t-1$, $m = 1, ..., q$.
(iii) The proofs will exploit the following results, which can be derived by exploiting the algebra of Berkes et al (2003) and Francq and Zakoian (2004). A particularly useful device, in this respect, is $x/(1+x) \leq x^s$, *for all* $x > 0$ *and any* $s \in (0,1)$.

$$\begin{aligned} |h_t^\infty - h_t| &= \sum_{i=t}^\infty \psi_i |a_{t-i}| \leq K \sum_{i=t}^\infty \rho^i |a_{t-i}| && \text{(A.2)} \\ \left| \frac{h_t^\infty - h_t}{h_t^\infty} \right| &\leq \sum_{i=t}^\infty \frac{\psi_i a_{t-i}}{\alpha_0 + \psi_i a_{t-i}} \leq K \sum_{i=t}^\infty \rho^i a_{t-i}^s, && \text{(A.3)} \end{aligned}$$

For $r = 1, 2, 3$,

$$\left| \frac{1}{(h_t^\infty)^r} - \frac{1}{(h_t)^r} \right| \leq C \left| \frac{h_t^\infty - h_t}{h_t^\infty} \right| \leq K \sum_{i=t}^\infty \rho^i a_{t-i}^s. \tag{A.5}$$

Let $\nabla_{\boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta\theta}}$ *denote first and second order differentiation, respectively; for example,* $\mathbf{x}_t^\infty = \frac{1}{h_t^\infty} \nabla_{\boldsymbol{\eta}} h_t^\infty$, $\nabla_{\boldsymbol{\varphi\eta}} h_t^\infty = \frac{\partial^2 h_t^\infty}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\eta}'}$, *etc. Then, by Assumption A3(ii),*

$$\begin{aligned} \|\mathbf{x}_t^\infty\| &\leq K \left\{ 1 + \sum_{i=1}^\infty i \rho^i a_{t-i}^s \right\}, \quad \|\mathbf{x}_t\| \leq K \left\{ 1 + \sum_{i=1}^{t-1} i \rho^i a_{t-i}^s \right\} && \text{(A.6)} \\ \left\| \frac{1}{h_t^\infty} \nabla_{\boldsymbol{\eta\eta}} h_t^\infty \right\| &\leq K \left\{ 1 + \sum_{i=2}^\infty i^2 \rho^i a_{t-i}^s \right\}, \quad \left\| \frac{1}{h_t} \nabla_{\boldsymbol{\eta\eta}} h_t \right\| \leq K \left\{ 1 + \sum_{i=2}^{t-1} i^2 \rho^i a_{t-i}^s \right\} && \text{(A.7)} \\ \frac{1}{h_t^\infty} \|\nabla_{\boldsymbol{\eta}} h_t^\infty - \nabla_{\boldsymbol{\eta}} h_t\| &\leq K \left\{ \rho^t + \sum_{i=t}^\infty i \rho^i a_{t-i}^s \right\}, \\ \frac{1}{h_t^\infty} \|\nabla_{\boldsymbol{\eta\eta}} h_t^\infty - \nabla_{\boldsymbol{\eta\eta}} h_t\| &\leq K \left\{ t \rho^t + \sum_{i=t}^\infty i^2 \rho^i a_{t-i}^s \right\}. && \text{(A.8)} \end{aligned}$$

Define $d_t = \sum_{k=1}^q \left\| \frac{\partial m_{t-k}}{\partial \boldsymbol{\varphi}} \right\| = \sum_{k=1}^q \|\mathbf{f}_{t-k}\|$ *and* $g_t = \sum_{k=1}^q \left\{ \|\mathbf{f}_{t-k-i}\|^2 + \left\| \frac{\partial^2 m_{t-k-i}}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \right\| \right\}$, *then by similar methods it can also be shown that*

$$\|\mathbf{c}_t^\infty\| \leq K\sum_{i=0}^{\infty}\rho^i d_{t-i};\ \ \|\mathbf{c}_t\| \leq K\sum_{i=0}^{t-1}\rho^i d_{t-i} \tag{A.9}$$

$$\left\|\frac{1}{h_t^\infty}\nabla_{\boldsymbol{\varphi\varphi}}h_t^\infty\right\| \leq K\sum_{i=0}^{\infty}\rho^i g_{t-i};\ \ \left\|\frac{1}{h_t}\nabla_{\boldsymbol{\varphi\varphi}}h_t\right\| \leq K\sum_{i=0}^{t-1}\rho^i g_{t-i} \tag{A.10}$$

$$\frac{1}{h_t^\infty}\|\nabla_{\boldsymbol{\varphi}}h_t^\infty - \nabla_{\boldsymbol{\varphi}}h_t\| \leq K\sum_{i=t}^{\infty}\rho^i d_{t-i},\ \ \frac{1}{h_t^\infty}\|\nabla_{\boldsymbol{\varphi\varphi}}h_t^\infty - \nabla_{\boldsymbol{\varphi\varphi}}h_t\| \leq K\sum_{i=t}^{\infty}\rho^i g_{t-i} \tag{A.11}$$

$$\left\|\frac{1}{h_t^\infty}\nabla_{\boldsymbol{\varphi\eta}}h_t^\infty\right\| \leq K\left\{d_t + \sum_{i=1}^{\infty}i\rho^i d_{t-i}\right\};\ \ \left\|\frac{1}{h_t}\nabla_{\boldsymbol{\varphi\eta}}h_t\right\| \leq K\left\{d_t + \sum_{i=1}^{t-1}i\rho^i d_{t-i}\right\} \tag{A.12}$$

$$\frac{1}{h_t^\infty}\|\nabla_{\boldsymbol{\varphi\eta}}h_t^\infty - \nabla_{\boldsymbol{\varphi\eta}}h_t\| \leq K\sum_{i=t}^{\infty}i\rho^i d_{t-i} \tag{A.13}$$

$$\begin{aligned}
\frac{1}{h_t^{\infty 2}}\left\|\frac{\partial h_t^\infty}{\partial\boldsymbol{\varphi}}\frac{\partial h_t^\infty}{\partial\boldsymbol{\varphi}'} - \frac{\partial h_t}{\partial\boldsymbol{\varphi}}\frac{\partial h_t}{\partial\boldsymbol{\varphi}'}\right\| &\leq K\sum_{i=t}^{\infty}\rho^i d_{t-i}\left\{\sum_{i=0}^{\infty}\rho^i d_{t-i}\right\}\\
\frac{1}{h_t^{\infty 2}}\left\|\frac{\partial h_t^\infty}{\partial\boldsymbol{\eta}}\frac{\partial h_t^\infty}{\partial\boldsymbol{\eta}'} - \frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}\right\| &\leq K\left\{\rho^t + \sum_{i=t}^{\infty}i\rho^i a_{t-i}^s\right\}\left\{1 + \sum_{i=1}^{\infty}i\rho^i a_{t-i}^s\right\}\\
\frac{1}{h_t^{\infty 2}}\left\|\frac{\partial h_t^\infty}{\partial\boldsymbol{\varphi}}\frac{\partial h_t^\infty}{\partial\boldsymbol{\eta}'} - \frac{\partial h_t}{\partial\boldsymbol{\varphi}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}\right\| &\leq K\Bigg[\sum_{i=t}^{\infty}\rho^i d_{t-i}\left\{1 + \sum_{i=1}^{\infty}i\rho^i a_{t-i}^s\right\}\\
&\quad + \left\{\rho^t + \sum_{i=t}^{\infty}i\rho^i a_{t-i}^s\right\}\sum_{i=0}^{\infty}\rho^i d_{t-i}\Bigg]
\end{aligned} \tag{A.14}$$

where in (A.14) we have used, $\mathbf{x}, \mathbf{y}, \mathbf{a},$ *and* $\mathbf{b}$,

$$\begin{aligned}
\left\|\mathbf{x}\mathbf{a}' - \mathbf{y}\mathbf{b}'\right\| &\leq \|\mathbf{x}-\mathbf{y}\|\,\|\mathbf{a}\| + \|\mathbf{a}-\mathbf{b}\|\,\|\mathbf{y}\|\\
\text{or,}\quad \left\|\mathbf{x}\mathbf{a}' - \mathbf{y}\mathbf{b}'\right\| &\leq \|\mathbf{a}-\mathbf{b}\|\,\|\mathbf{x}\| + \|\mathbf{x}-\mathbf{y}\|\,\|\mathbf{b}\|.
\end{aligned}$$

For $p = 1, 3/2, 2, 3,$ *conformable matrices* $\mathbf{A}$ *and* $\mathbf{B}$, *and since* $h_t/h_t^\infty \leq 1$,

$$\left\|\frac{1}{(h_t^\infty)^p}\mathbf{A} - \frac{1}{(h_t)^p}\mathbf{B}\right\| \leq K\left\{\frac{1}{(h_t^\infty)^p}\|\mathbf{A}-\mathbf{B}\| + \left\|\frac{1}{h_t^p}\mathbf{B}\right\|\left|\frac{h_t^\infty - h_t}{h_t^\infty}\right|\right\}. \tag{A.15}$$

Proposition 4 *Under Assumptions A and B, and exploiting (A.2) -(A.15), all the following are bounded uniformly in* t :

(a) $E\sup_{\boldsymbol{\theta}}|\varepsilon_t|^r\left\|\frac{1}{h_t^\infty}\nabla_{\boldsymbol{\theta}}h_t^\infty\right\|^2,\ 0 \leq r \leq 4$

(b) $E\sup_{\boldsymbol{\theta}}|\varepsilon_t|^r\left\|\frac{1}{h_t^\infty}\nabla_{\boldsymbol{\theta\theta}}h_t^\infty\right\|,\ r = 0, 2$
whilst the following are $O\left(\rho^t\right)$, *at most,*

(c) $E\sup_{\boldsymbol{\theta}}\|\varepsilon_t^r\mathbf{f}_t\|\left\|\frac{1}{(h_t^\infty)^p}\nabla_{\boldsymbol{\theta}}h_t^\infty - \frac{1}{h_t^p}\nabla_{\boldsymbol{\theta}}h_t\right\|,\ r = 0, 1,\ p = 1, 3/2, 2$

(d) $E\sup_{\boldsymbol{\theta}}|\varepsilon_t|^r\left\|\frac{1}{(h_t^\infty)^p}\nabla_{\boldsymbol{\theta\theta}}h_t^\infty - \frac{1}{h_t^p}\nabla_{\boldsymbol{\theta\theta}}h_t\right\|,\ r = 0, 2,\ p = 1, 2$

(e) $E\sup_{\boldsymbol{\theta}}|\varepsilon_t|^r\left\|\frac{1}{(h_t^\infty)^p}\left(\nabla_{\boldsymbol{\theta}}h_t^\infty\right)\left(\nabla_{\boldsymbol{\theta}}h_t^\infty\right)' - \frac{1}{h_t^p}\left(\nabla_{\boldsymbol{\theta}}h_t\right)'\left(\nabla_{\boldsymbol{\theta}}h_t\right)'\right\|,\ r = 0, 2,\ p = 2, 3.$

