



Goodness-of-fit tests for parametric specifications of conditionally heteroscedastic models

M. Dolores Jiménez-Gamero¹ · Sangyeol Lee² · Simos G. Meintanis^{3,4}

Received: 7 November 2018 / Accepted: 2 August 2019 / Published online: 12 August 2019

© Sociedad de Estadística e Investigación Operativa 2019

Abstract

We consider a goodness-of-fit test for certain parametrizations of conditionally heteroscedastic time series with unobserved components. Our test is quite general in that it can be employed to validate any given specification of arbitrary order and may even be invoked for testing not just GARCH models but also some related models such as autoregressive conditional duration models. The test statistic utilizes the characterization of Bierens (J Econom 20:105–134, 1982) and may be written down in a convenient closed-form expression. Consistency of the test is proved, and the asymptotic distribution of the test statistic under the null hypothesis is studied. Since this distribution depends on unknown quantities, two bootstrap resampling schemes are investigated and compared in order to approximate critical points and actually carry out the test. Finite-sample results are presented as well as applications of the proposed procedures to real data from the financial markets.

Keywords GARCH model · Goodness-of-fit test · Bierens characterization · Conditional duration model · Weighted bootstrap

Simos G. Meintanis: on sabbatical leave from the University of Athens.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11749-019-00676-0>) contains supplementary material, which is available to authorized users.

✉ M. Dolores Jiménez-Gamero
dolores@us.es

Sangyeol Lee
sylee@stats.snu.ac.kr

Simos G. Meintanis
simosmei@econ.uoa.gr

¹ Department of Statistics and Operations Research, University of Sevilla, Sevilla, Spain

² Department of Statistics, Seoul National University, Seoul, South Korea

³ Department of Economics, National and Kapodistrian University of Athens, Athens, Greece

⁴ Unit for Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa

Mathematics Subject Classification 62G10 · 62G20

1 Introduction

Consider a multiplicative strictly stationary process

$$x_t = \sigma_t \varepsilon_t, \quad (1)$$

where ε_t and σ_t are independent, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed (IID) random variables with mean zero and unit variance, and suppose $\sigma_t^2 = \text{Var}(x_t | \mathbb{I}_{t-1})$, where \mathbb{I}_t denotes the information available at time t . A GARCH model is a specification of the conditional variance which takes the general form

$$\sigma_t^2 = h(\mathbf{Y}_{t,p,q}), \quad (2)$$

where $\mathbf{Y}_{t,p,q} = (x_{t-1}, \dots, x_{t-p}, \log(\sigma_{t-1}^2), \dots, \log(\sigma_{t-q}^2))'$ and $h : \mathbb{R}^r \mapsto \mathbb{R}$, with $r = p + q$, being $p \geq 1$ and $q \geq 0$. The classical GARCH corresponds to

$$h(\mathbf{Y}_{t,p,q}) = \omega + \sum_{j=1}^p \gamma_j x_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z} \quad (3)$$

Although (3) is indeed the most popular specification of the function $h(\cdot)$ in (2) for the multiplicative time series in (1), there exist numerous alternative instances of (2), such as the asymmetric log-GARCH, the QGARCH and the EGARCH models, among many others, which have proved useful in applications. Some of these models are considered in our Monte Carlo study in Sect. 6. For more information, the reader is referred to Rachev and Mittnik (2000), Francq and Zakoïan (2010), Teräsvirta (2009) and Jondeau et al. (2007).

In this paper, we construct goodness-of-fit tests for correct specification of the function $h(\cdot)$ in (2) in the context of model (1). This problem has been considered by Berkes et al. (2003a), Hong and Lee (2003), Escanciano (2008), Halunga and Orme (2009), Leucht et al. (2015), Zheng et al. (2018) and Francq et al. (2018), among others.

A number of such tests are based on the following idea: If model (2) is correctly specified, then the sets $\{\varepsilon_t = x_t/h(\mathbf{Y}_{t,p,q})^{1/2}\}$ and $\{\varepsilon_t\}$ coincide; therefore, the random variables in $\{\varepsilon_t = x_t/h(\mathbf{Y}_{t,p,q})^{1/2}\}$ must be IID. For instance, the tests in Berkes et al. (2003a), Zheng et al. (2018) and Francq et al. (2018) are portmanteau-type tests based on a finite number of correlations of some functions of the residuals. It is well known that these tests are not globally consistent.

Now, when it comes to testing independence in a time series, methods based on characteristic functions such as the tests of Hong (1999) and the tests based on the so-called distance (auto)covariance originally suggested by Székely et al. (2007) and adapted to serial correlation by Fokianos and Pitsillou (2018) and Davis et al. (2018), all with observable data, are quite suitable. Davis et al. (2018) also applied such distance to the residuals of an autoregressive process. Hong and Lee (2003) adapted

the method in Hong (1999) for unobservable innovations $\{\varepsilon_t\}$ in linear and nonlinear time series models. This proposal is not consistent against all alternatives since it is centered on the pairwise independence of ε_t and ε_{t+s} , for all s , which is weaker than the independence of all innovations. We nevertheless revisit this test in our Monte Carlo study.

The relation to characteristic function-based methods with our method lies entirely in the weighting scheme which uses the complex exponential function, but further discussion on this issue is postponed till Remark 2. In fact, as it will be seen further down the paper, our method deals with the specification problem not via independence, but directly targeting correct specification of the conditional variance. As such, the starting point of our method is in the same spirit as the methods of Halunga and Orme (2009), Leucht et al. (2015) and Escanciano (2008). Nevertheless, Escanciano's method is in the frequency domain, which is rather different from ours. Halunga and Orme's approach is based on the fact that under model (2), ε_t and $\Upsilon_{t,p,q}$ must be uncorrelated, and their tests are not universally consistent. Finally, the test in Leucht et al. (2015) is restricted to the GARCH(1,1) case. On the other hand, our method is applicable to a wide range of models of general (fixed) order, and by virtue of Bierens characterization, it is consistent against all fixed alternative specifications of the model under test.

The remainder of the paper is outlined as follows. In Sect. 2, we introduce the new test statistic and discuss some computational aspects. Section 3 is devoted to the asymptotic properties of the proposed method. Specifically, we derive the limit (in probability) of the proposed test statistic as well as its asymptotic null distribution and show that the test which rejects the null hypothesis for large values of the test statistic is consistent against fixed alternatives. Because the asymptotic null distribution depends on certain unknown quantities, Section 4 studies in detail a consistent approximation to the null distribution. Specifically, we propose to approximate the null distribution by means of a weighted bootstrap estimator. Section 5 shows that the conditions required for the consistency of the weighted bootstrap estimator are met for certain popular specifications of (2), including conditional duration models. Section 6 summarizes a Monte Carlo experiment carried out to study the finite-sample properties of the method, along with empirical applications. Finally, we end in Section 7 with conclusions and discussion. All proofs are included in Supplementary Material.

Throughout the paper, we will make use of the following 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; if $A = (a_{jk})$ is a matrix, then $\|A\| = \sum_{j,k} |a_{jk}|$; for any complex number $x = a + ib$, $\bar{x} = a - ib$ and $|x| = \sqrt{a^2 + b^2} = \sqrt{x\bar{x}}$, with $i = \sqrt{-1}$; P_0 denotes probability by assuming that the null hypothesis is true; P_* , \mathbb{E}_* and Cov_* denote the conditional probability law, expectation and covariance, given the data, respectively; all limits in this paper are taken when $T \rightarrow \infty$, where T denotes the sample size; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; \xrightarrow{P} denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; L_w^2 stands for the space of all L^2 functions defined on the measure space $(\mathbb{R}^r, \mathcal{B}_r, \mu)$, where \mathcal{B}_r denotes the σ -field of Borel subsets of \mathbb{R}^r , and the measure μ has density w : $d\mu(\mathbf{u}) = w(\mathbf{u})d\mathbf{u}$, that is, $L_w^2 = \{f : \mathbb{R}^r \mapsto \mathbb{C} : \|f\|_w^2 = \int |f(\mathbf{u})|^2 w(\mathbf{u})d\mathbf{u} < \infty\}$, where an unspecified

integral denotes integration over \mathbb{R}^r ; $\langle \cdot, \cdot \rangle_w$ denotes the scalar product in the Hilbert space L^2_w .

2 The test statistic

Let $x_t, 1 \leq t \leq T$, be observations on the process (1). On the basis of these observations, we wish to test the null hypothesis

$$\mathcal{H}_0 : \sigma_t^2 = h_{\boldsymbol{\vartheta}}(\boldsymbol{\Upsilon}_{t,p,q}), \text{ for some } \boldsymbol{\vartheta} \in \Theta \subseteq \mathbb{R}^v, \tag{4}$$

where $h_{\boldsymbol{\vartheta}}(\cdot)$ denotes a specific function incorporating a multidimensional parameter $\boldsymbol{\vartheta}$.

