Discussion Papers
Department of Economics
University of Copenhagen

No. 21-06

Specification tests for GARCH processes

by

Giuseppe Cavaliere, Indeewara Perera
& Anders Rahbek

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Specification tests for GARCH processes

Giuseppe Cavaliere[*], Indeewara Perera[†] and Anders Rahbek[‡]

May 17, 2021

Abstract

This paper develops tests for the correct specification of the conditional variance function in GARCH models when the true parameter may lie on the boundary of the parameter space. The test statistics considered are of Kolmogorov-Smirnov and Cramér-von Mises type, and are based on a certain empirical process marked by centered squared residuals. The limiting distributions of the test statistics are not free from (unknown) nuisance parameters, and hence critical values cannot be tabulated. A novel bootstrap procedure is proposed to implement the tests; it is shown to be asymptotically valid under general conditions, irrespective of the presence of nuisance parameters on the boundary. The proposed bootstrap approach is based on shrinking of the parameter estimates used to generate the bootstrap sample toward the boundary of the parameter space at a proper rate. It is simple to implement and fast in applications, as the associated test statistics have simple closed form expressions. A simulation study demonstrates that the new tests: (i) have excellent finite sample behaviour in terms of empirical rejection probabilities under the null as well as under the alternative; (ii) provide a useful complement to existing procedures based on Ljung-Box type approaches. Two data examples are considered to illustrate the tests.

Keywords: GARCH model; Bootstrap; Specification test; Kolmogorov-Smirnov test; Cramér-von Mises test; Marked empirical process; Nuisance parameters on the boundary.

[*]Department of Economics, University of Exeter, UK and Department of Economics, University of Bologna, Italy (Email: giuseppe.cavaliere@unibo.it)

[†]Department of Economics, The University of Sheffield, UK (Email: i.perera@sheffield.ac.uk)

[‡]Department of Economics, University of Copenhagen, Denmark (Email: anders.rahbek@econ.ku.dk)

This research was supported by the Danish Council for Independent Research (DSF Grant 015-00028B) and by the Italian Ministry of University and Research (PRIN 2017 Grant 2017TA7TYC).

1 Introduction

Generalized autoregressive conditionally heteroskedastic (GARCH) models introduced by Bollerslev (1986) are widely used for modelling various financial time series processes. The data generation mechanism of a GARCH model requires the conditional variance to be always strictly positive, which is generally obtained by imposing a strictly positive intercept and non-negative GARCH coefficients in the conditional variance equation. Consequently, in GARCH models, the admissible parameter space typically needs to be inequality restricted. This represents an important difference between GARCH and other popular time series models, such as AR and ARMA models. Although omnibus specification testing in GARCH type models against unspecified alternatives has attracted considerable attention in the recent literature, a crucial weakness in the current theory remains the exclusion of the presence of nuisance parameters on the boundary. This paper contributes towards addressing this issue by developing new statistical methodology for specification testing in GARCH models.

There are a number of different GARCH models available in the literature and many of them are nonnested models (see Francq and Zakoïan, 2010). Therefore, in many cases, a sensible way to proceed when testing a specification of a GARCH model is to leave the alternative model unspecified, or to test the lack-of-fit. This type of tests, also known as omnibus tests, have their roots in the seminal work of Kolmogorov (1933) on testing for a specific probability distribution function, and Grenander and Rosenblatt (1957) on testing the hypothesis of white noise dependence. Several omnibus specification tests in GARCH type models have been proposed in the literature. These include tests based on weighted empirical processes of standardized residuals (Koul and Ling, 2006; Escanciano, 2010), spectral distributions based tests in the frequency domain (Hidalgo and Zaffaroni, 2007; Escanciano, 2008), residual based tests for nonnegative valued processes (Fernandes and Grammig, 2005; Koul *et al.*, 2012; Perera *et al.*, 2016), and Khmaladze type (Khmaladze, 1981) martingale transformations based tests (Bai 2003; Perera and Koul 2017), amongst others.

A key regularity condition imposed by the aforementioned specification tests is to restrict the true parameter to the interior of the null parameter space. Since the parameter space of a GARCH-type model is inequality restricted, this condition is not typically satisfied if some ARCH or GARCH coefficients are zero, because then the true parameter may lie on the boundary of the parameter space. Therefore, for the theory developed in the above cited papers, the true parameter being an interior point is essential; for example, the limiting process obtained in Theorem 2.1 in Hidalgo and Zaffaroni (2007) would not be Gaussian if, for instance, a GARCH(p, q) model is estimated when the underlying true process is a GARCH$(p-1, q)$, or a GARCH$(p, q-1)$ process. Similarly, the asymptotic properties of the other aforementioned papers would also not hold when some nuisance parameters lie on the boundary.

In this paper we contribute towards the literature of specification testing in GARCH models by developing a new class of tests for the correct specification of the conditional variance function while allowing the null model to have an unknown number of nuisance parameters on the boundary of the parameter space. Our test statistics are functionals of an empirical process marked by centered squared residuals and are easy to compute. The limiting distributions of the test statistics are not free from (unknown) nuisance parameters, and hence critical values cannot be tabulated for general use. We propose a bootstrap method to implement the tests and show that it is asymptotically valid under general conditions, irrespective of the presence of nuisance parameters on the boundary. The proposed bootstrap approach is simple to implement, and is based on a method of shrinkage of the parameter estimates used to generate the bootstrap sample toward the boundary of the parameter space at an appropriate rate. This approach is similar to the related bootstrap scheme advocated in Cavaliere *et al.* (2021), in a different context, for bootstrapping likelihood ratio statistics, and it also has its roots in the modified bootstrap approach considered in Chatterjee and Lahiri (2011) for bootstrapping Lasso-type estimators. Our bootstrap tests are shown to be consistent against fixed alternatives. We also separately consider the case the nuisance parameters lie in the interior of the parameter space for Kolmogorov-Smirnov and Cramér-von Mises type tests based on the aforementioned marked empirical process, and show that the bootstrap implementations of these tests under standard residual based bootstrap are asymptotically valid and consistent. Our tests can be implemented easily because the test statistics have simple closed form expressions. A simulation study shows that the proposed tests have desirable finite sample properties. We illustrate the testing procedure by considering two real data examples.

The rest of this paper is structured as follows. Section 2 formulates the problem, defines the estimators and test statistics. Section 3 provides the results relating to the asymptotic validity and consistency of the bootstrap tests when the parameters are in the interior of the parameter space. Section 4 considers inference when some components of the true parameter lie on the boundary of the parameter space. Section 5 describes a simulation study. Two empirical illustrations are discussed in Section 6. Section 7 concludes the paper. The proofs and some assumptions are relegated to Appendix A.

2 Formulation of the Problem

Let $(Y_1, Y_2, \ldots, Y_n)$ be a realization of an observable stationary process $\{Y_i\}$ satisfying

$$Y_i = h_i^{1/2}\varepsilon_i, \quad i \in \mathbb{Z} := \{0, \pm 1, \pm 2, \cdots\}, \tag{1}$$

where the errors ε_i, $i \in \mathbb{Z}$, are independent and identically distributed (i.i.d.) random variables (r.v.'s) having zero mean and unit variance with common cumulative distribution function (c.d.f.) F_0, and $h_i = \mathrm{E}[Y_i^2 \mid \mathcal{H}_{i-1}]$, where $\mathcal{H}_{i-1}$ denotes the information available up to time $i-1$ for forecasting Y_i, $i \in \mathbb{Z}$.

As is well-known, a GARCH(p_1, p_2) model for h_i takes the form

$$h_i = h_i(\boldsymbol{\phi}) = \omega + \sum_{j=1}^{p_1} \alpha_j Y_{i-j}^2 + \sum_{k=1}^{p_2} \beta_k h_{i-k}(\boldsymbol{\phi}), \quad i \in \mathbb{Z}; \tag{2}$$

the vector of parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{p_1+p_2+1})' = (\omega, \alpha_1, \ldots, \alpha_{p_1}, \beta_1, \ldots, \beta_{p_2})'$, usually belongs to a compact parameter space

$$\Phi \subset (0, \infty) \times [0, \infty)^{p_1+p_2} \tag{3}$$

with $\omega > 0$, $\alpha_k \geq 0\,(k = 1, \ldots, p_1)$, $\beta_k \geq 0\,(k = 1, \ldots, p_2)$, and in order to avoid well-known identification issues (see also Assumption (A3) below) one typically imposes $\sum_{k=1}^{p_1} \alpha_k \neq 0$.

Suppose we wish to test the adequacy of the above GARCH(p_1, p_2) model for h_i, i.e., to test the null hypothesis

$$\begin{aligned} \mathsf{H}_0: \quad & h_i = h_i(\boldsymbol{\phi}_0) = \omega_0 + \sum_{j=1}^{p_1} \alpha_{0j} Y_{i-j}^2 + \sum_{k=1}^{p_2} \beta_{0k} h_{i-k}(\boldsymbol{\phi}_0), \text{ a.s. for all } i, \text{ and} \\ & \text{for some } \boldsymbol{\phi}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p_1}, \beta_{01}, \ldots, \beta_{0p_2})' \in \Phi, \end{aligned} \tag{4}$$

against the alternative $\mathsf{H}_1 : \mathsf{H}_0$ *is not true.*

Since some ARCH or GARCH coefficients may be zero, the null model (4) allows some components of $\boldsymbol{\phi}_0$ to be on the boundary of the parameter space Φ.

Let $\hat{\boldsymbol{\phi}}$ denote the *Gaussian quasi maximum likelihood estimator* [QMLE] defined by

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi} \in \Phi} \sum_{i=1}^{n} \ell_i(\boldsymbol{\phi}), \quad \ell_i(\boldsymbol{\phi}) = \log h_i(\boldsymbol{\phi}) + [Y_i^2 / h_i(\boldsymbol{\phi})], \tag{5}$$

with $h_i(\boldsymbol{\phi})$ being defined recursively by (2) for $i = 1, 2, \ldots, n$. To simplify the exposition, the vector of initial values, $\varsigma_0 = (Y_0, \ldots, Y_{1-p_1}, h_0, \ldots, h_{1-p_2})' \in \mathbb{R}^{p_1} \times [0, \infty)^{p_2}$, is assumed to be fixed for the statistical analysis. The asymptotic results do not change if ς_0 is replaced by an arbitrarily chosen vector (e.g., by setting $Y_t = 0$ and $h_t = 0$, all $t \leq 0$); see, for example, the discussions in Straumann and Mikosch (2006), Perera and Silvapulle (2021) and Jensen and Rahbek (2004).

Let $\widehat{\varepsilon}_i := Y_i / \{h_i(\hat{\boldsymbol{\phi}})\}^{1/2}$, $i = 1, \ldots, n$, denote the estimated residuals. With $\boldsymbol{\phi} \in \Phi$, we propose an omnibus test statistic based on the marked empirical process:

$$\mathcal{U}_n(y, \boldsymbol{\phi}) := n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{Y_i^2}{h_i(\boldsymbol{\phi})} - 1 \right\} \mathbb{I}(Y_{i-1} \leq y), \quad y \in \mathbb{R},\, \boldsymbol{\phi} \in \Phi, \tag{6}$$

where $\mathbb{I}$ denotes the indicator function. We allow the domain of $\mathcal{U}_n(\cdot,\boldsymbol{\phi})$ to extend over the whole real line by letting $\mathcal{U}_n(-\infty,\boldsymbol{\phi}) := 0$ and $\mathcal{U}_n(\infty,\boldsymbol{\phi}) := n^{-1/2}\sum_{i=1}^{n}\{Y_i^2/h_i(\boldsymbol{\phi})-1\}$. Hence, $\mathcal{U}_n(\cdot,\boldsymbol{\phi})$ in (6) can be viewed as a process in the space of *cadlag* functions on $[-\infty,\infty]$, equipped with the uniform metric, which we denote by $\mathcal{D}(\mathbb{R})$. This process is an extension of the so-called cumulative sum process for the one sample setting to the current set up. Under H_0, $\mathrm{E}\mathcal{U}_n(y,\boldsymbol{\phi}_0) = 0$, for all y, but not under H_1. Hence, if H_0 is true, then we would expect $\mathcal{U}_n(y,\hat{\boldsymbol{\phi}})$ to be close to zero for all y, but not otherwise. Therefore, a suitable functional of $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ can potentially be used as a test statistic for testing H_0 against H_1.

The use of cumulative sum processes for specification testing similar to $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ goes back to von Neumann (1941), who proposed a test of constant regression based on an analog of this process. A motivation for basing inference in nonnegative valued processes on an analog of the process $\mathcal{U}_n(\cdot,\boldsymbol{\phi})$, $\boldsymbol{\phi}\in\Phi$, also appears in Koul *et al.* (2012). Similar tests have also been considered by Stute (1997) and Koul and Stute (1999) for certain regression and additive time series models. More recently, analogs of $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ have been used by several authors to propose asymptotically distribution free specification tests in related time series models; see, for example, Perera and Koul (2017) and Balakrishna *et al.* (2019). In the econometric analyses presented in these papers certain tests based on analogs of $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ have demonstrated desirable finite sample and asymptotic properties. Therefore, we find it of interest to develop specification tests based on similar statistics involving the process $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ for testing H_0 against H_1 in the current setup. In particular, we consider the Kolomogorov-Smirnov (KS) and Cramér-von Mises (CvM) type statistics which can be defined in terms of $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ as:

$$T_1 := \mathrm{KS} = \sup_y \left|\mathcal{U}_n(y,\hat{\boldsymbol{\phi}})\right|, \quad T_2 := \mathrm{CvM} = \int \mathcal{U}_n^2(y,\hat{\boldsymbol{\phi}})dG_n(y), \tag{7}$$

where $G_n(y) := n^{-1}\sum_{i=1}^{n}\mathbb{I}(Y_{i-1}\le y)$. Other suitable functionals of $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ may also be considered as possible test statistics (see D'Agostino and Stephens, 1986).

3 Inference when the parameters are in the interior of the parameter space

Before moving to the general case which includes possible parameters on the boundary of the parameter space, we here consider the case the true parameter $\boldsymbol{\phi}_0$ is in the interior of Φ.

The asymptotic distribution of $\mathcal{U}_n(\cdot,\boldsymbol{\phi}_0)$ under the null hypothesis H_0 can be derived by standard arguments, under the assumptions on the GARCH process discussed in the next subsection. Specifically, from a martingale central limit theorem [for example Hall and Heyde (1980), Corollary 3.1] and the Cramér-Wold device it follows that all finite dimensional distributions of $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ converge weakly to a multivariate normal distribution with mean vector zero and covariance matrix given by the covariance function

$$K(x, y) := \mathrm{E}(\varepsilon_i^2 - 1)^2 \mathbb{I}(Y_{i-1} \leq x \wedge y) = (\kappa_\varepsilon - 1) G(x \wedge y), \quad x, y \in \mathbb{R}, \tag{8}$$

where G denotes the (unconditional) distribution function (d.f.) of Y_0, $\kappa_\varepsilon := \mathrm{E}\varepsilon_i^4 < \infty$ and $x \wedge y = \min(x, y)$. Under H_0, G may depend on $\boldsymbol{\phi}_0$, but we do not exhibit this dependence. Then, since the function $\pi(x) := K(x, x) = (\kappa_\varepsilon - 1)G(x)$ is nondecreasing and nonnegative, tightness of the process $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ follows by e.g. Theorem 15.7 in Billingsley (1968), and therefore, under H_0, $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ converges weakly to the time-transformed Brownian motion $B \circ \pi$, in the space $\mathcal{D}(\mathbb{R})$ equipped with the uniform metric.

However, since $\boldsymbol{\phi}_0$ is replaced by $\hat{\boldsymbol{\phi}}$, the weak limit of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$ will not be of the form $B \circ \pi$; rather, it depends on $(\boldsymbol{\phi}_0, G)$. We derive this result in the next subsection, where weak convergence of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$ is derived for the case where the true value $\boldsymbol{\phi}_0$ lies in the interior of the parameter space.

3.1 Asymptotics for the original test statistics

First we introduce some notation to facilitate the presentation of the underlying assumptions for the asymptotic results. Let $\mathcal{A}_{\boldsymbol{\phi}}(z) = \sum_{i=1}^{p_1} \alpha_i z^i$ and $\mathcal{B}_{\boldsymbol{\phi}}(z) = 1 - \sum_{i=1}^{p_2} \beta_i z^i$ with $\mathcal{A}_{\boldsymbol{\phi}}(z) = 0$ if $p_1 = 0$ and $\mathcal{B}_{\boldsymbol{\phi}}(z) = 1$ if $p_2 = 0$. Furthermore, let

$$A_{0i} = \begin{pmatrix} \alpha_{01}\varepsilon_i^2 & \cdots & \alpha_{0p_1}\varepsilon_i^2 & \beta_{01}\varepsilon_i^2 & \cdots & \beta_{0p_2}\varepsilon_i^2 \\ & \boldsymbol{I}_{p_1-1} & \mathbf{0} & & \mathbf{0} & \\ \alpha_{01} & \cdots & \alpha_{0p_1} & \beta_{01} & \cdots & \beta_{0p_2} \\ & \mathbf{0} & & & \boldsymbol{I}_{p_2-1} & \mathbf{0} \end{pmatrix}, \quad i \geq 1,$$

with $\boldsymbol{I}_k$ denoting the $k \times k$ identity matrix.

In order to study the limiting behaviour of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$ we make the following assumptions on the process $\{Y_i\}_{i \in \mathbb{Z}}$ which satisfies (1)–(2).

(A1). *The parameter space Φ in equation (3) is a compact subset of $(0, \infty) \times [0, \infty)^{p_1+p_2}$, and contains a hypercube of the form $[\omega_L, \omega_U] \times [0, \epsilon]^{p_1+p_2}$, for some $\epsilon > 0$ and $\omega_U > \omega_L > 0$, which includes the true parameter $\boldsymbol{\phi}_0$.*

(A2). *The sequence of matrices $\boldsymbol{A}_0 = (A_{01}, A_{02}, \ldots)$ has a strictly negative top Lyapunov exponent; i.e., $\gamma(\boldsymbol{A}_0) = \lim_{i\to\infty} i^{-1} \log \|A_{0i}A_{0(i-1)} \ldots A_{01}\| < 0$, and $\sum_{j=1}^{p_2} \beta_j < 1$, $\forall \boldsymbol{\phi} \in \Phi$.*

(A3). *$\mathcal{A}_{\boldsymbol{\phi}_0}(1) \neq 0$, $\alpha_{0p_1} + \beta_{0p_2} \neq 0$, and the polynomials $\mathcal{A}_{\boldsymbol{\phi}_0}(z)$ and $\mathcal{B}_{\boldsymbol{\phi}_0}(z)$ have no common roots if $p_2 > 0$.*

(A4). *The errors* $\varepsilon_i, i \in \mathbb{Z}$, *are i.i.d. with zero mean and unit variance,* ε_i^2 *has a non-degenerate distribution, and* $\mathrm{E}|\varepsilon_i|^{4+d} < \infty$ *for some* $d > 0$.

The condition $\gamma(\boldsymbol{A}_0) < 0$ in (A2) ensures the existence of a unique strictly stationary solution $\{Y_i\}_{i\in\mathbb{Z}}$ to Model (1)–(2); see, e.g. Bougerol and Picard (1992a). Note that, in (A2), the strict stationarity condition $\gamma(\boldsymbol{A}_0) < 0$ is imposed only on the true value $\boldsymbol{\phi}_0$, but for $\boldsymbol{\phi} \neq \boldsymbol{\phi}_0$ we only impose the weaker restriction $\sum_{j=1}^{p_2} \beta_j < 1$. In Assumption (A3), the condition $\mathcal{A}_{\boldsymbol{\phi}_0}(1) \neq 0$ ensures that all the α_{0i} are not zero when $p_1 \neq 0$, and hence we do not allow the strictly stationary solution of (1)–(2) to be a strong white noise process. This in turn allows us to avoid certain identifiability issues when estimating the GARCH parameters with $p_2 \neq 0$ (see Francq and Zakoïan, 2010). Note that, in the ARCH case (i.e. when $p_2 = 0$), the Assumption (A3) is not required. In the general GARCH case when $p_2 > 0$, the Assumption (A3) allows for an overidentification of either the order of the ARCH parameters p_1 or the order of the GARCH parameters p_2, but not both. The condition $\mathrm{E}|\varepsilon_i|^{4+d} < \infty$ in Assumption (A4) is only required for the existence of the variance of the score vector $\partial\ell_i(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi}$; this is necessary for establishing the limiting distribution of the QMLE. Note that we do not assume that the true parameter $\boldsymbol{\phi}_0$ is in the interior of Φ. Thus, the assumptions do not exclude the cases where some α_i or β_j are zero. Assumptions similar to (A1)–(A4) have previously been discussed in the literature for establishing asymptotic properties of the QMLE; see, e.g., Francq and Zakoïan (2010) and Cavaliere *et al.* (2021).

Let

$$J(y, \boldsymbol{\phi}) := \mathrm{E}[\tau_1(\boldsymbol{\phi})\mathbb{I}(Y_0 \leq y)], \quad \tau_i(\boldsymbol{\phi}) := \frac{(\partial/\partial\boldsymbol{\phi})h_i(\boldsymbol{\phi})}{h_i(\boldsymbol{\phi})}, \quad i \in \mathbb{Z}, \boldsymbol{\phi} \in \Phi.$$

The next lemma provides an asymptotic uniform expansion for $\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})$. We make use of this expansion in the proof of establishing the weak convergence of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$.

Lemma 1. *Suppose that Assumptions (A1) and (A4) hold. Then, uniformly in* $y \in \mathbb{R}$,

$$\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}) = \mathcal{U}_n(y, \boldsymbol{\phi}_0) - n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)' J(y, \boldsymbol{\phi}_0) + o_p(1). \tag{9}$$

Unlike the process $\mathcal{U}_n(y, \boldsymbol{\phi}_0)$, the estimated process $\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})$ does not converge weakly to a time transformed Brownian motion, because the term $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)' J(y, \boldsymbol{\phi}_0)$ in (9), is of order $O_p(1)$ and hence is not asymptotically negligible. In fact, if Assumptions (A1)–(A3) are satisfied, then $\hat{\boldsymbol{\phi}}$ converges to $\boldsymbol{\phi}_0$ almost surely (a.s.), and additionally, if Assumption (A4) also holds and $\boldsymbol{\phi}_0$ is an interior point in Φ, then $\hat{\boldsymbol{\phi}}$ is asymptotically linear and satisfies

$$n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = -\Sigma_n^{-1}(\boldsymbol{\phi}_0) n^{-1/2} \sum_{i=1}^{n} (1 - \varepsilon_i^2)\tau_i(\boldsymbol{\phi}_0) + o_p(1), \tag{10}$$

where

$$\Sigma_n(\boldsymbol{\phi}) := n^{-1}\sum_{i=1}^{n}\tau_i(\boldsymbol{\phi})\tau_i(\boldsymbol{\phi})', \quad \tau_i(\boldsymbol{\phi}) := \frac{(\partial/\partial\boldsymbol{\phi})h_i(\boldsymbol{\phi})}{h_i(\boldsymbol{\phi})}, \quad \boldsymbol{\phi}\in\Phi;$$

see, for example, Berkes *et al.* (2003).

