Misspecification Tests for GARCH Models

Andreea G. Halunga and Chris D. Orme
Economic Studies
School of Social Sciences
University of Manchester
MANCHESTER M13 9PL
United Kingdom

October 10, 2004

Abstract

This paper develops a framework for the construction and analysis of misspecification tests for GARCH models and proposes new asymptotically valid and locally optimal tests of asymmetry and non-linearity. It is argued that the asymmetry test of Engle and Ng (1993) and non-linearity test of Lundbergh and Teräsvirta (2002) are neither asymptotically valid (since they ignore asymptotically non-negligible estimation effects) nor locally optimal (since they ignore the recursive nature of the conditional variance structure). The framework ecompasses conditional mean specifications estimated by either OLS, NLLS or QML, and it is shown that the GARCH misspecification tests can be asymptotically sensitive to unconsidered misspecification of the conditional mean. Monte Carlo results indicate the new tests are very powerful when compared with the previous tests proposed by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002).
JEL Classification: C12, C22

1 Introduction

Due to the widespread interest in stock market prices, a great deal of research has been undertaken on the specification of parametric models that can model the characteristics, and determinants, of stock prices. For example, the ARCH model, pioneered by Engle (1982), accommodates volatility clustering - a feature of stock returns first noticed by Mandelbrot (1963) and Fama (1970) and explained by the arrival and transmission of news. Subsequently, Bollerslev (1986) introduced the Generalized ARCH (GARCH) model, which nowadays represents a benchmark specification for all volatility models. The finance literature on volatility clustering has also documented an asymmetric behaviour of volatility to shocks. This feature manifests itself through negative shocks having higher impacts on volatility than equal positive shocks. The leverage effect suggested by Black (1976) and the risk premium investigated by French *et al.* (1987) are widely cited as possible explanations of this behaviour in financial markets. Several asymmetric and/or non-linear GARCH models have been proposed in order to model both volatility clustering and asymmetric effects of past shocks on volatility. Prominent among these models are: the EGARCH model of Nelson (1990); the GJR model of Glosten *et al* (1993); the TGARCH model of Zakoian (1994); and, the Smooth Transition GARCH (STGARCH) model of Hagerud (1997) and Gonzalez-Rivera (1998).

Since the GARCH specification is a parametric model in which a particular structure is imposed, it is important to perform misspecification tests to check that the model adequately represents the data. Bollerslev (1986) suggested a 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. Finally, Lundbergh and Teräsvirta (2002) recently proposed tests for remaining ARCH in standardised errors, linearity and parameter constancy. These test procedures, therefore, are important 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.

This paper develops a unifying framework for the construction and analysis of misspecification tests in GARCH models, based on the score principle and first order asymptotic analysis, from which new tests for asymmetry and non-linearity emerge. These tests represent developments over previous procedures in a number of respects. Firstly, either Ordinary Least Squares (OLS), Non-Linear Least Squares (NLLS) or Quasi-Maximum Likelihood (QML) estimation of the conditional mean is allowed for. Secondly, the tests are locally optimal, since they explicitly account for the recursive nature of the conditional heteroskedasticity, whereas those of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002), for example, do not. Thirdly, it is shown that there can be asymptotically non-negligible estimation effects, in the limit null distribution of the resultant test indicators, arising from the estimated conditional mean parameters. (This problem was addressed by Durbin, 1970, when testing for serial correlation with lagged dependent variables.) Importantly, these estimation effects can occur even when the estimated conditional mean parameters are orthogonal to estimated conditional heteroskedasticity parameters. Indeed, because of this orthogonality, such estimation effects appear to have been assumed away by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002), thus bringing into question the asymptotically invalidity of their test procedures. This also suggests that such tests will, therefore, be sensitive to misspecification of the conditional mean function. The results of a small Monte Carlo study reveal that, not only do, the new locally optimal tests have excellent power, compared with the tests of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002), but also that they will be sensitive to conditional mean misspecification.

The paper is organized as follows. Section 2 describes the null model, which is the GARCH specification, and briefly discusses estimation of the conditional mean by OLS, NLLS and QMLE which underpins the generic framework for the construction of misspecification tests in Section 3. In Section 4 the tests proposed by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002) are reviewed and new asymptotically valid and locally optimal tests for asymmetry and non-linearity are outlined using the methodology of Section 3. Section 5 presents a sensitivity analysis to local misspecification of the conditional mean. Section 6 presents some Monte Carlo evidence in support of the theoretical findings and Section 7 concludes.

2 The null GARCH model

The GARCH(1,1) model of Bollerslev (1986) represents the benchmark specification for modelling conditional volatility in financial and economic time series data and is employed as the null model for simplicity of derivations. Nevertheless, the results below can be easily generalized to a GARCH(p, q) model. The description of this null model and the assumptions imposed are presented below.

The conditional mean equation for the variable of interest, y_t, can contain lagged endogenous variables and/or strictly exogenous variables and is defined as:

$$y_t = f(\mathbf{w}_t; \boldsymbol{\varphi}) + \varepsilon_t \tag{1}$$

where $\mathbf{w}_t = (\mathbf{y}'_{t-1}, \mathbf{z}'_t)$, $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-l})' \in \Re^{l+1}$, exogenous variables are $\mathbf{z}_t = (z_{t1}, ..., z_{tk})' \in \Re^k$, $\boldsymbol{\varphi} = (\varphi_1, .., \varphi_r)'$ is a vector of mean parameters and $f(\mathbf{w}_t; \boldsymbol{\varphi})$ is at least twice continuously differentiable in $\boldsymbol{\varphi}$. The conditional mean of the observed time series y_t is $E[y_t|\mathcal{F}_{t-1}] = f(\mathbf{w}_t; \boldsymbol{\varphi})$, where $\mathcal{F}_{t-1}$ is a σ-field generated by the history of ε_t to date $t-1$. The innovation, ε_t, is the unanticipated shock at time t and is given by

$$\varepsilon_t = \xi_t h_t^{1/2} \tag{2}$$

where ξ_t is an $i.i.d$ sequence with mean zero and variance one, termed a standardised error process.

Assumption 1 The process y_t is strictly stationary and ergodic.

Assumption 2 (Conditional symmetry) $E\left[\xi_t^3\right] = 0$, i.e., $E\left[\varepsilon_t^3 | \mathcal{F}_{t-1}\right] = 0$.

The conditional symmetry assumption makes the estimated conditional mean and variance parameters asymptotically orthogonal, within a Quasi Maximum Likelihood (QML) framework; see Section 2.1. The conditional symmetry assumption can be tested using tests proposed by Bai and Ng (2001). The condition was also assumed by Lundbergh and Teräsvirta (2002).

Assumption 3 $E\left[\left(\xi_t^2-1\right)^2\right]=k_c-1$, where k_c is a finite constant.

Since the ξ_t are *i.i.d.*, the standardised errors are conditionally homokurtic with Assumption 3 yielding $E\left[\left(\frac{\varepsilon_t^2}{h_t}-1\right)^2|\mathcal{F}_{t-1}\right]=k_c-1$.

The conditional variance is specified as

$$h_t=\boldsymbol{\eta}'\mathbf{s}_t, \tag{3}$$

which is measurable with respect to $\mathcal{F}_{t-1}$, where $\boldsymbol{\eta}=(\alpha_0,\alpha_1,\beta_1)'$ and $\mathbf{s}_t=\left(1,\varepsilon_{t-1}^2,h_{t-1}\right)'$. Thus (1), (2) and (3) define the GARCH(1,1) model.

Assumption 4 The process h_t is strictly stationary and ergodic.

Sufficient conditions for h_t to be strictly positive are: $\alpha_0>0$, $\alpha_1\geq 0$, $\beta_1\geq 0$, whereas the necessary and sufficient condition for covariance stationarity of h_t is $\alpha_1+\beta_1<1$. Nelson (1990) showed that the necessary and sufficient condition for strict stationarity and ergodicity of the GARCH(1,1) model is $E\left[\ln\left(\alpha_1+\beta_1\xi_t^2\right)\right]<0$. However, this condition is weaker than covariance stationarity, but it also allows $\alpha_1+\beta_1=1$ and $\alpha_1+\beta_1$ slightly greater than one, and therefore it includes the IGARCH(1,1) model. Therefore the condition of covariance stationarity of h_t is sufficient for strict stationarity and ergodicity of the conditional variance process.

Assumption 5 $E\left(\xi_t^8\right)<\infty$.

This assumption is made by Comte and Lieberman (2003) in proving the asymptotic normality of the QML estimator of the GARCH model. Asymptotic theory for GARCH models was also considered by Bollerslev and Wooldridge (1992), Lee and Hansen (1994) and Lumsdaine (1996). For the present purposes, sufficient regularity is assumed so that appropriate Central Limit Theorems and a Uniform Law of Large Numbers can be applied in order to derive appropriate limiting distributions. Following Comte and Lieberman (2003) this requires, for example, the density of ξ_t to be absolutely continuous with respect to the Lebesgue measure and positive in the neighbourhood of the origin.

We define $\boldsymbol{\theta}' = (\boldsymbol{\varphi}', \boldsymbol{\eta}')$ and the quasi log-likelihood, based on normality, would be (ignoring constants)

$$L_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_t) + \frac{\varepsilon_t^2}{h_t}\right] \tag{4}$$

for a sample of T observations. Even if we assume normality in estimation, the true conditional distribution of ξ_t can be non-normal.

If the true distribution of ξ_t is symmetric then, as noted above, the estimated conditional mean and variance parameters of the GARCH model are asymptotically orthogonal. Engle (1982) presents a theorem for this result, and applied to the ARCH model, but since normality was assumed throughout his paper, the conditional symmetry assumption of ξ_t is not stated explicitly in the theorem. Asymptotic orthogonality implies that consistent estimation of $\boldsymbol{\eta}$ can be achieved based on any consistent estimator for $\boldsymbol{\varphi}$ and this suggests that tests designed to test the adequacy of h_t will not be influenced (asymptotically, at least) by the estimation of $\boldsymbol{\varphi}$; as when constructing tests for unconditional heteroskedasticity in the linear model, for example. However, this intuition is flawed for certain misspecification tests in the GARCH model; in particular, asymmetry and non-linearity tests.

The asymptotic properties of GARCH model parameter estimators based on OLS, NLLS and QML for estimating $\boldsymbol{\varphi}$, and QML for estimating $\boldsymbol{\eta}$, are briefly presented below with details being well documented elsewhere in the literature; see, for example, Weiss (1986) or Greene (2003). Throughout, the estimated parameter vector will be denoted $\hat{\boldsymbol{\theta}}' = \left(\hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\eta}}'\right)$, and true parameter value will be denoted by $\boldsymbol{\theta}_0' = (\boldsymbol{\varphi}_0', \boldsymbol{\eta}_0')$.

2.1 Parameter Estimation

2.1.1 OLS Estimation

Assuming a linear functional form for the conditional mean, i.e. $f(\mathbf{w}_t; \boldsymbol{\varphi}) = \mathbf{w}_t'\boldsymbol{\varphi}$, the conditional mean parameters can be estimated consistently by OLS. The asymptotic properties below are easily extended from the work of Weiss (1986) in which the errors of the linear conditional mean follow an ARCH process. (If the conditional mean is estimated by OLS, then Assumption 5 can be relaxed to $E\left(\xi_t^4\right) < \infty$; see also Gonçalves and Killian, 2004.) Standard OLS inferences are asymptotically valid if the $\mathbf{w}_t$ are fixed, otherwise (for example when lagged dependent variables are included) consistent standard errors for the conditional mean and variance parameters can be obtained by using the White's sandwich variance-covariance form, as suggested by Engle (1982). Note that the estimated conditional mean parameters, $\boldsymbol{\varphi}$, are consistent even without the conditional symmetry assumption, but this can result in asymptotically inefficient inferences.