In order to formulate our procedure, let $\sigma_t^2(\boldsymbol{\vartheta}) := h_{\boldsymbol{\vartheta}}(\boldsymbol{\Upsilon}_{t,p,q})$ and notice that, under the null hypothesis \mathcal{H}_0 , we have $\mathbb{E}(x_t^2 | \boldsymbol{\Upsilon}_{t,p,q}) = \sigma_t^2(\boldsymbol{\vartheta})$. In this connection, recall the characterization of Bierens (1982) that for real x and a given vector \mathbf{y} of dimension r , $\mathbb{E}(x|\mathbf{y}) = 0$ holds if and only if $\mathbb{E}(xe^{i\mathbf{u}'\mathbf{y}}) = 0$, for all $\mathbf{u} \in \mathbb{R}^r$. In view of this characterization, let $\widehat{\varepsilon}_t := x_t/\sigma_t(\widehat{\boldsymbol{\vartheta}}_T)$ be the residuals, where $\widehat{\boldsymbol{\vartheta}}_T$ is a consistent estimator of the parameter $\boldsymbol{\vartheta}$. Recall that $\sigma_t^2(\boldsymbol{\vartheta})$ depends on $\{x_k, -\infty < k \leq t - 1\}$, whereas we observe x_1, \dots, x_T . So, in order to calculate the residuals, instead of $\sigma_t^2(\widehat{\boldsymbol{\vartheta}}_T)$, we consider $\widetilde{\sigma}_t^2(\widehat{\boldsymbol{\vartheta}}_T)$ defined as $\sigma_t^2(\widehat{\boldsymbol{\vartheta}}_T)$ on the basis of initial values $\widetilde{\mathbf{Y}}_1 := (\widetilde{x}_0, \dots, \widetilde{x}_{1-p}, \log(\widetilde{\sigma}_0^2), \dots, \log(\widetilde{\sigma}_{1-q}^2))'$. Thus, in practice, instead of $\widehat{\varepsilon}_t$, we consider $\widetilde{\varepsilon}_t := x_t/\widetilde{\sigma}_t(\widehat{\boldsymbol{\vartheta}}_T), 1 \leq t \leq T$.

With this notation, our core test-process is defined as

$$S_T(\mathbf{u}) = \frac{1}{T - m} \sum_{t=m+1}^T (\widetilde{\varepsilon}_t^2 - 1) e^{i\mathbf{u}'\widetilde{\mathbf{Y}}_t}, \quad \mathbf{u} \in \mathbb{R}^r, \tag{5}$$

where $\widetilde{\mathbf{Y}}_t := (x_{t-1}, \dots, x_{t-p}, \log(\widetilde{\sigma}_{t-1}^2(\widehat{\boldsymbol{\vartheta}}_T)), \dots, \log(\widetilde{\sigma}_{t-q}^2(\widehat{\boldsymbol{\vartheta}}_T)))', m = \max\{p, q\}$. Under \mathcal{H}_0 , it is expected that $S_T(\mathbf{u})$ will be close to zero for large sample size $T, \forall \mathbf{u} \in \mathbb{R}^r$. We consider as test statistic the normalized integrated process

$$Q_{T,w} = (T - m) \int |S_T(\mathbf{u})|^2 w(\mathbf{u}) d\mathbf{u} = (T - m) \|S_T\|_w^2, \tag{6}$$

where $w(\cdot)$ denotes a weight function whose choice is discussed below. Rejection of the null hypothesis \mathcal{H}_0 is for large values of $Q_{T,w}$.

We would like to underline the computational simplicity of the test statistic in (6): by straightforward algebra, from (5)

$$|S_T(\mathbf{u})|^2 = \frac{1}{(T - m)^2} \sum_{t,s=m+1}^T (\widetilde{\varepsilon}_t^2 - 1) (\widetilde{\varepsilon}_s^2 - 1) \cos\{\mathbf{u}'(\widetilde{\mathbf{Y}}_t - \widetilde{\mathbf{Y}}_s)\},$$

which makes it possible to express the test statistic as

$$Q_{T,w} = \frac{1}{T - m} \sum_{t,s=m+1}^T (\widetilde{\varepsilon}_t^2 - 1) (\widetilde{\varepsilon}_s^2 - 1) I_w(\widetilde{\mathbf{Y}}_t - \widetilde{\mathbf{Y}}_s), \tag{7}$$

where

$$I_w(\mathbf{x}) = \int_{\mathbb{R}^r} \cos(\mathbf{u}'\mathbf{x}) w(\mathbf{u}) d\mathbf{u}. \tag{8}$$

The computation is further simplified by considering specific instances of the weight function $w(\cdot)$ in (8). For instance, if we choose the density of the spherical stable distribution [see Nolan (2013)], then we readily obtain from (8), $I_w(\mathbf{x}) = e^{-\|\mathbf{x}\|^\gamma}$, where $0 < \gamma \leq 2$ stands for the characteristic exponent of the spherical stable law. The resulting test statistic in (7) reduces to

$$Q_{T,\gamma} = \frac{1}{T-m} \sum_{t,s=m+1}^T (\tilde{\varepsilon}_t^2 - 1) (\tilde{\varepsilon}_s^2 - 1) e^{-\|\tilde{\mathbf{Y}}_t - \tilde{\mathbf{Y}}_s\|^\gamma}. \tag{9}$$

Here, we only employ the Gaussian density as weight function, leading to the test statistic in (9) with $\gamma = 2$, since this is the only member of the spherical stable class satisfying Assumption 7 which is instrumental in deriving the asymptotic results of the next section.

Remark 1 The null hypothesis \mathcal{H}_0 states, among other things, that σ_t^2 depends on \mathbb{I}_{t-1} through $\mathbf{Y}_{t,p,q}$. Obviously, $\mathbf{Y}_{t,p,q}$ could be replaced with any one-to-one transform of it (by properly modifying the expression of $h_{\boldsymbol{\vartheta}}(\cdot)$). The motivation of our choice for $\mathbf{Y}_{t,p,q}$ is of pure technical nature: a commonly used estimator for the parameter $\boldsymbol{\vartheta}$ in GARCH models is the quasimaximum likelihood estimator (QMLE) which minimizes, with respect to $\boldsymbol{\vartheta}$

$$\sum_{t=1}^T \tilde{\ell}_t(\boldsymbol{\vartheta}), \quad \text{where} \quad \tilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{x_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})} + \log \tilde{\sigma}_t^2(\boldsymbol{\vartheta}).$$

The conditions required for the QMLE to be a consistent and asymptotically normal estimator of $\boldsymbol{\vartheta}$ coincide with those required to derive asymptotic properties of the test statistic (6) with our choice of $\mathbf{Y}_{t,p,q}$.

3 Asymptotic properties

In the asymptotics, C (with $C > 0$) and ϱ (with $0 < \varrho < 1$) will denote generic constants whose exact values are unimportant and may vary across the text. Also let $\dot{\sigma}_t^2(\boldsymbol{\vartheta}) = \frac{\partial}{\partial \boldsymbol{\vartheta}} \sigma_t^2(\boldsymbol{\vartheta})$ and $\ddot{\sigma}_t^2(\boldsymbol{\vartheta}) = \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \sigma_t^2(\boldsymbol{\vartheta})$.

Assumption 1 $\{x_t\}_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic process.

Assumption 2 (a) $\hat{\boldsymbol{\vartheta}}_T \xrightarrow{P} \boldsymbol{\vartheta}_0$, for some $\boldsymbol{\vartheta}_0$ that belongs to the interior of Θ , $\boldsymbol{\vartheta}_0$ being the true parameter value when \mathcal{H}_0 is true.

(b) $\mathbb{E} \left[\left\{ \frac{x_t^2}{\sigma_t^2(\boldsymbol{\vartheta}_0)} \right\}^{(1+\delta)} \right] < \infty$, for some $\delta > 0$.

Assumption 3 $\sup_{\vartheta \in \Theta} \left| \frac{1}{\sigma_t^2(\vartheta)} - \frac{1}{\tilde{\sigma}_t^2(\vartheta)} \right| \leq C\rho^t, \sup_{\vartheta \in \Theta} \left| \log\{\sigma_t^2(\vartheta)\} - \log\{\tilde{\sigma}_t^2(\vartheta)\} \right| \leq C\rho^t.$

Assumption 4 $\mathbb{E}|x_0|^s < \infty$, for some $s > 0$.

Assumption 5 For any $k > 0$, there exists a neighborhood Θ_k of ϑ_0 such that

$$\mathbb{E} \left[\sup_{\vartheta \in \Theta_k} \left\{ \frac{\sigma_t^2(\vartheta_0)}{\sigma_t^2(\vartheta)} \right\}^k \right] < \infty.$$

Assumption 6 The function $\vartheta \mapsto \sigma_t^2(\vartheta)$ is twice continuously differentiable, and the derivatives satisfy: for any $k > 0$, there exists a neighborhood Θ_k of ϑ_0 such that

$$\mathbb{E} \left[\sup_{\vartheta \in \Theta_k} \left\| \frac{\dot{\sigma}_t^2(\vartheta)}{\sigma_t^2(\vartheta)} \right\|^k \right] < \infty, \quad \mathbb{E} \left[\sup_{\vartheta \in \Theta_k} \left\| \frac{\ddot{\sigma}_t^2(\vartheta)}{\sigma_t^2(\vartheta)} \right\|^k \right] < \infty.$$

Assumption 7 $w : \mathbb{R}^r \mapsto (0, \infty)$ is an even function: $w(\mathbf{u}) = w(-\mathbf{u}), \forall \mathbf{u} \in \mathbb{R}^r$, satisfying

$$\int (1 + \|\mathbf{u}\|^2)w(\mathbf{u})d\mathbf{u} < \infty.$$

Notice that the neighborhoods in Assumptions 5 and 6 may not coincide. Nevertheless, without loss of generality, we can assume that they do coincide and thus use the same symbol to denote both neighborhoods.

