

On the empirical characteristic function process of the residuals in GARCH models and applications

M. Dolores Jiménez Gamero

Received: 4 June 2013 / Accepted: 29 January 2014 / Published online: 19 February 2014
© Sociedad de Estadística e Investigación Operativa 2014

Abstract The class of generalized autoregressive conditional heteroscedastic (GARCH) models has been proved to be particularly valuable in modeling financial data. This paper is devoted to study the empirical characteristic function process of the residuals. Specifically, it is shown that such process uniformly converges to the population characteristic function (CF) of the innovations in compact sets. The weak convergence of this empirical process, suitably normalized, is also studied. The limit depends on the population CF of the innovations, the equation defining the GARCH model and the parameter estimators employed to calculate the residuals. Applications of the obtained results for testing symmetry and goodness-of-fit to the law of the innovations are given.

Keywords GARCH · Residuals · Empirical characteristic function · Symmetry · Goodness-of-fit · Bootstrap

Mathematics Subject Classification 60F05 · 60E05 · 62M10 · 62F40

1 Introduction

Empirical processes play a prominent role in Statistics, since statistical procedures often involve functionals of them. In certain settings, such as linear models or time series, some variables of interest, such as the errors or innovations, cannot be directly observed and the inference is based on the residuals, whose calculation involves the estimation of certain parameters in the model. GARCH models, introduced by [Bollerslev \(1986\)](#), belong to this class. The present paper is concerned with the study of the

M. D. Jiménez Gamero (✉)
Dpto. de Estadística e Investigación Operativa, Universidad de Sevilla, Sevilla, Spain
e-mail: dolores@us.es

empirical characteristic function (ECF) of the residuals of these models. This study is motivated by the fact that the last decades have witnessed an increasing number of statistical procedures based on functions of the ECF process in a wide range of models and settings: for example, in point estimation (Feuerverger and McDunnough 1981a,b), the k -sample problem (Hušková and Meintanis 2008; Alba Fernández et al. 2008) and goodness-of-fit (GOF) tests (Epps and Pulley 1983; Baringhaus and Henze 1988; Gürtler and Henze 2000; Meintanis 2004; Matsui and Takemura 2005, 2008; Jiménez-Gamero et al. 2009, for independent and identically distributed (IID) observations, and Hušková and Meintanis 2007, 2010 and Jiménez-Gamero et al. 2005, for the errors in regression models, among many others). Much of the appeal of these procedures is that the application usually requires weaker conditions than their analogues based on the empirical distribution function (EDF). Another advantage of the statistical procedures based on the ECF over those based on the EDF is that while the data dimension plays an important role in the later ones (for instance, the Cramér von Mises test cannot be readily calculated for d -dimensional data, for any $d \geq 2$), it plays no role for many ECF-based tests, since the Cramér Wold device (see, for example, Serfling 1980, pp 17–18) is automatically applied. A key step towards the development of statistical procedures based on the ECF for making inferences on GARCH models is to study the ECF process of the residuals.

Some other processes associated with the residuals of GARCH models have been previously studied. For example, Berkes and Horváth (2001) have studied the empirical process of the observations; Berkes and Horváth (2003) have studied the empirical process of the squared residuals; the results in this last paper inspired those in Horváth et al. (2004), where some GOF tests based on the EDF of the squared residuals were numerically studied; Kulperger and Yu (2005) have studied partial sums of k th powers of residuals, with applications to change-point problems and GOF; Koul and Ling (2006) have studied the empirical process of the residuals with applications to testing GOF for the distribution of the innovations; Horváth et al. (2008) have studied partial sums of the squared observations and of its EDF.

This paper is devoted to study the limit behavior of the ECF process of the residuals. Specifically, we study the convergence in the class of continuous functions defined on a compact set, as well as the convergence in the Hilbert space $L_2(w) = \{f : \mathbb{R} \rightarrow \mathbb{C} : \|f\|_w^2 = \int |f(t)|^2 w(t) dt < \infty\}$, for some nonnegative function w satisfying $0 < \int w(t) dt < \infty$. We also study the convergence in law to a Gaussian process. The covariance structure of the limit process depends on the distribution of the innovations, the estimators employed to approximate the parameters of the GARCH model and the equation defining the model. Applications of the obtained results are reported. Specifically, we consider the problem of testing symmetry, which is equivalent to testing that the imaginary part of the population characteristic function (CF) of the innovations is equal to 0. Surprisingly, the limiting null distribution of the considered test statistic coincides with that derived for IID data, which only depends on the population CF. Another application to the problem of testing GOF for the distribution of the innovations is also given. In both applications, the null distribution of the test statistic is approximated by a bootstrap algorithm. The consistency of these bootstrap estimators is proven.

The paper is organized as follows. Section 2 describes the model and summarizes some properties that will be used along the paper. The main results concerning the asymptotic behavior of the ECF process of the residuals are studied in Sect. 3. Section 4 provides two applications of the obtained results to testing symmetry and GOF for the distribution of the innovations. All proofs, as well as some intermediate results, are sketched in the Appendix.

Before ending this section, we introduce some notation: all vectors are column vectors; for any vector v , v_k denotes its k th coordinate, $\|v\|$ its Euclidean norm and v' its transpose; for any complex number $x = a + ib$, $\bar{x} = a - ib$ and $|x| = \sqrt{a^2 + b^2} = \sqrt{x\bar{x}}$; for any complex function $f(x)$, $\text{Re}f(t)$ and $\text{Im}f(t)$ denote the real and the imaginary parts of f , respectively, that is to say, $f(x) = \text{Re}f(t) + i\text{Im}f(x)$; 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, X_2, \dots, X_n , respectively; all limits in this paper are taken when $n \rightarrow \infty$; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; \xrightarrow{P} denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; an unspecified integral denotes integration over the whole real line \mathbb{R} ; $\langle \cdot, \cdot \rangle$ denotes the scalar product in the Hilbert space $L_2(w)$; without loss of generality it will be assumed along the paper that $\int w(t)dt = 1$.

2 The model

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

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

for $-\infty < j < \infty$, where $c > 0$, $a_k \geq 0$ and $b_l \geq 0$. If $q = 0$ then we get an autoregressive conditional heteroscedastic (ARCH) model, introduced by Engle (1982). Throughout this paper, it will be assumed that $\{X_j, -\infty < j < \infty\}$ satisfies (1) and (2), that it is stationary, that $\{\varepsilon_j, -\infty < j < \infty\}$ are IID variables with $E(\varepsilon_j) = 0$ and $E(\varepsilon_j^2) = 1$, and that ε_j is independent of $\{X_{j-k}, k \geq 1\}$.

Bougerol and Picard (1992a,b) have given necessary and sufficient conditions for the existence of a unique strictly stationary solution of (1) and (2). A necessary and sufficient condition for the process $\{X_j, -\infty < j < \infty\}$ to be (strictly) stationary with $E(X_j^2) < \infty$ is (see, for example, Theorem 4.4 in Fan and Yao 2003)

$$\sum_{k=1}^p a_k + \sum_{l=1}^q b_l < 1.$$

In this case, $E(X_j) = 0$ and

$$E(X_j^2) = c \left(1 - \sum_{k=1}^p a_k - \sum_{l=1}^q b_l \right)^{-1}.$$

Let \mathcal{F}_j be the σ -algebra generated by $\{\varepsilon_k, -\infty < k \leq j\}$. Since $E(X_j^2 | \mathcal{F}_{j-1}) = \sigma_j^2$, the expectations of X_j^2 and σ_j^2 coincide. If $E(\log \sigma_0^2) < \infty$, then Theorem 2.1 in Berkes et al. (2003) shows that σ_j^2 can be expressed as (see also Hall and Yao 2003)

$$\begin{aligned} \sigma_j^2 = \sigma_j^2(\theta) &= \frac{c}{1 - \sum_l b_l} + \sum_{k=1}^p a_k X_{j-k}^2 \\ &+ \sum_{k=1}^p a_k \sum_{v=1}^{\infty} \sum_{l_1}^q \dots \sum_{l_v}^q b_{l_1} \dots b_{l_v} X_{j-k-l_1-\dots-l_v}^2, \end{aligned}$$

where $\theta = (c, a_1, \dots, a_p, b_1, \dots, b_q)'$ and the multiple sum vanishes if $q = 0$. From Lemma 2.3 in Berkes et al. (2003), a sufficient condition for $E(\log \sigma_0^2) < \infty$ to hold is that $E(|\varepsilon_0^2|^\delta) < \infty$, for some $\delta > 0$. Since we assume that $E(\varepsilon_0^2) = 1$, then the above expansion for σ_j^2 holds. Let $r = 1 + p + q$ denote the dimension of θ , which is assumed to be fixed but unknown.

As in Berkes and Horváth (2003), it will be also assumed that $\theta \in \Theta_0 = \Theta(\rho_0, \rho_1, \rho_2) = \{u = (\gamma, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) : \beta_1 + \dots + \beta_q \leq \rho_0, \rho_1 \leq \min\{\gamma, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q\} \leq \max\{\gamma, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q\} \leq \rho_2\}$, 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 .

To estimate θ , it is often assumed that the errors ε_j are normally distributed. This estimator, $\hat{\theta}$, is called the Gaussian maximum likelihood estimator (GMLE). If

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

then $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normally distributed, even if the errors are not normally distributed (see Hall and Yao 2003; Francq and Zakoian 2004). Moreover, even if (3) does not hold then, under certain conditions, $n^\kappa(\hat{\theta} - \theta)$ is bounded in probability, for some $\kappa > 0$ (see Hall and Yao 2003). Although the GMLE has become the most popular estimator, other estimators have been proposed. Examples are the estimators in Peng and Yao (2003), which are asymptotically normally distributed without requiring (3), and those in Berkes and Horváth (2004), where a class of estimators including the GMLE is studied. From now on, we will denote through $\hat{\theta}$ to any estimator of θ .

