Article:

Perera, I. and Silvapulle, M. (2023) Bootstrap specification tests for dynamic conditional distribution models. Journal of Econometrics, 235 (2). pp. 949-971. ISSN 0304-4076

https://doi.org/10.1016/j.jeconom.2022.08.006

Bootstrap specification tests for dynamic conditional distribution models

Indeewara Perera

Department of Economics, The University of Sheffield, Sheffield S1 4DT, UK
(Email: i.perera@sheffield.ac.uk).

Mervyn J. Silvapulle

Department of Econometrics and Business Statistics
Monash University, Australia VIC 3145
(Email: Mervyn.Silvapulle@monash.edu).

September 17, 2022

Corresponding author: Indeewara Perera / Mervyn J. Silvapulle.

Abstract

This paper proposes bootstrap based tests for the specification of a given parametric conditional distribution in autoregressive time series with GARCH-type disturbances. The tests are based on an estimated residual empirical process and are implemented by parametric bootstrap. We show that the proposed tests are asymptotically valid, consistent, and have nontrivial asymptotic power against a large proportion of local alternatives. Our approach relies on non-primitive regularity conditions and certain properties of exponential almost sure convergence. The regularity conditions are shown to be satisfied by GARCH(p,q); this technique of verification is applicable to other models as well. In our Monte Carlo study, the proposed tests performed well and better than several competing tests, including the information matrix test. A real data example illustrates the testing procedure.

Keywords: GARCH; goodness-of-fit; residual empirical process; Kolmogorov-Smirnov test; lack-of-fit test; stochastic recurrence equations.

JEL Classifications: C12, C52.

1 Introduction

Time series models with conditionally heteroscedastic disturbance terms, in particular those in the family of Generalized AutoRegressive Conditional Heteroscedastic [GARCH] models, are widely used in empirical studies; examples include option pricing and currency exchange rate. In view of the practical importance of this class of models, there is a need for a range of specification tests in the econometricians' tool box. To this end, this paper develops methodology for testing the specification of a fully parametric GARCH-type model that includes a parametric error distribution. This is an area in which there is a scarcity of methodological developments although there is an extensive literature on GARCH-type models. To implement the proposed test, an easy to use bootstrap method is proposed and is shown to be asymptotically valid and consistent. The proposed methods are applicable to a wide range of models used in empirical studies, including linear GARCH(p,q), AR(k)-GARCH(p,q), and AR(k)-(Asymmetric)AGARCH(p,q).

To illustrate the nature of the testing problem, let us consider an example. Let X_i denote the financial return from an investment at time i. Consider the AR(1)-GARCH(1,1) model:

$$X_i = \mu_i(\boldsymbol{\phi}) + \Psi_i^{1/2}(\boldsymbol{\phi})\varepsilon_i, \quad \mu_i(\boldsymbol{\phi}) = \phi_1 + \phi_2 X_{i-1}, \quad \Psi_i(\boldsymbol{\phi}) = \phi_3 + \phi_4\{X_{i-1} - \mu_{i-1}(\boldsymbol{\phi})\}^2 + \phi_5 \Psi_{i-1}(\boldsymbol{\phi}),$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_5)^\top$ is an unknown parameter ($^\top$ denotes transpose), $\{\varepsilon_i\}$ are independent and identically distributed [iid] with zero mean and unit variance, and the common probability density function [pdf] $f_{\boldsymbol{\theta}}$ of $\{\varepsilon_i\}$ is skew-$t(\boldsymbol{\theta})$ (Hansen 1994). The methodology developed in this paper can be used for testing the null hypothesis that the parametric forms $\{\mu_i(\boldsymbol{\phi}), \Psi(\boldsymbol{\phi}), f_{\boldsymbol{\theta}}\}$ are correct.

Formulation of the problem: Let $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ denote the sets of non-negative integers, integers, and real numbers, respectively. For a given sequence $\{\upsilon_i\}_{i\in\mathbb{Z}}$, let $\boldsymbol{\upsilon}_{i-1}$ (in bold font) denote the vector of lagged values $(\upsilon_{i-1}, \ldots, \upsilon_{i-k})^\top$, for some known positive integer $k < \infty$; although the value of the lag-length k for different vectors could be different, we suppress it in our notation as it does not affect the derivations. Thus, in what follows, the bold terms $\boldsymbol{X}_{i-1}$ and $\boldsymbol{\mu}_{i-1}$ are vectors of known lengths consisting of lagged values of X_i and μ_i respectively; further, they may have different lengths. Let $\mathcal{H}_i$ denote the information set containing $\{\ldots, X_{-1}, \ldots, X_i\}$ at time i ($i \in \mathbb{Z}$).

Assume that the process $\{X_i\}_{i\in\mathbb{Z}}$ is strictly stationary and ergodic with finite second moment. Let the model $\mathcal{M}$ be defined as follows:

$$\mathcal{M}: \quad \begin{cases} X_i &= \mu_i + \{\Psi_i\}^{1/2}\varepsilon_i, \qquad \{\varepsilon_i : i \in \mathbb{Z}\} \text{ are iid}, \qquad \varepsilon_i \sim F^0 \\ \mu_i &= h(\boldsymbol{X}_{i-1}, \boldsymbol{\mu}_{i-1}; \boldsymbol{\phi}), \quad \Psi_i = g\{\boldsymbol{\Psi}_{i-1}, \boldsymbol{\eta}_{i-1}; \boldsymbol{\phi}\}, \quad F^0 = F_{\boldsymbol{\theta}}, \end{cases} \tag{1}$$

for some $\boldsymbol{\phi} \in \Phi \subset \mathbb{R}^r$ and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^s$, where F^0 is the common cumulative distribution function [cdf] of the error terms $\{\varepsilon_i\}_{i\in\mathbb{Z}}$, $E(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = 1$, $\eta_i = X_i - \mu_i$ ($i \in \mathbb{Z}$), $F_{\boldsymbol{\theta}}$ is a cdf ($\boldsymbol{\theta} \in \Theta$), the functional forms of $\{g, h, F_{\boldsymbol{\theta}}\}$ are known, and the finite dimensions r and s are known; thus $F^0 = F_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}_0$ denoting the true value of $\boldsymbol{\theta}$, $\mu_i = E(X_i \mid \mathcal{H}_{i-1})$, and $\Psi_i = \mathrm{Var}(X_i \mid \mathcal{H}_{i-1})$ $(i \in \mathbb{Z})$. The distinction between F^0 and $F_{\boldsymbol{\theta}}$ in (1) is important because we will also need to consider the model defined by (1) without the restriction $F^0 = F_{\boldsymbol{\theta}}$ for some $\boldsymbol{\theta} \in \Theta$. Let f^0 and $f_{\boldsymbol{\theta}}$ denote the pdf's corresponding to F^0 and $F_{\boldsymbol{\theta}}$ respectively.

This paper develops methods for testing

$$H_0 : \text{Model } \mathcal{M} \text{ is correct} \qquad \text{vs} \qquad H_1 : \text{Model } \mathcal{M} \text{ is not correct}, \tag{2}$$

where $\mathcal{M}$ is as in (1). The $\{AR(1)\text{-GARCH(1,1), skew-}t(\boldsymbol{\theta})\}$ model discussed earlier, is a special case of $\mathcal{M}$. Let us note that a test of the H_0 in (2) is also a test of

$$H_0^* : \; pr[X_{i+1} \le x \mid \mathcal{H}_i] = F_{\boldsymbol{\theta}}[\{x - \mu_i(\boldsymbol{\phi})\}/\Psi_i^{1/2}(\boldsymbol{\phi})], \; \forall x \in \mathbb{R}, \;\; \text{for some } (\boldsymbol{\phi}, \boldsymbol{\theta}) \in \Phi \times \Theta \; (i \in \mathbb{N}). \tag{3}$$

Related literature: Almost all of the papers in this literature fall into two categories, except the information matrix test of White (1982). The studies in the first group have sound theory but are applicable only to very special cases of $\mathcal{M}$, but not to the general testing problem studied in this paper. The tests in the second group are based on computer simulations for specific cases, but they lack rigorous methodology. Thus, there is a need for further methodological developments on the specific topic studied in this paper, namely test of H_0 against H_1 in (2).

In model $\mathcal{M}$, the distribution of X_i, conditional on the past $\{X_j, j < i\}$, depends on the *unobservable* set $\{\ldots, X_{-1}, X_0\}$ that extends to the infinite past. Estimation in such models is well developed; for example, see Lee and Hansen (1994), Hall and Yao (2003), Francq and Zakoïan (2004), Straumann and Mikosch (2006), Francq *et al.* (2011), and Conrad and Mammen (2016). Nevertheless, the theory for testing the specification of the corresponding conditional distribution model remains to be developed; in this context, most papers on testing focus only on the case the conditioning variables in the conditional distribution of X_i is fully observable, for example as in ARCH(p) models (Inglot and Stawiarski 2005; Koul and Ling 2006; Stawiarski 2009; Chen 2012; Koul and Mimoto 2012; Chen and Hong 2014). The features of $\mathcal{M}$ that cause the technical challenges in developing a bootstrap specification test have been recognized, but the problem remains unsolved (cf. Horváth *et al.*, 2004).

The methodology for specification tests are available for some special cases of model $\mathcal{M}$. For example, several tests are available when the conditioning variables, $\{\mu_j(\boldsymbol{\phi}), \Psi_j(\boldsymbol{\phi}), j = 1, \ldots, i\}$, are observable (cf. Corradi and Swanson 2006, Bierens and Wang 2017, Chen and Hong 2014, Troster and Wied 2021, Neumann and Paparoditis 2008). However, this is not the case for $\mathcal{M}$ because the conditional variance, and hence the distribution of X_i, conditional on $\mathcal{H}_{i-1}$, depends on the unobservable $\{\ldots, X_{-2}, X_{-1}, X_0\}$ extending back to infinite past (cf. Berkes *et al.* 2003, and Chap. 7 in Francq and Zakoïan 2010). Among the specification tests for $\mathcal{M}$ that allow for such unobservable conditioning variables, there are a few for testing the specification of only some components of the parametric specification (1). Examples include testing (a) the specification of $\Psi_i(\boldsymbol{\phi})$ in GARCH(1,1) (Leucht *et al.* 2015), (b) the specification of $\{\mu_i(\boldsymbol{\phi}), \Psi_i(\boldsymbol{\phi})\}$ (Escanciano 2010; Chen and Gao 2011; Chen *et al.* 2003), and (c) against no serial residual correlation (Andreou and Werker 2015). Tests for models with nonnegative error, for example those in Perera and Silvapulle (2017, 2021) and Perera and Koul (2017), are not applicable to the setting of this paper (see also Dette *et al.* 2009).

Several tests are available for testing the specification of the error distribution in some special cases; for example, Koul and Mimoto (2012), Chen (2012), Inglot and Stawiarski (2005); Stawiarski (2009), Bai (2003), Koul and Ling (2006). None of them is applicable to the general testing problem (2) for the model $\mathcal{M}$ in (1). The information matrix test of White (1982) is a general purpose test that can be used for testing (2) (Huo and Cho 2021). Its performance in the general context of (2) does not appear to have been evaluated.

Contribution of this paper: The main contributions of the paper can be summarised as follows:

(a) This paper develops methodology for testing the parametric specification of a large class of GARCH type models of the form $\mathcal{M}$ in (1). A simple and easy to use parametric bootstrap method for estimating the p-values of the tests is proposed and shown to be asymptotically valid.

(b) The main theorems show that the proposed tests are asymptotically valid and unbiased. Further, for certain sequences of local alternatives, we show that the asymptotic local power is a *strictly* increasing function of the distance between the null and the sequence of alternatives. Therefore, the asymptotic local power is larger than the asymptotic size of the test, which we term *non-trivial local power*. We show that the tests are consistent against a large proportion of fixed alternatives. We recognize that there are local alternatives and fixed alternatives against which the proposed tests do not have nontrivial power.

