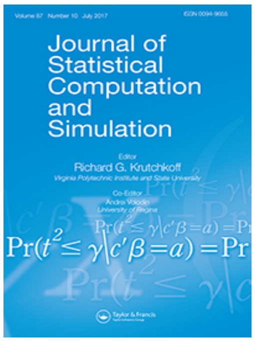

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Empirical characteristic function tests for GARCH innovation distribution using multipliers

María Dolores Jiménez-Gamero & Juan Carlos Pardo-Fernández

To cite this article: María Dolores Jiménez-Gamero & Juan Carlos Pardo-Fernández (2017) Empirical characteristic function tests for GARCH innovation distribution using multipliers, Journal of Statistical Computation and Simulation, 87:10, 2069-2093, DOI: 10.1080/00949655.2017.1313254

To link to this article: https://doi.org/10.1080/00949655.2017.1313254

Published online: 11 Apr 2017.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2017
VOL. 87, NO. 10, 2069–2093
http://dx.doi.org/10.1080/00949655.2017.1313254

Empirical characteristic function tests for GARCH innovation distribution using multipliers

María Dolores Jiménez-Gamero[a] and Juan Carlos Pardo-Fernández[b]

[a]Departamento de Estadística e I.O., Universidad de Sevilla, Spain; [b]Departamento de Estatística e I.O., Universidade de Vigo, Spain

ABSTRACT

Goodness-of-fit tests for the innovation distribution in GARCH models based on measuring deviations between the empirical characteristic function of the residuals and the characteristic function under the null hypothesis have been proposed in the literature. The asymptotic distributions of these test statistics depend on unknown quantities, so their null distributions are usually estimated through parametric bootstrap (PB). Although easy to implement, the PB can become very computationally expensive for large sample sizes, which is typically the case in applications of these models. This work proposes to approximate the null distribution through a weighted bootstrap. The procedure is studied both theoretically and numerically. Its asymptotic properties are similar to those of the PB, but, from a computational point of view, it is more efficient.

ARTICLE HISTORY
Received 18 July 2016
Accepted 27 March 2017

KEYWORDS
Characteristic function; consistency; GARCH model; goodness-of-fit; integral transformation; weighted bootstrap

1. Introduction

The class of generalized autoregressive conditional heteroscedastic (GARCH) models, introduced by Bollerslev [1], has been proved to be particularly valuable in modelling financial data. To estimate the parameters in a GARCH model, it is usually assumed that the errors or innovations have a normal distribution. Under certain not very restrictive conditions, the resultant estimator is strongly consistent and asymptotically normal, even if the errors are not normally distributed (see [2–5]). Nevertheless, for certain purposes, as observed in Klar et al. [6], Bai and Chen [7], Berkes and Horváth [8] and Koul and Ling [9], among many others, the knowledge of the true distribution of the innovations is quite convenient for several purposes (e.g. to evaluate the value at risk). Therefore, an important step in the analysis of GARCH models is to check if the data support the distributional hypothesis made on the innovations. Because of this reason, several goodness-of-fit (GOF) tests have been proposed for the innovation distribution.

The papers by Klar et al. [6] and Ghoudi and Rémillard [10] contain an extensive review of such tests as well as some numerical comparisons between them for the special case of testing for normality. In particular, Klar et al. [6] have numerically studied a test based on the empirical characteristic function (ECF) of the residuals and have compared it with other existing tests for the problem of checking normality. From the obtained numerical results, they conclude that the test based on the ECF is among of the most powerful ones. Some theoretical properties of that test have been studied in Jiménez-Gamero [11]. To approximate the null distribution of the test statistic, Klar et al. [6] have proposed to employ a PB. Although easy to implement, the PB can become very computationally expensive for large sample sizes, which is usually the case in financial data. This problem is not specific to the ECF

CONTACT María Dolores Jiménez-Gamero dolores@us.es

test in Klar et al. [6]. The same issue arises when one instead considers tests based on the empirical cumulative distribution function (ECDF). To overcome this difficulty for GOF tests based on the ECDF, Rémillard [12] has proposed to approximate the null distribution of the test statistics by a computationally more efficient estimator obtained by using a weighted bootstrap (WB), in the sense of Burke [13]. Ghoudi and Rémillard [10] have numerically compared the WB and the PB approximations for tests based on the ECDF for the problem of testing normality. They conclude that the tests based on the PB are, in general, more powerful. Nevertheless, the powers of the tests based on the PB and on the WB become quite similar for large sample sizes, but in this case the PB becomes extremely slow. Their findings coincide with those of Kojadinovic and Yan [14] and Jiménez-Gamero and Kim [15], who carried out a similar study for independent, identically distributed (IID) data for GOF tests based on the ECDF and the ECF, respectively.

In view of the good properties of the WB for ECDF-based tests for the innovations distribution in GARCH models and also for ECF-based tests for IID data, it is also expected to work well for approximating the null distribution of test statistics based on the ECF for the innovation distribution in GARCH models. Therefore, the purpose of this paper is to investigate, both theoretically and empirically, the use of a WB for approximating the null distribution of tests based on the ECF.

With this aim, the paper is organized as follows. Section 2 establishes the notation. Section 3 describes the model and the test. Section 4 is devoted to theoretically show the consistency of the WB approximation to the null distribution of the test statistic. Section 5 studies a test based on applying the integral transformation to the residuals. The approximations in the above sections are valid for a simple null hypothesis. Section 6 investigates the consistency of the WB null distribution estimator for testing a composite null hypothesis. Section 7 deals with some practical considerations such as the estimation of certain quantities required for the application of the WB approximation in practice. Section 8 displays the results of simulation experiments conducted to numerically compare the finite sample performance of the PB and the WB approximations as well as a real data application. Finally, Section 9 concludes and outlines possible extensions. Some technical results as well as the proofs are deferred to the appendices.

2. Notation

The notation employed in this paper is as follows: all vectors are column vectors; for any vector v, v_k denotes its kth coordinate and v' its transpose; if $A = (a_{jk})$ is a matrix, then $|A| = \sum_{j,k} |a_{jk}|$; for any complex number $x = a + \mathrm{i}b$, $\bar{x} = a - ib$ and $|x| = \sqrt{a^2 + b^2} = \sqrt{x\bar{x}}$; for any complex function $f(x)$, $\mathrm{Re} f(t)$ and $\mathrm{Im} f(t)$ denote the real and the imaginary parts of f, respectively, that is to say, $f(x) = \mathrm{Re} f(t) + \mathrm{i}\, \mathrm{Im} f(x)$; P_0, E_0 and Cov_0 denote probability, expectation and covariance, respectively, by assuming that the null hypothesis is true; P_*, E_* and Cov_* denote the conditional probability law, expectation and covariance, given $X_1, \dots, X_n$, respectively; all limits in this paper are taken when $n \to \infty$; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; $\xrightarrow{P}$ denotes convergence in probability; $\xrightarrow{\text{a.s.}}$ denotes the almost sure convergence; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; $L_2(w) = \{f : \mathbb{R} \to \mathbb{C} : \|f\|_w^2 = \int |f(t)|^2 w(t)\, \mathrm{d}t < \infty\}$, for some nonnegative function w satisfying $0 < \int w(t)\, \mathrm{d}t < \infty$; without loss of generality it will be assumed along the paper that $\int w(t)\, \mathrm{d}t = 1$; $\langle \cdot, \cdot \rangle$ denotes the scalar product in the Hilbert space $L_2(w)$; for any compact interval $S \subset \mathbb{R}$, $C(S)$ denotes the Banach space of continuous complex-valued functions on S with the usual sup-norm.

3. The model and the test statistic

Let $p, q \in \mathbb{N} \cup \{0\}$. A stochastic process $\{X_j,\ -\infty < j < \infty\}$ is said to follow a GARCH(p, q) model if it satisfies the equations

$$X_j = \sigma_j \varepsilon_j, \tag{1}$$

with

$$\sigma_j^2 = \sigma_j^2(\theta) = c + \sum_{k=1}^{p} a_k X_{j-k}^2 + \sum_{l=1}^{q} b_l \sigma_{j-l}^2, \tag{2}$$

for $-\infty < j < \infty$, where $\theta = (c, a_1, \ldots, a_p, b_1, \ldots, b_q)'$, with $c > 0$, $a_k \geq 0$ and $b_l \geq 0$. If $q = 0$ then we get an autoregressive conditional heteroscedastic model, introduced by Engle [16]. Bougerol and Picard [17,18] have given necessary and sufficient conditions for the existence of a unique strictly stationary ergodic solution of Equations (1) and (2). Throughout this paper it will be assumed that $\{X_j, -\infty < j < \infty\}$ satisfies Equations (1) and (2), that it is stationary, that $\{\varepsilon_j, -\infty < j < \infty\}$ are IID from a non-degenerate random variable ε, with $E(\varepsilon) = 0$ and $E(\varepsilon^2) = 1$, and that ε_j is independent of $\{X_{j-k}, k \geq 1\}$. We will also assume along the paper that the representation in Equation (2) is unique, which is ensured by assuming that the polynomials $\mathcal{A}(z) = \sum_{k=1}^{p} a_k z^k$ and $\mathcal{B}(z) = 1 - \sum_{l=1}^{q} b_l z^l$ (with $\mathcal{A}(z) = 0$ if $p = 0$ and $\mathcal{B}(z) = 0$ if $q = 0$) have no common roots (see [3]).

Let $r = 1+p+q$ denote the dimension of θ. θ is assumed to be fixed but unknown. It is also assumed that $\theta \in \Theta_0 = \Theta(\rho_0, \rho_1, \rho_2) = \{u = (\gamma, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) : \beta_1 + \cdots + \beta_q \leq \rho_0, \rho_1 \leq \min\{\gamma, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\} \leq \max\{\gamma, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\} \leq \rho_2\}$, for some constants ρ_0, ρ_1, ρ_2 satisfying $0 < \rho_0 < 1$, $0 < \rho_1 < \rho_2$, $q\rho_1 \leq \rho_0$. Note that this assumption requires p and q to be known and rules out zero coefficients in θ. This is required for the asymptotic normality of the estimator of θ discussed below.

Let $X_1, \ldots, X_n$ be a realization of length n of a GARCH(p,q) model. A commonly used estimator of θ is the Gaussian maximum likelihood estimator (GMLE), $\hat{\theta}_G$. If

$$E(\varepsilon^4) < \infty, \tag{3}$$

then $\sqrt{n}(\hat{\theta}_G - \theta)$ is asymptotically normally distributed, even if the errors are not normally distributed, see [2,4]. Moreover, even if (3) does not hold then, under certain conditions, $n^{\kappa}(\hat{\theta}_G - \theta)$ is bounded in probability, for some $\kappa > 0$, see [2]. Although the GMLE has become the most popular estimator, other estimators have been proposed. Examples are the estimators in Peng and Yao [19], which are asymptotically normally distributed without requiring (3), and those in Berkes and Horváth [8], where a class of estimators including the GMLE is studied. From now on, we will denote by $\hat{\theta}$ any estimator of θ. It will be assumed that $\hat{\theta}$ satisfies the following:

(A.1) $\hat{\theta}$ can be expressed as

$$\hat{\theta} = \theta + n^{-1} \sum_{j=1}^{n} L_j(\theta) + o_P(n^{-1/2}),$$

where $L_j(\theta) = (g_1(\varepsilon_j) l_1(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots), \ldots, g_r(\varepsilon_j) l_r(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots))'$, $1 \leq j \leq n$,

$$E\{g_u(\varepsilon_0)\} = 0, \quad E\{g_u(\varepsilon_0)^2\} < \infty, \quad E\{l_u(\varepsilon_{-1}, \varepsilon_{-2}, \ldots)^2\} < \infty, \quad 1 \leq u \leq r.$$

The GMLE as well as other often used estimators of θ satisfy (A.1) (see Section 3 of [8]). If $\hat{\theta}$ satisfies (A.1) then, by the martingale central limit theorem, $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N_r(0, \Sigma_\theta)$, an r-variate zero-mean normal law with variance-covariance matrix $\Sigma_\theta = \text{var}\{L_0(\theta)\} = (\varsigma_{uv})$, with $\varsigma_{uv} = E\{g_u(\varepsilon_0) g_v(\varepsilon_0)\} E\{l_u(\varepsilon_{-1}, \varepsilon_{-2}, \ldots) l_v(\varepsilon_{-1}, \varepsilon_{-2}, \ldots)\}$, $1 \leq u, v \leq r$.

In a GARCH model, the errors are not observable. Thus, to make inferences on the errors, we must approximate them by means of the residuals. With this aim, first we have to estimate $\sigma_j^2(\theta)$. Note that $\sigma_j^2(\theta)$ depends on $\{X_k, -\infty < k \leq j-1\}$, whereas we only observe $X_1, \ldots, X_n$. So, in order to calculate the residuals, instead of $\sigma_j^2(\hat{\theta})$, $1 \leq j \leq n$, we consider $\tilde{\sigma}_j^2(\hat{\theta})$, $1 \leq j \leq n$, recursively defined as in Equation (2) for arbitrary values $X_0, X_{-1}, \ldots, X_{1-p}, \tilde{\sigma}_0^2, \tilde{\sigma}_{-1}^2, \ldots, \tilde{\sigma}_{1-q}^2$ such as

$X_0^2 = \cdots = X_{1-p}^2 = \tilde{\sigma}_0^2 = \cdots = \tilde{\sigma}_{1-q}^2 = \hat{c}$ or $X_0^2 = \cdots = X_{1-p}^2 = \tilde{\sigma}_0^2 = \cdots = \tilde{\sigma}_{1-q}^2 = X_1^2$; another common choice is $\tilde{\sigma}_1 = \cdots = \tilde{\sigma}_m = \varsigma$, for some $\varsigma > 0$, $m = \max\{p, q\}$, and $\tilde{\sigma}_j$ following the recursion in Equation (2) for $j > m$. Let $\{\tilde{\varepsilon}_j = X_j/\tilde{\sigma}_j(\hat{\theta}),\ 1 \le j \le n\}$ be the residuals and let $\varphi_{n,\nu}(t)$ denote the ECF of the residuals $\tilde{\varepsilon}_{\nu+1}, \ldots, \tilde{\varepsilon}_n$

$$\varphi_{n,\nu}(t) = \frac{1}{n-\nu} \sum_{j=\nu+1}^{n} \exp\{it\tilde{\varepsilon}_j\},$$

for some integer $\nu \ge 1$. The reason for only considering the residuals $\tilde{\varepsilon}_{\nu+1}, \ldots, \tilde{\varepsilon}_n$, instead of all of them, $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$, is that for small j, $\tilde{\sigma}_j^2(\theta)$ is not a good approximation to $\sigma_j^2(\theta)$, and thus early terms in the series should be avoided for inferential purposes.