Assumption 1 is a condition on the data generation process. Assumption 2 will be required to derive the limit (in probability) of the statistic $(T - m)^{-1}Q_{T,w}$. This assumption is also employed by Leucht et al. (2015) for testing goodness-of-fit to the linear GARCH(1,1) model. It is required to get a closed-form expression of such limit under alternatives. Notice that under the null hypothesis, Assumption 2 typically holds true: part (a) is satisfied by the QMLE and part (b) entails that $\mathbb{E}\{\varepsilon^{2(1+\delta)}\} < \infty$, for some $\delta > 0$, which is usually assumed for the asymptotic normality of the QMLE. Assumption 3 is required to ensure that the choice of the initial values $\tilde{\Upsilon}_1 := (\tilde{x}_0, \dots, \tilde{x}_{1-p}, \log(\tilde{\sigma}_0^2), \dots, \log(\tilde{\sigma}_{1-q}^2))'$ is unimportant for the asymptotic behavior of the test statistic. Assumption 4 is a mild moment condition on the data generation process. Assumptions 5 and 6 are conditions on the model under the null hypothesis that must be checked for each particular model.

As mentioned in Sect. 2, rejection of the null hypothesis \mathcal{H}_0 is for large values of $Q_{T,w}$. The next result gives a sound justification for such assertion. Let $\Upsilon_t(\vartheta) = (x_{t-1}, \dots, x_{t-p}, \log\{\sigma_{t-1}^2(\vartheta)\}, \dots, \log\{\sigma_{1-q}^2(\vartheta)\})'$.

Theorem 1 *Suppose that Assumptions 1–7 hold. Then,*

$$\frac{Q_{T,w}}{T - m} \xrightarrow{P} \kappa = \int_{\mathbb{R}^r} |\mathcal{E}_0(\mathbf{u})|^2 w(\mathbf{u})d\mathbf{u}, \tag{10}$$

where $\mathcal{E}_0(\mathbf{u}) := \mathcal{E}_{\boldsymbol{\vartheta}_0}(\mathbf{u})$ with

$$\mathcal{E}_{\boldsymbol{\vartheta}}(\mathbf{u}) = \mathbb{E} \left\{ \left(\frac{x_t^2}{\sigma_t^2(\boldsymbol{\vartheta})} - 1 \right) e^{i\mathbf{u}'\boldsymbol{\Upsilon}_t(\boldsymbol{\vartheta})} \right\}.$$

Theorem 1 states that $\frac{1}{T-m} Q_{T,w}$ converges in probability to a nonnegative constant $\kappa \geq 0$ which, from Bierens characterization, satisfies that $\kappa = 0$ iff the null hypothesis is true. Thus, a reasonable test of \mathcal{H}_0 should reject the null hypothesis for large values of $Q_{T,w}$. Now, to determine what are large values we must calculate its null distribution, or at least an approximation to it. The exact null distribution of $Q_{T,w}$ is clearly unknown. As an approximation, we will derive its asymptotic null distribution. With this aim, a further assumption is needed, which is listed below.

Assumption 8 Under the null hypothesis,

$$\widehat{\boldsymbol{\vartheta}}_T = \boldsymbol{\vartheta}_0 + \frac{1}{T} \sum_{t=1}^T L_t(\boldsymbol{\vartheta}_0) + o_P(T^{-1/2}),$$

with $L_t(\boldsymbol{\vartheta}_0) = (g_1(\varepsilon_t)l_1(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \dots, g_v(\varepsilon_t)l_v(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots))'$, $1 \leq t \leq T$, satisfying $\mathbb{E}\{g_u(\varepsilon_0)\} = 0$, $\mathbb{E}\{g_u(\varepsilon_0)^2\} < \infty$, $\mathbb{E}\{l_u(\varepsilon_{-1}, \varepsilon_{-2}, \dots)^2\} < \infty$, $1 \leq u \leq v$.

First, next theorem derives an asymptotic approximation for $Q_{T,w}$.

Theorem 2 Suppose that Assumptions 1–8 hold and that $\mathbb{E}(\varepsilon_0^4) < \infty$. If \mathcal{H}_0 is true,

$$Q_{T,w} = \|W_T(\cdot; \boldsymbol{\vartheta}_0)\|_w^2 + o_P(1),$$

where

$$\begin{aligned} W_T(\mathbf{u}; \boldsymbol{\vartheta}) &= \frac{1}{\sqrt{T-m}} \sum_{t=m+1}^T H_t(\mathbf{u}; \boldsymbol{\vartheta}), \\ H_t(\mathbf{u}; \boldsymbol{\vartheta}) &= (\varepsilon_t^2 - 1) [\cos\{\mathbf{u}'\boldsymbol{\Upsilon}_t(\boldsymbol{\vartheta})\} + \sin\{\mathbf{u}'\boldsymbol{\Upsilon}_t(\boldsymbol{\vartheta})\}] - G(\mathbf{u}; \boldsymbol{\vartheta})' L_t(\boldsymbol{\vartheta}), \\ G(\mathbf{u}; \boldsymbol{\vartheta}) &= \mathbb{E} \left(\frac{\dot{\sigma}_0^2(\boldsymbol{\vartheta})}{\sigma_0^2(\boldsymbol{\vartheta})} [\cos\{\mathbf{u}'\boldsymbol{\Upsilon}_0(\boldsymbol{\vartheta})\} + \sin\{\mathbf{u}'\boldsymbol{\Upsilon}_0(\boldsymbol{\vartheta})\}] \right). \end{aligned}$$

Next, as a consequence of the result in Theorem 2, the following corollary provides the asymptotic null distribution of $Q_{T,w}$.

Corollary 1 Suppose that the assumptions of Theorem 2 hold. Then,

$$Q_{T,w} \xrightarrow{\mathcal{L}} \|Z\|_w^2,$$

where $\{Z(\mathbf{u}), \mathbf{u} \in \mathbb{R}^r\}$ is a zero-mean Gaussian random element in L_w^2 with covariance kernel $\varrho_0(\mathbf{u}, \mathbf{v}) = \mathbb{E}\{H_0(\mathbf{u}; \boldsymbol{\vartheta}_0)H_0(\mathbf{v}; \boldsymbol{\vartheta}_0)\}$, $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^r$.

As an immediate consequence of Theorem 1 and Corollary 1, it follows that the test that rejects the null hypothesis for large values of the test statistic $Q_{T,w}$ is consistent against any fixed alternative such that $E(x_t^2 | \Upsilon_{t,p,q}) \neq h_{\vartheta}(\Upsilon_{t,p,q})$.

Remark 2 In general, Bierens characterization involving $\mathbb{E}\{x g(\mathbf{y}, \mathbf{u})\}$, holds for a wide range of “instrument” functions $g(\cdot, \cdot)$. For a thorough discussion on the possible choice of instruments, the reader is referred to Stinchcombe and White (1998). Here, we have restricted ourselves to the exponential function $g(\mathbf{y}, \mathbf{u}) = e^{i\mathbf{u}'\mathbf{y}}$ which was the original instrument considered in Bierens (1982). There are several reasons for this choice such as computational convenience (see Sect. 2) and power considerations. In this connection, we refer to Escanciano (2008), where the resulting tests for this choice of g exhibit favorable performance in finite samples. Nevertheless, as it will become evident from the proofs in Supplementary Material, all results still hold true for arbitrary instruments, provided that the associated function $g(\cdot, \cdot)$ is bounded and satisfies

$$|g(\mathbf{x}, \mathbf{u}) - g(\mathbf{y}, \mathbf{u})| \leq C \|\mathbf{u}\| \|\mathbf{x} - \mathbf{y}\|, \quad (11)$$

for some positive constant $C > 0$.

Remark 3 As an anonymous referee observed, the test statistic is not scale invariant. The reason is that properly specified GARCH models contain a scale-free error term and thus model the scale of the data (either directly the variance, or otherwise some form of volatility) via a certain regression-type function. In turn, this regression function is in most cases itself not scale equivariant, and consequently, the model itself, as well as the corresponding parameter estimates, is not scale equivariant. Hence, the notion of scale equivariance does not apply to GARCH models, and therefore, the corresponding goodness-of-fit tests do depend on the scale of the data

Going back to Corollary 1, we note that besides its dependence on ϑ_0 and the corresponding estimator, the asymptotic null distribution of the test statistic $Q_{T,w}$ also depends on the unknown distribution of the innovations. Therefore, the asymptotic null distribution does not provide a useful approximation to the null distribution of $Q_{T,w}$. In the next section, we circumvent this problem by proposing a resampling scheme that consistently estimates the limit law of $Q_{T,w}$.