By using Lemma 1 and (10), when $\boldsymbol{\phi}_0$ is an interior point in Φ, one can show that $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ converges weakly to a centred Gaussian process. This result is stated in the next theorem.

Theorem 1. *Suppose that (A1)–(A4) are satisfied with $\boldsymbol{\phi}_0$ being an interior point in Φ. Let*

$$M_i(\boldsymbol{\phi}) := -\Sigma^{-1}(\boldsymbol{\phi})(1-\varepsilon_i^2)\tau_i(\boldsymbol{\phi}), \quad \Sigma(\boldsymbol{\phi}) := \mathrm{E}\{\tau_1(\boldsymbol{\phi})\tau_1(\boldsymbol{\phi})'\}, \quad \boldsymbol{\phi}\in\Phi,\ i\in\mathbb{Z}.$$

Then, the process $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ converges weakly to $\mathcal{U}_0$ in $\mathcal{D}(\mathbb{R})$, where $\mathcal{U}_0$ is a centred Gaussian process with covariance kernel

$$\begin{aligned}\mathrm{Cov}\{\mathcal{U}_0(x),\mathcal{U}_0(y)\} &= K(x,y) + J'(x,\boldsymbol{\phi}_0)\mathrm{E}[M_1(\boldsymbol{\phi}_0)M_1'(\boldsymbol{\phi}_0)]J'(y,\boldsymbol{\phi}_0)\\ &\quad -J'(x,\boldsymbol{\phi}_0)\mathrm{E}[(\varepsilon_1^2-1)M_1(\boldsymbol{\phi}_0)\mathbb{I}(Y_0\le y)]\\ &\quad -J'(y,\boldsymbol{\phi}_0)\mathrm{E}[(\varepsilon_1^2-1)M_1(\boldsymbol{\phi}_0)\mathbb{I}(Y_0\le x)],\end{aligned}$$

where $K(x,y)$ is as in (8).

In view of Theorem 1, the limiting distributions of KS and CvM statistics defined in (7) depend on the unknown $(\boldsymbol{\phi}_0, G)$ in a non-trivial way, despite the fact the true parameter is in the interior of Φ. Consequently, it does not appear that it would be possible to find a transformation that would lead to an asymptotically distribution free test, for example as in Bai (2003); Koul *et al.* (2012); Perera and Koul (2017); Escanciano *et al.* (2018). Hence, we proceed by considering bootstrap implementations of the tests.

3.2 Bootstrap implementation

In this section, we propose a bootstrap procedure for computing the critical values for the KS and CvM statistics in (7). We perform the resampling scheme under the null hypothesis and derive the asymptotic properties of the bootstrap statistics, irrespective of whether or not the data generating process satisfies the null hypothesis. To this end, we initially standardize the residuals $\widehat{\varepsilon}_i := Y_i/\{h_i(\hat{\boldsymbol{\phi}})\}^{1/2}$, $i=1,\ldots,n$, as

$$\check{\varepsilon}_i := \left\{n^{-1}\sum_{t=1}^{n}\bar{\varepsilon}_t^2\right\}^{-1/2}\bar{\varepsilon}_i, \quad \bar{\varepsilon}_i := \widehat{\varepsilon}_i - n^{-1}\sum_{t=1}^{n}\widehat{\varepsilon}_t, \quad i=1,\ldots,n, \tag{11}$$

and define the associated empirical distribution function of $\{\check{\varepsilon}_1,\ldots,\check{\varepsilon}_n\}$ as

$$\check{F}_n(x) := n^{-1}\sum_{i=1}^{n}\mathbb{I}(\check{\varepsilon}_i\le x), \quad x\in\mathbb{R}. \tag{12}$$

By construction, $\int_{\mathbb{R}} u\check{F}_n(u)du = 0$ and $\int_{\mathbb{R}} u^2\check{F}_n(u)du = 1$, hence a random variable with distribution function $\check{F}_n$ has zero mean and unit variance, therefore matching the first and second order moments of the error distribution F_0. From Lemma A.1 in Appendix A we obtain that $\check{F}_n$ converges to F_0 with probability one under the null hypothesis.

We next outline the bootstrap algorithm.

Bootstrap algorithm 1

Step 1: Compute $\{\hat{\boldsymbol{\phi}}, T_j\}$ on the original sample $\{Y_1, \dots, Y_n\}$, where T_j is the test statistic defined in (7) ($j = 1, 2$);
Step 2: Compute $\check{\varepsilon}_i$, $i = 1, \dots, n$ as in (11) and draw a random sample (with replacement) of size n, say $\{\varepsilon_1^*, \dots, \varepsilon_n^*\}$, independent of the original data, from the empirical distribution function $\check{F}_n(\cdot)$ in (12);
Step 3: Generate the bootstrap sample $\{Y_1^*, \dots, Y_n^*\}$ with bootstrap true values $(\hat{\boldsymbol{\phi}}, \check{F}_n)$ by

$$Y_i^* = \{h_i^*(\hat{\boldsymbol{\phi}})\}^{1/2}\varepsilon_i^*, \quad h_i^*(\hat{\boldsymbol{\phi}}) = \hat{\omega} + \sum_{j=1}^{p_1} \hat{\alpha}_j (Y_{i-j}^*)^2 + \sum_{k=1}^{p_2} \hat{\beta}_k h_{i-k}^*(\hat{\boldsymbol{\phi}}), \quad i \geq 1$$

initialized with $(Y_0^*, \dots, Y_{1-q}^*, h_0^*(\hat{\boldsymbol{\phi}}), \dots, h_{1-p}^*(\hat{\boldsymbol{\phi}}))' = \varsigma_0$, where ς_0 is an arbitrarily chosen vector (e.g. $Y_t^* = 0$ and $h_t^* = 0$, all $t \leq 0$);
Step 4: Using $\{Y_1^*, \dots, Y_n^*\}$, compute $\hat{\boldsymbol{\phi}}^*$, the bootstrap analog of $\hat{\boldsymbol{\phi}}$;
Step 5: Compute the bootstrap test statistic T_j^* as

$$T_1^* = \mathrm{KS}^* = \sup_y \left|\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*)\right|, \quad T_2^* = \mathrm{CvM}^* = \int \left\{\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*)\right\}^2 dG_n^*(y), \tag{13}$$

where $G_n^*(y)$ and $\mathcal{U}_n^*(y, \boldsymbol{\phi})$ are the bootstrap analogs of $G_n(y)$ and $\mathcal{U}_n(y, \boldsymbol{\phi})$, respectively. The bootstrap p-value is then defined as

$$p_n^* := P_n^*(T_j^* \geq T_j) \tag{14}$$

where P_n^* denotes the probability measure induced by the bootstrap (i.e., conditional on the original data). The bootstrap test corresponds to the decision rule:

$$\text{Reject } \mathsf{H}_0 \text{ at the nominal level } \alpha \text{ if the estimated p-value } p_n^* \text{ is less than } \alpha. \tag{15}$$

As is standard, p_n^* of (14) is unknown. It can be approximated with arbitrary accuracy by repeating steps 2–5 a large number of times, say B, and then setting $p_n^{*(B)}$ to be the fraction of times T_j^* exceeds T_j.

The above bootstrap algorithm is designed to mimic the null data generating process by replacing the unknown $(\boldsymbol{\phi}_0, F_0)$ by the estimators $(\hat{\boldsymbol{\phi}}, \check{F}_n)$. Therefore, to establish the validity of the bootstrap test in (15), we need to generalize the regularity assumptions of Lemma 1 and Theorem 1 allowing an arbitrary true value $(\boldsymbol{\phi}, F)$ in a neighbourhood of $(\boldsymbol{\phi}_0, F_0)$. The required conditions are introduced as Assumptions (B1)–(B5) in Appendix A. Theorem 2 below establishes the asymptotic validity of the bootstrap test (15) under (B2)–(B5).

In the bootstrap setup, we define $G_n^*(y) := n^{-1}\sum_{i=1}^n \mathbb{I}(Y_{i-1}^* \leq y)$, $y \in \mathbb{R}$. Similarly, the bootstrap analogue of the marked empirical process $\mathcal{U}_n(y, \boldsymbol{\phi})$ in (6) is defined by

$$\mathcal{U}_n^*(y, \boldsymbol{\phi}) := n^{-1/2}\sum_{i=1}^n \left\{\frac{(Y_i^*)^2}{h_i^*(\boldsymbol{\phi})} - 1\right\} \mathbb{I}(Y_{i-1}^* \leq y), \quad y \in \mathbb{R}, \boldsymbol{\phi} \in \Phi. \tag{16}$$

Let $O_p^*(1)$, in probability, $o_p^*(1)$, in probability, and E^* denote the usual stochastic orders of magnitude and expectation, respectively, with respect to P_n^* defined above. We denote convergence in distribution of bootstrap statistics as '$\xrightarrow{d^*}$'. That is, '$T_j^* \xrightarrow{d^*} g_j(\mathcal{U}_0)$ *in probability*' means that $P_n^*(T_j^* \leq \cdot) \xrightarrow{p} P\{g_j(\mathcal{U}_0) \leq \cdot\}$, at every continuity point of $P\{g_j(\mathcal{U}_0) \leq \cdot\}$.

Next theorem establishes the asymptotic validity of the bootstrap test (15) under H_0.

Theorem 2. *Suppose that Assumptions (A1)–(A4) and* H_0 *are satisfied and* $\boldsymbol{\phi}_0$ *is an interior point in* Φ. *Additionally, assume that Assumptions (B2)–(B4) hold with* $(\boldsymbol{\phi}_0^*, F_0^*) = (\boldsymbol{\phi}_0, F_0)$. *Let* $\mathcal{U}_0$ *be the limit process appearing in Theorem 1. Then, conditional on* $\{Y_1, \ldots, Y_n\}$,

1. $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ *converges weakly to* $\mathcal{U}_0$, *in probability.*

2. $g\{\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)\} \xrightarrow{d^*} g\{\mathcal{U}_0\}$, *in probability, for any continuous function* $g : \mathcal{D}(\mathbb{R}) \to \mathbb{R}$.

3. *There exists a continuous functional* $g_j : \mathcal{D}(\mathbb{R}) \to \mathbb{R}$ *such that* $T_j^* = g_j\{\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)\} + o_p^*(1)$, *in probability* $(j = 1, 2)$.

In view of Theorem 2, the bootstrap test (15) based on T_j is asymptotically valid under H_0 $(j = 1, 2)$. The next theorem shows that the bootstrap tests are consistent under H_1.

First we need to introduce some notation. Let $(\boldsymbol{\phi}_0^*, F_0^*)$ be the probability limit of $(\hat{\boldsymbol{\phi}}, \check{F}_n)$, such that $\hat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}_0^*$ and $d_2(\check{F}_n, F_0^*) \xrightarrow{p} 0$ as $n \to \infty$, where $d_2(F_X, F_Y)$ is the Mallows metric for the distance between two probability distributions F_X and F_Y (see also Lemma A.1 in Appendix A). Clearly, under the null hypothesis H_0, we have that $(\boldsymbol{\phi}_0^*, F_0^*) = (\boldsymbol{\phi}_0, F_0)$.

Theorem 3. *Suppose that* H_1 *holds. Assume that* $(\hat{\boldsymbol{\phi}}, \check{F}_n)$ *converges in probability to* $(\boldsymbol{\phi}_0^*, F_0^*)$, *the pseudo-true value under* H_1. *Additionally, assume that Assumptions (B1)–(B5) hold,* $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*) = O_p(1)$, *and (A6) holds if some components of* $\boldsymbol{\phi}_0^*$ *are zero. Then, conditional on* $\{Y_1, \ldots, Y_n\}$, *the bootstrap test (15) based on* T_j *has asymptotic power 1* $(j = 1, 2)$.

In view of Theorem 3, our tests have asymptotic power against a given alternative as long as Assumptions (B1)–(B5) hold, (A6) holds if some components of $\boldsymbol{\phi}_0^*$ are zero, and $n^{1/2}(\hat{\boldsymbol{\phi}} -$

$\phi_0^*) = O_p(1)$. Assumption (B1) introduces some regularity conditions in order to ensure the stationarity of the bootstrap data generating process under H_1. Assumptions (B2)–(B4) are the same as in Theorem 2 except that now $(\phi_0^*, F_0^*) \neq (\phi_0, F_0)$, and Assumption (B5) claims that there exists a $y \in \mathbb{R}$ such that $\mathrm{E}[\{h_1/h_1(\phi_0^*)-1\}\mathbb{I}(Y_0 \leq y)] \neq 0$, where $h_i = \mathrm{E}(Y_i^2 \mid \mathcal{H}_{i-1})$, $i \in \mathbb{Z}$. Since h_i is not of the form $h_i(\phi)$ under H_1 and (ϕ_0^*, F_0^*) is the pseudo-true value, the requirement $\mathrm{E}[\{h_1/h_1(\phi_0^*) - 1\}\mathbb{I}(Y_0 \leq y)] \neq 0$ is not very restrictive under H_1. However, in finite samples, the power of the tests can be sensitive to the form of the discrepancy between h_i and $h_i(\hat{\phi})$. More precisely, if $h_i(\hat{\phi})$ is significantly different from h_i such that the magnitude of the process $n^{-1/2}\sum_{i=1}^n\{Y_i^2/h_i(\hat{\phi}) - 1\}\mathbb{I}(Y_{i-1} \leq y)$ is 'large' for some y, then the KS and CvM functionals of $n^{-1/2}\sum_{i=1}^n\{Y_i^2/h_i(\hat{\phi}) - 1\}\mathbb{I}(Y_{i-1} \leq y)$ are likely to be significantly large compared to realizations from the empirical distributions of KS* and CvM*, respectively, leading to finite sample power of the bootstrap tests. Importantly, if the true conditional variance h_i is non-linear while the null parametric form $h_i(\phi)$ is linear, then our tests are likely to have better finite sample power compared to the case where h_i and $h_i(\phi)$ are both linear and the misspecification is only in terms of some missing lags, because in the latter case the KS and CvM functionals of $n^{-1/2}\sum_{i=1}^n\{Y_i^2/h_i(\hat{\phi}) - 1\}\mathbb{I}(Y_{i-1} \leq y)$ are likely to be smaller compared to the former.

For the validity of our bootstrap tests we have so far required the true value ϕ_0 to be an interior point of Φ under H_0. It is of interest to see whether the bootstrap implementation of $T_j (j = 1, 2)$ can be modified to obtain a consistent bootstrap test for the case ϕ_0 lies on the boundary of Φ under H_0. We consider this in the next section.

4 Inference when the true value is on the boundary

Heuristic arguments suggest that T_1 and T_2 in (7) could serve as possible test statistics for testing H_0 against H_1 regardless of whether ϕ_0 lies in the interior or on the boundary of the parameter space. In fact, from Lemma 1, under assumptions (A1) and (A4), we have

$$\mathcal{U}_n(y, \hat{\phi}) = \mathcal{U}_n(y, \phi_0) - n^{1/2}(\hat{\phi} - \phi_0)' J(y, \phi_0) + o_p(1), \tag{17}$$

uniformly in $y \in \mathbb{R}$, irrespective of whether ϕ_0 is in the interior or on the boundary of Φ, with $\mathcal{U}_n(\cdot, \phi_0)$ converging weakly to a time transformed Brownian motion. Therefore, the weak limit of $\mathcal{U}_n(\cdot, \hat{\phi})$, and hence the limiting distributions of T_1 and T_2, depend on the asymptotic behaviour of $n^{1/2}(\hat{\phi} - \phi_0)' J(\cdot, \phi_0)$. Hence, to establish the limiting distributions of the test statistics it is essential to study the large sample properties of $\hat{\phi}$ when ϕ_0 lies on the boundary of the parameter space. Several important results on this have already been obtained by Andrews (2001) and Francq and Zakoian (2007). For the ease of reference, in the next subsection, we summarize some of these results in the notation used in this paper.

4.1 Limiting distributions of the estimators

In this section, we summarize several technical results regarding the asymptotic behaviour of the QMLE $\hat{\boldsymbol{\phi}}$ in (5) when some components of $\boldsymbol{\phi}_0$ are allowed to be zero, and hence $\boldsymbol{\phi}_0$ could be on the boundary of Φ. First, we introduce the following additional regularity condition.

(A5). *$b_j(\boldsymbol{\phi}_0) > 0$ for all $j \geq 0$, where $h_i(\boldsymbol{\phi}_0) = \sum_{j=1}^{\infty} b_j(\boldsymbol{\phi}_0) Y_{i-j}^2$.*

Condition (A5) is equivalent to assuming that the ARCH coefficients being nonzero up to the order of the first GARCH coefficient that is zero. As shown in Berkes *et al.* (2003), a recursive formula may be obtained to compute $b_j(\boldsymbol{\phi})$ for any given j. Further, we have that $b_j(\boldsymbol{\phi}) \to 0$ exponentially fast as $j \to \infty$, uniformly in $\boldsymbol{\phi} \in \Phi$. This means that there exists some $0 < \nu < 1$ such that $\nu^{-j} \sup_{\boldsymbol{\phi}\in\Phi} b_j(\boldsymbol{\phi}) \to 0$ as $j \to \infty$.

Since the parameter $\boldsymbol{\phi}_0$ is allowed to contain zero components, by the assumption that Φ contains a hypercube (see (A1)), the space $n^{1/2}(\Phi - \boldsymbol{\phi}_0)$ increases to the convex cone

$$\Lambda = \Lambda(\boldsymbol{\phi}_0) = \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_{p_1+p_2+1},$$

where $\Lambda_1 = \mathbb{R}$, and for each $i = 2, \ldots, p_1 + p_2 + 1$, denoting $\boldsymbol{\phi}_0 = (\boldsymbol{\phi}_{01}, \ldots, \boldsymbol{\phi}_{0(1+p_1+p_2)})'$, $\Lambda_i = \mathbb{R}$ if $\boldsymbol{\phi}_{0i} \neq 0$ and $\Lambda_i = [0, \infty)$ if $\boldsymbol{\phi}_{0i} = 0$. Next lemma shows that, under (A1)–(A5), the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ can be represented as the projection of a normal vector distribution onto Λ; for further details on the nature of this projection, see Section 4 in Francq and Zakoian (2007).

Lemma 2. *Suppose that Assumptions (A1)-(A3) are satisfied. Then, $\hat{\boldsymbol{\phi}} \xrightarrow{a.s.} \boldsymbol{\phi}_0$, as $n \to \infty$. Additionally, assume that Assumptions (A4) and (A5) are also satisfied. Then,*

$$n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \xrightarrow{d} \lambda^{\Lambda} := \arg\inf_{\lambda\in\Lambda} (\lambda - Z)'\Sigma(\boldsymbol{\phi}_0)(\lambda - Z),$$

where $Z \sim \mathcal{N}(0, (\kappa_\varepsilon - 1)\Sigma^{-1}(\boldsymbol{\phi}_0))$, $\Sigma(\boldsymbol{\phi}) := \mathrm{E}\{\tau_1(\boldsymbol{\phi})\tau_1(\boldsymbol{\phi})'\}$, $\boldsymbol{\phi} \in \Phi$.

The proof of Lemma 2 follows from Francq and Zakoian (2007). If $\boldsymbol{\phi}_0$ is an interior point, then $\Lambda = \mathbb{R}^{p_1+p_2+1}$ and $\lambda^{\Lambda} = Z \sim \mathcal{N}(0, (\kappa_\varepsilon - 1)\Sigma^{-1}(\boldsymbol{\phi}_0))$, which is the same as the classical case (e.g., see Berkes and Horváth, 2004) as we also considered in the previous section.

4.2 Inconsistency of the standard bootstrap test with parameters on the boundary

The bootstrap true parameter value, say $\boldsymbol{\phi}_n^*$, plays a crucial role in defining the properties of any bootstrap test. For the standard bootstrap test in Section 3.2 we set $\boldsymbol{\phi}_n^*$ equal to $\hat{\boldsymbol{\phi}}$. In the proof of Theorem 2, under Assumptions (A1)–(A4) and (B2)–(B4), we obtain that the limiting behaviour of $n^{1/2}(\hat{\boldsymbol{\phi}}^* - \boldsymbol{\phi}_n^*)$, conditional on $(Y_1, \ldots, Y_n)$, is the same as that of $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$, under H_0, since $\boldsymbol{\phi}_0$ is an interior point and $\boldsymbol{\phi}_n^* = \hat{\boldsymbol{\phi}}$. This result plays a key role in the proof of establishing the validity of the bootstrap tests for the case the true parameter lies in the interior of the parameter space. Convergence results of this type have also been used in establishing the asymptotic validity of other bootstrap methods in similar contexts (see Hidalgo and Zaffaroni, 2007; Perera *et al.*, 2016).