The OLS estimator is given by $\hat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}_0 + (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\varepsilon}_0$, where $\boldsymbol{\varepsilon}_0$ has typical element $y_t - \mathbf{w}_t'\boldsymbol{\varphi}_0$ and $\mathbf{W}$ has rows $\mathbf{w}_t'$. Assuming that both

$$\mathbf{J}_{\varphi} = p\lim \frac{1}{T}\mathbf{W}'\mathbf{W} \tag{5}$$

and

$$\Omega_{11} = p\lim \frac{1}{T}\mathbf{W}'\mathbf{H}_0\mathbf{W}$$

are finite and positive definite, with $\mathbf{H}_0 = diag\left(\boldsymbol{\eta}_0'\mathbf{s}_t\right)$, and

$$\frac{1}{\sqrt{T}}\mathbf{W}'\boldsymbol{\varepsilon}_0 \xrightarrow{d} N\left(0, \Omega_{11}\right),$$

it follows that

$$\sqrt{T}\left(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{J}_{\varphi}^{-1}\Omega_{11}\mathbf{J}_{\varphi}^{-1}\right).$$

As noted by Engle (1982), since the variables, $\mathbf{w}_t\varepsilon_t$, are uncorrelated the asymptotic variance matrix can be estimated as $(\mathbf{W}'\mathbf{W})^{-1}\hat{\mathbf{W}}'\hat{\mathbf{E}}\hat{\mathbf{W}}(\mathbf{W}'\mathbf{W})^{-1}$, where $\hat{\mathbf{E}} = diag(\hat{\varepsilon}_t^2)$ with $\hat{\varepsilon}_t = y_t - \mathbf{w}_t'\hat{\boldsymbol{\varphi}}$; see, for example, White (2002, p.139) or Nicholls and Pagan (1983). However, an asymptotically more efficient variance matrix estimator could be obtained as $(\mathbf{W}'\mathbf{W})^{-1}\hat{\mathbf{W}}'\hat{\mathbf{H}}\hat{\mathbf{W}}(\mathbf{W}'\mathbf{W})^{-1}$, where $\hat{\mathbf{H}} = diag\left(\hat{\boldsymbol{\eta}}'\mathbf{s}_t\right)$ and $\hat{\boldsymbol{\eta}}$ is any consistent estimator for $\boldsymbol{\eta}$; see Section 2.1.3 below.

2.1.2 NLLS Estimation

Suppose we assume a non-linear functional form for the conditional mean. For example, Lundbergh and Teräsvirta (1999) proposed combining a non-linear specification for the conditional mean with a GARCH model for the conditional variance, i.e., the STAR-GARCH model, and used it in forecasting economic time series. The statistical properties of this model were investigated by Chan and McAleer (2002). When the form of transition function (including the transition variable) is known, and given the conditional symmetry assumption, the STAR model for the conditional mean can be estimated consistently by NLLS. This estimator minimizes $\sum_{t=1}^{T}\left(y_t - f\left(\mathbf{w}_t; \boldsymbol{\varphi}\right)\right)^2$ and shall also be denoted $\hat{\boldsymbol{\varphi}}$, where there is no ambiguity.

Here, adapt the notation so that $\mathbf{F}$ is the matrix with rows $\mathbf{f}_t' = \partial f\left(\mathbf{w}_t; \boldsymbol{\varphi}\right)/\partial\boldsymbol{\varphi}'$, with $\hat{\mathbf{F}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ where $\hat{\boldsymbol{\varepsilon}}$ has typical element $y_t - f\left(\mathbf{w}_t; \hat{\boldsymbol{\varphi}}\right)$. In general, the following Central Limit Theorem (CLT) will apply

$$\frac{1}{\sqrt{T}}\mathbf{F}_0'\boldsymbol{\varepsilon}_0 \xrightarrow{d} N\left(\mathbf{0}, \Omega_{11}^*\right),$$

where $\mathbf{F}_0$ denotes evaluation at $\boldsymbol{\varphi}_0$, and

$$\Omega_{11}^{*} = p\lim \frac{1}{T}\mathbf{F}_0'\mathbf{H}_0\mathbf{F}_0$$

is assumed to be finite and positive definite.

Therefore, after taking a mean value expansion of $\hat{\mathbf{F}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ about $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$, and noting that $E\left[\varepsilon_t|\mathcal{F}_{t-1}\right] = 0$, we obtain

$$\sqrt{T}\left(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{J}_{\varphi}^{*\,-1}\Omega_{11}^{*}\mathbf{J}_{\varphi}^{*\,-1}\right)$$

where

$$\mathbf{J}_{\varphi}^{*} = p\lim \frac{1}{T}\mathbf{F}_0'\mathbf{F}_0$$

is finite and positive definite. As before, due to conditional symmetry, the asymptotic variance of $\hat{\boldsymbol{\varphi}}$ can be consistently estimated by replacing $\boldsymbol{\eta}$ with any consistent estimator, such as the QML estimator, yielding (for example) $(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\hat{\mathbf{H}}\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}$.

2.1.3 Quasi-Maximum Likelihood Estimation

The conditional mean and variance parameters can be estimated jointly by quasi-maximum likelihood estimation. If the true conditional distribution of ξ_t is normal, then the QML estimator is interpreted as the Maximum Likelihood Estimator (MLE) and the estimators are asymptotically efficient. Under non-normality, conditional symmetry of ε_t is still maintained as is (for simplicity) a linear conditional mean. Ignoring constants of proportionality, let $s_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \dfrac{\partial L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, partitioned as $s_{\boldsymbol{\theta}}(\boldsymbol{\theta})' = \left(s_{\boldsymbol{\varphi}}(\boldsymbol{\theta})', s_{\boldsymbol{\eta}}(\boldsymbol{\theta})'\right)'$, in an obvious manner. In general, a CLT for the score function of the conditional mean parameters yields:

$$\sqrt{T}s_{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, \Omega_{11}^{**}\right)$$

where $s_{\boldsymbol{\varphi}}(\boldsymbol{\theta}) = T^{-1}\left(\mathbf{W}'\mathbf{H}^{-1}\boldsymbol{\varepsilon} + \frac{1}{2}\mathbf{C}'\boldsymbol{\vartheta}\right)$, $\mathbf{H}^{-1} = diag\left(h_t^{-1}\right)$ and $\mathbf{C}$ has rows $\mathbf{c}_t' = \dfrac{1}{h_t}\dfrac{\partial h_t}{\partial \boldsymbol{\varphi}'}$ with

$$\begin{aligned}\frac{\partial h_t}{\partial \boldsymbol{\varphi}'} &= -2\alpha_1\varepsilon_{t-1}\mathbf{w}_{t-1}' + \beta_1\frac{\partial h_{t-1}}{\partial \boldsymbol{\varphi}'} \\ &= -2\alpha_1\sum_{i=1}^{t}\beta_1^{i-1}\varepsilon_{t-i}\mathbf{w}_{t-i}'.\end{aligned}$$

The covariance matrix

$$\Omega_{11}^{**} = p\lim \frac{1}{T}\left(\mathbf{W}'\mathbf{H}_0^{-1}\mathbf{W} + \frac{(k_c-1)}{4}\mathbf{C}_0'\mathbf{C}_0\right)$$

is assumed finite and positive definite, with $\mathbf{C}_0$ denoting $\mathbf{C}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. A mean value expansion of $s_{\boldsymbol{\varphi}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, and conditional symmetry, yields

$$\sqrt{T}\left(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{J}_{\boldsymbol{\varphi}}^{**-1}\Omega_{11}^{**}\mathbf{J}_{\boldsymbol{\varphi}}^{**-1}\right)$$

where

$$\mathbf{J}_{\varphi}^{**} = p\lim \frac{1}{T}\left(\mathbf{W}'\mathbf{H}_0^{-1}\mathbf{W} + \frac{1}{2}\mathbf{C}_0'\mathbf{C}_0\right)$$

is assumed finite and positive definite.

2.1.4 GARCH Parameter Estimation

Exploiting the QML approach and consistency of $\hat{\boldsymbol{\varphi}}$ (the OLS, NLLS or QML estimator) a consistent estimator for $\boldsymbol{\eta}$ solves $s_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, where $s_{\boldsymbol{\eta}}(\boldsymbol{\theta}) = T^{-1}\mathbf{X}'\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}$ has typical element $\left\{\frac{\varepsilon_t^2}{h_t} - 1\right\}$, and $\mathbf{X}$ has rows

$$\begin{aligned} \mathbf{x}_t' &= \frac{1}{h_t}\frac{\partial h_t}{\partial \boldsymbol{\eta}'} \\ &= \frac{1}{h_t}\sum_{i=1}^{t}\beta_1^{i-1}\left(1, \varepsilon_{t-i}^2, h_{t-i}\right) \end{aligned}$$

assuming $h_0 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2$. Denoting $\mathbf{X}_0$ and $\boldsymbol{\vartheta}_0$ to be $\mathbf{X}$ and $\boldsymbol{\vartheta}$, respectively, evaluated at $\boldsymbol{\theta}_0$, a CLT for the score function of the conditional variance parameters yields:

$$\sqrt{T}s_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}}\mathbf{X}_0'\boldsymbol{\vartheta}_0 \xrightarrow{d} N\left(0, \Omega_{22}\right)$$

where

$$\Omega_{22} = p\lim \frac{k_c - 1}{T}\mathbf{X}_0'\mathbf{X}_0$$

is assumed to be finite and positive definite. Furthermore, and as stated previously, the asymptotic distribution of $\hat{\boldsymbol{\eta}}$ is not influenced by the choice for $\hat{\boldsymbol{\varphi}}$ (OLS, NLLS or QML). In order to inform the analysis that follows in Sections 3 and 4, it is worth briefly showing why this is so. Noting that, $\dfrac{\partial s_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'} - E\left[\dfrac{\partial s_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] \xrightarrow{p} \mathbf{0}$, conditional symmetry ensures that this expectation is zero, as follows:

$$\begin{aligned} E\left[\frac{\partial s_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{T}\sum_{t=1}^{T}E\left[E\left\{\frac{2\varepsilon_t}{h_t}\mathbf{x}_t\frac{\partial \varepsilon_t}{\partial \boldsymbol{\varphi}'} + \left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial \mathbf{x}_t}{\partial \boldsymbol{\varphi}'} - \frac{\varepsilon_t^2}{h_t^2}\mathbf{x}_t\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right|\mathcal{F}_{t-1}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= -\frac{1}{T}\sum_{t=1}^{T}E\left[\frac{1}{h_t}\mathbf{x}_t\mathbf{c}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{aligned}$$

where the subscript $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ denotes evaluation inside [.] at $\boldsymbol{\theta}_0$ before expectations are taken, and $E\left[\varepsilon_t|\mathcal{F}_{t-1}\right] = 0$, $E\left[\varepsilon_t^2|\mathcal{F}_{t-1}\right] = h_t$.

Now,

$$
\begin{aligned}
E\left[\frac{1}{h_t}\mathbf{c}_t\mathbf{x}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -2\alpha_1\sum_{i=1}^{t}\beta_1^{2(i-1)}E\left[\frac{1}{h_t^2}\left(\varepsilon_{t-i},\,\varepsilon_{t-i}^3,\,h_{t-i}\varepsilon_{t-i}\right)\mathbf{w}_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\quad -4\alpha_1\sum_{i=1}^{t}\sum_{j<i}^{t}\beta_1^{i+j-2}E\left[\frac{1}{h_t^2}\left(1,\,\varepsilon_{t-i}^2,\,h_{t-i}\right)\varepsilon_{t-j}\mathbf{w}_{t-j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}' \qquad (6)
\end{aligned}
$$

Consider the expectation on the right hand side of the first line above,

$$
E\left[\frac{1}{h_t^2}\left(1,\,\varepsilon_{t-i}^2,\,h_{t-i}\right)\varepsilon_{t-j}\mathbf{w}_{t-j}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left\{E\left[\frac{1}{h_t^2}\left(\varepsilon_{t-i},\,\varepsilon_{t-i}^3,\,h_{t-i}\varepsilon_{t-i}\right)\middle|\,\mathcal{F}_{t-i-1}\right]\mathbf{w}_{t-j}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
$$

which is zero if the conditional expectation, given $\mathcal{F}_{t-i-1}$, in the second line is zero. To establish the latter, follow Engle (1983) and treat this conditional expectation in two steps, observing that ε_{t-i-m}, $m = 1, 2, ...,$ is included in the conditioning set of $\mathcal{F}_{t-i-1}$ and therefore can be treated as non-random when taking this conditional expectation. First, construct the conditional expectation given $\mathcal{F}_{t-i}$, which is

$$
\left[\left(\varepsilon_{t-i},\,\varepsilon_{t-i}^3,\,h_{t-i}\varepsilon_{t-i}\right)E\left\{\frac{1}{h_t^2}\middle|\,\mathcal{F}_{t-i}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \equiv g(\varepsilon_{t-i}),
$$

where it is implicit that $g(.)$ is evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Since h_t is symmetric in ε_{t-i} and ε_{t-i}, ε_{t-i}^3, $h_{t-i}\varepsilon_{t-i}$ are all anti-symmetric, the elements in $h_t^{-2}\left(\varepsilon_{t-i},\,\varepsilon_{t-i}^3,\,h_{t-i}\varepsilon_{t-i}\right)$ are anti-symmetric in ε_{t-i}, which forms part of $\mathcal{F}_{t-i}$ and is, also, the only random element when, and at the second step, expectations are taken with respect to $\mathcal{F}_{t-i-1}$. Therefore, for the present argument, $h_t^{-2}\left(\varepsilon_{t-i},\,\varepsilon_{t-i}^3,\,h_{t-i}\varepsilon_{t-i}\right)$ can be regarded as a random function of (h_t, ε_{t-i}). Now, because h_t is symmetric in ε_{t-i} its conditional density given ε_{t-i} is also symmetric in ε_{t-i}. Therefore, by Engle (1983, Lemma p.1006), $g(\varepsilon_{t-i})$ is anti-symmetric in ε_{t-i}. Finally, the second step involves $E\left[g(\varepsilon_{t-i})|\mathcal{F}_{t-i-1}\right]$ which is zero, because the conditional density of ε_{t-i} given $\mathcal{F}_{t-i-1}$ is symmetric and $g(.)$ is anti-symmetric.