Let us consider the problem of testing for the null hypothesis

$$H_0: \text{ the CDF of } \varepsilon \text{ is } F_0,$$

where F_0 is a completely specified CDF, or equivalently

$$H_0: \text{ the CF of } \varepsilon_0 \text{ is } \varphi_0,$$

where φ_0 is the CF associated to F_0, $\varphi_0(t) = \int \exp(itu)\, dF_0(u)$. For this problem, Jiménez-Gamero [11] has shown that the test

$$\Psi = \Psi(X_1, \ldots, X_n) = \begin{cases} 1, & \text{if } R_{n,\nu} \ge r_\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

based on the test statistic

$$R_{n,\nu} = \| W_{n,\nu} \|_w^2,$$

where r_α is the $1-\alpha$ percentile of the null distribution of $R_{n,\nu}$, or a consistent approximation to it, is consistent against all fixed alternatives, whenever $w(t) > 0$, $\forall t \in \mathbb{R}$.

Observe that we can assume that the weight function w involved in the definition of the test statistic $R_{n,\nu}$ satisfies

$$w(t) = w(-t), \quad \forall t \in \mathbb{R}. \tag{4}$$

Otherwise, by defining $w_1(t) = 0.5\{w(t) + w(-t)\}$, which satisfies Equation (4), we have that $\| W_{n,\nu} \|_w = \| W_{n,\nu} \|_{w_1}$. Therefore, from now on we will assume that Equation (4) holds. In such a case, we can write

$$\| W_{n,\nu} \|_w^2 = \| \mathrm{Re}\, W_{n,\nu} + \mathrm{Im}\, W_{n,\nu} \|_w^2.$$

The exact null distribution of $R_{n,\nu}$ cannot be calculated. Moreover, its asymptotic null distribution cannot be used as an approximation because it depends on unknowns (see Remark 1 in [11]). To estimate the null distribution of $R_{n,\nu}$, [6] have proposed considering a PB algorithm. The consistency of this approximation has been derived in [11]. Nevertheless, from a computational point of view, it is rather inefficient, as it is very time consuming.

Since

$$R_{n,\nu} = \frac{1}{n-\nu} \sum_{j,k=\nu+1}^{n} h(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k), \quad h(x,y) = \int q(x,t) q(y,t) w(t)\, dt \tag{5}$$

with

$$q(x,t) = \cos(tx) - \mathrm{Re}\, \varphi_0(t) + \sin(tx) - \mathrm{Im}\, \varphi_0(t), \tag{6}$$

the test statistic $(1/(n-\nu))R_{n,\nu}$ is a degree-2 V-statistic evaluated on the residuals. In the statistical literature there are several papers dealing with the consistency of the WB distribution estimator of

U-statistics and V-statistics evaluated on IID data. Let $Z_1, \dots, Z_n$ be IID and let

$$V_n(\mathfrak{h}) = \frac{1}{n^2} \sum_{j,k=1}^{n} \mathfrak{h}(Z_j, Z_k)$$

be a degree-2 V-statistic. Assume that it is degenerate, that is, that $E\{\mathfrak{h}(Z_1, x)\} - E\{\mathfrak{h}(Z_1, Z_2)\} = 0$. Delhing and Mikosch [20] (see also Hušková and Janssen [21]) showed that if $\xi_1, \dots, \xi_n$ are IID from ξ with $E(\xi) = 0$ and $\text{var}(\xi) = 1$, independent of $Z_1, \dots, Z_n$, then the conditional distribution, given $Z_1, \dots, Z_n$, of

$$\frac{1}{n} \sum_{j,k=1}^{n} \mathfrak{h}(Z_j, Z_k)\xi_j\xi_k$$

consistently estimates that of $nV_n(\mathfrak{h})$. In the light of this result, since the residuals are an approximation for the innovations, which are IID variables, one may try to estimate the null distribution of $R_{n,\nu}$ by means of the conditional distribution, given $X_1, \dots, X_n$, of

$$R^*_{0,n,\nu} = \frac{1}{n-\nu} \sum_{j,k=\nu+1}^{n} h(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k)\xi_j\xi_k. \tag{7}$$

The next result gives the conditional asymptotic distribution, given $X_1, \dots, X_n$, of $R^*_{0,n,\nu}$. Unfortunately, this distribution does not approximate the null distribution of the test statistic $R_{n,\nu}$ properly, as discussed below.

Theorem 3.1: *Suppose that $\theta \in \Theta_0$, $\hat{\theta}$ satisfies* (A.1), *w is a non-negative function satisfying $\int t^2 w(t)\,\mathrm{d}t < \infty$ and $\nu = \nu(n)$ is an integer satisfying,*

$$\nu/n \to 0. \tag{8}$$

Then,

$$\sup_x |P_*(R^*_{0,n,\nu} \le x) - P(\|W\|_w^2 \le x)| \xrightarrow{P} 0,$$

where $\{W(t),\, t \in \mathbb{R}\}$ is a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_W(s,t) = E\{q(\varepsilon,t)q(\varepsilon,s)\}$, $\forall t, s \in \mathbb{R}$, and q is as defined in Equation (6).

Theorem 2 in [11] shows that if H_0 is true, ν satisfies Equation (8) and w is a non-negative function satisfying

$$\int t^4 w(t)\,\mathrm{d}t < \infty, \tag{9}$$

then $R_{n,\nu} \xrightarrow{\mathcal{L}} \|W_0\|_w^2$, where $W_0(t)$ is a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_{W_0}(s,t) = E_0[\{C(t) + S(t)\}\{C(s) + S(s)\}]$, $\forall t, s \in \mathbb{R}$, with

$$C(t) = \cos(t\varepsilon) - \text{Re}\,\varphi_0(t) - 0.5t\mu_c(t)\mu_{0A}(\theta)\}' L_1(\theta),$$
$$S(t) = \sin(t\varepsilon) - \text{Im}\,\varphi_0(t) - 0.5t\mu_s(t)\mu_{0A}(\theta)\}' L_1(\theta),$$

$\mu_c(t) = (\partial/\partial t)\,\text{Re}\,\varphi_0(t) = E_0\{-\varepsilon\sin(t\varepsilon)\}$, $\mu_s(t) = (\partial/\partial t)\,\text{Im}\varphi_0(t) = E_0\{\varepsilon\cos(t\varepsilon)\}$ and $\mu_{0A}(\theta) = E_0\{A_1(\theta)\} = \cdots = E_0\{A_n(\theta)\}$, where $\sigma_j^2(\theta)A_j(\theta)$ is the r-vector of derivatives of $\sigma_j^2(\theta)$ with respect to θ, that is, $A_j(\theta) = (1/\sigma_j^2(\theta))(\partial/\partial\theta)\sigma_j^2(\theta)$, for any j. Therefore, from Theorem 3.1, it clearly follows that the conditional distribution, given $X_1, \dots, X_n$, of $R^*_{0,n,\nu}$ does not provide a consistent estimator of the null distribution of $R_{n,\nu}$, because replacing θ by $\hat{\theta}$ has an effect on its asymptotic null distribution that it is not captured by the conditional distribution of $R^*_{0,n,\nu}$. As a consequence, the statistic to be bootstrapped should take into account such an effect. This is investigated in the next section.

4. The WB approximation

From the proof of Theorem 2 in [11], it follows that

$$R_{n,\nu} = R_{1,n,\nu} + o_P(1),$$

where

$$\begin{aligned} R_{1,n,\nu} &= \| W_{1,n,\nu} \|_w^2, \\ W_{1,n,\nu} &= \frac{1}{\sqrt{n-\nu}} \sum_{j=\nu+1}^{n} \{C_j(t) + S_j(t)\}, \\ C_j(t) &= \cos(t\varepsilon_j) - \operatorname{Re}\varphi_0(t) - 0.5t\mu_c(t)\mu_A(\theta)' L_j(\theta), \\ S_j(t) &= \sin(t\varepsilon_j) - \operatorname{Im}\varphi_0(t) - 0.5t\mu_s(t)\mu_A(\theta)' L_j(\theta), \quad \nu+1 \le j \le n, \\ \mu_A(\theta) &= E\{A_1(\theta)\} = \cdots = E\{A_n(\theta)\}. \end{aligned}$$

Let us consider the following WB version of $R_{1,n,\nu}$,

$$\begin{aligned} R^*_{2,n,\nu} &= \| W^*_{2,n,\nu} \|_w^2, \\ W^*_{2,n,\nu} &= \frac{1}{\sqrt{n-\nu}} \sum_{j=\nu+1}^{n} \{\hat{C}_j(t) + \hat{S}_j(t)\}\xi_j, \\ \hat{C}_j(t) &= \cos(t\tilde{\varepsilon}_j) - \operatorname{Re}\varphi_0(t) - 0.5t\mu_c(t)\widehat{\mu_A(\theta)}' \widehat{L_j(\theta)}, \\ \hat{S}_j(t) &= \sin(t\tilde{\varepsilon}_j) - \operatorname{Im}\varphi_0(t) - 0.5t\mu_s(t)\widehat{\mu_A(\theta)}' \widehat{L_j(\theta)}, \quad \nu+1 \le j \le n, \end{aligned}$$

$\xi_{\nu+1}, \ldots, \xi_n$ are IID from ξ with $E(\xi) = 0$ and $var(\xi) = 1$, independent of $X_1, \ldots, X_n$,

$$\widehat{\mu_A(\theta)} = \frac{1}{n} \sum_{j=1}^{n} \tilde{A}_j(\hat{\theta}), \quad \tilde{A}_j(\theta) = \frac{1}{\tilde{\sigma}_j^2(\theta)} \frac{\partial}{\partial\theta} \tilde{\sigma}_j^2(\theta),$$

and $\widehat{L_1(\theta)}, \ldots, \widehat{L_n(\theta)}$ satisfy

$$\frac{1}{n} \sum_{j=1}^{n} |L_j(\theta) - \widehat{L_j(\theta)}|^2 \xrightarrow{P} 0. \tag{10}$$

A candidate for $\widehat{L_j(\theta)}$ when θ is estimated by the GMLE will be discussed later in Section 7. The following result gives the conditional asymptotic distribution, given $X_1, \ldots, X_n$, of $R^*_{2,n,\nu}$.

Theorem 4.1: *Suppose that $\theta \in \Theta_0$, $\hat{\theta}$ satisfies* (A.1), *w satisfies Equation* (9) *and Equation* (10) *holds. Then*

$$\sup_x |P_*(R^*_{2,n,\nu} \le x) - P(\| W_1 \|_w^2 \le x)| \xrightarrow{P} 0,$$

where $\{W_1(t),\ t \in \mathbb{R}\}$ is a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_{W_1}(s,t) = E[\{C_1(t) + S_1(t)\}\{C_1(s) + S_1(s)\}]$, $\forall t, s \in \mathbb{R}$.

The result in Theorem 4.1 is valid whether or not the null hypothesis is true. Two immediate consequences follow.

Corollary 4.2: *If H_0 is true and the assumptions in Theorem* 4.1 *hold then*

$$\sup_x |P_*\{R^*_{2,n,\nu} \le x\} - P_0\{R_{n,\nu} \le x\}| \xrightarrow{P} 0.$$

Let $\alpha \in (0,1)$ and

$$\Psi^* = \Psi^*(X_1,\dots,X_n) = \begin{cases} 1, & \text{if } R_{n,\nu} \ge r^*_\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

where r^*_α is the $1-\alpha$ percentile of the conditional distribution of $R^*_{2,n,\nu}$, given $X_1,\dots,X_n$, or equivalently, $\Psi^* = 1$ if $p^* \le \alpha$, where $p^* = P_*\{R^*_{2,n,\nu} \ge R_{n,\nu,obs}\}$, $R_{n,\nu,obs}$ being the observed value of the test statistic $R_{n,\nu}$. The result in Corollary 4.2 states that the test Ψ^* is asymptotically correct, in the sense that its type I error is asymptotically equal to the nominal level α.

Corollary 4.3: *If H_0 is not true, the assumptions in Theorem* 4.1 *hold and w is such that*

$$\int |\varphi(t) - \varphi_0(t)|^2 w(t)\,dt > 0, \tag{11}$$

where φ denotes the CF of the innovations, then $P(\Psi^* = 1) \to 1$.

Since two distinct characteristic functions can be equal in a finite interval [22, p.479], a general way to ensure (11) whenever $\varphi \ne \varphi_0$ is to take w positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$. Thus, if $w(t) > 0, \forall t \in \mathbb{R}$, then Corollary 4.3 states that the test Ψ^* is consistent in the sense of being able to asymptotically detect any alternative.

Remark 4.4: The PB null distribution estimator of $R_{n,\nu}$ satisfies a result similar to that stated in Theorem 4.1 for the WB estimator, but the Gaussian process in the limit has covariance kernel $K_{W_0}(s,t)$ (see Theorem 5 in [11]). If H_0 is true then $K_{W_0}(s,t) = K_{W_1}(s,t)$, $\forall s,t$, but in general $K_{W_0}(s,t)$ and $K_{W_1}(s,t)$ may not coincide. Thus, although the test Ψ^* and the one obtained by approximating r_α by its PB estimator are both consistent against all fixed alternatives, their powers could differ for finite sample sizes.