Proof. *(a) and (b) follow in a straightforward manner from (A.6), (A.7), (A.9), (A.10) and (A.12), and the fact that the following moments are bounded:* $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t^4\mathbf{f}_{t-k}\mathbf{f}_{t-j}\right\|$, $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t^2\nabla_{\boldsymbol{\theta}}\mathbf{f}_{t-k}\right\|$, $E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^4\varepsilon_{t-j}^{2s}\varepsilon_{t-k}^{2s}\right|$, *s.* $\in(0,1)$. *For example, by an application of Holder's inequality and then Cauchy-Schwartz,*

$$E\sup_{\boldsymbol{\theta}}\varepsilon_t^4\varepsilon_{t-j}^{2s}\varepsilon_{t-k}^{2s}\leq\left(E\sup_{\boldsymbol{\theta}}\varepsilon_t^{4(1+s)}\right)^{1/(1+s)}\left(E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-j}\varepsilon_{t-k}\right|^{2(1+s)}\right)^{s/(1+s)}<\infty$$

because, for some $s\in(0,1)$, $E\sup_{\boldsymbol{\theta}}\varepsilon_t^{4(1+s)}<\infty$. *For (c) use (A.15), (A.11), (A.8), (A.9), (A.6), (A.3) and the fact that* $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\mathbf{f}_t\mathbf{f}_{t-k}'\right\|$, $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\mathbf{f}_t\mathbf{f}_{t-k}\varepsilon_{t-j}^{2s}\right\|$, $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\mathbf{f}_t\varepsilon_{t-k}^{2s}\right\|$, $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\mathbf{f}_t\varepsilon_{t-k}^{2s}\varepsilon_{t-j}^{2s}\right\|$ *are all bounded, by Cauchy-Schwartz, for* $s\in(0,1)$. *In particular,*

$$\left[E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\mathbf{f}_t\varepsilon_{t-k}^{2s}\varepsilon_{t-j}^{2s}\right\|\right]^2\leq E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\mathbf{f}_t\right\|^2E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-k}\varepsilon_{t-j}\right|^{4s}<\infty,$$

since, by Holder's inequality, and $4s\left(1+s\right)\leq4\left(1+s\right)$, $s\in(0,1)$,

$$E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-k}\varepsilon_{t-j}\right|^{4s}\leq\left(E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-k}\right|^{4s(1+s)}\right)^{1/(1+s)}\left(E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-j}\right|^{4(1+s)}\right)^{s/(1+s)}<\infty.$$

Similarly (d) holds since all the following are bounded: $E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t\right|^2\left|\varepsilon_{t-j}\right|^{2s}\left\|\mathbf{f}_{t-k}\right\|^2$, $E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t\right|^2\left|\varepsilon_{t-j}\right|^{2s}\left\|\nabla_{\boldsymbol{\varphi}}\mathbf{f}_{t-k}\right\|$, $s\in(0,1)$. *In particular,*

$$\begin{aligned}\left[E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t\right|^2\left|\varepsilon_{t-j}\right|^{2s}\left\|\mathbf{f}_{t-k}\right\|^2\right]^2&\leq E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t^2\mathbf{f}_{t-k}\right\|^2E\sup_{\boldsymbol{\theta}}\left\|\left|\varepsilon_{t-j}\right|^{2s}\mathbf{f}_{t-k}\right\|^2<\infty\\\left[E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t\right|^2\left|\varepsilon_{t-j}\right|^{2s}\left\|\nabla_{\boldsymbol{\varphi}}\mathbf{f}_{t-k}\right\|\right]^2&\leq E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t\nabla_{\boldsymbol{\varphi}}\mathbf{f}_{t-k}\right\|^2E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t\varepsilon_{t-j}^{2s}\right|^2<\infty.\end{aligned}$$

Finally, (e) can be shown to hold, in a similar manner, noting that the following are bounded: $E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t^2\mathbf{f}_{t-k}\varepsilon_{t-j}^{2s}\varepsilon_{t-l}^{2s}\right\|$, $E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^2\varepsilon_{t-j}^{2s}\varepsilon_{t-k}^{2s}\varepsilon_{t-l}^{2s}\right|$, $s\in(0,1)$. *In particular,*

$$\begin{aligned}\left[E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t^2\mathbf{f}_{t-k}\varepsilon_{t-j}^{2s}\varepsilon_{t-l}^{2s}\right\|\right]^2&\leq\left(E\sup_{\boldsymbol{\theta}}\left\|\varepsilon_t^2\mathbf{f}_{t-k}\right\|^2\right)\left(E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-k}\varepsilon_{t-l}\right|^{4s}\right)<\infty\\\left[E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^2\varepsilon_{t-j}^{2s}\varepsilon_{t-k}^{2s}\varepsilon_{t-l}^{2s}\right|\right]^2&\leq\left(E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^4\varepsilon_{t-j}^{4s}\right|\right)\left(E\sup_{\boldsymbol{\theta}}\left|\varepsilon_{t-k}\varepsilon_{t-l}\right|^{4s}\right)<\infty\end{aligned}$$

and, by Holder's Inequality, $E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^4\varepsilon_{t-j}^{4s}\right|\leq\left(E\sup_{\boldsymbol{\theta}}\varepsilon_t^{4(1+s)}\right)^{1/(1+s)}\left(E\sup_{\boldsymbol{\theta}}\varepsilon_{t-j}^{4(1+s)}\right)^{s/(1+s)}$.
■

The following three Propositions follow the approach of Ling and McAleer (2003), Berkes et al (2003) and Francq and Zakoian (2004), and are used to establish the consistency and asymptotic normality of the QMLE estimator $\hat{\boldsymbol{\theta}}$.

Proposition 5 *Under Assumptions A, B1, B2, B3a:*

(a) $E\left[l_t^\infty(\boldsymbol{\theta})\right]$ *exists for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

(b) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|L_T^\infty(\boldsymbol{\theta})-E\left[l_t^\infty(\boldsymbol{\theta})\right]\right|=o_p(1)$.

(c) $E\left[l_t^\infty(\boldsymbol{\theta})\right]$ *achieves a unique maximum at* $\boldsymbol{\theta}_0$.

(d) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|L_T^\infty(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})\right|=o_p(1)$.

Proof.
(a) First, by Assumption A3(ii), $h_t^\infty\geq\lambda>0$, *uniformly in* $\boldsymbol{\theta}$; *therefore* $E\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\varepsilon_t^2/h_t^\infty\right|\leq\lambda^{-1}E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t\right|^2<\infty$, *by Proposition 3. Second,by Assumption A3,* $\left|h_t^\infty\right|\leq K\sum_{i=1}^\infty\rho^i\left|a_{t-i}\right|$. *Thus,* $E\sup_{\boldsymbol{\theta}}\left|h_t^\infty\right|<\infty$, *and by Jensen's inequality* $E\sup_{\boldsymbol{\theta}}\left|\ln\left|h_t^\infty\right|\right|\leq\left|\ln E\sup_{\boldsymbol{\theta}}\left|h_t^\infty\right|\right|<\infty$,

so that $E\left[l_t^\infty(\boldsymbol{\theta})\right]$ *exists for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

(b) By a Uniform Law of Large Numbers (ULLN) (e.g., Theorem 3.1 of Ling and McAleer 2003, p.287), it follows that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|L_T^\infty(\boldsymbol{\theta})-E\left[l_t^\infty(\boldsymbol{\theta})\right]\right|=o_p(1)$.

(c) Write

$$\begin{aligned}2E\left[l_t^\infty(\boldsymbol{\theta})\right] &= \left\{-E\left[\ln(h_t^\infty)\right]-E\left[\varepsilon_{0t}^2/h_t^\infty\right]\right\}-\left\{E\left[(m_t-m_{0t})^2/h_t^\infty\right]\right\}\\ &= \left\{L_1(\boldsymbol{\theta})\right\}+\left\{L_2(\boldsymbol{\theta})\right\}\end{aligned}$$

since $E\left[\varepsilon_{0t}m_t/h_t^\infty|\mathcal{F}_{t-1}\right]=0$. *Firstly,* $L_2(\boldsymbol{\theta})=-\left\{E\left[(m_t-m_{0t})^2/h_t^\infty\right]\right\}$ *achieves a maximum value of* 0 *only when* $m_t=m_{0t}$, *for all* t *almost surely, which, by Assumption B2, holds only if* $\boldsymbol{\varphi}=\boldsymbol{\varphi}_0$. *Secondly, (and as argued by Ling and McAleer, 2003, Lemma 4.4) using Proposition 2(ii) and given* $\boldsymbol{\varphi}=\boldsymbol{\varphi}_0$, $L_1(\boldsymbol{\theta})$ *achieves a maximum only if* $\boldsymbol{\eta}=\boldsymbol{\eta}_0$. *Thus* $E\left[l_t^\infty(\boldsymbol{\theta})\right]$ *achieves its unique maximum at* $\boldsymbol{\theta}=\boldsymbol{\theta}_0$.

(d) We have

$$2\left|L_T^\infty(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})\right|\le T^{-1}\sum_{t=1}^T\left|\ln\frac{h_t^\infty}{h_t}\right|+T^{-1}\sum_{t=1}^T\left|\varepsilon_t^2\left(\frac{1}{h_t^\infty}-\frac{1}{h_t}\right)\right|.$$

(A.2) and $\ln(x)\le x-1$, *for all* $x>0$, *yield* $E\sup_{\boldsymbol{\theta}\in\Theta}\left|\ln(h_t^\infty/h_t)\right|\le\lambda^{-1}E\sup_{\boldsymbol{\theta}\in\Theta}\left|h_t^\infty-h_t\right|=O(\rho^t)$, *at most. Therefore* $T^{-1}\sum_{t=1}^T E\sup_{\boldsymbol{\theta}\in\Theta}\left|\ln(h_t^\infty/h_t)\right|=o(1)$, *implying (by Markov's inequality)* $T^{-1}\sum_{t=1}^T\sup_{\boldsymbol{\theta}}\left|\ln(h_t^\infty/h_t)\right|=o_p(1)$. *Next,* $T^{-1}\sum_{t=1}^T\sup_{\boldsymbol{\theta}\in\Theta}\left|\varepsilon_t^2\left(\frac{1}{h_t^\infty}-\frac{1}{h_t}\right)\right|=T^{-1}\sum_{t=1}^T X_t u_t$ *where* $X_t=\sup_{\boldsymbol{\theta}}\varepsilon_t^2$ *is strictly stationary and ergodic with* $E\left|X_t\right|<\infty$, *and* $|u_t|\le 2\lambda^{-1}$. *By (A.5),* $E\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{h_t^\infty}-\frac{1}{h_t}\right|=O\left(\rho^t\right)$, *so that* $T^{-1}\sum_{t=1}^T a_t=o_p(1)$ *and* $T^{-1}\sum_{t=1}^T\sup_{\boldsymbol{\theta}\in\Theta}\left|\varepsilon_t^2\left(\frac{1}{h_t^\infty}-\frac{1}{h_t}\right)\right|=o_p(1)$, *applying Lemma 4.5 of Ling and McAleer (2003, p.288). This completes the proof.* ■

Proposition 6 *Under Assumptions A, B1, B2, B3a&b:*

(a) $\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}\equiv\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0)$ *is finite and positive definite, where* $\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})=E\left[\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta})\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta})'\right]$.

(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^T\left\|\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta}_0)-\mathbf{d}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0)\right\|=o_p(1)$.

(c) $\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}\right)$

Proof.

(a) We first show that $\Omega_{\boldsymbol{\theta\theta}}=E\left[\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta}_0)\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta}_0)'\right]$ is finite. Denoting $\zeta_{0t}^\infty=\left(\frac{\varepsilon_{0t}^2}{h_{0t}^\infty}-1\right)$, we have

$$\mathbf{d}_{\boldsymbol{\theta}t}^{\infty\prime}(\boldsymbol{\theta}_0)=\frac{1}{2}\zeta_{0t}^\infty\frac{1}{h_{0t}^\infty}\frac{\partial h_{0t}^\infty}{\partial\boldsymbol{\theta}}+\begin{bmatrix}\varepsilon_{0t}\frac{1}{h_{0t}^\infty}\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\\ \mathbf{0}\end{bmatrix}$$

and it is sufficient to show that $E\left\|\zeta_{0t}^{\infty 2}\left\{\frac{1}{h_{0t}^\infty}\right\}^2\frac{\partial h_{0t}^\infty}{\partial\boldsymbol{\theta}}\frac{\partial h_{0t}^\infty}{\partial\boldsymbol{\theta}'}\right\|$ and $E\left\|\varepsilon_{0t}^2\left\{\frac{1}{h_{0t}^\infty}\right\}^2\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}'}\right\|$ are both finite. Since $h_{0t}^\infty\ge\lambda>0$ for all t and $\xi_t=\varepsilon_{0t}/\sqrt{h_{0t}^\infty}$, $E\left[\xi_t^2|\mathcal{F}_{t-1}\right]=1$, this follows immediately from Assumption B3(ii) and Proposition 4.