However, in the current setup, the parameter $\boldsymbol{\phi}_0$ is allowed to contain zero components, and hence we require additional conditions to ensure that the bootstrap tests are consistent. In particular, a crucial requirement for the validity of the bootstrap tests is to have the following rate of consistency for the bootstrap true value $\boldsymbol{\phi}_n^* = (\boldsymbol{\phi}_{n1}^*, \ldots, \boldsymbol{\phi}_{n(1+p_1+p_2)}^*)'$:

$$n^{1/2}(\boldsymbol{\phi}_{ni}^* - \boldsymbol{\phi}_{0i}) = \begin{cases} o_p(1), & \text{if } \boldsymbol{\phi}_{0i} = 0 \\ O_p(1), & \text{if } \boldsymbol{\phi}_{0i} > 0 \end{cases}, \quad i = 1, 2, \ldots, 1 + p_1 + p_2. \tag{18}$$

This requirement has previously been introduced in Cavaliere *et al.* (2021) for establishing the validity of a bootstrap based inference procedure in a different context. In the current setup, the requirement (18) ensures that the bootstrap method based on $\boldsymbol{\phi}_n^*$ replicates the unknown limiting distribution of T_j under the null, while being of order $O_p^*(1)$, in probability, under the alternative $(j = 1, 2)$, as is established in Theorems 4 and 5 below. If we set $\boldsymbol{\phi}_n^* = \hat{\boldsymbol{\phi}}$, then it only holds that $n^{1/2}(\boldsymbol{\phi}_{ni}^* - \boldsymbol{\phi}_{0i}) = O_p(1)$ for $i = 1, 2, \ldots, 1 + p_1 + p_2$, and hence (18) is not satisfied. Therefore, the standard bootstrap test outlined in Section 3.2 is not consistent when some parameters lie on the boundary of Φ. Hence, in the next subsection, instead of the standard bootstrap, we propose a new bootstrap method based on using a different mechanism in choosing the bootstrap true values $\boldsymbol{\phi}_{ni}^*$, $i = 1, 2, \ldots, 1 + p_1 + p_2$.

4.3 Consistent bootstrap implementations

In this section we propose a modified bootstrap testing procedure based on shrinking the parameter estimators in the bootstrap data generation. The main idea is that, instead of using $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}_n = (\hat{\boldsymbol{\phi}}_{n1}, \ldots, \hat{\boldsymbol{\phi}}_{n(1+p_1+p_2)})'$ as the true value in the bootstrap data generation, we make use of a transformed version of $\hat{\boldsymbol{\phi}}$, denoted $\hat{\boldsymbol{\phi}}^\dagger = \hat{\boldsymbol{\phi}}_n^\dagger = (\hat{\boldsymbol{\phi}}_{n1}^\dagger, \ldots, \hat{\boldsymbol{\phi}}_{n(1+p_1+p_2)}^\dagger)'$ defined by

$$\hat{\boldsymbol{\phi}}_{ni}^\dagger := \hat{\boldsymbol{\phi}}_{ni} \mathbb{I}(\hat{\boldsymbol{\phi}}_{ni} > c_n) \quad i = 1, 2, \ldots, 1 + p_1 + p_2, \tag{19}$$

where c_n is a scalar sequence converging to zero at an appropriate rate:

$$c_n \to 0, \quad \text{and} \quad n^{1/2} c_n \to \infty \quad \text{as } n \to \infty. \tag{20}$$

This approach has its roots in the Hodges-Le Cam super-efficient type estimators, see e.g. Bickel *et al.* (1998), Chatterjee and Lahiri (2011) and Cavaliere *et al.* (2021).

In view of the parameter restrictions in (3), denoting $\boldsymbol{\phi}_0 = (\boldsymbol{\phi}_{01}, \ldots, \boldsymbol{\phi}_{0(1+p_1+p_2)})'$, we have that $\boldsymbol{\phi}_{01} = \omega_0 > 0$, $\boldsymbol{\phi}_{0i} = \alpha_{0(i-1)} \geq 0$ $(i = 2, \ldots, 1+p_1)$, and $\boldsymbol{\phi}_{0i} = \beta_{0(i-1-p_1)} \geq 0$ $(i = 2+p_1, \ldots, 1+p_1+p_2)$. Thus, $\boldsymbol{\phi}_{01}$ is always in the interior, and $\boldsymbol{\phi}_{0j}$ is on the boundary of the parameter space only if $\boldsymbol{\phi}_{0j} = 0$ for some $j \in \{2, 3, \ldots, 1+p_1+p_2\}$; i.e. some ARCH or GARCH coefficient is zero. Since $\hat{\boldsymbol{\phi}}$ is root-n consistent, the proposed shrinkage in terms of the c_n sequence ensures that $P(\hat{\boldsymbol{\phi}}^{\dagger}_{nj} = 0) \to 1$ as $n \to \infty$ whenever $\operatorname{plim} \hat{\boldsymbol{\phi}}_{nj} = 0$, $j \in \{2, 3, \ldots, 1+p_1+p_2\}$, where 'plim' is the probability limit as $n \to \infty$. Hence, unlike $\hat{\boldsymbol{\phi}}_{nj}$, in large samples, the transformed estimator $\hat{\boldsymbol{\phi}}^{\dagger}_{nj}$ lies on the boundary of the parameter space with large probability whenever $\boldsymbol{\phi}_{0j}$ is on the boundary; i.e. $\boldsymbol{\phi}_{0j} = 0$. Since $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = O_p(1)$ and c_n converges at a rate slower than $n^{-1/2}$, this ensures that the requirement (18) is satisfied by the parameter $\hat{\boldsymbol{\phi}}^{\dagger}_n$ defined by (19)–(20). Hence, as established in Theorems 4 and 5 below, the bootstrap based on $\boldsymbol{\phi}^*_n = \hat{\boldsymbol{\phi}}^{\dagger}_n$ allows us to replicate the unknown limiting distributions of T_1 and T_2 under H_0, while being of order $O^*_p(1)$, in probability, under the alternative.

We next provide a step-by-step guide of the proposed modified bootstrap approach.

Bootstrap algorithm 2 (shrinking parameter estimators approach)

Step 1: Compute $\{\hat{\boldsymbol{\phi}}, T_j\}$ on the original sample $\{Y_1, \ldots, Y_n\}$;
Step 2: Compute $\breve{\varepsilon}_i$, $i = 1, \ldots, n$ as in (11) and draw a random sample (with replacement) of size n, say $\{\varepsilon^*_1, \ldots, \varepsilon^*_n\}$, independent of the original data, from the empirical distribution function $\breve{F}_n(\cdot) := n^{-1}\sum_{i=1}^{n} \mathbb{I}(\breve{\varepsilon}_i \leq \cdot)$;
Step 3: Generate the bootstrap sample $\{Y^*_1, \ldots, Y^*_n\}$ with bootstrap true values $(\hat{\boldsymbol{\phi}}^{\dagger}, \breve{F}_n)$ as

$$Y^*_i = \{h^*_i(\hat{\boldsymbol{\phi}}^{\dagger})\}^{1/2}\varepsilon^*_i, \quad h^*_i(\hat{\boldsymbol{\phi}}^{\dagger}) = \hat{\omega}^{\dagger} + \sum_{j=1}^{p_1} \hat{\alpha}^{\dagger}_j (Y^*_{i-j})^2 + \sum_{k=1}^{p_2} \hat{\beta}^{\dagger}_k h^*_{i-k}(\hat{\boldsymbol{\phi}}^{\dagger}), \quad i \geq 1$$

initialized with $(Y^*_0, \ldots, Y^*_{1-p_1}, h^*_0(\hat{\boldsymbol{\phi}}^{\dagger}), \ldots, h^*_{1-p_2}(\hat{\boldsymbol{\phi}}^{\dagger}))' = \varsigma_0$, where ς_0 is an arbitrarily chosen vector (e.g., set $Y^*_t = 0$ and $h^*_t = 0$, all $t \leq 0$);
Step 4: Using $\{Y^*_1, \ldots, Y^*_n\}$, compute $\{\hat{\boldsymbol{\phi}}^*, T^*_j\}$ the bootstrap analogs of $\{\hat{\boldsymbol{\phi}}, T_j\}$.
The bootstrap test then corresponds to the decision rule:

$$\text{Reject } \mathsf{H}_0 \text{ at the nominal level } \alpha \text{ if the estimated p-value } p^*_n \text{ is less than } \alpha. \tag{21}$$

The bootstrap p-value p^*_n is defined as in (14), and it can be approximated with arbitrary accuracy by repeating steps 2–4 a large number of times, say B, and then setting $p^{*(B)}_n$ to be the fraction of times T^*_j exceeds T_j $(j = 1, 2)$.

Note that, the limiting distribution of $n^{1/2}(\hat{\boldsymbol{\phi}}^{\dagger} - \boldsymbol{\phi}_0)$ is the same as that of $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ whenever $\boldsymbol{\phi}_0$ is in the interior of Φ. Hence, the bootstrap test (21) collapses into the bootstrap method outlined in Section 3.2 as $n \to \infty$, whenever $\boldsymbol{\phi}_0$ is in the interior of Φ.

4.4 Asymptotic validity

In this section we establish the asymptotic validity of the bootstrap based on the shrinking parameter estimators approach introduced in the previous subsection.

Note that the bootstrap analogue of the marked empirical process $\mathcal{U}_n(y, \boldsymbol{\phi})$ for the bootstrap test (21) is defined as in (16), with

$$\mathcal{U}_n^*(y, \boldsymbol{\phi}) := n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{(Y_i^*)^2}{h_i^*(\boldsymbol{\phi})} - 1 \right\} \mathbb{I}(Y_{i-1}^* \leq y), \quad y \in \mathbb{R},\, \boldsymbol{\phi} \in \Phi,$$

except that Y_i^* and $h_i^*(\boldsymbol{\phi})$ are now based on the bootstrap method outlined in Section 4.3. The next theorem establishes the asymptotic validity of the bootstrap test (21). First, we introduce the following additional assumption. Recall that F_0 is the c.d.f of ε_i in (1).

(A6). *The Assumptions (A1)–(A5) continue to hold when* $\zeta_0 = (\boldsymbol{\phi}_0, F_0)$ *is replaced by* $\zeta_n = (\boldsymbol{\phi}_n, F_n)$*, where* $\zeta_n \to (\boldsymbol{\phi}_0^*, F_0^*) := \operatorname{plim}(\hat{\boldsymbol{\phi}}, \check{F}_n)$ *as* $n \to \infty$*.*

Since Assumptions (A1)–(A5) correspond to the original data generating process, the underlying true parameter value $\zeta_0 = (\boldsymbol{\phi}_0, F_0)$ is fixed. However, in the bootstrap data generation the true parameter $(\hat{\boldsymbol{\phi}}^\dagger, \check{F}_n)$ is not fixed but converges to $(\boldsymbol{\phi}_0^*, F_0^*)$ as $n \to \infty$. Therefore, it is not adequate to assume only (A1)–(A5) in order to establish the validity of the bootstrap tests. Assumption (A6) ensures that (A1)–(A5) hold for triangular arrays, and hence allows us to extend the arguments in the proof of Lemma 2 to a triangular array setup, which in turn is essential for establishing that the limiting distribution of $n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}_n^\dagger)$, conditional on $(Y_1, \ldots, Y_n)$, is the same as that of $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ under H_0, while being of order $O_p^*(1)$, in probability, under H_1. This result plays a key role in the proofs of the asymptotic validity and consistency of the bootstrap tests obtained in the next two theorems.

Theorem 4. *Suppose that Assumptions (A1)–(A6) and (B2)–(B4) hold. Then, under* H_0*, the conditional weak limit of* $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ *is the same as that of* $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$*, in probability, and hence the bootstrap test* (21) *based on* T_j *is asymptotically valid* ($j = 1, 2$)*.*

The next theorem establishes the consistency of the bootstrap test (21).

Theorem 5. *Suppose that* H_1 *holds. Assume that the estimator* $\hat{\boldsymbol{\phi}}$ *converges in probability to some point in* Φ*, and* $\check{F}_n$ *in* (12) *converges in probability with respect to the Mallows metric. Suppose that Assumptions (B1)–(B4) hold with* $(\boldsymbol{\phi}_0^*, F_0^*) := \operatorname{plim}(\hat{\boldsymbol{\phi}}, \check{F}_n)$*. Additionally, assume that Assumption (A6) holds,* $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*) = O_p(1)$ *and there exists a* $y \in \mathbb{R}$*, with* $h_i = \mathrm{E}(Y_i^2 \mid \mathcal{H}_{i-1})$*,* $i \in \mathbb{Z}$*, such that* $\mathrm{E}[\{h_1/h_1(\boldsymbol{\phi}_0^*) - 1\}\mathbb{I}(Y_0 \leq y)] \neq 0$*. Then, conditional on* $\{Y_1, \ldots, Y_n\}$*, the bootstrap test (21) based on* T_j *has asymptotic power 1* ($j = 1, 2$)*.*

Theorem 4 shows that the proposed shrinkage in terms of the c_n sequence, or more generally, the requirement (18) ensures that the bootstrap test statistics T_1^* and T_2^* based on (21) replicate the unknown limiting distributions of T_1 and T_2 under the null hypothesis. Theorem 5 establishes that T_1^* and T_2^* are of order $O_p^*(1)$, in probability, under the alternative; that is, the proposed bootstrap method is also consistent even if it is unknown whether any of the nuisance parameters are on the boundary or not.

5 Numerical Study

In this section we carry out a Monte Carlo simulation study to evaluate the finite sample performance of the KS and CvM tests based on the bootstrap method (21) in Section 4.3. Our main focus is the case where the true parameter value $\boldsymbol{\phi}_0$ of the data generating process lies on the boundary of the parameter space. For comparison, we also consider the case where $\boldsymbol{\phi}_0$ is an interior point. Several data generating processes under the alternative hypothesis are also considered in order to investigate the finite sample power properties of the tests. Although there are several other tests that can be applied for testing the conditional variance specification in GARCH-type models, as mentioned in the introduction, the theory for their validity does not hold when the true parameter is on the boundary. Hence, in these simulations, we compare the proposed tests with the general purpose Ljung-Box Q test which tests the significance of the serial dependence of the squared residuals estimated from the fitted model. We denote the Ljung-Box Q test for a lag length ℓ by LBQ(ℓ).

5.1 Design of the simulation study

Tests are evaluated when the parametric form $h_i(\boldsymbol{\phi})$ under H_0 is

$$\begin{aligned}
&\mathsf{H}_0^A \text{ [GARCH(1,1)]:} \qquad h_i(\boldsymbol{\phi}) = \omega + \alpha Y_{i-1}^2 + \beta h_{i-1}(\boldsymbol{\phi}),\\
&\mathsf{H}_0^B \text{ [GARCH(1,2)]:} \qquad h_i(\boldsymbol{\phi}) = \omega + \alpha Y_{i-1}^2 + \beta_1 h_{i-1}(\boldsymbol{\phi}) + \beta_2 h_{i-2}(\boldsymbol{\phi}).
\end{aligned}$$

For the error distribution, we consider the standard normal distribution.

For the conditional variance h_i of the true data generating process [DGP] we consider the following 9 cases:

DGP1 [ARCH(1)]: $h_i = 0.20 + 0.7Y_{i-1}^2$, [$\mathsf{H}_0^A$ and H_0^B both true]

DGP2 [GARCH(1,1)]: $h_i = 0.10 + 0.20Y_{i-1}^2 + 0.70h_{i-1}$, [$\mathsf{H}_0^A$ and H_0^B both true]

DGP3 [GARCH(1,2)]: $h_i = 0.10 + 0.10Y_{i-1}^2 + 0.15h_{i-1} + 0.70h_{i-2}$, [$\mathsf{H}_0^A$ false and H_0^B true]

DGP4 [GJR-GARCH(1,1)][1]: [H_0^A and H_0^B both false]

$h_i = 0.10 + 0.1Y_{i-1}^2 + 0.5h_{i-1} + 0.3Y_{i-1}^2\mathbb{I}(Y_{i-1} < 0)$,

[1]The Glosten-Jagannathan-Runkle GARCH model of Glosten *et al.* (1993).

DGP5 [GARCH(2,2)]:
$h_i = 0.10 + 0.05Y_{i-1}^2 + 0.6Y_{i-2}^2 + 0.1h_{i-1} + 0.2h_{i-2}$, [$\mathsf{H}_0^A$ and H_0^B both false]
DGP6 [EGARCH(1,1)]:
$\ln h_i = 0.1 + 0.4 \ln h_{i-1} + 0.2(|\varepsilon_{i-1}| - \mathrm{E}|\varepsilon_{i-1}|) - 0.2\varepsilon_{i-1}$, [$\mathsf{H}_0^A$ and H_0^B both false]
DGP7 [i.i.d.]: $h_i = 1$, [H_0^A and H_0^B both true, but Assumption (A3) false]
DGP8 [Threshold GARCH(1,1)]:
$h_i = 0.10 + 0.1Y_{i-1}^2 + 0.5h_{i-1} + 0.3h_{i-1}\mathbb{I}(Y_{i-1} < 0)$, [$\mathsf{H}_0^A$ and H_0^B both false]
DGP9 [T-CHARM][2]: $h_i = \mathbb{I}(Y_{i-1} \leq 0) + 1.2\mathbb{I}(Y_{i-1} > 0)$. [$\mathsf{H}_0^A$ and H_0^B both false.]

DGPs 1, 2, 3, 5 and 7 are linear models. We use these DGPs to evaluate the size and power properties of the tests, focusing particularly on the cases where the true parameter lies on the boundary. The DGPs 4, 6, 8 and 9 are nonlinear models, and hence the linear models specified under the null hypotheses H_0^A and H_0^B are misspecified for each of DGPs 4, 6, 8 and 9. We use these four DGPs to evaluate the finite sample power properties of the tests.

In the next two subsections we consider testing for the two null models H_0^A [GARCH(1,1)] and H_0^B [GARCH(1,2)] separately. The results are based on 2000 Monte Carlo replications. For each replication and data generating process, we first compute the QMLE $\hat{\boldsymbol{\phi}}$ and compute the test statistics KS, CvM, and LBQ(ℓ), $\ell = 3, 5, 10, 15, 20$. To implement the proposed KS and CvM tests we use the bootstrap method (21) outlined in Section 4.3 with $c_n = n^{-1/3}/50$, while adopting the 'Warp-Speed' Monte Carlo method of Giacomini *et al.* (2013) in order to reduce the computational burden. The results are presented in Figures 1–6. In these figures the results of the LBQ(ℓ) tests are presented for only $\ell = 5, 10$ and 15; the patterns of the results for $\ell = 3$ and $\ell = 20$ are similar to those for $\ell = 5, 10$ and 15 and hence are omitted.

5.2 Testing for H_0^A [GARCH(1,1)]

The DGP1 [ARCH(1)] and DGP2 [GARCH(1,1)] are members of the null family under H_0^A. For the DGP1, the GARCH coefficient is zero, and hence the true parameter lies on the boundary of the parameter space. For the DGP2, the true parameter is an interior point. The DGP7 [i.i.d.] is a member of the null family H_0^A with the true value on the boundary. However, since there are no ARCH or GARCH coefficients, the DGP 7 does not satisfy Assumption (A3) and hence the theory developed in this paper does not apply in this case. The DGPs 3, 4, 5, 6, 8 and 9 are considered in order to evaluate the empirical power properties of the tests; these DGPs are part of the alternative hypothesis. A summary of the results are given in Figures 1, 2, 5 and 6, and are discussed in Section 5.4.

[2]T-CHARM refers to the conditionally heteroscedastic AR models proposed by Chan *et al.* (2014).

5.3 Testing for H_0^B [GARCH(1,2)]

The DGPs 1, 2 and 3 are members of the null family H_0^B [GARCH(1,2)]. For each of DGP1 and DGP2, one of ARCH or GARCH coefficients is zero, and hence the true parameter lies on the boundary of the parameter space. The true parameter of the DGP3 [GARCH(1,2)] is an interior point. Since there are no ARCH or GARCH coefficients, when testing for H_0^B, the DGP7 does not satisfy Assumption (A3) and hence as in the previous case, the theory developed in this paper does not apply. The DGPs 4, 5, 6, 8 and 9 continue to be under the alternative hypothesis when testing for GARCH(1,2).

The results are given in Figures 3, 4, 5 and 6, and are discussed in the next subsection.

5.4 Summary of the results

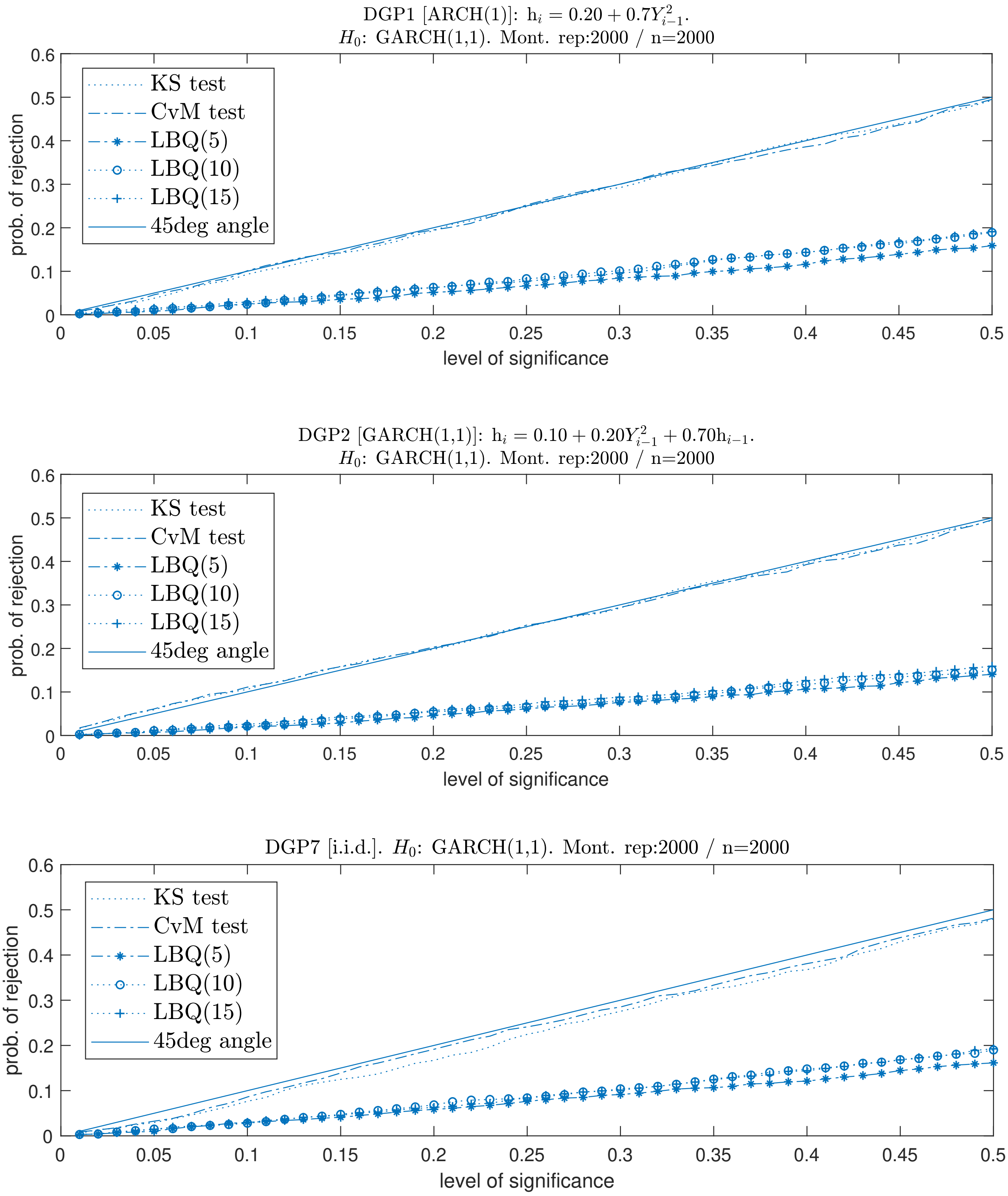
The main observations of the simulation results are the following:

(i) The KS and CvM bootstrap tests proposed in this paper perform consistently well in terms of the Type-I error rate. By contrast the LBQ test does not perform well in terms of finite sample size at any of the lag lengths considered. In particular, for every instance in which the DGP is under the null hypothesis, the LBQ test is significantly undersized at each of the lag lengths considered.