(c) A method for verifying the regularity conditions is illustrated for GARCH(p,q). This method of verification has a general structure and is also applicable to a range of models, including AR(k)-GARCH(p,q) and AR(k)-AGARCH(p,q). These verifications may turn out to be long for complex models, but they follow a well-defined set of mathematical steps that can be performed independent of the methodological developments in this paper. Thus, the main paper and the verifications for GARCH(p,q) in the supplement are well separated and they may be read almost independently.

General comments: Specification tests typically have the assumed model in the null hypothesis, as in (2). By contrast, in *superiority tests* the desired model is in the alternative hypothesis and everything else in the null model (Silvapulle and Sen 2005). An anonymous reviewer indicated that such a superiority test with $\mathcal{M}$ in the alternative hypothesis would be useful and suggested a possible formulation. Such a test would certainly be useful, but it would require significant new methodology.

Since the statistical inference problem (2) is formulated as a test of hypothesis, it cannot be used for concluding that the model $\mathcal{M}$ is correct; all that it can do is to show whether or not there is sufficient statistical evidence to conclude that the model $\mathcal{M}$ does not fit the data — this is well known. Further, it cannot detect the source of model violation when the model $\mathcal{M}$ in H_0 is rejected. Nevertheless, the following strategy is useful. The main scarcity in the specification testing tool box for models of the form $\mathcal{M}$ is for testing the specification of the assumed parametric form $F_{\boldsymbol{\theta}}$ for the error distribution. Specification tests for $\{\mu_i(\cdot), \mathit{\Psi}_i(\cdot)\}$, without involving $F_{\boldsymbol{\theta}}$, have been developed (Escanciano *et al.* 2018). Therefore, we recommend that such tests and diagnostics be applied first to choose suitable forms for $\{\mu_i(\cdot)$ and $\mathit{\Psi}_i(\cdot)\}$. Then apply the method in this paper that is suitable for testing the entire specification $\mathcal{M}$, including the error distribution and the iid assumption of the error terms.

Outline of the paper: Section 2 introduces the test statistics and states the bootstrap implementation. Section 3 establishes the weak convergence of the test statistics. Section 4 provides the general results on the validity of the bootstrap tests. Asymptotic power of the tests against fixed and local alternatives is discussed in Section 5. Section 6 provides a brief discussion on the applicability of the method for testing the specification of GARCH(p,q); detailed verification of the regularity conditions to ensure that our tests are applicable to GARCH(p,q), are provided in the online supplement to this paper. The results of a simulation study are presented in Section 7, and Section 8 contains a brief illustrative data example. The Appendix at the end of this paper provides the main steps for the proofs of the theorems.

2 The test and its bootstrap implementation

Let the observable approximations $\{\widetilde{\mathit{\Psi}}_i, \widetilde{\mu}_i, \widetilde{\eta}_i\}$ of $\{\mathit{\Psi}_i, \mu_i, \eta_i\}$ be defined recursively by

$$\widetilde{\mu}_i(\boldsymbol{\phi}) = h\big(\boldsymbol{X}_{i-1}, \widetilde{\boldsymbol{\mu}}_{i-1};\ \boldsymbol{\phi}\big), \quad \widetilde{\eta}_i(\boldsymbol{\phi}) = X_i - \widetilde{\mu}_i(\boldsymbol{\phi}), \quad \widetilde{\mathit{\Psi}}_i(\boldsymbol{\phi}) = g\big(\widetilde{\boldsymbol{\Psi}}_{i-1}, \widetilde{\boldsymbol{\eta}}_{i-1};\ \boldsymbol{\phi}\big) \quad (i \geq 1; \boldsymbol{\phi} \in \Phi), \quad (4)$$

conditional on the initial values for $\{(X_i, \widetilde{\mathit{\Psi}}_i, \widetilde{\mu}_i) : i \leq 0\}$ chosen by the user; in this paper, we use $(X_i, \widetilde{\mathit{\Psi}}_i, \widetilde{\mu}_i) = (\bar{X}, 1, \bar{X})$ for $i \leq 0$, where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. In a theorem established later in this paper, we show that the asymptotic properties of our tests do not depend on the initial values. Let

$$\ell_i(\boldsymbol{\phi}) = \log \Psi_i(\boldsymbol{\phi}) + [\eta_i^2(\boldsymbol{\phi})/\Psi_i(\boldsymbol{\phi})], \quad \widetilde{\ell}_i(\boldsymbol{\phi}) = \log \widetilde{\Psi}_i(\boldsymbol{\phi}) + [\widetilde{\eta}_i^{\,2}(\boldsymbol{\phi})/\widetilde{\Psi}_i(\boldsymbol{\phi})] \quad (\boldsymbol{\phi} \in \Phi). \tag{5}$$

The unobservable Gaussian quasi-loglikelihood and its observable approximations are $\{-\sum_i \ell_i(\boldsymbol{\phi})\}$ and $\{-\sum_i \widetilde{\ell}_i(\boldsymbol{\phi})\}$, respectively.

To estimate the model under H_0, we use the observable *Gaussian quasi maximum likelihood estimator* [QMLE] $\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}\in\Phi} \sum_{i=1}^n \widetilde{\ell}_i(\boldsymbol{\phi})$. Let $\widetilde{\varepsilon}_i = \widetilde{\eta}_i(\hat{\boldsymbol{\phi}})/\{\widetilde{\Psi}_i(\hat{\boldsymbol{\phi}})\}^{1/2}$ $(i = 1, \ldots, n)$ denote the observable residuals. Under the null hypothesis H_0, $\{\widetilde{\varepsilon}_i, i = 1, \ldots, n\}$ are expected to behave almost as a simple random sample from $F_{\boldsymbol{\theta}}$. Therefore, we propose to estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}} := \arg\max_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta})$, where $Q_n(\boldsymbol{\theta}) = \sum_{i=1}^n \hbar(\boldsymbol{\theta}; \widetilde{\varepsilon}_i)$ and $\hbar$ is a suitably chosen function. For example, $\hbar(\boldsymbol{\theta}; \varepsilon)$ could be $\log f_{\boldsymbol{\theta}}(\varepsilon)$, in which case $\hat{\boldsymbol{\theta}}$ is expected to be close to the maximum likelihood estimator of $\boldsymbol{\theta}$ based on the unobservable sample $\{\varepsilon_i : i = 1, \ldots, n\}$.

To simplify the statements of the main results, we assume that $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ converges in probability, and define $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0) = \text{plim}(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ where 'plim' denotes the probability limit. If H_0 is true, then the regularity conditions imposed later entail that $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$ is the true parameter value of the null model $\mathcal{M}$ in (1); if H_0 is not true then $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$ is treated as a *pseudo-true value.* The bootstrap test is valid even if we do not make the foregoing assumption that $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ converges in probability under H_1, but the results would be unnecessarily complicated. Next, with I denoting the indicator function, we define the following two empirical distributions and their corresponding empirical processes:

$$F_n(x) = n^{-1} \textstyle\sum_{i=1}^n I(\varepsilon_i \le x), \qquad W_n(x) = \sqrt{n}\big\{F_n(x) - F_{\boldsymbol{\theta}_0}(x)\big\}, \tag{6}$$

$$\widetilde{F}_n(x) = n^{-1} \textstyle\sum_{i=1}^n I(\widetilde{\varepsilon}_i \le x), \qquad \widetilde{W}_n(x) = \sqrt{n}\big\{\widetilde{F}_n(x) - F_{\hat{\boldsymbol{\theta}}}(x)\big\}. \tag{7}$$

If H_0 is true then we expect $n^{-1/2}W_n(\cdot)$ and $n^{-1/2}\widetilde{W}_n(\cdot)$ to be uniformly close to zero. The latter is observable, but not the former. Therefore, a test statistic can be based on the latter. If H_0 is not true, then we expect any mis-specification to result in $\widetilde{F}_n$ being not uniformly close to $F_{\hat{\boldsymbol{\theta}}}$. For example, suppose that the true error distribution F^0 is not of the form $F_{\boldsymbol{\theta}}$ and that the rest of the specification in $\mathcal{M}$ is correct. Then, $\widetilde{F}_n$ is consistent for F^0, but $F_{\hat{\boldsymbol{\theta}}}$ is not consistent for F^0. Therefore, $\sup_x |\widetilde{W}_n(x)|$ is expected to be large. In fact, as shown in Section 5, if any one of the three parametric forms $\{F_{\boldsymbol{\theta}}, \mu_i(\boldsymbol{\phi}), \Psi_i(\boldsymbol{\phi})\}$ for the specification of model $\mathcal{M}$ is incorrect, then $\sup_x |\widetilde{W}_n(x)|$ is expected to be large for almost any type of misspecification.

These heuristics suggest that we may use a suitably chosen functional of $\widetilde{W}_n$ and $\hat{\boldsymbol{\theta}}$, denoted $T := \mathcal{T}(\widetilde{W}_n; \hat{\boldsymbol{\theta}})$, as a statistic for testing (2). In this paper, we consider the following: *Kolmogorov–Smirnov* [*KS*], $T_1 = \sup_x \big|\widetilde{W}_n(x)\big|$; *Cramér–von Mises* [*CvM*], $T_2 = \int \widetilde{W}_n^2(x) dF_{\hat{\boldsymbol{\theta}}}(x)$;

Kuiper $[Ku]$, $T_3 = \sup_x \widetilde{W}_n(x) + \sup_x\{-\widetilde{W}_n(x)\}$; *Anderson–Darling* $[A^2]$, $T_4 = \int \widetilde{W}_n^2(x)[F_{\hat{\boldsymbol{\theta}}}(x)\{1-F_{\hat{\boldsymbol{\theta}}}(x)\}]^{-1}dF_{\hat{\boldsymbol{\theta}}}(x)$; *Watson* $[U^2]$, $T_5 = \int \left\{\widetilde{W}_n(x) - \int[\widetilde{W}_n(x)]dF_{\hat{\boldsymbol{\theta}}}(x)\right\}^2 dF_{\hat{\boldsymbol{\theta}}}(x)$.

Simpler formulas are available for computing the statistics $T_1, \dots T_5$, as indicated below. Since $\widetilde{F}_n(x)$ is a nondecreasing step function and $F_{\hat{\boldsymbol{\theta}}}(x)$ is non-decreasing, it follows that the maximum and the minimum of $\widetilde{W}_n(x)$ may occur only at the n jump points, $\{\widetilde{\varepsilon}_1, \dots, \widetilde{\varepsilon}_n\}$. Therefore, Kolmogorov–Smirnov and Kuiper statistics can be computed easily. For the integral statistics, simpler formulae are derived by changing the variable of integration to $y = F_{\hat{\boldsymbol{\theta}}}(x)$. Let $\widetilde{\varepsilon}_{(1)} \le \cdots \le \widetilde{\varepsilon}_{(n)}$ denote the ordered values of $\{\widetilde{\varepsilon}_1, \dots, \widetilde{\varepsilon}_n\}$, $D^+ = \sqrt{n}\max_i \left\{i/n - F_{\hat{\boldsymbol{\theta}}}(\widetilde{\varepsilon}_{(i)})\right\}$, and $D^- = \sqrt{n}\max_i \left\{F_{\hat{\boldsymbol{\theta}}}(\widetilde{\varepsilon}_{(i)}) - [(i-1)/n]\right\}$. Then, we obtain the following simple computational formulas:

$KS = \max\{D^+, D^-\}$, $\qquad CvM = \sum_{i=1}^n \left\{F_{\hat{\theta}}(\widetilde{\varepsilon}_{(i)}) - (2n)^{-1}(2i-1)\right\}^2 + (12n)^{-1}$,

$Ku = D^+ + D^-$, $\qquad A^2 = -n - n^{-1}\sum_{i=1}^n (2i-1)[\log\{F_{\hat{\theta}}(\widetilde{\varepsilon}_{(i)})\} + \log\{1 - F_{\hat{\theta}}(\widetilde{\varepsilon}_{(n+1-i)})\}]$,

$U^2 = CvM - n\{n^{-1}\sum_{i=1}^n F_{\hat{\theta}}(\widetilde{\varepsilon}_{(i)}) - 0.5\}^2$.