Furthermore, $\Omega_{\boldsymbol{\theta\theta}}$ is positive definite since $E\left[\zeta_{0t}^{\infty 2}\right]=k_c-1>0$ independent of h_{0t}^∞, and, by Proposition 2, for any vectors $\mathbf{c}$, $\mathbf{b}$ of the same dimension of $\boldsymbol{\varphi}$ and $\boldsymbol{\eta}$, respectively, $\mathbf{c}'\mathbf{d}_{\boldsymbol{\varphi}t}^\infty(\boldsymbol{\theta}_0)=0$, for all t almost surely, only if $\mathbf{c}=\mathbf{0}$, and $\mathbf{b}'\mathbf{d}_{\boldsymbol{\eta}t}^\infty(\boldsymbol{\theta}_0)=0$, for all t almost surely, only if $\mathbf{b}=\mathbf{0}$.

(b) The proof is similar to that of Proposition 5. Firstly, with the notation above and $\tilde{\zeta}_{0t}=\varepsilon_{0t}^2/\tilde{h}_{0t}-1$, where $\tilde{h}_{0t}=h_t(\boldsymbol{\theta}_0)$, to distinguish it from $h_{0t}^\infty=h_t^\infty(\boldsymbol{\theta}_0)$, $\mathbf{d}_{\boldsymbol{\varphi}t}^\infty(\boldsymbol{\theta}_0)-\mathbf{d}_{\boldsymbol{\varphi}t}(\boldsymbol{\theta}_0)=\frac{1}{2}\left\{\zeta_{0t}^\infty\frac{1}{h_{0t}^\infty}\frac{\partial h_{0t}^\infty}{\partial\boldsymbol{\varphi}}-\tilde{\zeta}_{0t}\frac{1}{\tilde{h}_{0t}}\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\varphi}}\right\}+\left\{\varepsilon_{0t}\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\left(\frac{1}{h_{0t}^\infty}-\frac{1}{\tilde{h}_{0t}}\right)\right\}$, so that

$$\begin{aligned}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\mathbf{d}_{\boldsymbol{\varphi}t}^\infty(\boldsymbol{\theta}_0)-\mathbf{d}_{\boldsymbol{\varphi}t}(\boldsymbol{\theta}_0)\right)\right\| &\le \frac{1}{2}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\left\{\zeta_{0t}^\infty\frac{1}{h_{0t}^\infty}\frac{\partial h_{0t}^\infty}{\partial\boldsymbol{\varphi}}-\tilde{\zeta}_{0t}\frac{1}{\tilde{h}_{0t}}\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\varphi}}\right\}\right\|\\ &\quad+\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\left\{\varepsilon_{0t}\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\left(\frac{1}{h_{0t}^\infty}-\frac{1}{\tilde{h}_{0t}}\right)\right\}\right\|\\ &= \frac{1}{2}\left\|R_T\right\|+\left\|Q_T\right\|.\end{aligned}$$

It is sufficient to show that $E\|Q_T\| = o(1)$ and $E\|R_T\| = o(1)$. By Assumption A3, and since $(i \geq 1)$ $h_t^{\infty} \geq \alpha_0 + \psi_i a_{t-i}$, and $\xi_t = \varepsilon_{0t}/\sqrt{h_{0t}^{\infty}}$, $iid\ (0,1)$

$$
\begin{aligned}
E\|Q_T\| &\leq \lambda^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left\|\xi_t\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\right\|\left|\frac{h_{0t}^{\infty}-\tilde{h}_{0t}}{\sqrt{h_{0t}^{\infty}}}\right| \\
&\leq C\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left[\left\|\xi_t\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\right\|\left\{\sum_{i=t}^{\infty}\frac{\psi_{i0}a_{0,t-i}}{\sqrt{\alpha_{00}+\psi_{i0}a_{0,t-i}}}\right\}\right] \\
&\leq K\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left[\left\|\xi_t\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\right\|\left\{\sum_{i=t}^{\infty}\rho^i\left|a_{0,t-i}\right|^{1/2}\right\}\right]
\end{aligned}
$$

since $0 < \alpha_{00} < \infty$ and $x/\sqrt{1+x} \leq \sqrt{x}$, for all $x \geq 0$. Now, by Cauchy-Schwartz, iterative expectations and Assumption B3(ii), $E\left[\left\|\xi_t\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}a_{0,t-i}^{1/2}\right\|\right] \leq \sqrt{E\left\|\frac{\partial m_{0t}}{\partial\boldsymbol{\varphi}}\right\|^2 E\left|a_{0,t-i}\right|} < \infty$, so that

$$
E\|Q_T\| \leq O(1)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} O\left(\rho^t\right) = o(1).
$$

Next, by (A.15) and $E\left[\varepsilon_{0t}^2|\mathcal{F}_{t-1}\right] = h_{0t}^{\infty}$,

$$
\begin{aligned}
E\|R_T\| &\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left\{\varepsilon_{0t}^2\left\|\frac{1}{\left(h_{0t}^{\infty}\right)^2}\frac{\partial h_{0t}^{\infty}}{\partial\boldsymbol{\varphi}}-\frac{1}{\left(\tilde{h}_{0t}\right)^2}\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\varphi}}\right\|+\left\|\frac{1}{h_{0t}^{\infty}}\frac{\partial h_{0t}^{\infty}}{\partial\boldsymbol{\varphi}}-\frac{1}{\tilde{h}_{0t}}\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\varphi}}\right\|\right\} \\
&\leq K\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left\{\frac{1}{h_{0t}^{\infty}}\left\|\frac{\partial h_{0t}^{\infty}}{\partial\boldsymbol{\varphi}}-\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\varphi}}\right\|+\left\|\frac{1}{\tilde{h}_{0t}}\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\varphi}}\right\|\left|h_{0t}^{\infty}-\tilde{h}_{0t}\right|\right\}.
\end{aligned}
$$

It follows from (A.11), (A.9), (A.2) and similar arguments to Proposition 4 that $E\|R_T\| = o(1)$
Secondly, and in a similar fashion by and (A.8), (A.6

$$
E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{d}_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta}_0)-\mathbf{d}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0)\right)\right\| \leq K\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left\{\frac{1}{h_t^{\infty}}\left\|\frac{\partial h_{0t}^{\infty}}{\partial\boldsymbol{\eta}}-\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\eta}}\right\|+\left\|\frac{1}{\tilde{h}_{0t}}\frac{\partial\tilde{h}_{0t}}{\partial\boldsymbol{\eta}}\right\|\left|h_{0t}^{\infty}-\tilde{h}_{0t}\right|\right\} = o(1).
$$

Thus, $\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{d}_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta}_0)-\mathbf{d}_{\boldsymbol{\eta}t}(\boldsymbol{\theta}_0)\right)\right\| = o_p(1)$ by Markov's inequality.
(c) As in Lemma 5.2 of Ling and McAleer (2003), a martingale difference CLT yields $\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}^{\infty}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}})$ so that (b) yields $\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}})$. ■

Proposition 7 *Under Assumptions A, B1, B2, B3a,b&c:*

- *(a)* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\mathbf{P}_{\boldsymbol{\theta\theta}T}^{\infty}(\boldsymbol{\theta})-\mathbf{J}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})\right\| = o_p(1)$, *where* $\mathbf{J}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = -E\left[\frac{\partial\mathbf{d}_{\boldsymbol{\theta}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]$ *is finite for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ *and* $\mathbf{J}_{\boldsymbol{\theta\theta}} = \mathbf{J}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0)$ *is positive definite.*
- *(b)* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\mathbf{P}_{\boldsymbol{\theta\theta}T}^{\infty}(\boldsymbol{\theta})-\mathbf{P}_{\boldsymbol{\theta\theta}T}(\boldsymbol{\theta})\right\| = o_p(1)$.

Proof.
(a) We first show that $\mathbf{J}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = -E\left[\frac{\partial\mathbf{d}_{\boldsymbol{\theta}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]$ is finite for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$; it is then straightforward to show that $\mathbf{J}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0)$ is positive definite. We have

$$
\begin{aligned}
\frac{\partial d_{\boldsymbol{\varphi}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'} &= -\frac{1}{h_t^{\infty}}\frac{\partial m_t}{\partial\boldsymbol{\varphi}}\frac{\partial m_t}{\partial\boldsymbol{\varphi}'}-\frac{\varepsilon_t}{(h_t^{\infty})^2}\frac{\partial m_t}{\partial\boldsymbol{\varphi}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}'}-\frac{\varepsilon_t}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}}\frac{\partial m_t}{\partial\boldsymbol{\varphi}'}+\frac{\varepsilon_t}{h_t^{\infty}}\frac{\partial^2 m_t}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} \\
&\quad -\frac{1}{2}\left\{\left(2\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}'}-\left(\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)\left(\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'}\right)\right\}
\end{aligned}
$$

$$\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}'} = -\frac{1}{2}\left\{\left(2\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}'}-\left(\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)\left(\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\right)\right\}$$

$$\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}'} = -\frac{\varepsilon_t}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}}\frac{\partial m_t}{\partial \boldsymbol{\varphi}'}-\frac{1}{2}\left\{\left(2\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}'}-\left(\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)\left(\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\eta}\partial \boldsymbol{\varphi}'}\right)\right\}$$

Thus

$$\begin{aligned}\left\|\frac{\partial d_{\boldsymbol{\varphi}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}'}\right\| &\leq K\left\{\left\|\frac{\partial m_t}{\partial \boldsymbol{\varphi}}\right\|^2+(\varepsilon_t^2+1)\left(\left\|\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}}\right\|^2+\left\|\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\|\right)\right.\\&\left.+\left\|\varepsilon_t\frac{\partial m_t}{\partial \boldsymbol{\varphi}}\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}'}\right\|+\left\|\varepsilon_t\frac{\partial^2 m_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\|\right\}\end{aligned}$$

and by Assumption B3(ii), Proposition 4 (a &b) and Cauchy-Schwartz it can be seen that $E\sup_{\boldsymbol{\theta}}\left\|\frac{\partial d_{\boldsymbol{\varphi}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}'}\right\|<\infty$

Similarly, and by the same arguments

$$E\sup_{\theta}\left\|\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}'}\right\|\leq KE\sup_{\theta}\left\{(\varepsilon_t^2+1)\left(\left\|\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}}\right\|^2+\left\|\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\right\|\right)\right\}<\infty,$$

$$\begin{aligned}E\sup_{\theta}\left\|\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}'}\right\| &\leq C\left\{(\varepsilon_t^2+1)\left(\left\|\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}}\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}'}\right\|+\left\|\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\eta}\partial \boldsymbol{\varphi}'}\right\|\right)\right.\\&\left.+\left\|\varepsilon_t\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\eta}}\frac{\partial m_t}{\partial \boldsymbol{\varphi}'}\right\|\right\}<\infty.\end{aligned}$$