(ii) No significant difference in performance of the KS and CvM tests can be identified, in terms of Type-I error rates, irrespective of whether the true parameters are on the boundary or not. Thus, the simulation results indicate that the bootstrap method based on the shrinking parameter estimators approach outlined in Section 4.3 performs well irrespective of whether the true value of $\boldsymbol{\phi}$ is on the boundary or in the interior.

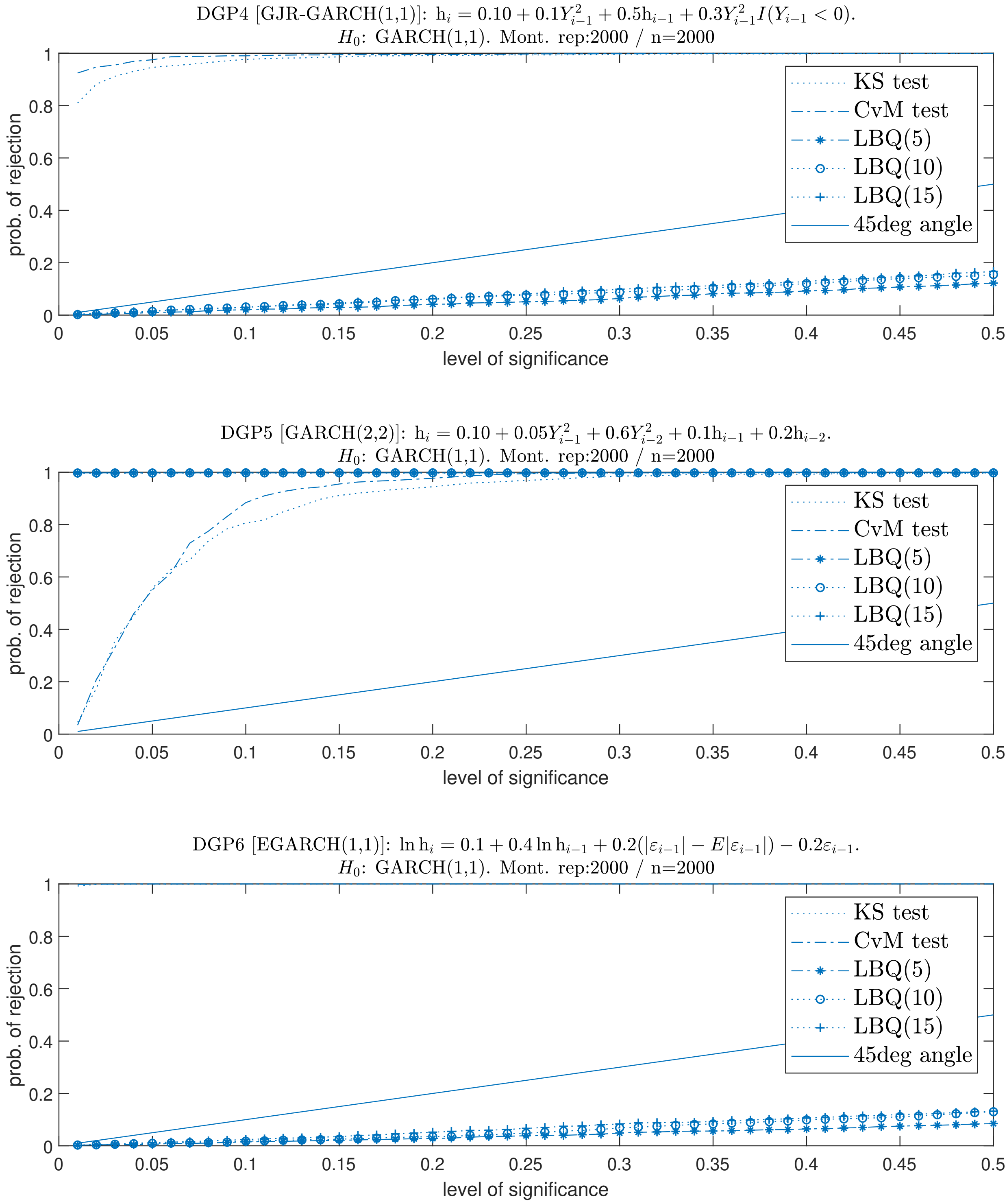
(iii) Both KS and CvM tests exhibit good overall power properties. In particular, they exhibit excellent power gains compared to the LBQ test when the DGP is non-linear. Although the LBQ test performs better than KS and CvM when the DGP is linear and the misspecification is in terms of some missing lags, the power gains are not as significant as in the cases where KS and CvM outperform LBQ. In particular, the LBQ test does not exhibit any empirical power against the nonlinear DGPs, DGP4, DGP6, DGP8 and DGP9, when testing for both H_0^A [GARCH(1,1)] and H_0^B [GARCH(1,2)]. As expected, the empirical power of KS and CvM tests increase with the significance level α and the sample size n. Overall, in terms of empirical power, the CvM test performs marginally better than the KS test.

Figure 1: Empirical rejection rates for testing H_0^A [GARCH(1,1)]



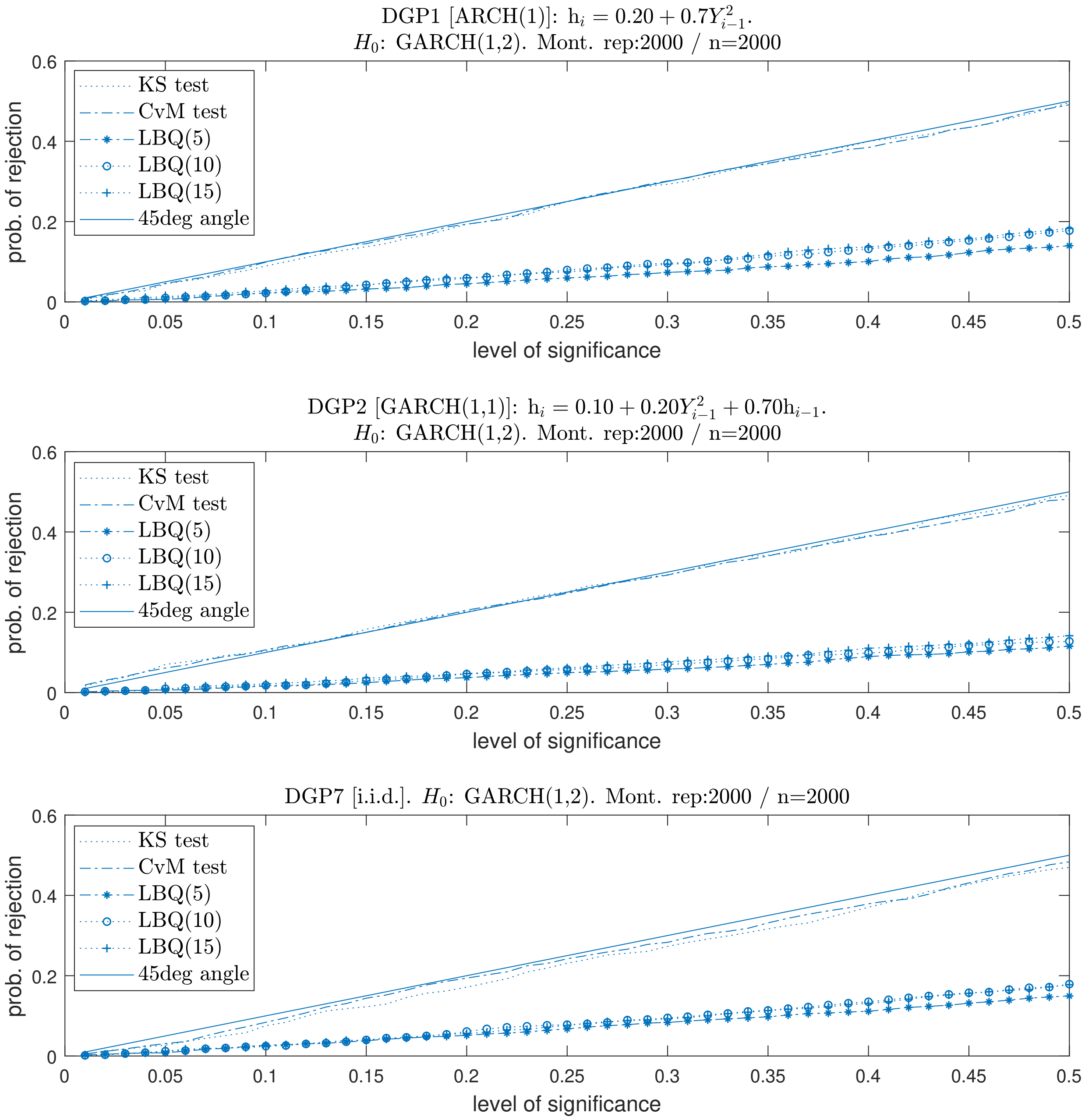
Notes: DGPs 1 and 2 are under H_0^A. The true parameter for the DGP1 is a boundary point. For the DGP2, the true parameter is an interior point. Although the DGP7 [i.i.d.] is a member of the null family H_0^A, with the true value on the boundary, it does not satisfy all the conditions assumed for establishing the validity of the bootstrap tests.

Figure 2: Empirical power of KS and CvM for testing H_0^A [GARCH(1,1)]



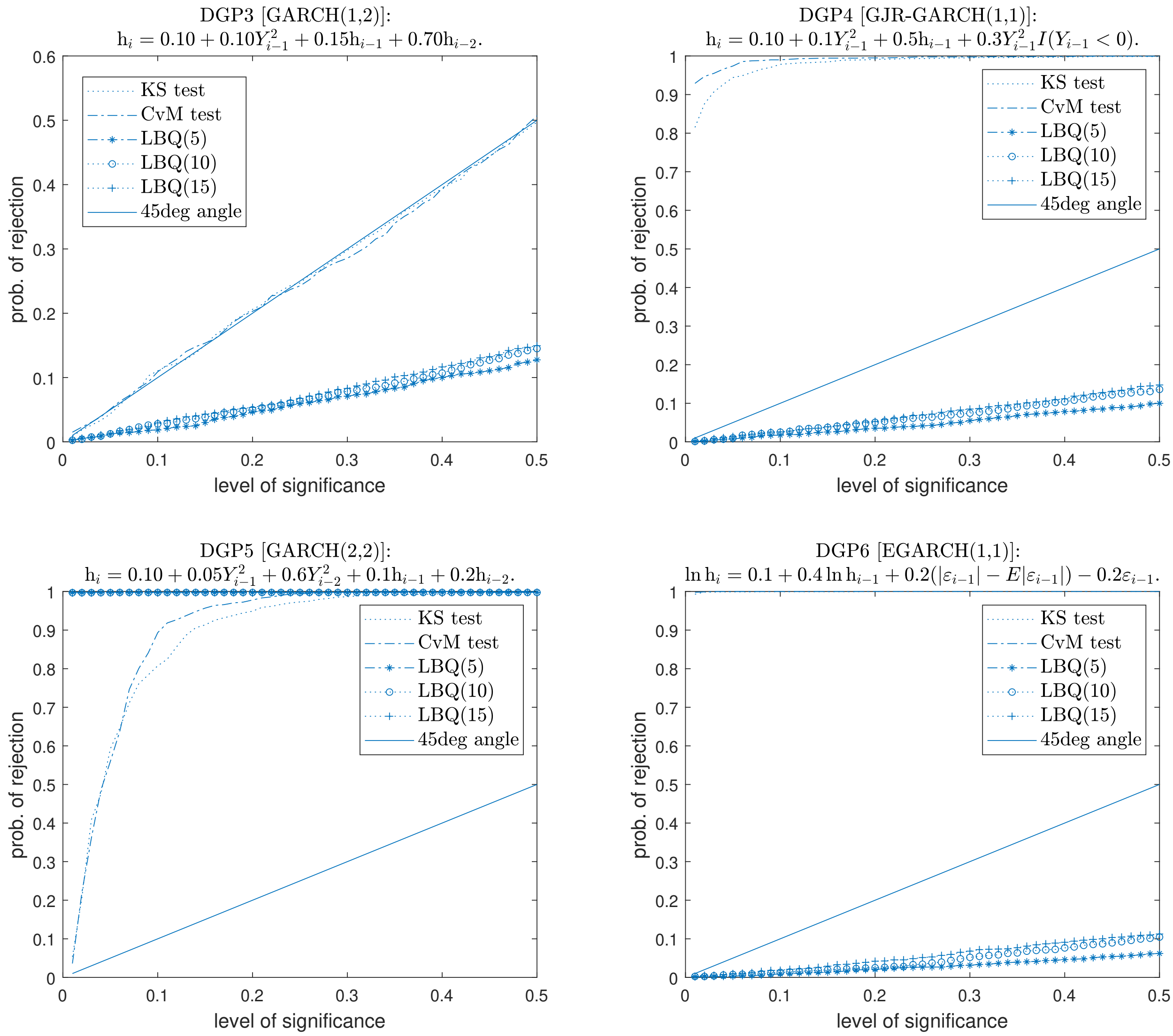
Notes: The DGPs 4–6 are considered to evaluate the power of the tests.

Figure 3: Empirical rejection rates for testing H_0^B [GARCH(1,2)] when the true parameter lies on the boundary of the parameter space.



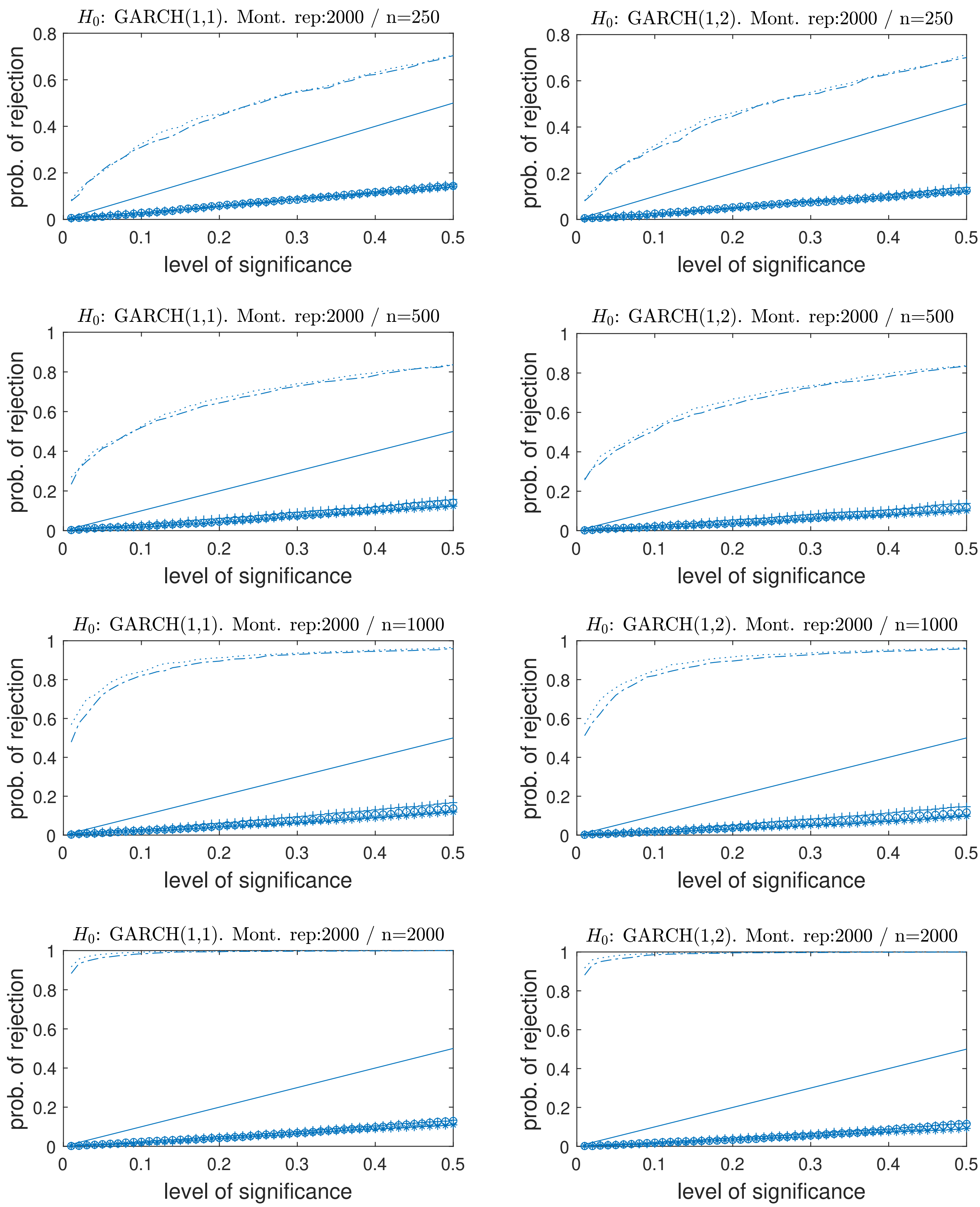
Notes: For DGPs 1 and 2, the true parameter under H_0^B lies on the boundary of the parameter space.

Figure 4: Empirical size and power for testing H_0^B [GARCH(1,2)]. The sample size $n = 2000$.



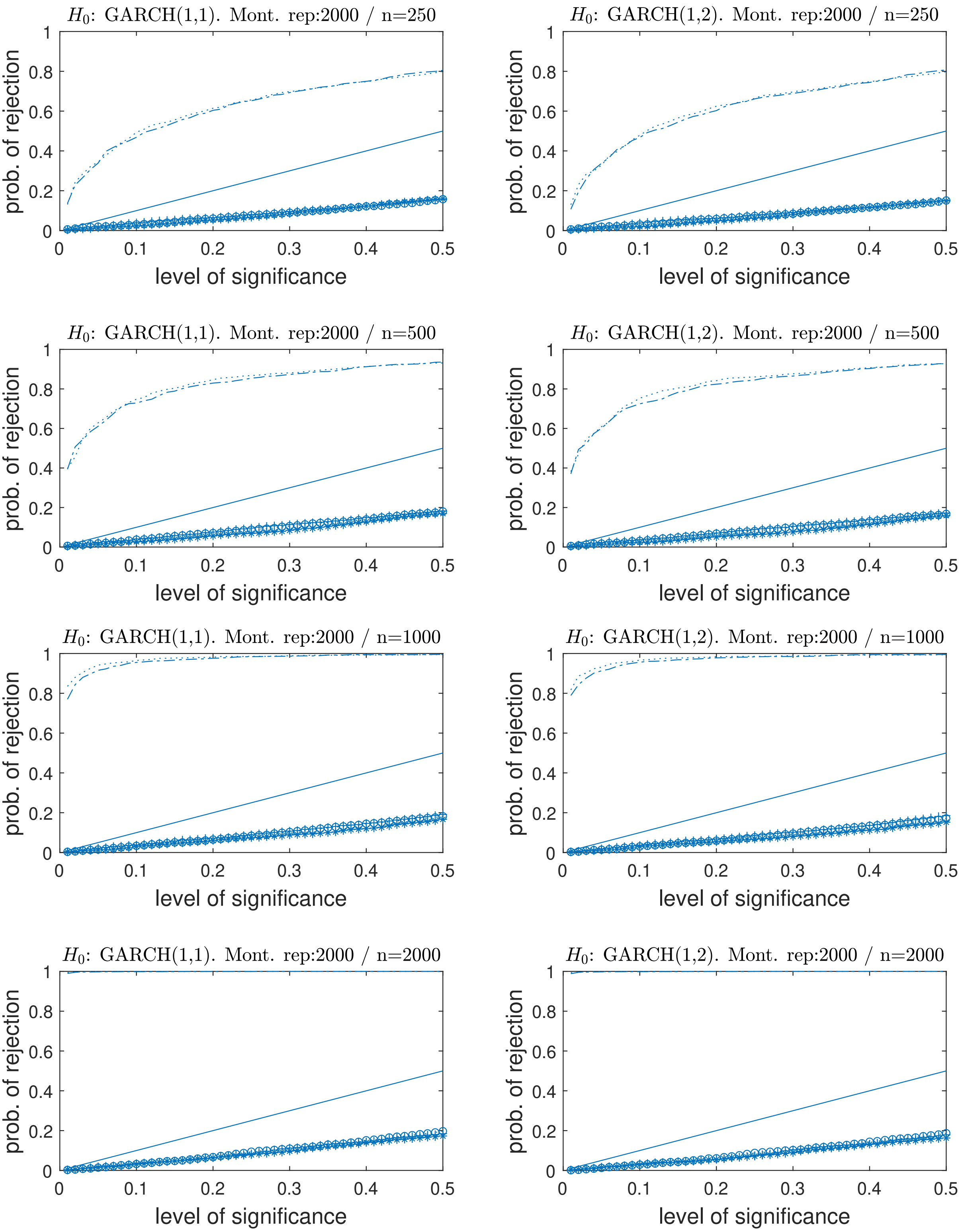
Notes: The DGP3 [GARCH(1,2)] is under H_0^B; the true parameter of the DGP3 is an interior point of the parameter space. The DGPs 4, 5, and 6 are part of the alternative.