3 Main results

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, we must first estimate $\sigma_j^2(\theta)$. Note that $\sigma_j^2(\theta)$ depends on $\{X_k, -\infty < k \leq j - 1\}$, whereas

we observe X_1, \dots, X_n . So, in order to calculate the residuals, instead of $\sigma_j^2(\hat{\theta})$, we consider $\tilde{\sigma}_j^2(\hat{\theta})$, where

$$\begin{aligned} \tilde{\sigma}_j^2(\theta) &= \frac{c}{1 - \sum_l b_l} + \sum_{k=1}^{\min\{p, j-1\}} a_k X_{j-k}^2 \\ &+ \sum_{k=1}^p a_k \sum_{v=1}^{\infty} \sum_{l_1}^q \dots \sum_{l_v}^q b_{l_1} \dots b_{l_v} X_{j-k-l_1-\dots-l_v}^2 I(j-k-l_1-\dots-l_v) \end{aligned}$$

and $I(S)$ denotes the indicator function of the set S , which only depends on the observations X_1, \dots, X_{j-1} . Let $\{\tilde{\varepsilon}_j = X_j/\tilde{\sigma}_j(\hat{\theta}), 1 \leq j \leq n\}$ be the residuals and let $\varphi_{n,v}(t)$ denote the ECF of the residuals $\tilde{\varepsilon}_{v+1}, \dots, \tilde{\varepsilon}_n$

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

for some integer $v \geq 1$. The reason for only considering the residuals $\tilde{\varepsilon}_{v+1}, \dots, \tilde{\varepsilon}_n$, instead of all of them, $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$, 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 IID data, it is well known that the ECF of the data estimates consistently the population CF and that the ECF process converges to a complex Gaussian process in finite intervals (see, for example, [Feuerverger and Mureika 1977](#); [Csörgő 1981a,b](#); [Marcus 1981](#)). The next theorems state similar results for the ECF of the residuals. Let $\varphi(t)$ denote the CF of ε_0 .

Theorem 1 Assume that $\theta \in \Theta_0$ and $n^\kappa(\hat{\theta} - \theta) = O_P(1)$, for some $\kappa > 0$. Let $v = v(n)$ be an integer satisfying

$$v/n \rightarrow 0. \tag{4}$$

Then,

- (a) $\sup_{t \in S} |\varphi_{n,v}(t) - \varphi(t)| \xrightarrow{P} 0, \forall S$ compact interval.
- (b) $\|\varphi_{n,v} - \varphi\|_w \xrightarrow{P} 0$.

Next, we study the convergence in law of the ECF process $Y_{n,v}(t) = \sqrt{n-v} \{\varphi_{n,v}(t) - \varphi(t)\}$ and of its $L_2(w)$ -norm, $\|Y_{n,v}\|_w$. With this aim, we will assume that $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal. Specifically, we will assume 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 $L_j(\theta) = (g_1(\varepsilon_j)l_1(\varepsilon_{j-1}, \varepsilon_{j-2}, \dots), \dots, g_r(\varepsilon_j)l_r(\varepsilon_{j-1}, \varepsilon_{j-2}, \dots))'$, $1 \leq j \leq n$,

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

The GMLE as well as other often used estimators of θ satisfy (A.1) (see Sect. 3 of [Berkes and Horváth 2003](#)). If $\hat{\theta}$ satisfies (A.1) then, by the Martingale Central Limit Theorem (see, for example, [Kundu et al. 2000](#)), $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N_r(0, \Sigma_\theta)$, an r -variate zero mean normal law with variance matrix $\Sigma_\theta = \text{var}\{L_0(\theta)\} = (\varsigma_{uv})$, where

$$\varsigma_{uv} = E\{g_u(\varepsilon_0)g_v(\varepsilon_0)\}E\{l_u(\varepsilon_{-1}, \varepsilon_{-2}, \dots)l_v(\varepsilon_{-1}, \varepsilon_{-2}, \dots)\}, \quad 1 \leq u, v \leq r.$$

Let $\mu_c(t) = \frac{\partial}{\partial t} \text{Re}\varphi(t) = E\{-\varepsilon_0 \sin(t\varepsilon_0)\}$ and $\mu_s(t) = \frac{\partial}{\partial t} \text{Im}\varphi(t) = E\{\varepsilon_0 \cos(t\varepsilon_0)\}$. Observe that these derivatives exist because we assume that the innovations have finite first moment. Finally, let $\sigma_j^2(\theta)A_j(\theta)$ be the r -vector of derivatives of $\sigma_j^2(\theta)$ with respect to θ , that is, $A_j(\theta) = \frac{1}{\sigma_j^2(\theta)} \frac{\partial}{\partial \theta} \sigma_j^2(\theta)$.

Theorem 2 *Assume that $\theta \in \Theta_0$ and that $\hat{\theta}$ satisfies (A.1). Let $v = v(n)$ be an integer satisfying (4). Let $Y_{n,v}(t) = \sqrt{n-v}\{\varphi_{n,v}(t) - \varphi(t)\}$ and let $Y(t)$ be a zero mean complex valued Gaussian process with covariance structure*

$$\begin{aligned} \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\} &= \text{Cov}\{C(t), C(s)\}, \\ \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\} &= \text{Cov}\{C(t), S(s)\}, \\ \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\} &= \text{Cov}\{S(t), S(s)\}, \end{aligned}$$

$\forall t, s \in \mathbb{R}$, where $C(t) = \cos(t\varepsilon_0) - \text{Re}\varphi(t) - \frac{1}{2}t\mu_c(t)E\{A_0(\theta)\}'L_0(\theta)$, $S(t) = \sin(t\varepsilon_0) - \text{Im}\varphi(t) - \frac{1}{2}t\mu_s(t)E\{A_0(\theta)\}'L_0(\theta)$. Then,

- (a) $Y_{n,v}(t)$ converges weakly to $Y(t)$ in every compact interval.
- (b) If

$$\int t^4 w(t) dt < \infty, \tag{5}$$

then $\|Y_{n,v}\|_{2w}^2 \xrightarrow{\mathcal{L}} \|Y\|_{2w}^2$.

Remark 1 Let $m_c(t) = \text{Cov}\{\cos(t\varepsilon_0), L_0(\theta)\}$ and $m_s(t) = \text{Cov}\{\sin(t\varepsilon_0), L_0(\theta)\}$. Note that,

$$\begin{aligned} \text{Cov}\{C(t), C(s)\} &= \frac{1}{2}\{\text{Re}\varphi(t+s) + \text{Re}\varphi(t-s)\} - \text{Re}\varphi(t)\text{Re}\varphi(s) \\ &\quad - \frac{1}{2}t\mu_c(t)E\{A_0(\theta)\}'m_c(s) - \frac{1}{2}s\mu_c(s)E\{A_0(\theta)\}'m_c(t) \\ &\quad + \frac{1}{4}ts\mu_c(t)\mu_c(s)E\{A_0(\theta)\}'\Sigma_\theta E\{A_0(\theta)\}, \end{aligned}$$

$$\begin{aligned} \text{Cov}\{C(t), S(s)\} &= \frac{1}{2} \{ \text{Im}\varphi(t+s) + \text{Im}\varphi(t-s) \} - \text{Re}\varphi(t)\text{Im}\varphi(s) \\ &\quad - \frac{1}{2} t\mu_c(t)E\{A_0(\theta)\}'m_s(s) - \frac{1}{2} s\mu_s(s)E\{A_0(\theta)\}'m_c(t) \\ &\quad + \frac{1}{4} ts\mu_c(t)\mu_s(s)E\{A_0(\theta)\}'\Sigma_\theta E\{A_0(\theta)\}, \\ \text{Cov}\{S(t), S(s)\} &= \frac{1}{2} \{ -\text{Re}\varphi(t+s) + \text{Re}\varphi(t-s) \} - \text{Im}\varphi(t)\text{Im}\varphi(s) \\ &\quad - \frac{1}{2} t\mu_s(t)E\{A_0(\theta)\}'m_s(s) - \frac{1}{2} s\mu_s(s)E\{A_0(\theta)\}'m_s(t) \\ &\quad + \frac{1}{4} ts\mu_s(t)\mu_s(s)E\{A_0(\theta)\}'\Sigma_\theta E\{A_0(\theta)\}, \end{aligned}$$

$\forall t, s \in \mathbb{R}$. Therefore, in contrast to the IID case, the limit law of the ECF process depends not only on the CF of the innovations, but also on the estimator of θ employed, through $\Sigma_\theta, m_c(t)$ and $m_s(t)$, and on the equation defining the GARCH model through $E\{A_0(\theta)\}$.