Similar tests have been studied in the literature for models that are either simpler or different (cf. Andrews 1997; Horváth *et al.* 2004). Our tests complement the literature on empirical processes based tests such as those in Andrews (1997), Khmaladze and Koul (2004), Koul and Ling (2006), Escanciano *et al.* (2018), and Delgado and Stute (2008).

The main results in Section 3 show that the limiting null distributions of test statistics such as $T_1, \dots, T_5$ depend on the unknown parameter $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$. The covariance kernel of the limiting process of $\widetilde{W}_n$ turns out to be complicated; 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), and Escanciano *et al.* (2018). Therefore, in this paper, we propose a bootstrap method to estimate the p-values of the test statistics and implement the tests; for a detailed account of bootstrapping dependent processes, see Lahiri (2003).

2.1 The bootstrap algorithm for implementing the test

Since the population model is fully parametric under H_0, we propose a parametric bootstrap method based on $T := \mathcal{T}(\widetilde{W}_n; \hat{\boldsymbol{\theta}})$. The following algorithm uses the first m periods as burn-in.

Step 1: Compute $\{\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, T\}$ for the realized sample $\{X_1, \dots, X_n\}$.

Step 2: Generate $n + m + 1$ independent observations $\{\varepsilon^*_{-m}, \dots, \varepsilon^*_n\}$ from $F_{\hat{\boldsymbol{\theta}}}$.

Step 3: Generate $\{X^*_{-m}, \dots, X^*_n\}$, under H_0, with $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ as the true population value.

[*Remark:* In this step, first choose the initial values $(X^*_i, \Psi^*_i, \mu^*_i) = (\bar{X}, 1, \bar{X})$, for $i < -m$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. Then generate X^*_i recursively using the model equation (1), for $i = -m, \dots, n$.]

Step 4: Based on the bootstrap sample $\{X^*_1, \dots, X^*_n\}$, compute the analogs $\{\hat{\boldsymbol{\phi}}^*, \hat{\boldsymbol{\theta}}^*, T^*\}$ of $\{\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, T\}$.

Step 5: Repeat steps 2–4 a large number of times and estimate the p-value by $\hat{p}$, the proportion of times T^* exceeds the sample value T. Finally, the proposed bootstrap test is

$$\text{Reject } H_0 \text{ at level } \alpha \text{ if } \hat{p} \leq \alpha. \tag{8}$$

Under the assumptions of Theorem 2 in Section 4.2, it can be shown that the effect of the initialization in Step 3 is negligible in large samples. Therefore, to simplify the exposition, we assume that the above bootstrap method has no initialization effect.

3 Weak convergence of the test statistics under the null

The main result of this section shows that, under H_0, the test statistic T_j converges in distribution to a non-degenerate continuous random variable ($j = 1, \ldots, 5$). This theorem imposes several high-level regularity conditions that are stated as Assumptions 1-7. Since almost all of them have appeared in the literature, we indicate where they have appeared and make only brief comments.

First, let us introduce some notation. For a function of a parameter and real variable, let 'dot' and 'prime' denote differentiations with respect to the parameter and the real variable, respectively. For example, $\dot{\hbar}(\boldsymbol{\theta}, x) = (\partial/\partial\boldsymbol{\theta})\hbar(\boldsymbol{\theta}, x)$, $\hbar'(\boldsymbol{\theta}, x) = (\partial/\partial x)\hbar(\boldsymbol{\theta}, x)$, $\dot{\Psi}_i(\boldsymbol{\phi}) = (\partial/\partial\boldsymbol{\phi})\Psi_i(\boldsymbol{\phi})$, $\ddot{\Psi}_i(\boldsymbol{\phi}) = (\partial^2/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top)\Psi_i(\boldsymbol{\phi})$, $\dot{F}_{\boldsymbol{\theta}}(x) = (\partial/\partial\boldsymbol{\theta})F_{\boldsymbol{\theta}}(x)$, and $F'_{\boldsymbol{\theta}}(x) = (\partial/\partial x)F_{\boldsymbol{\theta}}(x)$. Let

$$\tau_i(\boldsymbol{\phi}) = \dot{\Psi}_i(\boldsymbol{\phi})/\Psi_i(\boldsymbol{\phi}), \quad \varphi_i(\boldsymbol{\phi}) = \dot{\mu}_i(\boldsymbol{\phi})/\sqrt{\Psi_i(\boldsymbol{\phi})}, \quad (\boldsymbol{\phi} \in \Phi). \tag{9}$$

We say that a sequence of random variables $\{Y_i\}_{i\in\mathbb{N}}$ converges to zero *exponentially almost surely*, $Y_i \overset{e.a.s.}{\to} 0$, if there exists $\gamma > 1$ such that $\gamma^i Y_i \overset{a.s.}{\to} 0$ as $i \to \infty$. Let $\overset{d}{\to}$ and $\overset{p}{\to}$ denote convergence in distribution and convergence in probability, respectively. Let '$\overset{w}{\Rightarrow}$' denote weak convergence.

3.1 Assumptions

In the following assumptions, B denotes a generic compact neighborhood of either $\boldsymbol{\phi}_0$ or $\boldsymbol{\theta}_0$.

Assumption 1. *(1.1). The parameter space Φ is compact. The model (1) admits a unique stationary and ergodic solution $\{\Psi_i(\boldsymbol{\phi}_0), \mu_i(\boldsymbol{\phi}_0)\}_{i\in\mathbb{Z}}$ with $\boldsymbol{\phi}_0$ as an interior point of Φ.*
(1.2). $\mu_0(\boldsymbol{\phi}) \overset{a.s.}{=} \mu_0(\boldsymbol{\phi}_0)$ and $\Psi_0(\boldsymbol{\phi}) \overset{a.s.}{=} \Psi_0(\boldsymbol{\phi}_0)$ imply that $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.
(1.3). There exists an $\alpha_L > 0$ such that $g(\cdot) > \alpha_L$ [g is defined in (1)].
(1.4). $\mu_i, \Psi_i, \widetilde{\mu}_i$, and $\widetilde{\Psi}_i$ are twice continuously differentiable ($i \in \mathbb{Z}$), and $E[|X_0|^{4+d}]$, $E[|\mu_0(\boldsymbol{\phi}_0)|^{4+d}]$, $E[\|\varphi_0(\boldsymbol{\phi}_0)\|^{4+d}]$, $E[|\Psi_0(\boldsymbol{\phi}_0)|^{2+d}]$, $E[\|\tau_0(\boldsymbol{\phi}_0)\|^{2+d}]$ are finite for some $d > 0$.

Assumption 2. *(2.1). $\sup_{\boldsymbol{\phi}\in\Phi} |\widetilde{\Psi}_i(\boldsymbol{\phi}) - \Psi_i(\boldsymbol{\phi})|$, $\sup_{\boldsymbol{\phi}\in\Phi} |\widetilde{\mu}_i(\boldsymbol{\phi}) - \mu_i(\boldsymbol{\phi})| \overset{e.a.s.}{\to} 0$ as $i \to \infty$.*
(2.2). $\sup_{\boldsymbol{\phi}\in B} \|\dot{\widetilde{\Psi}}_i(\boldsymbol{\phi}) - \dot{\Psi}_i(\boldsymbol{\phi})\|$, $\sup_{\boldsymbol{\phi}\in B} \|\dot{\widetilde{\mu}}_i(\boldsymbol{\phi}) - \dot{\mu}_i(\boldsymbol{\phi})\| \overset{e.a.s.}{\to} 0$ as $i \to \infty$.

Assumption 3. *(3.1). The objective function* $M(\boldsymbol{\phi}) := E[-\ell_0(\boldsymbol{\phi})]$, *where* $\ell_i(\boldsymbol{\phi}) := \log \Psi_i(\boldsymbol{\phi}) + \{X_i - \mu_i(\boldsymbol{\phi})\}^2/\Psi_i(\boldsymbol{\phi})$, *is uniquely maximized at* $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.
(3.2). $E[\dot{\ell}_0(\boldsymbol{\phi}_0)\dot{\ell}_0(\boldsymbol{\phi}_0)^\top]$ *and* $E[\ddot{\ell}_0(\boldsymbol{\phi}_0)]$ *are invertible,* $\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\| = o_p(1)$, *and* $\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) - n^{-1/2}A(\boldsymbol{\phi}_0)\sum_{i=1}^n \dot{\ell}_i(\boldsymbol{\phi}_0) = o_p(1)$, *where* $A(\boldsymbol{\phi}) = -[E\ddot{\ell}_0(\boldsymbol{\phi})]^{-1}$.

Assumption 4. *The following results hold for every constant* $M > 0$:
(4.1). $\sup \sqrt{n} \mid \mu_i(\boldsymbol{t}) - \mu_i(\boldsymbol{s}) - (\boldsymbol{t} - \boldsymbol{s})^\top \dot{\mu}_i(\boldsymbol{s}) \mid \{\Psi_i(\boldsymbol{\phi}_0)\}^{-1/2} = o_p(1)$,
(4.2). $\sup \sqrt{n} \mid \{\Psi_i(\boldsymbol{t})\}^{1/2} - \{\Psi_i(\boldsymbol{s})\}^{1/2} - 2^{-1}(\boldsymbol{t} - \boldsymbol{s})^\top \dot{\Psi}_i(\boldsymbol{s})\{\Psi_i(\boldsymbol{s})\}^{-1/2} \mid \{\Psi_i(\boldsymbol{\phi}_0)\}^{-1/2} = o_p(1)$,
where the supremum is taken over $1 \le i \le n$ *and* $\{(\boldsymbol{t}, \boldsymbol{s}) : \boldsymbol{t}, \boldsymbol{s} \in B, \sqrt{n}\|\boldsymbol{t} - \boldsymbol{s}\| \le M\}$.

Assumption 5. $\max_{1\le i\le n} \|\tau_i(\boldsymbol{\phi}_0)\|$ *and* $\max_{1\le i\le n} \|\varphi_i(\boldsymbol{\phi}_0)\|$ *are* $o_p(n^{1/2})$.

Assumption 6. *(6.1). The parameter space* Θ *is compact with* $\boldsymbol{\theta}_0$ *being an interior point of* Θ, $Q(\boldsymbol{\theta}) := E[\hbar(\boldsymbol{\theta}; \varepsilon_0)]$ *is uniquely maximized at* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = o_p(1)$ *and* $E[\ddot{\hbar}(\boldsymbol{\theta}_0; \varepsilon_0)]$ *is invertible.*
(6.2). Let $h_{\boldsymbol{\theta}}(\varepsilon) = \{h_{\boldsymbol{\theta}1}(\varepsilon), \ldots, h_{\boldsymbol{\theta}q}(\varepsilon)\}^\top := -E[\ddot{\hbar}(\boldsymbol{\theta}; \varepsilon_0)]^{-1}\dot{\hbar}(\boldsymbol{\theta}; \varepsilon)$. *Then,* $E[h_{\boldsymbol{\theta}_0}(\varepsilon_0)] = 0$ *under* H_0. *There exists a* $d > 0$ *such that* $\sup_{\boldsymbol{\theta}\in B} \int \|h_{\boldsymbol{\theta}}(t)\|^{2+d} f_{\boldsymbol{\theta}}(t)\, dt < \infty$.