Remark 4.5: It is also worth observing that the assumptions in [11] for the PB to work, in the sense of providing a consistent approximation of the null distribution, are stronger than those assumed in Theorem 4.1 for the validity of the WB.

Remark 4.6: A problem with the PB is that when $\hat{a}_1 + \cdots + \hat{a}_p + \hat{b}_1 + \cdots + \hat{b}_q$ is close to 1, then for a high percentage of bootstrap samples it will happen that $\hat{a}^*_1 + \cdots + \hat{a}^*_p + \hat{b}^*_1 + \cdots + \hat{b}^*_q > 1$ thus leading to a non-stationary behaviour. This problem is avoided by using the WB, since this mechanism does not require to estimate the GARCH parameters from the resamples.

Remark 4.7: The results stated so far keep on being true if the raw multipliers, $\xi_{\nu+1},\dots,\xi_n$, are replaced by the centred multipliers, $\xi_{\nu+1} - \bar{\xi},\dots,\xi_n - \bar{\xi}$, as suggested in [13,14], where $\bar{\xi} = (1/(n-\nu))\sum_{j=\nu+1}^n \xi_j$.

5. A test based on transformation

Another problem related to the test Ψ is the calculation of the test statistic $R_{n,\nu}$. From expression (5) it follows that closed–form expressions for $R_{n,\nu}$ would be possible only for certain distributions and certain choices of w. For example, if the distribution in H_0 is the standard normal and w is taken as the probability density function (PDF) of a normal law, then the kernel h in Equation (5) has a closed expression, and thus $R_{n,\nu}$ can be easily calculated, see [6]. An example of interest in finance is testing if the innovations have a t_g-distribution, for some fixed g (see, e.g. [7]), but, unfortunately, it is rather difficult to find a weight function w so that the kernel h has a closed expression. In order to alleviate this problem, Meintanis et al. [23] have proposed to transform the original data in such a way that the transformed data follow a distribution for which the kernel h may be easily calculated. Specifically, assuming that the data are continuous and univariate, they propose to apply the integral transformation: if the random variable ε has a continuous CDF F, then $U = F(\varepsilon)$ has a uniform distribution on the interval $(0, 1)$. In our setting, the innovations are not observable, and thus we must transform the residuals. Let

$$U_j = F_0(\tilde{\varepsilon}_j),$$

$1 \le j \le n$, and consider the test statistic

$$T_{n,\nu} = \|Z_{n,\nu}\|_w^2,$$

where $Z_{n,\nu}(t) = \sqrt{n-\nu}\{\phi_{n-\nu}(t) - \phi_0(t)\}$, $\phi_{n-\nu}(t) = (1/(n-\nu))\sum_{j=\nu+1}^{n} \exp\{\mathrm{i}tU_j\}$ and $\phi_0(t)$ is the CF of a uniform distribution on $(0, 1)$. Some properties of the ECF of the transformed data $\phi_{n-\nu}(t)$ are studied in Appendix A.1.

The expression of the test statistic $T_{n,\nu}$ can be easily calculated for several weight functions. Although the null distribution of $T_{n,\nu}$ could be approximated by means of a PB, because of the same reasons argued for $R_{n,\nu}$, we next investigate a WB approximation. A similar reasoning to that employed in Section 4 leads us to define

$$\begin{aligned}
T^*_{2,n,\nu} &= \|Z^*_{2,n,\nu}\|_w^2,\\
Z^*_{2,n,\nu}(t) &= \frac{1}{\sqrt{n-\nu}}\sum_{j=\nu+1}^{n}\{\hat{\mathcal{C}}_j(t) + \hat{\mathcal{S}}_j(t)\}\xi_j,\\
\hat{\mathcal{C}}_j(t) &= \cos\{tF_0(\tilde{\varepsilon}_j)\} - \mathrm{Re}\phi_0(t) - 0.5t\mu_R(t)\widehat{\mu_A(\theta)}'\widehat{L_j(\theta)},\\
\hat{\mathcal{S}}_j(t) &= \sin\{tF_0(\tilde{\varepsilon}_j)\} - \mathrm{Im}\phi_0(t) - 0.5t\mu_I(t)\widehat{\mu_A(\theta)}'\widehat{L_j(\theta)}, \quad \nu+1 \le j \le n,
\end{aligned}$$

where $\mu_R(t) = -E[\varepsilon f_0(\varepsilon)\sin\{tF_0(\varepsilon)\}]$, $\mu_I(t) = E[\varepsilon f_0(\varepsilon)\cos\{tF_0(\varepsilon)\}]$, f_0 is the PDF associated to F_0 and $\xi_{\nu+1}, \dots, \xi_n$ are as before. The next result gives the conditional asymptotic distribution of $T^*_{2,n,\nu}$, given $X_1, \dots, X_n$.

Theorem 5.1: *Suppose that $\theta \in \Theta_0$, $\hat{\theta}$ satisfies* (A.1), *ν satisfies Equation* (8), *F_0 is a continuous CDF with bounded PDF f_0, f_0 has a bounded derivative, w satisfies Equation* (9) *and Equation* (10) *holds. Then*

$$\sup_x |P_*(T^*_{2,n,\nu} \le x) - P(\|Z_1\|_w^2 \le x)| \xrightarrow{P} 0,$$

where $\{Z_1(t),\ t \in \mathbb{R}\}$ is a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_{Z_1}(s,t) = E[\{\mathcal{C}_1(t) + \mathcal{S}_1(t)\}\{\mathcal{C}_1(s) + \mathcal{S}_1(s)\}]$, $\forall t, s \in \mathbb{R}$, with

$$\begin{aligned}
\mathcal{C}_1(t) &= \cos\{tF_0(\varepsilon_1)\} - \mathrm{Re}\phi_0(t) - 0.5t\mu_R(t)\mu_A(\theta)'L_1(\theta),\\
\mathcal{S}_1(t) &= \sin\{tF_0(\varepsilon_1)\} - \mathrm{Im}\phi_0(t) - 0.5t\mu_I(t)\mu_A(\theta)'L_1(\theta).
\end{aligned}$$

As a consequence of Theorem 5.1, similar corollaries and remarks to those stated after Theorem 4.1 can be given now. To save space we omit them and only underline that the test

$$\Upsilon^* = \Upsilon^*(X_1, \dots, X_n) = \begin{cases} 1, & \text{if } T_{n,\nu} \geq t^*_\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

is asymptotically correct and consistent against all fixed alternatives, whenever $w(t) > 0$, $\forall t \in \mathbb{R}$, where t^*_α is the $1-\alpha$ percentile of the conditional distribution of $T^*_{2,n,\nu}$, given $X_1, \dots, X_n$, or equivalently, $\Upsilon^* = 1$ if $p^* \leq \alpha$, where $p^* = P_*\{T^*_{2,n,\nu} \geq T_{n,\nu,\text{obs}}\}$, $T_{n,\nu,\text{obs}}$ being the observed value of the test statistic $T_{n,\nu}$.

6. Composite null hypothesis

So far we have studied the case of a simple null hypothesis. This section shows how the obtained results can be extended for testing GOF to a composite null hypothesis. Let $\mathcal{F} = \{F(\cdot;\gamma),\ \gamma \in \Gamma \subseteq \mathbb{R}^m\}$ be a parametric family of CDFs. Let $\varphi(\cdot;\gamma)$ denote the CF associated with the CDF $F(\cdot;\gamma)$. By analogy with the case of a simple null hypothesis, to test for

$$H_0 : F \in \mathcal{F},$$

we could consider the test statistic $R_{n,\nu}(\hat{\gamma}) = \|W_{n,\nu}(\cdot;\hat{\gamma})\|^2_w$, with $W_{n,\nu}(t;\gamma) = \sqrt{n-\nu}\{\varphi_{n,\nu}(t) - \varphi(t;\gamma)\}$ and $\hat{\gamma} = \hat{\gamma}(\tilde{\varepsilon}_{\nu+1}, \dots, \tilde{\varepsilon}_n)$ consistently estimates γ. Because of the reasons explained in Section 5, we instead consider $T_{n,\nu}(\hat{\gamma})$ defined as $T_{n,\nu}(\hat{\gamma}) = \|Z_{n,\nu}(\cdot;\hat{\gamma})\|^2_w$, with $Z_{n,\nu}(t;\gamma) = \sqrt{n-\nu}\{\phi_{n,\nu}(t;\gamma) - \phi_0(t)\}$, $\phi_{n,\nu}(t;\gamma) = \frac{1}{n-\nu}\sum_{j=\nu+1}^n \exp\{itU_j(\gamma)\}$ and $U_j(\gamma) = F(\tilde{\varepsilon}_j;\gamma)$, $\nu+1 \leq j \leq n$.

To study the behaviour of $T_{n,\nu}(\hat{\gamma})$, some assumptions will be required on $\hat{\gamma}$ and on the CDFs in the family $\mathcal{F}$, which are listed below.

(A.2) $\hat{\gamma} \xrightarrow{P} \gamma_0$, for some $\gamma_0 \in \text{int}\Gamma$.

(A.3) $(\partial/\partial x)F(x;\gamma) = f(x;\gamma)$ and $(\partial/\partial\gamma)F(x;\gamma) = D_1F(x;\gamma)$ exist and are bounded $\forall x$, $\forall\gamma \in \Gamma_0 \subset \Gamma$, where Γ_0 is an open neighbourhood of γ_0.

(A.4) The second-order derivatives of $F(x;\gamma)$ with respect to x and γ exist and are bounded $\forall x$, $\forall\gamma \in \Gamma_0 \subset \Gamma$, where Γ_0 is an open neighbourhood of γ_0.

(A.5) When H_0 is true, $\hat{\gamma} = \gamma_0 + n^{-1}\sum_{j=1}^n l(\varepsilon_j;\gamma_0) + M(\theta,\gamma_0)(\hat{\theta}-\theta) + o_P(n^{-1/2})$, with $E_0\{l(\varepsilon_j;\gamma_0)\} = 0$, $E_0\{\|l(\varepsilon_j;\gamma_0)\|^2\} < \infty$ and $M(\theta,\gamma_0)$ is a $m \times p$-matrix of constants that may depend on θ and γ_0.

Since the innovations are not observable, it seems reasonable to treat the residuals as if they were the true errors and then apply some method designed for IID data to estimate γ. In Section 7 it will be seen that if the considered method is maximum likelihood, then the resulting estimator satisfies (A.5).

Some asymptotic properties of $T_{n,\nu}(\hat{\gamma})$ are given in Appendix A.2. Next we investigate a WB approximation to the null distribution of $T_{n,\nu}(\hat{\gamma})$. A similar reasoning to that employed in Section 4 leads us to define

$$T^*_{2,n,\nu}(\gamma) = \|Z^*_{2,n,\nu}(\cdot;\gamma)\|^2_w,$$

$$Z^*_{2,n,\nu}(t;\gamma) = \frac{1}{\sqrt{n-\nu}}\sum_{j=\nu+1}^n \{\hat{C}_j(t;\gamma) + \hat{S}_j(t;\gamma)\}\xi_j,$$

$$\hat{C}_j(t;\gamma) = \cos\{tF(\tilde{\varepsilon}_j;\gamma)\} - \text{Re}\phi_0(t) - 0.5t\mu_{1R}(t;\gamma)\widehat{\mu_A(\theta)}'\widehat{L_j(\theta)}$$
$$+ t\mu_{2R}(t;\gamma)'\{\hat{l}(\tilde{\varepsilon}_j;\gamma) + \widehat{M(\theta,\gamma)}\widehat{L_j(\theta)}\},$$

$$\hat{\mathcal{S}}_j(t;\gamma) = \sin\{tF(\tilde{\varepsilon}_j;\gamma)\} - \mathrm{Im}\phi_0(t) - 0.5t\mu_{1I}(t;\gamma)\widehat{\mu_A(\theta)}'\widehat{L_j(\theta)} + t\mu_{2I}(t;\gamma)'\{\hat{l}(\tilde{\varepsilon}_j;\gamma) + \widehat{M(\theta,\gamma)}\widehat{L_j(\theta)}\}, \quad \nu+1 \le j \le n,$$

where $\xi_{\nu+1},\ldots,\xi_n$ are as before, $\mu_{1R}(t;\gamma) = -E[\varepsilon f(\varepsilon;\gamma)\sin\{tF(\varepsilon;\gamma)\}]$, $\mu_{1I}(t;\gamma) = E[\varepsilon f(\varepsilon;\gamma) \cos\{tF(\varepsilon;\gamma)\}]$, $\mu_{2R}(t;\gamma) = -E[\sin\{tF(\varepsilon;\gamma)\}D_1F(\varepsilon;\gamma)]$, $\mu_{2I}(t;\gamma) = E[\cos\{tF(\varepsilon;\gamma)\}D_1F(\varepsilon;\gamma)]$, $\widehat{M(\theta,\gamma)}$ satisfies

$$\widehat{M(\theta,\gamma)} \xrightarrow{P} M(\theta,\gamma_0), \tag{12}$$

and $\hat{l}(\varepsilon_j;\gamma)$ is such that

$$\frac{1}{n}\sum_{j=1}^{n}\|\hat{l}(\tilde{\varepsilon}_j;\hat{\gamma}) - l_1(\varepsilon_j;\gamma_0)\|^2 \xrightarrow{P} 0,$$
$$\text{with } E\{\|l_1(\varepsilon;\gamma_0)\|^2\} < \infty \quad \text{and} \quad l_1(\varepsilon;\gamma_0) = l(\varepsilon;\gamma_0) \quad \text{if } H_0 \text{ is true.} \tag{13}$$

The next result gives the conditional asymptotic distribution of $T^*_{2,n,\nu}(\hat{\gamma})$, given $X_1,\ldots,X_n$.