4 Applications

4.1 Testing for symmetry

Many commonly used packages allow the practitioner to choose between several symmetric distributions for obtaining the (quasi) maximum likelihood estimator of the parameter θ , usually: normal (obtaining the GMLE), Laplace and Student t . The two later distributions let us model tails which are heavier than those of the normal law, a fact frequently observed in financial time series (see [Rydberg 2000](#)). Note that all of these distributions are symmetric, a hypothesis questioned by several authors, in the light of certain practical applications (see also [Rydberg 2000](#)). So, one could wish to test if the hypothesis of symmetry is supported by the data. This hypothesis is equivalent to the following

$$H_{0S} : \text{the law of the errors is symmetric} \iff H_{0S} : \text{Im}\varphi(t) = 0, \forall t.$$

As a consequence of Theorem 1, under the assumptions in this theorem, if w is a weight function such that $w(t) > 0, \forall t \in \mathbb{R}$, then

$$T_{n,v} = T_{n,v}(X_1, \dots, X_n) = \|\text{Im}\varphi_{n,v}\|_w^2 \xrightarrow{P} \|\text{Im}\varphi\|_w^2 \geq 0, \tag{6}$$

with $\|\text{Im}\varphi\|_w = 0$ if and only if H_{0S} is true. So a reasonable test for testing H_{0S} should reject the null hypothesis for “large” values of $T_{n,v}$. This statistic (with $v = 0$) was first proposed by [Feuerverger and Mureika \(1977\)](#) for testing symmetry in the IID case (see also [Henze et al. 2003](#)). Now, to determine what are large values of $T_{n,v}$, we should calculate the null distribution of $T_{n,v}$, or at least a consistent approximation to it. Clearly, the null distribution of $T_{n,v}$ is unknown. A classical way to approximate

the null distribution of a test statistic is through its asymptotic null distribution. As a consequence of Theorem 2, under H_{0S} ,

$$(n - \nu)T_{n,\nu} \xrightarrow{\mathcal{L}} W_{0S} = \|Y_{0S}\|_w^2, \tag{7}$$

where $Y_{0S}(t) = \text{Im}Y(t)$, $Y(t)$ being as defined in Theorem 2. From Theorem 2 and Remark 1, since under H_{0S} we have that $\text{Im}\varphi(t) = 0$ and $\mu_s(t) = \frac{\partial}{\partial t}\text{Im}\varphi(t) = 0, \forall t$, it follows that the Covariance structure of $Y_{0S}(t)$ is given by

$$K(s, t) = E\{Y_{0S}(t), Y_{0S}(s)\} = \frac{1}{2}\{-\varphi(t + s) + \varphi(t - s)\}. \tag{8}$$

Note that the asymptotic null distribution of $(n - \nu)T_{n,\nu}$ depends neither on the estimator of θ employed nor on the equation defining the GARCH model governing the data, but only on the population CF of the innovations. In fact, the asymptotic null distribution of $(n - \nu)T_{n,\nu}$ coincides with that obtained in Feuerverger and Mureika (1977) for the IID case. In other words, under H_{0S} , $\sqrt{n - \nu}\varphi_{n,\nu}(t)$ asymptotically behaves just like $\sqrt{n}\varphi_n(t)$ in the sense that both processes have the same weak limit, where $\varphi_n(t) = \frac{1}{n} \sum_{j=1}^n \varepsilon_j$. Let $0 < \alpha < 1$. The limit (7) tells us that $(n - \nu)T_{n,\nu} = O_P(1)$, and thus from (6), it follows that the test function for testing H_{0S}

$$\Psi_S = \Psi_S(X_1, X_2, \dots, X_n) = \begin{cases} 1, & \text{if } (n - \nu)T_{n,\nu} \geq t_\alpha, \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

where t_α is the $1 - \alpha$ percentile of the null distribution of $(n - \nu)T_{n,\nu}$, or a consistent approximation to it, is consistent against fixed alternatives, that is to say, it rejects H_{0S} with probability tending to one when it is false.

As observed before, the null distribution of $(n - \nu)T_{n,\nu}$ cannot be exactly calculated. The asymptotic null distribution of $(n - \nu)T_{n,\nu}$ cannot be used to approximate its null distribution, because it depends on the unknown CF of the innovations. Thus, we have to resort to other methods to approximate the null distribution of the test statistic.

The test (9) has been numerically investigated by Klar et al. (2012). To approximate the null distribution of the test statistic, these authors have employed the following bootstrap algorithm, which is quite similar to the bootstrap schemes employed in Hall and Yao (2003); Horváth et al. (2004) and Pascual et al. (2006).

Algorithm 1 (i) On the basis of X_1, \dots, X_n , compute $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = (\hat{c}, \hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q)'$.

(ii) Compute the residuals $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$.

(iii) Define the bootstrap observations

$$X_{n,j}^* = \tilde{\sigma}_j^*(\hat{\theta})\varepsilon_j^*,$$

where

$$\tilde{\sigma}_j^{*2}(\hat{\theta}) = \hat{c} + \sum_{k=1}^{\min\{p,j-1\}} \hat{a}_k X_{n,j-k}^{*2} + \sum_{l=1}^{\min\{q,j-1\}} \hat{b}_l \tilde{\sigma}_{j-l}^{*2}(\hat{\theta})$$

and $\varepsilon_j^* = v_j \tilde{\varepsilon}_j$, $j = 1, \dots, n$, v_1, \dots, v_n are IID with $P(v_j = -1) = P(v_j = 1) = 0.5$, and v_1, \dots, v_n are also independent of $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$.

- (iv) Based on the bootstrap data, $\mathbf{X}_n^* = (X_{n,1}^*, \dots, X_{n,n}^*)$ calculate the test statistic, obtaining $T_{n,v}^* = T_{n,v}(X_{n,1}^*, \dots, X_{n,n}^*)$. Approximate the null distribution of $(n - v)T_{n,v}^*$ through the conditional distribution of $(n - v)T_{n,v}^*$, given the data.

The above algorithm can be slightly modified by generating the bootstrap innovations from a symmetrization of the EDF of the residuals. We call Algorithm 2 to the resulting bootstrap algorithm.

Algorithm 2 Steps (i), (ii) and (iv) are as in Algorithm 1.

- (iii) The bootstrap observations are defined as in Algorithm 1, but now $\varepsilon_{v+1}^*, \dots, \varepsilon_n^*$ are IID from the EDF of $\tilde{\varepsilon}_{v+1}, \dots, \tilde{\varepsilon}_n, -\tilde{\varepsilon}_{v+1}, \dots, -\tilde{\varepsilon}_n$.

In practice, the bootstrap estimation of the null distribution of $(n - v)T_{n,v}$ in step (iv) in Algorithms 1 and 2 must be carried out by simulation, that is, by generating a high number of bootstrap samples, say $\mathbf{X}_n^{*1}, \dots, \mathbf{X}_n^{*B}$, and then approximating the null distribution of $(n - v)T_{n,v}$ through the EDF of $(n - v)T_{n,v}^{*1} = (n - v)T_{n,v}(\mathbf{X}_n^{*1}), \dots, (n - v)T_{n,v}^{*B} = (n - v)T_{n,v}(\mathbf{X}_n^{*B})$. This requires the calculation of $\hat{\theta}^{*1} = \hat{\theta}(\mathbf{X}_n^{*1}), \dots, \hat{\theta}^{*B} = \hat{\theta}(\mathbf{X}_n^{*B})$ as well as the bootstrap residuals, $\tilde{\varepsilon}_j^{*b}$, $v + 1 \leq j \leq n$, $1 \leq b \leq B$.

We can considerably save computing time by taking advantage of the property that, under H_{0S} , $\sqrt{n - v}\varphi_{n,v}(t)$ asymptotically behaves the same as $\sqrt{n}\varphi_n(t)$. With this aim, we treat the residuals $\tilde{\varepsilon}_{v+1}, \dots, \tilde{\varepsilon}_n$ as if they were the true errors $\varepsilon_{v+1}, \dots, \varepsilon_n$ and then act as in the IID setting (following for example the approach in Henze et al. 2003). This way we elude the calculation of $(n - v)T_{n,v}^{*1} = (n - v)T_{n,v}(\mathbf{X}_n^{*1}), \dots, (n - v)T_{n,v}^{*B} = (n - v)T_{n,v}(\mathbf{X}_n^{*B})$. Algorithms 3 and 4 give two bootstrap null distribution estimators that make use of this fact.

Algorithm 3 Steps (i) and (ii) are as in Algorithm 1.

- (iii) Let $\varepsilon_j^* = v_j \tilde{\varepsilon}_j$, $j = 1, \dots, n$, where v_1, \dots, v_n are IID with $P(v_j = -1) = P(v_j = 1) = 0.5$, and v_1, \dots, v_n are also independent of $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$.
- (iv) Approximate the null distribution of $(n - v)T_{n,v}$ through the conditional distribution of $(n - v)T_{n,v}^*$, given the data, where $T_{n,v}^* = \|\mathcal{S}_{n,v}^*\|_w^2$ and $\mathcal{S}_{n,v}^*(t) = \frac{1}{n-v} \sum_{j=v+1}^n \sin(t\varepsilon_j^*)$.

Algorithm 4 Steps (i), (ii) and (iv) are as in Algorithm 3.

- (iii) $\varepsilon_{v+1}^*, \dots, \varepsilon_n^*$ are IID from the EDF of $\tilde{\varepsilon}_{v+1}, \dots, \tilde{\varepsilon}_n, -\tilde{\varepsilon}_{v+1}, \dots, -\tilde{\varepsilon}_n$.

We next show the consistency of the distribution estimators yielded by Algorithms 3 and 4. Observe that no additional assumption is assumed to prove such consistency.

Before stating these results, we want to remark that to derive the convergence in (7), it is not necessary to assume that $\hat{\theta}$ satisfies (A.1), but only that $\sqrt{n}(\hat{\theta} - \theta) = O_P(1)$ (this is evident from the proof of Theorem 2).

Theorem 3 *Assume that $\theta \in \Theta_0$, that (5) holds and that $\sqrt{n}(\hat{\theta} - \theta) = O_P(1)$. Let $v = v(n)$ be an integer satisfying (4). If $T_{n,v}^*$ is as defined in Algorithm 3 or Algorithm 4, then*

$$\sup_x |P_*\{(n - v)T_{n,v}^* \leq x\} - P(W_{0S} \leq x)| \xrightarrow{P} 0,$$

where W_{0S} is as defined in (7).