Figure 5: Empirical power for testing H_0^A [GARCH(1,1)] and H_0^B [GARCH(1,2)] for the DGP8 [Threshold GARCH(1,1)]: $h_i = 0.10 + 0.1Y_{i-1}^2 + 0.5h_{i-1} + 0.3h_{i-1}\mathbb{I}(Y_{i-1} < 0)$. KS $\cdots$, CvM $- \cdot - \cdot -$, LBQ(5) $- \cdot * \cdot -$, LBQ(10) $- \cdot \circ \cdot -$, LBQ(15) $- \cdot + \cdot -$, 45deg $-$.



Notes: GARCH(1,1) and GARCH(1,2) are both misspecified for the DGP8.

Figure 6: Empirical power for testing H_0^A [GARCH(1,1)] and H_0^B [GARCH(1,2)] for the DGP9 [T-CHARM]: $h_i = \mathbb{I}(Y_{i-1} \leq 0) + 1.2\mathbb{I}(Y_{i-1} > 0)$. KS $\cdots$, CvM $- \cdot - \cdot -$, LBQ(5) $- \cdot * \cdot -$, LBQ(10) $- \cdot \mathrm{o} \cdot -$, LBQ(15) $- \cdot + \cdot -$, 45deg angle $-$.



Notes: GARCH(1,1) and GARCH(1,2) are both misspecified for the DGP9.

6 Empirical illustrations

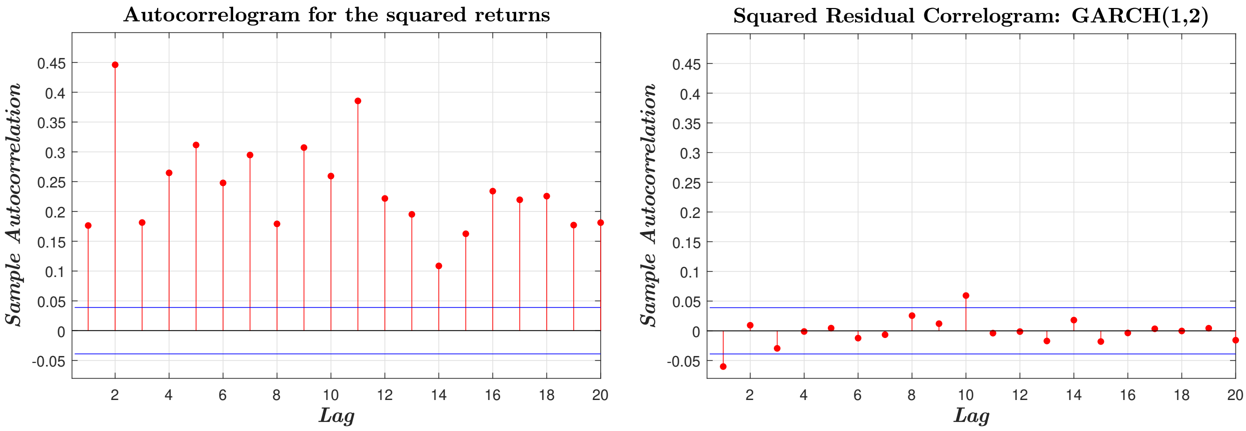
To illustrate the bootstrap testing procedure, we briefly discuss two real data examples.

Example 1

We first consider a data example based on the daily log returns of the SPDR exchange-traded fund (ETF) for the S&P 500 index. This ETF is usually denoted by the tick symbol SPY. The data spans the period January 3, 2007 to June 30, 2017 and contains 2640 observations. A shorter version of this data set was previously studied by Tsay and Chen (2018), and by using some preliminary diagnostics, they concluded that a GJR-GARCH(1,1) model provides a good fit. In their empirical analysis, Tsay and Chen (2018) concluded that the leverage effect of the fitted GJR-GARCH(1,1) model is statistically significant at the 5% level. This indicates that if one specifies a GARCH(1,2) model for the conditional variance then that may not provide a good fit for the data. In order to investigate this, in this empirical illustration, we employ the proposed KS and CvM bootstrap tests to test the adequacy of the GARCH(1,2) specification, expecting that the proposed tests would be able to detect a misspecification. For comparison the LBQ test considered in the simulations in the previous section is also considered.

For the GARCH(1,2) specification, the p-values of the KS and CvM tests are both zero up to 3 decimal places, whereas the p-value for the LBQ(20) turns out to be 0.123, and those for LBQ(15) and LBQ(5) are 0.032 and 0.033, respectively. Thus, the KS and CvM tests proposed in this paper clearly reject the GARCH(1,2) specification, but the LBQ(20) fails to reject the GARCH(1,2) specification at the 10% level of significance. Note that the Ljung-Box Q test is designed to check the significance of the autocorrelations of the squared residuals at multiple lags jointly. Figure 7 shows the sample autocorrelations for both the squared values of the observed time series and the squared residuals estimated from the fitted GARCH(1,2). As expected, squared SPY log returns are significantly serially correlated, but the correlogram of squared residuals suggests no significant serial correlations except for some minor ones at lags 1 and 10. This explains the relatively large p-values of the LBQ test. However, squared residuals can be serially uncorrelated, but dependent, and hence it appears that the tests proposed in this paper are better suited than the Ljung-Box Q test in detecting the misspecification of the conditional variance specification in this case.

Figure 7: Autocorrelogram of the squared SPY log returns time series (first panel), and the squared residual correlogram for the fitted GARCH(1,2) model (second panel). The sample period is from January 3, 2007 to June 30, 2017.



Example 2

In this illustrative example we consider a data set from the Caterpillar stock traded on the New York Stock Exchange. The variable of interest is the daily log return of the Caterpillar stock, defined by $Y_t = 100(\log P_t - \log P_{t-1})$ where P_t is the stock price at time t. The sample contains 2515 observations and spans the period Jan 02, 2001 to Dec 31, 2010. Tsay (2013) analyzed this data set by applying several diagnostic methods, and fitted a GARCH(1,1) model (see Table 5.1 in Tsay, 2013). When we fit a GARCH(1,2) model to this data set, the estimated GARCH(2) coefficient turns out to be statistically insignificant, practically at any level of significance. This indicates that, when testing the GARCH(1,2) specification, one component of the true parameter could potentially be a boundary point of the parameter space, whereas when the null model is GARCH(1,1) the true parameter could potentially be an interior point. Of course we do not have any certainty that this is actually true. But, as an illustration, we employ the proposed KS and CvM bootstrap tests to test GARCH(1,1) and GARCH(1,2) specifications. For comparison the LBQ test is also considered. The p-values of the tests are given in Table 1. As expected the tests support both GARCH(1,1) and GARCH(1,2) specifications with large p-values. In the simulations in the previous section, the LBQ test was undersized when testing for the correct specification. Thus, the large p-values of the LBQ test in Table 1 are consistent with the simulation results reported in the previous section.

Table 1: The p-values of the specification tests for testing GARCH(1,1) and GARCH(1,2) specifications for the conditional variance of the daily log-return of the Caterpillar stock. The data spans the period Jan 02, 2001 to Dec 31, 2010.

	Tests						
Null model	KS	CvM	LBQ(3)	LBQ(5)	LBQ(10)	LBQ(15)	LBQ(20)
GARCH(1,1)	0.363	0.332	0.943	0.995	0.999	0.999	0.999
GARCH(1,2)	0.491	0.427	0.880	0.978	0.999	0.998	0.999

7 Conclusion

This paper contributes to advance the current statistical methodology for inference in GARCH models by developing bootstrap based omnibus specification tests while allowing parameters on the boundary of the parameter space. In particular, Kolmogorov-Smirnov and Cramér-von Mises type test statistics are proposed based on a certain empirical process marked by centered squared residuals. We first derive the asymptotic null distributions of the proposed test statistics when the true parameter is in the interior of the parameter space. Since the limiting distributions of the test statistics are not free from (unknown) nuisance parameters, we propose a bootstrap method to implement the tests and establish that the proposed bootstrap method is asymptotically valid and consistent. However, when some components of the nuisance parameters lie on the boundary of the parameter space, this bootstrap testing procedure is not consistent. Hence, as an alternative, we also propose a modified version of the bootstrap by employing a method of shrinkage of the parameter estimates in the bootstrap data generation. We show that the modified bootstrap procedure is asymptotically valid and consistent, regardless of the presence of nuisance parameters on the boundary. Our bootstrap methods can be implemented easily under fairly general and easily verifiable assumptions and have desirable finite sample properties in terms of empirical size and power.

Our results can be extended in several directions. For instance, it is of interest to see if the methods we propose in this paper can be extended to models beyond the standard GARCH(p_1, p_2). To this end, consider the model $\mathcal{M}$ defined by

$$\mathcal{M}: \quad Y_i \ = h_i^{1/2}\varepsilon_i, \quad h_i = g_{\boldsymbol{\phi}}(Y_{i-1}, \cdots, Y_{i-p_1}, h_{i-1}, \cdots, h_{i-p_2}), \quad i \in \mathbb{Z}, \tag{22}$$

for some $\boldsymbol{\phi} \in \Phi \subset \mathbb{R}^{p_1+p_2}$, where $\{g_{\boldsymbol{\phi}}; \boldsymbol{\phi} \in \Phi\}$ is a parametric family of nonnegative functions on $\mathbb{R}^{p_1+p_2}$, and the error terms $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are i.i.d. with zero mean and unit variance. Thus,

$h_i = h_i(\boldsymbol{\phi}) = \mathrm{Var}(Y_i \mid \mathcal{H}_{i-1}),\ i \in \mathbb{Z}$. Consider the hypothesis testing problem

$$\mathsf{H}_0 : \text{Model } \mathcal{M} \text{ is correct} \qquad \text{vs} \qquad \mathsf{H}_1 : \text{Model } \mathcal{M} \text{ is not correct.} \tag{23}$$

The GARCH(p_1, p_2) model is a special case of $\mathcal{M}$. Another example is the asymmetric AGARCH(p_1, p_2) model defined by $h_i = h_i(\boldsymbol{\phi}) = \alpha_0 + \sum_{j=1}^{p_1} \alpha_j \big(|Y_{i-j}| - \gamma Y_{i-j}\big)^2 + \sum_{k=1}^{q} \beta_k h_{i-k}$, where $\boldsymbol{\phi} = (\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma)'$, $\alpha_0 > 0, \alpha_j \geq 0,\ \beta_k \geq 0$ $(i \in \mathbb{Z},\ 1 \leq j \leq p,\ 1 \leq k \leq q)$. Similarly, several other extensions of the standard GARCH model can also be written in the general form (22).

Heuristic arguments suggest that the bootstrap tests proposed in this paper for ARCH(p) and GARCH(p_1, p_2) models can also be extended to this general setup. In fact, the bootstrap algorithm outlined in Section 4.3 can be readily applied to any model of the form (22), based on a suitable estimator for $\boldsymbol{\phi}$. However, parameter estimation, when the true value is on the boundary of the parameter space, in the family of models in (22), has not yet been studied, and therefore it is not a trivial task to extend the methods developed in this paper to a general setup of the form (22); this would provide a potential direction for a possible extension of the paper. Furthermore, our testing procedures can also be potentially extended to Poisson autoregressions with exogenous covariates as considered in Agosto *et al.* (2016).

A APPENDIX: Assumptions and Proofs

A.1 Some notations and assumptions

In this appendix we introduce some additional notation, and state Assumptions (B1)–(B5) required for the main theorems. First, we generalize the data generating process specified by model (1)–(4) for an arbitrary $\boldsymbol{\phi} \in \Phi$ and a given innovation distribution F with zero mean and unit variance. To this end we need to first introduce the following regularity assumption.

(B1). *The process $\{Y_i\}_{i \in \mathbb{Z}}$ is strictly stationary and ergodic and obeys model (1) under the alternative hypothesis H_1. The parameter space Φ is a compact subset of $(0, \infty) \times [0, \infty)^{p_1 + p_2}$ and contains a hypercube of the form $[\omega_L, \omega_U] \times [0, \epsilon]^{p_1 + p_2}$, for some $\epsilon > 0$ and $\omega_U > \omega_L > 0$, which includes $\boldsymbol{\phi}_0^* = (\omega_0^*, \alpha_{01}^*, \ldots, \alpha_{0p_1}^*, \beta_{01}^*, \ldots, \beta_{0p_2}^*)'$, where $\boldsymbol{\phi}_0^*$ is the pseudo-true parameter value under H_1, defined by $\boldsymbol{\phi}_0^* := \operatorname{plim} \hat{\boldsymbol{\phi}}$, where 'plim' is the probability limit as $n \to \infty$. Further, $\sum_{i=1}^{p_1} \alpha_{0i}^* + \sum_{j=1}^{p_2} \beta_{0j}^* < 1$ for $p_1 \geq 1$, $p_2 \geq 0$, $\sum_{i=1}^{p_1} \alpha_{0i}^* \neq 0$.*

The strict stationarity of the process $\{Y_i : i \in \mathbb{Z}\}$ obeying (1)–(4), which follows from (A1), (A2) and (A3), ensures that the true parameter $\boldsymbol{\phi}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p_1}, \beta_{01}, \ldots, \beta_{0p_2})'$ under the null hypothesis H_0 satisfies $\sum_{i=1}^{p_1} \alpha_{0i} + \sum_{j=1}^{p_2} \beta_{0j} < 1$ (see Bougerol and Picard, 1992a,b). Assumption (B1) assumes that this continues to hold when $\boldsymbol{\phi}_0^*$ is the pseudo true value under the alternative hypothesis H_1. Since $(\boldsymbol{\phi}_0^*, F_0^*) := \text{plim}(\hat{\boldsymbol{\phi}}, \check{F}_n)$, under the null hypothesis H_0, we have that $(\boldsymbol{\phi}_0^*, F_0^*) = (\boldsymbol{\phi}_0, F_0)$, and under the alternative hypothesis H_1, $(\boldsymbol{\phi}_0^*, F_0^*)$ is the pseudo-true value of $(\boldsymbol{\phi}, F)$. Therefore, if either H_0 holds under (A1)–(A3) or H_1 holds under (B1), regardless of whether $\boldsymbol{\phi}_0^*$ is in the interior or on the boundary of Φ, for all sufficiently small $\epsilon > 0$ and $\omega_U > \omega_L > 0$, there exists a hypercube of the form

$$\bar{\Phi} := [\omega_L, \omega_U] \times [0, \epsilon]^{p_1+p_2} \subset \Phi \tag{A.1}$$

including $\boldsymbol{\phi}_0^*$, such that, for every $\boldsymbol{\phi} \in \bar{\Phi}$ and c.d.f. F (with mean 0 and variance 1), the model defined by

$$\begin{aligned} Y_i^{(\boldsymbol{\phi},F)} &= \{h_i^{(\boldsymbol{\phi},F)}(\boldsymbol{\phi})\}^{1/2} \varepsilon_i^{(F)}, \\ h_i^{(\boldsymbol{\phi},F)}(\boldsymbol{\phi}) &= \omega + \sum_{j=1}^{p_1} \alpha_j \{Y_{i-j}^{(\boldsymbol{\phi},F)}\}^2 + \sum_{j=1}^{p_2} \beta_j h_{i-j}^{(\boldsymbol{\phi},F)}(\boldsymbol{\phi}), \end{aligned} \tag{A.2}$$

has a unique strictly stationary and ergodic solution with $\mathrm{E}[\{Y_0^{(\boldsymbol{\phi},F)}\}^2] < \infty$, where $\varepsilon_i^{(F)} = F^{-1}(U_i) := \inf\{y \in \mathbb{R} : F(y) \geq U_i\}$ and $\{U_i, i \in \mathbb{Z}\}$ are i.i.d. uniform(0,1) random variables, for example, by Theorem 2.1 of Chen and An (1998).

For $(\boldsymbol{\phi}, F) = (\boldsymbol{\phi}_0, F_0)$ the model (A.2) is equivalent to the DGP defined by (1)–(4). Usually, $\boldsymbol{\phi}_0$ and F_0 are unknown. Hence, in order to generate data from a model that mimics (1)–(4), one needs to replace $(\boldsymbol{\phi}_0, F_0)$ by some known $(\boldsymbol{\phi}_n, F_n)$ which is sufficiently close to $(\boldsymbol{\phi}_0, F_0)$. Let $(\boldsymbol{\phi}_n, F_n)$ be such a sequence in the product space $\bar{\Phi} \times \mathcal{D}(\mathbb{R})$, F_n is a c.d.f with zero mean and unit variance ($n \in \mathbb{N}$), such that $(\boldsymbol{\phi}_n, F_n) \to (\boldsymbol{\phi}_0^*, F_0^*)$ as $n \to \infty$, with $\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_0^*\| \to 0$ and $d_2(F_n, F_0^*) \to 0$ as $n \to \infty$, where $d_2(F_X, F_Y)$ denotes the Mallows metric for the distance between two probability distributions F_X and F_Y (see Lemma A.1). Note that, since $(\boldsymbol{\phi}_0^*, F_0^*) = \text{plim}(\hat{\boldsymbol{\phi}}, \check{F}_n)$, we have $(\boldsymbol{\phi}_0^*, F_0^*) = (\boldsymbol{\phi}_0, F_0)$ under H_0, and $(\boldsymbol{\phi}_0^*, F_0^*)$ is the pseudo-true value under H_1. In what follows, when the DGP (A.2) corresponds to $(\boldsymbol{\phi}_n, F_n)$ instead of using $h_i^{(\phi_n,F_n)}(\cdot)$ and $\tau_i^{(\phi_n,F_n)}(\cdot)$, we let the analogs of $h_i(\cdot)$ and $\tau_i(\cdot)$ be denoted by $h_{ni}(\cdot)$ and $\tau_{ni}(\cdot)$, respectively. Note that, under H_0, the probability laws of $h_i^{(\phi_0,F_0)}(\cdot)$ and $\tau_i^{(\phi_0,F_0)}(\cdot)$ are identical to those of $h_i(\cdot)$ and $\tau_i(\cdot)$, respectively.

Next, let us introduce some notation. Let 'dot' denote differentiation:

$$\dot{h}_i(\boldsymbol{\phi}) = (\partial/\partial\boldsymbol{\phi}) h_i(\boldsymbol{\phi}), \quad \ddot{h}_i(\boldsymbol{\phi}) = (\partial/\partial\boldsymbol{\phi}) \dot{h}_i(\boldsymbol{\phi}).$$

Let $\mathcal{F}$ denote the set of all c.d.f.'s with zero mean and unit variance, i.e,

$$\mathcal{F} := \{F \in \mathcal{D}(\mathbb{R}) : F \text{ is a c.d.f. with mean 0 and variance 1}\}.$$

For any given constant $\delta > 0$, let $\mathcal{F}_\delta^0 := \{F \in \mathcal{F} : d_2(F, F_0^*) \leq \delta\}$.

We say that a sequence of random variables $\{Z_i\}_{i \in \mathbb{N}}$ *converges to zero exponentially almost surely*, denoted $Z_i \overset{e.a.s.}{\to} 0$, if there exists $\gamma > 1$ such that $\gamma^i Z_i \overset{a.s.}{\to} 0$ as $i \to \infty$. The norm $\|\cdot\|_\Lambda$ for a continuous matrix-valued function H on a compact set $\Lambda \subset \mathbb{R}^{r_1}$, that is $H \in \mathbb{C}[\Lambda, \mathbb{R}^{r_2\times r_3}]$, is defined by $\|H\|_\Lambda := \sup_{s\in\Lambda}\|H(s)\|$, when r_1, r_2, r_3 are known positive integers. If H is real valued, then $\|H\|_\Lambda = \sup_{s\in\Lambda}|H(s)|$. We let $\bar{\mathbb{R}} := [-\infty, \infty]$.

In order to establish the asymptotic validity of the bootstrap testing procedure we also introduce the following additional assumptions. Recall that $(\boldsymbol{\phi}_0^*, F_0^*) = \text{plim}(\hat{\boldsymbol{\phi}}, \check{F}_n)$.

(B2). *There exist $\delta > 0$ with $\mathrm{E}([\sup_{F\in\mathcal{F}_\delta^0}\{F^{-1}(U_i)\}^2]^{2+d}) < \infty$ for some $d > 0$.*

(B3). *If $\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_0^*\| \to 0$ and $d_2(F_n, F_0^*) \to 0$ as $n \to \infty$, then for every $y \in \bar{\mathbb{R}}$, we have that $\mathrm{E}\mathbb{I}(Y_1^{(\boldsymbol{\phi}_n,F_n)} \le y) \to \mathrm{E}\mathbb{I}(Y_1^{(\boldsymbol{\phi}_0^*,F_0^*)} \le y)$ as $n \to \infty$.*

(B4). *For every nonrandom sequence $\zeta_n := (\boldsymbol{\phi}_n, F_n) \to \zeta_0^* := (\boldsymbol{\phi}_0^*, F_0^*)$, where $\zeta_n \in \bar{\Phi} \times \mathcal{F}$, we have that $\mathrm{E}[\tau_{n1}(\boldsymbol{\phi}_n)] \to \mathrm{E}[\tau_1^{(\boldsymbol{\phi}_0^*,F_0^*)}(\boldsymbol{\phi}_0^*)]$, $\mathrm{E}[\tau_{n1}(\boldsymbol{\phi}_n)\tau_{n1}(\boldsymbol{\phi}_n)'] \to \mathrm{E}[\tau_1^{(\boldsymbol{\phi}_0^*,F_0^*)}(\boldsymbol{\phi}_0^*)\tau_1^{(\boldsymbol{\phi}_0^*,F_0^*)}(\boldsymbol{\phi}_0^*)']$ as $n \to \infty$. Recall that $\tau_{ni}(\boldsymbol{\phi}) := \dot{h}_{ni}(\boldsymbol{\phi})/h_{ni}(\boldsymbol{\phi})$ with $h_{ni}(\boldsymbol{\phi}) = h_i^{(\boldsymbol{\phi}_n,F_n)}(\boldsymbol{\phi})$.*

The next assumption is used in the proof of Theorem 3 to establish the consistency of the bootstrap test (15) based on T_j $(j = 1, 2)$.

(B5). *There exists a $y \in \mathbb{R}$, with $h_i = \mathrm{E}(Y_i^2 \mid \mathcal{H}_{i-1})$, $i \in \mathbb{Z}$, such that $\mathrm{E}[\{h_1/h_1(\boldsymbol{\phi}_0^*)-1\}\mathbb{I}(Y_0 \le y)] \neq 0$ under H_1, where $\boldsymbol{\phi}_0^*$ is the pseudo-true value under H_1.*

A.2 Some preliminary results

In this subsection we obtain several preliminary lemmas required for the main proofs.