Theorem 6.1: *Suppose that $\theta \in \Theta_0$, ν satisfies Equation* (8), *w satisfies Equation* (9) *and* (A.1)–(A.5), (10), (12) *and* (13) *hold. Then*

$$\sup_x |P_*(T^*_{2,n,\nu}(\hat{\gamma}) \le x) - P(\|Z_1(\cdot;\gamma_0)\|^2_w \le x)| \xrightarrow{P} 0,$$

where $\{Z_1(t;\gamma_0),\ t\in\mathbb{R}\}$ is a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_{Z_1}(s,t;\gamma_0) = E[\{\mathcal{C}_1(t;\gamma_0)+\mathcal{S}_1(t;\gamma_0)\}\{\mathcal{C}_1(s;\gamma_0)+\mathcal{S}_1(s;\gamma_0)\}]$, $\forall t,s\in\mathbb{R}$, with

$$\mathcal{C}_1(t;\gamma) = \cos\{tF(\varepsilon_1;\gamma)\} - \mathrm{Re}\phi(t) - 0.5t\mu_{1R}(t;\gamma)\mu_A(\theta)'L_1(\theta) + t\mu_{2R}(t;\gamma)'\{l_1(\varepsilon_1;\gamma) + M(\theta,\gamma)L_1(\theta)\},$$
$$\mathcal{S}_1(t;\gamma) = \sin\{tF(\varepsilon_1;\gamma)\} - \mathrm{Im}\phi(t) - 0.5t\mu_{1I}(t;\gamma)\mu_A(\theta)'L_1(\theta) + t\mu_{2I}(t;\gamma)'\{l_1(\varepsilon_1;\gamma) + M(\theta,\gamma)L_1(\theta)\}.$$

As a consequence of Theorem 6.1, similar corollaries and remarks to those stated after Theorem 4.1 can be given now. For the sake of brevity, we omit them and only underline that the test

$$\Omega^* = \Omega^*(X_1,\ldots,X_n) = \begin{cases} 1, & \text{if } T_{n,\nu}(\hat{\gamma}) \ge t^*_\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

is asymptotically correct and consistent against all fixed alternatives, whenever $w(t) > 0$, $\forall t\in\mathbb{R}$, where t^*_α is the $1-\alpha$ percentile of the conditional distribution of $T^*_{2,n,\nu}(\hat{\gamma})$, given $X_1,\ldots,X_n$, or equivalently, $\Omega^* = 1$ if $p^* \le \alpha$, where $p^* = P_*\{T^*_{2,n,\nu}(\hat{\gamma}) \ge T_{n,\nu,obs}\}$, $T_{n,\nu,obs}$ being the observed value of the test statistic $T_{n,\nu}(\hat{\gamma})$.

7. Some practical considerations

7.1. On the estimation of $L_j(\theta)$ when $\hat{\theta}$ is the GLME

After stating (A.1) we mentioned that this assumption is satisfied by the GLME as well as other estimators of θ. Since the GLME is calculated by most statistical packages and programming languages, this subsection deals with the estimation of $L_j(\theta)$ by $\widehat{L_j(\theta)}$ so that Equation (10) holds, for this estimator. In such a case (see, e.g. [3,4]) the expansion in (A.1) holds with

$$L_j = L_j(\theta) = (\varepsilon_j^2 - 1)A_j(\theta)J^{-1}, \quad 1 \le j \le n,$$

where $J = E\{(\varepsilon_1^2 - 1)^2\}E\{A_1(\theta)A_1(\theta)'\}$. All unknown quantities in the expression of L_j must be replaced by adequate estimators. Let

$$\hat{L}_j = (\tilde{\varepsilon}_j^2 - 1)\tilde{A}_j(\hat{\theta})\hat{J}^{-1}, \quad 1 \le j \le n,$$

where $\hat{J} = \frac{1}{n}\sum_{j=1}^{n}(\tilde{\varepsilon}_j^2 - 1)^2\tilde{A}_j(\hat{\theta})\tilde{A}_j(\hat{\theta})'$. The next result shows that $\hat{L}_j$ provides a suitable approximation for L_j, in the sense that Equation (10) holds.

Proposition 7.1: *If $E(\varepsilon^4) < \infty$ and $\hat{\theta}$ is the GLME, then $\{\hat{L}_j,\ 1 \le j \le n\}$ satisfy Equation* (10).

7.2. On the calculation of $\widehat{\mu_A(\theta)}$ and $\hat{L}_j$

Observe that when $\hat{\theta}$ is the GMLE and $\{\hat{L}_j,\ 1 \le j \le n\}$ are as in the previous subsection, the practical calculation of $\widehat{\mu_A(\theta)}$ and $\{\hat{L}_j,\ 1 \le j \le n\}$ can be done as follows:

(1) Calculate the GMLE, $\hat{\theta}$.
(2) Take, for instance, $\tilde{\sigma}_1 = \cdots = \tilde{\sigma}_m = \varsigma$, for some $\varsigma > 0$, $m = \max\{p, q\}$, and $\tilde{\sigma}_j$ following the recursion in Equation (2) for $j > m$ with $\theta = \hat{\theta}$.
(3) Calculate $\tilde{\varepsilon}_j = X_j/\tilde{\sigma}_j$, $1 \le j \le n$.
(4) Recursively calculate

$$d_j = \begin{cases} (0, \ldots, 0)' \in \mathbb{R}^r, & \text{for } j \le m \\ (1, X_{j-1}^2, \ldots, X_{1-p}^2, \tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-q}^2)' + \sum_{l=1}^{q} \hat{\beta}_l d_{l-j}, & \text{for } j > m \end{cases}$$
$$\tilde{A}_j(\hat{\theta}) = d_j/\tilde{\sigma}_j^2,$$
$$\lambda_j = (\tilde{\varepsilon}_j^2 - 1)\tilde{A}_j(\hat{\theta}), \quad 1 \le j \le n.$$

(5) Finally, take $\widehat{\mu_A(\theta)} = (1/n)\sum_{j=1}^{n}\tilde{A}_j(\hat{\theta})$ and $\hat{L}_j = \lambda_j\hat{J}^{-1}$, with $\hat{J} = (1/n)\sum_{j=1}^{n}\lambda_j\lambda_j'$.

7.3. On the estimation of γ

As observed in Section 6, in the case of composite null hypothesis and with the aim of estimating the unknown parameter γ, it seems reasonable to treat the residuals as if they were the true errors and then apply some method designed for IID data. This subsection deals the case where the method considered is maximum likelihood, that is, γ is estimated by

$$\hat{\gamma} = \arg\max \sum_{j=1}^{n} \ell(\tilde{\varepsilon}_j; \gamma)$$

with $\ell(x;\gamma) = \log f(x;\gamma)$. The next result gives conditions for the consistency, that is, for Assumption (A.2), as well as for an asymptotic expansion which meets Assumption (A.5), implying its asymptotic normality.

Theorem 7.2: (a) *Suppose that* $\ell(x;\gamma)$ *is continuous as a function of* $\gamma \in \Gamma$, $\forall x \in \mathbb{R}$, Γ *is compact,* $E\{\sup_{\gamma\in\Gamma} |\ell(\varepsilon;\gamma)|\} < \infty$, $E\{\ell(\varepsilon;\gamma)\}$ *has a unique minimum at* γ_0 *and* $(\partial/\partial x)\ell(x;\gamma)$ *is bounded* $\forall x \in \mathbb{R}$, $\forall \gamma \in \Gamma$. *Then* $\hat{\gamma}$ *satisfies Assumption* (A.2).

(b) If, in addition, all second-order derivatives of $(\partial/\partial\gamma)\ell(x;\gamma)$ *exist and are bounded* $\forall x \in \mathbb{R}, \forall\gamma \in \Gamma_0 \subseteq \Gamma$, *where* Γ_0 *is an open neighbourhood of* γ_0, $C(\gamma_0) = E\{(\partial^2/\partial\gamma\partial\gamma')\ell(\varepsilon;\gamma_0)\}$ *exists and is nonsingular and* $E\{\varepsilon(\partial^2/\partial x\partial\gamma)\ell(\varepsilon;\gamma_0)\}$ *and* $E\{\|(\partial/\partial\gamma)\ell(\varepsilon;\gamma_0)\|^2\}$ *exist, then* $\hat{\gamma}$ *satisfies Assumption* (A.5) *with* $l(\varepsilon;\gamma_0) = -C(\gamma_0)^{-1}(\partial/\partial\gamma)\ell(\varepsilon;\gamma_0)$ *and* $M(\theta,\gamma_0) = 0.5C(\gamma_0)^{-1}E\{\varepsilon(\partial^2/\partial x\partial\gamma)\ell(\varepsilon;\gamma_0)\}\mu_A(\theta)'$.

Once it has been shown that the maximum likelihood estimator based on the residuals of γ satisfies the required assumptions, we next must find estimators for $M(\theta,\gamma_0)$ and $l(\varepsilon;\gamma_0)$ fulfilling Equations (12) and (13), respectively. The next result deals with this issue.

Proposition 7.3: *Suppose that assumptions in Theorem* 7.2 *(a) hold,* $E\{\|(\partial/\partial\gamma)\ell(\varepsilon;\gamma_0)\|^2\}$ *exist,* $(\partial^2/\partial\gamma\partial\gamma')\ell(x;\gamma)$, $(\partial^2/\partial x\partial\gamma)\ell(x;\gamma)$ *and the derivatives of* $(\partial^2/\partial\gamma\partial\gamma')\ell(x;\gamma)$ *exist and are bounded* $\forall x \in \mathbb{R}$, $\forall\gamma \in \Gamma_0 \subseteq \Gamma$, *where* Γ_0 *is an open neighbourhood of* γ_0 *and* $C(\gamma) = E\{(\partial^2/\partial\gamma\partial\gamma')\ell(\varepsilon;\gamma)\}$ *exists and is nonsingular* $\forall\gamma \in \Gamma_0$. *Let*

$$\hat{l}(\tilde{\varepsilon}_j;\hat{\gamma}) = -\hat{C}^{-1}\frac{\partial}{\partial\gamma}\ell(\tilde{\varepsilon}_j;\hat{\gamma}), \quad \widehat{M(\theta,\gamma)} = 0.5\hat{C}^{-1}\frac{1}{n}\sum_{j=1}^{n}\tilde{\varepsilon}_j\frac{\partial^2}{\partial x\partial\gamma}\ell(\tilde{\varepsilon}_j;\hat{\gamma})\widehat{\mu_A(\theta)}',$$

with

$$\hat{C} = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial\gamma\partial\gamma'}\ell(\tilde{\varepsilon}_j;\hat{\gamma}),$$

then $\widehat{M(\theta,\gamma)}$ *and* $\hat{l}(\tilde{\varepsilon}_j;\hat{\gamma})$ *satisfy Equations* (12) *and* (13), *respectively.*

7.4. On the calculation of the WB approximation

In practice, to calculate the WB approximation to the null distribution of $R_{n,\nu}$ (similarly for $T_{n,\nu}$ or $T_{n,\nu}(\hat{\gamma})$) proceed as follows:

(1) Estimate θ through $\hat{\theta}$.
(2) Compute the observed value of the test statistic, $R_{n,\nu,obs}$.
(3) Calculate $m_{jk} = \int\{\hat{C}_j(t) + \hat{S}_j(t)\}\{\hat{C}_k(t) + \hat{S}_k(t)\}w(t)\,dt$, $\nu + 1 \le j \le k \le n$. Note that $m_{jk} = m_{kj}$.
(4) For some large integer B, repeat the following steps for every $b \in \{1,\ldots,B\}$: Generate $n-\nu$ IID variables $\xi_{\nu+1},\ldots,\xi_n$ with mean 0 and variance 1. Calculate $R^{*b}_{2,n,\nu} = (1/(n-\nu))\sum_{\nu+1\le j,k\le n} m_{jk}\xi_j\xi_k$ or, as noted in Remark 4.7, $R^{*b}_{2,n,\nu} = (1/(n-\nu))\sum_{\nu+1\le j,k\le n} m_{jk}(\xi_j - \bar{\xi})(\xi_k - \bar{\xi})$.
(5) Approximate the p-value by $\hat{p} = (1/B)\sum_{b=1}^{B} I\{R^{*b}_{2,n,\nu} > R_{n,\nu,obs}\}$.

One major advantage of the WB over the PB is that the former does not re-estimate the GARCH parameters and the residuals at each iteration. For the WB approximation, most of the work is done before starting simulations, at steps (1)–(3). Once the set $\{m_{jk},\ \nu+1 \le j \le k \le n\}$ is computed, the WB replicates $R^{*1}_{2,n,\nu},\ldots,R^{*B}_{2,n,\nu}$ are calculated very fast.

8. Numerical results

In order to study the finite sample performance of the proposed procedure and to compare it with the PB, we carry out three numerical experiments with simulated data. The first experiment deals with testing for a standard normal distribution for the innovations. In the second experiment, the null hypothesis states a t_5 distribution for the innovations and the test is based on the integral transformation. Finally, in the third experiment, the problem of testing a composite null hypothesis with the skew-normal distribution is considered. An application to a real data set is also provided. Explicit formulas for the test statistics and m_{jk} are given in Appendix A.3.

8.1. Simulation Experiment 1

As in [6,10], in this experiment we consider the problem of testing normality for the innovation distribution. Specifically, we consider the following GARCH(1,1) model

$$X_j = \sigma_j \varepsilon_j, \quad \sigma_j^2 = 0.1 + 0.3X_{j-1}^2 + 0.3\sigma_{j-1}^2.$$

We generate a sample of size $n = 400$. The parameter $\theta = (c, a_1, b_1)$ is estimated by the GMLE by using the function `garch` of the R package `tseries`. Then we calculate the residuals and the test statistic $R_{n,\nu}$ with weight function w the PDF of a standard normal law. The p-value of the observed value of the test statistic is estimated: (a) by means of the PB, following Algorithm 6 in [11] and considering $B = 200$ bootstrap samples, as in [6,10] (denoted in the tables as PB); (b) by means of the WB with $\xi_1, \dots, \xi_n$ IID standard normal variables and $B = 1000$, as in [10] (denoted in the tables as WB); (c) by means of the WB as in (b), but with centred multipliers $\xi_1 - \bar{\xi}, \dots, \xi_n - \bar{\xi}$ (denoted in the tables as WBC). This experiment is repeated 2,500 times for several innovation distributions, as indicated in Table 1. Table 2 summarizes the obtained results for $\nu = 10$. As for the level, all methods give satisfactory results; as for the power, in the tried cases we observe that PB is a bit less powerful than WB, which is a little bit less powerful than WBC.