Let $\{\lambda_j\}$ be the set of eigenvalues of operator \mathcal{A} defined on $L_2(w)$ by

$$\mathcal{A}v(y) = \int K(x, y)v(x)w(x)dx.$$

The random variate W_{0S} is distributed as a (infinite) sum of independent chi-squared variates with one degree of freedom, χ_1^2 , multiplied by the eigenvalues of \mathcal{A} , $\sum_j \lambda_j \chi_{1j}^2$. The set $\{\lambda_j\}$ is unknown because $K(t, s)$ is unknown. Nevertheless, $K(s, t)$ can be consistently estimated by

$$\begin{aligned} K_n(t, s) &= \frac{1}{n - v} \sum_{j=v+1}^n \sin(t\tilde{\varepsilon}_j) \sin(s\tilde{\varepsilon}_j) \\ &= \frac{1}{2(n - v)} \sum_{j=v+1}^n [\cos\{(t - s)\tilde{\varepsilon}_j\} - \cos\{(t + s)\tilde{\varepsilon}_j\}]. \end{aligned}$$

From Lemma 6 in Sect. 5, $K_n(t, s) \xrightarrow{P} K(t, s), \forall s, t \in \mathbb{R}$. Thus, we can approximate the distribution of W_{0S} , and thus the null distribution of $(n - v)T_{n,v}$, by means of

$$W_n = \sum_j \hat{\lambda}_j \chi_{1j}^2, \tag{10}$$

where $\{\hat{\lambda}_j\}$ are the eigenvalues of operator \mathcal{A}_n defined by $\mathcal{A}_n v(y) = \int K_n(x, y)v(x)w(x)dx$. Routine calculations show that $\{\hat{\lambda}_j\}$ are the eigenvalues of the $(n - v) \times (n - v)$ -matrix $M = (m_{jk})$ with

$$m_{jk} = \frac{1}{2(n - v)} \{I_w(\tilde{\varepsilon}_j - \tilde{\varepsilon}_k) - I_w(\tilde{\varepsilon}_j + \tilde{\varepsilon}_k)\},$$

where $I_w(t) = \int \cos(tx)w(x)dx$. Therefore, the set $\{\hat{\lambda}_j\}$ can be easily calculated using most statistical and mathematical programming languages. W_n is also a bootstrap estimator of the null distribution of $(n - v)T_{n,v}$. It is usually called a “bootstrap in

the limit” estimator, since it has been built by replacing all unknown quantities in the limit distribution of the test statistic by appropriate estimators. The next result shows that W_n estimates consistently the null distribution of $(n - \nu)T_{n,\nu}$.

Theorem 4 *Under assumptions in Theorem 3,*

$$\sup_x |P_{*}\{W_n \leq x\} - P(W_{0S} \leq x)| \xrightarrow{P} 0,$$

where W_n and W_{0S} are defined in (10) and (7), respectively.

We will call Algorithm 5 to the bootstrap approximation to the null distribution of $(n - \nu)T_{n,\nu}$ in Theorem 4.

Algorithm 5 Steps (i) and (ii) are as in Algorithm 1.

- (iii) Calculate the eigenvalues $\{\hat{\lambda}_j\}$ of matrix M
- (iv) Approximate the null distribution of $(n - \nu)T_{n,\nu}$ through the conditional distribution of $W_n = \sum_j \hat{\lambda}_j \chi_{1j}^2$, given the data.

Remark 2 Using the trigonometric identity $2 \sin(a) \sin(b) = \cos(a - b) - \cos(a + b)$, it can be easily derived the following alternative expression for $(n - \nu)T_{n,\nu}$,

$$(n - \nu)T_{n,\nu} = \sum_{j,k=\nu+1}^n m_{jk},$$

which is useful from a computational point of view.

Remark 3 In practice, the bootstrap distribution estimators in Algorithms 1–4 must be approximated by simulation. As for the null distribution estimator in Algorithms 5, since the distribution of a linear combination of χ^2 variates is unknown, the conditional distribution of W_n must be approximated either by simulation or by some numerical method (see, for example, Kotz et al. 1967; Castaño-Martínez and López-Blázquez 2005).

We have presented five bootstrap algorithms to estimate the null distribution of $(n - \nu)T_{n,\nu}$. To compare their finite sample performance, we carried out a small simulation experiment. We generated data from a GARCH(1,1) model with $c = 0.1$, $a_1 = 0.3$, $b_1 = 0.3$ and several symmetric distributions for the innovations, namely, normal, Laplace and t_5 . The sample size we took was $n = 400$ and $\nu = 10$. We took as weight function w in the definition of test statistic $T_{n,\nu}$ the density of the standard normal distribution. As in Klar et al. (2012), in order to approximate the bootstrap p value of the observed value of the test statistic, we generated $B = 200$ bootstrap samples for Algorithms 1–4. The conditional distribution of W_n was also approximated by simulation. This experiment was repeated 1,000 times. The parameters in the GARCH model were estimated through the GMLE. To calculate the parameter estimators as well as the residuals, we used the package `tseries` of the R language. Table 1 reports the number of bootstrap p values less than or equal to α , for $\alpha = 0.05, 0.10$, which are

Table 1 Estimated probabilities of type I errors and relative CPU

	Normal		Laplace		t_5		$CPU_i/CPU_1(\%)$
	0.05	0.10	0.05	0.10	0.05	0.10	
Algorithm 1	0.051	0.100	0.062	0.109	0.050	0.104	100.00
Algorithm 2	0.055	0.101	0.058	0.110	0.059	0.104	122.48
Algorithm 3	0.052	0.090	0.058	0.103	0.049	0.106	81.50
Algorithm 4	0.052	0.093	0.057	0.105	0.052	0.102	85.83
Algorithm 5	0.050	0.088	0.053	0.094	0.047	0.095	2.19

Table 2 Estimated powers for nominal level $\alpha = 0.05$

	Skewed normal		Skewed Laplace		Skewed t_5		
	γ :	Skewness:	1.3	1.9	1.1	1.2	1.1
Algorithm 1	0.538	0.987	0.371	0.877	0.226	0.556	
Algorithm 2	0.542	0.986	0.378	0.891	0.232	0.554	
Algorithm 3	0.541	0.987	0.372	0.860	0.220	0.542	
Algorithm 4	0.530	0.982	0.371	0.877	0.224	0.544	
Algorithm 5	0.523	0.985	0.362	0.867	0.213	0.535	

the estimated type I errors. Looking at this table, we see that the estimated type I errors are quite close to the nominal values in all cases. We also compared the algorithms in terms of the CPU consumed. Last column in Table 1 displays the obtained results. Algorithm 5 emerges as the cheapest in terms of computing time.

The power of the test Ψ_S , when the null distribution of the test statistic is estimated by means of Algorithm 1, has been numerically investigated by Klar et al. (2012). To study if the method of approximating the null distribution has any impact on the power for finite sample size, we repeated the above experiment with samples from skewed versions of the symmetric distributions in Table 1. Such skewed versions were obtained by applying the skewing mechanism proposed in Fernández and Steel (1998), namely, the density of the skewed distribution, indexed by a scalar $\gamma \in (0, \infty)$, is generated from the symmetric density f as follows

$$f_\gamma(t) = \frac{2}{\gamma + 1/\gamma} \{f(t/\gamma)I(t \geq 0) + f(\gamma t)I(t < 0)\}.$$

For $\gamma = 1$, we obtain the symmetric density f , for $\gamma > 1$ ($\gamma < 1$) f_γ is skewed to the right (left). Since $f_\gamma(t) = f_{1/\gamma}(-t)$, it is sufficient to consider values $\gamma > 1$. As in Klar et al. (2012), the values of γ were chosen so that the value of the skewness coefficient (that in our case coincides with the third moment because $E(\varepsilon_j) = 0$ and $E(\varepsilon_j^2) = 1$) has comparable values across the different distributions. Table 2 displays the obtained results for nominal level $\alpha = 0.05$. Looking at this table, we conclude that the method of estimating the null distribution of the test statistic has little effect on the power, since all estimated powers are quite close.

Summarizing, since the levels and the powers yielded by the five algorithms are very close, and Algorithm 5 is, from a computational point of view, the cheapest, we recommend its use.

4.2 Testing goodness-of-fit for the distribution of the innovations

To estimate the parameters of a GARCH model, it is usually assumed that the errors or innovations are normally distributed. Under certain not very restrictive conditions, the resultant estimator is normally distributed, even if the errors are not normally distributed (see [Hall and Yao 2003](#); [Berkes et al. 2003](#); [Francq and Zakoian 2004](#); [Escanciano 2009](#)). Nevertheless, as shown in [Berkes and Horváth \(2004\)](#) and numerically observed by [Huang et al. \(2008\)](#), the choice of the correct likelihood leads to more accurate estimates of the parameters. In addition, as argued in [Angelidis et al. \(2004\)](#) and [Koul and Ling \(2006\)](#), among many others, the knowledge of the error distribution plays an important role in evaluating the Value at Risk (VaR), a quantity very useful in economics and finance, whose calculation involves the distribution of the innovations. Hence, for certain purposes, a very important step in the analysis of GARCH models is to check if the data support the distributional hypothesis made on the innovations.

Some tests have been proposed for testing GOF for the innovations distribution. Since the innovations or errors are not observable, all these tests are necessarily based on the estimated errors or residuals. The proposed tests are “residual versions” for testing GOF for IID data. For example, [Horváth et al. \(2004\)](#) have numerically studied some GOF tests based on the EDF of the squared residuals for testing GOF for normality; [Kulperger and Yu \(2005\)](#) have proposed a Jarque–Bera type normality test; [Koul and Ling \(2006\)](#) and [Bai and Chen \(2008\)](#) have proposed a Kolmogorov–Smirnov type GOF test for testing a simple null hypothesis; [Horváth and Zytikis \(2006\)](#), [Mimoto \(2008\)](#) and [Koul and Mimoto \(2012\)](#) have proposed GOF tests for testing a simple null hypothesis, which are based on a kernel-type density estimator calculated from the residuals.

In a recent paper, [Klar et al. \(2012\)](#) have numerically studied a test based on the ECF of the residuals, comparing it with some of the tests cited above, for the problem of testing normality. From the obtained numerical results, they conclude that the test based on the ECF is one of the most powerful. The test statistic based on the ECF considered in [Klar et al. \(2012\)](#) is just $R_{n,v} = \|Y_{n,v}\|_w^2$, with $\varphi(t) = \varphi_0(t)$ the CF of the normal law and w the density of the standard normal distribution. Thus, the results in Sect. 3 provide a theoretical basis for this test. Specifically, for testing

$$H_{0G} : \text{the CF of } \varepsilon_0 \text{ is } \varphi_0(t),$$

for some $\varphi_0(t)$ totally specified, from Theorems 1(b) and 2(b) it follows that the test

$$\Psi_G = \Psi_G(X_1, X_2, \dots, X_n) = \begin{cases} 1, & \text{if } R_{n,v} \geq r_\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

where r_α is the $1 - \alpha$ percentile of the null distribution of $R_{n,v}$, or a consistent approximation to it, is consistent against fixed alternatives, that is to say, it rejects H_{0G} with probability tending to one when it is false, whenever $w(t) > 0, \forall t \in \mathbb{R}$. The null distribution of $R_{n,v}$ cannot be exactly calculated. The asymptotic null distribution of $R_{n,v}$ cannot be used to approximate its null distribution because it depends on unknowns (recall Remark 1). To approximate the null distribution of $R_{n,v}$, Klar et al. (2012) have employed the following bootstrap algorithm.