The expectation in the cross-product term in (6) is

$$
E\left\{E\left[\frac{1}{h_t^2}\left(1,\,\varepsilon_{t-i}^2,\,h_{t-i}\right)\varepsilon_{t-j}\mathbf{w}_{t-j}\middle|\,\mathcal{F}_{t-i}\right]\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
$$

and, because $j < i$, the conditional expectation given $\mathcal{F}_{t-i}$ can be expressed as,

$$
E\left[\frac{1}{h_t^2}\left(1,\,\varepsilon_{t-i}^2,\,h_{t-i}\right)\varepsilon_{t-j}\mathbf{w}_{t-j}\middle|\,\mathcal{F}_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = E\left[\left(1,\,\varepsilon_{t-i}^2,\,h_{t-i}\right)E\left\{\frac{1}{h_t^2}\varepsilon_{t-j}\mathbf{w}_{t-j}\middle|\,\mathcal{F}_{t-i}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.
$$

Since $\mathcal{F}_{t-i} \subset \mathcal{F}_{t-j}$, and for $i = j+1$,

$$\begin{aligned} E\left[\frac{1}{h_t^2}\varepsilon_{t-j}\mathbf{w}_{t-j}\right|\left.\mathcal{F}_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \mathbf{w}_{t-j}E\left[E\left\{\frac{1}{h_t^2}\varepsilon_{t-j}\right|\left.\mathcal{F}_{t-j}\right\}\right|\left.\mathcal{F}_{t-j-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \mathbf{w}_{t-j}E\left[E\left\{g(\varepsilon_{t-j})|\mathcal{F}_{t-j-1}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \end{aligned}$$

Previous arguments reveal that $g(\varepsilon_{t-j})$ is anti-symmetric in ε_{t-j}, so that conditional symmetry yileds $E\left[g(\varepsilon_{t-j})|\mathcal{F}_{t-j-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$. More generally, for $j < i-1$,

$$\begin{aligned} E\left[\frac{1}{h_t^2}\varepsilon_{t-j}\mathbf{w}_{t-j}\right|\left.\mathcal{F}_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= E\left[\mathbf{w}_{t-j}\ E\left\{\frac{1}{h_t^2}\varepsilon_{t-j}\right|\left.\mathcal{F}_{t-j-1}\right\}\right|\left.\mathcal{F}_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= E\left[\mathbf{w}_{t-j}E\left\{g(\varepsilon_{t-j})|\mathcal{F}_{t-j-1}\right\}|\mathcal{F}_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \end{aligned}$$

which is again zero since $E\left[g(\varepsilon_{t-j})|\mathcal{F}_{t-j-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$.

Similar analysis also reveals that, under conditional symmetry,

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2}{h_t}-1\right)^2\mathbf{x}_t\mathbf{c}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= (k_c-1)\,E\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{c}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \mathbf{0} \end{aligned}$$

making $s_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0)$ and $s_{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)$ asymptotically uncorrelated, although this is not needed to establish that the limit null distribution of $\hat{\boldsymbol{\eta}}$ is not influenced by $\hat{\boldsymbol{\varphi}}$.

Using this result, a standard mean value expansion of $s_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ about $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ (where $\hat{\boldsymbol{\varphi}}$ is the OLS, NLLS or QML estimator) yields

$$\sqrt{T}\left(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{J}_{\boldsymbol{\eta}}^{-1}\Omega_{22}\mathbf{J}_{\boldsymbol{\eta}}^{-1}\right)$$

where

$$\mathbf{J}_{\boldsymbol{\eta}} = p\lim\frac{1}{T}\mathbf{X}_0'\mathbf{X}_0 \tag{7}$$

is finite and positive definite. Although estimated separately, and as pointed out by Engle (1982), the process could be iterated without affecting the asymptotic distribution of the resulting estimators for $\boldsymbol{\varphi}$ and $\boldsymbol{\eta}$.

Having obtained a consistent estimator for $\boldsymbol{\eta}$ in this way, consistent estimation of the asymptotic variance of $\hat{\boldsymbol{\varphi}}$ (including OLS and NLLS) follows from previous discussion. When $\hat{\boldsymbol{\varphi}}$ is the QML estimator, this can be estimated as

$$\left(\mathbf{W}'\hat{\mathbf{H}}^{-1}\mathbf{W}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}\left(\mathbf{W}'\hat{\mathbf{H}}^{-1}\mathbf{W}+\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{4T}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)\left(\mathbf{W}'\hat{\mathbf{H}}^{-1}\mathbf{W}+\frac{1}{2}\hat{\mathbf{C}}'\hat{\mathbf{C}}\right)^{-1}$$

where $\hat{\mathbf{H}}^{-1}$and $\hat{\mathbf{C}}$ denote $\mathbf{H}$ and $\mathbf{C}$, respectively, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, and (k_c-1) is consistently estimated by $\hat{k}_c - 1 = T^{-1}\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}$, where $\hat{\boldsymbol{\vartheta}}$ denotes evaluation at $\hat{\boldsymbol{\theta}}$.

3 Generic Misspecification Test Statistic

In this section, a general misspecification test for GARCH models is developed. A certain intellectual and notational economy is achieved by employing this approach, since it provides a unifying framework for generating misspecification tests. In particular, asymptotically valid and locally optimal tests for asymmetry and non-linearity emerge as special cases. To proceed, define $\xi_t = \varepsilon_t/\sqrt{h_t}$, the standardised error.

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 conditional moment tests, of the form:

$$E\left[\left(\xi_t^2 - 1\right)\mathbf{r}\left(\mathcal{F}_{t-1}\right)\right] = 0 \tag{8}$$

where $\mathbf{r}$ is a measurable function of $\mathcal{F}_{t-1}$. The intuition, here, is that if the GARCH model is appropriate, then the squared standardised errors should be serially uncorrelated with any past information. For example, Lundbergh and Teräsvirta (2002) employ a similar approach in order to test for no remaining ARCH effects, in a GARCH model, but where the implicit null is $E\left[\left(\xi_t^2 - 1\right)|\mathcal{G}_{t-1}\right] = 0$, with $\mathcal{G}_{t-1} = \sigma\left(\xi_{t-1}^2, .., \xi_{t-m}^2\right)$; see the discussion in Section 3.1 of Lundbergh and Teräsvirta (2002). However, this moment condition does not generally guarantee that $E\left[\left(\xi_t^2 - 1\right)|\mathcal{F}_{t-1}\right] = 0$, which implies that the Lundbergh and Teräsvirta test could have lower power than a test for which the null is $E\left[\left(\xi_t^2 - 1\right)|\mathcal{F}_{t-1}\right] = 0$.

For the present purposes, consider the construction of a rather general test procedure, and statistic, designed to assess whether the GARCH(1,1) model, presented in Section 2, is misspecified. Specifically, the misspecified GARCH model is conceived as:

$$\begin{aligned} y_t &= f\left(\mathbf{w}_t;\boldsymbol{\varphi}\right) + \varepsilon_t \\ \varepsilon_t &= \varsigma_t\left(h_t^a\right)^{1/2} \qquad (9) \\ h_t^a &= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + g\left(\mathbf{v}_t;\boldsymbol{\pi}\right) + \beta_1 h_{t-1}^a \qquad (10) \end{aligned}$$

where ς_t are *i.i.d.* (zero mean and unit variance) random variables, and $g\left(\mathbf{v}_t;\boldsymbol{\pi}\right) = \boldsymbol{\pi}'\mathbf{v}_t$ is a non-linear and/or asymmetric function of ε_{t-1}, with $\mathbf{v}_t$ being a vector of omitted variables; for example $\mathbf{v}_t' = \left(\varepsilon_{t-1}, \varepsilon_{t-1}^3\right)$. The null hypothesis is $H_0 : \boldsymbol{\pi} = \mathbf{0}$, which must imply $g\left(\mathbf{v}_t;\boldsymbol{\pi}\right) = 0$. Observe that the alternative specification includes lagged h_t^a, which would appear to be appropriate for the more general model.

For the moment, define the conditional quasi log-likelihood under the alternative hypothesis as:

$$L_T^A(\boldsymbol{\theta};\boldsymbol{\pi}) = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_t^a) + \frac{\varepsilon_t^2}{h_t^a}\right]$$

where $\boldsymbol{\theta}' = (\boldsymbol{\varphi}', \boldsymbol{\eta}')$ as previously. The generic test indicator emerges from the score principle and has the form of (ignoring constants of proportionality)

$$d_T(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right)\hat{\mathbf{r}}_t\right] = \frac{1}{T}\hat{\mathbf{R}}'\hat{\boldsymbol{\vartheta}} \tag{11}$$

where the matrix $\mathbf{R}$ has rows

$$\begin{aligned} \mathbf{r}_t &= \left[\frac{1}{h_t^a}\frac{\partial h_t^a}{\partial \pi}\right]_{\pi=0} \\ &= \frac{1}{h_t}\sum_{i=1}^{t}\beta_1^{i-1}\mathbf{v}_{t-i+1} \end{aligned} \tag{12}$$

which are the test variables (described later in Section 4 for each source of misspecification) and where "hats" denote that everything is evaluated at the consistent null parameter estimator, $\hat{\boldsymbol{\theta}}$, estimated according to OLS, NLLS or QML methods, as discussed previously. Although constructed from a score principle, the test indicator in (11) is simply the sample analogue of (8). Thus assessing the statistical significance of (11) provides the basis for a test procedure, following all three methods of estimation.