Thus, by Theorem 3.1 of Ling and McAleer (2003), (a) holds and this completes the proof.
(b) Note that $\sup_{\boldsymbol{\theta}}\left\|\mathbf{P}_{\boldsymbol{\theta\theta}T}^{\infty}(\boldsymbol{\theta})-\mathbf{P}_{\boldsymbol{\theta\theta}T}(\boldsymbol{\theta})\right\|\leq T^{-1}\sum_t\sup_{\boldsymbol{\theta}}\left\|\frac{\partial^2 l_t^{\infty}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\frac{\partial^2 l_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|$, and we consider the latter. First,

$$\begin{aligned}\left\|\frac{\partial d_{\boldsymbol{\varphi}t}^{\infty}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}'}-\frac{\partial d_{\boldsymbol{\varphi}t}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}'}\right\| &\leq \left\{\left\|\frac{\partial m_t}{\partial \boldsymbol{\varphi}}\right\|^2+\left\|\varepsilon_t\frac{\partial^2 m_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\|\right\}\left|\frac{1}{h_t^{\infty}}-\frac{1}{h_t}\right|+2\left\|\varepsilon_t\frac{\partial m_t}{\partial \boldsymbol{\varphi}}\right\|\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}}-\frac{1}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\right\|\\&+\varepsilon_t^2\left\|\frac{1}{(h_t^{\infty})^3}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}'}-\frac{1}{h_t^3}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right\|\\&+\frac{1}{2}\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}}\frac{\partial h_t^{\infty}}{\partial \boldsymbol{\varphi}'}-\frac{1}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\varphi}}\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right\|+\frac{1}{2}\varepsilon_t^2\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}-\frac{1}{h_t^2}\frac{\partial^2 h_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\|\\&+\frac{1}{2}\left\|\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}-\frac{1}{h_t}\frac{\partial^2 h_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\|\\&= \sum_{j=1}^{4}\|R_{jt}\|\end{aligned}$$

Consider $1/T\sum_{t=1}^{T}\sup_{\theta\in\Theta}\|R_{1t}\|$, where $R_{1t}=X_t a_t$, with $a_t=\sup_{\theta}\left|\frac{1}{h_t^{\infty}}-\frac{1}{h_t}\right|$, and apply Lemma 4.5 of Ling & McAleer, 2003. We know that $a_t<2\lambda^{-1}$ and $T^{-1}\sum_{t=1}^{T}a_t=o_p(1)$, and since $E\sup_{\theta}\left\{\left\|\frac{\partial m_t}{\partial \boldsymbol{\varphi}}\right\|^2+\left\|\varepsilon_t\frac{\partial^2 m_t}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\right\|\right\}<\infty$, by Assumption B3(ii)&(iii), we have $1/T\sum_{t=1}^{T}\sup_{\theta\in\Theta}\|R_{1t}\|=o_p(1)$. By Proposition 4 $E\sup_{\boldsymbol{\theta}}\|R_{jt}\|=O(\rho^t)$, $j=2,\ldots 6$, so that $1/T\sum_{t=1}^{T}\sup_{\theta\in\Theta}\|R_{jt}\|=o_p(1)$, by Markov's inequality.

Similarly, $\sup_{\boldsymbol{\theta}\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\partial d_{\eta t}^{\infty}(\theta)}{\partial\varphi'}-\frac{\partial d_{\eta t}(\theta)}{\partial\varphi'}\right\|=o_p(1)$, since

$$\begin{aligned}\left\|\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'}-\frac{\partial d_{\boldsymbol{\eta}t}(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'}\right\| &\leq \varepsilon_t^2\left\|\frac{1}{(h_t^{\infty})^3}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}'}-\frac{1}{h_t^3}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\varphi}'}\right\| \\ &+\frac{1}{2}\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}'}-\frac{1}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\varphi}'}\right\|+\frac{1}{2}\varepsilon_t^2\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial^2 h_t^{\infty}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\varphi}'}-\frac{1}{h_t^2}\frac{\partial^2 h_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\varphi}'}\right\| \\ &+\frac{1}{2}\left\|\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\varphi}'}-\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\varphi}'}\right\|+2\left\|\varepsilon_t\frac{\partial m_t}{\partial\boldsymbol{\varphi}}\right\|\left\|\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}}-\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\right\| \\ &= K\sum_{j=1}^{5}\|R_{jt}\|,\end{aligned}$$

and by Proposition 4 $E\sup_{\boldsymbol{\theta}}\|R_{jt}\|=O(\rho^t)$, $j=2,...6$.
Finally, and analogously

$$\begin{aligned}\left\|\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}-\frac{\partial d_{\boldsymbol{\eta}t}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}\right\| &\leq \varepsilon_t^2\left\|\frac{1}{(h_t^{\infty})^3}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}'}-\frac{1}{h_t^3}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}\right\|+\frac{1}{2}\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}'}-\frac{1}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\eta}}\frac{\partial h_t}{\partial\boldsymbol{\eta}'}\right\| \\ &+\frac{1}{2}\varepsilon_t^2\left\|\frac{1}{(h_t^{\infty})^2}\frac{\partial^2 h_t^{\infty}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}-\frac{1}{h_t^2}\frac{\partial^2 h_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right\|\frac{1}{2}\left\|\frac{1}{h_t^{\infty}}\frac{\partial^2 h_t^{\infty}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}-\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right\|\end{aligned}$$

so that, by Proposition 4, $E\sup_{\boldsymbol{\theta}}\left\|\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}-\frac{\partial d_{\boldsymbol{\eta}t}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}\right\|=O\left(\rho^t\right)$.
This completes the proof. ■

Proof of Theorem 1. By Proposition 5, as in Ling and McAleer (2003), $\hat{\boldsymbol{\theta}}=\arg\max_{\boldsymbol{\theta}}L_T(\boldsymbol{\theta})$ is consistent. The limit distribution then follows from standard mean value expansion of $\mathbf{D}_{\boldsymbol{\theta}T}(\hat{\boldsymbol{\theta}})=\mathbf{0}$, exploiting Propositions 6 and 7, as follows. Firstly, $\mathbf{0}=\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0)-\mathbf{P}_{\boldsymbol{\theta\theta}T}\left(\tilde{\boldsymbol{\theta}}\right)\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$, where $\tilde{\boldsymbol{\theta}}$ is the usual "mean value" satisfying $\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}_0+o_p(1)$. By Propositions 6 and7, $\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0)=O_p(1)$ and $\mathbf{P}_{\boldsymbol{\theta\theta}T}\left(\tilde{\boldsymbol{\theta}}\right)=O_p(1)$, so that $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=O_p(1)$. Second, by Proposition 7, and the triangle inequality, $\mathbf{P}_{\boldsymbol{\theta\theta}T}\left(\tilde{\boldsymbol{\theta}}\right)=\mathbf{J}_{\boldsymbol{\theta\theta}}+o_p(1)$. Thirdly, since $\mathbf{J}_{\boldsymbol{\theta\theta}}$ is positive definite, $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\mathbf{J}_{\boldsymbol{\theta\theta}}^{-1}\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0)+o_p(1)$, and the result follows from Proposition 6. Finally, the expressions for $\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}$ and $\mathbf{J}_{\boldsymbol{\theta\theta}}$ are easily obtained from the previous results in Lemmas 6 and 7. ■

Proof of Lemma 1. The proof follows from the results given previously. We know from these results and/or assumptions made that $E\sup_{\theta}\|\mathbf{q}_t^{\infty}\mathbf{q}_t^{\infty\prime}\|<\infty$, for $\mathbf{q}_t^{\infty\prime}=\left(\frac{1}{\sqrt{h_t^{\infty}}}\mathbf{f}_t',\mathbf{c}_t^{\infty\prime},\mathbf{x}_t^{\infty\prime}\right)$. Moreover, $T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}\mathbf{q}_t^{\infty\prime}-\mathbf{q}_t\mathbf{q}_t'\|=o_p(1)$ since

$$T^{-1}\sum_{t=1}^{T}\left\|\mathbf{q}_t^{\infty}\mathbf{q}_t^{\infty\prime}-\mathbf{q}_t\mathbf{q}_t'\right\|\leq 2T^{-1}\sum_{t=1}^{T}\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\|\,\|\mathbf{q}_t^{\infty}\|+T^{-1}\sum_{t=1}^{T}\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\|^2.$$

It is readily shown, using Proposition 4 and related results, that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\|=o_p(1)$ so that $T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\|^2=o_p(1)$ (since, $\frac{1}{T}\sum_{t=1}^{T}z_t^2\leq\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_t\right\}^2$, when $z_t\geq 0$ for all t). In addition,

$$T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\|\,\|\mathbf{q}_t^{\infty}\|\leq\sqrt{T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\|^2\,T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}\|^2}=o_p(1)$$

since $T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{q}_t^{\infty}\|^2=O_p(1)$. Therefore, by a ULLN and the triangle inequality, $T^{-1}\sum_{t=1}^{T}(\mathbf{q}_t\mathbf{q}_t')_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}-E\left[\mathbf{q}_t^{\infty}\mathbf{q}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=o_p(1)$.
We also need to show that $\hat{k}_c=k_c+o_p(1)$ and $\hat{v}_c=v_c+o_p(1)$. By similar arguments, $E\sup_{\boldsymbol{\theta}}\left(\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)^2<\infty$ and

$$\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left\|\left(\frac{\varepsilon_t^2}{h_t^{\infty}}-1\right)^2-\left(\frac{\varepsilon_t^2}{h_t}-1\right)^2\right\|\leq K\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^4+\varepsilon_t^2\right|\left|\frac{1}{h_t^{\infty}}-\frac{1}{h_t}\right|=o_p(1)$$

by (A.5) and Lemma 4.5 of Ling & McAleer (2003). Finally, $E\sup_{\boldsymbol{\theta}}\left(\frac{\varepsilon_t}{\sqrt{h_t^\infty}}\right)^3<\infty$ and by exactly the same reasoning

$$\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left\|\left(\frac{\varepsilon_t}{\sqrt{h_t^\infty}}\right)^3-\left(\frac{\varepsilon_t}{\sqrt{h_t}}\right)^3\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^3\right|\left|\frac{1}{h_t^{\infty 3/2}}-\frac{1}{h_t^{3/2}}\right|=o_p(1),$$

since $\left|\frac{1}{h_t^{\infty 3/2}}-\frac{1}{h_t^{3/2}}\right|\leq K\left|h_t^\infty-h_t\right|$. ■

8 Appendix B

Proof of Theorem 2. We establish the following:
(a) $\boldsymbol{\Omega}=E\left[\mathbf{d}_t^\infty\left(\boldsymbol{\theta}_0\right)\mathbf{d}_t^\infty\left(\boldsymbol{\theta}_0'\right)\right]$ is finite and positive definite, where $\mathbf{d}_t^\infty(\boldsymbol{\theta})'=\left(\mathbf{d}_{\boldsymbol{\varphi}t}^{\infty\prime}(\boldsymbol{\theta}),\mathbf{d}_{\boldsymbol{\eta}t}^{\infty\prime}(\boldsymbol{\theta}),\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta})'\right)$;
(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left\|\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta})-\mathbf{d}_{\boldsymbol{\pi}t}(\boldsymbol{\theta})\right\|=o_p(1)$;
(c) $\mathbf{J}_{\boldsymbol{\pi\theta}}\left(\boldsymbol{\theta}\right)$ is finite for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, so that $\sup_\theta\left\|\mathbf{P}_{\boldsymbol{\pi\theta}T}^\infty(\boldsymbol{\theta})-\mathbf{J}_{\boldsymbol{\pi\theta}}\left(\boldsymbol{\theta}\right)\right\|=o_p\left(1\right)$, where $\mathbf{P}_{\boldsymbol{\pi\theta}T}^\infty(\boldsymbol{\theta})=\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$;
Given (a) and similarly to Proposition 6, $\sqrt{T}\mathbf{D}_T^\infty\left(\boldsymbol{\theta}_0\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Omega}\right)$, where $\mathbf{D}_T^\infty\left(\boldsymbol{\theta}\right)=T^{-1}\sum\mathbf{d}_t^\infty\left(\boldsymbol{\theta}\right)$, $\mathbf{d}_t^\infty\left(\boldsymbol{\theta}\right)'=\left(\mathbf{d}_{\boldsymbol{\varphi}t}^{\infty\prime}(\boldsymbol{\theta}),\mathbf{d}_{\boldsymbol{\eta}t}^{\infty\prime}(\boldsymbol{\theta}),\mathbf{d}_{\boldsymbol{\pi}t}^I(\boldsymbol{\theta})'\right)$. $\boldsymbol{\Omega}$ is positive definite provided $\mathbf{r}_t^\infty$ does not contain redundant terms (eg, linear combinations of $\mathbf{c}_t^\infty$ and/or $\mathbf{x}_t^\infty$). By (b)