Algorithm 6 (i) On the basis of X_1, \dots, X_n , compute $\hat{\theta} = (\hat{c}, \hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q)'$.

(ii) Define the bootstrap data

$$X_{n,j}^* = \sigma_j^*(\hat{\theta})\varepsilon_j^*$$

where $\{\varepsilon_j^*, -\infty < j < \infty\}$ are IID with common CF $\varphi_0(t)$ and

$$\sigma_j^{*2}(\hat{\theta}) = \hat{c} + \sum_{k=1}^p \hat{a}_k X_{n,j-k}^{*2} + \sum_{l=1}^q \hat{b}_l \sigma_{j-l}^{*2}(\hat{\theta}), \quad j \in \mathbb{Z}.$$

(iii) Approximate the null distribution of $R_{n,v} = R_{n,v}(X_1, \dots, X_n)$ through the conditional distribution of $R_{n,v}^* = R_{n,v}(X_1^*, \dots, X_n^*)$, given X_1, \dots, X_n ,

To prove that the above bootstrap scheme provides a consistent null distribution estimator of $R_{n,v}$, we will assume that $\hat{\theta}^* = \hat{\theta}(X_1^*, \dots, X_n^*)$ satisfies the following assumption, which is equal to assumption (A.1) plus a Lindeberg condition to ensure that $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ is asymptotically normal, plus a continuity condition to ensure that when H_0 is true $\sqrt{n}(\hat{\theta} - \theta)$ and $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ both converge in law to the same limit.

(A.2) (a) $\hat{\theta}^*$ can be expressed as

$$\hat{\theta}^* = \hat{\theta} + n^{-1} \sum_{j=1}^n L_j(\hat{\theta}) + r^*,$$

with $r^* = o_{P_*}(n^{-1/2})$ in probability, that is to say, with probability tending to 1, and

$$L_j(\hat{\theta}) = (g_1(\varepsilon_j^*)l_1(\varepsilon_{j-1}^*, \varepsilon_{j-2}^*, \dots), \dots, g_r(\varepsilon_j^*)l_r(\varepsilon_{j-1}^*, \varepsilon_{j-2}^*, \dots))', \quad 1 \leq j \leq n.$$

(b) $E_*\{g_u(\varepsilon_0^*)\} = 0, E_*\{g_u(\varepsilon_0^*)^2\} < \infty, E_*\{l_u(\varepsilon_{-1}^*, \varepsilon_{-2}^*, \dots)^2\} < \infty, 1 \leq u \leq r$, in probability.

(c) For every $b \in \mathbb{R}^r, \frac{1}{n} \sum_{j=1}^n E_* \left[\{b' L_j(\hat{\theta})\}^2 \mid \mathcal{F}_{0j-1} \right] \xrightarrow{P_*} b' \Sigma_{0\theta} b$, with probability tending to 1, where \mathcal{F}_{0j} is the σ -algebra generated by $\{\varepsilon_k^*, -\infty < k \leq j\}$ and $\Sigma_{0\theta} = \text{Cov}_0\{L_0(\theta)\}$.

- (d) $\lim L_n(\epsilon, e_k) = 0$ for every $\epsilon > 0$ and every $1 \leq k \leq r$, in probability, where $\{e_1, \dots, e_r\}$ is any basis of \mathbb{R}^r and for $b \in \mathbb{R}^r$,

$$L_n(\epsilon, b) = \frac{1}{n} \sum_{j=1}^n E_* \left[\{b' L_j(\hat{\theta})\}^2 I\{|b' L_j(\hat{\theta})| > \epsilon\} \right].$$

- (e) For every $b \in \mathbb{R}^r$, $\frac{1}{n} \sum_{j=1}^n E_* \left\{ \cos(t \epsilon_j^*) b' L_j(\hat{\theta}) \mid \mathcal{F}_{0j-1} \right\} \xrightarrow{P_*} E_0\{\cos(t \epsilon_0) b' L_0(\theta)\}$ and $\frac{1}{n} \sum_{j=1}^n E_* \left\{ \sin(t \epsilon_j^*) b' L_j(\hat{\theta}) \mid \mathcal{F}_{0j-1} \right\} \xrightarrow{P_*} E_0\{\sin(t \epsilon_0) b' L_0(\theta)\}$ in probability, $\forall t \in \mathbb{R}$.

If $\hat{\theta}^*$ satisfies (A.2)(a)–(d) then, from Theorem 1.3 in Kundu et al. (2000), it follows that

$$\sup_x \left| P_*\{\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \leq x\} - P(Z \leq x) \right| \xrightarrow{P} 0,$$

where $Z \sim N_r(0, \Sigma_{0\theta})$. If in addition $\hat{\theta}$ satisfies (A.1) and H_0 is true, then $\sqrt{n}(\hat{\theta} - \theta)$ and $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ both converge in law to the same limit.

Let $\mu_{0c}(t) = \frac{\partial}{\partial t} \text{Re}\varphi_0(t)$ and $\mu_{0s}(t) = \frac{\partial}{\partial t} \text{Im}\varphi_0(t)$. The next result shows the consistency of the bootstrap approximation in Algorithm 6 as an estimator of the null distribution of the test statistic $R_{n,v}$.

Theorem 5 Assume that $\theta \in \Theta_0$, that $\sqrt{n}(\hat{\theta} - \theta) = O_P(1)$ and that $\hat{\theta}^*$ satisfies (A.2). Let $v = v(n)$ be an integer satisfying (4). Let $Y_0(t)$ be a zero mean complex valued Gaussian process with Covariance structure

$$\begin{aligned} \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\} &= \text{Cov}_0\{C_0(t), C_0(s)\} \\ \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\} &= \text{Cov}_0\{C_0(t), S_0(s)\} \\ \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\} &= \text{Cov}_0\{S_0(t), S_0(s)\} \end{aligned}$$

$\forall t, s \in \mathbb{R}$, where $C_0(t) = \cos(t \epsilon_0) - \text{Re}\varphi_0(t) - \frac{1}{2}t\mu_{0c}(t)E_0\{A_0(\theta)\}'L_0(\theta)$, $S_0(t) = \sin(t \epsilon_0) - \text{Im}\varphi_0(t) - \frac{1}{2}t\mu_{0s}(t)E_0\{A_0(\theta)\}'L_0(\theta)$. Let w be a non-negative function satisfying (5) and let $W_0 = \|Y_0\|_w^2$. Then

$$\sup_x \left| P_*(R_{n,v}^* \leq x) - P(W_0 \leq x) \right| \xrightarrow{P} 0.$$

The result in Theorem 5 holds whether or not H_0 is true. If H_0 is true and $\hat{\theta}$ satisfies (A.1), then the conditional distribution of $R_{n,v}^*$, given X_1, \dots, X_n , and the distribution of $R_{n,v}$ are close in the sense that both converge to the same limit.

Remark 4 The following alternative expression of $R_{n,v}$, which can be easily derived using elementary formulas for the sine and the cosine of a sum, is useful from a

computational point of view,

$$R_{n,v} = \frac{1}{n-v} \sum_{j=v}^n \sum_{k=v}^n h(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k),$$

where $h(x, y) = I_w(x - y) - I_{w0}(x) - I_{w0}(y) + I_{w00}$, with $I_{w0}(x) = \int I_w(x - y)dF_0(y)$ and $I_{w00} = \iint I_w(x - y)dF_0(x)dF_0(y)$, F_0 being the cumulative distribution function corresponding to φ_0 .

Acknowledgments The author thanks the anonymous referees for their constructive comments and suggestions which helped to improve the presentation. The author also acknowledges financial support from grant UJA2013/08/01.

5 Appendix: proofs

Before proving the results in Sects. 3 and 4, we state some preliminary results.

Lemma 1 *There exist $\rho_{01} \leq \rho_1$ and $\rho_{02} \leq \rho_2$, with $0 < \rho_{01} < \rho_{02}$, such that $\theta \in \Theta_1 = \Theta(\rho_0, \rho_{01}, \rho_{02})$ and*

$$E \left\{ \sup_{u \in \Theta_1} \frac{\sigma_0^{2\xi}(\theta)}{\sigma_0^{2\xi}(u)} \right\} < \infty, \quad \xi = 1, 2.$$

Proof Recall that we assume that $\{X_j, -\infty < j < \infty\}$ is stationary. From Proposition 1 in Francq and Zakoian (2004), this implies that there exists $s > 0$ such that $E(X_0^{2s}) < \infty$.

If $u_1, u_2 \in \Theta_1$, with $u_j = (\gamma_j, \alpha_{j1}, \dots, \alpha_{jp}, \beta_{j1}, \dots, \beta_{jq}), j = 1, 2$, then

$$\frac{\beta_{1k}}{\beta_{2k}} \leq \frac{\rho_{02}}{\rho_{01}} := 1 + \delta, \tag{11}$$

$1 \leq k \leq q$. From the proof of Theorem 2.2 in Francq and Zakoian (2004),

$$\sup_{u \in \Theta_1} \frac{\sigma_j^2(u)}{\sigma_j^2(\theta)} \leq D \left(1 + \sum_{k=1}^p \sum_{v=1}^{\infty} (1 + \delta)^v \rho^{v\zeta} X_{j-k-v}^{2\zeta} \right),$$

for some positive constant D and some $0 < \rho < 1, \forall \zeta \in (0, 1)$. Choosing ζ such that $E(X_0^{4\zeta}) < \infty$ and ρ_{01}, ρ_{02} such that $\delta = (1 - \rho^\zeta)/(2\rho^\zeta)$, we get the result for $\xi = 2$, which implies the result for $\xi = 1$. □

A detailed reading of the proof of Lemma 1 reveals that the result is still true if instead of keeping $\theta \in \Theta_0$ fixed, we consider a sequence of values for θ , say $\{\theta_m\}$, all of them in Θ_1 , since the bound in (11) is valid for all $u_1, u_2 \in \Theta_1$. Next lemma states this fact.