Table 3 compares PB and WB in terms of the required CPU time. This table shows the CPU consumed in seconds to get one p-value for the testing problem studied in this experiment for $n = 400(200)1000$. The figures in the table clearly show the computational efficiency of the WB in comparison to the PB. The difference when using raw multipliers and centred multipliers for the WB is negligible.

Table 1. Innovation distributions considered in Experiment 1.

H_0	$\varepsilon_j \sim N(0,1)$,
H_1	$\varepsilon_j = u_j/\sqrt{6/4}, u_j \sim t_6$,
H_2	$\varepsilon_j = u_j/\sqrt{2}, u_j \sim$Laplace,
H_3	$\varepsilon_j \sim N(0,1), j \le n/2, \varepsilon_j = u_j/\sqrt{2}, u_j \sim$Laplace$, j > n/2$,
H_4	$\varepsilon_j \sim N(0,1), j \le n/2, \varepsilon_j = u_j/\sqrt{7/5}, u_j \sim t_7, j > n/2$,
H_5	$\varepsilon_j = \{\Phi^{-1}(u_j^{1/2}) - \frac{1}{\pi}\}/\sqrt{\pi/(\pi-1)}, u_j \sim U(0,1)$, Φ CDF of a $N(0,1)$.

Table 2. Empirical percentages of rejection, α denotes the nominal level.

	H_0			H_1			H_2		
α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
PB	1.36	4.56	9.24	16.68	42.96	61.68	76.76	93.80	97.04
WB	1.04	4.32	9.04	21.12	49.76	65.68	93.04	99.20	99.88
WBC	1.12	4.52	9.24	21.68	50.24	65.84	93.20	99.20	99.88
	H_3			H_4			H_5		
α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
PB	10.64	35.36	55.12	1.96	8.64	16.20	94.28	98.52	98.96
WB	14.48	41.12	59.28	1.92	9.48	17.80	99.60	99.92	99.96
WBC	14.80	41.16	59.60	2.24	9.84	18.12	99.64	99.92	99.96

Table 3. CPU time consumed for the calculation of one p-value (in seconds).

	PB	PB	WB	WBC
B	200	1000	1000	1000
$n = 400$	10.28	49.52	1.81	1.97
$n = 600$	20.64	96.33	3.85	3.89
$n = 800$	34.33	160.82	6.87	6.75
$n = 1000$	50.28	240.96	10.75	10.52

8.2. Simulation Experiment 2

As recognized in [7], one of the most frequently used distributions when modelling conditional volatility for financial variables as in GARCH is the t_g-distribution. The degrees of freedom are taken $g = 5$ because this value is considered to be appropriate for financial data and is widely used in empirical analysis. Motivated by this fact, in our second experiment we consider the problem of testing $H_0 : \varepsilon_j \sim t_5$. As argued in Section 5, for testing H_0 it is convenient to use the test statistic $T_{n,\nu}$ based on the transformed residuals. So we repeat the experiment in Section 8.1 but using $T_{n,\nu}$ instead of $R_{n,\nu}$. As asserted in Section 5, there are several weight functions providing easily computable expressions for $T_{n,\nu}$. In our experiment we consider the following ones: the PDF of standard normal distribution and

$$w(t) = \frac{1}{\pi}\frac{1-\cos(t)}{t^2}, \tag{14}$$

which is the choice for w recommended in Epps and Pulley [24] (see also Section 4 in [25]). Table 4 displays the distributions for the innovations considered in this second experiment. Tables 5 and 6 show a summary of the results obtained for both weight functions and $\nu = 10$. Looking at these tables we can see that the levels are quite close to the nominal values in all cases; as for the power, it is clear that the second weight function gives better results. For this weight function, the PB gives a little bit

Table 4. Innovation distributions considered in Experiment 2.

H_0	$\varepsilon_j = u_j/\sqrt{5/3}, u_j \sim t_5$,
H_1	$\varepsilon_j \sim N(0,1)$,
H_2	$\varepsilon_j = u_j/\sqrt{2}, u_j \sim$Laplace,
H_3	$\varepsilon_j = \{\Phi^{-1}(u_j^{1/2}) - \frac{1}{\pi}\}/\sqrt{\pi/(\pi-1)}, u_j \sim U(0,1)$, Φ CDF of a $N(0,1)$,
H_4	$\varepsilon_j \sim SN(0,1,0.7)$[a],
H_5	$\varepsilon_j \sim SN(0,1,0.8)$[b].

[a] The skew-normal distribution with mean 0, variance 1 and skewness $\gamma_1 = 0.7$.
[b] Similar with $\gamma_1 = 0.8$.

Table 5. Empirical percentages of rejection, α denotes the nominal level.

	H_0			H_1			H_2		
α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
PB	1.08	5.52	10.20	1.84	7.08	14.52	1.08	5.12	10.24
WB	1.04	4.84	9.08	1.16	4.64	11.32	1.36	5.76	10.76
WBC	1.04	4.88	9.24	1.16	4.72	11.52	1.32	5.52	10.56
	H_3			H_4			H_5		
α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
PB	99.84	99.96	100.00	24.00	43.44	55.80	32.68	53.88	65.60
WB	99.64	99.96	100.00	17.88	38.84	51.96	25.96	48.72	61.12
WBC	99.64	99.96	99.96	18.48	39.28	52.04	26.84	48.96	61.08

Note: The weight function is the PDF of a standard normal.

Table 6. Empirical percentages of rejection, α denotes the nominal level.

	H_0			H_1			H_2		
α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
PB	1.28	4.52	9.68	5.16	32.76	57.96	1.84	10.48	22.60
WB	1.20	5.56	10.24	4.48	28.28	56.12	3.76	15.64	28.28
WBC	1.20	5.52	10.40	4.52	29.08	56.52	3.88	16.04	28.60
	H_3			H_4			H_5		
α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
PB	99.96	100.00	100.00	47.12	82.60	93.40	60.40	91.00	97.60
WB	99.96	100.00	100.00	53.24	84.00	93.72	70.24	93.24	97.68
WBC	99.98	100.00	100.00	54.16	84.24	93.76	70.36	93.40	97.76

Note: The weight function is as defined in Equation (14).

better results than WB and WBC when the innovations are Gaussian, for the rest of the cases the opposite is observed.

8.3. Simulation Experiment 3

In our third simulation experiment we consider the problem of testing for a composite null hypothesis. Specifically, we considered the problem of testing GOF for a skew-normal distribution (see Azzalini [26]),

$$H_0 : \varepsilon_j \sim SN(0, 1, \gamma_1), \quad \text{for some } \gamma_1 \in [-0.9953, 0.9953],$$

parametrized with the centred parameters, that is, with mean 0, variance 1 and skewness γ_1, where $\gamma_1 = E[\{X - E(X)\}^3]/\text{var}(X)^{3/2}$ (see Azzalini and Capitanio [27], for the definition and relationships between the centred and direct parametrizations). This is an interesting hypothesis since the skew-normal family has the appealing property of strictly including the normal law, as well as a wide variety of skewed densities. This capability to accommodate asymmetry is useful for modelling financial data since asymmetry is one of the stylized features of this sort of data (see Rydberg [28]).

We repeated the experiment in the previous section but, taking into account that the test with normal weight is less powerful than the one with the weight defined in Equation (14), we only considered the latter. The parameter was estimated by its maximum likelihood estimator. $M(\theta, \gamma)$ and $l(\cdot; \gamma)$ were estimated by using the estimators in Proposition 7.3.

In this experiment we considered two instances of H_0: $\gamma_1 = 0.70$ (the corresponding direct parameters are location $= -1.18$, scale $= 1.54$ and shape $= 3.23$) and $\gamma_1 = 0.85$ (the corresponding direct parameters are location $= -1.25$, scale $= 1.61$ and shape $= 4.98$). Table 7 shows the innovation distributions considered in this third experiment. As in the previous experiments, the sample size is $n = 400$. Nevertheless, since the results for the WB are a bit oversized for $\gamma_1 = 0.85$, we also perform the experiment with $n = 700$. Table 8 displays the obtained results for significance level $\alpha = 0.05$. Looking at the results for H_0 we see that, although for $n = 400$ the empirical levels are

Table 7. Innovation distributions considered in Experiment 3.

H_0	$\varepsilon_j \sim SN(0, 1, \gamma_1)$,
H_1	$\varepsilon_j = (u_j - 6)/\sqrt{12}$, $u_j \sim \chi^2_6$,
H_2	$\varepsilon_j = (u_j - 3)/\sqrt{6}$, $u_j \sim \chi^2_3$,
H_3	$\varepsilon_j \sim st(0, 1, 0.7, 5)^a$,
H_4	$\varepsilon_j = u_j/\sqrt{2}$, $u_j \sim$ Laplace.

[a] The skew-t distribution in [29] with mean 0, variance 1, $\gamma_1 = 0.7$ and 5 degrees of freedom.

Table 8. Empirical percentages of rejection for $\alpha = 0.05$.

	H_0 with $\gamma_1 = 0.7$		H_0 with $\gamma_1 = 0.85$		H_1	
n	400	700	400	700	400	700
PB	5.20	5.36	5.36	5.04	7.80	8.24
WB	5.24	5.40	6.20	5.36	9.96	11.24
WBC	5.48	5.52	6.36	5.48	10.08	11.28
	H_2		H_3		H_4	
n	400	700	400	700	400	700
PB	38.56	74.19	31.35	64.64	91.00	99.62
WB	55.00	85.18	25.40	51.16	86.60	99.91
WBC	55.64	85.28	25.75	51.64	86.96	99.91

reasonably close to the theoretical values, as the sample size increases the closeness grows. As for the power, we see that, as in the previous experiment and as observed in Remark 4.4, no test is most powerful for all alternatives.

8.4. A real data set

As an example, we apply the proposed technique to the time series of log returns of the Spanish stock market index IBEX35 from January 1997 to December 2003. This index is a market capitalization weighted index comprising the 35 most liquid Spanish stocks traded in the Madrid Stock Exchange General Index. The series consists of the daily closing prices of the IBEX35 index from January 1997 to December 2003, with $n = 1746$ observations. Figure 1 displays the time series of log returns. Figure 2 displays the sample autocorrelation function of the log returns (left) and of the squared log returns (right). This figure shows that there is almost no significant autocorrelation in the log return series $\{X_t\}$, but such an autocorrelation does exist for in the squared series $\{X_t^2\}$, as it should happen in a GARCH model.

Next we fitted a GARCH(1,1) model to the log returns, obtaining the following estimates: $\hat{c} = 8.078 \times 10^{-6}$, $\hat{a}_1 = 1.104 \times 10^{-1}$ and $\hat{b}_1 = 8.666 \times 10^{-1}$. We first tested for normality of the innovations, $H_{0N} : \varepsilon_j \sim N(0, 1)$. Proceeding as in Section 8.1 with $B = 2000$, we got the p-values 0.0600 (centred multiplies) and 0.0595 (raw multipliers), indicating that H_{0N} might not be supported by the data. Since the histogram of the residuals (see Figure 3) reveals that the innovations were generated by an asymmetric distribution, proceeding as in Section 8.3 with $B = 2000$, we tested $H_{0SN} : \varepsilon_j \sim SN(0, 1, \gamma_1)$, for some $\gamma_1 \in [-0.9953, 0.9953]$, obtaining the p-values 0.5055 (centred multiplies) and

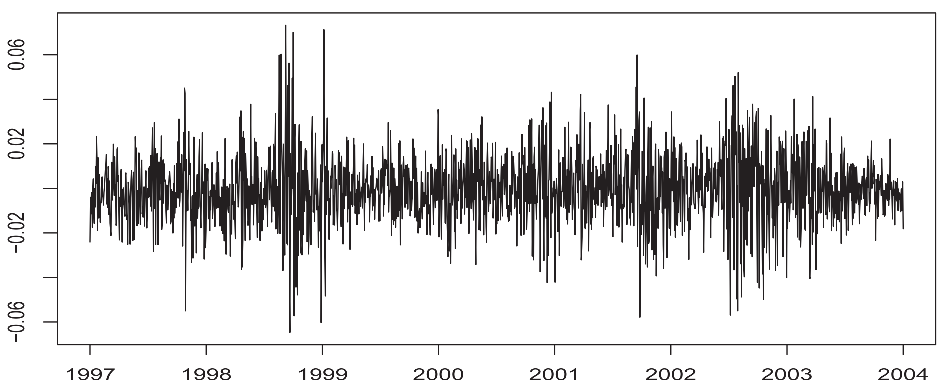

Figure 1. Time series of log returns of the IBEX35 index from January 1997 to December 2003.

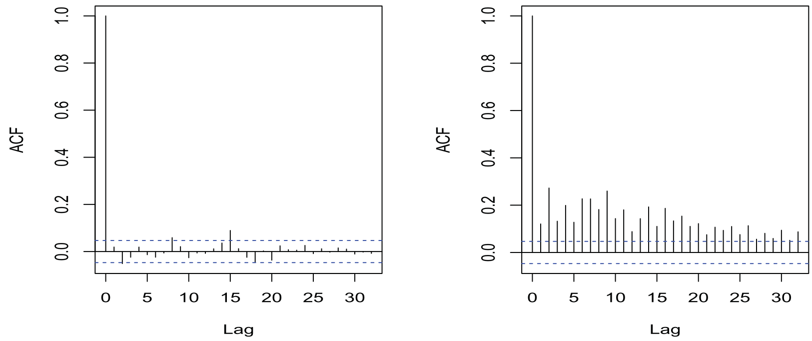

Figure 2. Sample autocorrelation function of the log returns (left) and of the squared log returns (right).