Assumption 7. *(7.1). For all* $\boldsymbol{\theta} \in \Theta$, $F_{\boldsymbol{\theta}}$ *has an a.e. positive density* $f_{\boldsymbol{\theta}}$.
(7.2). $f_{\boldsymbol{\theta}}(y)$ *and* $F_{\boldsymbol{\theta}}(y)$ *are twice continuously differentiable in* $(\boldsymbol{\theta}, y)$.
(7.3). There exist real numbers $a > 0$ *and* $d > 0$ *such that* $\int |y|^{4+d} f_{\boldsymbol{\theta}}(y)dy < \infty$ *and* $\sup_{y\in\mathbb{R}, |u|<a}(1 + y^2)|f'_{\boldsymbol{\theta}}\{(1 + u)y\}| < \infty$ *for* $\boldsymbol{\theta} \in B$. *Further, there exists a nonnegative function* K *with* $\int K(y)dy < \infty$, *such that* $\sup_{\boldsymbol{\theta}\in B} \|\dot{f}_{\boldsymbol{\theta}}(y)\| \le K(y)$ *for all* $y \in \mathbb{R}$.
(7.4). $\sup_{\boldsymbol{\theta}\in B,\, y\in\mathbb{R}}(1 + |y|)f_{\boldsymbol{\theta}}(y)$, $\sup_{\boldsymbol{\theta}\in B,\, y\in\mathbb{R}} \|(\partial^2/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T)F_{\boldsymbol{\theta}}(y)\|^{2+d}$, $\sup_{\boldsymbol{\theta}\in B,\, y\in\mathbb{R}} \|\dot{F}_{\boldsymbol{\theta}}(y)\|$, *and* $\sup_{\boldsymbol{\theta}\in B} \int |1 - x|^{2+d} f_{\boldsymbol{\theta}}(x)dx$ *are finite.*
(7.5). There exists a uniformly continuous function $r : \mathbb{R} \mapsto [0, \infty)$, *such that* (a) $E[r^2(\varepsilon_0)] < \infty$ *and* $\sup_{\boldsymbol{\theta}\in\Theta} \int r^2(x) f_{\boldsymbol{\theta}}(x) < \infty$, (b) $|\hbar'(\boldsymbol{\theta}; x)| \le r(x)$ *and* $|\hbar''(\boldsymbol{\theta}; x)| \le r(x)$ *for* $\boldsymbol{\theta} \in \Theta$, *and* (c) $\sup_{\boldsymbol{\theta}\in B} \|u(\boldsymbol{\theta}; x)\| \le r(x)$ *for* $u = \dot{\hbar}', \ddot{\hbar}$ *and* $\ddot{\hbar}'$.

Assumptions 1–5 involve only the semiparametric model $X_i = \mu_i(\boldsymbol{\phi}) + \Psi_i^{1/2}(\boldsymbol{\phi})\varepsilon_i$ leaving the error distribution F^0 unspecified. This model has been studied extensively in the literature for various special cases, such as GARCH(p,q) and AGARCH(p,q), and Assumptions 1–5 have appeared as lemmas in those studies. Therefore, these assumptions are well-entrenched in the literature.

Assumption 1 has been used in the literature to prove the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$, which requires $E|X_0|^4 < \infty$ (cf. Lee and Hansen 1994; Berkes and Horváth 2004; Francq and Zakoïan 2004; Ling 2007; Straumann and Mikosch 2006). If $E|X_0|^4 = \infty$ then $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ does not converge to a normal distribution but to a stable law (cf. Hall and Yao 2003). Our assumption

$E|X_0|^{(4+d)} < \infty$ for some $d > 0$, is practically equivalent to $E|X_0|^4 < \infty$. Assumption 1 also requires the true value $\boldsymbol{\phi}_0$ to be an interior point. This condition may not allow overspecifications of $\Psi_i(\boldsymbol{\phi})$ under the null hypothesis. As an example, suppose that the true DGP is GARCH(1,1) and the null hypothesis specifies GARCH(2,2) with $\Psi_i(\boldsymbol{\phi}) = \phi_1 + \phi_2 X_{i-1}^2 + \phi_3 X_{i-2}^2 + \phi_3 \Psi_{i-1}(\boldsymbol{\phi}) + \phi_4 \Psi_{i-2}(\boldsymbol{\phi})$, $\phi_1 > \alpha_L > 0,\ \phi_2, \phi_3, \phi_4, \phi_5 \geq 0$. Then, the true parameter $\boldsymbol{\phi}_0$ lies on the boundary of the parameter space, and hence Assumption 1 is not satisfied (see, e.g, Remark 7.2 in Francq and Zakoïan 2010).

The $e.a.s$ rate of convergence assumed in Assumption 2 plays a central role throughout this paper. It is satisfied by many models of the form $\mathcal{M}$, including GARCH(p,q) and AGARCH(p,q) (see Straumann and Mikosch 2006). In general, it is satisfied by models that are defined by *stochastic recurrence equations* as in (1) that satisfy a certain *contractive on average* property (see Straumann and Mikosch 2006). The $e.a.s.$ rate of convergence in Assumption 2 helps to show that the remainder terms that appear in various Taylor series expansions, converge to zero faster than $n^{-1/2}$. Consequently, their effects become asymptotically negligible. Another use of the $e.a.s.$ rate of convergence in Assumption 2 is that the effect of the initial values on inference based on the QMLE, and the effects of substituting $(\widetilde{\Psi}_i, \widetilde{\mu}_i)$ for (Ψ_i, μ_i) become asymptotically negligible.

An expansion of the form in Assumption 3 is typically satisfied by the Gaussian QMLE (cf. Ling 2007; Francq and Zakoïan 2004; Lee and Hansen 1994; Berkes and Horváth 2004). Assumption 4 is a slightly stronger version of conditions (8.3.2) and (8.3.3) on page 381 in Koul (2002). Assumptions 3, 4, and 5 are satisfied if (a) Assumptions 1 and 2 are satisfied, (b) $E[\ddot{\ell}_0(\boldsymbol{\phi}_0)]$ and $E[\dot{\ell}_0(\boldsymbol{\phi}_0)\dot{\ell}_0(\boldsymbol{\phi}_0)^\top]$ are invertible, and (c) there exist a compact neighbourhood $K_{\boldsymbol{\phi}}(\subset \Phi)$ of $\boldsymbol{\phi}_0$ and a $d > 0$, such that $\sup_{\bar{\boldsymbol{\phi}} \in K_{\boldsymbol{\phi}}} E\|Y_0(\bar{\boldsymbol{\phi}})\|^{2+d} < \infty$ for $Y_0 \in \{\tau_0, \varphi_0, \ddot{\Psi}_0\{\Psi_0\}^{-1/2}, \ddot{\mu}_0\{\Psi_0\}^{-1/2}\}$. Thus, Assumptions 3, 4, and 5 follow essentially from moment conditions previously used in the literature.

Assumptions 6 and 7 are about the parametric form $F_{\boldsymbol{\theta}}$ for the error distribution. The only high-level condition in Assumptions 6 and 7 is that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ as $n \to \infty$. This is a mild condition since it turns out that $\widetilde{\varepsilon}_i$ and ε_i are close for large i. The other parts in Assumptions 6 and 7 can be verified directly for a given $F_{\boldsymbol{\theta}}$. For most parametric families, Assumption 6 holds when $\hat{\boldsymbol{\theta}}$ is based on $\hbar(\boldsymbol{\theta}; x) = \log f_{\boldsymbol{\theta}}(x)$; this is not surprising since $\hat{\boldsymbol{\theta}}$ is the MLE in the iid setting. The condition $\int |y|^{4+d} dF_{\boldsymbol{\theta}}(y) < \infty$ as in Assumption 7.3 is mild since the asymptotic normality of the QMLE $\hat{\boldsymbol{\phi}}$ requires $E|\varepsilon_1|^4 < \infty$. Conditions similar to Assumption 7.1 have been used in Koul and Ossiander (1994) and Koul and Ling (2006). Assumption 7.5 is typically satisfied when $\hbar(\boldsymbol{\theta}; x) = \log f_{\boldsymbol{\theta}}(x)$. For example, for the case $f_{\boldsymbol{\theta}}$ is the normal density and $\hbar(\boldsymbol{\theta}; x) = \log f_{\boldsymbol{\theta}}(x)$, Assumption 7.5 holds with $r(x) = |x|$. Similarly, this condition is also satisfied by many other choices for $f_{\boldsymbol{\theta}}$.

3.2 Weak convergence of the test statistic under the null hypothesis

Let $D[0,1]$ and $D[-\infty,\infty]$ denote the spaces of *cadlag* functions on $[0,1]$ and $[-\infty,\infty]$ respectively, equipped with the uniform metric. The next theorem says that, under the null hypothesis, each test statistic converges weakly to a distribution that depends on unknown nuisance parameters.

Theorem 1. *Suppose that H_0 holds and Assumptions 1-7 are satisfied. Then, we have the following:*

1. $\widetilde{W}_n \stackrel{w}{\Rightarrow} W_0$ *in* $D[-\infty,\infty]$*, where* W_0 *is a centered Gaussian process.*

2. *There exists a continuous functional* $\mathfrak{g}_j : D[0,1] \to \mathbb{R}$ *such that* $T_j = \mathfrak{g}_j\{\widetilde{W}_n \circ F_{\boldsymbol{\theta}_0}^{-1}\} + o_p(1)$ *and* $T_j \stackrel{d}{\to} \mathfrak{g}_j\{G_0\}$ *as* $n \to \infty$ *where* G_0 *is a centered Gaussian process with the same covariance kernel as for* $W_0 \circ F_{\boldsymbol{\theta}_0}^{-1}$ *and* W_0 *is the process in the previous part* ($j = 1, \ldots, 5$).

The proof is given in the Appendix at the end of this paper. The definition of the centered Gaussian process W_0 is provided in the online supplement. The method that we adopt for showing that $\widetilde{W}_n \stackrel{w}{\Rightarrow} W_0$ is not standard, and involves some new techniques/results based on stochastic recurrence equations, $\stackrel{e.a.s.}{\to}$ rate of convergence, and asymptotic uniform expansions of weighted empirical processes. The results in Escanciano (2010), in particular Theorem A1 therein, have paved the way to the development of our new techniques. The proof of part 2 of our theorem follows from the continuous mapping theorem and the weak convergence of $\widetilde{W}_n$.

4 Asymptotic validity of the bootstrap test

This section establishes the asymptotic validity of the bootstrap test (8).

4.1 Regularity conditions

Let $\bar{\boldsymbol{\omega}} := (\bar{\boldsymbol{\phi}}^\top, \bar{\boldsymbol{\theta}}^\top)^\top$ denote an arbitrary point in $\Phi \times \Theta$. Let $\{X_i(\bar{\boldsymbol{\omega}}), \mathit{\Psi}_i(\bar{\boldsymbol{\omega}}), \mu_i(\bar{\boldsymbol{\omega}})\}_{i\in\mathbb{Z}}$ denote the stationary process defined by model $\mathcal{M}$ except that the population parameter value of the data generating process [DGP] is $\bar{\boldsymbol{\omega}}$, which may be different from the true value $\boldsymbol{\omega}_0$ of the target population. In the previous section, $\bar{\boldsymbol{\omega}}$ was fixed at $\boldsymbol{\omega}_0 := (\boldsymbol{\phi}_0^\top, \boldsymbol{\theta}_0^\top)^\top$ and we suppressed $\boldsymbol{\omega}_0$; for example, we wrote X_i for $X_i(\boldsymbol{\omega}_0)$. Now, to study bootstrap, we need to explicitly show the population parameter $\bar{\boldsymbol{\omega}}$ for the DGP. Let $\mu_i(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}})$ and $\mathit{\Psi}_i(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}})$ based on $\{X_i(\bar{\boldsymbol{\omega}})\}_{i\geq 0}$ be defined as

$$\mu_i(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}}) = h\big\{\boldsymbol{X}_{i-1}(\bar{\boldsymbol{\omega}}), \boldsymbol{\mu}_{i-1}(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}}); \boldsymbol{\phi}\big\}, \quad \mathit{\Psi}_i(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}}) = g\big\{\boldsymbol{\mathit{\Psi}}_{i-1}(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}}), \boldsymbol{\eta}_{i-1}(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}}); \boldsymbol{\phi}\big\}, \tag{10}$$

where $\eta_i(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}}) = X_i(\bar{\boldsymbol{\omega}}) - \mu_i(\boldsymbol{\phi}; \bar{\boldsymbol{\omega}})$, ($i \in \mathbb{Z}$; $\boldsymbol{\phi} \in \Phi$); note that $\boldsymbol{\phi} \in \Phi$ and $\bar{\boldsymbol{\phi}} \in \Phi$ play different roles.