The next lemma shows that $\check{F}_n$ in (12) converges to F_0 with probability 1.

Lemma A.1. *Let $d_2(F_X, F_Y)$ denote the Mallows metric for the distance between two probability distributions F_X and F_Y defined by $d_2(F_X, F_Y) = \inf\{\mathrm{E}|X-Y|^2\}^{1/2}$, where the infimum is over all square integrable random variables X and Y with marginal distributions F_X and F_Y.* (a) *Suppose that Assumptions (A1)–(A4) and H_0 hold, and $\boldsymbol{\phi}_0$ is an interior point in Φ. Then, $d_2(\check{F}_n, F_0) \overset{a.s.}{\to} 0$ as $n \to \infty$.* (b) *Additionally, assume that Assumption (A5) is also satisfied, then $d_2(\check{F}_n, F_0) \overset{a.s.}{\to} 0$ as $n \to \infty$, irrespective of whether $\boldsymbol{\phi}_0$ is in the interior of Φ.*

Proof of Lemma A.1. Under Assumptions (A1)–(A3), $\hat{\boldsymbol{\phi}}$ converges to $\boldsymbol{\phi}_0$ (a.s.), irrespective of whether $\boldsymbol{\phi}_0$ is in the interior of the parameter space (see Lemma 2). If, in addition, (A4) is also satisfied and $\boldsymbol{\phi}_0$ is in the interior of Φ, then $\hat{\boldsymbol{\phi}}$ is asymptotically linear and satisfies (10), and hence $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = O_p(1)$. If Assumptions (A1)–(A5) are satisfied, then $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = O_p(1)$ by Lemma 2, irrespective of whether $\boldsymbol{\phi}_0$ is in the interior of Φ.

Proof of Part (a): Assumptions (A1)–(A4) are satisfied, and $\boldsymbol{\phi}_0$ is an interior point in Φ.

Let $H_n(x) := n^{-1}\sum_{i=1}^{n}\mathbb{I}(\varepsilon_i \le x)$, be the empirical distribution function of the unobserved errors $\{\varepsilon_1, \ldots, \varepsilon_n\}$. From the triangular inequality we have that

$$d_2(\widehat{F}_n, F_0) \le d_2(\widehat{F}_n, H_n) + d_2(H_n, F_0),$$

where $H_n(x) := n^{-1}\sum_{i=1}^{n}\mathbb{I}(\varepsilon_i \le x)$, $x \in \mathbb{R}$, is the empirical distribution function of the unobserved errors $\{\varepsilon_1, \ldots, \varepsilon_n\}$.

We already have that $d_2(H_n, F_0) \overset{a.s.}{\to} 0$ as $n \to \infty$ (see, for example, Lemma 8.4 of Bickel and Freedman, 1981). Thus, it suffices to show that $d_2(\check{F}_n, H_n) \overset{a.s.}{\to} 0$ as $n \to \infty$. To this end, let J be a random variable having Laplace distribution on $\{1, \ldots, n\}$, with $P(J = i) = 1/n$ for each $i = 1, \ldots, n$. Define two random variables $X^{(1)}$ and $Y^{(1)}$ by

$$X^{(1)} = \varepsilon_J \text{ and } Y^{(1)} = \hat{\varepsilon}_J.$$

Then, $X^{(1)}$ and $Y^{(1)}$ have the marginal distributions H_n and $\check{F}_n$ respectively. Therefore,

$$\begin{aligned}\{d_2(\check{F}_n, H_n)\}^2 &= \inf\{\mathrm{E}|X - Y|^2\} \le \mathrm{E}\{X^{(1)} - Y^{(1)}\}^2 \\ &= n^{-1}\sum_{i=1}^{n}(\varepsilon_i - \check{\varepsilon}_i)^2 = (n\hat{\sigma}_n^2)^{-1}\sum_{i=1}^{n}\{\hat{\sigma}_n\varepsilon_i - (\widehat{\varepsilon}_i - n^{-1}\sum_{j=1}^{n}\widehat{\varepsilon}_j)\}^2,\end{aligned} \tag{A.3}$$

where $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n}\{\widehat{\varepsilon}_i - n^{-1}\sum_{j=1}^{n}\widehat{\varepsilon}_j\}^2$.

Since $\hat{\boldsymbol{\phi}} \overset{a.s.}{\to} \boldsymbol{\phi}_0$ and $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = O_p(1)$, it follows that $\hat{\sigma}_n^2 \overset{a.s.}{\to} 1$. Hence, for some constant $K > 0$, (A.3) is bounded from above by

$$Kn^{-1}\sum_{i=1}^{n}(\widehat{\varepsilon}_i - \varepsilon_i)^2 + Kn^{-2}\left(\sum_{i=1}^{n}\varepsilon_i\right)^2 + M_n \tag{A.4}$$

where M_n is a random variable that converges to zero with probability one. Here we have used some arguments from the proof of Lemma 6 in Perera and Silvapulle (2021).

Furthermore, for some $K < \infty$, we have that

$$n^{-1}\sum_{i=1}^{n}(\widehat{\varepsilon}_i - \varepsilon_i)^2 \;\le\; Kn^{-1}\sum_{i=1}^{n}\varepsilon_i^2\left[\{h_i(\boldsymbol{\phi}_0)\}^{1/2} - \{h_i(\hat{\boldsymbol{\phi}})\}^{1/2}\right]^2.$$

From Proposition A.1, we obtain that

$$\{h_i(\hat{\boldsymbol{\phi}})\}^{1/2} - \{h_i(\boldsymbol{\phi}_0)\}^{1/2} = 2^{-1}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'\dot{h}_i(\boldsymbol{\phi}_0)/\{h_i(\boldsymbol{\phi}_0)\}^{1/2} + o_p(n^{-1/2}). \tag{A.5}$$

Because $\hat{\boldsymbol{\phi}} \overset{a.s.}{\to} \boldsymbol{\phi}_0$ and $\|\mathrm{E}\varepsilon_1^2\dot{h}_i(\boldsymbol{\phi}_0)/\{h_i(\boldsymbol{\phi}_0)\}^{1/2}\| < \infty$, then we have that

$$n^{-1}\sum_{i=1}^{n}\varepsilon_i^2[\{h_i(\hat{\boldsymbol{\phi}})\}^{1/2} - \{h_i(\boldsymbol{\phi}_0)\}^{1/2}] \overset{a.s.}{\to} 0 \text{ as } n \to \infty.$$

Consequently, $d_2(\check{F}_n, H_n) \overset{a.s.}{\to} 0$ and hence $d_2(\check{F}_n, F_0) \overset{a.s.}{\to} 0$.

Proof of part (b): Under Assumptions (A1)–(A5), from Proposition A.1 below, we obtain that (A.5) holds irrespective of whether $\boldsymbol{\phi}_0$ is in the interior of Φ. Since $\hat{\boldsymbol{\phi}} \overset{a.s.}{\to} \boldsymbol{\phi}_0$, and $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = O_p(1)$, then it follows that $d_2(\check{F}_n, F_0) \overset{a.s.}{\to} 0$ by repeating the arguments of the proof of part (a). □

Lemma A.2. *Suppose that either Assumption (A1) or Assumption (B1) is satisfied. Additionally, assume that Assumption (B2) holds. Then, for every* $\zeta = (\boldsymbol{\phi}, F) \in \bar{\Phi} \times \mathcal{F}_\delta^0$

a) *The model (A.2) has a unique stationary ergodic solution* $\{Y_i^{(\zeta)} : i \in \mathbb{Z}\}$,

b) $\mathrm{E}[|Y_i^{(\zeta)}|^{4+d}]$, $\mathrm{E}[|h_i^{(\zeta)}(\boldsymbol{\phi})|^{2+d}]$ *and* $\mathrm{E}[\|\tau_i^{(\zeta)}(\boldsymbol{\phi})\|^{2+d}]$ *are finite for some* $d > 0$.

Proof. Let $\zeta = (\boldsymbol{\phi}, F) \in \bar{\Phi} \times \mathcal{F}_\delta^0$ be fixed and arbitrary. Since F has zero mean and unit variance, the condition $\sum_{i=1}^{p_1} \alpha_i + \sum_{j=1}^{p_2} \beta_j < 1$ is necessary and sufficient for the process $\{Y_i^{(\zeta)}; i \in \mathbb{Z}\}$ to be strictly stationary and have finite second moments with $\mathrm{E}(Y_i^{(\zeta)}) = 0$ and $\mathrm{E}[\{Y_i^{(\zeta)}\}^2] = \omega/(1 - \sum_{i=1}^{p_1} \alpha_i - \sum_{j=1}^{p_2} \beta_j)$; see for example Nelson (1990) and Chen and An (1998). Therefore, if H_0 holds, then part (a) follows from Assumption (A1), and if H_1 holds, then it follows from Assumption (B1).

Since $\varepsilon_i^{(F)} = F^{-1}(U_i)$ with $\{U_i, i \in \mathbb{Z}\}$ being i.i.d. uniform(0,1) random variables, from Assumption (B2), we have $\mathrm{E}([\{\varepsilon_i^{(F)}\}^2]^{2+d}) < \infty$, and hence part (b) also follows. □

Lemma A.3. *Suppose that either Assumption (A1) or Assumption (B1) is satisfied. Additionally, assume that Assumptions (B2)–(B4) hold. Then, for every nonrandom sequence* $\zeta_n := (\boldsymbol{\phi}_n, F_n) \to \zeta_0^* := (\boldsymbol{\phi}_0^*, F_0^*)$, *where* $\zeta_n \in \bar{\Phi} \times \mathcal{F}$, *we have that*

$$| \, n^{1/2}(\hat{\boldsymbol{\phi}}_{nn} - \boldsymbol{\phi}_n) - \Sigma_{nn}^{-1}(\boldsymbol{\phi}_n) n^{-1/2} \sum_{i=1}^{n} \tau_{ni}(\boldsymbol{\phi}_n)(\varepsilon_{ni}^2 - 1) \, | = o_p(1), \tag{A.6}$$

with $\Sigma_{nn}(\boldsymbol{\phi}) := n^{-1} \sum_{i=1}^{n} \tau_{ni}(\boldsymbol{\phi}) \tau_{ni}(\boldsymbol{\phi})'$, *and* $\varepsilon_{ni} = F_n^{-1}(U_i)$, *where* $\hat{\boldsymbol{\phi}}_{nn}$ *is the analogue of* $\hat{\boldsymbol{\phi}}$ *for the data generating process at* $(\boldsymbol{\phi}_n, F_n)$, *and* $\boldsymbol{U} = \{U_i, i \in \mathbb{Z}\}$ *denote a sequence of i.i.d. random variables from the uniform(0,1) distribution.*

Proof. The proof follows from arguing as in the proof of Lemma 4(a) in Perera and Silvapulle (2020) for the GARCH(p_1, p_2) setup. □

Lemma A.4. *Suppose that either Assumption (A1) or Assumption (B1) is satisfied. Additionally, assume that Assumptions (B2)–(B4) hold. Then, for every constant* $C < \infty$,
i) $\sup |h_{ni}^{1/2}(\boldsymbol{b}) - h_{ni}^{1/2}(\boldsymbol{a}) - 2^{-1}(\boldsymbol{b} - \boldsymbol{a})' \dot{h}_{ni}(\boldsymbol{a}) h_{ni}^{-1/2}(\boldsymbol{a})| h_{ni}^{-1/2}(\boldsymbol{a}) = o_p(n^{-1/2})$,
ii) $\sup | \, h_{ni}(\boldsymbol{b}) - h_{ni}(\boldsymbol{a}) - (\boldsymbol{b} - \boldsymbol{a})' \dot{h}_{ni}(\boldsymbol{a}) \, | \, h_{ni}^{-1}(\boldsymbol{a}) = o_p(n^{-1/2})$,
where the supremum is taken over $1 \le i \le n$ *and over* $\{(\boldsymbol{b}, \boldsymbol{a}) : \boldsymbol{b}, \boldsymbol{s} \in \bar{\Phi}, \sqrt{n}\|\boldsymbol{b} - \boldsymbol{a}\| \le C\}$.

Proof of Lemma A.4. Let $D := \bar{\Phi} \times \mathcal{F}_\delta^0$, $p = p_1 + p_2 + 1$. Let $\wp_{ni}(\boldsymbol{\phi}) = \{h_{ni}(\boldsymbol{\phi})\}^{1/2}$ for $\boldsymbol{\phi} \in \bar{\Phi}$. Let $\Delta_{ni}(\boldsymbol{a}, \boldsymbol{b}) := h_{ni}^{1/2}(\boldsymbol{b}) - h_{ni}^{1/2}(\boldsymbol{a}) - 2^{-1}(\boldsymbol{b} - \boldsymbol{a})' \dot{h}_{ni}(\boldsymbol{a}) h_{ni}^{-1/2}(\boldsymbol{a})$. Let $\boldsymbol{a}, \boldsymbol{b} \in \bar{\Phi}$ be fixed but arbitrary. Then, for each $n \in \mathbb{N}$, there exists $\boldsymbol{\delta}_{n1} \in \mathbb{R}^p$, such that $\boldsymbol{b} = \boldsymbol{a} + n^{-1/2}\boldsymbol{\delta}_{n1}$. Hence, $\Delta_{ni}(\boldsymbol{a}, \boldsymbol{b}) = \wp_{ni}(\boldsymbol{b}) - \wp_{ni}(\boldsymbol{a}) - n^{-1/2}\boldsymbol{\delta}_{n1}' \dot{\wp}_{ni}(\boldsymbol{a})$. Therefore, by the Mean Value Theorems for functions from $\mathbb{R}^p$ to $\mathbb{R}$, and $\mathbb{R}^p$ to $\mathbb{R}^p$, with right partial derivatives, for every $n \in \mathbb{N}$, there exist $\boldsymbol{\delta}_{n2}, \boldsymbol{\delta}_{n3} \in \mathbb{R}^p$ with $\|\boldsymbol{\delta}_{n3}\| \le \|\boldsymbol{\delta}_{n2}\| \le \|\boldsymbol{\delta}_{n1}\|$, such that

$$\Delta_{ni}(\boldsymbol{a}, \boldsymbol{b}) \;=\; n^{-1/2}\boldsymbol{\delta}_{n1}' \left[\dot{\wp}_{ni}(\boldsymbol{a} + n^{-1/2}\boldsymbol{\delta}_{n2}) - \dot{\wp}_{ni}(\boldsymbol{a})\right] \;=\; n^{-1}\boldsymbol{\delta}_{n1}' \ddot{\wp}_{ni}(\boldsymbol{a} + n^{-1/2}\boldsymbol{\delta}_{n3})\boldsymbol{\delta}_{n2},$$

where

$$\ddot{\wp}_{ni}(\boldsymbol{\phi}) = \frac{1}{2}\frac{\ddot{h}_{ni}(\boldsymbol{\phi})}{\{h_{ni}(\boldsymbol{\phi})\}^{1/2}} - \frac{1}{4}\frac{\tau_{ni}(\boldsymbol{\phi})\dot{h}_{ni}'(\boldsymbol{\phi})}{\{h_{ni}(\boldsymbol{\phi})\}^{1/2}}.$$

Therefore, for any given constant $C > 0$, w.p. 1,

$$\max_{1\le i\le n} \sup_{\boldsymbol{a},\boldsymbol{b}\in\bar{\Phi}, \sqrt{n}\|\boldsymbol{b}-\boldsymbol{a}\|\le C} n^{1/2}|\Delta_{ni}(\boldsymbol{a}, \boldsymbol{b})| \;\le\; n^{-1/2}C^2 \max_{1\le i\le n} \|\ddot{\wp}_{ni}\|_{\bar{\Phi}}. \tag{A.7}$$

For example, for the GARCH(1,1) case, by differentiation, we have

$$\dot{h}_{ni}(\boldsymbol{\phi}) = \{(1-\beta)^{-1}, \sum_{j=1}^{\infty} \beta^{j-1} Y_{n(i-j)}^2, \omega(1-\beta)^{-2} + \alpha \sum_{j=2}^{\infty} (j-1)\beta^{j-2} Y_{n(i-j)}^2\}',$$

and the 3×3 matrix $\ddot{h}_{ni}(\boldsymbol{\phi})$ is given by

$$\begin{aligned}
&[\ddot{h}_{ni}]_{rk}(\boldsymbol{\phi}) = 0 \text{ for } r = 1, 2,\ k = 1, 2, \quad [\ddot{h}_{ni}]_{13}(\boldsymbol{\phi}) = [\ddot{h}_{ni}]_{31}(\boldsymbol{\phi}) = (1-\beta)^{-2},\\
&[\ddot{h}_{ni}]_{23}(\boldsymbol{\phi}) = [\ddot{h}_{ni}]_{32}(\boldsymbol{\phi}) = \sum_{j=2}^{\infty}(j-1)\beta^{j-2}Y_{n(i-j)}^2, \quad \text{and}\\
&[\ddot{h}_{ni}]_{33}(\boldsymbol{\phi}) = 2\omega(1-\gamma)^{-3} + \alpha\sum_{j=3}^{\infty}(j-2)(j-1)\beta^{j-3}Y_{n(i-j)}^2.
\end{aligned}$$

One can similarly obtain the derivatives for the GARCH(p_1, p_2) for any given p_1, p_2.

Since $\sup_{\zeta\in D} \|h_i^{(\zeta)}(\cdot)\|_{\bar{\Phi}} > \omega_L > 0$, then part (i) follows under (B2)–(B4), by arguing as in the proof of Lemma 4 in Perera and Silvapulle (2021) applying Chebyshev's inequality. The part (ii) follows similarly under (B2)–(B4). $\square$

The next lemma follows from the proof of Theorem 13.1 in Billingsley (1968) and an application of the Cauchy-Schwarz inequality; see also Lemma 5.1 in Stute (1997). We restate this result here for the ease of reference. This lemma is useful for establishing the tightness of certain processes.

Lemma A.5. *Let $\{(a_i, b_i); 1 \le i \le n\}$ be i.i.d. square-integrable bivariate random vectors with $\mathrm{E}(a_i) = \mathrm{E}(b_i) = 0$, $1 \le i \le n$. Then we have that $\mathrm{E}\{(\sum_{i=1}^n a_i)^2(\sum_{j=1}^n b_j)^2\} \le n\mathrm{E}(a_1^2 b_1^2) + 3n(n-1)\mathrm{E}(a_1^2)\mathrm{E}(b_1^2)$.*

For the proof of Lemma 1 we make use of the following proposition.

Proposition A.1. *Suppose that the assumptions of Lemma 1 hold. Then, for every $K < \infty$,*
(a) $n^{1/2} \sup \mid h_i^{1/2}(t) - h_i^{1/2}(s) - 2^{-1}(t-s)'\dot{h}_i(s)h_i^{-1/2}(s) \mid h_i^{-1/2}(\boldsymbol{\phi}_0) = o_p(1)$,
(b) $n^{1/2} \sup \mid h_i(t) - h_i(s) - (t-s)'\dot{h}_i(s) \mid h_i^{-1}(\boldsymbol{\phi}_0) = o_p(1)$, *where the supremum is taken over* $1 \le i \le n$ *and over* $\{(t,s) : t, s \in \Phi, \sqrt{n}\|t-s\| \le K\}$.

Proof. The proof follows as a special cases of the proof of Lemma A.4. $\square$

A.3 Initialization effect

In this section we obtain several technical results to establish that the effect of initialization in the bootstrap data generation is asymptotically negligible. First let us introduce some notation.

Notation A: If data are generated from (A.2) for $i \ge -m$, conditional on a vector of starting values $\varsigma_0 = (y_0, \cdots, y_{1-q}, s_0, \cdots, s_{1-p})'$, then we use the superscript "(m,ζ)" instead of "(ζ)", where $\zeta = (\boldsymbol{\phi}, F)$. For example, $h_i^{(m,\zeta)}(\cdot)$ and $\tau_i^{(m,\zeta)}(\cdot)$ are the analogues of $h_i^{(\zeta)}(\cdot)$ and $\tau_i^{(\zeta)}(\cdot)$, respectively, when the data generating model obeys (A.2) for $i \ge -m$, conditional on the starting values ς_0.

Lemma A.6. *Suppose that either Assumption (A1) or Assumption (B1) is satisfied. Additionally, assume that Assumption (B2) holds. Then, there exists a compact set $K_1 \subseteq \bar{\Phi}$, which contains $\boldsymbol{\phi}_0^*$, such that the following hold for some $\delta > 0$ with $K = K_1 \times \mathcal{F}_\delta^0$:*

a) $\sup_{\zeta \in K} \|h_i^{(m,\zeta)} - h_i^{(\zeta)}\|_{K_1}$, $\sup_{\zeta \in K} \|\dot{h}_i^{(m,\zeta)} - \dot{h}_i^{(\zeta)}\|_{K_1}$, $\overset{e.a.s.}{\to} 0$ *as* $i \to \infty$;

b) $\sup_{\zeta \in K} \mathrm{E}\|\ddot{h}_0^{(\zeta)}\|_{K_1}^{2+d}$ *and* $\sup_{\zeta \in K} \mathrm{E}\|\tau_0^{(\zeta)}\|_{K_1}^{2+d}$ *are finite for some* $d > 0$.