Lemma 2 *There exist $\rho_{01} \leq \rho_1$ and $\rho_{02} \leq \rho_2$, with $0 < \rho_{01} < \rho_{02}$, such that $\theta \in \Theta_1 = \Theta(\rho_0, \rho_{01}, \rho_{02})$ and for any $\{\theta_m\} \subset \Theta_1$*

$$\sup_m E \left\{ \sup_{u \in \Theta_1} \frac{\sigma_0^{2\xi}(\theta_m)}{\sigma_0^{2\xi}(u)} \right\} < \infty, \quad \xi = 1, 2.$$

Recall that $\sigma_j^2(\theta)A_j(\theta)$ is the r -vector of derivatives of $\sigma_j^2(\theta)$ with respect to θ . 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, $B_j(\theta) = \frac{1}{\sigma_j^2(\theta)} \frac{\partial^2}{\partial\theta\partial\theta'} \sigma_j^2(\theta)$. The following result is in the proof of Lemma 5.6 in [Berkes et al. \(2003\)](#).

Lemma 3 (a) $E \left(\sup_{u \in \Theta_0} |A_0(u)|^\zeta \right) < \infty$, for any $\zeta > 0$,
 (b) $E \left(\sup_{u \in \Theta_0} |B_0(u)|^\zeta \right) < \infty$, for any $\zeta > 0$,

where $|V|$ denotes the largest of the absolute values of the elements of the vector (or matrix) V .

Let $\{\hat{\varepsilon}_j = X_j/\sigma_j(\hat{\theta}), 1 \leq j \leq n\}$ be the non-truncated version of the residuals. The following result is Eq. (5.6) in [Koul and Mimoto \(2012\)](#). Here, we state it under weaker assumptions.

Lemma 4 *Assume that $\hat{\theta} \in \Theta_1$, where Θ_1 is as in Lemma 1. Then*

$$\sum_{j=1}^\infty |\tilde{\varepsilon}_j - \hat{\varepsilon}_j| = O(1), \quad a.s. \tag{12}$$

Proof To derive (12), [Koul and Mimoto \(2012\)](#) use Lemma 5.1 in [Berkes et al. \(2003\)](#), which proves that

$$E \left\{ \sup_{u \in \Theta_0} \frac{\sigma_0^2(\theta)}{\sigma_0^2(u)} \right\} < \infty, \tag{13}$$

provided that $E(|\varepsilon_0|^{2(1+\xi)}) < \infty$ for some $\xi > 0$. The Lemma also assumes that $\lim_{x \rightarrow 0} x^{-\mu} P(\varepsilon_0^2 \leq x) = 0$, for some $\mu > 0$. Nevertheless, [Francq and Zakoian \(2004\)](#) showed that (13) holds by only assuming that $\{X_j, -\infty < j < \infty\}$ is stationary if instead of taking the supremum in $u \in \Theta_0$, it is taken in an adequately chosen subset $\Theta_1 \subseteq \Theta_0$ (recall the proof of Lemma 1). □

Lemma 5 *Assume that $\theta \in \Theta_1$, where Θ_1 is as in Lemma 1, and that $n^\kappa(\hat{\theta} - \theta) = O_P(1)$, for some $\kappa > 0$. Let $v = v(n)$ be an integer satisfying (4). Then*

- (a) $\frac{1}{n-v} \sum_{j=v+1}^n |\hat{\varepsilon}_j - \varepsilon_j| = o_P(1)$.
- (b) If $\kappa > 0.25$, $\frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n (\hat{\varepsilon}_j - \varepsilon_j)^2 = o_P(1)$.

Proof Let $D > 0$ be a constant. By the Mean Value Theorem and Markov inequality,

$$P \left(\frac{1}{n-v} \sum_{j=v+1}^n |\hat{\varepsilon}_j - \varepsilon_j| > D \|\hat{\theta} - \theta\| \right) \leq \frac{1}{D} \frac{1}{n-v} \sum_{j=v+1}^n E \left(\frac{|X_j|}{\sigma_j(\tilde{\theta}_j)} \|A_j(\tilde{\theta}_j)\| \right),$$

with $\tilde{\theta}_j = \tau_j \hat{\theta} + (1 - \tau_j)\theta$, for some $0 < \tau_j < 1$. By Hölder inequality,

$$E \left(\frac{|X_j|}{\sigma_j(\tilde{\theta}_j)} \|A_j(\tilde{\theta}_j)\| \right) \leq E^{1/2} \left(\frac{X_j^2}{\sigma_j^2(\tilde{\theta}_j)} \right) E^{1/2} \left(\|A_j(\tilde{\theta}_j)\|^2 \right).$$

$$E \left(\frac{X_j^2}{\sigma_j^2(\tilde{\theta}_j)} \right) = E \left(\varepsilon_j^2 \frac{\sigma_j^2(\theta)}{\sigma_j^2(\tilde{\theta}_j)} \right) \leq E \left(\varepsilon_j^2 \sup_{u \in \Theta_1} \frac{\sigma_j^2(\theta)}{\sigma_j^2(u)} \right) = E \left(\sup_{u \in \Theta_1} \frac{\sigma_0^2(\theta)}{\sigma_0^2(u)} \right) < \infty,$$

where the last equality comes from the fact that $\sup_{u \in \Theta_1} \frac{\sigma_j^2(\theta)}{\sigma_j^2(u)}$ only depends on $\{X_k, k \leq j - 1\}$, the independence between ε_j and $\{X_k, k \leq j - 1\}$ and $E(\varepsilon_j^2) = 1$. On the other hand, if $\hat{\theta} \in \Theta_1$, from Lemma 3,

$$E \left(\|A_j(\tilde{\theta}_j)\|^2 \right) \leq E \left(\sup_{u \in \Theta_1} \|A_j(u)\|^2 \right) = E \left(\sup_{u \in \Theta_1} \|A_0(u)\|^2 \right) < \infty.$$

Thus, if $\hat{\theta} \in \Theta_1$, $\frac{1}{n-v} \sum_{j=v+1}^n |\hat{\varepsilon}_j - \varepsilon_j| = O_p(\|\hat{\theta} - \theta\|)$. Since we assume that $n^\kappa(\hat{\theta} - \theta) = O_p(1)$, for some $\kappa > 0$, which implies that $\|\hat{\theta} - \theta\| = o_p(1)$ and that $P(\hat{\theta} \in \Theta_1) \rightarrow 1$, it readily follows the result in (a).

Part (b) can be proven following similar steps to those given in the proof of part (a). □

Note that the results in Lemmas 1, 2, 4 (and thus also Lemma 5) depend of a subset $\Theta_1 \subset \Theta_0$ containing θ . To simplify notation, from now on we will assume that $\Theta_1 = \Theta_0$.

Proof of Theorem 1 (a) Let $\hat{\varepsilon}_j = X_j/\sigma_j(\hat{\theta})$, $1 \leq j \leq n$, and $c_{n,v}(t) = \frac{1}{n-v} \sum_{j=v+1}^n e^{it\varepsilon_j}$, which satisfies

$$\sup_{t \in S} |c_{n,v}(t) - \varphi(t)| \xrightarrow{P} 0, \tag{14}$$

for any fixed compact interval S . We have

$$|\varphi_{n,v}(t) - c_{n,v}(t)| \leq 2|t| \left\{ \frac{1}{n-v} \sum_{j=v+1}^n |\tilde{\varepsilon}_j - \hat{\varepsilon}_j| + \frac{1}{n-v} \sum_{j=v+1}^n |\hat{\varepsilon}_j - \varepsilon_j| \right\}. \tag{15}$$

The result follows from (14), (15) and Lemmas 4 and 5.

(b) Let $\delta > 0$. There exists a compact interval $S = S(\delta)$ such that $\int_S w(t)dt \geq 1 - \delta$. Since $|\varphi_{n,v}(t) - \varphi(t)|^2 \leq 4$, we have that $\|\varphi_{n,v} - \varphi\|_w^2 \leq 4\delta + \sup_{t \in S} |\varphi_{n,v}(t) - \varphi(t)|$ and, therefore, the result follows from part (a). \square