$$\sup_{\boldsymbol{\theta}}\left\|\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty\left(\boldsymbol{\theta}\right)-\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}\left(\boldsymbol{\theta}\right)\right\|\leq\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left\|\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta})-\mathbf{d}_{\boldsymbol{\pi}t}(\boldsymbol{\theta})\right\|=o_p(1)$$

so that $\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})=\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty(\hat{\boldsymbol{\theta}})+o_p(1)$ and we can deal with $\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty(\hat{\boldsymbol{\theta}})$. A mean value expansion of $\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})$ about $\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0$ yields

$$\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty(\hat{\boldsymbol{\theta}})=\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty(\boldsymbol{\theta}_0)-\mathbf{P}_{\boldsymbol{\pi\theta}T}^\infty(\tilde{\boldsymbol{\theta}})\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)$$

where $\tilde{\boldsymbol{\theta}}$ is the usual "mean value" satisfying $\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}_0+o_p(1)$. Since $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$, the triangle inequality and (c) ensure that $\mathbf{P}_{\boldsymbol{\pi\theta}T}^\infty(\tilde{\boldsymbol{\theta}})=\mathbf{J}_{\boldsymbol{\pi\theta}}+o_p(1)$ and, substituting $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\mathbf{J}_{\boldsymbol{\theta\theta}}^{-1}\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}^\infty(\boldsymbol{\theta}_0)+o_p(1)$ from Theorem 1, yields

$$\begin{aligned}\sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty(\hat{\boldsymbol{\theta}}) &= \sqrt{T}\mathbf{D}_{\boldsymbol{\pi}T}^\infty(\boldsymbol{\theta}_0)-\mathbf{J}_{\boldsymbol{\pi\theta}}\mathbf{J}_{\boldsymbol{\theta\theta}}^{-1}\sqrt{T}\mathbf{D}_{\boldsymbol{\theta}T}^\infty(\boldsymbol{\theta}_0)+o_p(1)\\ &= \mathbf{A}\sqrt{T}\mathbf{D}_T^\infty(\boldsymbol{\theta}_0)+o_p(1),\end{aligned}$$

and the result follows.
For the particular class of tests, characterised by indicator (9),

$$\begin{aligned}\frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'} &= -2\frac{\varepsilon_t}{h_t^\infty}\frac{\partial m_t}{\partial\boldsymbol{\varphi}'}\mathbf{r}_t^\infty-\frac{\varepsilon_t^2}{h_t^\infty}\mathbf{r}_t^\infty\left(\frac{1}{h_t^\infty}\frac{\partial h_t^\infty}{\partial\boldsymbol{\varphi}'}\right)+\left(\frac{\varepsilon_t^2}{h_t^\infty}-1\right)\frac{\partial\mathbf{r}_t^\infty}{\partial\boldsymbol{\varphi}'}\\ \frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'} &= -\frac{\varepsilon_t^2}{h_t^\infty}\mathbf{r}_t^\infty\left(\frac{1}{h_t^\infty}\frac{\partial h_t^\infty}{\partial\boldsymbol{\eta}'}\right)+\left(\frac{\varepsilon_t^2}{h_t^\infty}-1\right)\frac{\partial\mathbf{r}_t^\infty}{\partial\boldsymbol{\eta}'}\end{aligned}$$

so that $\mathbf{J}_{\boldsymbol{\pi\varphi}}=E\left[\mathbf{r}_t^\infty\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $\mathbf{J}_{\boldsymbol{\pi\eta}}=E\left[\mathbf{r}_t^\infty\mathbf{x}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and similarly, from expressions for $\mathbf{d}_{\boldsymbol{\theta}t}^\infty(\boldsymbol{\theta}_0)$ in the proof of Lemma 6 and $\mathbf{d}_{\boldsymbol{\pi}t}^\infty(\boldsymbol{\theta}_0)=\zeta_{0t}^\infty\mathbf{r}_{0t}^\infty$, where $\zeta_{0t}^\infty=\varepsilon_{0t}^2/h_{0t}^\infty-1$,

$$\begin{aligned}\boldsymbol{\Omega}_{\boldsymbol{\pi\varphi}} &= \nu_c E\left[\frac{1}{\sqrt{h_t^\infty}}\mathbf{r}_t^\infty\mathbf{f}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}+\frac{(k_c-1)}{2}E\left[\mathbf{r}_t^\infty\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},\\ \boldsymbol{\Omega}_{\boldsymbol{\pi\eta}} &= \frac{(k_c-1)}{4}E\left[\mathbf{r}_t^\infty\mathbf{x}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},\\ \boldsymbol{\Omega}_{\boldsymbol{\pi\pi}} &= (k_c-1)\,E\left[\mathbf{r}_t^\infty\mathbf{r}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.\end{aligned}$$

We now establish that (a)-(c), above hold:
(a) Since $\boldsymbol{\Omega}_{\boldsymbol{\theta\theta}}$ is finite (Proposition 6), by Cauchy-Schwartz, we only have to show that $E\left[\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta}_0)\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta}_0)'\right]$ is finite. The latter is true since $E\sup_{\boldsymbol{\theta}}\|\mathbf{r}_t^{\infty}\|^2<\infty$, so that $E\left\|(\zeta_{0t}^{\infty})^2\,\mathbf{r}_{0t}^{\infty}\mathbf{r}_{0t}^{\infty\prime}\right\| = (k_c-1)\,E\left\|\mathbf{r}_{0t}^{\infty}\right\|^2<\infty$.
(b) It can be shown that

$$\begin{aligned}
\left\|\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta})-\mathbf{d}_{\boldsymbol{\pi}t}(\boldsymbol{\theta})\right\| &\leq K\left\{\left|\varepsilon_t^2+1\right|\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|+\varepsilon_t^2\left\|\mathbf{r}_t^{\infty}\right\|\left|\frac{h_t^{\infty}-h_t}{h_t^{\infty}}\right|\right. \\
&\qquad\left.\varepsilon_t^2\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|\left|\frac{h_t^{\infty}-h_t}{h_t^{\infty}}\right|\right\} \\
&= K\sum_{j=1}^{3}R_{jt}.
\end{aligned}$$

By assumption, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E\sup_{\boldsymbol{\theta}}R_{1t}=o(1)$. Secondly, since $E\sup_{\boldsymbol{\theta}}\varepsilon_t^2\left\|\mathbf{r}_t^{\infty}\right\|\left|\varepsilon_{t-i}^{2s}\right|\leq\sqrt{E\sup_{\boldsymbol{\theta}}\left\|\mathbf{r}_t^{\infty}\right\|^2E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^4\varepsilon_{t-i}^{4s}\right|}<\infty$, by (A.3), $E\sup_{\theta}R_{2t}=O\left(\rho^t\right)$, so that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E\sup_{\boldsymbol{\theta}}R_{2t}=o(1)$. Finally, note that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}R_{3t} &\leq \sqrt{\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\varepsilon_t^4\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|^2\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left|\frac{h_t^{\infty}-h_t}{h_t^{\infty}}\right|^2} \\
&\leq \left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\varepsilon_t^2\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|\right\}\left\{\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left|\frac{h_t^{\infty}-h_t}{h_t^{\infty}}\right|\right\}=o_p(1)
\end{aligned}$$

since, $\frac{1}{T}\sum_{t=1}^{T}z_t^2\leq\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_t\right\}^2$, when $z_t\geq 0$ for all t, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\varepsilon_t^2\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|=o_p(1)$, by assumption, and $\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\left|\frac{h_t^{\infty}-h_t}{h_t^{\infty}}\right|=O_p(1)$ by previous results.
By Markov's Inequality, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}R_{jt}=o_p(1)$, $j=1,2,3$, and the result follows.
(c) In a similar manner to the proof of Proposition 7, with

$$\begin{aligned}
\left\|\frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'}\right\| &\leq K\left\{\left\|\varepsilon_t\frac{\partial m_t}{\partial\boldsymbol{\varphi}}\mathbf{r}_t^{\infty}\right\|+\left|\varepsilon_t^2+1\right|\left(\left\|\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\varphi}}\right\|\left\|\mathbf{r}_t^{\infty}\right\|+\left\|\frac{\partial\mathbf{r}_t^{\infty}}{\partial\boldsymbol{\varphi}'}\right\|\right)\right\} \\
\left\|\frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}\right\| &\leq K\left\{\left|\varepsilon_t^2+1\right|\left(\left\|\frac{1}{h_t^{\infty}}\frac{\partial h_t^{\infty}}{\partial\boldsymbol{\eta}'}\right\|\left\|\mathbf{r}_t^{\infty}\right\|+\left\|\frac{\partial\mathbf{r}_t^{\infty}}{\partial\boldsymbol{\eta}'}\right\|\right)\right\}.
\end{aligned}$$

Then since $h_t^{\infty}\geq\lambda>0$ for all t and $\boldsymbol{\theta}$, using Cauchy-Schwartz inequality, the assumptions of Theorem 2 and previous results, $E\sup_{\boldsymbol{\theta}}\left\|\frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'}\right\|<\infty$ and $E\sup_{\boldsymbol{\theta}}\left\|\frac{\partial\mathbf{d}_{\boldsymbol{\pi}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}\right\|<\infty$. These are sufficient for a ULLN to apply, ensuring $\sup_{\theta}\left\|\mathbf{P}_{\boldsymbol{\pi\theta}T}^{\infty}(\boldsymbol{\theta})-\mathbf{J}_{\boldsymbol{\pi\theta}}(\boldsymbol{\theta})\right\|=o_p(1)$.
Finally we briefly show that these additional assumptions of Theorem 2 are satisfied for the test variables considered in this paper, in Section 4.
First, consider the Lundbergh and Teräsvirta (2002) nonlinearity (scalar) test variable $r_t=\varepsilon_{t-k}^3=r_t^{\infty}$. Trivially, assumption (i) is satisfied. For (ii) we require $E\left[\varepsilon_{0t}^6\right]<\infty$; for (iii), $E\sup_{\boldsymbol{\theta}\in\Theta}\left\|\varepsilon_t^2\frac{\partial\mathbf{r}_t^{\infty}}{\partial\boldsymbol{\varphi}}\right\|=3E\sup_{\boldsymbol{\theta}}\left|\varepsilon_t^2\varepsilon_{t-k}^2\frac{\partial m_{t-k}}{\partial\boldsymbol{\varphi}}\right|<\infty$, by Cauchy-Schwartz. Similarly for the asymmetry test variable of $I_{t-1}\varepsilon_{t-1}$, but taking into account (20).
For the non-linearity test variable, with $\|\mathbf{r}_t^{\infty}\|$ having typical element $r_{kt}^{\infty}=\frac{1}{h_t^{\infty}}\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-k-i}^3$, $k=1,...,q$, and noting $h_t^{\infty}\geq\alpha_0+\psi_i\alpha_k\varepsilon_{t-k-i}^2$,

$$|r_{kt}^{\infty}|\leq\frac{1}{\sqrt{\alpha_k}}\sum_{i=0}^{\infty}\frac{\sqrt{\psi_i\alpha_k}\,|\varepsilon_{t-k-i}|}{\alpha_0+\psi_i\alpha_k\varepsilon_{t-k-i}^2}\sqrt{\psi_i}\varepsilon_{t-k-i}^2\leq K\sum_{i=0}^{\infty}\rho^i\varepsilon_{t-k-i}^2$$

$$|r_{kt}^{\infty}-r_{kt}|\leq K\sum_{i=t}^{\infty}\rho^i\varepsilon_{t-k-i}^2$$

$$
\begin{aligned}
\left\|\frac{\partial r_{kt}^{\infty}}{\partial \boldsymbol{\varphi}}\right\| &\leq 3\frac{1}{h_t^{\infty}}\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-k-i}^2\left\|\nabla_{\varphi}\mathbf{f}_{t-k-i}\right\|+\frac{1}{h_t^{\infty}}\sum_{i=0}^{\infty}\psi_i\left\|\varepsilon_{t-k-i}^3\mathbf{c}_t^{\infty}\right\| \\
&\leq 3\frac{1}{\sqrt{\alpha_k}}\sum_{i=0}^{\infty}\frac{\sqrt{\psi_i\alpha_k}\left|\varepsilon_{t-k-i}\right|}{\alpha_0+\psi_i\alpha_k\varepsilon_{t-k-i}^2}\sqrt{\psi_i}\left\|\varepsilon_{t-k-i}\nabla_{\varphi}\mathbf{f}_{t-k-i}\right\|+\frac{1}{\sqrt{\alpha_k}}\sum_{i=0}^{\infty}\frac{\sqrt{\psi_i\alpha_k}\left|\varepsilon_{t-k-i}\right|}{\alpha_0+\psi_i\alpha_k\varepsilon_{t-k-i}^2}\sqrt{\psi_i}\varepsilon_{t-k-i}^2\left\|\mathbf{c}_t^{\infty}\right\| \\
&\leq K\left\{\sum_{i=0}^{\infty}\rho^i\left\|\varepsilon_{t-k-i}\nabla_{\varphi}\mathbf{f}_{t-k-i}\right\|+\sum_{i=0}^{\infty}\rho^i\varepsilon_{t-k-i}^2\left\|\mathbf{c}_t^{\infty}\right\|\right\} \\
\left\|\frac{\partial r_{kt}^{\infty}}{\partial \boldsymbol{\eta}}\right\| &\leq K\left\{\sum_{i=1}^{\infty}i\rho^i\varepsilon_{t-k-i}^2+\sum_{i=0}^{\infty}\rho^i\varepsilon_{t-k-i}^2\left\|\mathbf{x}_t^{\infty}\right\|\right\}.
\end{aligned}
$$