Proof. The proof follows from arguing as in the verifications of Conditions (M1) and (M2) in Perera and Silvapulle (2020) for the GARCH(p_1, p_2) setup, with the set $K_{\boldsymbol{\theta}}$ being replaced by $\mathcal{F}_\delta^0$. Sine everything follows after suitable modifications of the arguments already developed in Perera and Silvapulle (2020), we omit the details. $\square$

In view of Notation A above, conditional on $(Y_1, \ldots, Y_n)$, we have

$$\{Y_i^*, h_i^*(\boldsymbol{\phi}), \tau_i^*(\boldsymbol{\phi})\} \equiv \{Y_i^{(0,\zeta^*)}, h_i^{(0,\zeta^*)}(\boldsymbol{\phi}), \tau_i^{(0,\zeta^*)}(\boldsymbol{\phi})\}, \quad \boldsymbol{\phi} \in \Phi, i \in \mathbb{N}, \tag{A.8}$$

where $\zeta^* = (\hat{\boldsymbol{\phi}}, \check{F}_n)$ for the bootstrap method in Section 3.2, and $\zeta^* = (\hat{\boldsymbol{\phi}}^\dagger, \check{F}_n)$ for the bootstrap method in Section 4.3 with $\hat{\boldsymbol{\phi}}^\dagger$ given by (19). Similarly, let the bootstrap process generated by (A.2), conditional on $(Y_1, \ldots, Y_n)$, without any initialization, be defined as

$$\{Y_i^{*(\infty)}, h_i^{*(\infty)}(\boldsymbol{\phi}), \tau_i^{*(\infty)}(\boldsymbol{\phi})\} = \{Y_i^{(\zeta^*)}, h_i^{(\zeta^*)}(\boldsymbol{\phi}), \tau_i^{(\zeta^*)}(\boldsymbol{\phi})\}, \ \boldsymbol{\phi} \in \Phi, i \in \mathbb{Z}, \tag{A.9}$$

where $\zeta^* = (\hat{\boldsymbol{\phi}}, \check{F}_n)$ and $\zeta^* = (\hat{\boldsymbol{\phi}}^\dagger, \check{F}_n)$ for the bootstraps in Section 3.2 and Section 4.3, respectively. Thus, $\{Y_i^{*(\infty)}, h_i^{*(\infty)}(\boldsymbol{\phi}), \tau_i^{*(\infty)}(\boldsymbol{\phi})\}$ represents a hypothetical (non-operational) bootstrap process generated by (A.2), without any initialization. Further, let

$$\hat{\boldsymbol{\phi}}^{*(\infty)} = \arg\min_{\boldsymbol{\phi}\in\Phi} \sum_{i=1}^n \ell_i^{*(\infty)}(\boldsymbol{\phi}), \quad \ell_i^{*(\infty)}(\boldsymbol{\phi}) = \log h_i^{*(\infty)}(\boldsymbol{\phi}) + \frac{\{Y_i^{*(\infty)}\}^2}{h_i^{*(\infty)}(\boldsymbol{\phi})},$$

$$\mathcal{U}_n^{*(\infty)}(y, \boldsymbol{\phi}) = n^{-1/2} \sum_{i=1}^n \left(\frac{\{Y_i^{*(\infty)}\}^2}{h_i^{*(\infty)}(\boldsymbol{\phi})} - 1 \right) \mathbb{I}(Y_{i-1}^{*(\infty)} \le y), \quad y \in \mathbb{R}, \ \boldsymbol{\phi} \in \Phi.$$

Since no initialization is used in the bootstrap data generation in (A.9), the marked empirical process $\mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)})$, unlike $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*)$, is not subject to any initialization error.

The next lemma shows that, conditional on $(Y_1, \ldots, Y_n)$, for the bootstrap method in Section 3.2, $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*)$ and $\mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)})$ are uniformly close, in probability.

Lemma A.7. *Suppose that the assumptions of Theorem 2 are satisfied with $\boldsymbol{\phi}_0$ being an interior point in Φ, or the assumptions of Theorem 3 are satisfied. Then, conditional on $(Y_1, \ldots, Y_n)$, $\sup_{y\in\mathbb{R}} | \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)}) | = o_p^*(1)$, in probability.*

Proof. Assume without loss of generality that $Y_i^* \le Y_i^{*(\infty)}$. Then, we have

$$\begin{aligned} & \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)}) \\ = \ & n^{-1/2} \sum_{i=1}^n \left(\frac{\{Y_i^*\}^2}{h_i^*(\hat{\boldsymbol{\phi}}^*)} - \frac{\{Y_i^{*(\infty)}\}^2}{h_i^{*(\infty)}(\hat{\boldsymbol{\phi}}^{*(\infty)})} \right) \mathbb{I}(Y_{i-1}^{*(\infty)} \le y) \\ & + n^{-1/2} \sum_{i=1}^n \left(\frac{\{Y_i^*\}^2}{h_i^*(\hat{\boldsymbol{\phi}}^*)} - 1 \right) \mathbb{I}(Y_{i-1}^* \le y < Y_{i-1}^{*(\infty)}) \\ = \ & \mathbf{I} + \mathbf{II}, \quad \text{say.} \end{aligned} \tag{A.10}$$

Since Lemma A.6(a) shows that $|h_i^*(\hat{\boldsymbol{\phi}}) - h_i^{*(\infty)}(\hat{\boldsymbol{\phi}})| \overset{e.a.s.}{\to} 0$ as $i \to \infty$, from Lemma 2.1 in Straumann and Mikosch (2006), $|\{h_{i-1}^*(\hat{\boldsymbol{\phi}})\}^{1/2} - \{h_{i-1}^{*(\infty)}(\hat{\boldsymbol{\phi}})\}^{1/2}| \overset{e.a.s.}{\to} 0$, as $i \to \infty$, and hence from Lemma 2.3 in Straumann and Mikosch (2006), it follows that

$$|Y_{i-1}^* - Y_{i-1}^{*(\infty)}| = |\varepsilon_{i-1}^*| |\{h_{i-1}^*(\hat{\boldsymbol{\phi}})\}^{1/2} - \{h_{i-1}^{*(\infty)}(\hat{\boldsymbol{\phi}})\}^{1/2}| \overset{e.a.s.}{\to} 0, \quad \text{as } i \to \infty,$$

and hence the sum $\mathbf{II}$ in (A.10) is of order $o_p^*(1)$, in probability, uniformly in $y \in \mathbb{R}$.

The first sum in (A.10) is bounded as

$$\begin{aligned} |\mathbf{I}| \ & \le \ n^{-1/2} \sum_{i=1}^n \left| \frac{\{Y_i^*\}^2}{h_i^*(\hat{\boldsymbol{\phi}}^*)} - \frac{\{Y_i^{*(\infty)}\}^2}{h_i^{*(\infty)}(\hat{\boldsymbol{\phi}}^{*(\infty)})} \right| \\ & \le \ n^{-1/2} \sum_{i=1}^n \{Y_i^*\}^2 \left| \frac{1}{h_i^*(\hat{\boldsymbol{\phi}}^*)} - \frac{1}{h_i^{*(\infty)}(\hat{\boldsymbol{\phi}}^{*(\infty)})} \right| + n^{-1/2} \omega_L^{-1} \sum_{i=1}^n \left| \{Y_i^*\}^2 - \{Y_i^{*(\infty)}\}^2 \right| \\ & = \ \mathbf{I}_A + \mathbf{I}_B, \quad \text{say.} \end{aligned}$$

Since $|\{Y_i^*\}^2 - \{Y_i^{*(\infty)}\}^2| = |\varepsilon_i^*|^2 |h_i^*(\hat{\boldsymbol{\phi}}) - h_i^{*(\infty)}(\hat{\boldsymbol{\phi}})| \overset{e.a.s.}{\to} 0$, as $i \to \infty$, $\sum_{i=1}^n |\{Y_i^*\}^2 - \{Y_i^{*(\infty)}\}^2|$ converges a.s. as $n \to \infty$, by Lemma 2.1 in Straumann and Mikosch (2006), and hence sum $\mathbf{I}_B$ is of order $o_p^*(1)$.

From the proof of Lemma 8 in Perera and Silvapulle (2017), $n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}^{*(\infty)}) = o_p^*(1)$. By Lemma A.6, $\sup_{\boldsymbol{\phi} \in K_1} |h_i^*(\boldsymbol{\phi}) - h_i^{*(\infty)}(\boldsymbol{\phi})| \overset{e.a.s.}{\to} 0$ for some compact set $K_1 \subseteq \bar{\Phi}$. Hence, Lemma 2.3 in Straumann and Mikosch (2006) yields that sum $\mathbf{I}_A$ is also of order $o_p^*(1)$. Therefore, $\sup_{y \in \mathbb{R}} | \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)}) |= o_p^*(1)$, in probability. □

The next lemma shows that, conditional on $(Y_1, \dots, Y_n)$, for the bootstrap method in Section 4.3, $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*)$ and $\mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)})$ are uniformly close, in probability.

Lemma A.8. *Suppose that either assumptions of Theorem 4 or Theorem 5 are satisfied. Then, conditional on* $\{Y_1, \dots, Y_n\}$, $\sup_{y \in \mathbb{R}} | \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}^{*(\infty)}) |= o_p^*(1)$, *in probability.*

Proof. The proof follows by arguing as in the proof of Lemma A.7 with $\hat{\boldsymbol{\phi}}$ replaced by $\hat{\boldsymbol{\phi}}^\dagger$. □

By Lemmas A.7 and A.8 we obtain that the effect of initialization in the bootstrap data generation is asymptotically negligible. Hence, in the next section we only focus on $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$.

A.4 Main proofs

This section provides the proofs of the main results stated in the paper.

Proof of Lemma 1. First, partition $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$ as follows.

$$\begin{aligned}
&\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}) - \mathcal{U}_n(y, \boldsymbol{\phi}_0) \\
&= n^{-1/2} \sum_{i=1}^n \{Y_i^2 / h_i(\hat{\boldsymbol{\phi}}) - Y_i^2 / h_i(\boldsymbol{\phi}_0))\} \mathbb{I}(Y_{i-1} \le y) \\
&= n^{-1/2} \sum_{i=1}^n \varepsilon_i^2 \{h_i(\boldsymbol{\phi}_0)/h_i(\hat{\boldsymbol{\phi}}) - 1\} \mathbb{I}(Y_{i-1} \le y) \\
&= n^{-1/2} \sum_{i=1}^n -\varepsilon_i^2 \left(\frac{h_i(\hat{\boldsymbol{\phi}}) - h_i(\boldsymbol{\phi}_0)}{h_i(\boldsymbol{\phi}_0)} \right) \mathbb{I}(Y_{i-1} \le y) \\
&\quad + n^{-1/2} \sum_{i=1}^n \varepsilon_i^2 \left(h_i(\hat{\boldsymbol{\phi}}) - h_i(\boldsymbol{\phi}_0) \right) \left(\frac{1}{h_i(\boldsymbol{\phi}_0)} - \frac{1}{h_i(\hat{\boldsymbol{\phi}})} \right) \mathbb{I}(Y_{i-1} \le y).
\end{aligned} \tag{A.11}$$

Since $\mathrm{E}(\varepsilon_0^2) = 1$ and $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = O_p(1)$, by applying Proposition A.1 and the Ergodic Theorem to the expansion (A.11), we obtain that, uniformly in $y \in \mathbb{R}$,

$$\begin{aligned}
\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}) &= \mathcal{U}_n(y, \boldsymbol{\phi}_0) - n^{-1} \sum_{i=1}^n \{\varepsilon_i^2 n^{1/2} (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)' \boldsymbol{\tau}_i(\boldsymbol{\phi}_0)\} \mathbb{I}(Y_{i-1} \le y) + o_p(1) \\
&= \mathcal{U}_n(y, \boldsymbol{\phi}_0) - n^{1/2} (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)' \mathrm{E}[\boldsymbol{\tau}_1(\boldsymbol{\phi}_0) \mathbb{I}(Y_0 \le y)] + o_p(1).
\end{aligned}$$

$\square$

For the proof of Theorem 1 we introduce the following additional notation. For $d \geq 1$, let $\mathbb{C}^d \equiv \mathcal{C}([-\infty, \infty], \mathbb{R}^d)$ denote the space of continuous functions from $[-\infty, \infty]$ into $\mathbb{R}^d$. A sequence of d-dimensional stochastic processes (with *cadlag* paths) is said to be $\mathcal{C}$-tight if it has associated laws that are tight and whose limit points are concentrated on the set of continuous paths $\mathbb{C}^d$.

Proof of Theorem 1. From Lemma 1 and (10), we obtain that, uniformly in $y \in \mathbb{R}$,

$$\begin{aligned}\mathcal{U}_n(y, \hat{\phi}) &= \mathcal{U}_n(y, \phi_0) \\ &\quad - \mathrm{E}[\tau_1'(\phi_0)\mathbb{I}(Y_0 \leq y)]\Sigma_n^{-1}(\phi_0) n^{-1/2} \sum_{i=1}^{n} (\varepsilon_i^2 - 1)\tau_i(\phi_0) + o_p(1). \end{aligned} \tag{A.12}$$

By the Ergodic Theorem (e.g., Theorem 2.5.2 of Giraitis *et al.*, 2012) $\Sigma_n(\phi_0) \overset{a.s.}{\to} \Sigma(\phi_0)$, as $n \to \infty$. Hence, by using the above asymptotic uniform expansion of $\mathcal{U}_n(\cdot, \hat{\phi})$, we derive that

$$\begin{aligned}\mathrm{Cov}\{\mathcal{U}_n(x, \hat{\phi}), \mathcal{U}_n(y, \hat{\phi})\} &= K(x, y) + J'(x, \phi_0)\mathrm{E}[M_1(\phi_0)M_1'(\phi_0)]J'(y, \phi_0) \\ &\quad - J'(x, \phi_0)\mathrm{E}[(\varepsilon_1^2 - 1)M_1(\phi_0)\mathbb{I}(Y_0 \leq y)] \\ &\quad - J'(y, \phi_0)\mathrm{E}[(\varepsilon_1^2 - 1)M_1(\phi_0)\mathbb{I}(Y_0 \leq x)] + o(1), \quad x, y \in \mathbb{R}.\end{aligned}$$

Hence, $\mathrm{Cov}\{\mathcal{U}_n(x, \hat{\phi}), \mathcal{U}_n(y, \hat{\phi})\} = \mathrm{Cov}\{\mathcal{U}_0(x), \mathcal{U}_0(y)\} + o(1)$, $x, y \in \mathbb{R}$, where $\mathcal{U}_0$ is the centred Gaussian process in Theorem 1. The convergence of finite dimensional distributions of $\mathcal{U}_n(\cdot, \hat{\phi})$ can be derived by e.g. an application of Theorem 18.3 in Billingsley (1999).

To show that $\mathcal{U}_n(y, \hat{\phi})$ is tight, let $G^{-1}(u) := \inf\{y \in \mathbb{R} : G(y) \geq u\}$ and

$$\bar{\mathcal{U}}_n(u, \phi) := n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{Y_i^2}{h_i(\phi)} - 1 \right\} \mathbb{I}(Y_{i-1} \leq G^{-1}(u)), \quad u \in [0, 1], \ \phi \in \Phi.$$

Then, by standard quantile representation, we have that $\mathcal{U}_n(y, \phi) = \bar{\mathcal{U}}_n(G(y), \phi)$ for $y \in \bar{\mathbb{R}}$. Let $0 \leq u_1 \leq u \leq u_2 \leq 1$ be fixed but arbitrary. Set

$$\begin{aligned} a_i &= \{\varepsilon_i^2 - 1\}\mathbb{I}(G^{-1}(u_1) < Y_{i-1} \leq G^{-1}(u)), \\ b_i &= \{\varepsilon_i^2 - 1\}\mathbb{I}(G^{-1}(u) < Y_{i-1} \leq G^{-1}(u_2)). \end{aligned}$$

Note that $\mathrm{E}(a_i) = \mathrm{E}(b_i) = 0$ and $a_i b_i = 0$. Further, $\bar{\mathcal{U}}_n(u, \phi_0) - \bar{\mathcal{U}}_n(u_1, \phi_0) = n^{-1/2} \sum_{i=1}^{n} a_i$ and $\bar{\mathcal{U}}_n(u_2, \phi_0) - \bar{\mathcal{U}}_n(u, \phi_0) = n^{-1/2} \sum_{j=1}^{n} b_j$. Therefore, from Lemma A.5, we obtain that

$$\begin{aligned} &\mathrm{E}[\{\bar{\mathcal{U}}_n(u, \phi_0) - \bar{\mathcal{U}}_n(u_1, \phi_0)\}\{\bar{\mathcal{U}}_n(u_2, \phi_0) - \bar{\mathcal{U}}_n(u, \phi_0)\}] \\ &= n^{-2}\mathrm{E}\{(\sum_{i=1}^{n} a_i)^2 (\sum_{j=1}^{n} b_j)^2\} \ \leq \ [n\mathrm{E}(a_1^2 b_1^2) + 3n(n-1)\mathrm{E}(a_1^2)\mathrm{E}(b_1^2)]/n^2 \\ &= 3n(n-1)n^{-2}[\mathrm{E}\{\varepsilon_1^2 - 1\}^2]^2 (u - u_1)(u_2 - u) \\ &\leq 3(\kappa_\varepsilon - 1)^2 (u_2 - u_1)^2. \end{aligned}$$

Since u_1 and u_2 are arbitrary, then it follows that $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ is $\mathcal{C}$-tight, e.g. by Theorem 15.7 of Billingsley (1968). Since G is continuous, the last term in (A.12),

$$\mathrm{E}[\tau_1'(\boldsymbol{\phi}_0)\mathbb{I}(Y_0 \leq y)]c\Sigma_n^{-1}(\boldsymbol{\phi}_0)n^{-1/2}\sum_{i=1}^n \tau_i(\boldsymbol{\phi}_0)\varrho(\varepsilon_i),$$

is asymptotically $\mathcal{C}$-tight, and hence $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}}_0)$ is also asymptotically $\mathcal{C}$-tight. From the latter fact and the convergence of finite dimensional distributions of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$ to those of $\mathcal{U}_0(\cdot)$, it follows that $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}}) \overset{w}{\Rightarrow} \mathcal{U}_0(\cdot)$ in $\mathcal{D}(\mathbb{R})$, where '$\overset{w}{\Rightarrow}$' denotes weak convergence of processes. □

We next state the proof of Theorem 2. In view of Lemmas A.7 and A.8, and the continuous mapping theorem, the effect of initialization in the bootstrap data generation is asymptotically negligible. Therefore, in the next proof, and in the sequel, we do not distinguish between $\{Y_i^{*(\infty)}, h_i^{*(\infty)}(\boldsymbol{\phi}), \tau_i^{*(\infty)}(\boldsymbol{\phi})\}$ in (A.9), and the operational bootstrap process $\{Y_i^*, h_i^*(\boldsymbol{\phi}), \tau_i^*(\boldsymbol{\phi})\}$ in (A.8).

Proof of Theorem 2. By extending the arguments of Lemma 1 to a triangular array setup, we obtain that, conditional on $\{Y_1, \ldots, Y_n\}$, uniformly over $y \in \mathbb{R}$,

$$\begin{aligned}
\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*) &= \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}) - \frac{1}{n}\sum_{i=1}^n (\varepsilon_i^*)^2 n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}})'\tau_i^*(\hat{\boldsymbol{\phi}})\mathbb{I}(Y_{i-1}^* \leq y) + o_p^*(1), \\
&= \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}) - \frac{1}{\sqrt{n}}\sum_{i=1}^n (\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}})'\tau_i^*(\hat{\boldsymbol{\phi}})\mathbb{I}(Y_{i-1}^* \leq y) + o_p^*(1),
\end{aligned}$$

in probability. Since $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}) = n^{-1/2}\sum_{i=1}^n (\varepsilon_i^* - 1)^2\mathbb{I}(Y_{i-1}^* \leq y)$, by using Assumption (B3), for every $x, y \in \mathbb{R}$, with $x \wedge y := \min(x, y)$, we derive that

$$\begin{aligned}
\mathrm{cov}^*\{\mathcal{U}_n^*(x, \hat{\boldsymbol{\phi}}), \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}})\} &= n^{-1}\sum_{i=1}^n \mathrm{E}^*(\varepsilon_i^* - 1)^2\mathbb{I}(Y_{i-1}^* \leq x \wedge y) \\
&= (\kappa_\varepsilon - 1)\mathrm{E}\mathbb{I}(Y_{i-1} \leq x \wedge y) + o_p(1) \\
&= K(x, y) + o_p(1).
\end{aligned}$$

Further, by arguing as for Theorem 1 in a triangular array, we also obtain that

$$\mathrm{Cov}^*\{\mathcal{U}_n^*(x, \hat{\boldsymbol{\phi}}^*), \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*)\} = K(x, y) + g^*(x, y, \boldsymbol{\phi}_0) + o_p(1), \quad x, y \in \mathbb{R},$$

where

$$\begin{aligned}
g^*(x, y, \boldsymbol{\phi}_0) &= J'(x, \boldsymbol{\phi}_0)\mathrm{E}[M_1(\boldsymbol{\phi}_0)M_1'(\boldsymbol{\phi}_0)]J'(y, \boldsymbol{\phi}_0) \\
&\quad -J'(x, \boldsymbol{\phi}_0)\mathrm{E}[(\varepsilon_1^2 - 1)M_1(\boldsymbol{\phi}_0)\mathbb{I}(Y_0 \leq y)] \\
&\quad -J'(y, \boldsymbol{\phi}_0)\mathrm{E}[(\varepsilon_1^2 - 1)M_1(\boldsymbol{\phi}_0)\mathbb{I}(Y_0 \leq x)].
\end{aligned}$$

By using the Cramer-Wold device and a CLT for triangular arrays of row-wise independent mean zero r.v.'s (e.g., Corollary 3.3.1 of Hall and Heyde, 1980) we obtain that the finite dimensional distributions of $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ converge to those of $\mathcal{U}_0$, in probability, where $\mathcal{U}_0$ is the centred Gaussian process in Theorem 1. Further, by extending the arguments of Theorem 1 to a triangular array we also obtain that $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ is asymptotically $\mathcal{C}$-tight. Hence the part 1 follows. The part 2 follows from an application of the continuous mapping theorem. Since $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ converges weakly, and $n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}) = O_p^*(1)$, in probability, part 3 also holds for the KS and CvM functional forms. □