A mean value expansion of the test indicator $\sqrt{T}d_T(\hat{\boldsymbol{\theta}})$ in (11) about $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$, yields:

$$\sqrt{T}d_T(\hat{\boldsymbol{\theta}}) = \sqrt{T}d_T(\boldsymbol{\theta}_0) + \frac{\partial d_T\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\eta}'}\sqrt{T}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) + \frac{\partial d_T\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\varphi}'}\sqrt{T}\left(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\right) \tag{13}$$

where $\bar{\boldsymbol{\theta}}$ denotes the usual mean value "between" $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, so that $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + o_p(1)$ under the null hypothesis. A Uniform Law of Large Numbers ensures that $\frac{\partial d_T(\bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} - E\left[\frac{\partial d_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] \xrightarrow{p} 0$, where unless stated otherwise all expectations are taken under the null. Correspondingly define

$$\mathbf{D}_{\boldsymbol{\varphi}} = E\left[\frac{\partial d_T\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\varphi}'}\right], \qquad \mathbf{D}_{\boldsymbol{\eta}} = E\left[\frac{\partial d_T\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\eta}'}\right].$$

Using similar arguments to those of Section 2.1.4,

$$\begin{aligned} \mathbf{D}_{\boldsymbol{\varphi}} &= \frac{1}{T}\sum_{t=1}^{T}E\left[\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial \mathbf{r}_t}{\partial \boldsymbol{\varphi}'} - \frac{2\varepsilon_t}{h_t}\mathbf{r}_t\mathbf{w}_t' - \frac{\varepsilon_t^2}{h_t^2}\mathbf{r}_t\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \frac{1}{T}\sum_{t=1}^{T}E\left[-\frac{\varepsilon_t^2}{h_t^2}\mathbf{r}_t\frac{\partial h_t}{\partial \boldsymbol{\varphi}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \frac{1}{T}\sum_{t=1}^{T}E\left[-\mathbf{r}_t\mathbf{c}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= -p\lim\frac{1}{T}\mathbf{R}_0'\mathbf{C}_0, \end{aligned} \tag{14}$$

and

$$
\begin{aligned}
\mathbf{D}_{\boldsymbol{\eta}} &= \frac{1}{T}\sum_{t=1}^{T} E\left[\left(\frac{\varepsilon_t^2}{h_t}-1\right)\frac{\partial \mathbf{r}_t}{\partial \boldsymbol{\eta}'}-\frac{\varepsilon_t^2}{h_t^2}\mathbf{r}_t\frac{\partial h_t}{\partial \boldsymbol{\eta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left[-\mathbf{r}_t\mathbf{x}_t'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= -p\lim \frac{1}{T}\mathbf{R}_0'\mathbf{X}_0. \qquad (15)
\end{aligned}
$$

In Section 4, we investigate $\mathbf{D}_{\varphi}$ and $\mathbf{D}_{\boldsymbol{\eta}}$ for the particular test indicators, through choice of $\mathbf{v}_t$, designed to detect asymmetries and non-linearities. Before doing so, though, the asymptotically valid test statistic is now constructed according to the method of estimation under the null. For this, consistent estimation of both $\mathbf{D}_{\varphi}$ and $\mathbf{D}_{\boldsymbol{\eta}}$ may be required in order to capture potential asymptotically non-negligible estimation effects. Where necessary, this can be achieved by employing $-\frac{1}{T}\hat{\mathbf{R}}'\hat{\mathbf{C}}$ and $-\frac{1}{T}\hat{\mathbf{R}}'\hat{\mathbf{X}}$, respectively.

3.1 OLS Estimation

Assuming that the conditional mean is estimated by OLS under the null, and exploiting conditional symmetry, a suitable CLT yields

$$
\frac{1}{\sqrt{T}}\begin{pmatrix}\mathbf{W}'\boldsymbol{\varepsilon}_0\\ \mathbf{X}_0'\boldsymbol{\vartheta}_0\\ \mathbf{R}_0'\boldsymbol{\vartheta}_0\end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix},\begin{pmatrix}\Omega_{11} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \Omega_{22} & \Omega_{32}'\\ \mathbf{0} & \Omega_{32} & \Omega_{33}\end{pmatrix}\right)
$$

where $\mathbf{R}_0$ denotes evaluation at $\boldsymbol{\theta}_0$ and

$$
\begin{aligned}
\Omega_{32} &= (k_c-1)\, p\lim \frac{1}{T}\mathbf{R}_0'\mathbf{X}_0 \qquad (16)\\
\Omega_{33} &= (k_c-1)\, p\lim \frac{1}{T}\mathbf{R}_0'\mathbf{R}_0. \qquad (17)
\end{aligned}
$$

Standard first order asymptotic theory, the discussion in Section 2.1.1 and (13) yields

$$
\begin{aligned}
\sqrt{T}d_T(\hat{\boldsymbol{\theta}}) &= \frac{1}{\sqrt{T}}\mathbf{R}_0'\boldsymbol{\vartheta}_0 \\
&\quad +\mathbf{D}_{\varphi}\mathbf{J}_{\varphi}{}^{-1}\frac{1}{\sqrt{T}}\mathbf{W}'\boldsymbol{\varepsilon}_0 \\
&\quad +\mathbf{D}_{\boldsymbol{\eta}}\mathbf{J}_{\boldsymbol{\eta}}{}^{-1}\frac{1}{\sqrt{T}}\mathbf{X}_0'\boldsymbol{\vartheta}_0 + o_p(1),
\end{aligned}
$$

which implies

$$
\sqrt{T}d_T(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{T}}\hat{\mathbf{R}}'\hat{\boldsymbol{\vartheta}} \xrightarrow{d} N(\mathbf{0},\Sigma)
$$

where

$$\begin{aligned}
\Sigma &= \mathbf{A}\Omega\mathbf{A}' \qquad (18)\\
\Omega &= \begin{bmatrix} \Omega_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} & \Omega_{32}' \\ \mathbf{0} & \Omega_{32} & \Omega_{33} \end{bmatrix}\\
\mathbf{A} &= \begin{bmatrix} \mathbf{D}_{\varphi}\mathbf{J}_{\varphi}{}^{-1}, & \mathbf{D}_{\boldsymbol{\eta}}\mathbf{J}_{\boldsymbol{\eta}}{}^{-1} & \mathbf{I}_m \end{bmatrix}
\end{aligned}$$

and $\mathbf{I}_m$ is the identity matrix of rank $m = rank(\Omega_{33})$, and it is assumed that Ω is positive definite. Therefore, the general form of the misspecification test statistic is

$$Td_T(\hat{\boldsymbol{\theta}})'\hat{\Sigma}^{-1}d_T(\hat{\boldsymbol{\theta}})$$

which converges in distribution to that of a χ^2_m random variable under the null, where $\hat{\Sigma}$ is any consistent estimator of Σ, i.e. $\hat{\Sigma} = \Sigma + o_p(1)$. Such a consistent estimator can be obtained by estimating $\mathbf{D}_{\varphi}$ and $\mathbf{D}_{\boldsymbol{\eta}}$ as described above, estimating Ω_{11}, Ω_{22}, $\mathbf{J}_{\varphi}$ and $\mathbf{J}_{\boldsymbol{\eta}}$ in the manner described in Section 2.1, and estimating Ω_{32} as $\left(\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\right)\frac{1}{T}\hat{\mathbf{R}}'\hat{\mathbf{X}}$ and Ω_{33} as $\left(\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\right)\frac{1}{T}\hat{\mathbf{R}}'\hat{\mathbf{R}}$.

3.2 NLLS Estimation

If the conditional mean is estimated by NLLS, then the following CLT is required

$$\frac{1}{\sqrt{T}}\begin{pmatrix} \mathbf{F}_0'\boldsymbol{\varepsilon}_0 \\ \mathbf{X}_0'\boldsymbol{\vartheta}_0 \\ \mathbf{R}_0'\boldsymbol{\vartheta}_0 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Omega_{11}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} & \Omega_{32}' \\ \mathbf{0} & \Omega_{32} & \Omega_{33} \end{pmatrix}\right).$$

Again, standard first order theory arguments yield

$$\sqrt{T}d_T(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{T}}\hat{\mathbf{R}}'\hat{\boldsymbol{\vartheta}} \xrightarrow{d} N(\mathbf{0}, \Sigma^*)$$

where

$$\begin{aligned}
\Sigma^* &= \mathbf{A}^*\Omega^*\mathbf{A}^{*\prime} \qquad (19)\\
\Omega^* &= \begin{bmatrix} \Omega_{11}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} & \Omega_{32}' \\ \mathbf{0} & \Omega_{32} & \Omega_{33} \end{bmatrix}\\
\mathbf{A}^* &= \begin{bmatrix} \mathbf{D}_{\varphi}\mathbf{J}_{\varphi}^*{}^{-1}, & \mathbf{D}_{\boldsymbol{\eta}}\mathbf{J}_{\boldsymbol{\eta}}{}^{-1} & \mathbf{I}_m \end{bmatrix}.
\end{aligned}$$

Therefore, the generic misspecification test, and its limit distribution under the null, is:

$$Td_T(\hat{\boldsymbol{\theta}})'\hat{\Sigma}^{*-1}d_T(\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2_m$$

where $\hat{\Sigma}^*$ is any consistent estimator of Σ, which can be obtained in an obvious manner following previous discussions.

3.3 QML Estimation

Finally, if the conditional mean is estimated by QML, then the following CLT is exploited

$$\frac{1}{\sqrt{T}}\left(\begin{array}{c}\mathbf{W}'\mathbf{H}_0^{-1}\boldsymbol{\varepsilon}_0+\frac{1}{2}\mathbf{C}_0'\boldsymbol{\vartheta}_0\\ \mathbf{X}_0'\boldsymbol{\vartheta}_0\\ \mathbf{R}_0'\boldsymbol{\vartheta}_0\end{array}\right)\xrightarrow{d}N\left(\left(\begin{array}{c}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{array}\right),\left(\begin{array}{ccc}\Omega_{11}^{**} & \mathbf{0} & \Omega_{31}^{**\prime}\\ \mathbf{0} & \Omega_{22} & \Omega_{32}'\\ \Omega_{31}^{**} & \Omega_{32} & \Omega_{33}\end{array}\right)\right)$$

with

$$\Omega_{31}^{**}=p\lim\frac{1}{T}\frac{k_c-1}{2}\mathbf{R}_0'\mathbf{C}_0.$$

(The fact that the elements of $\mathbf{W}'\mathbf{H}_0^{-1}\boldsymbol{\varepsilon}_0+\frac{1}{2}\mathbf{C}_0'\boldsymbol{\vartheta}_0$ and $\mathbf{X}_0'\boldsymbol{\vartheta}_0$ are asymptotically uncorrelated is discussed in Section 2.1.4, although there may exist asymptotic correlation between $\mathbf{W}'\mathbf{H}_0^{-1}\boldsymbol{\varepsilon}_0+\frac{1}{2}\mathbf{C}_0'\boldsymbol{\vartheta}_0$ and $\mathbf{R}_0'\boldsymbol{\vartheta}_0$.) Standard first order theory yields

$$\sqrt{T}d_T(\hat{\boldsymbol{\theta}})=\frac{1}{\sqrt{T}}\hat{\mathbf{R}}'\hat{\boldsymbol{\vartheta}}\xrightarrow{d}N(\mathbf{0},\Sigma^{**})$$

where

$$\begin{aligned}\Sigma^{**} &= \mathbf{A}^{**}\Omega^{**}\mathbf{A}^{**\prime} \qquad (20)\\ \Omega^{**} &= \begin{bmatrix}\Omega_{11}^{**} & \mathbf{0} & \Omega_{31}^{**\prime}\\ \mathbf{0} & \Omega_{22} & \Omega_{32}'\\ \Omega_{31}^{**} & \Omega_{32} & \Omega_{33}\end{bmatrix}\\ \mathbf{A}^{**} &= \begin{bmatrix}\mathbf{D}_{\varphi}\mathbf{J}_{\varphi}^{**-1}, & \mathbf{D}_{\eta}\mathbf{J}_{\eta}^{-1} & \mathbf{I}_m\end{bmatrix}.\end{aligned}$$

The test statistic and limit null distribution follows in an analogous manner to the previous two cases, with Ω_{31}^{*} being consistently estimated by $\left(\frac{\hat{\vartheta}'\hat{\vartheta}}{T}\right)\frac{1}{T}\hat{\mathbf{R}}'\hat{\mathbf{C}}$.

The next section of the paper describes how previously proposed misspecification test statistics relate to the above unifying procedure and also offers some analysis on both their asymptotic validity and local optimality. In particular, notice that if $\mathbf{D}_{\varphi}$ can not be guaranteed to be zero, then there are potential estimation effects from the conditional mean that have to be taken into account when constructing asymptotically valid tests statistics, even though $\hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{\eta}}$ are (asymptotically) orthogonal. If so, the limit distribution of the test indicator will be different according to the method of estimation employed, through the introduction of either $\mathbf{J}_{\varphi}$, $\mathbf{J}_{\varphi}^{*}$ or $\mathbf{J}_{\varphi}^{**}$.