4 Approximating the null distribution using multipliers

The tests proposed in Leucht et al. (2015) for testing goodness-of-fit to a linear GARCH(1,1) model is, in a sense, similar to our proposal (it can be seen as a particular case of our proposal if instead of taking the exponential function, a general function $g(\cdot, \cdot)$ satisfying (11) is considered). These authors proposed a residual-based bootstrap to approximate the null distribution of their test statistic. We could also consider a residual-based bootstrap procedure to approximate the null distribution of our test statistic but, as noted in Jeong (2017), the residual-based bootstrap actively employs the specific dependence structure of the model. In addition, the residual-based bootstrap

is very time-consuming, since the parameter must be estimated with each resample. Because of these reasons, next we prove the consistency of a widely applicable bootstrap method, the weighted bootstrap, also called the multipliers method. This method has been used effectively in the context of testing goodness-of-fit for the error distribution in linear GARCH models for methods based on the empirical distribution function in Ghoudi and Rémillard (2014, 2018), as well as for methods based on the empirical characteristic function in Jiménez-Gamero and Pardo-Fernández (2017).

In order to discuss the implementation of the multipliers method for approximating the null distribution of $Q_{T,w}$, recall that Theorem 2 states that $Q_{T,w} = Q_{1,T,w} + o_p(1)$, with

$$Q_{1,T,w} = \|W_T(\cdot; \boldsymbol{\vartheta}_0)\|_w^2.$$

Let ξ_{m+1}, \dots, ξ_T be IID copies of a random variable ξ with $\mathbb{E}(\xi) = 0$ and $\text{Var}(\xi) = 1$, independent of x_1, \dots, x_T . Let us consider the following weighted bootstrap version of $Q_{1,T,w}$,

$$Q_{1,T,w}^* = \|W_{1,T}^*(\cdot; \boldsymbol{\vartheta}_0)\|_w^2,$$

with

$$W_{1,T}^*(\mathbf{u}; \boldsymbol{\vartheta}_0) = \frac{1}{\sqrt{T-m}} \sum_{t=m+1}^T H_t(\mathbf{u}; \boldsymbol{\vartheta}_0) \xi_t.$$

The following result gives the large-sample conditional distribution of $Q_{1,T,w}^*$, given x_1, \dots, x_T .

Theorem 3 *Suppose that Assumptions 1 and 6–8 hold and that $\mathbb{E}(\varepsilon_0^4) < \infty$. Then,*

$$\sup_x \left| P_* (Q_{1,T,w}^* \leq x) - P \left(\|Z_1\|_w^2 \leq x \right) \right| \xrightarrow{P} 0,$$

where $\{Z_1(\mathbf{u}), \mathbf{u} \in \mathbb{R}^r\}$ is a zero-mean Gaussian random element in L_w^2 having covariance kernel $\varrho(\mathbf{u}, \mathbf{v}) = \mathbb{E}\{H_0(\mathbf{u}; \boldsymbol{\vartheta}_0)H_0(\mathbf{v}; \boldsymbol{\vartheta}_0)\}$, $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^r$.

As an immediate consequence of Theorem 3, the following corollary states that when \mathcal{H}_0 is true, the conditional distribution of $Q_{1,T,w}^*$, given x_1, \dots, x_T , provides a consistent estimator of the null distribution of $Q_{T,w}$.

Corollary 2 *Suppose that the assumptions of Theorem 2 hold. Then,*

$$\sup_x \left| P_* (Q_{1,T,w}^* \leq x) - P_0 (Q_{T,w} \leq x) \right| \xrightarrow{P} 0.$$

Although the conditional distribution, given x_1, \dots, x_T , of $Q_{1,T,w}^*$ provides a consistent estimator of the null distribution of $Q_{T,w}$, it presents the same drawbacks

as the asymptotic null distribution of $Q_{T,w}$, since this distribution depends on the unknown values of $H_{m+1}(\mathbf{u}; \boldsymbol{\vartheta}_0), \dots, H_T(\mathbf{u}; \boldsymbol{\vartheta}_0)$. These quantities depend on $\boldsymbol{\vartheta}_0$, on the function $G(\mathbf{u}; \boldsymbol{\vartheta}_0)$, as well as on the unobservable innovations. In addition, $L_t(\boldsymbol{\vartheta})$ in Assumption 8 is usually unknown. Thus, $Q_{1,T,w}^*$ cannot be used in practice. To overcome this problem, we replace ε_t by $\tilde{\varepsilon}_t$, ($m + 1 \leq t \leq T$), $G(\mathbf{u}; \boldsymbol{\vartheta}_0)$ by

$$\tilde{G}(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T) = \frac{1}{T - m} \sum_{t=m+1}^T \tilde{A}_t(\hat{\boldsymbol{\vartheta}}_T) \{ \cos(\mathbf{u}'\tilde{\mathbf{Y}}_t) + \sin(\mathbf{u}'\tilde{\mathbf{Y}}_t) \},$$

with

$$\tilde{A}_t(\boldsymbol{\vartheta}) = \frac{\partial}{\partial \boldsymbol{\vartheta}} \log\{\tilde{\sigma}_t^2(\boldsymbol{\vartheta})\}, \tag{12}$$

and $L_t(\boldsymbol{\vartheta}_0)$ by $\widehat{L}_t(\hat{\boldsymbol{\vartheta}}_0)$ satisfying

$$\frac{1}{T - m} \sum_{t=m+1}^T \|L_t(\boldsymbol{\vartheta}_0) - \widehat{L}_t(\hat{\boldsymbol{\vartheta}}_0)\|^2 \xrightarrow{P} 0. \tag{13}$$

A candidate for $\widehat{L}_t(\hat{\boldsymbol{\vartheta}}_0)$ will be discussed in Sect. 5 for some special cases of specifications in \mathcal{H}_0 . So, instead of $Q_{1,T,w}^*$ we consider

$$Q_{T,w}^* = \|W_T^*(\cdot; \hat{\boldsymbol{\vartheta}}_T)\|_w^2,$$

with

$$W_T^*(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T) = \frac{1}{\sqrt{T - m}} \sum_{t=m+1}^T \tilde{H}_t(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T)\xi_t,$$

and

$$\tilde{H}_t(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T) = (\tilde{\varepsilon}_t^2 - 1) \{ \cos(\mathbf{u}'\tilde{\mathbf{Y}}_t) + \sin(\mathbf{u}'\tilde{\mathbf{Y}}_t) \} - \tilde{G}(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T)' \widehat{L}_t(\hat{\boldsymbol{\vartheta}}_0).$$

The next result shows that $Q_{T,w}^*$ and $Q_{1,T,w}^*$ both have the same conditional limit distribution, given the observations. The main advantage of $Q_{T,w}^*$ over $Q_{1,T,w}^*$ is that the former does not depend on unknown quantities, and thus, at least in principle, its distribution can be calculated. This practical issue will be dealt with at the end of the section. From now on, the conditional distribution of $Q_{T,w}^*$, given the data, will be called the weighted bootstrap distribution of $Q_{T,w}$.

Assumption 9

$$\frac{1}{T - m} \sum_{t=m+1}^T \|A_t(\boldsymbol{\vartheta}_0) - \tilde{A}_t(\hat{\boldsymbol{\vartheta}}_T)\| \xrightarrow{P} 0.$$

Theorem 4 *Suppose that Assumptions 1–9 and (13) hold. Then*

$$\sup_x |P_*(Q_{T,w}^* \leq x) - P(Q_{1,T,w}^* \leq x)| \xrightarrow{P} 0.$$

Note that the result in Theorem 4 is valid whether or not the null hypothesis is true. Two immediate consequences follow.

Corollary 3 *If \mathcal{H}_0 is true and the assumptions of Theorem 4 hold, then*

$$\sup_x |P_*(Q_{T,w}^* \leq x) - P_0(Q_{T,w} \leq x)| \xrightarrow{P} 0.$$

In view of Corollary 3 and for $\alpha \in (0, 1)$, we consider the test function

$$\Psi^* = \Psi^*(x_1, \dots, x_T) = \begin{cases} 1, & \text{if } Q_{T,w} \geq q_\alpha^*, \\ 0, & \text{otherwise,} \end{cases}$$

where q_α^* is the $1 - \alpha$ percentile of the conditional distribution of $Q_{T,w}^*$, given x_1, \dots, x_T , or equivalently, $\Psi^* = 1$ if $p^* \leq \alpha$, where $p^* = P_*\{Q_{T,w}^* \geq Q_{T,w,obs}\}$, $Q_{T,w,obs}$ being the observed value of the test statistic $Q_{T,w}$. The result in Corollary 3 states that the test Ψ^* is asymptotically correct, in the sense that its type I error is asymptotically equal to the nominal level α . In the next corollary, we prove that Ψ^* is asymptotically able to detect any fixed alternative.