Proof of Theorem 3. If some components of $\boldsymbol{\phi}_0^*$ lie on the boundary of the parameter space; i.e., $\phi_{0i}^* = 0$ for some $i = 2, \ldots, p_1 + p_2 + 1$, then the proof follows from arguing as in the proof of Theorem 5. Hence, here we only consider the case $\boldsymbol{\phi}_0^*$ is in the interior of Φ. Since Assumptions (B1)–(B4) hold, and $(\hat{\boldsymbol{\phi}}, \check{F}_n) \xrightarrow{p} (\boldsymbol{\phi}_0^*, F_0^*)$, except that $(\boldsymbol{\phi}_0^*, F_0^*)$ is the pseudo-true value under H_1, by arguing as in the proof of Theorem 2, conditional on $\{Y_1, \ldots, Y_n\}$, the process $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ converges weakly to the centred Gaussian process $\mathcal{U}_0^\dagger(\cdot)$ specified by the covariance kernel

$$
\begin{aligned}
Cov\{\mathcal{U}_0^\dagger(x), \mathcal{U}_0^\dagger(y)\} \;=\;& \mathrm{E}(\{F_0^{*-1}(U_i)\}^2 - 1)^2 \mathbb{I}(Y_{i-1} \le x \wedge y) \\
&+ J'(x, \boldsymbol{\phi}_0^*) \mathrm{E}[V_i(\boldsymbol{\phi}_0) V_i'(\boldsymbol{\phi}_0^*)] J'(y, \boldsymbol{\phi}_0^*) \\
&- J'(x, \boldsymbol{\phi}_0^*) \mathrm{E}[(\{F_0^{*-1}(U_i)\}^2 - 1) V_i(\boldsymbol{\phi}_0^*) \mathbb{I}(Y_{i-1} \le y)] \\
&- J'(y, \boldsymbol{\phi}_0^*) \mathrm{E}[(\{F_0^{*-1}(U_i)\}^2 - 1) V_i(\boldsymbol{\phi}_0^*) \mathbb{I}(Y_{i-1} \le x)],
\end{aligned}
$$

in probability, where $\boldsymbol{U} = \{U_i, i \in \mathbb{Z}\}$ are i.i.d. uniform(0,1) random variables,

$$
V_i(\boldsymbol{\phi}) := -\Sigma^{-1}(\boldsymbol{\phi})[1 - \{F_0^{*-1}(U_i)\}^2] \tau_i(\boldsymbol{\phi}), \quad \boldsymbol{\phi} \in \Phi.
$$

Therefore, it suffices to show that $n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})| = O_p(1)$, for some $y \in \mathbb{R}$ satisfying Assumption (B5). Fix such a y. Under H_1, $\varepsilon_i = Y_i/\sqrt{h_i}$, where $h_i = \mathrm{E}[Y_i^2 \mid \mathcal{H}_{i-1}]$, $i \in \mathbb{Z}$. Therefore, we have that

$$
\begin{aligned}
n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})| \;\le\;& \Big| n^{-1} \sum_{i=1}^n \varepsilon_i^2 h_i \big[h_i^{-1}(\hat{\boldsymbol{\phi}}) - h_i^{-1}(\boldsymbol{\phi}_0^*)\big] \mathbb{I}(Y_{i-1} \le y) \Big| \\
&+ \Big| n^{-1} \sum_{i=1}^n \big\{ \varepsilon_i^2 \big(h_i / h_i(\boldsymbol{\phi}_0^*)\big) - 1 \big\} \mathbb{I}(Y_{i-1} \le y) \Big|. \qquad \text{(A.13)}
\end{aligned}
$$

Lemma A.4 holds under Assumptions (B1)–(B4), with $\boldsymbol{\phi}_0^*$ denoting the pseudo-true value under H_1. Therefore, on the set $n^{1/2}\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*\| \le K$, the first term on the right hand side of (A.13) is bounded from the above by $\omega_L^{-1} n^{-1} \sum_{i=1}^n Y_i^2 \big|(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*)' \tau_i(\boldsymbol{\phi}_0^*)\big| + o_p(n^{-1/2}) = o_p(1)$, by the Ergodic Theorem, and because $\hat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}_0^*$. Since $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*) = O_p(1)$, then by using an extended Glivenko-Cantelli type argument, we obtain that

$$
n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})| \;=\; \big|\mathrm{E}\big([h_1/h_1(y, \boldsymbol{\phi}_0^*) - 1] \mathbb{I}(Y_0 \le y)\big)\big| + o_p(1), \text{ under } \mathsf{H}_1.
$$

Hence, the proof follows from Assumption (B5). $\square$

Proof of Theorem 4. Irrespective of whether $\boldsymbol{\phi}_0$ is in the interior or on the boundary of the parameter space, by arguing as in the proof of Lemma 1, we obtain that

$$\sup_{y\in\mathbb{R}} \mid \mathcal{U}_n(y,\hat{\boldsymbol{\phi}}) - \mathcal{U}_n(y,\boldsymbol{\phi}_0) + n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0)' J(y,\boldsymbol{\phi}_0) \mid = o_p(1), \tag{A.14}$$

with $\mathcal{U}_n(\cdot,\boldsymbol{\phi}_0)$ converging weakly to a centred Gaussian process with covariance kernel $K(x,y) = (\kappa_\varepsilon - 1)G(x\wedge y),\ x,y\in\mathbb{R}$.

To establish the validity of the bootstrap tests we first consider the case $\boldsymbol{\phi}_0$ is in the interior of the parameter space. Since $\hat{\boldsymbol{\phi}}$ is the QMLE in (5),

$$n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0) = Z_n + o_p(1), \tag{A.15}$$

where

$$Z_n := -\Sigma_n^{-1}(\boldsymbol{\phi}_0) n^{-1/2} \sum_{i=1}^{n} (1-\varepsilon_i^2)\tau_i(\boldsymbol{\phi}_0), \quad \Sigma_n(\boldsymbol{\phi}) := n^{-1}\sum_{i=1}^{n} \tau_i(\boldsymbol{\phi})\tau_i(\boldsymbol{\phi})',$$

and hence, with the asymptotic uniform expansion of $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ in (A.14), it follows as in the proof of Theorem 1 that $\mathcal{U}_n(\cdot,\hat{\boldsymbol{\phi}})$ converges weakly to $\mathcal{U}_0(\cdot)$ in $\mathcal{D}(\mathbb{R})$, where $\mathcal{U}_0$ is the centred Gaussian process given in Theorem 1.

In the method of bootstrap data generation outlined in Section 4.3 the transformed estimator $\hat{\boldsymbol{\phi}}^{\dagger}$ plays the role of the true parameter $\boldsymbol{\phi}_0$; recall that $\hat{\boldsymbol{\phi}}^{\dagger} = \hat{\boldsymbol{\phi}}_n^{\dagger} = (\hat{\boldsymbol{\phi}}_{n1}^{\dagger},\ldots,\hat{\boldsymbol{\phi}}_{n(1+p_1+p_2)}^{\dagger})'$ where $\hat{\boldsymbol{\phi}}_{ni}^{\dagger} := \hat{\boldsymbol{\phi}}_{ni}\mathbb{I}(\hat{\boldsymbol{\phi}}_{ni} > c_n)$, $i = 1,2,\ldots,1+p_1+p_2$, and (c_n) is a non-random sequence with $c_n \to 0$ and $n^{1/2}c_n \to \infty$ as $n\to\infty$. Let $A_{ni} = \{\hat{\boldsymbol{\phi}}_{ni} > c_n\},\ i = 1,2,\ldots,1+p_1+p_2$. Since $\boldsymbol{\phi}_0$ is an interior point, we have $\boldsymbol{\phi}_{0j} > 0$, $j = 1,\ldots,1+p_1+p_2$. Further, as c_n converges to 0 at a rate slower than $n^{-1/2}$ and $n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0) = O_p(1)$, we obtain that $P(\cap_{i=1}^{1+p_1+p_2} A_{ni}) \to 1$ as $n\to\infty$. Since $\hat{\boldsymbol{\phi}}^{\dagger} = \hat{\boldsymbol{\phi}}$ on the set $\cap_{i=1}^{1+p_1+p_2} A_{ni}$, then the asymptotic validity of the bootstrap tests follows from the same arguments used in the proof of Theorem 2.

Next, we consider the validity of the bootstrap tests for the case some components of $\boldsymbol{\phi}_0$ lie on the boundary of the parameter space; i.e., $\phi_{0i} = 0$ for some $i = 2,\ldots,p_1+p_2+1$. Since $\boldsymbol{\phi}_0$ is not an interior point, in this case, the limiting behaviour of $n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0)$ is not linear as in (A.15), and as in the proof of Theorem 2 of Francq and Zakoian (2007), we obtain

$$n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0) = \lambda_n^{\Lambda} + o_p(1), \quad \lambda_n^{\Lambda} := \arg\inf_{\lambda\in\Lambda} (Z_n-\lambda)'\Sigma_n(\boldsymbol{\phi}_0)(Z_n-\lambda). \tag{A.16}$$

The vector λ_n^{Λ} is the orthogonal projection of Z_n on the convex set Λ for the inner product $\langle x,y\rangle := x'\Sigma_n(\boldsymbol{\phi}_0)y$, and it is characterized by $\lambda_n^{\Lambda}\in\Lambda$, $\langle Z_n-\lambda_n^{\Lambda}, \lambda_n^{\Lambda}-\lambda\rangle \geq 0$, $\forall\lambda\in\Lambda$; see e.g. Lemma 1.1 in Zarantonello (1971). Thus, by arguing as in the proof of Theorem 2 in Francq and Zakoian (2007) we obtain that $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \xrightarrow{d} \lambda^{\Lambda} := \arg\inf_{\lambda \in \Lambda}(\lambda - Z)'\Sigma(\boldsymbol{\phi}_0)(\lambda - Z)$. Therefore, for a continuous G, the term $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'J(y, \boldsymbol{\phi}_0)$ in (A.14) is asymptotically $\mathcal{C}$-tight, as is $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ by Lemma 1 and Theorem 1.

Next, consider the bootstrap data generation as outlined in Section 4.3. Since $\hat{\boldsymbol{\phi}}^{\dagger}$ converges to $\boldsymbol{\phi}_0$, a.s., by a triangular array version of the proof of Lemma 1 replacing $\hat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}_0$ by $\hat{\boldsymbol{\phi}}^*$ and $\hat{\boldsymbol{\phi}}^{\dagger}$, respectively, we obtain that conditional on $\{Y_1, \ldots, Y_n\}$, uniformly in $y \in \mathbb{R}$,

$$\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^*) = \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^{\dagger}) - n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}^{\dagger})'J^*(y, \hat{\boldsymbol{\phi}}^{\dagger}) + o_p^*(1), \tag{A.17}$$

in probability, where

$$J^*(y, \boldsymbol{\phi}) = \mathrm{E}^*[\tau_1^*(\boldsymbol{\phi})\mathbb{I}(Y_0^* \leq y)], \quad \tau_i^*(\boldsymbol{\phi}) := \frac{(\partial/\partial\boldsymbol{\phi})h_i^*(\boldsymbol{\phi})}{h_i^*(\boldsymbol{\phi})}, \quad \boldsymbol{\phi} \in \Phi.$$

Since $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^{\dagger}) = n^{-1/2}\sum_{i=1}^n(\varepsilon_i^* - 1)^2\mathbb{I}(Y_{i-1}^* \leq y)$, by using Assumption (B3), for every $x, y \in \mathbb{R}$, with $x \wedge y := \min(x, y)$, we obtain that

$$\begin{aligned}
\mathrm{cov}^*\{\mathcal{U}_n^*(x, \hat{\boldsymbol{\phi}}^{\dagger}), \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^{\dagger})\} &= n^{-1}\sum_{i=1}^n \mathrm{E}^*(\varepsilon_i^* - 1)^2\mathbb{I}(Y_{i-1}^* \leq x \wedge y) \\
&= (\kappa_\varepsilon - 1)\mathrm{E}\mathbb{I}(Y_{i-1} \leq x \wedge y) + o_p(1) \\
&= \mathrm{cov}\{\mathcal{U}_n(x, \boldsymbol{\phi}_0), \mathcal{U}_n(y, \boldsymbol{\phi}_0)\} + o_p(1).
\end{aligned} \tag{A.18}$$

Further, it follows from Assumption (B3) that $J^*(y, \boldsymbol{\phi}) = \mathrm{E}[\tau_1(\boldsymbol{\phi})\mathbb{I}(Y_0 \leq y)] + o_p(1)$.

Therefore, in order to establish that the conditional weak limit of $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ is the same as that of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$, in probability, we need to first show that the conditional limiting distribution of $n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}^{\dagger})$ is the same as that of $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$, in probability. To this end, it suffices to show that, conditional on $\{Y_1, \ldots, Y_n\}$,

$$n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}^{\dagger}) = \lambda_n^{\Lambda} + o_p^*(1), \text{ in probability.} \tag{A.19}$$

In order to obtain (A.19) we consider a triangular array version of the proof of (A.16).

In the proof of (A.16) in Francq and Zakoian (2007), first λ_n^{Λ} is represented as the orthogonal projection of Z_n on the convex set Λ, for the inner product $\langle x, y\rangle := x'\Sigma_n(\boldsymbol{\phi}_0)y$, and then this projection is approximated by that of Z_n on the set $n^{1/2}(\Phi - \boldsymbol{\phi}_0)$. Since Φ contains a hypercube which includes $\boldsymbol{\phi}_0$, see (A.1), the set $n^{1/2}(\Phi - \boldsymbol{\phi}_0)$ increases to Λ as $n \to \infty$. This plays a key role in the proof of (A.16). Recall that,

$$\Lambda = \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_{p_1+p_2+1},$$

where $\Lambda_1 = \mathbb{R}$, and for each $i = 2, \ldots, p_1 + p_2 + 1$, denoting $\boldsymbol{\phi}_0 = (\boldsymbol{\phi}_{01}, \ldots, \boldsymbol{\phi}_{0(1+p_1+p_2)})'$, $\Lambda_i = \mathbb{R}$ if $\boldsymbol{\phi}_{0i} \neq 0$ and $\Lambda_i = [0, \infty)$ if $\boldsymbol{\phi}_{0i} = 0$. In order to extend the proof of (A.16) to the triangular array setup of the bootstrap data generation, we need to replace $\boldsymbol{\phi}_0$ by the bootstrap true parameter $\hat{\boldsymbol{\phi}}^\dagger$. However, to ensure that $n^{1/2}(\Phi - \hat{\boldsymbol{\phi}}^\dagger)$ increases to Λ, we need to show that $\hat{\boldsymbol{\phi}}^\dagger$ satisfies two important conditions. First, to allow the hypercube in (A.1) to contain $\hat{\boldsymbol{\phi}}^\dagger$ with probability converging to one, we need to have

$$\hat{\boldsymbol{\phi}}^\dagger \to \boldsymbol{\phi}_0 \text{ in probability as } n \to \infty. \tag{A.20}$$

Further, $\hat{\boldsymbol{\phi}}^\dagger$ should satisfy the following rate of consistency property.

$$n^{1/2}(\hat{\boldsymbol{\phi}}^\dagger_{ni} - \boldsymbol{\phi}_{0i}) = \begin{cases} o_p(1), & \text{if } \boldsymbol{\phi}_{0i} = 0 \\ O_p(1), & \text{if } \boldsymbol{\phi}_{0i} > 0 \end{cases}, \quad i = 1, 2, \ldots, 1 + p_1 + p_2. \tag{A.21}$$

Since at least one component of $\boldsymbol{\phi}_0$ is zero, the rate of consistency (A.21) ensures that $n^{1/2}(\Phi - \hat{\boldsymbol{\phi}}^\dagger) = n^{1/2}(\Phi - \boldsymbol{\phi}_0) - n^{1/2}(\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}_0)$ converges to Λ in probability.

The consistency of $\hat{\boldsymbol{\phi}}^\dagger$ follows from that of $\hat{\boldsymbol{\phi}}$, and hence (A.20) holds. Since c_n converges to 0 at a rate slower than $n^{-1/2}$, (A.21) follows by arguing as in the proof of Lemma 1 in Cavaliere *et al.* (2021). Hence, by a triangular array extension of the proof of Lemma 2, under Assumption (A6), we obtain that (A.19) holds, for example by arguing as in the proof of Proposition 3.2 in Hidalgo and Zaffaroni (2007); see also the discussion under Assumption E2 in Andrews (1997). Therefore, from (A.14)–(A.17), and the asymptotic tightness of $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^\dagger)$, we obtain that the conditional weak limit of $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}^*)$ is the same as that of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}})$, in probability. Hence, the continuous mapping theorem yields that the bootstrap test (21) based on T_j is asymptotically valid ($j = 1, 2$). □

Proof of Theorem 5. If $\boldsymbol{\phi}_0^*$ is an interior point, then the proof follows from Theorem 3. Therefore, we only consider the case some components of $\boldsymbol{\phi}_0^*$ lie on the boundary of the parameter space; i.e., $\phi_{0i}^* = 0$ for some $i = 2, \ldots, p_1 + p_2 + 1$. Since Assumptions (B1), (B2), (B3), (B4) and (A6) continue to hold, although $(\boldsymbol{\phi}_0^*, F_0^*)$ is the pseudo-true value under H_1, by arguing as in the proof of Theorem 4, for every $y \in \mathbb{R}$, we have that $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}^\dagger) = O_p^*(1)$, in probability. Therefore, it suffices to show that $n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})| = O_p(1)$, for some $y \in \mathbb{R}$ satisfying $\mathrm{E}[\{h_1/h_1(\boldsymbol{\phi}_0^*) - 1\}\mathbb{I}(Y_0 \leq y)] \neq 0$. Fix such a y. Under H_1, $\varepsilon_i = Y_i/\sqrt{h_i}$, where $h_i = \mathrm{E}[Y_i^2 \mid \mathcal{H}_{i-1}]$. Hence, by using Assumptions (B1), (B2), (B3), and (B4), and arguing as in the proof of Lemma A.4, on the set $n^{1/2}\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*\| \leq C$, where $0 < C < \infty$, we obtain that

$$\begin{aligned} &\Big|n^{-1}\sum_{i=1}^n Y_i^2\big[h_i^{-1}(\hat{\boldsymbol{\phi}}) - h_i^{-1}(\boldsymbol{\phi}_0^*)\big]\mathbb{I}(Y_{i-1} \leq y)\Big| \\ \leq\ & \omega_L^{-1} n^{-1}\sum_{i=1}^n Y_i^2\big|(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*)'\tau_i(\boldsymbol{\phi}_0^*)\big| + o_p(n^{-1/2}). \end{aligned}$$

By the Ergodic Theorem, and because $\hat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}_0^*$, the sum in the above upper bound is $o_p(1)$, and hence $n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}})|$ is bounded from the above by $|n^{-1}\sum_{i=1}^n \{\varepsilon_i^2(h_i/h_i(\boldsymbol{\phi}_0^*)) -$

$1\}\mathbb{I}(Y_{i-1} \leq y)|$ up to a term of order $o_p(1)$. Since $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0^*) = O_p(1)$, then the proof follows by an extended Glivenko-Cantelli type argument as in the proof of Theorem 3. $\square$

References

Agosto, A., Cavaliere, G., Kristensen, D., and Rahbek, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance*, **38**, 640–663.

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica*, **65**(5), 1097–1128.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, **69**(3), 683–734.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics*, **85**(3), 531–549.

Balakrishna, N., Koul, H. L., Sakhanenko, L., and Ossiander, M. (2019). Fitting a pth order parametric generalized linear autoregressive multiplicative error model. *Sankhya B*, **81**(1, suppl.), S103–S122.

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, **32**(2), 633–655.

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

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**(6), 1196–1217.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag, New York. Reprint of the 1993 original.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons Inc., New York.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3), 307–327.

Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, **52**(1-2), 115–127.

Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, **20**(4), 1714–1730.

Cavaliere, G., Nielsen, H. B., Pedersen, R. S., and Rahbek, A. (2021). Bootstrap Inference On The Boundary Of The Parameter Space With Application To Conditional Volatility Models. *Journal of Econometrics*, **in press**.

Chan, K.-S., Li, D., Ling, S., and Tong, H. (2014). On conditionally heteroscedastic AR models with thresholds. *Statistica Sinica*, **24**(2), 625–652.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, **106**(494), 608–625.

Chen, M. and An, H. Z. (1998). A note on the stationarity and the existence of moments of the GARCH model. *Statist. Sinica*, **8**(2), 505–510.

D'Agostino, R. B. and Stephens, M. A., editors (1986). *Goodness-of-fit techniques*, volume 68 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York.

Escanciano, J. C. (2008). Joint and marginal specification tests for conditional mean and variance models. *Journal of Econometrics*, **143**(1), 74–87.

Escanciano, J. C. (2010). Asymptotic distribution-free diagnostic tests for heteroskedastic time series models. *Econometric Theory*, **26**(3), 744–773.

Escanciano, J. C., Pardo-Fernández, J. C., and Keilegom, I. V. (2018). Asymptotic distribution-free tests for semiparametric regressions with dependent data. *The Annals of Statistics*, (in press).

Fernandes, M. and Grammig, J. (2005). Nonparametric specification tests for conditional duration models. *Journal of Econometrics*, **127**(1), 35–68.

Francq, C. and Zakoian, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*, **117**(9), 1265–1284.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH models: structure, statistical inference and financial applications*. Wiley. John Wiley & Sons Ltd.

Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, **29**(3), 567–589.

Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. Imperial College Press, London.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48**(5), 1779–1801.

Grenander, U. and Rosenblatt, M. (1957). *Statistical analysis of stationary time series*. John Wiley & Sons, New York; Almqvist & Wiksell, Stockholm.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.

Hidalgo, J. and Zaffaroni, P. (2007). A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics*, **141**(2), 835–875.

Jensen, S. r. T. and Rahbek, A. (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory*, **20**(6), 1203–1226.

Khmaladze, È. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Theory of Probability and Its Applications*, **26**(2), 240–257.

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, **4**, 83–91.

Koul, H. L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *The Annals of Statistics*, **34**(2), 994–1012.

Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *The Annals of Statistics*, **27**(1), 204–236.

Koul, H. L., Perera, I., and Silvapulle, M. J. (2012). Lack-of-fit testing of the conditional mean function in a class of Markov multiplicative error models. *Econometric Theory*, **28**(6), 1283–1312.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory*, **6**(3), 318–334.

Perera, I. and Koul, H. L. (2017). Fitting a two phase threshold multiplicative error model. *Journal of Econometrics*, **197**(2), 348–367.

Perera, I. and Silvapulle, M. J. (2017). Specification tests for multiplicative error models. *Econometric Theory*, **33**(2), 413–438.

Perera, I. and Silvapulle, M. J. (2020). Specification tests for dynamic conditional distribution models. *Discussion paper.*

Perera, I. and Silvapulle, M. J. (2021). Bootstrap based probability forecasting in multiplicative error models. *Journal of Econometrics*, **221**(1), 1–24.

Perera, I., Hidalgo, J., and Silvapulle, M. J. (2016). A goodness-of-fit test for a class of autoregressive conditional duration models. *Econometric Reviews*, **35**(6), 1111–1141.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, **34**(5), 2449–2495.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, **25**(2), 613–641.

Tsay, R. and Chen, R. (2018). *Nonlinear Time Series Analysis.* Wiley Series in Probability and Statistics. Wiley.

Tsay, R. S. (2013). *An introduction to analysis of financial data with R.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ.

von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, **12**, 367–395.

Zarantonello, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory. I. Projections on convex sets. In *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, pages 237–341.