4 Testing for Asymmetry and Non-linearity

4.1 Engle and Ng Test

The most popular asymmetry tests are those proposed by Engle and Ng (1993). In order to confirm the asymmetric behaviour of financial series, they construct score type tests using the indicator function

$$I_{t-1} = \begin{cases} 1 \text{ if } \varepsilon_{t-1} \leq 0 \\ 0 \text{ otherwise.} \end{cases}$$

For purposes of exposition, consider the negative size bias test which examines the significance of the test variable $I_{t-1}\varepsilon_{t-1}$ in order to assess if important negative shocks have more impact on volatility than important positive shocks. The analysis of other tests proposed by Engle and Ng (1993) follows that of the symmetry test described below, except that the test variable will be defined differently. Specifically, Engle and Ng (1993) propose that under the alternative:

$$\log\left(h_t^a\right) = \log\left(h_t\left(\mathbf{s}_t; \eta\right)\right) + g\left(\mathbf{v}_t; \boldsymbol{\pi}\right) \tag{21}$$

where $h_t\left(\mathbf{s}_t; \boldsymbol{\eta}\right)$ is the model under the null (i.e. the GARCH model) and

$$g\left(\mathbf{v}_t; \boldsymbol{\pi}\right) = \pi I_{t-1}\varepsilon_{t-1}. \tag{22}$$

The test statistic proposed by Engle and Ng (1993) has the following form

$$T_{EN} = T \times \frac{\hat{\boldsymbol{\vartheta}}'\hat{\mathbf{Z}}\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\hat{\boldsymbol{\vartheta}}}{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}, \tag{23}$$

where $\hat{\mathbf{Z}}$ has rows $\hat{\mathbf{z}}_t' = \left(\hat{I}_{t-1}\hat{\varepsilon}_{t-1}, \hat{\mathbf{x}}_t'\right)$, and is assumed to be asymptotically χ_1^2 under the null. 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 (and the existence of appropriate higher order moments).

However, notice that the alternative model in (21) does not appear to represent an optimal form for the asymmetric and/or non-linear GARCH models proposed in the literature, since the conditional variance under the alternative h_t^a does not contain lagged h_t^a on the right hand side of (21). Rather, the alternative model of Engle and Ng (1993) represents a hybrid model, entailing, first, the GARCH model under the null and then the addition of test variables that allows for the asymmetric behaviour to h_t. This suggests that the approach of Engle and Ng (1993) will not yield locally optimal tests.

Moreover, the Engle and Ng (1993) tests are also *potentially* asymptotically invalid under the null hypothesis, since the construction of the test statistic does not take account of the possible estimation effects from the conditional mean equation. This can be seen below where $\mathbf{D}_{\varphi}$ is constructed for this particular case, from the discussion in Section 3, assuming a linear conditional mean function. For the test variable of Engle and Ng (1993), which is $\hat{\mathbf{r}}_t = \hat{I}_{t-1}\hat{\varepsilon}_{t-1}$,

$$\begin{aligned}
\mathbf{D}_{\varphi} &= 2\alpha_1 \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\beta_1^{(i-1)} E\left[\frac{1}{h_t} I_{t-1}\varepsilon_{t-1}\varepsilon_{t-i}\mathbf{w}'_{t-i}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= 2\alpha_1 \frac{1}{T}\sum_{t=1}^{T} E\left\{E\left[\frac{1}{h_t} I_{t-1}\varepsilon_{t-1}^2 \,|\mathcal{F}_{t-2}\right]\mathbf{w}'_{t-1}\right. \\
&\qquad \left. + \sum_{i=2}^{t}\beta_1^{(i-1)} E\left[\frac{1}{h_t} I_{t-1}\varepsilon_{t-1}\varepsilon_{t-i}\,|\mathcal{F}_{t-i-1}\right]\mathbf{w}'_{t-i}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
\end{aligned}$$

which is not necessarily non-zero (certainly, $E\left[\frac{1}{h_t} I_{t-1}\varepsilon_{t-1}^2|\mathcal{F}_{t-2}\right]$ is non-negative) and implies that there are, potentially, asymptotically non-negligible estimation effects from the conditional mean, thus rendering the Engle and Ng (1993) asymmetry tests asymptotically invalid.

4.2 Lundbergh and Teräsvirta Test

Lundbergh and Teräsvirta (2002) entertained an alternative in which ε_t follows a smooth transition GARCH (1,1) model,

$$h_t^a = h_t + g(\mathbf{v}_t; \boldsymbol{\pi})$$

where

$$g(\mathbf{v}_t; \boldsymbol{\pi}) = \left(\alpha_{01} + \alpha_{11}\varepsilon_{t-1}^2\right) F_n(\varepsilon_{t-1}; \gamma; \mathbf{c}).$$

Here, non-linearity is introduced in the intercept and the term containing the squared past errors via a smooth transition function $F_n(\varepsilon_{t-1}; \gamma; \mathbf{c})$, where

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

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. The smooth transition GARCH model is proposed by Hagerud (1997) and Gonzalez-Rivera (1998). The alternative model includes as a special case the GJR model of Glosten *et al.* (1993) when $\gamma \to \infty$.

Under the null of $\gamma = 0$, it follows that $F_n = 0$, and taking a first-order Taylor expansion of F_n around $\gamma = 0$, for $n = 1$ in (24) yields

$$g(\mathbf{v}_t; \boldsymbol{\pi}) = \boldsymbol{\pi}'\mathbf{v}_t \tag{25}$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2)'$, $\mathbf{v}_t = (\varepsilon_{t-1}, \varepsilon_{t-1}^3)'$. The null hypothesis now becomes $H_0 : \boldsymbol{\pi} = \mathbf{0}$, implying linearity, so that $g(\mathbf{v}_t; \boldsymbol{\pi}) = 0$. Lundbergh and Teräsvirta (2002) construct the following non-linearity test statistic

$$T_{LT} = T\frac{\hat{\boldsymbol{\vartheta}}'\hat{\mathbf{Z}}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\boldsymbol{\vartheta}}}{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}} \tag{26}$$

where $\hat{\mathbf{Z}}$ is a matrix with rows $\hat{\mathbf{z}}_t' = (\hat{\mathbf{x}}_t', \hat{\mathbf{v}}_t')$ and $\hat{\mathbf{v}}_t = \left(\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-1}^3\right)'$. The test is assumed to be asymptotically distributed as a χ_2^2 random variable under the null. Lundbergh and Teräsvirta (2002) also define an alternative regression based procedure, following Wooldridge (1991), which they suggest 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. However, heterokurticity is ruled out by Assumption 3.

As with the Engle and Ng (1993) approach, Lundbergh and Teräsvirta (2002) ignore the fact that the model under the alternative h_t^a should have lagged h_{t-i}^a on the right hand side implying that their non-linearity test is not locally optimal. Furthermore, it is also asymptotically invalid, since it ignores non-negligible estimation effects. Consider $\mathbf{D}_\varphi$, defined at (14), but employing the test variables considered by Lundbergh and Teräsvirta (2002), i.e. $\hat{\mathbf{r}}_t = \left(\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-1}^3\right)'$. Assuming conditional symmetry and, for simplicity, a linear functional form for the conditional mean,

$$\begin{aligned}
\mathbf{D}_\varphi &= 2\alpha_1 \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{1}{h_t}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-1}^3\end{pmatrix}\sum_{i=1}^{t}\beta_1^{(i-1)}\varepsilon_{t-i}\mathbf{w}_{t-i}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= 2\alpha_1 \frac{1}{T}\sum_{t=1}^{T} E\left\{E\left[\frac{1}{h_t}\begin{pmatrix}\varepsilon_{t-1}^2\\ \varepsilon_{t-1}^4\end{pmatrix}\middle|\mathcal{F}_{t-2}\right]\mathbf{w}_{t-1}'\right. \\
&\quad \left. + \frac{1}{T}\sum_{i=2}^{t}\beta_1^{i-1}E\left[\frac{1}{h_t}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-1}^3\end{pmatrix}\varepsilon_{t-i}\middle|\mathcal{F}_{t-i-1}\right]\mathbf{w}_{t-i}'\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= 2\alpha_1 \frac{1}{T}\sum_{t=1}^{T} E\left\{E\left[\frac{1}{h_t}\begin{pmatrix}\varepsilon_{t-1}^2\\ \varepsilon_{t-1}^4\end{pmatrix}\middle|\mathcal{F}_{t-2}\right]\mathbf{w}_{t-1}'\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
\end{aligned}$$

which again is non-zero, in general. The the second term (in the second line) is zero because, for $i \geq 2$,

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

and

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

where $E\left[h_t^{-1}\varepsilon_{t-1}^m|\mathcal{F}_{t-1}\right] = g(\varepsilon_{t-1})$, $m = 1, 3$, which is anti-symmetric in ε_{t-1}, so that $E\left[g(\varepsilon_{t-1})|\mathcal{F}_{t-2}\right] = 0$ since the conditional density of ε_{t-1} given $\mathcal{F}_{t-2}$ is symmetric.

Additionally, Lundbergh and Teräsvirta (2002) employ an inefficient asymptotic variance. Specifically, for the test variables considered by Lundbergh and Teräsvirta (2002),

$$\begin{aligned} \mathbf{D}_{\boldsymbol{\eta}} &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\beta_1^{i-1}E\left[\frac{1}{h_t}\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-1}^3 \end{pmatrix}\left(1, \varepsilon_{t-i}^2, h_{t-i}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= -\frac{1}{T}\sum_{t=1}^{T}E\left\{E\left[\frac{1}{h_t}\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-1}^3 \end{pmatrix}\left(1, \varepsilon_{t-1}^2, h_{t-1}\right)\middle|\mathcal{F}_{t-2}\right]\right. \\ &\quad \left.+\frac{1}{T}\sum_{i=2}^{t}\beta_1^{i-1}\left(1, \varepsilon_{t-i}^2, h_{t-i}\right)E\left[\frac{1}{h_t}\begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-1}^3 \end{pmatrix}\middle|\mathcal{F}_{t-2}\right]\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \end{aligned}$$

Arguments similar to those employed above, and in Section 2.1.4, imply that $\mathbf{D}_{\boldsymbol{\eta}}$ is the null vector. This is contrary to Lundbergh and Teräsvirta (2002) who, by accounting for the estimation effects of $\hat{\boldsymbol{\eta}}$, employ an inefficient variance in constructing the test statistic.

It can be shown that the test for remaining ARCH effects and the parameter constancy test proposed by Lundbergh and Teräsvirta (2002) are asymptotically valid as the estimation effects from the conditional mean are asymptotically negligible in these tests. However, the parameter constancy test of Lundbergh and Teräsvirta (2002) is not the locally most powerful test, since the alternative model defined by them does not contain h_{t-1}^a on the right hand of the conditional variance equation, under the alternative.

4.3 Testing for Asymmetry

Exploiting the general methodology of Section 3, new asymmetry tests that are guaranteed asymptotically valid and locally optimal are now provided. With asymmetry characterised by (10) with $g(\mathbf{v}_t; \boldsymbol{\pi}) = \pi I_{t-1}\varepsilon_{t-1}$, the generic misspecification test indicator (11), has test variable (12)

$$\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=1}^{t}\hat{\beta}_1^{i-1}\hat{I}_{t-i}\hat{\varepsilon}_{t-i} \tag{27}$$

which is evaluated at the null parameter estimator $\hat{\boldsymbol{\theta}}$. Observe how this differs from the Engle and Ng test variable of $\hat{I}_{t-1}\hat{\varepsilon}_{t-1}$, although arguments similar to those of Section 4.1 show that, here also, the corresponding expressions for $\mathbf{D}_{\boldsymbol{\varphi}}$, (14), and $\mathbf{D}_{\boldsymbol{\eta}}$, (15) can not be guaranteed to be zero, as follows.

Assuming conditional symmetry and, for simplicity of derivations, a linear conditional mean,

$$\begin{aligned}
\mathbf{D}_{\boldsymbol{\varphi}} &= 2\alpha_1\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\beta_1^{2(i-1)}E\left[\frac{1}{h_t^2}I_{t-i}\varepsilon_{t-i}^2\mathbf{w}_{t-i}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\
&\quad +4\alpha_1\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\sum_{j<i}\beta_1^{i+j-2}E\left[\frac{1}{h_t^2}I_{t-i}\varepsilon_{t-i}\varepsilon_{t-j}\mathbf{w}_{t-j}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
\end{aligned}$$

which is, in general, non-zero, as is

$$\begin{aligned}
\mathbf{D}_{\boldsymbol{\eta}} &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\beta_1^{2(i-1)}E\left[\frac{1}{h_t^2}I_{t-i}\varepsilon_{t-i}\left(1,\varepsilon_{t-i}^2,h_{t-i}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\
&\quad -\frac{2}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\sum_{j<i}\beta_1^{i+j-2}E\left[\frac{1}{h_t^2}I_{t-i}\varepsilon_{t-i}\left(1,\varepsilon_{t-j}^2,h_{t-j}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.
\end{aligned}$$

The discussion in Section 2.1 provides the following test statistic, following OLS estimation of the conditional mean

$$T_A = Td_T(\hat{\boldsymbol{\theta}})'\hat{\Sigma}^{-1}d_T(\hat{\boldsymbol{\theta}}) \tag{28}$$

where $\hat{\Sigma}$ is any consistent estimator of Σ, i.e. $\hat{\Sigma} = \Sigma + o_p(1)$, and

$$\begin{aligned}
\Sigma &= p\lim\frac{1}{T}\left[(k_c-1)\mathbf{R}_0'\mathbf{M}_{\mathbf{X}}\mathbf{R}_0 + \mathbf{R}_0'\mathbf{C}_0(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{H}_0\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{C}_0'\mathbf{R}_0\right]\\
\mathbf{M}_{\mathbf{X}} &= \mathbf{I}_{\mathbf{X}} - \mathbf{X}_0\left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'
\end{aligned}$$

and where $\mathbf{I}_{\mathbf{X}}$ an identity matrix of order $\dim(\eta)$. A consistent estimator of the asymptotic variance Σ is given by

$$\hat{\Sigma}^* = \frac{1}{T}\left[\frac{\hat{\boldsymbol{\vartheta}}'\hat{\boldsymbol{\vartheta}}}{T}\hat{\mathbf{R}}'\hat{\mathbf{M}}_{\mathbf{X}}\hat{\mathbf{R}} + \hat{\mathbf{R}}'\hat{\mathbf{C}}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\hat{\mathbf{H}}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\hat{\mathbf{C}}'\hat{\mathbf{R}}\right]$$

where $\mathbf{M}_{\mathbf{X}} = \mathbf{I}_{\mathbf{X}} - \hat{\mathbf{X}}\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'$. Under the null, $T_A \xrightarrow{d} \chi_1^2$.