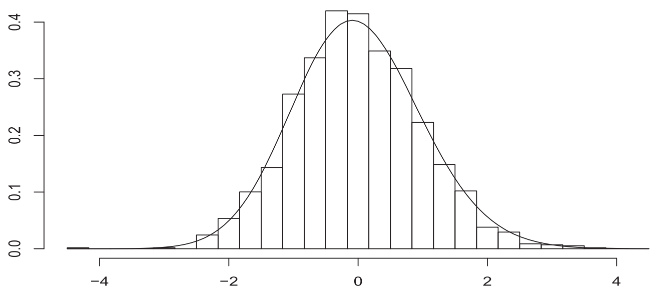

Figure 3. Histogram of the residuals and PDF of the fitted SN law.

0.5085 (raw multipliers) and thus H_{0SN} cannot be rejected. Figure 3 graphs the histogram of the residuals with the fitted skew-normal PDF superimposed ($\hat{\gamma}_1 = 1.733 \times 10^{-1}$). Looking at this figure we see that resulting PDF yields a quite reasonable fit.

9. Discussion and further research

In this piece of research GOF tests based on the CF for the innovations in a GARCH model are studied. Both the simple null hypothesis and the composite null hypothesis are considered. The asymptotic null distributions cannot be used to find critical values as they depend on unknowns. WB versions of the test statistics are analysed in detail in order to estimate their null distribution. The numerical experiments show a correct performance in practice. The main advantage of the WB approximation over the classical PB is the computational efficiency.

The research in this paper is limited to the univariate linear GARCH model. The studied methodology could be extended to testing GOF for the innovation distribution in other univariate GARCH models, such as log-GARCH models (see, e.g. [30]), or to multivariate GARCH models, such as CCC-GARCH models [31], or to other multiplicative error models, such as autoregressive conditional duration models (see, e.g. [32]).

Acknowledgments

The authors thank the anonymous referee for their constructive comments and suggestions which helped to improve the presentation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research in this manuscript has been financially supported by grant MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness

References

[1] Bollerslev T. Generalized, autoregressive conditional heteroskedasticity. J Econometrics. 1986;31:307–327.
[2] Hall P, Yao Q. Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica. 2003;71:285–317.
[3] Berkes I, Horváth L, Kokoszka P. GARCH processes: structure and estimation. Bernoulli. 2003;9:201–227.
[4] Francq C, Zakoïan JM. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli. 2004;10:605–637.
[5] Escanciano JC. Quasimaximum likelihood estimation of semi-strong GARCH models. Econometric Theory. 2009;25:561–570.
[6] Klar B, Lindner F, Meintanis SG. Specification tests for the error distribution in GARCH models. Comput Statist Data Anal. 2012;56:3587–3598.
[7] Bai J, Chen Z. Testing multivariate distributions in GARCH models. J Econometrics. 2008;143:19–36.
[8] Berkes I, Horváth L. The efficiency of the estimators of the parameters in GARCH processes. Ann Statist. 2004;32:633–655.
[9] Koul HL, Ling S. Fitting an error distribution in some heteroscedastic time series models. Ann Statist. 2006;34:994–1012.
[10] Ghoudi K, Rémillard B. Comparison of specification tests for GARCH models. Comput Statist Data Anal. 2014;76:291–300.
[11] Jiménez-Gamero MD. On the empirical characteristic function process of the residuals in GARCH models and applications. Test. 2014;23:409–432.
[12] Rémillard B. Specification tests for dynamic models using multipliers. Technical Report; 2012. SSRN Working Paper Series No. 2028558.
[13] Burke MD. Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Statist Probab Lett. 2000;46:13–20.
[14] Kojadinovic I, Yan J. Goodness-of-fit testing based on a weighted bootstrap: a fast large sample alternative to the parametric bootstrap. Canad J Statist. 2012;40:480–500.
[15] Jiménez-Gamero MD, Kim HM. Fast goodness-of-fit tests based on the characteristic function. Comput Statist Data Anal. 2015;89:172–191.
[16] Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 1982;50:987–1007.
[17] Bougerol P, Picard N. Strict stationarity of generalized autoregressive process. Ann Probab. 1992;52:115–127.
[18] Bougerol P, Picard N. Stationarity of GARCH processes and of some nonnegative time series. J Econometrics. 1992;20:1714–1730.
[19] Peng L, Yao Q. Least absolute deviations for ARCH and GARCH models. Biometrika. 2003;90:967–975.
[20] Delhing H, Mikosch T. Random quadratic forms and the bootstrap for U-statistics. J Multivariate Anal. 1994;51:392–413.
[21] Hušková M, Janssen P. Consistency of the generalized bootstrap for degenerate U-statistics. Ann Statist. 1993;21:1811–1823.
[22] Feller W. An introduction to probability theory and its applications, Vol. 2. New York: Wiley; 1971.
[23] Meintanis SG, Jiménez-Gamero MD, Alba-Fernández V. A class of goodness-of-fit tests based on transformation. Comm Statist Theory Methods. 2014;43:1708–1735.
[24] Epps TW, Pulley LB. A test for normality based on the empirical characteristic function. Biometrika. 1983;70:723–726.
[25] Jiménez-Gamero MD, Alba-Fernández V, Muñoz-García J, et al. Goodness-of-fit tests based on empirical characteristic functions. Comput Statist Data Anal. 2009;53:3957–3971.
[26] Azzalini A. A class of distributions which includes the normal ones. Scand J Statist. 1985;12:171–178.
[27] Azzalini A, Capitanio A. The skew-normal and related families. IMS monographs series, New York: Cambridge University Press; 2014.
[28] Rydberg TH. Realistic statistical modelling of financial data. Int Stat Rev. 2000;68:233–258.
[29] Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J R Stat Soc B. 2003;65:367–389.
[30] Francq C, Wintenberger O, Zakoïan JM. GARCH models without positivity constraints: exponential or log GARCH? J Econometrics. 2013;177:34–46.
[31] Francq C, Zakoïan JM. GARCH models: structure statistical, inference and applications. London: Wiley; 2010.
[32] Engle RF, Russell JR. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica. 1998;66:1127–1162.

[33] Kosorok MR. Introduction to empirical processes and semiparametric inference. New York: Springer; 2008.
[34] Billingsley P. Convergence of probability measures. New York: Wiley; 1968.
[35] Henze N, Hlávka Z, Meintanis SG. Testing for spherical symmetry via the empirical characteristic function. Statistics. 2014;48:1282–1296.
[36] Newey WK, McFadden D. Large sample estimation and hypothesis testing. In: Engle R, McFadden D, editors. Handbook of econometrics, vol. IV. Amsterdam: North-Holland; 1994. p. 2112–2245.

Appendix 1. Further technical and simulation results

A.1 A test based on transformation: technical results

This section studies some properties of the ECF of the transformed data, $\phi_{n-\nu}(t)$. Let $\phi(t) = E[\exp\{itF_0(\varepsilon)\}]$ and let $\{Z(t),\ t \in \mathbb{R}\}$ be a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_Z(s,t) = E[\{\mathcal{C}(t) + \mathcal{S}(t)\}\{\mathcal{C}(s) + \mathcal{S}(s)\}]$, $\forall t, s \in \mathbb{R}$, with

$$\mathcal{C}(t) = \cos\{tF_0(\varepsilon)\} - \operatorname{Re}\phi(t) - 0.5t\mu_R(t)\mu_A(\theta)'L_1(\theta),$$
$$\mathcal{S}(t) = \sin\{tF_0(\varepsilon)\} - \operatorname{Im}\phi(t) - 0.5t\mu_I(t)\mu_A(\theta)'L_1(\theta).$$

Theorem A.1: *Suppose that $\theta \in \Theta_0$, $\hat{\theta}$ satisfies* (A.1), *ν satisfies Equation* (8) *and F_0 is a continuous CDF with bounded PDF f_0. Then,*

(a) $\sup_{t\in S}|\phi_{n,\nu}(t) - \phi(t)| \xrightarrow{P} 0$, *for every compact interval S.*
(b) $\|\phi_{n,\nu} - \phi\|_w \xrightarrow{P} 0$.
(c) *If f_0 has a bounded derivative then $\{Z_{n,\nu}(t), t \in S\}$ converges weakly on $C(S)$ to $\{Z(t),\ t \in S\}$, in every compact interval S.*
(d) *If in addition w satisfies Equation* (9), *we also have that $\|Z_{n,\nu}\|_w^2 \xrightarrow{\mathcal{L}} \|Z\|_w^2$.*

A.2 Composite null hypothesis: technical results

This section studies some asymptotic properties of $T_{n,\nu}(\hat{\gamma})$. Let $\phi(t;\gamma) = E[\exp\{itF(\varepsilon;\gamma)\}]$ and let $\{Z(t;\gamma_0);,\ t \in \mathbb{R}\}$ be a zero-mean Gaussian process on $L_2(w)$ having covariance kernel $K_Z(s,t;\gamma_0) = E[\{\mathcal{C}(t;\gamma_0) + \mathcal{S}(t;\gamma_0)\}\{\mathcal{C}(s;\gamma_0) + \mathcal{S}(s;\gamma_0)\}]$, $\forall t, s \in \mathbb{R}$, with

$$\begin{aligned}\mathcal{C}(t;\gamma_0) &= \cos\{tF(\varepsilon;\gamma_0)\} - \operatorname{Re}\phi(t) - 0.5t\mu_{1R}(t;\gamma_0)\mu_A(\theta)'L_1(\theta)\\ &\quad + t\mu_{2R}(t;\gamma_0)'\{l(\varepsilon;\gamma_0) + M(\theta,\gamma_0)L_1(\theta)\},\\ \mathcal{S}(t;\gamma_0) &= \sin\{tF(\varepsilon;\gamma_0)\} - \operatorname{Im}\phi(t) - 0.5t\mu_{1I}(t;\gamma_0)\mu_A(\theta)'L_1(\theta)\\ &\quad + t\mu_{2I}(t;\gamma_0)'\{l(\varepsilon;\gamma_0) + M(\theta,\gamma_0)L_1(\theta)\}.\end{aligned}$$

Theorem A.2: *Suppose that $\theta \in \Theta_0$, $\hat{\theta}$ satisfies* (A.1), *ν satisfies Equation* (8), *the family $\mathcal{F}$ satisfies* (A.3) *and $\hat{\gamma}$ is an estimator of γ satisfying* (A.2). *Then,*

(a) $\sup_{t\in S}|\phi_{n,\nu}(t;\hat{\gamma}) - \phi(t;\gamma_0)| \xrightarrow{P} 0$, *for every compact interval S.*
(b) $\|\phi_{n,\nu} - \phi\|_w \xrightarrow{P} 0$.

If, in addition, (A.4) *holds and $\hat{\gamma}$ satisfies* (A.5), *then*

(c) *$\{Z_{n,\nu}(t;\hat{\gamma}), t \in S\}$ converges weakly on $C(S)$ to $\{Z(t;\gamma_0),\ t \in S\}$, in every compact interval S.*
(d) *If w satisfies Equation* (9), *we also have that $\|Z_{n,\nu}(\cdot;\hat{\gamma})\|_w^2 \xrightarrow{\mathcal{L}} \|Z(\cdot;\gamma_0)\|_w^2$.*

A.3 Expressions used in the numerical experiments

For Simulation Experiment 1 the test statistic has the following expression

$$R_{n,\nu} = \frac{1}{n-\nu}\sum_{j,k=\nu+1}^{n}\exp\{-0.5(\tilde{\varepsilon}_j - \tilde{\varepsilon}_k)^2\} - \sqrt{2}\sum_{j=\nu+1}^{n}\exp\{-0.25\tilde{\varepsilon}_j^2\} + \frac{n-\nu}{\sqrt{3}}$$

and

$$m_{jk} = \exp\{-0.5(\tilde{\varepsilon}_j - \tilde{\varepsilon}_k)^2\} - \frac{1}{\sqrt{2}}\exp\{-0.25\tilde{\varepsilon}_j^2\} - \frac{1}{\sqrt{2}}\exp\{-0.25\tilde{\varepsilon}_k^2\} + \frac{1}{\sqrt{3}}$$
$$- 0.5v_j \frac{1}{4\sqrt{2}}(\tilde{\varepsilon}_k^2 - 2)\exp\{-0.25\tilde{\varepsilon}_k^2\} - 0.5v_k \frac{1}{4\sqrt{2}}(\tilde{\varepsilon}_j^2 - 2)\exp\{-0.25\tilde{\varepsilon}_j^2\}$$
$$- 0.5(v_j + v_k)\frac{1}{3\sqrt{3}} + 0.25 v_j v_k \frac{1}{3\sqrt{3}},$$

where $v_j = \widehat{\mu_A(\theta)}' \widehat{L_j(\theta)}$, with $\widehat{L_j(\theta)}$ as in Proposition 7.1, $\nu + 1 \le j, k \le n$.

For Simulation Experiment 2 the expressions for the test statistic and m_{jk} become

$$T_{n,\nu} = \frac{1}{n-\nu}\sum_{j,k=\nu+1}^{n} \varphi_w\{F_0(\tilde{\varepsilon}_j) - F_0(\tilde{\varepsilon}_k)\} - 2\sum_{j=\nu+1}^{n} I_{1j} + (n-\nu)I_{00}$$

and

$$m_{jk} = \varphi_w\{F_0(\tilde{\varepsilon}_j) - F_0(\tilde{\varepsilon}_k)\} - I_{1j} - I_{1k} + I_{00} + 0.25v_jv_kI_{01} + 0.5(v_j+v_k)I_{02} + 0.5v_jI_{2k} + 0.5v_kI_{2j},$$

respectively, where v_j is as before, $\varphi_w(x) = \int \cos(tx)w(t)\,dt$,

$$I_{00} = \int_{-1}^{1} \varphi_w(x)\{1 - |x|\}\,dx, \quad I_{01} = -\int xyf_0^2(x)f_0^2(y)\varphi_w''\{F_0(x) - F_0(y)\}\,dx\,dy,$$
$$I_{02} = \int xf_0^2(x)[\varphi_w\{F_0(x)\} - \varphi_w\{1 - F_0(x)\}]\,dx,$$
$$I_{1j} = \int_{-F_0(\tilde{\varepsilon}_j)}^{1-F_0(\tilde{\varepsilon}_j)} \varphi_w(x)\,dx, \quad I_{2j} = \int xf_0^2(x)\varphi_w'\{F_0(\tilde{\varepsilon}_j) - F_0(x)\}\,dx,$$

$1 \le j \le n$, with $\varphi_w'(x) = (d/dx)\varphi_w(x)$, $\varphi_w''(x) = (d^2/dx^2)\varphi_w(x)$.