Proof of Theorem 2 (a) Let $S \subset \mathbb{R}$ be any compact set. We have

$$\left| \frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n e^{it\hat{\varepsilon}_j} - \frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n e^{it\tilde{\varepsilon}_j} \right| \leq 2|t| \frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n |\hat{\varepsilon}_j - \tilde{\varepsilon}_j| = |t|o_P(1), \tag{16}$$

where the last equality follows from Lemma 4. Since $\hat{\varepsilon}_j = \varepsilon_j + (\hat{\varepsilon}_j - \varepsilon_j)$, by Taylor expansion,

$$\frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n \{\cos(t\hat{\varepsilon}_j) - \operatorname{Re}\varphi(t)\} = V_n(t) + tW_n(t) + t^2Z_n(t),$$

where

$$V_n(t) = \frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n \{\cos(t\varepsilon_j) - \operatorname{Re}\varphi(t)\}, \quad W_n(t) = \frac{-1}{\sqrt{n-v}} \sum_{j=v+1}^n \sin(t\varepsilon_j)(\hat{\varepsilon}_j - \varepsilon_j),$$

and $|Z_n(t)| \leq Z_{n1}, \forall t \in \mathbb{R}$, with $Z_{n1} = (n-v)^{-1/2} \sum_{j=v+1}^n (\hat{\varepsilon}_j - \varepsilon_j)^2 = o_P(1)$, from Lemma 5. By Taylor expansion of $\hat{\theta}$ around θ , $W_n(t) = W_{1n}(t) + W_{2n}(t) + W_{3n}(t) + W_{4n}(t) + W_{5n}(t)$, with

$$\begin{aligned} W_{1n}(t) &= \frac{1}{2}\sqrt{n}(\hat{\theta} - \theta) \frac{1}{\sqrt{(n-v)n}} \sum_{j=v+1}^n \{\sin(t\varepsilon_j)\varepsilon_j + \mu_c(t)\} A_j(\theta), \\ W_{2n}(t) &= \frac{1}{2}\sqrt{n}(\hat{\theta} - \theta)\mu_c(t) \frac{1}{\sqrt{(n-v)n}} \sum_{j=v+1}^n [A_j(\theta) - E\{A_0(\theta)\}], \\ W_{3n}(t) &= -\frac{1}{8}\sqrt{n}(\hat{\theta} - \theta) \frac{1}{n\sqrt{n-v}} \sum_{j=v+1}^n \sin(t\varepsilon_j)\varepsilon_j A_j(\tilde{\theta}_j) A_j(\tilde{\theta}_j)' \frac{\sigma_j(\theta)}{\sigma_j(\tilde{\theta}_j)} \sqrt{n}(\hat{\theta} - \theta)', \\ W_{4n}(t) &= \frac{1}{4}\sqrt{n}(\hat{\theta} - \theta) \frac{1}{n\sqrt{n-v}} \sum_{j=v+1}^n \sin(t\varepsilon_j)\varepsilon_j B_j(\tilde{\theta}_j) \frac{\sigma_j(\theta)}{\sigma_j(\tilde{\theta}_j)} \sqrt{n}(\hat{\theta} - \theta)', \\ W_{5n} &= \frac{-1}{2}\sqrt{n}(\hat{\theta} - \theta) \left(\frac{n-v}{n}\right)^{1/2} \mu_c(t) E\{A_0(\theta)\}, \end{aligned}$$

where $\tilde{\theta}_j = \alpha_j \hat{\theta} + (1 - \alpha_j)\theta$, for some $\alpha_j \in (0, 1), \forall j$. Recall that \mathcal{F}_j denotes the σ -algebra generated by $\{\varepsilon_k, -\infty < k \leq j\}$. For each fixed $t \in \mathbb{R}$,

$\{\sin(t\varepsilon_j)\varepsilon_j A_j(\theta) - \mu_c(t)A_j(\theta)\}$ is a real-valued martingale difference with respect to $\{\mathcal{F}_j\}$. Thus, from the Central Limit Theorem for real-valued martingale differences

$$\frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n \{\sin(t\varepsilon_j)\varepsilon_j - \mu_c(t)\} A_j(\theta) = o_P(1).$$

Taking into account that $\sqrt{n}(\hat{\theta} - \theta) = o_P(1)$, we get that $W_{1n}(t) = o_P(1)$ for each fixed $t \in \mathbb{R}$. From the Ergodic Theorem and taking into account that $\sqrt{n}(\hat{\theta} - \theta) = o_P(1)$, we get that $W_{2n}(t) = o_P(1)$ for each fixed $t \in \mathbb{R}$. We also have $W_{3n}(t) = o_P(1)$ and $W_{4n}(t) = o_P(1)$ for each fixed $t \in \mathbb{R}$. Since $W_{1n}(t) + W_{2n}(t) + W_{3n}(t) + W_{4n}(t)$ is continuous as a function of t , we conclude that

$$\sup_{t \in S} |t \{W_{1n}(t) + W_{2n}(t) + W_{3n}(t) + W_{4n}(t)\}| = o_P(1).$$

From (A.1),

$$\sup_{t \in S} \left| t W_{5n}(t) + \frac{1}{2} t \mu_c(t) E\{A_0(\theta)\}' \frac{1}{\sqrt{n}} \sum_{j=1}^n L_j(\theta) \right| = o_P(1).$$

Because $v/n \rightarrow 0$, it is easy to see that $\sup_{t \in S} |V_n(t) - V_{n1}(t)| = o_P(1)$, where $V_{n1}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{\cos(t\varepsilon_j) - \text{Re}\varphi(t)\}$. Putting all above results together, we get

$$\sup_{t \in S} \left| \frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n \{\cos(t\varepsilon_j) - \text{Re}\varphi(t)\} - \frac{1}{\sqrt{n}} \sum_{j=1}^n C_j(t) \right| = o_P(1),$$

with $C_j(t) = \cos(t\varepsilon_j) - \text{Re}\varphi(t) - \frac{1}{2} t \mu_c(t) E\{A_0(\theta)\}' L_j(\theta)$, $1 \leq j \leq n$. Analogously,

$$\sup_{t \in S} \left| \frac{1}{\sqrt{n-v}} \sum_{j=v+1}^n \{\sin(t\varepsilon_j) - \text{Im}\varphi(t)\} - \frac{1}{\sqrt{n}} \sum_{j=1}^n S_j(t) \right| = o_P(1),$$

with $S_j(t) = \sin(t\varepsilon_j) - \text{Im}\varphi(t) - \frac{1}{2} t \mu_s(t) E\{A_0(\theta)\}' L_j(\theta)$, $1 \leq j \leq n$. Let $Y_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{C_j(t) + iS_j(t)\}$. This process satisfies: $E\{Y_n(t)\} = 0, \forall t \in \mathbb{R}$ and

$$\begin{aligned} \text{Cov}\{\text{Re}Y_n(t), \text{Re}Y_n(s)\} &= \text{Cov}\{C_1(t), C_1(s)\} = \text{Cov}\{\text{Re}Y(t), \text{Re}Y(s)\}, \\ \text{Cov}\{\text{Im}Y_n(t), \text{Im}Y_n(s)\} &= \text{Cov}\{S_1(t), S_1(s)\} = \text{Cov}\{\text{Im}Y(t), \text{Im}Y(s)\}, \\ \text{Cov}\{\text{Im}Y_n(t), \text{Re}Y_n(s)\} &= \text{Cov}\{S_1(t), C_1(s)\} = \text{Cov}\{\text{Im}Y(t), \text{Re}Y(s)\}, \quad \forall t, s \in \mathbb{R}. \end{aligned}$$

The Central Limit Theorem for real-valued martingale differences and the Cramér Wold device prove that for all finite collections s_1, \dots, s_v , the random vector $(Y_n(s_1), \dots, Y_n(s_v))$ converges in distribution to $(Y(s_1), \dots, Y(s_v))$. To finish the

proof, we must prove that the measures $\text{Re}Y_n(t)$ and $\text{Im}Y_n(t)$ are tight. Since $E(\varepsilon^2) = 1 < \infty$, $\frac{1}{\sqrt{n}} \sum_{j=1}^n \{\cos_j(t\varepsilon_j) - \text{Re}\varphi(t)\}$ is tight in every finite interval (see Csörgő 1981a). Because $E\{A_0(\theta)\}$ is finite, $\frac{1}{\sqrt{n}} \sum_{j=1}^n L_j(\theta)$ is bounded in probability and $t\mu_c(t)$ is a continuous function, we get that $t\mu_c(t)E\{A_0(\theta)\}' \frac{1}{\sqrt{n}} \sum_{j=1}^n L_j(\theta)$ is tight in every finite interval. Thus, $\text{Re}Y_n(y)$ is tight in every finite interval. Analogously, $\text{Im}Y_n(y)$ is tight in every finite interval. These facts imply the uniform convergence of $Y_{n,v}(t)$ to $Y(t)$ in finite intervals.

(b) From the proof of part (a): $|Y_{n,v}(t) - Y_n(t)| \leq a_{0n} + a_{1n}t + a_{2n}t^2$, with $a_{jn} = o_P(1)$, $j = 0, 1, 2$, which implies $\|Y_{n,v} - Y\|_w^2 = o_P(1)$. The result follows from the this equality by taking into account that $\|Y_{n,v}\|_w^2 = \|Y\|_w^2 + \|Y_{n,v} - Y\|_w^2 + 2C_n$, with $|C_n|^2 \leq \|Y\|_w^2 \|Y_{n,v} - Y\|_w^2 = o_P(1)$. □

To prove Theorem 3, we will use the following auxiliary result.

Lemma 6 *Suppose assumptions in Theorem 3 hold. Let*

$$K_n(t, s) = \frac{1}{n - v} \sum_{j=v+1}^n \sin(t\tilde{\varepsilon}_j) \sin(s\tilde{\varepsilon}_j).$$

Then

$$\sup_{(t,s) \in S} |K_n(t, s) - K(t, s)| \xrightarrow{P} 0,$$

$\forall S \subset \mathbb{R}^2$ compact set, where $K(t, s)$ is as defined in (8).