Next, we modify (4) to define the corresponding quantities at the population parameter $\bar{\boldsymbol{\omega}}$ instead of at $\boldsymbol{\omega}_0$. Let $\widetilde{\mu}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})$ and $\widetilde{\Psi}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})$ corresponding to $\{X_i(\bar{\boldsymbol{\omega}})\}_{i\geq 0}$ be defined recursively as

$$\widetilde{\mu}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) = h\big\{\boldsymbol{X}_{i-1}(\bar{\boldsymbol{\omega}}), \widetilde{\boldsymbol{\mu}}_{i-1}(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}); \boldsymbol{\phi}\big\}, \quad \widetilde{\Psi}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) = g\big\{\widetilde{\boldsymbol{\Psi}}_{i-1}(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}), \widetilde{\boldsymbol{\eta}}_{i-1}(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}); \boldsymbol{\phi}\big\}, \tag{11}$$

where $\widetilde{\eta}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) = X_i(\bar{\boldsymbol{\omega}}) - \widetilde{\mu}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})$, conditional on $(X_i(\bar{\boldsymbol{\omega}}), \widetilde{\Psi}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}), \widetilde{\mu}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})) = (\bar{X}, 1, \bar{X})$ for $i \leq 0$.

Let $K_{\boldsymbol{\omega}}$ denote a 'small' set containing an open neighbourhood of $\boldsymbol{\omega}_0 \equiv (\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$; the precise choice of $K_{\boldsymbol{\omega}}$ is made at the time of verifying the regularity conditions. Therefore, $\hat{\boldsymbol{\omega}}$ lies in $K_{\boldsymbol{\omega}}$ with probability approaching one. To establish bootstrap validity, we will consider DGPs with true parameter $\bar{\boldsymbol{\omega}}$, where $\bar{\boldsymbol{\omega}} \in K_{\boldsymbol{\omega}}$. Recall that a regularity condition to establish the weak convergence of the test statistic is that $\sup_{\boldsymbol{\phi}\in\Phi} |\widetilde{\Psi}_i(\boldsymbol{\phi};\boldsymbol{\omega}_0) - \Psi_i(\boldsymbol{\phi};\boldsymbol{\omega}_0)| \overset{e.a.s.}{\to} 0$ as $i \to \infty$ (see Assumption 2). To establish validity of the bootstrap, we need the corresponding stronger uniform convergence, $\sup_{\bar{\boldsymbol{\omega}}\in K_{\boldsymbol{\omega}}} \sup_{\boldsymbol{\phi}\in\Phi} |\widetilde{\Psi}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) - \Psi_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})| \overset{e.a.s.}{\to} 0$ as $i \to \infty$. The crux of Condition M, stated below for validity of bootstrap is that Assumptions 1–7 hold uniformly over $\bar{\boldsymbol{\omega}} \in K_{\boldsymbol{\omega}}$.

Condition M. There exist compact neighbourhoods $K_{\boldsymbol{\phi}}(\subset \Phi)$ and $K_{\boldsymbol{\theta}}(\subset \Theta)$ of $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$, respectively, such that the following conditions hold with $K_{\boldsymbol{\omega}} = K_{\boldsymbol{\phi}} \times K_{\theta}$:

(M1). (a) *At every $\bar{\boldsymbol{\omega}} \in K_{\boldsymbol{\omega}}$, there exists a unique stationary ergodic process $\{X_i(\bar{\boldsymbol{\omega}}), \Psi_i(\bar{\boldsymbol{\omega}}), \mu_i(\bar{\boldsymbol{\omega}})\}_{i\in\mathbb{Z}}$, satisfying model $\mathcal{M}$ defined by (1).* (b) $\sup_{(\boldsymbol{\phi},\bar{\boldsymbol{\omega}})\in\Phi\times K_{\boldsymbol{\omega}}} |\widetilde{Y}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) - Y_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})| \overset{e.a.s.}{\to} 0$ *as $i \to \infty$ for $Y_i \in \{\mu_i, \Psi_i\}$. [The symbol $\widetilde{Y}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})$ for $Y_i = \mu_i$, denotes $\widetilde{\mu}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})$.]*
(c) *For $Y_i \in \{\dot{\mu}_i, \dot{\Psi}_i, \ddot{\Psi}_i, \ddot{\mu}_i\}$, we have* $\sup_{(\boldsymbol{\phi},\bar{\boldsymbol{\omega}})\in K_{\boldsymbol{\phi}}\times K_{\boldsymbol{\omega}}} \|\widetilde{Y}_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) - Y_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})\| \overset{e.a.s.}{\to} 0$ *as $i \to \infty$.*

(M2). *For the process $\{X_i(\bar{\boldsymbol{\omega}}), \Psi_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}), \mu_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})\}$, let the analog of $\{\tau_i(\boldsymbol{\phi}), \varphi_i(\boldsymbol{\phi}), \ell_i(\boldsymbol{\phi})\}$ be denoted by $\{\tau_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}), \varphi_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}), \ell_i(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})\}$. Let $E^{(\bar{\boldsymbol{\omega}})}$ and $P^{(\bar{\boldsymbol{\omega}})}$ denote the corresponding expectation and probability measure, respectively. Then, we have the following:*
(a) *For some $d > 0$,* $\sup_{(\boldsymbol{\phi},\bar{\boldsymbol{\omega}})\in K_{\boldsymbol{\phi}}\times K_{\boldsymbol{\omega}}} E^{(\bar{\boldsymbol{\omega}})}\|Y_0(\boldsymbol{\phi};\bar{\boldsymbol{\omega}})\|^{2+d} < \infty$ *for $Y_0 \in \{\tau_0, \varphi_0, \ddot{\Psi}_0\Psi_0^{-1/2}, \ddot{\mu}_0\Psi_0^{-1/2}\}$.*
(b) $E^{(\bar{\boldsymbol{\omega}})}[\dot{\ell}_0(\bar{\boldsymbol{\phi}};\bar{\boldsymbol{\omega}})\dot{\ell}_0(\bar{\boldsymbol{\phi}};\bar{\boldsymbol{\omega}})^\top]$, $E^{(\bar{\boldsymbol{\omega}})}[\ddot{\ell}_0(\bar{\boldsymbol{\phi}};\bar{\boldsymbol{\omega}})]$, *and* $E^{(\bar{\boldsymbol{\omega}})}[\ddot{\hbar}(\bar{\boldsymbol{\theta}};\varepsilon_0(\bar{\boldsymbol{\omega}}))]$ *are nonsingular ($\bar{\boldsymbol{\omega}} \in K_{\boldsymbol{\omega}}$).*

(M3). (a) *Assumptions 6 and 7 in Section 3.1 hold with $B = K_{\boldsymbol{\theta}}$.*
(b) *Let $\boldsymbol{\omega}_n := (\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)$ denote a given non-random sequence such that $\boldsymbol{\omega}_n \to \boldsymbol{\omega}_0$ as $n \to \infty$. Let the analog of $\{\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}\}$, based on $\{X_i(\boldsymbol{\omega}_n), \Psi_i(\boldsymbol{\phi};\boldsymbol{\omega}_n), \mu_i(\boldsymbol{\phi};\boldsymbol{\omega}_n) : i = 1, \ldots, n\}$, be denoted by $\{\hat{\boldsymbol{\phi}}(\boldsymbol{\omega}_n), \hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_n)\}$. Let $h_{\boldsymbol{\theta}}^{(\boldsymbol{\omega})}(t) = \{h_{\boldsymbol{\theta}1}^{(\boldsymbol{\omega})}(t), \ldots, h_{\boldsymbol{\theta}q}^{(\boldsymbol{\omega})}(t)\}^\top := -E^{(\boldsymbol{\omega})}[\ddot{\hbar}(\boldsymbol{\theta};\varepsilon_0(\boldsymbol{\theta}))]^{-1}\dot{\hbar}(\boldsymbol{\theta};t)$. Then, there exists $d > 0$ such that $\sup_{\boldsymbol{\omega}\in K_{\boldsymbol{\omega}}} \int \|h_{\boldsymbol{\theta}}^{(\boldsymbol{\omega})}(t)\|^{2+d} f_{\boldsymbol{\theta}}(t)\, dt < \infty$. Further, $\|\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_n) - \boldsymbol{\theta}_n\| = o_{P(\boldsymbol{\omega}_n)}(1)$ and*

$$\hat{\theta}_j(\boldsymbol{\omega}_n) - \theta_{nj} = \frac{1}{n}\sum_{i=1}^{n} h_{\boldsymbol{\theta}_n j}^{(\boldsymbol{\omega}_n)}(\varepsilon_i(\boldsymbol{\theta}_n)) - [\hat{\boldsymbol{\phi}}(\boldsymbol{\omega}_n) - \boldsymbol{\phi}_n]^\top \mathcal{Q}_{nnj} + o_{P(\boldsymbol{\omega}_n)}(n^{-1/2}) \quad (j = 1, \ldots, q),$$

where $\mathcal{Q}_{nnj} = n^{-1}\sum_{i=1}^{n}[\varphi_i(\boldsymbol{\phi}_n;\boldsymbol{\omega}_n) + \tau_i(\boldsymbol{\phi}_n;\boldsymbol{\omega}_n)\varepsilon_i(\boldsymbol{\omega}_n)/2] h'^{(\boldsymbol{\omega}_n)}_{\boldsymbol{\theta}_n j}(\varepsilon_i(\boldsymbol{\theta}_n))$.

(M4). (a) *Assumption 1.3 in Section 3.1 holds.* (b) *For any* $\bar{\boldsymbol{\omega}} \in K_{\boldsymbol{\omega}}$*, if* $\mu_0(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) \overset{a.s.}{=} \mu_0(\bar{\boldsymbol{\phi}};\bar{\boldsymbol{\omega}})$ *and* $\Psi_0(\boldsymbol{\phi};\bar{\boldsymbol{\omega}}) \overset{a.s.}{=} \Psi_0(\bar{\boldsymbol{\phi}};\bar{\boldsymbol{\omega}})$ *then* $\boldsymbol{\phi} = \bar{\boldsymbol{\phi}}$*.*

Next, we comment on the foregoing conditions and then state the validity of the bootstrap test in the next subsection. Condition (M1)(a) is essentially the same as Assumption (1.1), except that the latter is about the existence of a solution to the defining equations at the single value $\boldsymbol{\omega}_0$ while (M1)(a) is for any fixed value in a neighbourhood of $\boldsymbol{\omega}_0$. In practice, there is no difference between these two conditions since $\boldsymbol{\omega}_0$ is arbitrary and unknown. Conditions (M1)(b) and (M1)(c) correspond to Assumption 2; this is the core of Condition M. The difference is that Condition (M1) requires the convergence in Assumption 2 to hold uniformly in a small neighbourhood of $\boldsymbol{\omega}_0$ rather than only at the fixed point $\boldsymbol{\omega}_0$. If Condition (M1) is satisfied then Assumption 2 holds when the true DGP corresponds to either (a) $\boldsymbol{\omega}_n$ with $\boldsymbol{\omega}_n \to \boldsymbol{\omega}_0$, or (b) $\hat{\boldsymbol{\omega}} \in K_{\boldsymbol{\omega}}$, conditional on $\{X_1, \ldots, X_n\}$. The main purpose of the stochastic recurrence equation method, developed and illustrated in the Supplement for GARCH(p, q), is to verify (M1)(b) and (M1)(c).

Condition (M2) is mild and most of it is already contained in Assumptions 1–5. Finiteness of the moments relating to τ_0 and φ_0 follow from Assumption (1.4). Condition (M2)(b) is a slight variation of Assumptions (3.2) and (6.1). Condition (M3)(b) is essentially an extension of Proposition 3 in the Appendix to the triangular array setup. Once (M2) has been established, (M3)(b) should follow without much difficulty in much the same way as for the Quasi-loglikelihood function. Condition (M4)(b) is essentially the same identifiability requirement as Assumption (1.2).