Corollary 4 *If \mathcal{H}_0 is not true and the assumptions of Theorem 4 hold, then $P(\Psi^* = 1) \rightarrow 1$.*

The calculation of the exact weighted bootstrap distribution of $Q_{T,w}$ is, from a practical point of view, not feasible. Nevertheless, the weighted bootstrap p -value can be approximated by simulation, following the steps in next algorithm. Let $M = (M_{t,s})$ denote the $(T - m) \times (T - m)$ -matrix with elements

$$M_{t,s} = \int \tilde{H}_{m+t}(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T) \tilde{H}_{m+s}(\mathbf{u}; \hat{\boldsymbol{\vartheta}}_T) w(\mathbf{u}) d\mathbf{u}, \quad 1 \leq t, s \leq T - m.$$

Algorithm

1. Compute the test statistic $Q_{T,w,obs}$.
2. For some large integer B , repeat for every $b \in \{1, \dots, B\}$:
 - (a) Generate $\xi^b = (\xi_{m+1}^b, \dots, \xi_T^b)'$.
 - (b) Calculate $Q_{T,w}^{*b} = \xi^{b'} M \xi^b / (T - m)$.
3. Approximate the p -value corresponding to $Q_{T,w,obs}$ by $\hat{p}^* = \frac{1}{B} \sum_{b=1}^B I\{Q_{T,w}^{*b} > Q_{T,w,obs}\}$.

Before moving to applications, we reiterate that in the weighted bootstrap algorithm, the parameter $\boldsymbol{\vartheta}$ needs to be estimated only once, while the residual-based bootstrap counterpart requires estimation of this parameter for each resample.

5 Applications

The theory so far developed for testing \mathcal{H}_0 is very general. This section discusses the applicability of the proposed procedure to testing for some special models.

5.1 Testing for the linear GARCH(p,q) model

Let us now consider the problem of testing for the linear GARCH model defined in (3), for some $p \geq 1, q \geq 0$. The consistency and the asymptotic normality of the QMLE in this model were first studied in Berkes et al. (2003b). Then, Francq and Zakoian (2004) derived those properties under somewhat weaker assumptions, so we will follow the approach in the latter paper.

Francq and Zakoian (2004) have given a necessary and sufficient condition for Assumption 1 to hold. If such a condition holds, then Assumption 4 is met (ibid. Proposition 1). Let $\boldsymbol{\vartheta} = (\omega, \gamma_1, \dots, \gamma_q, \beta_1, \dots, \beta_q)' \in \Theta \subset (0, \infty) \times [0, \infty)^{p+q}$, for some compact set Θ . Assume that $\boldsymbol{\vartheta}_0 = (\omega_0, \gamma_{01}, \dots, \gamma_{0q}, \beta_{01}, \dots, \beta_{0q})'$ belongs to the interior of Θ , ε_t^2 has a non-degenerate distribution satisfying $\mathbb{E}(\varepsilon_0^4) < \infty$, $\{x_t\}$ is stationary and ergodic, $\sum_{j=1}^q \beta_{0j} < 1$ and that the parameter $\boldsymbol{\vartheta}_0$ is identifiable. Then, the QMLE of $\boldsymbol{\vartheta}$ satisfies Assumption 8 with

$$\begin{aligned} L_j &= L_j(\boldsymbol{\vartheta}) = (\varepsilon_j^2 - 1)A_t(\boldsymbol{\vartheta})J^{-1}, \quad 1 \leq t \leq T, \\ J &= E\{A_1(\boldsymbol{\vartheta})A_1(\boldsymbol{\vartheta})'\} \\ A_t(\boldsymbol{\vartheta}) &= \frac{\dot{\sigma}_t^2(\boldsymbol{\vartheta})}{\sigma_t^2(\boldsymbol{\vartheta})}, \quad 1 \leq t \leq T, \end{aligned} \tag{14}$$

(ibid. Theorem 2.2). Under these conditions, it follows that Assumptions 3, 5 and 6 hold (ibid. Theorems 2.1 and 2.2).

For the consistency of the weighted bootstrap distribution estimator of $Q_{T,w}$, we also need $\widehat{L}_t(\widehat{\boldsymbol{\vartheta}}_0)$ satisfying (13). Let

$$\begin{aligned} \widehat{L}_t &= (\widehat{\varepsilon}_j^2 - 1)\widetilde{A}_j(\widehat{\boldsymbol{\vartheta}}_T)\widehat{J}^{-1}, \quad 1 \leq t \leq T, \\ \widehat{J} &= \frac{1}{n} \sum_{t=1}^T \widetilde{A}_t(\widehat{\boldsymbol{\vartheta}}_T)\widetilde{A}_t(\widehat{\boldsymbol{\vartheta}}_T)', \end{aligned} \tag{15}$$

and $\widetilde{A}_j(\boldsymbol{\vartheta})$ as defined in (12). Proposition 7.1 in Jiménez-Gamero and Pardo-Fernández (2017) shows that $\{\widehat{L}_t, 1 \leq t \leq T\}$ satisfies (13), while Lemma B.1 (f) of the same paper guarantees that Assumption 9 holds. Therefore, our procedure can be applied to testing for the linear GARCH model when the parameter is estimated by means of the QMLE.

We note that asymptotic results in Leucht et al. (2015) for the special case of the linear GARCH(1,1) model are obtained under the same assumptions, but the corresponding residual-based bootstrap resampling scheme employed requires stronger assumptions than our proposal. As it was mentioned in Introduction, the approach in Escanciano (2008) is rather different from ours, even the required assumptions are not

comparable: although that paper does not assume that the innovations are IID, when the test in that paper is applied to test for \mathcal{H}_0 , it requires $\mathbb{E}\{(x_0^2 - \sigma_0^2)^2\} < \infty$, which is ensured if $\mathbb{E}\{x_0^4\} < \infty$ and $\mathbb{E}\{\sigma_0^4\} < \infty$. From Theorem 2.9 in Francq and Zakoian (2010), the existence of such moments entails certain restrictions on the parametric space Θ .

5.2 Testing for the asymmetric log-GARCH(p,q) model

Although the linear GARCH model is the most popular, it possesses some limitations such as (i) the positivity of the components of ϑ which complicates estimation procedures, (ii) the assumption that volatility is bounded from below and (iii) the lack of the leverage effect whereby negative observations tend to have more impact on volatility than positive observations of the same magnitude. Because of these reasons, some alternative models have been suggested. Here, we consider the asymmetric log-GARCH(p,q) model, whose specification of the function $h(\cdot)$ in (2) is given by

$$h(\Upsilon_{t,p,q}) = \exp \left\{ \omega + \sum_{j=1}^p [\gamma_{j+} I(x_{t-j} > 0) + \gamma_{j-} I(x_{t-j} < 0)] \log(x_{t-j}^2) + \sum_{j=1}^q \beta_j \log(\sigma_{t-j}^2) \right\}, \quad \forall t = 1, 2, \dots$$

For recent works on this class of models, the reader is referred to Francq et al. (2013, 2018), and the references therein.

Theorem 2.1 in Francq et al. (2013) gives a sufficient condition for Assumption 1 to hold. Theorem 3.1 (see also Proposition 3.3 for the case $p = q = 1$) in that paper gives a sufficient condition for Assumption 4 to hold. Expressions (7.6) and (7.8) in the proof of Theorem 4.1 in Francq et al. (2013) show that, under the assumptions of that theorem, Assumption 3 is also met. Under those assumptions, our Assumptions 5 and 6 also hold (ibid. Theorem 4.2). Also the QMLE of ϑ satisfies Assumption 8 with $L_j(\vartheta)$ as in (14).

For the consistency of the weighted bootstrap distribution estimator of $Q_{T,w}$, we also need $\widehat{L}_t(\vartheta_0)$ satisfying (13). With this aim, we can use the approximation defined in (15). Then,

$$\sup_{\vartheta} \left\| \frac{\partial}{\partial \vartheta} \log\{\sigma_t^2(\vartheta)\} - \frac{\partial}{\partial \vartheta} \log\{\tilde{\sigma}_t^2(\vartheta)\} \right\| \leq C \rho^t,$$

(ibid. Theorem 4.2), which implies that Assumption 9 holds. Then, Proposition 7.1 in Jiménez-Gamero and Pardo-Fernández (2017) shows that $\{\hat{L}_t, 1 \leq t \leq T\}$ satisfies (13), since the proof is based on the assumptions made, and not on the particular specification of the linear GARCH. Therefore, our procedure can be applied to testing for the asymmetric log-GARCH model when the parameter is estimated by means of the QMLE.

5.3 Testing for specification of duration models

Engle and Russell (1998) proposed the linear autoregressive conditional duration model to analyze high-frequency data on durations between successive financial market trades. Later on, since the linear specification is so restrictive in practice, as pointed out by Dufour and Engle (2000), various alternative nonlinear conditional duration models have been proposed (for example, Bauwens and Giot (2003) proposed logarithmic autoregressive conditional duration models).