The asymmetry test statistic proposed in (28) remains asymptotically valid under conditional non-normality of ξ_t, provided ξ_t satisfies the assumptions of Section 2. Following Engle and Ng (1993), we can also test asymmetry for more extreme values of ε_{t-1}. The asymptotic distribution of the test in this case, is the same as the previous one, except that the test indicator is $\hat{\mathbf{r}}_t = \hat{h}_t^{-1} \sum_{i=1}^{t} \hat{\beta}_1^{i-1} \hat{I}_{t-i} \hat{\varepsilon}_{t-i}^2$.

When employing different consistent estimators for the conditional mean parameter vector $\boldsymbol{\varphi}$, the asymptotic distribution of the test statistic will change accordingly, as noted in Section 3. If the conditional mean is nonlinear, for example a STAR or SETAR model, and estimated by NLLS, then, given the conditional symmetry assumption, the test statistic becomes

$$T_A^* = T d_T(\hat{\boldsymbol{\theta}})'(\hat{\Sigma}^*)^{-1} d_T(\hat{\boldsymbol{\theta}})$$

where $\hat{\Sigma}^* = \Sigma^* + o_p(1)$, where

$$\Sigma^* = p\lim \frac{1}{T} \left[(k_c - 1) \mathbf{R}_0' \mathbf{M}_\mathbf{X} \mathbf{R}_0 + \mathbf{R}_0' \mathbf{C}_0 (\mathbf{F}_0' \mathbf{F}_0)^{-1} \mathbf{F}_0' \mathbf{H}_0 \mathbf{F}_0 (\mathbf{F}_0' \mathbf{F}_0)^{-1} \mathbf{C}_0' \mathbf{R}_0 \right]$$

which can be consistently estimated by

$$\hat{\Sigma}^* = \frac{1}{T} \left[\frac{\hat{\boldsymbol{\vartheta}}' \hat{\boldsymbol{\vartheta}}}{T} \hat{\mathbf{R}}' \hat{\mathbf{M}}_\mathbf{X} \hat{\mathbf{R}} + \hat{\mathbf{R}}' \hat{\mathbf{C}} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \hat{\mathbf{H}} \mathbf{F} (\mathbf{F}'\mathbf{F})^{-1} \hat{\mathbf{C}}' \hat{\mathbf{R}} \right].$$

Now, if the conditional mean is estimated by QMLE, then

$$T_A^{**} = T d_T(\hat{\boldsymbol{\theta}})'(\hat{\Sigma}^{**})^{-1} d_T(\hat{\boldsymbol{\theta}})$$

where $\hat{\Sigma}^{**} = \Sigma^{**} + o_p(1)$ with

$$\begin{aligned} \Sigma^{**} = \; & p\lim \frac{1}{T} \Bigg[(k_c - 1) \mathbf{R}_0' \mathbf{M}_\mathbf{X} \mathbf{R}_0 - (k_c - 1) \mathbf{R}_0' \mathbf{C}_0 \left(\mathbf{W}' \mathbf{H}_0^{-1} \mathbf{W} + \frac{1}{2} \mathbf{C}_0' \mathbf{C}_0 \right)^{-1} \mathbf{C}_0' \mathbf{R}_0 \\ & + \mathbf{R}_0' \mathbf{C}_0 \left(\mathbf{W}' \mathbf{H}_0^{-1} \mathbf{W} + \frac{1}{2} \mathbf{C}_0' \mathbf{C}_0 \right)^{-1} \left(\mathbf{W}' \mathbf{H}_0^{-1} \mathbf{W} + \frac{(k_c - 1)}{4} \mathbf{C}_0' \mathbf{C}_0 \right) \\ & \left(\mathbf{W}' \mathbf{H}_0^{-1} \mathbf{W} + \frac{1}{2} \mathbf{C}_0' \mathbf{C}_0 \right)^{-1} \mathbf{C}_0' \mathbf{R}_0 \Bigg] \end{aligned}$$

which can be consistently estimated by

$$\begin{aligned} \hat{\Sigma}^{**} = \; & p\lim \frac{1}{T} \Bigg[\frac{\hat{\boldsymbol{\vartheta}}' \hat{\boldsymbol{\vartheta}}}{T} \hat{\mathbf{R}}' \hat{\mathbf{M}}_\mathbf{X} \hat{\mathbf{R}} - \frac{\hat{\boldsymbol{\vartheta}}' \hat{\boldsymbol{\vartheta}}}{T} \hat{\mathbf{R}}' \hat{\mathbf{C}} \left(\mathbf{W}' \hat{\mathbf{H}}^{-1} \mathbf{W} + \frac{1}{2} \hat{\mathbf{C}}' \hat{\mathbf{C}} \right)^{-1} \hat{\mathbf{C}}' \hat{\mathbf{R}} \\ & + \hat{\mathbf{R}}' \hat{\mathbf{C}} \left(\mathbf{W}' \hat{\mathbf{H}}^{-1} \mathbf{W} + \frac{1}{2} \hat{\mathbf{C}}' \hat{\mathbf{C}} \right)^{-1} \left(\mathbf{W}' \hat{\mathbf{H}}^{-1} \mathbf{W} + \frac{\hat{\boldsymbol{\vartheta}}' \hat{\boldsymbol{\vartheta}}}{4T} \hat{\mathbf{C}}' \hat{\mathbf{C}} \right) \\ & \left(\mathbf{W}' \hat{\mathbf{H}}^{-1} \mathbf{W} + \frac{1}{2} \hat{\mathbf{C}}' \hat{\mathbf{C}} \right)^{-1} \hat{\mathbf{C}}' \hat{\mathbf{R}} \Bigg]. \end{aligned}$$

4.4 Testing for non-linearity

Following the generic test procedure introduced in Section 3, the function of omitted variables is derived from a first-order Taylor expansion of (24). It transpires that the score indicator for testing for non-linearity is $d_T(\hat{\boldsymbol{\theta}})$ in (11) where the test variables take the form

$$\mathbf{r}_t = \frac{1}{h_t}\sum_{i=1}^{t}\beta_1^{i-1}\begin{pmatrix}\varepsilon_{t-i}\\ \varepsilon_{t-i}^3\end{pmatrix}. \tag{29}$$

The test variables of Lundbergh and Teräsvirta (2002) are simply $\mathbf{v}_t = \left(\varepsilon_{t-1}, \varepsilon_{t-1}^3\right)'$, which omits a multiplicative h_t^{-1} term and as well as ignoring the recursive behaviour of the conditional variance. Similar arguments to those of Section 4.2 imply that $\mathbf{D}_{\boldsymbol{\varphi}}$ can not be guaranteed to be zero, whilst analysis similar to that of Section 2.1.4 reveals that $\mathbf{D}_{\boldsymbol{\eta}} = \mathbf{0}$, which implies that the influence of $\hat{\boldsymbol{\eta}}$ is asymptotically negligible and need not be considered in constructing the asymptotic variance of the test. To see this, observe that (again, assuming that the conditional mean is linear)

$$\begin{aligned}
\mathbf{D}_{\boldsymbol{\varphi}} &= 2\alpha_1\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\beta_1^{2(i-1)}E\left[\frac{1}{h_t^2}\begin{pmatrix}\varepsilon_{t-i}\\ \varepsilon_{t-i}^3\end{pmatrix}\varepsilon_{t-i}\mathbf{w}_{t-i}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\
&\quad +4\alpha_1\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\sum_{j<i}\beta_1^{i+j-2}E\left[\frac{1}{h_t^2}\begin{pmatrix}\varepsilon_{t-i}\\ \varepsilon_{t-i}^3\end{pmatrix}\varepsilon_{t-j}\mathbf{w}_{t-j}'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\
&= 2\alpha_1\frac{1}{T}\sum_{t=1}^{T}E\left\{\sum_{i=1}^{t}\beta_1^{2(i-1)}E\left[\frac{1}{h_t^2}\begin{pmatrix}\varepsilon_{t-i}^2\\ \varepsilon_{t-i}^4\end{pmatrix}\middle|\mathcal{F}_{t-i-1}\right]\mathbf{w}_{t-i}'\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},
\end{aligned}$$

which is non-zero, in general, although observe the expectations of the cross-product terms will be zero following the line of argument presented in Section 2.1.4.

On the other hand,

$$\begin{aligned}
\mathbf{D}_{\boldsymbol{\eta}} &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\beta_1^{2(i-1)}E\left[\frac{1}{h_t^2}\begin{pmatrix}\varepsilon_{t-i}\\ \varepsilon_{t-i}^3\end{pmatrix}\left(1, \varepsilon_{t-i}^2, h_{t-i}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\
&\quad -\frac{2}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\sum_{j<i}\beta_1^{i+j-2}E\left[\frac{1}{h_t^2}\begin{pmatrix}\varepsilon_{t-i}\\ \varepsilon_{t-i}^3\end{pmatrix}\left(1, \varepsilon_{t-j}^2, h_{t-j}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}
\end{aligned}$$

and arguments similar to those used in Section 2.1.4 show that $\mathbf{D}_{\boldsymbol{\eta}} = \mathbf{0}$.

Similar analysis also shows that $\Omega_{32} = \mathbf{0}$, in this case. Therefore, the forms of the asymptotic variance of our non-linearity test will differ from those of our asymmetry test proposed in Section 4.3 by not including terms to account for the estimation effects from the conditional variance under the null.

Thus, the non-linearity test statistic is

$$T_N = T d_T(\hat{\boldsymbol{\theta}})' \hat{\Sigma}^{-1} d_T(\hat{\boldsymbol{\theta}}) \tag{30}$$

which is asymptotically distributed χ^2_m under the null, where $m = \dim(\mathbf{v}_t)$, $\hat{\Sigma} = \Sigma + o_p(1)$ and

$$\Sigma = p\lim \frac{1}{T}\left[(k_c - 1)\,\mathbf{R}_0'\mathbf{R}_0 + \mathbf{R}_0'\mathbf{C}_0(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{H}_0\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{C}_0'\mathbf{R}_0\right]. \tag{31}$$

As with the asymmetry test, the non-linearity test proposed here is asymptotically valid under non-normality provided ξ_t satisfies the assumptions of Section 2. Furthermore, the asymptotic variance matrix in the limit distribution will change according to the choice of $\hat{\boldsymbol{\varphi}}$. Following NLLS estimation, the asymptotic variance becomes:

$$\Sigma^* = p\lim \frac{1}{T}\left[(k_c - 1)\,\mathbf{R}_0'\mathbf{R}_0 + \mathbf{R}_0'\mathbf{C}_0(\mathbf{F}_0'\mathbf{F}_0)^{-1}\mathbf{F}_0'\mathbf{H}_0\mathbf{F}_0(\mathbf{F}_0'\mathbf{F}_0)^{-1}\mathbf{C}_0'\mathbf{R}_0\right],$$

whilst if $\hat{\boldsymbol{\varphi}}$ is estimated by QML,

$$\begin{aligned}\Sigma^{**} \;=\; & p\lim \frac{1}{T}\Bigg[(k_c - 1)\,\mathbf{R}_0'\mathbf{R}_0 - (k_c - 1)\,\mathbf{R}_0'\mathbf{C}_0\left(\mathbf{W}'\mathbf{H}_0^{-1}\mathbf{W} + \frac{1}{2}\mathbf{C}_0'\mathbf{C}_0\right)^{-1}\mathbf{C}_0'\mathbf{R}_0 \\ & + \mathbf{R}_0'\mathbf{C}_0\left(\mathbf{W}'\mathbf{H}_0^{-1}\mathbf{W} + \frac{1}{2}\mathbf{C}_0'\mathbf{C}_0\right)^{-1}\left(\mathbf{W}'\mathbf{H}_0^{-1}\mathbf{W} + \frac{(k_c - 1)}{4}\mathbf{C}_0'\mathbf{C}_0\right) \\ & \left(\mathbf{W}'\mathbf{H}_0^{-1}\mathbf{W} + \frac{1}{2}\mathbf{C}_0'\mathbf{C}_0\right)^{-1}\mathbf{C}_0'\mathbf{R}_0\Bigg].\end{aligned}$$

As before, the various asymptotic variances can be consistently estimated using $\hat{\Sigma}$, $\hat{\Sigma}^*$ or $\hat{\Sigma}^{**}$ with the appropriate redefinition of $\hat{\mathbf{R}}$.