4.2 The main result on the asymptotic validity of the bootstrap test

First, we recall some standard notation. Let $O_{p_n^*}$, $o_{p_n^*}$, and E_* denote the usual stochastic orders of magnitude and expectation, respectively, with respect to the bootstrap law, P_n^*, conditional on $\{X_1, \ldots, X_n\}$. The convergence in distribution of bootstrap statistics is denoted by '$\xrightarrow{d^*}$'. For example, '$T_j^* \xrightarrow{d^*} \mathfrak{g}_j(G_0)$, *in probability*' means that $P_n^*(T_j^* \leq z) \xrightarrow{p} P\{\mathfrak{g}_j(G_0) \leq z\}$, at every continuity point z of $P\{\mathfrak{g}_j(G_0) \leq z\}$. The convergence of bootstrapped values such as T_j^* and of processes such as $\widetilde{W}_n^*\{F_{\hat{\boldsymbol{\theta}}}^{-1}(\cdot)\}$ presented below are '*in probability*', and they are valid irrespective of whether or not H_0 is true.

Theorem 2. *Suppose that Conditions (M1)–(M4) are satisfied. Additionally, assume that* $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ *converges in probability to* $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$*, and that* $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$ *is an interior point in* $\Phi \times \Theta$*. Let* G_0 *be the limit process introduced in Theorem 1 except that* $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$ *is the pseudo-true value if* H_0 *is not satisfied. Then, conditional on* $\{X_1, \ldots, X_n\}$*,*

1. $\widetilde{W}_n^* \circ F_{\hat{\boldsymbol{\theta}}}^{-1}$ *converges weakly to* G_0*, in probability.*

2. $\mathfrak{g}\{\widetilde{W}_n^* \circ F_{\hat{\boldsymbol{\theta}}}^{-1}\} \xrightarrow{d^*} \mathfrak{g}\{G_0\}$*, in probability, for any continuous functional* $\mathfrak{g} : D[0,1] \to \mathbb{R}$*.*

3. *There exists a continuous functional* $\mathfrak{g}_j : D[0,1] \to \mathbb{R}$ *such that* $T_j^* = \mathfrak{g}_j(\widetilde{W}_n^* \circ F_{\hat{\boldsymbol{\theta}}}^{-1}) + o_{p_n^*}(1)$*, in probability* $(j = 1, \ldots, 5)$*.*

4. *The* $\hat{p}$ *used in (8) is a valid large sample estimate of the p-value, and the bootstrap test (8) based on* T_j *is asymptotically valid* $(j = 1, \ldots, 5)$*.*

We close this section with a few comments on the possibility of extending the methodology developed in this paper to the multivariate GARCH family. To establish the asymptotic validity of our tests we show that the asymptotic approximations derived for a fixed true value $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$, see, e.g., Straumann and Mikosch (2006), hold uniformly over populations having the true parameter in a neighbourhood of $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$. In order to extend the methodology developed in this paper to the multivariate GARCH we need to first extend the aforementioned type of uniform convergence results to a multivariate setup. A particularly important property of the stochastic recurrence equations defining the GARCH family that we use in our proofs is that it is "contractive on average", which is established by showing that the top Lyapunov exponent is less than one. It is encouraging that some multivariate GARCH models (see (A2) on page 295 in Francq and Zakoïan, 2010) satisfy such a condition. This indicates that the asymptotic approximations that we derive for a fixed true value may also be extended to a multivariate setting. However, it is not clear whether such expansions would also hold uniformly over a local parameter space in a multivariate setting. In addition, to establish bootstrap validity, we develop crucial results in Section S3 in the Supplement, in particular Theorem S1. This theorem provides a uniform approximation of weighted empirical processes in a triangular array setting. If we were to extend our proofs to multivariate GARCH, we need to extend this theorem to the multivariate setting. However, the theoretical foundation of uniform convergence of multivariate weighted empirical processes is not yet well-developed, and hence it is unclear whether or not such a result would hold in the multivariate setting. Thus, although there are some promising indications that our methodology could potentially be extended to multivariate GARCH, there are several significant technical challenges that need to be resolved first.

5 Asymptotic power of the tests

Since the model under the null hypothesis is fully parametric and the test does not involve kernel type nonparametric components, one would expect that the proposed tests would have non-trivial limiting power against certain local alternatives converging to the null hypothesis at the parametric rate $O(n^{-1/2})$. We consider two such sequences of local alternatives, each corresponding to violating one assumption in H_0. The first assumes that only the error distribution is misspecified, and the second assumes that only the conditional mean and/or conditional variance is misspecified.

5.1 Local power when only the error distribution is misspecified

Let $\widetilde{F}$ denote a cdf satisfying the conditions imposed on the true error distribution. Let $\delta \geq 0$ be fixed but arbitrary. Define $m(\cdot) := \{\widetilde{F}(\cdot) - F_{\boldsymbol{\theta}_0}(\cdot)\} - [\int h_{\boldsymbol{\theta}_0}(\varepsilon)\ d\widetilde{F}(\varepsilon)]^\top \dot{F}_{\boldsymbol{\theta}_0}(\cdot)$ and $F_{(n)} := (1 - n^{-1/2}\delta)F_{\boldsymbol{\theta}_0} + n^{-1/2}\delta\widetilde{F}$ $(n > \delta^{-2})$. Suppose that $F_{(n)} \notin \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$. Consider the sequence of local alternatives, H_{1n} : *The model* (1) *holds except that the common error distribution is* $F_{(n)}$.

Theorem 3. *Suppose that Assumptions 1-7 hold. Further, assume that* $\sup_{x\in\mathbb{R}}[\tilde{f}(x)/f_{\boldsymbol{\theta}_0}(x)] < \infty$, $\dot{F}_{\boldsymbol{\theta}_0}(\cdot)$ *is continuous on* $\mathbb{R}$, *and* $\dot{F}_{\boldsymbol{\theta}_0}(-\infty) = \mathbf{0} = \dot{F}_{\boldsymbol{\theta}_0}(\infty)$. *Then, under* H_{1n}, $\widetilde{W}_n(\cdot) \stackrel{w}{\Rightarrow} \delta m(\cdot) + W_0(\cdot)$, *where* W_0 *is the centered Gaussian process in Theorem 1. In consequence, the asymptotic local power of* T_j $(j = 1, \ldots, 5)$ *strictly increases with* δ; *further, the asymptotic distribution of the test statistic under* H_{1n} *strictly stochastically dominates that under* H_0. *Additionally, suppose that (M1)–(M4) are also satisfied. Then, conditional on* $\{X_1, \ldots, X_n\}$, *the aforementioned asymptotic local power properties of* T_j *carry over to the corresponding bootstrap test in* (8), *in probability* $(j = 1, \ldots, 5)$.

To prove this theorem, we apply an extension of the Anderson lemma to Gaussian processes (Lewandowski *et al.* 1995). Andrews (1997) studied Kolmogorov–Smirnov type specification tests for a special case of the setting in this paper. Andrews conjectured that Kolmogorov type tests have local asymptotic power strictly larger than the asymptotic size of the test. Our foregoing theorem shows that the result that Andrews (1997) conjectured is correct. Finally, let us note that to prove the strict stochastic dominance result in Theorem 3, an approach based on the Kac–Siegert representation appears difficult.

5.2 Local power when the conditional mean and/or variance is misspecified

Let $r_i = r(X_{i-1}, X_{i-2}, \ldots)$ and $s_i = s(X_{i-1}, X_{i-2}, \ldots)$ be stationary processes satisfying the moment conditions of μ_i and Ψ_i, respectively. Further, suppose that the following probability limits exist: $R_r = \text{plim}\ n^{-1}\sum_{i=1}^n r_i/\Psi_i^{1/2}$, $R_{r\varphi} = \text{plim}\ n^{-1}\sum_{i=1}^n(r_i\varphi_i/\Psi_i^{1/2})$, $R_s = \text{plim}\ n^{-1}\sum_{i=1}^n s_i/\Psi_i$, and $R_{s\tau} = \text{plim}\ n^{-1}\sum_{i=1}^n(s_i\tau_i)/\Psi_i$, where τ_i and φ_i are as in (9). Let $\delta > 0$ be fixed but arbitrary. Further, let H_{2n} : $\quad X_i = (\mu_i + n^{-1/2}\delta r_i) + (\Psi_i + n^{-1/2}\delta s_i)^{1/2}\varepsilon_i \quad (i \in \mathbb{N})$ *with the rest of the model specification in* $\mathcal{M}$ *being correct*; this setting is similar to that in Ling and Tong (2011). Thus, the only misspecification is in the conditional mean and/or variance. Let λ_n denote the likelihood ratio for H_0 against H_{2n}. Then, we have $\log(\lambda_n) = \delta n^{-1/2}\sum_{i=1}^{n} K_i + R_n$, for some $\{K_i\}_{i\in\mathbb{N}}$ and R_n. As in the proof of Theorem 3 in the previous section on local power, it may be verified that $\log(\lambda_n)$ is asymptotically $N(-2^{-1}\sigma^2, \sigma^2)$ for some $\sigma > 0$, and H_0 is contiguous to $\{H_{2n}\}_{n\in\mathbb{N}}$. Hence the techniques in the previous subsection can be applied for deriving the asymptotic local power against H_{2n}. The proof of the next theorem is similar to that of Theorem 3, and hence is omitted.

To simplify notation, let us write f, F and $\dot{F}$ for $f_{\boldsymbol{\theta}_0}, F_{\boldsymbol{\theta}_0}$, and $\dot{F}_{\boldsymbol{\theta}_0}$ respectively. Let $A = A(\boldsymbol{\phi}_0)$, where $A(\boldsymbol{\phi}) = -[E\ddot{\ell}_0(\boldsymbol{\phi})]^{-1}$ (see Assumption 3). Let $h(\varepsilon) = h_{\boldsymbol{\theta}_0}(\varepsilon)$, where $h_{\boldsymbol{\theta}}(\varepsilon)$ is defined in Assumption 6. Let $\tau_i = \dot{\Psi}_i(\boldsymbol{\phi}_0)/\Psi_i(\boldsymbol{\phi}_0)$, $\varphi_i = \dot{\mu}_i(\boldsymbol{\phi}_0)/\sqrt{\Psi_i(\boldsymbol{\phi}_0)}$, and $\mathcal{Q} = \text{plim}[n^{-1}\sum_{i=1}^{n}\mathcal{Q}_i]$, where $\mathcal{Q}_i = h'(\varepsilon_i)\{\varphi_i + 2^{-1}\varepsilon_i\tau_i\}^\top$; the matrix $\mathcal{Q}_i$ arises in the proof of Theorem 1. Let

$$
\begin{aligned}
M_\mu(x) &= R_r\left[\left\{\int f'(t)h^\top(t)\,dt\right\}\dot{F}(x) - f(x)\right] - 2R_{r\varphi}^\top A\left[\mathcal{Q}^\top\dot{F}(x) + \left\{\frac{x}{2}E[\tau_1] + E[\varphi_1]\right\}f(x)\right]\\
M_\psi(x) &= \frac{R_s}{2}\left[\left\{\int tf'(t)h^\top(t)\,dt\right\}\dot{F}(x) - xf(x)\right] - R_{s\tau}^\top A\left[\mathcal{Q}^\top\dot{F}(x) + \left\{\frac{x}{2}E[\tau_1] + E[\varphi_1]\right\}f(x)\right].
\end{aligned}
$$

Then, we have the following result about the asymptotic local power of the tests.