Let us consider the autoregressive conditional duration model

$$\begin{aligned}y_t &= \psi_t \epsilon_t, \\ \psi_t &= h_{\vartheta}(\mathbf{\Lambda}_{t,p,q}),\end{aligned}$$

where $\mathbf{\Lambda}_{t,p,q} = (y_{t-1}, \dots, y_{t-p}, \log \psi_{t-1}, \dots, \log \psi_{t-q})'$, and $\{h_{\vartheta} : \vartheta \in \Theta\}$ denotes a specific parametric family of nonnegative functions that involves a parameter vector ϑ , and $\{\epsilon_i\}_{i \in \mathbb{Z}}$ is a sequence of IID positive random variables with $\mathbb{E}\epsilon_0 = 1$ and $\mathbb{E}\epsilon_0^2 < \infty$.

In practice, y_t is the duration between two consecutive events such as the quote, price change with ψ_t being the conditional expected value given past observations.

Notice that an autoregressive conditional duration model has the dynamics of the square of a GARCH model. Since our procedure, initially designed for testing model specification in GARCH models, depends on the square of the data and on the specification for σ_t^2 , all results carry over to the present model, and thus, specification testing may be carried out just by replacing x_t^2 and σ_t^2 by y_t and ψ_t , respectively.

6 Monte Carlo and real data applications

6.1 Monte Carlo results

This section is devoted to the study of the finite-sample performance of the proposed tests in terms of level approximation and power. With this purpose, a Monte Carlo simulation study was conducted. In all cases, the innovations are sequences of IID variables having a standard normal distribution and the parameters were estimated by means of the QMLE. The results displayed in the tables are the empirical powers obtained by generating in each case 1000 samples of size $T = 500$. As weight function $w(\cdot)$, we took the density function of a $(p + q)$ -variate normal distribution which yields $I_w(\mathbf{x}) = \exp\{-(1/a)\|\mathbf{x}\|^2\}$ for several values of $a > 0$, which meets Assumption 7. The nominal significance levels are $\alpha = 0.05$ and $\alpha = 0.10$. With the aim of comparing the behavior of the approximations provided by the weighted bootstrap (WB in the tables) studied in Sect. 4 and the residual-based bootstrap (RB in the tables) as defined in Leucht et al. (2015), we calculated both of them with $B = 1000$ for the weighted bootstrap and by generating $B = 500$ bootstrap samples for the residual-based bootstrap, as in Leucht et al. (2015). All computations in this paper have been performed by using programs written in the R language (R Core Team 2018).

Table 1 Percentages of rejections under $\mathcal{H}_{0,\text{lin}}$ for $T = 500$ (level)

(ω, γ, β)	Proposed test					Portm.			HL		
	a	WB		RB		k	5%	10%	p	5%	10%
		5%	10%	5%	10%						
(0.1, 0.1, 0.3)	0.5	4.4	10.6	4.9	10.1	8	2.8	5.0	3	3.6	6.1
	1.0	4.7	10.6	4.4	9.8	16	3.8	6.7	6	3.0	6.1
	1.5	5.0	10.9	4.6	10.0	24	4.2	7.6	9	3.3	6.1
(0.1, 0.1, 0.5)	0.5	4.4	9.5	4.5	9.3	8	3.0	5.4	3	3.6	6.2
	1.0	4.7	10.2	4.5	9.3	16	4.4	7.2	6	3.1	6.3
	1.5	5.4	10.2	4.9	9.4	24	4.5	8.1	9	3.1	6.7
(0.1, 0.1, 0.7)	0.5	4.6	9.1	4.0	7.9	8	4.4	8.4	3	3.6	6.1
	1.0	4.3	9.6	4.0	8.2	16	4.7	7.2	6	3.3	6.4
	1.5	4.4	9.9	4.0	8.5	24	4.9	8.9	9	2.8	6.8
(0.1, 0.1, 0.8)	0.5	7.1	11.3	3.4	7.1	8	14.0	18.6	3	3.8	6.5
	1.0	7.9	12.7	3.5	6.2	16	12.0	16.5	6	4.1	7.0
	1.5	7.8	13.3	3.2	6.3	24	11.3	15.0	9	4.6	8.4

We first considered the problem of testing goodness-of-fit for a linear GARCH(1,1) model, that is,

$$\mathcal{H}_{0,\text{lin}} : \sigma_t^2 = \omega + \gamma x_{t-1}^2 + \beta \sigma_{t-1}^2.$$

The parameter (ω, γ, β) was estimated by the QMLE by using the function `garch` of the R package `tseries`. In addition to the proposal in this paper, we tried the portmanteau-type test in Berkes et al. (2003a) (headed as Portm. in the tables) for several values of the number of lags considered, k , and the test in Hong and Lee (2003) (headed as HL in the tables) for several values of the bandwidth, p . Table 1 displays the results for the level. For the test proposed in this paper, when $\beta \leq 0.7$, both approximations give actual levels which are quite close to the nominal ones; for $\beta = 0.8$, the weighted bootstrap approximation is a bit liberal, and the residual-based bootstrap approximation becomes very conservative. Unreported simulation results indicate that for both methods a sample size as large as $T = 1000$ is required in order to obtain empirical levels close to the nominal values for $\beta = 0.8$ and that even higher samples sizes are required for more persistent models. Motivated by a real data set application, we also tried the weighted bootstrap approximation for $T = 3300$, $\omega = 0.02$, $\gamma = 0.04$ and $\beta = 0.95$. For this sample size, the residual-based bootstrap approximation becomes unaffordable. Table 2 displays the results, showing that when the model is very persistent, but the sample size is large, the weighted bootstrap provides an adequate approximation.

The test in Berkes et al. (2003a) is conservative for small values of β and rather liberal for large β . In unreported simulations, we observed that such behavior for large β persists even for larger sample sizes. We also observe that the actual level

Table 2 Percentages of rejections under $\mathcal{H}_{0,\text{lin}}$ for $T = 3300$ with the weighted bootstrap approximation (level)

(ω, γ, β)	α (%)	a		
		0.5	1.0	1.5
(0.02, 0.04, 0.95)	5	5.2	4.9	4.4
	10	8.8	8.4	8.9

strongly depends on the number of lags considered. The test in Hong and Lee (2003) is conservative for all bandwidths considered, and its level is robust to the choice of the bandwidth, as also observed by the authors in their paper.

As alternatives, we considered the QGARCH(1,1) defined through Eq. (1) with

$$\sigma_t^2 = \omega + \gamma x_{t-1}^2 + \beta \sigma_{t-1}^2 + \delta x_{t-1},$$

and the asymmetric log-GARCH(1,1), as defined in Sect. 5.2. Table 3 displays the obtained results for the power. It also includes the test in Leucht et al. (2015) (labeled as LNK in Table 3) with the same choices for the functions involved in the definition of their test statistic as those considered in their simulations. For the proposal in this paper and QGARCH alternative, the test based on the weighted bootstrap approximation and the one based on the residual-based bootstrap approximation exhibit quite close powers, but for the asymmetric log-GARCH alternative the test based on the weighted bootstrap approximation is clearly more powerful. Although under the null hypothesis both approximations should be close (at least for large sample sizes), under alternatives they may be rather different, and thus, the powers may considerably differ. In all tried cases, the test proposed in this paper outperforms the ones in Leucht et al. (2015). Berkes et al. (2003a) and Hong and Lee (2003).

We also considered the problem of testing goodness-of-fit for the asymmetric log-GARCH(1,1) model, that is,

$$\mathcal{H}_{0,\text{alog}} : \log(\sigma_t^2) = \omega + \{\gamma_+ I(x_{t-1} > 0) + \gamma_- I(x_{t-1} < 0)\} \log(x_{t-j}^2) + \beta \log(\sigma_{t-1}^2).$$

As before, in addition to the proposal in this paper, we tried (an adaptation of) the portmanteau-type test in Francq et al. (2018) (headed as Portm. in the tables) for several values of the number of lags considered, k , and again the test in Hong and Lee (2003) (headed as HL in the tables) for several values of the bandwidth, p . Table 4 displays the results for the level. As for the test proposed in this paper, both approximations give actual levels which are quite close to the nominal ones. Motivated by a real data set application, we also tried the weighted bootstrap approximation for $T = 3300$, $\omega = 0.04$, $\gamma_+ = 0.03$, $\gamma_- = 0.02$ and $\beta = 0.96$. For this sample size, the residual-based bootstrap approximation becomes unaffordable. Table 5 displays the results obtained. The empirical levels are quite close to the nominal values. The portmanteau-type test gives actual levels close to the nominal values for large k . The test in Hong and Lee (2003) is conservative for all bandwidths considered.