It is then straightforward to show that assumptions (i)-(iii) are satisfied. Similarly for the asymmetry test variable $r_t^{\infty}=\frac{1}{h_t^{\infty}}\sum_{i=0}^{\infty}\psi_i I_{t-k}\varepsilon_{t-k-i}$, but taking into account (20).
■

Proof of Lemma 2. The proof is similar to that of Lemma 1. We can show that $E\sup_{\theta}\|\mathbf{r}_t^{\infty}\mathbf{q}_t^{\infty\prime}\|<\infty$, so that $T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{r}_t^{\infty}\mathbf{q}_t^{\infty\prime}-E\left[\mathbf{r}_t^{\infty}\mathbf{q}_t^{\infty\prime}\right]\|=o_p(1)$, by previous arguments. It remains to establish that $T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}}\|\mathbf{r}_t^{\infty}\mathbf{q}_t^{\infty\prime}-\mathbf{r}_t\mathbf{q}_t'\|=o_p(1)$, since this then ensures that $T^{-1}\sum_{t=1}^{T}\left(\mathbf{r}_t\mathbf{q}_t'\right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}-E\left[\mathbf{r}_t^{\infty}\mathbf{q}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=o_p(1)$. Note that

$$
T^{-1}\sum_{t=1}^{T}\left\|\mathbf{r}_t^{\infty}\mathbf{q}_t^{\infty\prime}-\mathbf{r}_t\mathbf{q}_t'\right\|\leq T^{-1}\sum_{t=1}^{T}\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|\left\|\mathbf{q}_t^{\infty}\right\|+T^{-1}\sum_{t=1}^{T}\left\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\right\|\left\|\mathbf{r}_t^{\infty}\right\|+T^{-1}\sum_{t=1}^{T}\left\|\mathbf{r}_t^{\infty}-\mathbf{r}_t\right\|\left\|\mathbf{q}_t^{\infty}-\mathbf{q}_t\right\|
$$

and it can easily be shown that each term on the right hand side is $o_p(1)$. We also know that $\hat{k}_c=k_c+o_p(1)$ and $\hat{v}_c=v_c+o_p(1)$, from the proof of Lemma 1, and the result follows. ■

Proof of Lemma 3. (i) Firstly, from the expression for $\frac{\partial d_{\boldsymbol{\eta}t}^{\infty}(\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}'}$ in Proposition 7, it is easy to see that $\mathbf{J}_{\boldsymbol{\eta}\boldsymbol{\varphi}}=\frac{1}{2}E\left[\mathbf{x}_t^{\infty}\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.
Now,

$$
E\left[\mathbf{x}_t^{\infty}\mathbf{c}_t^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=-2\sum_{k=1}^{q}\alpha_{0k}E\left[\frac{1}{h_t^{\infty 2}}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\mathbf{s}_{t-1-i}^{\infty}\varepsilon_{t-k-j}\mathbf{f}_{t-k-j}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \qquad (30)
$$

which exists, provided $E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}\mathbf{s}_{t-m}^{\infty}\mathbf{f}_{t-l}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ exists (for all l,m), since $\left|\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\right|=\left|\sum_{i=0}^{\infty}\psi_i\right|^2\leq\left\{\sum_{i=0}^{\infty}|\psi_i|\right\}^2<\infty$. Thus $E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}\mathbf{s}_{t-m}^{\infty}\mathbf{f}_{t-l}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ has to be examined for the cases $l=m$, $l<m$ and $l>m$, where $\mathbf{s}_{t-m}^{\infty}=\left(1,\varepsilon_{t-m}^2,..,\varepsilon_{t-m-q+1}^2,h_{t-m}^{\infty},..,h_{t-m-p+1}^{\infty}\right)'$.
Specifically, for $l=m$, $E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}\mathbf{s}_{t-l}^{\infty}\mathbf{f}_{t-l}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is

$$
E\left\{E\left[\frac{1}{h_t^{\infty 2}}\left(\varepsilon_{t-l},\,\varepsilon_{t-l}^3,..,\varepsilon_{t-l}\varepsilon_{t-l-q+1}^2,\,\varepsilon_{t-l}h_{t-l}^{\infty},..,\varepsilon_{t-l}h_{t-l-p+1}^{\infty}\right)'\middle|\mathcal{F}_{t-l-1}\right]\mathbf{f}_{t-l}'\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
$$

which is zero if the expression for the conditional expectation, given $\mathcal{F}_{t-l-1}$, above is zero. To establish the latter, follow Engle (1982) and treat this conditional expectation in two steps, observing that ε_{t-l-n}, $n=1,2,...$, are $\mathcal{F}_{t-l-1}$- measurable. First, construct the conditional expectation given $\mathcal{F}_{t-l}$, which is

$$
\left[\left(\varepsilon_{t-l},\,\varepsilon_{t-l}^3,..,\varepsilon_{t-l}\varepsilon_{t-l-q+1}^2,\,\varepsilon_{t-l}h_{t-l}^{\infty},..,\varepsilon_{t-l}h_{t-l-p+1}^{\infty}\right)'E\left\{\frac{1}{h_t^{\infty 2}}\middle|\mathcal{F}_{t-l}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\equiv\phi(\varepsilon_{t-l}),
$$

where it is implicit that $\phi(.)$ is evaluated at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. Since h_t^{∞} is symmetric in ε_{t-l} and the elements in $\varepsilon_{t-l}s_{t-l}^{\infty}$ are all anti-symmetric in ε_{t-l}, the elements in $\left(h_t^{\infty}\right)^{-2}\varepsilon_{t-l}s_{t-l}^{\infty}$ are anti-symmetric in ε_{t-l}, which forms part of $\mathcal{F}_{t-l}$ and, at the second step, expectations with respect to $\mathcal{F}_{t-l-1}$ are taken only with random elements. Now, because h_t^{∞} is symmetric in ε_{t-l}, its conditional density given ε_{t-l} is also symmetric in ε_{t-l}. Therefore, by Engle (1982, Lemma p.1006), $\phi(\varepsilon_{t-l})$ is anti-symmetric in ε_{t-l}. Finally, the second step involves $E\left[\phi(\varepsilon_{t-l})|\mathcal{F}_{t-l-1}\right]$ which is zero, because the conditional density of ε_{t-l} given $\mathcal{F}_{t-l-1}$ is symmetric and $\phi(.)$ is

anti-symmetric.
The other typical expectation in (30) for $l < m$ and $l > m$ is

$$E\left\{E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-m}\mathbf{s}_{t-l}^{\infty}\middle|\mathcal{F}_{t-m-1}\right]\mathbf{f}'_{t-m}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

which is zero if the conditional expectation, given $\mathcal{F}_{t-m-1}$, is zero. The later can be expressed as

$$E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-m}\mathbf{s}_{t-l}^{\infty}\middle|\mathcal{F}_{t-m-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left\{E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-m}\mathbf{s}_{t-l}^{\infty}\middle|\mathcal{F}_{t-m}\right]\middle|\mathcal{F}_{t-m-1}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

For $l > m$, the elements of $\mathbf{s}_{t-l}^{\infty}$ belong to $\mathcal{F}_{t-m-1}$ and the preceding arguments show that $E\left[(h_t^{\infty})^{-2}\varepsilon_{t-m}\middle|\mathcal{F}_{t-m-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$. For $m > l$, note that the elements of $(h_t^{\infty})^{-2}\mathbf{s}_{t-l}^{\infty}$ are symmetric in ε_{t-m}, so that $E\left[(h_t^{\infty})^{-2}\varepsilon_{t-m}\mathbf{s}_{t-l}^{\infty}\middle|\mathcal{F}_{t-m}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \equiv \phi(\varepsilon_{t-m})$ is anti-symmetric in ε_{t-m} and, again, $E\left[\phi(\varepsilon_{t-m})|\,\mathcal{F}_{t-m-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$, where elements included in the conditioning set $\mathcal{F}_{t-m-1}$ are treated as non-random when taking the conditional expectation. It follows that $\mathbf{J}_{\boldsymbol{\eta}\boldsymbol{\varphi}} = \mathbf{0}$.
Since $v_c = 0$, (ii), (iii) and (iv) follow immediately, given previous definitions. ■

Proof of Lemma 4. Note that $\mathbf{c}_t^{\infty\prime} = (h_t^{\infty})^{-1}\sum_{i=0}^{\infty}\beta_1^i\varepsilon_{t-1-i}\mathbf{f}'_{t-1-i}$ and $\mathbf{r}_t = \left(\varepsilon_{t-1}, \varepsilon_{t-1}^3\right)'$ so that $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\varphi}}$ can be written as

$$\begin{aligned}
\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\varphi}} &= -2\alpha_{01}E\left\{E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}^2\\ \varepsilon_{t-1}^4\end{pmatrix}\middle|\mathcal{F}_{t-2}\right]\mathbf{f}'_{t-1}\right.\\
&\qquad\left.+\sum_{i=1}^{\infty}\beta_1^i E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-1}^3\end{pmatrix}\varepsilon_{t-1-i}\middle|\mathcal{F}_{t-2}\right]\mathbf{f}'_{t-1-i}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\
&= -2\alpha_{01}E\left\{E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}^2\\ \varepsilon_{t-1}^4\end{pmatrix}\middle|\mathcal{F}_{t-2}\right]\mathbf{f}'_{t-1}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
\end{aligned}$$

which is non-zero, in general, since $E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}^2\\ \varepsilon_{t-1}^4\end{pmatrix}\middle|\mathcal{F}_{t-2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} > 0$ almost surely. The second term (after the second equality) is zero because, for $j \geq 2$,

$$E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-1}^3\end{pmatrix}\varepsilon_{t-j}\middle|\mathcal{F}_{t-2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left(\varepsilon_{t-j}E\left[\frac{1}{h_t^{\infty}}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-1}^3\end{pmatrix}\middle|\mathcal{F}_{t-2}\right]\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and

$$E\left[\frac{1}{h_t^{\infty}}\varepsilon_{t-1}^s\middle|\mathcal{F}_{t-2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left[\phi\left(\varepsilon_{t-1}\right)|\mathcal{F}_{t-2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

where $E\left[(h_t^{\infty})^{-1}\varepsilon_{t-1}^s|\mathcal{F}_{t-1}\right] = \phi(\varepsilon_{t-1})$, $s = 1, 3$, which is anti-symmetric in ε_{t-1}, so that $E\left[\phi(\varepsilon_{t-1})|\mathcal{F}_{t-2}\right] = 0$ because the conditional density of ε_{t-1} given $\mathcal{F}_{t-2}$ is symmetric. Thus, in general, $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\varphi}} \neq \mathbf{0}$.
Second, with $\mathbf{x}_t^{\infty\prime} = (h_t^{\infty})^{-1}\sum_{i=0}^{\infty}\beta_1^i\mathbf{s}_{t-1-i}^{\infty\prime}$, $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\eta}}$ can be written as

$$\begin{aligned}
\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\eta}} &= E\left\{E\left[\frac{1}{h_t^{\infty}}\binom{\varepsilon_{t-1}}{\varepsilon_{t-1}^3}\mathbf{s}_{t-1}^{\infty\prime}\middle|\mathcal{F}_{t-2}\right]\right.\\
&\qquad\left.+\sum_{i=1}^{\infty}\beta_1^i E\left[\frac{1}{h_t^{\infty}}\binom{\varepsilon_{t-1}}{\varepsilon_{t-1}^3}\middle|\mathcal{F}_{t-2}\right]\mathbf{s}_{t-1-i}^{\infty\prime}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.
\end{aligned}$$