Theorem 4. *Suppose that Assumptions 1-7 are satisfied. Further, suppose that $\dot{F}(\cdot)$ is continuous on $\mathbb{R}$, and $\dot{F}(-\infty) = \mathbf{0} = \dot{F}(\infty)$. Then, under H_{2n}, $\widetilde{W}_n(\cdot) \overset{w}{\Rightarrow} \delta M(\cdot) + W_0(\cdot)$, in $D[-\infty,\infty]$, where W_0 is the centered Gaussian process in Theorem 1, $M(x) = M_\mu(x) + M_\psi(x)$. Suppose that $M(x) \neq 0$ on a set of positive measure. Then, the asymptotic local power of T_j $(j = 1, \ldots, 5)$ strictly increases with δ; further, the asymptotic distribution of the test statistic under H_{2n} strictly stochastically dominates that under H_0. Additionally, suppose that Conditions (M1)–(M4) are also satisfied. Then, conditional on $\{X_1, \ldots, X_n\}$, the aforementioned asymptotic local power properties of T_j carry over to the corresponding bootstrap test in (8), in probability $(j = 1, \ldots, 5)$.*

This theorem shows that the test has asymptotic local power larger than the asymptotic size of the test if and only if $M(x) \neq 0$ for $x \in D$, where D is some set with positive measure. Due to the nature of the above explicit form of $M(x)$ it is difficult to provide general primitive conditions to characterize the cases when $M(\cdot) = 0$ and hence the local asymptotic power in Theorem 4 is zero. Nevertheless, it is instructive to examine the expression for $M(x)$ in a simple example.

Let us consider the ARCH(1) model $X_i = \Psi_i^{1/2}\varepsilon_i$, where $\Psi_i = \phi_0 + \phi_1 X_{i-1}^2$. We consider the simpler case when the error distribution $F_{\boldsymbol{\theta}} = F^0$ is free of unknown parameters; therefore, there is no parameter $\boldsymbol{\theta}$, and $f(x) > 0$ $(x \in \mathbb{R})$. Then $\dot{F}_{\boldsymbol{\theta}_0}(x) = 0$, $\varphi_i(\boldsymbol{\phi}) = \dot{\mu}_i/\Psi_i = 0$, and $h_{\boldsymbol{\theta}}(\varepsilon) = 0$. Consider the sequence of local alternatives $H_{2n} : X_i = (\Psi_i + n^{-1/2}\delta s_i)^{1/2}\varepsilon_i$, where $s_i = s(X_{i-1}, X_{i-2}, \ldots) = \alpha_0 + \sum_{j=1}^{p_0}\alpha_j X_{i-j}^2$ for some $p_0 \in \mathbb{N}$, $\alpha_0 > 0$, and $0 \leq \alpha_j < 1$ $(j =$

$1,\ldots,p_0$). Then $\{s_i\}$ is strictly stationary and ergodic and $0 < E(s_1) < \infty$. Further, $M_\mu = 0$ and $M_\psi(x) = Pxf(x)$, where $P = -2^{-1}\text{plim}\ n^{-1}\sum_{i=1}^{n}[1+\tau_i^\top AE(\tau_1)]s_i/\Psi_i = -2^{-1}[E(s_1/\Psi_1) + E(\tau_1^\top As_1/\Psi_1)E(\tau_1)]$, $\tau_i = (1, X_{i-1}^2)^\top \Psi_i^{-1}$, and $A_{[2\times 2]} = 2E\{[1, X_{i-1}^2;\ X_{i-1}^2, X_{i-1}^4]\Psi_{i-1}^2\}$. Note that s_i is strictly positive, the elements of τ_i are positive with probability one, the elements of $E(\tau_1)$ and A are positive, and $P = -2^{-1}[E(s_1/\Psi_1)+E(\tau_1^\top As_1/\Psi_1)E(\tau_1)] > 0$. Therefore, $M(x) = Pxf(x) \neq 0$ for every $x \neq 0$. Therefore, our bootstrap test has asymptotic local power larger than its asymptotic size. Similarly, our test has non-trivial asymptotic power against H_{2n} for any stationary sequence $\{s_i\}$ satisfying $E(s_1/\Psi_1) + E(\tau_1^\top As_1/\Psi_1)E(\tau_1) \neq 0$. However, if $E(s_1/\Psi_1) = -E(\tau_1^\top As_1/\Psi_1)E(\tau_1)$ then our test does not have nontrivial local power.

5.3 Consistency of the tests against fixed alternatives

We say that the bootstrap test based on T_j is *consistent* if, conditional on $\{X_1,\ldots,X_n\}$, the p-value $\hat{p}$ in (8) converges to 0, in probability, as $n\to\infty$, under the fixed alternative. Suppose that the true DGP is different from that in H_0, which would be the case if one of the parametric forms $\{\Psi_i(\boldsymbol{\phi}) : \boldsymbol{\phi}\in\Phi\}$, $\{\mu_i(\boldsymbol{\phi}) : \boldsymbol{\phi}\in\Phi\}$, or $\{F_{\boldsymbol{\theta}} : \boldsymbol{\theta}\in\Theta\}$ is not correctly specified. For simplicity, we consider only the Kolmogorov–Smirnov type statistic $T_1 := n^{1/2}\|\widetilde{F}_n - F_{\hat{\boldsymbol{\theta}}}\|_\infty$. In general, we would expect $\widetilde{F}_n$, being the empirical distribution of the residuals $\{\widetilde{\varepsilon}_1,\ldots,\widetilde{\varepsilon}_n\}$, to converge to a distribution function under some conditions, which may include that the true DGP is stationary. Let us suppose that this is the case, and there exists a distribution function G_a such that $\|\widetilde{F}_n - G_a\|_\infty = o_p(1)$.

Under the assumptions of Theorem 2, we have $\|F_{\hat{\boldsymbol{\theta}}} - F_{\boldsymbol{\theta}_0}\|_\infty = o_p(1)$ and hence $n^{-1/2}T_1 := \|\widetilde{F}_n - F_{\hat{\boldsymbol{\theta}}}\|_\infty \xrightarrow{p} \|G_a - F_{\boldsymbol{\theta}_0}\|_\infty$. In general, we do not know the functional form of G_a or the value of $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty$. If $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty \neq 0$ then $T_1 = n^{1/2}\|\widetilde{F}_n - F_{\hat{\boldsymbol{\theta}}}\|_\infty \xrightarrow{p} \infty$ as $n\to\infty$, and hence the bootstrap test based on T_1 would be consistent. Next, we consider two fixed alternatives, show that $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty \neq 0$, and deduce the consistency of the bootstrap test.

Suppose that the assumptions of Theorem 2 are satisfied, the error distribution F^0 is not of the form $F_{\boldsymbol{\theta}}$ ($\boldsymbol{\theta}\in\Theta$), and the rest of the specifications in H_0 are correct. Since F^0 is not of the form $F_{\boldsymbol{\theta}}$ ($\boldsymbol{\theta}\in\Theta$) and Θ is compact, we have $\|F^0 - F_{\boldsymbol{\theta}_0}\|_\infty > 0$. Since the only misspecification is in the error distribution, it follows that the Gaussian QMLE $\hat{\boldsymbol{\phi}}$ is consistent for the true value $\boldsymbol{\phi}_0$, $\|\widetilde{F}_n - F^0\|_\infty = o_p(1)$ (see the proof of Theorem 1), the aforementioned G_a may be chosen as F^0, and $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty = \|F^0 - F_{\boldsymbol{\theta}_0}\|_\infty > 0$. Therefore, the bootstrap test based on T_1 is consistent.

Next, we consider consistency of the test when the conditional variance is misspecified; for simplicity, we consider the case when there is no mean function μ_i. Suppose that the null model is $H_0 : X_i = \{\Psi_i(\boldsymbol{\phi})\}^{1/2}\varepsilon_i$, and that the true DGP is $X_i = g_i^{1/2}e_i$, where $\{X_i, g_i, e_i\}$ is stationary and $\{e_i\}$ is iid with common continuous distribution function $F^e(\cdot)$. Let $B_i = \Psi_i^{1/2}(\boldsymbol{\phi}_0)g_i^{-1/2}$ and $G_a(x) = E[F^e(B_i x)]$ $(x \in \mathbb{R})$, where the expectation $E(\cdot)$ is with respect to the true DGP.

To study consistency of the test, let us suppose that g_i is not of the form $\Psi_i(\cdot)$, and hence the conditional variance is misspecified. Then, $\{B_i\}_{i\in\mathbb{N}}$ forms a sequence of identically distributed nondegenerate random variables, and $G_a(x) := E[F^e(B_i x)]$ is a continuous cdf. Further, since $\widetilde{F}_n(x) = n^{-1}\sum_{i=1}^{n} I(\widetilde{\varepsilon}_i \le x)$ $(x \in \mathbb{R})$, where $\widetilde{\varepsilon}_i = \widetilde{\Psi}_i^{-1/2}(\hat{\boldsymbol{\phi}})g_i^{1/2}e_i$ $(i = 1,\ldots,n)$, it can be shown using the regularity conditions of the bootstrap tests and by applying the Ergodic theorem that $\widetilde{F}_n(x) = G_a(x) + o_p(1)$ $(x \in \mathbb{R})$. Then, by using a Glivenko-Cantelli type argument, we obtain that the uniform convergence $\|\widetilde{F}_n - G_a\|_\infty = o_p(1)$ also holds, and hence $\|\widetilde{F}_n - F_{\hat{\boldsymbol{\theta}}}\|_\infty \xrightarrow{p} \|G_a - F_{\boldsymbol{\theta}_0}\|_\infty$. Therefore, if $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty \neq 0$, then $T_1 \xrightarrow{p} \infty$ as $n \to \infty$ and hence the bootstrap test based on T_1 is consistent. Since $\{B_i\}_{i\in\mathbb{N}}$ is nondegenerate, one would expect in this scenario to have $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty \neq 0$, irrespective of whether or not the error distribution is correctly specified; the main feature that drives the consistency of the test in this case is the fact that $\{B_i\}_{i\in\mathbb{N}}$ is nondegenerate under misspecifications of the conditional variance.

Similar type of arguments hold when the null model is the more general form $\mathcal{M}$ that includes the mean function $\mu_i(\cdot)$, and the misspecification may be in any part of the specification in $\mathcal{M}$. Further, such arguments are also applicable to the tests based on $\{T_2,\ldots,T_5\}$ with the norm $\|\cdot\|_\infty$ replaced by that corresponds to the particular statistic. The following proposition summarizes these results.

Proposition 1. *Suppose that the assumptions of Theorem 2 are satisfied, and one of the three parametric specifications $\{\Psi_i(\boldsymbol{\phi}), \mu_i(\boldsymbol{\phi}), F_{\boldsymbol{\theta}}\}$ in H_0 is not correct. Additionally, assume that there exists a distribution function G_a such that $\|\widetilde{F}_n - G_a\|_\infty = o_p(1)$ and $\|G_a - F_{\boldsymbol{\theta}_0}\|_\infty > 0$. Then, the bootstrap test (8) based on T_j is consistent ($j = 1,\ldots,5$).*

In view of Proposition 1 our tests may not have power against misspecifications that allow $\|\operatorname{plim}\widetilde{F}_n - F_{\boldsymbol{\theta}_0}\|_\infty = 0$. The set of such fixed alternatives appear to be small.

6 Verification of the high-level assumptions for GARCH(p, q)

Since the asymptotic validity of the proposed bootstrap tests are established under a set of high-level conditions, it is essential to know how these conditions could be verified for a given model. In this section, we indicate the formulation for verifying the regularity conditions for GARCH(p,q), and the detailed verifications are relegated to the online supplement. The method of verification therein is a step-by-step exercise in mathematics, and it is applicable to a range of models.

Consider the GARCH(p,q) model with error distribution $F_{\boldsymbol{\theta}}$:

$$\mathcal{M}: \quad \begin{cases} X_i &= \{\Psi_i\}^{1/2}\varepsilon_i, \quad \{\varepsilon_i : i \in \mathbb{Z}\} \text{ are iid}, \quad \varepsilon_i \overset{d}{\sim} F_{\boldsymbol{\theta}}, \\ \Psi_i &= \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{i-j}^2 + \sum_{j=1}^{q} \beta_j \Psi_{i-j} \end{cases} \tag{12}$$

for some $(\boldsymbol{\phi}, \boldsymbol{\theta})$ where $\boldsymbol{\phi}^\top = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$. Let $F_{\boldsymbol{\theta}}$ be the skew$-t(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (v, \lambda)^\top$, $v \geq 5$. The details provided below would remain largely unchanged for many other distributions. Let $\{U_i\}_{i\in\mathbb{Z}}$ be iid uniform(0,1) random variables, and $\varepsilon_i(\boldsymbol{\theta}) = F_{\boldsymbol{\theta}}^{-1}(U_i)$, $(\boldsymbol{\theta} \in \Theta; i \in \mathbb{Z})$. Without loss of generality, assume that $\varepsilon_i = \varepsilon_i(\boldsymbol{\theta}_0)$. A typical assumption made in empirical studies involving GARCH(p,q) is that $\{1 - \sum_j \beta_{j0} - \sum_j \alpha_{j0}\} > 0$ and $E\{|\varepsilon_1|^4\} < \infty$ (Francq and Zakoïan 2010).