As alternative, we considered the QGARCH(1,1) model. Table 6 displays the obtained power results. As noticed before, for the proposal in this paper, the tests

Table 3 Percentages of rejections under alternatives for testing $\mathcal{H}_{0,\text{lin}}$ for $T = 500$ (power)

		Proposed test				Portm.			HL		
QGARCH(1,1)											
		WB		RB							
$(\omega, \gamma, \beta, \delta)$	a	5%	10%	5%	10%	k	5%	10%	p	5%	10%
(0.1, 0.1, 0.3, 0.05)	0.5	34.7	46.6	34.7	45.7	8	2.6	5.4	3	11.1	14.4
	1.0	35.7	47.7	35.9	46.6	16	4.0	6.9	6	8.3	11.8
	1.5	36.8	48.2	35.0	46.5	24	4.7	8.2	9	6.2	10.4
	LKN			26.2	41.0						
(0.1, 0.1, 0.3, 0.10)	0.5	89.4	94.6	89.0	94.3	8	2.9	5.7	3	47.6	55.1
	1.0	91.0	95.1	90.5	94.9	16	3.8	7.0	6	36.9	46.1
	1.5	91.2	95.2	90.9	95.0	24	4.5	7.6	9	29.8	37.8
	LKN			81.3	91.1						
Asymmetric log-GARCH(1,1)											
		WB		RB							
$(\omega, \gamma_+, \gamma_-, \beta)$	a	5%	10%	5%	10%	k	5%	10%	p	5%	10%
(0.1, 0.1, 0.15, 0.3)	0.5	31.0	44.5	21.1	33.3	8	2.6	5.0	3	5.6	7.9
	1.0	22.5	35.0	12.8	21.9	16	3.0	5.7	6	4.1	7.4
	1.5	18.4	29.4	8.2	16.6	24	4.1	7.6	9	4.2	7.5
	LKN			2.3	7.9						
(0.1, 0.1, 0.15, 0.5)	0.5	28.8	41.0	16.9	25.2	8	3.8	6.8	3	6.0	8.3
	1.0	22.1	32.2	10.9	19.4	16	4.1	7.4	6	5.5	9.5
	1.5	19.2	27.6	7.2	15.7	24	4.8	8.0	9	5.6	9.4
	LKN			2.7	7.9						

based on the considered approximations to the null distribution may have different powers. In this case, the test based on the residual-based bootstrap approximation is a bit more powerful than the one based on the weighted bootstrap approximation. In all tried cases, the new test outperforms the portmanteau-type test and the one in Hong and Lee (2003).

We also compared both approximations in terms of the required CPU. Table 7 shows the average CPU time in seconds to get one p -value for testing $\mathcal{H}_{0,\text{lin}}$ for $T = 500, 1000$, by generating $B = 1000$ resamples in each case. The quantities in the table clearly show the computational efficiency of the weighted bootstrap in comparison with the residual-based bootstrap. This advantage of the weighted bootstrap over the residual-based bootstrap comes from the fact that the former does not re-estimate the GARCH parameters and the residuals at each iteration.

6.2 Real data fits to GARCH and duration models

This subsection reports the results of three real data set applications.

Table 4 Percentages of rejections under $\mathcal{H}_{0,\text{alog}}$ for $T = 500$ (level)

$(\omega, \gamma_+, \gamma_-, \beta)$	Proposed test					Portm.			HL		
	a	WB		RB		k	5%	10%	p	5%	10%
		5%	10%	5%	10%						
(0.1, 0.1, 0.15, 0.3)	0.5	4.5	9.0	5.6	10.2	8	3.4	5.9	3	3.5	5.9
	1.0	5.5	9.5	5.9	10.3	16	5.1	8.0	6	3.1	6.0
	1.0	5.5	10.0	6.1	10.4	24	5.8	10.4	9	2.8	6.4
(0.1, 0.1, 0.15, 0.5)	0.5	4.3	9.2	6.1	10.3	8	3.5	6.0	3	3.4	6.0
	1.0	5.1	9.6	6.3	10.5	16	4.8	7.7	6	3.4	6.3
	1.5	5.7	10.2	5.9	11.2	24	5.7	10.2	9	3.1	7.2
(0.1, 0.1, 0.15, 0.7)	0.5	3.2	6.6	4.5	9.7	8	3.7	6.6	3	3.7	6.2
	1.0	4.2	9.0	5.4	10.1	16	4.6	7.1	3	3.5	6.3
	1.5	4.5	9.4	5.5	10.2	24	6.2	10.4	3	3.1	6.7
(0.1, 0.1, 0.15, 0.8)	0.5	2.9	7.2	3.3	8.1	8	3.9	7.1	3	3.6	6.0
	1.0	3.5	8.7	4.1	9.6	16	4.7	8.3	6	3.3	6.2
	1.5	4.1	9.4	4.6	8.9	24	6.4	10.1	9	2.9	6.3

Table 5 Percentages of rejections under $\mathcal{H}_{0,\text{alog}}$ for $T = 3300$ with the weighted bootstrap approximation (level)

$(\omega, \gamma_+, \gamma_-, \beta)$	α	a		
		0.5	1.0	1.5
(0.04, 0.03, 0.02, 0.96)	5%	5.0	4.8	4.6
	10%	9.0	9.4	9.0

Table 6 Percentages of rejections under alternatives for testing $\mathcal{H}_{0,\text{alog}}$ for $T = 500$ (power)

$(\omega, \gamma, \beta, \delta)$	Proposed test					Portm.			HL		
	a	WB		RB		k	5%	10%	p	5%	10%
		5%	10%	5%	10%						
(0.1, 0.1, 0.3, 0.02)	0.5	11.1	18.6	15.1	24.2	8	9.0	14.1	3	4.9	6.7
	1.0	10.6	17.8	15.2	24.3	16	8.8	13.4	6	3.6	7.0
	1.5	10.0	17.6	15.3	23.3	24	9.3	14.6	9	3.3	6.2
(0.1, 0.1, 0.3, 0.05)	0.5	29.8	42.9	36.3	48.3	8	9.4	14.3	3	7.8	12.0
	1.0	31.1	43.0	39.3	49.9	16	8.7	14.6	6	6.2	10.0
	1.5	30.6	42.9	39.9	50.6	24	9.9	15.1	9	4.8	9.2
(0.1, 0.1, 0.3, 0.10)	0.5	80.4	88.0	85.4	90.4	8	9.5	14.9	3	29.7	36.3
	1.0	81.1	89.0	87.3	91.7	16	8.9	15.3	6	23.0	29.6
	1.5	82.3	89.3	87.5	92.4	24	9.5	16.2	9	18.2	25.1

Table 7 CPU time consumed for the calculation of one p -value (in seconds)

	WB		RB	
T	500	1000	500	1000
Time	1.37	5.96	106.61	400.09

6.2.1 S&P500 daily stock index

First, we applied our methodology to the extensively studied S&P500 daily stock index. Specifically, we considered two periods: the first from January 2, 1990, to December 31, 1993, and the second from January 2, 1998, to August 28, 2002. Figures 1 and 2 in Supplementary Material display the original time series and the time series of the log returns for each period. The number of observations in these periods is 1013 and 1170, respectively. The reason for choosing such periods is that, according to the results obtained in Hlávka et al. (2017), the S&P500 return series associated with each period is a martingale difference sequence. So we could go a step beyond and test for some GARCH-type specification.

We tested the null hypothesis $\mathcal{H}_{0,\text{lin}}$. The QMLE of the parameters are: $(\hat{\omega}, \hat{\gamma}, \hat{\beta}) = (7.0185 \times 10^{-9}, 1.7753 \times 10^{-2}, 9.8129 \times 10^{-1})$, for the first period, and $(\hat{\omega}, \hat{\gamma}, \hat{\beta}) = (7.1723 \times 10^{-6}, 9.4489 \times 10^{-2}, 8.6811 \times 10^{-1})$, for the second one. Since from the simulations we know that for the sample size at hand, the resulting test is a bit liberal under persistent models, we refrain from applying the weighted bootstrap approximation to calculate the p -value for the fitted GARCH(1,1) model of the first period. The residual-based bootstrap p -values for the proposed test and the test in Leucht et al. (2015) are 0.014 (for $a = 0.5, 1.0$ and 1.5) and 0.080, respectively, showing that the GARCH(1,1) specification for the conditional variance is not adequate for the first period. For the second period, we got the weighted bootstrap p -values: 0.581, 0.564, 0.544, and the residual-based bootstrap p -values: 0.302, 0.334 and 0.350 for $a = 0.5, 1.0$ and 1.5 , respectively, and 0.204 for the test in Leucht et al. (2015). No test rejects $\mathcal{H}_{0,\text{lin}}$ for the second period. A visual inspection of Figure 1 in Supplementary Material suggests a possibility of dispersion change for the first period, whereas Figure 2 in Supplementary Material shows more stability for the second period.