Similar arguments to those employed previously, imply that $\mathbf{J}_{\boldsymbol{\pi}\boldsymbol{\eta}}$ is the null vector. ■

Proof of Proposition 1. The method of proof follows very closely that of Godfrey (1996). Consider the negative size bias test of Engle and Ng (1993) in which $\hat{r}_t = \hat{I}_{t-1}\hat{\varepsilon}_{t-1}$

and for simplicity, in this case, $m(\mathbf{w}_t;\boldsymbol{\varphi}) = \mathbf{w}_t'\boldsymbol{\varphi}$. Define the following dummy variables, which will be employed in the ensuing asymptotic analysis

$$\begin{array}{lllllll}
\mathcal{D}_{t1} & = & 1, & \text{if} \quad \varepsilon_{0,t-1} \leq 0 & \text{and} \quad \hat{\varepsilon}_{t-1} \leq 0, & \mathcal{D}_{t1} = 0, & \text{otherwise} \\
\mathcal{D}_{t2} & = & 1, & \text{if} \quad \varepsilon_{0,t-1} > 0 & \text{and} \quad \hat{\varepsilon}_{t-1} \leq 0, & \mathcal{D}_{t2} = 0, & \text{otherwise} \\
\mathcal{D}_{t3} & = & 1, & \text{if} \quad \varepsilon_{0,t-1} \leq 0 & \text{and} \quad \hat{\varepsilon}_{t-1} > 0, & \mathcal{D}_{t3} = 0, & \text{otherwise} \\
\mathcal{D}_{t4} & = & 1, & \text{if} \quad \varepsilon_{0,t-1} > 0 & \text{and} \quad \hat{\varepsilon}_{t-1} > 0, & \mathcal{D}_{t4} = 0, & \text{otherwise}
\end{array}$$

for $t = 1, ..., T$. Note that both $\Pr(\mathcal{D}_{t2} = 1)$ and $\Pr(\mathcal{D}_{t3} = 1)$ tend to zero as $T \to \infty$, under fairly general conditions on $\mathbf{w}_t$, since $\hat{\varepsilon}_{t-1} - \varepsilon_{0,t-1} = -\mathbf{w}_{t-1}'(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$ and $\hat{\boldsymbol{\varphi}}$ is root-T consistent for $\boldsymbol{\varphi}_0$.
Then, noting that $\hat{r}_t - r_{0t} = 0$ when $\mathcal{D}_{t4} = 1$, the difference between $\sqrt{T}D_{\boldsymbol{\pi}T}(\hat{\boldsymbol{\theta}})$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right) r_{0t}\right]$ can be expressed as

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)(\hat{r}_t - r_{0t})\right] &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\left\{\mathcal{D}_{t1}(\hat{\varepsilon}_{t-1} - \varepsilon_{0,t-1}) + \mathcal{D}_{t2}\hat{\varepsilon}_{t-1} - \mathcal{D}_{t3}\varepsilon_{0,t-1}\right\}\right] \\
&= \Gamma_1 + \Gamma_2 + \Gamma_3
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_1 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathcal{D}_{t1}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)(\hat{\varepsilon}_{t-1} - \varepsilon_{0,t-1})\right] = -\frac{1}{T}\sum_{t=1}^{T}\left[\mathcal{D}_{t1}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\mathbf{w}_{t-1}'\right]\sqrt{T}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0) \\
\Gamma_2 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathcal{D}_{t2}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\hat{\varepsilon}_{t-1}\right] \\
\Gamma_3 &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathcal{D}_{t3}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\varepsilon_{0,t-1}\right].
\end{aligned}$$

It can now be shown that $\Gamma_j = o_p(1)$, for $j = 1, 2, 3$, which is sufficient for $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)(\hat{r}_t - r_{0t})\right] = o_p(1)$. For example, Γ_3 can be written as

$$\begin{aligned}
\Gamma_3 &= -\left(\sum_{t=1}^{T}\mathcal{D}_{t3}/T\right)^{1/2}\left[\left(\sum_{t=1}^{T}\mathcal{D}_{t3}\right)^{-1/2}\sum_{t=1}^{T}\mathcal{D}_{t3}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\varepsilon_{0,t-1}\right] \\
&= -(\mathcal{M}_3/T)^{1/2}\left[(\mathcal{M}_3)^{-1/2}\sum_{t\in\mathcal{T}_3}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\varepsilon_{0,t-1}\right]
\end{aligned}$$

where $\mathcal{M}_3 = \sum_{t=1}^{T}\mathcal{D}_{t3}$ is the number of observations for which $\mathcal{D}_{t3} = 1$ and $\mathcal{T}_3$ denotes the subsample of observations with $\mathcal{D}_{t3} = 1$. Now, $\mathcal{M}_3/T$ is the proportion of sample observations for which $\mathcal{D}_{t3} = 1$. Since $\Pr(\mathcal{D}_{t3} = 1) \to 0$, $\mathcal{M}_3/T$ is thus $o_p(1)$. Similar to the preceding analysis, since $\mathcal{M}_3 \to \infty$, a mean value expansion of $(\mathcal{M}_3)^{-1/2}\sum_{t\in\mathcal{T}_3}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\varepsilon_{0,t-1}$ reveals that it is $O_p(1)$. Therefore, $\Gamma_3 = o_p(1)$ and, in a similar fashion, it can be shown that $\Gamma_2 = o_p(1)$.
Turning to Γ_1,

$$\Gamma_1 = -(\mathcal{M}_1/T)\left[\mathcal{M}_1^{-1}\sum_{t\in\mathcal{T}_1}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\mathbf{w}_{t-1}'\right]\sqrt{T}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$$

where $\mathcal{M}_1 = \sum_{t=1}^{T}\mathcal{D}_{t1}$ is the number of observations for which $\mathcal{D}_{t1} = 1$ and $\mathcal{T}_1$ denotes the subsample of observations with $\mathcal{D}_{t1} = 1$. Now, $\mathcal{M}_1/T$ is the proportion of sample observations for which $\mathcal{D}_{t1} = 1$. In this case, $\Pr(\mathcal{D}_{t1} = 1) \to 1$, so that $\mathcal{M}_1/T \xrightarrow{p} 1$, and a mean value expansion of $\mathcal{M}_1^{-1}\sum_{t\in\mathcal{T}_1}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\mathbf{w}_{t-1}'$ reveals that it is $o_p(1)$. Hence, $\Gamma_1 = o_p(1)$, also. ■

Proof of Lemma 5. Specifically, $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ can be written as (with $\mathbf{c}_t^{\infty\prime} = (h_t^\infty)^{-1}\sum_{i=0}^{\infty}\beta_1^i\varepsilon_{t-1-i}\mathbf{f}'_{t-1-i}$),

$$\begin{aligned}\mathbf{J}_{\boldsymbol{\pi\varphi}} &= -2\alpha_{01}E\left\{E\left[\frac{1}{h_t^\infty}I_{t-1}\varepsilon_{t-1}^2\,|\mathcal{F}_{t-2}\right]\mathbf{f}'_{t-1}\right.\\ &\left.+\sum_{i=1}^{\infty}\beta_1^iE\left[\frac{1}{h_t^\infty}I_{t-1}\varepsilon_{t-1}\varepsilon_{t-1-i}\,|\mathcal{F}_{t-2}\right]\mathbf{f}'_{t-1-i}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\end{aligned}$$

which is non-zero (certainly, $E\left[\frac{1}{h_t^\infty}I_{t-1}\varepsilon_{t-1}^2|\mathcal{F}_{t-2}\right]$ is non-negative). ■

Proof of Lemma 6. The proof is similar to that of Lemma 3. Firstly, for non-negligible estimation effects from the conditional mean,

$$\mathbf{J}_{\boldsymbol{\pi\varphi}} = -2\sum_{k=1}^{q}\alpha_{0k}E\left[\frac{1}{h_t^{\infty 2}}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\varepsilon_{t-k-j}\mathbf{v}_{t-1-i}\mathbf{f}'_{t-k-j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

is non-zero, in general, if at least one element in $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ is non-zero. This amounts to examining the typical expectation $E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\varepsilon_{t-m}\mathbf{f}'_{t-m}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ for $l = m$, $l < m$ and $l > m$, where $s = 1, 3$. Firstly, for $l = m$,

$$E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\varepsilon_{t-l}\mathbf{f}'_{t-l}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left\{E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\varepsilon_{t-l}\middle|\mathcal{F}_{t-l-1}\right]\mathbf{f}'_{t-l}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

with the conditional expectation given by $E\left[(h_t^\infty)^{-2}\varepsilon_{t-l}^u|\mathcal{F}_{t-l-1}\right]$ for $u = 2, 4$. Similar to the arguments in Lemma 4, $E\left[(h_t^\infty)^{-2}\varepsilon_{t-l}^u|\mathcal{F}_{t-l-1}\right] > 0$ almost surely and thus $E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\varepsilon_{t-l}\mathbf{f}'_{t-l}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is non-zero. Further, for $l < m$,

$$E\left\{\mathbf{f}'_{t-m}E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\varepsilon_{t-m}\middle|\mathcal{F}_{t-m-1}\right]\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left\{\mathbf{f}'_{t-m}E\left[\varepsilon_{t-m}E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\middle|\mathcal{F}_{t-l-1}\right]\middle|\mathcal{F}_{t-m-1}\right]\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and similar arguments to those employed in the Proof of Lemma 3 establish that $E\left[(h_t^\infty)^{-2}\varepsilon_{t-l}^s|\mathcal{F}_{t-l-1}\right] = 0$ for $s = 1, 3$. For $l > m$, since $\mathcal{F}_{t-l-1} \subseteq \mathcal{F}_{t-m-1}$,

$$E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\varepsilon_{t-m}\mathbf{f}'_{t-m}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left\{\mathbf{f}'_{t-m}\varepsilon_{t-l}^sE\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-m}\middle|\mathcal{F}_{t-m-1}\right]\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

where the elements ε_{t-l}^s, $s = 1, 3$ belong to $\mathcal{F}_{t-m-1}$ and previous arguments show that $E\left[(h_t^\infty)^{-2}\varepsilon_{t-m}|\mathcal{F}_{t-m-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$.