5 Sensitivity Analysis

The tests proposed above employ an implicit alternative, since they are expected to be locally most powerful against the alternatives for which they are designed. In this section, asymptotic local analysis is employed to investigate whether the asymmetry and non-linearity tests proposed have power against alternatives for which they are not designed, i.e. unconsidered alternatives. Intuitively, since the test indicators are asymptotically sensitive to the conditional mean estimation, one might expect the tests to be sensitive to misspecification of the mean.

Suppose that the true conditional mean is represented by

$$E\left[y_t|\mathcal{F}_{t-1}\right] = m\left(\mathbf{w}_t\right) = \mathbf{w}_t'\boldsymbol{\varphi} + l(\mathbf{w}_t)'\frac{\boldsymbol{\delta}}{\sqrt{T}}$$

where $l\left(\mathbf{w}_t\right)$ is a bounded function, which can be non-linear, $\boldsymbol{\delta}'\boldsymbol{\delta} < \infty$, and, for simplicity, the null conditional mean is assumed linear with $\boldsymbol{\varphi}$ being estimated by OLS (although, qualitatively, the results below will remain valid under NLLS or QML estimation). Then, the true process is

$$y_t = m\left(\mathbf{w}_t\right) + \varepsilon_t.$$

In order to analyse the asymptotic behaviour of the non-linearity tests under local misspecification of the conditional mean, the approach of Godfrey and Orme (1996) is adopted. Define $\hat{\boldsymbol{\theta}}' = \left(\hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\eta}}', \mathbf{0}'\right)$, under the null, with $\left(\hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\eta}}'\right)$ as previously and $\boldsymbol{\theta}_0' = \left(\boldsymbol{\varphi}_0', \boldsymbol{\eta}_0', \boldsymbol{\delta}_T'\right)$ as the true parameter vector, where $\boldsymbol{\delta}_T = T^{-1/2}\boldsymbol{\delta}'$. Proceeding generally, a mean value expansion of $\sqrt{T}d_T(\hat{\boldsymbol{\theta}})$ defined in (11) about $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ yields

$$\begin{aligned}\sqrt{T}d_T(\hat{\boldsymbol{\theta}}) \quad &= \quad \sqrt{T}d_T(\boldsymbol{\theta}_0) + \frac{\partial d_T\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\eta}'}\sqrt{T}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) + \frac{\partial d_T\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\varphi}'}\sqrt{T}\left(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\right) \\ &\quad -\frac{\partial d_T\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\delta}_T'}\boldsymbol{\delta}\end{aligned}$$

where $\bar{\boldsymbol{\theta}}$ denotes the usual mean value between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, so that $\bar{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$. As previously, assume Uniform Laws of Large Numbers apply, but now under local alternatives (see, for example, Newey and McFadden (1994)). The analysis of Godfrey and Orme (1996) then yields

$$\begin{aligned}\sqrt{T}d_T(\hat{\boldsymbol{\theta}}) \quad = \quad & \frac{1}{\sqrt{T}}\mathbf{R}_0'\boldsymbol{\vartheta}_0 \\ & +\mathbf{D}_{\varphi}\mathbf{J}_{\varphi}{}^{-1}\frac{1}{\sqrt{T}}\mathbf{W}'\boldsymbol{\varepsilon}_0 \\ & +\mathbf{D}_{\boldsymbol{\eta}}\mathbf{J}_{\boldsymbol{\eta}}{}^{-1}\frac{1}{\sqrt{T}}\mathbf{X}_0'\boldsymbol{\vartheta}_0 \\ & -\mathbf{D}_{\boldsymbol{\delta}}\boldsymbol{\delta} + o_p(1).\end{aligned}$$

where $\dfrac{\partial d_T\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\delta}_T'} - \mathbf{D}_{\boldsymbol{\delta}} \xrightarrow{p} \mathbf{0}$..

Under the true process, and using similar arguments as before,

$$\begin{aligned}
\mathbf{D}_{\boldsymbol{\delta}} &= \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{2\varepsilon_t}{h_t}\mathbf{r}_t\frac{\partial\varepsilon_t}{\partial\boldsymbol{\delta}_T'} + \left(\frac{\varepsilon_t^2}{h_t}-1\right)\frac{\partial\mathbf{r}_t}{\partial\boldsymbol{\delta}_T'} - \frac{\varepsilon_t^2}{h_t^2}\mathbf{r}_t\frac{\partial h_t}{\partial\boldsymbol{\delta}_T'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left[-\frac{\varepsilon_t^2}{h_t^2}\mathbf{r}_t\frac{\partial h_t}{\partial\boldsymbol{\delta}_T'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left[-\frac{1}{h_t}\mathbf{r}_t\frac{\partial h_t}{\partial\boldsymbol{\delta}_T'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= -p\lim\frac{1}{T}\mathbf{R}_0'\mathbf{G}_0
\end{aligned}$$

where $\mathbf{G}$ has rows

$$\begin{aligned}
\mathbf{g}_t' &= \frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\delta}_T'} \\
&= 2\alpha_1\frac{1}{h_t}\sum_{i=1}^{t}\beta_1^{i-1}\varepsilon_{t-i}l\left(\mathbf{w}_{t-i}\right)'.
\end{aligned}$$

Now, following the arguments used in Section 2.1.4, it can be shown that for both the asymmetry test T_A proposed in Section 4.3, with $\mathbf{r}_t = h_t^{-1}\sum_{i=1}^{t}\beta_1^{i-1}I_{t-i}\varepsilon_{t-i}$, and for the non-linearity test proposed in Section 4.4, with test variable $\mathbf{r}_t = \dfrac{1}{h_t}\sum_{i=1}^{t}\beta_1^{i-1}\begin{pmatrix}\varepsilon_{t-i}\\ \varepsilon_{t-i}^3\end{pmatrix}$, $\mathbf{D}_{\boldsymbol{\delta}}$ is, in general, non-zero.

This implies that the asymmetry and non-linearity tests are asymptotically sensitive to local misspecification of the conditional mean. The results of Godfrey and Orme (1996) show that under local misspecification of the conditional mean, the test is asymptotically distributed as non-central χ^2 with the noncentrality parameter given by

$$\lambda = \boldsymbol{\delta}'\mathbf{D}_{\boldsymbol{\delta}}'\Sigma^{-1}\mathbf{D}_{\boldsymbol{\delta}}\boldsymbol{\delta}.$$

Thus, for example, neglected non-linearities in the conditional mean may lead to misleading conclusions about non-linearities in the conditional variance.

Modelling non-linearities in the conditional mean with GARCH errors has recently become of interest. Li and Li (1996) modelled non-linearities in both the conditional mean and variance through a SETAR type conditional mean and a Threshold ARCH (TARCH) for the conditional variance. The STAR-STGARCH model of Lundbergh and Teräsvirta (1999) models non-linearities in both the conditional mean and variance parameters by allowing the switch between regimes to be smooth. In the light of the preceding analysis, misspecification tests of the conditional mean should be conducted before assessing the ARCH/GARCH specification of the errors. Indeed, misspecification tests of the conditional mean (i.e., tests of $E\left[\varepsilon_t|\mathbf{w}_t\right]=0$) can be made asymptotically robust to misspecification of the conditional variance. Therefore, the strategy in misspecification testing of GARCH models is first to estimate the conditional mean where the errors follow a GARCH process and test misspecification of the conditional mean by using, say, the a parametric test and the bootstrap method of Gonçalves and Killian (2004), or a consistent test in the spirit of Fan and Li (1999). Then, after estimating the GARCH process, misspecification testing of the conditional variance can be performed using the tests proposed in Section 4.3 and 4.4.

Chan and McAleer (2002) examine the effects on the MLE of misspecifying a STAR-GARCH model as an AR-GARCH model. They report a bias in the α_1 and β_1 estimates of the conditional variance in finite samples and argue that the magnitude of the finite samples bias depends on the functional form of the transition function. Here it has been verified that neglected non-linearities in the conditional mean may lead to misleading inferences about the conditional variance equation.

6 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. Furthermore, the sensitivity of the tests to local misspecification of the conditional mean is examined, in finite samples.

The Monte Carlo experiment for assessing the size properties of the tests is based on an AR(1)-GARCH(1,1) data generation process, namely

$$\begin{aligned} y_t &= \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t}\xi_t \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned}$$

where $\xi_t \sim N(0,1)$ or $\xi_t \sim t(10)$ (standardised Student t-distribution with 10 degrees of freedom).

We consider the following sets of parameter values for the conditional mean:

Model (1) : $y_t = \varepsilon_t$
Model (2) : $y_t = 1 + 0.1 y_{t-1} + \varepsilon_t$

The conditional variance equation follows Engle and Ng (1993), such that the unconditional variance of ε_t equals one without loss of generality, where

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): $h_t = 0.2 + 0.05\varepsilon_{t-1}^2 + 0.75h_{t-1}$

Combining the conditional mean and variance equations yields six models to generate. For this purpose, a series of 1000 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. This leaves an effective sample size of 800 observations. Each model is replicated and estimated 1000 times. Because of computational costs, we estimate the conditional mean parameters by OLS and then the conditional variance parameters by QML, as described in Section 2.1. The test statistics considered were T_A of (28) with $\hat{\mathbf{r}}_t = \frac{1}{\hat{h}_t}\sum_{i=1}^{t}\hat{\beta}_1^{i-1}\hat{I}_{t-i}\hat{\varepsilon}_{t-i}$; T_N of (30) with $\mathbf{r}_t = \frac{1}{h_t}\sum_{i=1}^{t}\beta_1^{i-1}\varepsilon_{t-i}^3$; the Engle and Ng statistic, T_{EN}, of (23); and,.the Lundbergh and Teräsvirta statistic, T_{LT}, of (26) with $v_t = \hat{\varepsilon}_{t-1}^3$.

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, i.e. no mean estimation for Model (1) and OLS estimation for Model (2). 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%. However, the empirical literature typical reports high persistence volatility models for financial time series. When there are estimation effects from the conditional mean generated as an AR process, T_{EN} tends to be slightly undersized for all volatility models relative to T_A, which is slightly oversized. Note that under high persistence and an AR conditional mean specification, T_{EN} is quite undersized which is consistent with the analysis of Section 4.1 which implies that the variance estimator employed can be expected to be biased downwards, in general.

Table 1. Empirical size

			$N(0,1)$				$t(10)$			
	φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
H	-	-	4.9	4.6	5.8	3.1	5.7	4.1	4.5	1.4
(0.09,0.9)	1	0.1	4.8	3.9	5.0	2.3	4.9	4.3	3.9	1.5
M	-	-	5.2	4.9	4.8	4.0	4.7	3.8	4.1	4.0
(0.05,0.9)	1	0.1	5.2	4.5	5.2	4.3	5.5	4.1	4.2	3.0
L	-	-	6.0	4.8	4.2	4.1	4.5	4.6	4.1	2.9
(0.05,0.75)	1	0.1	5.5	4.6	5.2	4.0	4.7	4.6	4.8	2.6

The empirical size of the non-linearity test, T_N, is close to the nominal size, whereas T_{LT} is undersized in all the experiments, especially for a high persistence volatility model and under $t(10)$ errors. When the conditional mean is generated as an AR process, the empirical size of T_N is close to the nominal size, except for a low persistence volatility model, whereas that of T_{LT} is lower than the nominal size of 5% for all volatility models examined. Again, by ignoring asymptotically non-negligible estimation effects, the T_{LT} statistic employs a variance estimator which is biased downwards, thus providing and under-sized test procedure.