6.2.2 Daily exchange rates

We consider returns series of the daily exchange rates of the American Dollar (USD), the Japanese Yen (JPY), the British Pound (GBP), the Swiss Franc (CHF) and Canadian Dollar (CAD) with respect to the Euro, from January 5, 1999, to January 18, 2012, which corresponds to 3344 observations. The data were obtained from the web site <http://www.ecb.int/stats/exchange/eurofxref/html/index.en.html>. These data were analyzed in Francq et al. (2013), where the authors fitted a log-GARCH(1,1) model to each exchange rate. Here, we test for the adequacy of such a model as well as of the (linear) GARCH(1,1). Table 8 displays the parameter estimators for each model and each exchange rate (although Table 1 in Francq et al. (2013) displays the fitted parameters for the log-GARCH(1,1) model, we decided to include them here for the

Table 8 Fitted parameter values for the GARCH(1,1) and the log-GARCH(1,1) models on daily returns of exchange rates and weighted bootstrap *p*-values for testing goodness-of-fit to these models

GARCH(1,1)						
	$\hat{\omega}$	$\hat{\gamma}$	$\hat{\beta}$	<i>p</i> -value		
				<i>a</i> = 0.5	<i>a</i> = 1.0	<i>a</i> = 1.5
USD	0.0016	0.0304	0.9663	0.0045	0.0060	0.0120
JPY	0.0032	0.0601	0.9363	0.0000	0.0000	0.0000
GBP	0.0015	0.0537	0.9416	0.2370	0.2885	0.3070
CHF	0.0008	0.0914	0.9050	0.0915	0.0960	0.1305
CAD	0.0037	0.0320	0.9599	0.2800	0.1095	0.0705

log-GARCH(1,1)							
	$\hat{\omega}$	$\hat{\gamma}_+$	$\hat{\gamma}_-$	$\hat{\beta}$	<i>p</i> -value		
					<i>a</i> = 0.5	<i>a</i> = 1.0	<i>a</i> = 1.5
USD	0.0242	0.0271	0.0160	0.9705	0.0595	0.0775	0.1060
JPY	0.0506	0.0372	0.0424	0.9517	0.0015	0.0025	0.0025
GBP	0.0316	0.0299	0.0290	0.9643	0.5070	0.2540	0.1785
CHF	0.0574	0.0461	0.0355	0.9543	0.0000	0.0010	0.0010
CAD	0.0210	0.0246	0.0175	0.9685	0.0440	0.0170	0.0105

sake of completeness) and the weighted bootstrap *p*-values for testing goodness-of-fit for each model based on 2000 replications. Looking at the results in this table, we conclude that the log-GARCH(1,1) provides an adequate fit only for the USD and the GBP series. On the other hand, the GARCH(1,1) model is adequate for the GBP, the CHF and the CAD series, while no model provides a good description of the JPY data.

6.2.3 Exxon price duration

In our third application, we use NYSE price duration data to assess the performance of the linear ACD(1,1) model

$$\psi_t = \omega + \gamma y_{t-1} + \beta \psi_{t-1},$$

to the adjusted Exxon price durations, which comprises $n = 2716$ observations (see Fernandes and Gramming (2005) for a detailed description of the data). Figure 3 in Supplementary Material displays duration plot and histogram. Fernandes and Gramming (2005) and Meintanis et al. (2018) tested goodness-of-fit for the innovation distribution by assuming a linear ACD(1,1) model. In estimating the parameters, we employed the QMLE with a unit exponential distribution innovation distribution. The resulting estimates are $(\hat{\omega}, \hat{\gamma}, \hat{\beta}) = (0.0695, 0.0507, 0.8801)$, yielding weighted bootstrap *p*-values: 0.453, 0.544 and 0.545, and the corresponding residual-based bootstrap *p*-values: 1.0, 1.0 and 1.0 for $a = 0.5, 1.0$ and 1.5, respectively, indicating that a linear ACD(1,1) model cannot be rejected.

7 Conclusion

We suggest a goodness-of-fit test for general conditional moment specification models of arbitrary (but finite) memory, which is consistent against any deviation from the null model and which may be readily computed without approximate numerical techniques. The new procedure was implemented by suitable bootstrap resampling and found to perform well in finite samples for GARCH models as well as for conditional duration models. Thus, our approach suits well practitioners' purposes since it can be readily be adapted to a wide range of GARCH-type models with applications in diverse fields.

Although not handled here, our method can be readily extended to Poisson autoregressive models for time series of counts, wherein the conditional mean-adjusted observations play the same role as the error terms in GARCH models. An entirely different question is whether the present method can be adapted to test goodness-of-fit to continuous volatility models such as the models tested in Monsalve-Cobis et al. (2011) and González-Manteiga et al. (2017). This issue, however, requires separate investigation.

Acknowledgements The authors are grateful to the anonymous referees for their constructive and useful comments. MD Jiménez-Gamero has been financially supported by Grant MTM2017-89422-P (Ministerio de Economía, Industria y Competitividad, Spain, and ERDF). S Lee has been supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT No. 2018R1A2A2A05019433. Part of the research of SG Meintanis was carried out at the Department of Economics and was supported by Grant Number 11699 of the Special Account for Research Grants (EAKE) of the National and Kapodistrian University of Athens.

References

- Bauwens L, Giot P (2003) Asymmetric ACD models: introducing price information in ACD models. *Empir Econ* 28:709–731
- Berkes I, Horváth L, Kokoszka PS (2003a) Asymptotics for garch squared residual correlations. *Econom Theor* 19:515–540
- Berkes I, Horváth L, Kokoszka PS (2003b) GARCH processes: structure and estimation. *Bernoulli* 9:201–227
- Bierens HJ (1982) Consistent model specification tests. *J Econom* 20:105–134
- Davis RA, Matsui M, Mikosch T, Wan P (2018) Applications of distance correlation to time series. *Bernoulli* 24:3087–3116
- Dufour A, Engle R (2000) Time and the price impact of a trade. *J Finance* 55:2467–2498
- Engle R, Russell J (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66:1127–1162
- Escanciano JC (2008) Joint and marginal specification tests for conditional mean and variance models. *J Econom* 143:74–87
- Fernandes M, Gramming J (2005) Nonparametric specification tests for conditional duration models. *J Econom* 127:16–38
- Fokianos K, Pitsillou M (2018) Testing independence for multivariate time series via the auto-distance correlation matrix. *Biometrika* 105:337–352
- Françq C, Zakoïan JM (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10:605–637
- Françq C, Zakoïan JM (2010) GARCH models: structure, statistical inference and financial applications. Wiley, New York
- Françq C, Winterberger O, Zakoïan JM (2013) GARCH model without positivity constraints: exponential or log GARCH? *J Econom* 177:34–46

- Franco C, Winterberger O, Zakoian JM (2018) Goodness-of-fit tests for Log-GARCH and EGARCH models. *Test* 27:27–51
- Ghousi K, Rémillard B (2014) Comparison of specification tests for GARCH models. *Comput Stat Data Anal* 76:291–300
- Ghousi K, Rémillard B (2018) Serial independence tests for innovations of conditional mean and variance models. *Test* 27:3–26
- González-Manteiga W, Zubelli JP, Monsalve-Cobis A, Febrero-Bande M (2017) Goodness-of-fit test for stochastic volatility models. In: Ferger D et al (eds) *From statistics to mathematic finance. Festschrift in Honour of Winfried Stute*. Springer, New York, pp 89–104
- Halunga AG, Orme CD (2009) First-order asymptotic theory for parametric misspecification tests of GARCH models. *Econom Theor* 25:364–410
- Hlávka Z, Hušková M, Kirch C, Meintanis SG (2017) Fourier-type tests involving martingale difference processes. *Econom Rev* 36:468–492
- Hong Y (1999) Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *J Am Stat Assoc* 94:1201–1220
- Hong Y, Lee TH (2003) Diagnostic checking for the adequacy of nonlinear time series models. *Econom Theor* 19:1065–1121
- Jeong M (2017) Residual-based GARCH bootstrap and second order asymptotic refinement. *Econom Theor* 33:779–790
- Jiménez-Gamero MD, Pardo-Fernández JC (2017) Empirical characteristic function tests for GARCH innovation distribution using multipliers. *J Stat Comput Simul* 87:2069–2093
- Jondeau E, Poon SH, Rockinger M (2007) *Financial modeling under non-Gaussian distributions*. Springer, London
- Leucht A, Kreiss JP, Neumann MH (2015) A model specification test for GARCH(1,1) processes. *Scand J Stat* 42:1167–1193
- Meintanis SG, Milošević B, Obradović M (2018) Goodness-of-fit tests in conditional duration models. *Stat Pap*. <https://doi.org/10.1007/s00362-017-0930-8>
- Monsalve-Cobis A, González-Manteiga W, Febrero-Bande M (2011) Goodness-of-fit test for interest rate models: an approach based on empirical processes. *Comput Stat Data Anal* 55:3073–3092
- Nolan JP (2013) Multivariate elliptically contoured stable distribution: theory and estimation. *Comput Stat* 28:2067–2089
- R Core Team (2018) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 7 Nov 2018
- Rachev S, Mittnik S (2000) *Stable Paretian models in finance*. Wiley, New York
- Stinchcombe MB, White H (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econom Theor* 14:295–325
- Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. *Ann Stat* 35:1769–2794
- Teräsvirta T (2009) An introduction to univariate GARCH models. In: Andersen TG et al (eds) *Handbook of financial time series*. Springer, Berlin, pp 17–42
- Zheng Y, Li WK, Li G (2018) A robust goodness-of-fit test for generalized autoregressive conditionally heteroscedastic models. *Biometrika* 105:73–89