Secondly, for $\mathbf{J}_{\boldsymbol{\pi\eta}}$, the expectation to be examined is $E\left[\frac{1}{h_t^{\infty 2}}\varepsilon_{t-l}^s\mathbf{s}_{t-m}^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, $s = 1, 3$ for $l = m$, $l > m$ and $l < m$ and arguments similar to those used in Lemma 3 show that $\mathbf{J}_{\boldsymbol{\pi\eta}} = \mathbf{0}$. In particular, notice that for $s = 1$, the above expectation was shown to be zero for all three cases in Lemma 3. Similar arguments also apply for $s = 3$. ■

Proof of Lemma 7. Firstly, with $\mathbf{c}_t^{\infty\prime} = (h_t^\infty)^{-1}\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-1-i}\mathbf{f}'_{t-1-i}$ and $\mathbf{r}_t^\infty = (h_t^\infty)^{-1}\sum_{i=0}^{\infty}\psi_i\mathbf{v}_{t-1-i}$, $\mathbf{J}_{\boldsymbol{\pi\varphi}}$ can be written as

$$\mathbf{J}_{\boldsymbol{\pi\varphi}} = -2\sum_{k=1}^{q}\alpha_{0k}E\left[\frac{1}{h_t^{\infty 2}}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\varepsilon_{t-k-j}\mathbf{v}_{t-1-i}\mathbf{f}'_{t-k-j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

For a typical element in $\mathbf{J}_{\boldsymbol{\pi\varphi}}$, the expectation to be examined is $E\left[\frac{1}{h_t^{\infty 2}}I_{t-l}\varepsilon_{t-l}\varepsilon_{t-m}\mathbf{f}'_{t-m}\right]$ for the cases $l = m$, $l < m$ and $l > m$. Consider just $l = m$; here we have $E\left[\frac{1}{h_t^{\infty 2}}I_{t-l}\varepsilon_{t-l}^2\mathbf{f}'_{t-l}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} =$

$E\left\{E\left[\frac{1}{h_t^2}I_{t-l}\varepsilon_{t-l}^2\middle|\mathcal{F}_{t-l-1}\right]\mathbf{f}'_{t-l}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is certainly non-zero.
Secondly, with $\mathbf{x}_t^{\infty\prime}=(h_t^{\infty})^{-1}\sum_{i=0}^{\infty}\psi_i\mathbf{s}_{t-1-i}^{\infty\prime}$, $\mathbf{J}_{\boldsymbol{\pi\eta}}$ in (11) can be written as

$$\mathbf{J}_{\boldsymbol{\pi\eta}}=E\left[\frac{1}{h_t^{\infty 2}}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\mathbf{v}_{t-1-i}\mathbf{s}_{t-1-j}^{\infty\prime}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

For a typical element in $\mathbf{J}_{\boldsymbol{\pi\eta}}$, the expectation to be examined is $E\left[\frac{1}{h_t^{\infty 2}}I_{t-l}\varepsilon_{t-l}\mathbf{s}_{t-m}^{\infty\prime}\right]$ for the cases $l=m$, $l<m$ and $l>m$. Similar arguments to those employed previously show that this is non-zero in general. ■

References

[1] Bai, J., Ng, S. (2001) A consistent test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225-258.

[2] Berkes, I., Horváth, L., Kokoszka, P. (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201-227.

[3] Black, F. (1976) Studies of stock prices volatility changes. *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, 177-181.

[4] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[5] Bollerslev, T., Wooldridge, J.M. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, 143-172.

[6] Chan, F., McAleer, M. (2002) Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence. *Journal of Applied Econometrics* 17, 509-534.

[7] Chan, F., McAleer, M. (2003) On the structure, asymptotic theory and applications of STAR-GARCH models. CIRJE-F-216 Discussion Paper, University of Western Australia.

[8] Comte, F., Lieberman, O. (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61-84.

[9] Cox, D.R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical. Society, Series B*, 49, 1-39 (with discussion).

[10] Durbin, J. (1970) Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica* 38, 410-421.

[11] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1008.

[12] Engle, R.F., Ng, V.K. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749-1777.

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

[14] French, K.R., Schwert, G.W., Stambaugh, R.F. (1987) Expected stock returns and volatility. *Journal of Financial Economics* 19, 3-30.

[15] Glosten, L., Jagannathan, R., Runkle, D. (1993) On the relation between expected value and the volatility of nominal excess returns on stocks. *Journal of Finance* 48, 1779-1801.

[16] Godfrey, L.G. (1996) Some results on the Glejser and Koenker tests for heteroskedasticity. *Journal of Econometrics* 72, 275-299.

[17] Godfrey, L.G., Orme, C.D., 1996. On the behavior of conditional moment tests in the presence of unconsidered local alternatives. *International Economic Review* 37, 263-281.

[18] Gonzalez-Rivera, G. (1998) Smooth Transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 3, 61-78.

[19] Greene, W.H. (2003) *Econometric Analysis* (5th Edition). Prentice-Hall.

[20] Hagerud, G.E. (1997) A smooth transition ARCH model for asset returns. SSE/EFI Working Paper Series in Economics and Finance, No. 162.

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

[22] Li, W.K., Mak, T.K. (1994) On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15, 627-636.

[23] Li, C.W., Li, W.K. (1996) On a double-threshold autoregressive heteroskedastic time series model. *Journal of Applied Econometrics* 11, 253-274.

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

[25] Lundbergh, S., Teräsvirta, T. (2002) Evaluating GARCH models. *Journal of Econometrics* 110, 417-435.

[26] Lundbergh, S., Teräsvirta, T. (1999) Modelling economic high frequency time series with STAR-STGARCH models. SSE/EFI Working Paper Series in Economics and Finance, No. 390.

[27] Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347-370.

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

[29] Newey, W.K. (1985) Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 53, 1047-1070.

[30] Newey, W.K., McFadden, D. (1994) Large sample estimation and hypothesis testing. In: Engle, R.F. and McFadden, D.L. (Eds.), *Handbook of Econometrics*, Vol. 4, pp. 2111-2245 (Chapter 36).

[31] Orme, C.D. (1991) The small-sample performance of the information-matrix test. *Journal of Econometrics* 46, 309-331.

[32] Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and Testing. *Econometric Theory* 2, 107-131.

[33] White, H. (2001) *Asymptotic Theory for Econometricians.* Academic Press: California

[34] Wooldridge, J. (1991) On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* 47, 5-46.

[35] Zakoian, J-M. (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931-955.

Table 1. Empirical size

	$N(0,1)$				$t(7)$				$t(5)$				$t(3)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1-H	4.4	5.2	5.1	2.6	4.2	4.1	4.0	1.7	5.8	4.4	2.2	1.0	5.9	2.7	2.4	0.5
M2-H	4.5	4.9	5.6	2.1	4.4	3.6	3.8	1.5	6.2	4.1	2.5	0.9	7.2	2.7	2.8	0.5
M3-H	4.6	5.0	5.4	2.5	4.4	3.7	3.7	1.6	6.7	4.0	2.5	1.0	7.2	2.6	2.7	0.5
M4-H	4.9	5.0	5.4	2.6	4.0	3.7	3.7	1.4	6.1	4.0	2.4	1.0	8.0	2.5	2.6	0.5
M1-M	4.8	4.7	4.8	4.0	4.9	4.0	4.2	3.1	5.6	4.1	2.9	2.0	5.5	2.9	2.8	1.1
M2-M	5.0	4.4	5.2	4.3	4.4	4.1	3.8	3.4	5.1	3.7	2.8	2.0	5.9	2.9	2.7	1.1
M3-M	5.0	4.6	5.4	4.3	4.5	3.9	4.0	3.4	5.0	4.0	2.8	2.0	6.0	2.9	2.7	1.1
M4-M	5.1	4.6	5.4	4.0	4.4	3.8	4.2	3.3	5.3	4.0	3.1	2.1	6.7	3.2	2.6	1.1
M1-L	6.0	4.7	4.2	4.1	4.6	3.8	3.8	2.6	4.8	4.1	2.5	1.4	6.5	2.0	1.9	0.5
M2-L	5.7	4.7	5.3	3.9	4.7	3.8	4.0	2.8	4.0	4.0	2.6	1.5	6.4	1.5	2.3	0.6
M3-L	5.7	4.4	5.2	4.0	4.8	4.0	4.3	2.5	4.0	3.8	2.6	1.6	6.1	1.8	2.3	0.7
M4-L	5.7	4.6	5.2	3.9	4.6	4.0	3.7	2.7	4.6	4.0	2.5	1.5	5.8	1.9	2.1	0.6

Table 1. Empirical power

GJR (1,1) model

$h_t = 0.005 + 0.23\left[|\varepsilon_{t-1}| - 0.23\varepsilon_{t-1}\right]^2 + 0.7h_{t-1}$

	$N(0,1)$				$t(7)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1	88.7	47.1	89.2	16.5	62.2	32.7	64.9	14.4
M2	84.6	38.3	87.4	14.5	59.0	27.1	59.0	12.2
M3	84.1	38.5	87.0	14.2	59.9	27.2	58.8	12.6
M4	84.6	37.9	87.8	16.6	58.7	28.0	59.7	12.8

$h_t = 0.005 + .23\left[|\varepsilon_{t-1}| - 0.17\varepsilon_{t-1}\right]^2 + 0.7h_{t-1}$

	$N(0,1)$				$t(7)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1	67.5	30.0	67.0	12.4	40.4	22.5	42.8	11.6
M2	62.1	26.1	65.3	10.5	39.7	19.7	37.3	10.2
M3	61.3	26.6	64.6	10.6	40.5	19.3	37.1	10.6
M4	61.3	24.5	66.2	12.2	40.2	20.0	37.6	10.9

STGARCH (1,1) model

$$h_t = 0.005 + 0.136\varepsilon_{t-1}^2 - 0.212F(\varepsilon_{t-1})\,\varepsilon_{t-1}^2 + 0.7h_{t-1}$$
$$F(\varepsilon_{t-1}) = \frac{1}{1+\exp(-100\varepsilon_{t-1})} - \frac{1}{2}$$

	$N(0,1)$				$t(7)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1	95.7	69.0	96.7	49.3	78.0	46.4	81.5	29.6
M2	95.8	64.7	97.0	45.9	75.0	43.4	80.6	27.1
M3	95.5	64.5	96.7	45.7	74.8	43.9	80.3	26.7
M4	95.2	63.9	96.5	47.3	75.9	45.5	81.3	26.9

$$h_t = 0.005 + 0.136\varepsilon_{t-1}^2 - 0.17F(\varepsilon_{t-1})\,\varepsilon_{t-1}^2 + 0.7h_{t-1}$$
$$F(\varepsilon_{t-1}) = \frac{1}{1+\exp(-100\varepsilon_{t-1})} - \frac{1}{2}$$

	$N(0,1)$				$t(7)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1	86.1	51.5	88.7	36.6	60.7	34.5	62.3	20.6
M2	84.7	46.9	88.2	41.8	59.0	34.0	62.0	18.9
M3	83.8	48.0	87.4	34.3	57.2	33.6	60.8	19.2
M4	84.1	46.9	87.2	35.2	59.1	34.2	62.3	19.3

EGARCH (1,1) model

$\log(h_t) = -0.23 + 0.9\log(h_{t-1}) + 0.25\left[|\xi_{t-1}| - 0.3\xi_{t-1}\right]$

	$N(0,1)$				$t(7)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1	83.9	29.7	76.5	35.9	64.9	23.3	52.8	23.8
M2	82.8	27.2	75.7	33.7	64.4	20.1	49.4	20.3
M3	82.1	27.0	75.5	34.2	63.1	19.9	49.3	20.9
M4	83.2	27.0	75.1	32.4	63.0	19.7	49.3	21.7

TGARCH (1,1) model

$\sqrt{h_t} = 0.07 + 0.081(1 - I_{t-1})|\varepsilon_{t-1}| + 0.193I_{t-1}|\varepsilon_{t-1}| + 0.831\sqrt{h_{t-1}}$

	$N(0,1)$				$t(7)$			
	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
M1	98.3	46.3	97.4	54.5	91.1	31.2	81.9	33.0
M2	98.8	40.4	96.7	48.5	90.1	26.6	80.6	29.3
M3	98.3	41.5	96.8	49.5	89.4	26.6	80.5	29.2
M4	98.9	40.3	96.9	49.9	89.6	27.1	81.0	29.7