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 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, which has similar parameter values as the GJR model except that 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 $\mathbf{v}_t = \varepsilon_{t-1}^3$ when the data is generated from the GJR and STGARCH models, but $\mathbf{v}_t = \varepsilon_{t-1}$ for the EGARCH and TGARCH models. (Again, for each non-linear model 1000 observations are drawn from the standard normal distribution, as well as from the standardised Student t-distribution with 10 degrees of freedom. The first 200 observations are disregarded in order to avoid initialization problems. Each design is carried out with 1000 replications. Again, for each model the conditional mean equation is designed as previously.)

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, when the asymmetry parameter is 0.212, and under normality, the simulated power for the non-linearity test T_N is 88.9%, whereas that of the test proposed by Lundbergh and Teräsvirta (2002), T_{LT}, is 7.7%, when there are no estimation effects from the conditional mean. This implies, that T_{LT} is relatively insensitive to this alternative model. Similar conclusions can be drawn when the coefficient that measures the asymmetry is reduced to 0.17.

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 and the model with larger asymmetry is examined the power of T_N is 96.4% whereas that of T_{LT} is 40.6%. Similarly, the asymmetry test T_A attains a simulated power of 95.6%, whereas the actual rejection frequency of T_{EN} is 67.2%. For the non-normal distribution, the differences are also significant.

Table 2. Empirical power

GJR (1,1) model

$$h_t = 0.005 + 0.136\varepsilon_{t-1}^2 + 0.212 I\left(\varepsilon_{t-1}\right)\varepsilon_{t-1}^2 + 0.7h_{t-1}$$

$$I\left(\varepsilon_{t-1}\right) = \begin{cases} 1 \text{ if } \varepsilon_{t-1}<0 \\ 0 \text{ otherwise} \end{cases}$$

		$N(0,1)$				$t(10)$			
φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
-	-	85.8	47.8	88.9	7.7	65.2	30.0	68.6	1.9
1	0.1	81.6	45.3	86.2	6.4	63.0	30.4	68.3	2.6

$$h_t = 0.005 + 0.136\varepsilon_{t-1}^2 + 0.17 I\left(\varepsilon_{t-1}\right)\varepsilon_{t-1}^2 + 0.7h_{t-1}$$

$$I\left(\varepsilon_{t-1}\right) = \begin{cases} 1 \text{ if } \varepsilon_{t-1}<0 \\ 0 \text{ otherwise} \end{cases}$$

		$N(0,1)$				$t(10)$			
φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
-	-	75.4	35.7	75.9	7.6	53.4	25.6	53.4	1.9
1	0.1	69.5	36.3	73.9	7.0	49.4	24.3	54.8	2.3

STGARCH (1,1) model

$$h_t = 0.005 + 0.136\varepsilon_{t-1}^2 - 0.212 F\left(\varepsilon_{t-1}\right)\varepsilon_{t-1}^2 + 0.7h_{t-1}$$

$$F\left(\varepsilon_{t-1}\right) = \frac{1}{1+\exp(-100\varepsilon_{t-1})} - \frac{1}{2}$$

		$N(0,1)$				$t(10)$			
φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
-	-	95.4	68.7	96.6	41.3	81.7	49.7	88.0	18.9
1	0.1	95.6	67.2	96.4	40.6	81.4	49.3	88.0	18.6

$$h_t = 0.005 + 0.136\varepsilon_{t-1}^2 - 0.17 F\left(\varepsilon_{t-1}\right)\varepsilon_{t-1}^2 + 0.7h_{t-1}$$

$$F\left(\varepsilon_{t-1}\right) = \frac{1}{1+\exp(-100\varepsilon_{t-1})} - \frac{1}{2}$$

		$N(0,1)$				$t(10)$			
φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
-	-	85.3	51.3	87.5	30.5	64.7	36.6	70.6	12.4
1	0.1	83.6	51.0	87.1	29.2	63.0	36.0	70.4	12.3

EGARCH (1,1) model

$$\log(h_t) = -0.23 + 0.9\log(h_{t-1}) + 0.25\left[\left|\xi_{t-1}\right| - 0.3\xi_{t-1}\right]^2$$

		$N(0,1)$				$t(10)$			
φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
-	-	84.7	29.9	75.2	34.3	68.9	20.8	57.8	25.3
1	0.1	82.2	22.0	73.2	34.1	66.6	12.1	54.6	21.6

TGARCH (1,1) model

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

		$N(0,1)$				$t(10)$			
φ_0	φ_1	T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
-	-	98.4	45.8	97.3	52.0	93.0	30.5	86.8	36.8
1	0.1	98.1	44.9	95.9	51.7	91.3	30.4	84.3	36.5

Notice that the asymmetry tests have higher power against the STGARCH (1,1) model compared with the corresponding GJR (1,1) model. 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 excellent power when compared with T_{EN} and T_{LT}. Moreover, the simulations reveal these tests can be interpreted 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.

Finally, Monte Carlo results are reported with illustrate the sensitivity of these tests to unconsidered misspecification in the conditional mean, as analysed in Section 5. The results are obtained by employing SETAR and STAR models as true data generation processes for the conditional mean, with the conditional variance structures being given, as before, by Models H, M and L. The precise specifications employed to generate the data are as follows:

SETAR1: $y_t = \begin{cases} 1 - 0.1y_{t-1} + \varepsilon_t & \text{if} \quad y_{t-1} < 1 \\ 0.5 + 0.7y_{t-1} + \varepsilon_t & \text{if} \quad y_{t-1} \geq 1 \end{cases}$

SETAR2: $y_t = \begin{cases} 0.5 - 0.7y_{t-1} + \varepsilon_t & \text{if} \quad y_{t-1} < 1 \\ 0.5 + 0.7y_{t-1} + \varepsilon_t & \text{if} \quad y_{t-1} \geq 1 \end{cases}$

LSTAR1: $y_t = 0.9 + 0.1y_{t-1} + (1.5 - 0.7y_{t-1})\, F_{t-1} + \varepsilon_t$ where

$$F_{t-1} = \{1 + \exp\left[-7\left(y_{t-1} - 0.05\right)\right]\}^{-1}$$

ESTAR1: $y_t = 0.9 + 0.1y_{t-1} + (1.5 - 0.7y_{t-1})\, F_{t-1} + \varepsilon_t$ where

$$F_{t-1} = 1 - \exp\left[-7(y_{t-1} - 0.05)^2\right]$$

In order to investigate the finite sample sensitivity of the test procedures to conditional mean misspecification, the assumed specification is given by an $AR(1)$ namely: $y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t$, whereas for the conditional variance specification the assumed and the true specifications are given by the GARCH (1,1) model. In these experiments, as before, for each model 1000 observations are drawn from the standard normal distribution, in which first 200 observations are discarded. Each design is carried out with 1000 replications.

Table 3 reports the rejection frequencies of the tests when the conditional mean is misspecified. The tests do indeed exhibit some "power" in this case, although T_A and T_N appear relatively more robust than T_{EN} and T_{LT}, respectively. Note, also, that T_N is actually quite robust to LSTAR1 and ESTAR1 misspecifications. Thus, neglected non-linearities in the conditional mean may lead to misleading conclusions about non-linearities in the conditional variance. Therefore, it is important to perform misspecification tests of the conditional mean which are robust to conditional heteroskedasticity.

Table 3. Rejection frequencies for misspecified conditional mean models

Model		$N(0,1)$				$t(10)$			
		T_A	T_{EN}	T_N	T_{LT}	T_A	T_{EN}	T_N	T_{LT}
SETAR1	**H** (0.09,0.9)	4.2	24.3	9.3	17.5	3.9	20.7	8.5	14.4
	M (0.05,0.9)	9.4	22.5	15.6	25.8	6.8	21.1	13.2	21.2
	L (0.05,0.75)	12.2	16.7	19.8	22.6	10.1	14.2	17.0	18.7
SETAR2	**H** (0.09,0.9)	71.8	99.1	79.0	91.0	66.7	98.3	76.8	86.4
	M (0.05,0.9)	92.2	99.3	97.6	99.3	81.3	98.0	93.5	96.9
	L (0.05, 0.75)	93.4	99.5	99.8	99.3	82.7	96.0	98.4	96.2
LSTAR1	**H** (0.09,0.9)	12.8	57.9	3.2	16.1	10.5	46.2	1.6	10.8
	M (0.05,0.9)	15.2	56.2	5.8	21.1	10.8	42.2	6.0	9.5
	L (0.05,0.75)	26.6	46.0	12.8	19.1	17.7	32.9	8.4	7.0
ESTAR1	**H** (0.09,0.9)	10.8	50.1	3.6	7.0	9.1	33.9	3.7	3.2
	M (0.05, 0.9)	13.3	49.5	5.3	11.6	9.6	33.9	3.2	4.8
	L (0.05,0.75)	26.6	46.6	7.5	11.3	16.3	29.5	4.8	4.8

7 Conclusion

This paper has provided some unifying theoretical results for misspecification testing in GARCH models, which have practical implications for empirical research. New asymptotically valid and locally optimal tests for asymmetry and non-linearity for GARCH models have been proposed. Moreover, it has been argued that the asymmetry and non-linearity tests proposed by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002) , respectively, are neither asymptotically valid (since they ignore asymptotically non-negligible estimation effects) nor locally optimal (since they ignore the recursive nature of the conditional variance structure).

In addition to linear mean specifications estimated by OLS, theoretical results are also presented for the asymptotic distribution of asymmetry and non-linearity tests when the conditional mean is estimated by NLLS and QML. These results, therefore, provide the framework for misspecification tests to be undertaken in a variety of circumstances. For example, misspecification tests for the STAR-GARCH model have not been considered in the literature to date.

Most importantly, the Monte Carlo results clearly suggest that the power of the new tests is excellent when compared with the previous tests proposed by Engle and Ng (1993) and Lundbergh and Teräsvirta (2002). Moreover, the tests are powerful against various non-linear models proposed in the literature, suggesting that they can be used as general misspecification tests in GARCH models against non-linearity and/or asymmetry. Finally, and contrary to previous understanding, the theoretical results here show that GARCH misspecification tests are asymptotically sensitive to unconsidered misspecification of the conditional mean, which suggests that robust tests of the conditional mean specification must be undertaken before tests of the GARCH specification are performed. Such tests of the conditional mean might employ the recently proposed bootstrap scheme by Gonçalves and Killian (2004) and this is left for future research.

References

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

[2] Bera, A.K., Higgins, M.L., 1993. ARCH models: properties, estimation and testing. Journal of Economic Surveys 7, 304-362.

[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] Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models. Handbook of Econometrics Vol. 4, Ch. 49, 2659-3038.

[7] 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.

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

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

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

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

[12] Fama, E.F., 1970. Efficient capital markets: a review of the theory and empirical work. Journal of Finance 25, 383-416.

[13] Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. Nonparametric Statistics 10, 245-271.

[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., 1988. Misspecification tests in econometrics: the Lagrange Multiplier principle and others procedures. Cambridge University Press.

[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] Gonçalves, S., Killian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics, 123, 89-120.

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

[20] Green, W.H., 2003. Econometric Analysis (5th Edition). Prentice-Hall.

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

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

[23] 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.

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

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

[26] Lundbergh, S., Teräsvirta, T., 2002. Evaluating GARCH models. Journal of Econometrics 110, 417-435.

[27] 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.

[28] Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 36, 394-419.

[29] Nelson, D.B., 1990. Stationary and persistence in the GARCH(1,1) model. Econometric Theory 6, 318-334.

[30] Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 307-346.

[31] 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).

[32] Nicholls, D.F., Pagan, A.R., 1983. Heteroskedasticity in models with lagged dependent variables. Econometrica 51, 1233-1242.

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

[34] Weiss, A..A.., 1986. Asymptotic theory for ARCH models: estimation and testing. Econometric Theory 2, 107-131.

[35] White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1-26.

[36] White, H. 2001. Asymptotic Theory for Econometricians. Academic Press: California

[37] 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.

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