For Simulation Experiment 3 the expression of the test statistic is the same as that in the Simulation Experiment 2 with $f_0(x)$ and $F_0(x)$ replaced by $f(x;\hat{\gamma})$ and $F(x;\hat{\gamma})$, respectively. The expression of m_{jk} becomes

$$m_{jk} = \varphi_w\{F_0(\tilde{\varepsilon}_j) - F_0(\tilde{\varepsilon}_k)\} - I_{1j} - I_{1k} + I_{00} + 0.5(v_j+v_k)I_{02} + 0.5v_jI_{2k} + 0.5v_kI_{2j} + 2Y_jI_{3k} + 2Y_kI_{3j},$$

where v_j, I_{1j}, I_{2j}, I_{00} and I_{02} are as defined above for Simulation Experiment 2 with $f_0(x)$ and $F_0(x)$ replaced by $f(x;\hat{\gamma})$ and $F(x;\hat{\gamma})$, respectively, $Y_j = \hat{l}(\tilde{\varepsilon}_j;\hat{\gamma}) + \widehat{M(\theta,\gamma)}\widehat{L_j(\theta)}$ and

$$I_{3j} = -\int_{\tilde{\varepsilon}_j}^{\infty} f(x;\hat{\gamma})D_1F(x;\hat{\gamma})\,dx + \int f(x;\hat{\gamma})F(x;\hat{\gamma})D_1F(x;\hat{\gamma})\,dx.$$

Notice that $I_{01} = 0$ for the considered weight function.

Appendix 2. Proofs

Before proving the results in the previous sections we state a preliminary lemma. Some of the results in this lemma are known, but we prefer to include them to facilitate the reading of our proofs. Along this section K and ρ are generic constants taking many different values $K > 0$ and $0 < \rho < 1$. Let $\sigma_j^2(\theta)B_j(\theta)$ be the $r \times r$-matrix of second-order derivatives of $\sigma_j^2(\theta)$ with respect to θ, that is, $B_j(\theta) = (1/\sigma_j^2(\theta))(\partial^2/\partial\theta\partial\theta')\sigma_j^2(\theta)$. Let $\{\hat{\varepsilon}_j = X_j/\sigma_j(\hat{\theta}),\ 1 \le j \le n\}$ denote the non-truncated version of the residuals.

Lemma A.3: (a) *Let $k \in \mathbb{N}$. There exists $\Theta_k \subseteq \Theta_0$ such that $\theta \in \operatorname{int}\Theta_k$ and $E\{\sup_{u_1,u_2\in\Theta_k}(\sigma_0^{2k}(u_1)/\sigma_0^{2k}(u_2))\} < \infty$.*
(b) *$E(\sup_{u\in\Theta_0}|A_1(u)|^{\zeta}) < \infty$, for any $\zeta > 0$.*
(c) *$E(\sup_{u\in\Theta_0}|B_1(u)|^{\zeta}) < \infty$, for any $\zeta > 0$.*
(d) *$\sup_{u\in\Theta_0}|\sigma_j^2(u) - \tilde{\sigma}_j^2(u)| \le K\rho^j$.*
(e) *$\sup_{u\in\Theta_0}|(\partial/\partial\theta)\sigma_j^2(u) - (\partial/\partial\theta)\tilde{\sigma}_j^2(u)| \le K\rho^j$.*
(f) *If $\hat{\theta} \xrightarrow{a.s.(P)} \theta$ and $f:\mathbb{R}\to\mathbb{R}$ is such that $E\{|f(\varepsilon)|\} < \infty$, then*

$$\frac{1}{n}\sum_{j=1}^{n}|A_j(\theta) - \tilde{A}_j(\hat{\theta})|^k f(\varepsilon_j) \xrightarrow{a.s.(P)} 0, \quad \forall k \in \mathbb{N}.$$

(g) *If* $\hat{\theta} \xrightarrow{a.s.(P)} \theta$, *then* $\widehat{\mu_A(\theta)} \xrightarrow{a.s.(P)} \mu_A(\theta)$.

(h) *If* $\hat{\theta} \xrightarrow{a.s.(P)} \theta$, *then* $\sum_{j\geq 1} |\hat{\varepsilon}_j - \tilde{\varepsilon}_j|^k = O(1)$ *a.s. (in probability)*, $k = 1, 2$, *and* $(1/(n-\nu)) \sum_{j=\nu+1}^{n} |\hat{\varepsilon}_j - \varepsilon_j| \xrightarrow{a.s.(P)} 0$. *If* $\sqrt{n}(\hat{\theta} - \theta) = O_P(1)$, *then* $(1/\sqrt{n-\nu}) \sum_{j=\nu+1}^{n} (\hat{\varepsilon}_j - \varepsilon_j)^2 \xrightarrow{P} 0$.

Proof: (a) The proof closely follows the lines of the proof of (4.26) in [4], so we omit it. For (b) and (c), see (4.29) in [4]. For (d), see (4.6) in [4]. For (e), see (4.33) in [4].

(f) We have that

$$A_j(\theta) - \tilde{A}_j(\hat{\theta}) = \frac{1}{\sigma_j^2(\theta)} \frac{\partial}{\partial\theta} \sigma_j^2(\theta) \pm \frac{1}{\sigma_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) \pm \frac{1}{\tilde{\sigma}_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) - \frac{1}{\tilde{\sigma}_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \tilde{\sigma}_j^2(\hat{\theta}). \tag{A1}$$

From the mean value theorem we get

$$\frac{1}{\sigma_j^2(\theta)} \frac{\partial}{\partial\theta} \sigma_j^2(\theta) - \frac{1}{\sigma_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) = \{B_j(\hat{\theta}_j) - A_j(\hat{\theta}_j) A_j(\hat{\theta}_j)'\}(\hat{\theta} - \theta),$$

where $\hat{\theta}_j = \alpha_j \theta + (1 - \alpha_j)\hat{\theta}$, for some $\alpha_j \in (0, 1)$. Note that for n large enough, $\hat{\theta} \in \Theta_0$ a.s. (in probability). Thus,

$$\frac{1}{n} \sum_{j=1}^{n} \left| \frac{1}{\sigma_j^2(\theta)} \frac{\partial}{\partial\theta} \sigma_j^2(\theta) - \frac{1}{\sigma_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) \right|^k |f(\varepsilon_j)|$$
$$\leq \frac{1}{n} \sum_{j=1}^{n} \left\{ \sup_{u\in\Theta_0} |B_j(u)| + \sup_{u\in\Theta_0} |A_j(u)|^2 \right\}^k |f(\varepsilon_j)| |\hat{\theta} - \theta|^k.$$

From Lemma A.3 (b) and (c), the ergodic theorem and $\hat{\theta} \xrightarrow{a.s.(P)} \theta$, it follows that the right-hand side of the above expression converges a.s. (in probability) to 0. We also have that

$$\frac{1}{n} \sum_{j=1}^{n} \left| \frac{1}{\sigma_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) - \frac{1}{\tilde{\sigma}_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) \right|^k |f(\varepsilon_j)|$$
$$\leq K \frac{1}{n} \sum_{j=1}^{n} |f(\varepsilon_j)| \sup_{u\in\Theta_0} |A_j(u)|^k \sup_{u\in\Theta_0} |\sigma_j^2(u) - \tilde{\sigma}_j^2(u)|^k \leq \frac{1}{n} \sum_{j=1}^{n} |f(\varepsilon_j)| \sup_{u\in\Theta_0} |A_j(u)|^k \rho^j,$$

where the last inequality follows from Lemma A.3 (d). Since $E\{|f(\varepsilon_j)| \sup_{u\in\Theta_0} |A_j(u)|^k\} = E\{|f(\varepsilon_j)|\} E\{\sup_{u\in\Theta_0} |A_j(u)|^k\} < \infty$, by Lemma 2.2 in [3], the right-hand side of the above expression converges a.s. (in probability) to 0. Finally,

$$\frac{1}{n} \sum_{j=1}^{n} \left| \frac{1}{\tilde{\sigma}_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \sigma_j^2(\hat{\theta}) - \frac{1}{\tilde{\sigma}_j^2(\hat{\theta})} \frac{\partial}{\partial\theta} \tilde{\sigma}_j^2(\hat{\theta}) \right|^k |f(\varepsilon_j)|$$
$$\leq K \frac{1}{n} \sum_{j=1}^{n} \sup_{u\in\Theta_0} \left| \frac{\partial}{\partial\theta} \sigma_j^2(u) - \frac{\partial}{\partial\theta} \tilde{\sigma}_j^2(u) \right|^k |f(\varepsilon_j)| \leq K \frac{1}{n} \sum_{j=1}^{n} |f(\varepsilon_j)| \rho^j,$$

where the last inequality follows from Lemma A.3(e). Thus, by reasoning as before, the right-hand side of the above expression converges a.s. (in probability) to 0. All above facts imply the result.

(g) From the ergodic theorem,

$$\mu_A(\theta) - \frac{1}{n} \sum_{j=1}^{n} A_j(\theta) \xrightarrow{a.s.} 0. \tag{A2}$$

The result follows from Equation (A2) and the result in part (f).

(h) We have

$$\hat{\varepsilon}_j - \tilde{\varepsilon}_j = \frac{X_j}{\sigma_j(\hat{\theta}) \tilde{\sigma}_j(\hat{\theta}) \{\sigma_j(\hat{\theta}) + \tilde{\sigma}_j(\hat{\theta})\}} \{\tilde{\sigma}_j^2(\hat{\theta}) - \sigma_j^2(\hat{\theta})\}.$$

Thus, from the result in part (d),

$$|\hat{\varepsilon}_j - \tilde{\varepsilon}_j|^k \leq K |\varepsilon_j|^k \sup_{u_1, u_2 \in \Theta_1} \frac{\sigma_j^{2k}(u_1)}{\sigma_j^{2k}(u_2)} \rho^{jk}, \quad k = 1, 2,$$

a.s. (in probability). Since $E\{|\varepsilon_j|^k \sup_{u_1,u_2\in\Theta_1} \frac{\sigma_j^{2k}(u_1)}{\sigma_j^{2k}(u_2)}\} < \infty$, Lemma 2.2 in [3] implies that $\sum_{j\geq 1} |\hat{\varepsilon}_j - \tilde{\varepsilon}_j|^k = O(1)$ a.s. (in probability), $k = 1{,}2$. We also have that

$$\hat{\varepsilon}_j - \varepsilon_j = \frac{\varepsilon_j}{\sigma_j(\hat{\theta})\{\sigma_j(\hat{\theta}) + \sigma_j(\theta)\}}\{\sigma_j^2(\theta) - \sigma_j^2(\hat{\theta})\}.$$

Now, from the mean value theorem,

$$|\hat{\varepsilon}_j - \varepsilon_j| \leq K|\varepsilon_j| \sup_{u_1,u_2\in\Theta_1} \frac{\sigma_j^2(u_1)}{\sigma_j^2(u_2)} \sup_{u\in\Theta_0} |A_j(u)||\hat{\theta} - \theta|,$$

a.s. (in probability). Since $E\{|\varepsilon_j| \sup_{u_1,u_2\in\Theta_1}(\sigma_j^2(u_1)/\sigma_j^2(u_2)) \sup_{u\in\Theta_0} |A_j(u)|\} < \infty$, the ergodic theorem implies that $(1/(n-\nu))\sum_{j=\nu+1}^n |\hat{\varepsilon}_j - \varepsilon_j| \xrightarrow{a.s.(P)} 0$. If $\sqrt{n}(\hat{\theta} - \theta) = O_P(1)$, then again by the ergodic theorem we get that $(1/\sqrt{n-\nu})\sum_{j=\nu+1}^n (\hat{\varepsilon}_j - \varepsilon_j)^2 \xrightarrow{P} 0$. ■

Proof of Theorem 3.1**:** We have that $R^*_{0,n,\nu} = R^*_{1,n,\nu} + R^*_{2,n,\nu} + 2R^*_{3,n,\nu}$, with $R^*_{1,n,\nu} = \|Z_1^*\|_w^2$, $R^*_{2,n,\nu} = \|Z_2^*\|_w^2$, $R^{*2}_{3,n,\nu} \leq R^*_{1,n,\nu}R^*_{2,n,\nu}$, $Z_1^*(t) = (1/(n-\nu))\sum_{j=\nu+1}^n q(\varepsilon_j, t)\xi_j$ and

$$Z_2^*(t) = \frac{1}{n-\nu}\sum_{j=\nu+1}^n \{q(\tilde{\varepsilon}_j, t) - q(\varepsilon_j, t)\}\xi_j.$$