Proof The result is a consequence of Theorem 1 (a) since, by applying elementary trigonometric identities, we have the following alternative expression for $K_n(t, s)$

$$K_n(t, s) = \frac{1}{2(n - v)} \sum_{j=v+1}^n [\cos\{(t - s)\tilde{\varepsilon}_j\} - \cos\{(t + s)\tilde{\varepsilon}_j\}].$$

□

Proof of Theorem 3 To prove the result, we apply Theorem 1.1 in Kundu et al. (2000). So we will check that conditions (i)–(iii) in such Theorem hold. With this aim, we first note that $E_*\{\sin(t\varepsilon_j^*)\} = 0$, $v + 1 \leq j \leq n$, $\forall n$ and

$$\text{Cov}_*\{S_{n,v}^*(t), S_{n,v}^*(s)\} = \frac{1}{n - v} \sum_{j=v+1}^n \sin(t\tilde{\varepsilon}_j) \sin(s\tilde{\varepsilon}_j) = K_n(t, s),$$

for both algorithms. From Lemma 6, $K_n(t, s) \xrightarrow{P} K(t, s)$, $\forall s, t \in \mathbb{R}$. Note also that $|K(t, s)| \leq 1$, $\forall s, t \in \mathbb{R}$. Let $\{e_k, k \geq 0\}$ be an orthonormal basis of $L_2(w)$. Since

$K_n(t, s) \xrightarrow{P} K(t, s)$ and $|K(t, s)| \leq 1$, Dominated Convergence Theorem yields

$$\lim \langle R_n e_k, e_l \rangle = \lim \int K_n(t, s) e_k(t) e_l(s) w(t) w(s) dt ds = \langle R e_k, e_l \rangle,$$

in probability. This proves that condition (i) holds. To verify condition (ii), using Monotone Convergence Theorem, Parseval’s relation and Dominated Convergence Theorem, we get

$$\begin{aligned} \lim \sum_{k=0}^{\infty} \langle C_n e_k, e_k \rangle &= \lim \sum_{k=0}^{\infty} E_* \{ \langle R_n, e_k \rangle^2 \} \\ &= \lim E_* \{ \|R_n\|_w^2 \} = \int \lim K_n(t, t) w(t) dt \\ &= \int K(t, t) w(t) dt = E \|Y_{0S}\|_w < \infty, \end{aligned}$$

in probability. Let $W_{nj} = (n - v)^{-1/2} \sin(t\varepsilon_j^*)$. To prove condition (iii), we first notice that $|\langle W_{nj}, e_k \rangle| \leq 1/\sqrt{n - v}$, whence $E (\langle W_{nj}, e_k \rangle^2 I \{ |\langle W_{nj}, e_k \rangle| > \epsilon \}) = 0$ for sufficiently large n . Thus, $\lim \sum_{j=1}^n E (\langle W_{nj}, e_k \rangle^2 I \{ |\langle W_{nj}, e_k \rangle| > \epsilon \}) = 0$, for every $\epsilon > 0$ and every $k \geq 0$. \square

Proof of Theorem 4 Let $\epsilon > 0$ be arbitrary but fixed. Let $a \in L_2(w)$ such that $\|a\|_w = 1$. There exists $m = m(\epsilon) > 0$ such that $\int_C a^2(t) w(t) dt \geq 1 - \epsilon$, where $C = [-m, m]$. Let \bar{C} denote the complementary set of C . Let $y_n = \sup_{s, t \in C} |K_n(t, s) - K(s, t)|$. Taking into account that $\int |a(t)| w(t) dt \leq 1$, $\int_{\bar{C}} |a(t)| w(t) dt \leq \sqrt{\epsilon}$ and $|K_n(t, s) - K(s, t)| \leq 2$, we get

$$\int |K_n(t, s) - K(s, t)| a(t) a(s) w(t) w(s) ds dt \leq y_n + 2\epsilon.$$

The above inequality and Lemma 6 both imply that $\int |K_n(t, s) - K(s, t)| a(t) a(s) w(t) w(s) ds dt = o_P(1)$. By Corollary XI.9.4 (a) in Dunford and Schwartz (1963, p. 1090), this implies that $|\lambda_j - \hat{\lambda}_j| = o_P(1), \forall j$, where $\lambda_1 \geq \lambda_2 \geq \dots$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots$, and hence

$$\sup_x |P_*(W_n \leq x) - P(W_{0S} \leq x)| \xrightarrow{P} 0.$$

The above limit and (7) both imply the result. \square

Proof of Theorem 5 The proof closely follows that of Theorem 2, so we omit it. \square

References

Alba Fernández V, Jiménez-Gamero MD, Muñoz-García J (2008) A test for the two-sample problem based on empirical characteristic functions. *Comput Stat Data Anal* 52:3730–3748

- Angelidis T, Benos A, Degiannakis S (2004) The use of GARCH models in VaR estimation. *Stat Methodol* 1:105–128
- Bai J, Chen Z (2008) Testing multivariate distributions in GARCH models. *J Econom* 143:19–36
- Baringhaus L, Henze N (1988) A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* 35:339–348
- Berkes I, Horváth L (2001) Strong approximation of the empirical process of GARCH sequences. *Ann Appl Probab* 11:789–809
- Berkes I, Horváth L (2003) Limit results for the empirical process of squared residuals in GARCH models. *Stoch Process Appl* 105:271–298
- Berkes I, Horváth L (2004) The efficiency of the estimators of the parameters in GARCH processes. *Ann Stat* 32:633–655
- Berkes I, Horváth L, Kokoszka P (2003) GARCH processes: structure and estimation. *Bernoulli* 9:201–227
- Bollerslev T (1986) Generalized, autoregressive conditional heteroskedasticity. *J. Econom* 31:307–327
- Bougerol P, Picard N (1992a) Strict stationarity of generalized autoregressive process. *Ann Probab* 52:115–127
- Bougerol P, Picard N (1992b) Stationarity of GARCH processes and of some nonnegative time series. *J Econom* 52:115–127
- Castaño-Martínez A, López-Blázquez F (2005) Distribution of a sum of weighted central chi-square variables. *Commun Stat Theory Methods* 34:515–524
- Csörgő S (1981a) Limit behaviour of the empirical characteristic function. *Ann Probab* 9:130–144
- Csörgő S (1981b) Multivariate empirical characteristics functions. *Z Wahrsch Verw Gebiete* 55:203–229
- Dunford N, Schwartz JT (1963) Linear operators. Spectral theory. Self adjoint operators in Hilbert space. Part II. Wiley, NJ
- Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1007
- Epps TW, Pulley LB (1983) A test for normality based on the empirical characteristic function. *Biometrika* 70:723–726
- Escanciano JC (2009) Quasimaximum likelihood estimation of semi-strong GARCH models. *Econom Theory* 25:561–570
- Fan J, Yao Q (2003) Nonlinear time series. Springer, New York
- Fernández C, Steel M (1998) On Bayesian modeling of fat tails and skewness. *J Am Stat Assoc* 93:359–371
- Feuerverger A, McDunnough P (1981a) On some Fourier methods for inference. *J Am Stat Assoc* 76:379–387
- Feuerverger A, McDunnough P (1981b) On the efficiency of empirical characteristic function procedures. *J Roy Stat Soc Ser B* 43:20–27
- Feuerverger A, Mureika RA (1977) The empirical characteristic function and its applications. *Ann Stat* 5:88–97
- Françq C, Zakoïan JM (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10:605–637
- Gürtler N, Henze N (2000) Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann Inst Stat Math* 52:267–286
- Hall P, Yao Q (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71:285–317
- Henze N, Klar B, Meintanis SG (2003) Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *J Multivar Anal* 87:275–297
- Horváth L, Kokoszka P, Teyssièrè G (2004) Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *J Stat Comput Simul* 74:485–4965
- Horváth L, Zytikis R (2006) Testing goodness of fit based on densities of GARCH innovations. *Econom Theory* 22:457–482
- Horváth L, Kokoszka P, Zytikis R (2008) Distributional analysis of empirical volatility in Garch processes. *J Stat Plan Inference* 138:3578–3589
- Huang D, Wang H, Yao Q (2008) Estimating GARCH models: when to use what? *Econom J* 11:27–38
- Hušková M, Meintanis SG (2007) Omnibus tests for the error distribution in the linear regression model. *Statistics* 41:363–376
- Hušková M, Meintanis SG (2008) Tests for the multivariate k-sample problem based on the empirical characteristic function. *J Nonparametr Stat* 20:263–277

- Hušková M, Meintanis SG (2010) Tests for the error distribution in nonparametric possibly heteroscedastic regression models. *Test* 19:92–112
- Jiménez-Gamero MD, Alba-Fernández V, Muñoz-García J, Chalco-Cano Y (2009) Goodness-of-fit tests based on empirical characteristic functions. *Comput Stat Data Anal* 53:3957–3971
- Jiménez-Gamero MD, Muñoz-García J, Pino-Mejías R (2005) Testing goodness of fit for the distribution of errors in multivariate linear models. *J Multivar Anal* 95:301–322
- Klar B, Lindner F, Meintanis SG (2012) Specification tests for the error distribution in GARCH models. *Comput Stat Data Anal* 56:3587–3598
- Kotz S, Johnson NL, Boyd DW (1967) Series representations of quadratic forms in normal variables. I. Central case. *Ann Math Stat* 38:823–837
- Koul HL, Ling S (2006) Fitting an error distribution in some heteroscedastic time series models. *Ann Stat* 34:994–1012
- Koul HL, Mimoto N (2012) A goodness-of-fit test for GARCH innovation density. *Metrika* 75:127–149
- Kulperger R, Yu H (2005) High moment partial sum processes of residuals in GARCH models and their applications. *Ann Stat* 33:2395–2422
- Kundu S, Majumdar S, Mukherjee K (2000) Central limit theorems revisited. *Stat Probab Lett* 47:27–265
- Matsui M, Takemura A (2005) Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Ann Inst Stat Math* 57:183–199
- Matsui M, Takemura A (2008) Goodness-of-fit tests for symmetric stable distributions-Empirical characteristic function approach. *Test* 17:546–566
- Marcus MB (1981) Weak convergence of the empirical characteristic function. *Ann Probab* 9:194–201
- Meintanis SG (2004) Goodness-of-fit tests for the logistic distribution based on empirical transformations. *Sankhyā* 66:306–326
- Mimoto N (2008) Convergence in distribution for the sup-norm of a kernel density estimator for GARCH innovations. *Stat Probab Lett* 78:915–923
- Pascual L, Romo J, Ruiz E (2006) Bootstrap prediction for returns and volatilities in GARCH models. *Comput Stat Data Anal* 50:2293–2312
- Peng L, Yao Q (2003) Least absolute deviations for ARCH and GARCH models. *Biometrika* 90:967–975
- Rydberg TH (2000) Realistic statistical modelling of financial data. *Int Stat Rev* 68(3):233–258
- Serfling RJ (1980) *Approximation theorems of mathematical statistics*. Wiley