Let $\delta > 0$, by the Markov inequality and the mean value theorem,

$$P_*\{|Z_2^*(t)| > \delta\} \leq \frac{1}{\delta^2}t^2\frac{1}{n-\nu}\sum_{j=\nu+1}^n (\tilde{\varepsilon}_j - \varepsilon_j)^2 = t^2 o_P(1),$$

where the last equality follows from Lemma A.3(h). Therefore $W_2^* = o_{P_*}(1)$ (in probability). Now, by the conditional multiplier central limit theorem for IID Euclidean data (see, e.g. Lemma 10.5 in [33]), the finite dimensional distributions of the process $\{Z_1^*(t),\ t \in \mathbb{R}\}$, $(Z_1^*(t_1), \ldots, Z_1^*(t_r))'$, converge to a zero-mean normal law with variance-covariance matrix $(K_W(t_j, t_k))_{1\leq j,k\leq r}$ (a.s.). Let $s, t \in \mathbb{R}$,

$$E_*\{Z_1^*(t) - Z_1^*(s)\}^2 = \frac{1}{n-\nu}\sum_{j=\nu+1}^n \{q(\varepsilon_j, t) - q(\varepsilon_j, s)\}^2$$

$$\leq 4|t-s|^2\frac{1}{n-\nu}\sum_{j=\nu+1}^n (|\varepsilon_j| + 1)^2 \leq K|t-s|^2, \quad \text{a.s.}$$

Hence, from Theorem 12.3 in [34], conditional on $X_1, \ldots, X_n$, $\{Z_1^*(t),\ t \in S\}$ is tight (a.s.), for any compact interval $S \subset \mathbb{R}$, and therefore, conditional on $X_1, \ldots, X_n$, $Z_1^*(t)$ converges weakly to $W(t)$ in every compact interval (a.s.). Now, from the continuous mapping theorem, conditional on $X_1, \ldots, X_n$,

$$\int_{-R}^{R} Z_1^*(t)^2 w(t)\mathrm{d}t \xrightarrow{\mathcal{L}} \int_{-R}^{R} W(t)^2 w(t)\,\mathrm{d}t,$$

for every $R > 0$ (a.s.). Now, taking into account that

$$E_*\{Z_1^*(t)^2\} = \frac{1}{n-\nu}\sum_{j=\nu+1}^n q(\varepsilon_j, t)^2 \leq 16, \quad E\{W(t)^2\} = K_W(t,t) \leq 16,$$

proceeding as in the proof of Corollary 3.2 of [35], we get that, conditional on $X_1, \ldots, X_n$,

$$\int Z_1^*(t)^2 w(t)\,\mathrm{d}t \xrightarrow{\mathcal{L}} \int W(t)^2 w(t)\,\mathrm{d}t,$$

(a.s.) and thus the result follows. ■

Let $R^*_{1,n,\nu} = \|W^*_{1,n,\nu}\|_w^2$, where

$$W^*_{1,n,\nu} = \frac{1}{\sqrt{n-\nu}}\sum_{j=\nu+1}^n \{C_j(t) + S_j(t)\}\xi_j.$$

The following result gives the conditional asymptotic distribution, given $X_1, \ldots, X_n$, of $R^*_{1,n,\nu}$.

Lemma A.4: *Suppose that* $\theta \in \Theta_0$, $\hat{\theta}$ *satisfies* (A.1), *w satisfies Equation* (9) *and* ν *satisfies Equation* (8). *Then,*

$$\sup_x |P_*(R^*_{1,n,\nu} \leq x) - P(\|W_1\|^2_w \leq x)| \xrightarrow{P} 0,$$

where $\{W_1(t),\ t \in \mathbb{R}\}$ *is as defined in Theorem* 4.1.

Proof: From Lemma B.1.1 in [12], the finite dimensional distributions of the process $\{W^*_{1,n}(t),\ t \in \mathbb{R}\}$, $(W^*_{1,n}(t_1), \dots, W^*_{1,n}(t_r))'$, converge to a zero-mean normal law with variance-covariance matrix $(K_{W_1}(t_j, t_k))_{1\leq j,k\leq r}$ (a.s.). Let $s, t \in \mathbb{R}$,

$$E_*\{W^*_{1,n}(t) - W^*_{1,n}(s)\}^2 = \frac{1}{n-\nu}\sum_{j=\nu+1}^{n}\{C_j(t) + S_j(t) - C_j(s) - S_j(s)\}^2.$$

Taking into account that $|\cos(t\varepsilon_j) - \cos(s\varepsilon_j)| \leq |\varepsilon_j||t-s|$, $|\sin(t\varepsilon_j) - \sin(s\varepsilon_j)| \leq |\varepsilon_j||t-s|$, $|\mathrm{Re}\varphi_0(t) - \mathrm{Re}\,\varphi_0(s)| \leq K|t-s|$, $|\mathrm{Im}\,\varphi_0(t) - \mathrm{Im}\,\varphi_0(s)| \leq K|t-s|$, and that the continuity of the functions $t\mu_c(t)$, $t\mu_s(t)$ implies that $|t\mu_c(t) - s\mu_c(s)| \leq K|t-s|$, $|t\mu_s(t) - s\mu_s(s)| \leq K|t-s|$, it follows that

$$|C_j(t) + S_j(t) - C_j(s) - S_j(s)| \leq K|\varepsilon_j||t-s| + K|t-s| + kK|t-s||L_j(\theta)|.$$

Since $(1/n)\sum_{j=1}^n \varepsilon_j^2 \to 1$ (a.s.) and $(1/n)\sum_{j=1}^n |L_j(\theta)|^2$ also have a finite limit (a.s.), we conclude that

$$E_*\{W^*_{1,n}(t) - W^*_{1,n}(s)\}^2 \leq K|t-s|^2 \quad \text{a.s.}$$

Now the proof follows similar steps to those given in the one of Theorem 3.1, so we omit it. ■

***Proof of Theorem* 4.1:** We first show that replacing ε_j by $\tilde{\varepsilon}_j$ in the expression of $R^*_{1,n}$ has an asymptotically negligible effect. Let $W^*_{1,1,n}(t) = (1/\sqrt{n})\sum_{j=1}^n\{C_{1j}(t) + S_{1j}(t)\}\xi_j$ with $C_j - C_{1j} = \cos(t\varepsilon_j) - \cos(t\tilde{\varepsilon}_j)$, $S_j - S_{1j} = \sin(t\varepsilon_j) - \sin(t\tilde{\varepsilon}_j)$, $1 \leq j \leq n$. Let $\eta > 0$. Since

$$E_*\left(\left[\frac{1}{\sqrt{n}}\sum_{j=1}^n\{\cos(t\varepsilon_j) - \cos(t\tilde{\varepsilon}_j)\}\xi_j\right]^2\right)$$
$$= \frac{1}{n}\sum_{j=1}^n\{\cos(t\varepsilon_j) - \cos(t\tilde{\varepsilon}_j)\}^2 \leq t^2\frac{1}{n}\sum_{j=1}^n(\varepsilon_j - \tilde{\varepsilon}_j)^2 = t^2 o_P(1),$$

where the last equality comes from Lemma A.3 (h), and similarly

$$E_*\left(\left[\frac{1}{\sqrt{n}}\sum_{j=1}^n\left\{\sin(t\varepsilon_j) - \sin(t\tilde{\varepsilon}_j)\right\}\xi_j\right]^2\right) = t^2 o_P(1),$$

we conclude that

$$P_*(\|W^*_{1,n} - W^*_{1,1,n}\|^2_w > \eta) \to 0, \tag{A3}$$

in probability. From Equation (A3) and the result in Lemma A.4, it follows that

$$\|W^*_{1,n}\|^2_w = \|W^*_{1,1,n}\|^2_w + o_{P_*}(1), \tag{A4}$$

in probability. Now we show that replacing $L_j(\theta)$ by $\widehat{L_j(\theta)}$, $1 \leq j \leq n$, satisfying Equation (10) in the expression of $\|W^*_{1,1,n}\|^2_w$ asymptotically has no effect. Let $W^*_{1,2,n}(t) = (1/\sqrt{n})\sum_{j=1}^n\{C_{2j}(t) + S_{2j}(t)\}\xi_j$ with $C_{1j} - C_{2j} = 0.5t\mu_c(t)E\{A_0(\theta)\}'\{L_j(\theta) - \widehat{L_j(\theta)}\}$, $S_{1j} - S_{2j} = 0.5t\mu_s(t)E\{A_0(\theta)\}'\{L_j(\theta) - \widehat{L_j(\theta)}\}$, $1 \leq j \leq n$. Let $\mu(t)$ denote $\mu_c(t)$ or $\mu_s(t)$. Since

$$E_*\left(\left[\frac{1}{\sqrt{n}}\sum_{j=1}^n t\mu(t)E\{A_0(\theta)\}'\{L_j(\theta) - \widehat{L_j(\theta)}\}\xi_j\right]^2\right) \leq Kt^2\frac{1}{n}\sum_{j=1}^n|L_j(\theta) - \widehat{L_j(\theta)}|^2 = t^2 o_P(1),$$

by reasoning as before we get

$$\|W^*_{1,1,n}\|^2_w = \|W^*_{1,2,n}\|^2_w + o_{P_*}(1), \tag{A5}$$

in probability. Analogously, by Lemma A.3(g), we get

$$\|W^*_{1,2,n}\|^2_w = R^*_{2,n} + o_{P_*}(1), \tag{A6}$$

in probability. The result follows from Lemma A.4 and (A4)–(A6). ■

Proof of Corollary* 4.3*: The result follows from Theorem 4.1 and Theorem 1(b) in [11]. ■

Proof of Theorems* A.1 *and* A.2*: The proof of parts (a) and (b) closely follows the proof of Theorem 1 in [11]; the proof of parts (c) and (d) closely follows the proof of Theorem 2 in [11]. ■

Proof of Theorems* 5.1 *and* 6.1*: The proof closely follows the steps in the ones of Lemma A.4 and Theorem 4.1. ■

Proof of Proposition* 7.1*: We have that

$$(\varepsilon_j^2-1)A_j(\theta)-(\tilde{\varepsilon}_j^2-1)\tilde{A}_j(\hat{\theta})=(\varepsilon_j^2-1)\{A_j(\theta)-\tilde{A}_j(\hat{\theta})\}+(\varepsilon_j^2-\tilde{\varepsilon}_j^2)A_j(\theta)-(\varepsilon_j^2-\tilde{\varepsilon}_j^2)\{A_j(\theta)-\tilde{A}_j(\hat{\theta})\}.$$

From Lemma A.3 (f),

$$\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^2-1)|A_j(\theta)-\tilde{A}_j(\hat{\theta})|\xrightarrow{a.s.}0.$$

Since under the assumed conditions $\hat{\theta}\xrightarrow{\text{a.s.}}\theta$ (see [4]) and, following the proof of Lemma A.3(f),

$$|\hat{\varepsilon}_j^2-\tilde{\varepsilon}_j^2|\le K\varepsilon_j^2\sup_{u_1,u_2\in\Theta_2}\frac{\sigma_j^2(u_1)}{\sigma_j^2(u_2)}\rho^j,\quad |\hat{\varepsilon}_j^2-\varepsilon_j^2|\le K\varepsilon_j^2\sup_{u_1,u_2\in\Theta_2}\frac{\sigma_j^2(u_1)}{\sigma_j^2(u_2)}\sup_{u\in\Theta_0}|A_j(u)||\hat{\theta}-\theta|, \tag{A7}$$

a.s., it follows that

$$\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^2-\tilde{\varepsilon}_j^2)|A_j(\theta)|\xrightarrow{\text{a.s.}}0.$$

Taking into account (A1) and (A7), similar steps to those given in the proof of Lemma A.3(f) show that

$$\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^2-\tilde{\varepsilon}_j^2)|A_j(\theta)-\tilde{A}_j(\hat{\theta})|\xrightarrow{\text{a.s.}}0.$$

Thus, we have shown that

$$\left|\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^2-1)A_j(\theta)-\frac{1}{n}\sum_{j=1}^{n}(\tilde{\varepsilon}_j^2-1)\tilde{A}_j(\hat{\theta})\right|\xrightarrow{\text{a.s.}}0. \tag{A8}$$

To prove the result we must show that

$$\hat{J}\xrightarrow{\text{a.s.}}J. \tag{A9}$$

With this aim we first observe that, from the ergodic theorem,

$$J-\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^2-1)^2A_j(\theta)A_j(\theta)'\xrightarrow{\text{a.s.}}0.$$

Therefore, to prove Equation (A9) it suffices to see that

$$\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j^2-1)^2A_j(\theta)A_j(\theta)'-\frac{1}{n}\sum_{j=1}^{n}(\tilde{\varepsilon}_j^2-1)^2\tilde{A}_j(\hat{\theta})\tilde{A}_j(\hat{\theta})'\xrightarrow{\text{a.s.}}0.$$

The proof of the above convergence follows similar steps to those given to show Equation (A8). ■

Proof of Theorem* 7.2*: (a) From the mean value theorem and the assumptions made, we get

$$\frac{1}{n}\sum_{j=1}^{n}\ell(\tilde{\varepsilon}_j;\gamma)=\frac{1}{n}\sum_{j=1}^{n}\ell(\varepsilon_j;\gamma)+o_P(1).$$

Now the result follows from Lemma 2.4 and Theorem 2.1 in [36].

(b) The estimator $\hat{\gamma}$ satisfies

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\gamma}\ell(\tilde{\varepsilon}_j;\hat{\gamma})=0.$$

By applying a second-order Taylor expansion to the term on the left-hand side of the above equality and taking into account the assumptions made, the result in part (a) and Lemma A.3, we get

$$0=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\gamma}\ell(\varepsilon_j;\gamma_0)+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial^2}{\partial x\partial\gamma}\ell(\varepsilon_j;\gamma_0)(\tilde{\varepsilon}_j-\varepsilon_j)$$
$$+\{C(\gamma_0)+o_P(1)\}\sqrt{n}(\hat{\gamma}-\gamma_0)+o_P(1).$$

Proceeding as in the proof of Theorem 2 in [11] we get

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial^2}{\partial x\partial\gamma}\ell(\varepsilon_j;\gamma_0)(\tilde{\varepsilon}_j-\varepsilon_j)=-0.5E\left\{\varepsilon\frac{\partial^2}{\partial x\partial\gamma}\ell(\varepsilon;\gamma_0)\right\}\mu_A(\theta)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}L_j(\theta)+o_P(1)$$

and the result follows. ■

Proof of Proposition 7.3***:*** The result can be proven by Taylor expansion. ■