Let $Z_i = \sup_{\bar{\boldsymbol{\theta}} \in K_{\boldsymbol{\theta}}} |\varepsilon_i(\bar{\boldsymbol{\theta}})|$, where $K_{\boldsymbol{\theta}}$ denotes a closed ball in Θ containing the true value $\boldsymbol{\theta}_0$ as an interior point. If $\{1 - \sum_j \beta_{j0} - \sum_j \alpha_{j0}\} > 0$ and $E(|Z_i|^{4+\delta}) < \infty$ for some $\delta > 0$ and $K_{\boldsymbol{\theta}}$, then the conditions of Theorem 2 for the consistency of the bootstrap test are also satisfied (see the Supplement). Therefore, we conclude that the bootstrap test (8) for testing the specification of model $\mathcal{M}$ in (12) is consistent. The additional requirement $E(|Z_1|^{4+\delta}) < \infty$, for some $\delta > 0$, is only slightly stronger than the condition $E\{|\varepsilon_1|^4\} < \infty$.

In the method of verification presented in the Supplement, we represent both the model under the null hypothesis and the operational bootstrap processes simultaneously in a system of *stochastic recurrence equations* [SRE]. These equations are defined such that the observable random quantities, such as X_i and X_i^* $(i = 1, \ldots, n)$, are generated simultaneously by the same SRE. Then, we invoke results from the literature on SRE.

The regularity conditions in the main theorems of this paper on the consistency of the bootstrap are weak enough to ensure that the results are applicable to a broad class of models that are likely to be used in empirical studies. Some of the steps in the verifications are in fact generalizations of the steps to show that the Gaussian QMLE is asymptotically normal, using the SRE approach. These observations lead us to conjecture that, for a given specific model of the form $\mathcal{M}$, if the asymptotic normality of the Gaussian QMLE $\hat{\boldsymbol{\phi}}$ can be established using the SRE approach (for example, as in Straumann and Mikosch 2006), it is likely that the conditions (M1)–(M4) are also satisfied and they can be verified using the method illustrated in the Supplement for GARCH(p,q). Conversely, if the asymptotic normality of $\hat{\boldsymbol{\phi}}$ cannot be established using the SRE approach then it is likely to be difficult to verify (M1)–(M4) using the method illustrated in the Supplement.

7 Simulation Study

This section presents the results of a simulation study that evaluates the finite sample performance of the proposed tests and comparisons with the information matrix test of White (1982). For the null model, we consider (a) GARCH(1,1) and (b) AR(1)-GARCH(1,1), each combined with a range of choices for $F_{\boldsymbol{\theta}}$. We include the test statistics, KS, CvM, K_u, A^2, and U^2 defined in Section 2.

Since our tests are developed for a broad class of null models of the form $\mathcal{M}$, it is desirable to include some special cases of model $\mathcal{M}$ and tests that are specifically developed for the special cases. To this end, we consider the test proposed by Koul and Mimoto (2012), denoted KM hereafter, which is developed for GARCH(p,q) with the error distribution being free of unknown parameters. In our simulation study, the KM test turns out to have large type-I error rates. Therefore, we also consider a non-operational size corrected version of KM, denoted KMe. This test uses the true value of $\boldsymbol{\omega}$ in the computations instead of the estimator $\hat{\boldsymbol{\omega}}$, and computes the critical value by simulation. Thus, KM and KMe are neither operational nor competitors to our tests, but they provide useful information to assess the performance of our tests against some benchmark. With this in mind, we consider several cases for which the KM and KMe tests are applicable, e.g., the GARCH(1,1) model with standard normal error distribution. For the null error distribution of standard normal, we also include the Jarque-Bera test for normality (cf. Bera and Jarque, 1981).

Empirical power of the tests are evaluated against various misspecifications of the conditional mean, conditional variance, error distribution, and violations of the iid error assumption. We consider sample sizes ranging from 500 to 15000, with 2000 Monte Carlo replications, and adopt the "Warp-Speed" Monte Carlo method of Giacomini *et al.* (2013).

The choices for the error distribution

F1: Standard normal [*N*]; pdf: $f_N(x) = (2\pi)^{-1/2}\exp(-x^2/2)$.

F2: Laplace [*L*]; pdf: $f_L(x) = 2^{-1/2}\exp(-\sqrt{2}|x|)$.

F3: Standardized t with d degrees of freedom [Std $t(d)$].

F4: Generalized error distribution [GED]; pdf: $f_{GE}(x,\boldsymbol{\theta}) = \{2\Gamma(\boldsymbol{\theta}^{-1})\}^{-1}\boldsymbol{\theta}C_{\boldsymbol{\theta}}\exp(-|C_{\boldsymbol{\theta}}x|^{\boldsymbol{\theta}})$, where Γ is the gamma function, $C_{\boldsymbol{\theta}} = \{\Gamma(3\boldsymbol{\theta}^{-1})/\Gamma(\boldsymbol{\theta}^{-1})\}^{1/2}$, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^+$.
This family encompass several distributions, for example Laplace ($\boldsymbol{\theta} = 1$) and normal ($\boldsymbol{\theta} = 2$).

F5: Standardized extreme value distribution [SEVD]; pdf: $f_{EV}(x) = \beta^{-1}\exp\{z - \exp(z)\}$, where $z = (x-m)/b$, $b = \sqrt{6}/\pi$, $m = b\gamma_e$, and γ_e is the Euler-Mascheroni constant.

F6: Skewed-$t(v,\lambda)$ distribution, where v and λ are the degrees of freedom and skewness parameters respectively (cf. Hansen 1994).

F7: Normal-Laplace mixture [NL(ρ)]; pdf: $f_{NL}(x) = (1-\rho)f_N(x) + \rho f_L(x)$.

The choices for the data generating process

M1: G-F [GARCH(1,1) model]: $X_i = \{\Psi_i\}^{1/2}\varepsilon_i,\ \Psi_i = 0.1 + 0.2X_{i-1}^2 + 0.7\Psi_{i-1},\ \varepsilon_i \overset{iid}{\sim} F$.

[In the abbreviation G-F, F denotes the error distribution, and it may be any of the aforementioned 7 distributions; for example, G-N denotes GARCH(1,1) with standard normal error distribution.]

For the DGPs *M2* – *M4* below, we used the GARCH(1,1) model, $\Psi_i = 0.1 + 0.2e_{i-1}^2 + 0.7\Psi_{i-1}$.

M2: ARG-F [AR(1)-GARCH(1,1)]: $X_i = 0.1 + 0.2X_{i-1} + e_i, \quad e_i = \sqrt{\Psi_i}\varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} F$.

M3: ARG$^{(a)}$-F [AR(1)-GARCH(1,1) with non-iid errors]: $X_i = 0.1 + 0.2X_{i-1} + e_i$,
$e_i = \sqrt{\Psi_i}\varepsilon_i,\ \varepsilon_i = (c_0\varepsilon_{i-1} + c_1)a_i, \quad a_i \overset{iid}{\sim} F, \quad c_0 = 0.3, \quad c_1 = (1-c_0^2)^{1/2}$.

M4: TARG-F [Threshold AR(1)-GARCH(1,1)]:

$$X_i := \begin{cases} 0.5 + 0.8X_{i-1} + e_i, & X_{i-1} \le 0.8, \\ 0.05 + 0.06X_{i-1} + e_i, & X_{i-1} > 0.8, \end{cases} \quad e_i = \sqrt{\Psi_i}\varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} F.$$

M5: EGARCH-F [EGARCH(1,1) model]: $X_i = \{\Psi_i\}^{1/2}\varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} F$,
$\ln\Psi_i = 0.1 + 0.5\ln\Psi_{i-1} + 0.3(|\varepsilon_{i-1}| - E|\varepsilon_{i-1}|) - 0.8\varepsilon_{i-1}$.

M6: BIL-F [Bilinear model]: $X_i = 0.1 + 0.6X_{i-1} + 0.7\varepsilon_{i-1}X_{i-2} + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} F$.

M7: NLMA-F [Nonlinear moving average model]: $X_i = 0.8\varepsilon_{i-1}^2 + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} F$.

M8: AR2G22-F [AR(2)-GARCH(2,2)]: $X_i = 0.1 + 0.3X_{i-1} + 0.4X_{i-2} + e_i, \quad e_i = \sqrt{\Psi_i}\varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} F$,
$\Psi_i = 0.1 + 0.1e_{i-1}^2 + 0.2e_{i-2}^2 + 0.2\Psi_{i-1} + 0.4\Psi_{i-2}$.

Design for comparison with the information matrix test of White (1982)

The Information Matrix (IM) test introduced by White (1982) is a well-known general purpose specification test for parametric models. Under White's original formulation of the IM test, which was based on iid observations, the asymptotic null distribution of the IM-test statistic was shown to be χ^2. Its asymptotic null distribution has not been derived for the general setting of this paper. Implementation of the information matrix test in our setting is challenging because it relies on the third order derivatives of the log likelihood function for model $\mathcal{M}$ that involves stochastic recurrence equations. Nevertheless, it appears that one may still estimate the finite sample null distribution of the IM-test statistic by bootstrap; see Huo and Cho (2021) and Cho and White (2014).

We compare the IM test with our tests for two sequences of local hypotheses. Both sequences use AR(1)-GARCH(1,1) for the mean-variance functions. For the error distributions, we use the following mixture distributions with $f_{SkT}(x, v, \lambda)$ denoting the pdf of Skewed-$t(v, \lambda)$:

$$\text{SkewTL}(\delta),\ \text{pdf}:\ f_{SkTL}(x) = (1-\delta)f_{SkT}(x, 7, -0.1) + \delta f_L(x), \tag{13}$$

$$\text{GESkewT}(\delta),\ \text{pdf}:\ f_{GESkT}(x) = (1-\delta)f_{GE}(x, 1.5) + \delta f_{SkT}(x, 7, -0.1), \tag{14}$$

with $n = 1000, 2000, 3000$ and $\delta = 0, 0.2, 0.4, 0.6, 0.8, 1$. For each (δ, n), and the error distribution (13), we test the null hypothesis that the data were generated from an AR(1)-GARCH(1,1) model with Skewed-$t(v, \lambda)$ distribution for some (v, λ). Similarly, for the DGP with error distribution (14), we test the null hypothesis that the data are generated from an AR(1)-GARCH(1,1) model with GED($\boldsymbol{\theta}$) distribution. To obtain the p-values of the IM-test statistic, we applied the parametric bootstrap method in Section 4 of Huo and Cho (2021).

Results of the simulation study

A summary of the main results are presented in Table 1 and Figure 1. Additional simulation results are presented in Section S4 in the Supplement. The results and discussions for comparing our tests with KMe and Jarque-Bera [JB] tests are also provided in the Supplement; recall that these tests are not competitors to our tests but were included to provide some benchmarks that are different from the IM test. The effect of model *overspecification* under the null hypothesis (i.e when the true DGP is strictly nested in the null model) is also evaluated in Section S4 in the supplement. The main observations are the following:

(a) The tests proposed in this paper, namely KS, Ku, CvM, U^2, and A^2, performed consistently well throughout in terms of empirical type-I error rates. We also observed that our tests performed well in the over-specified settings, in the sense that the type I error rates of our tests were close to the nominal level (see Figures S7 and S8 in the Supplement).

(b) Our tests outperformed the information matrix test.

(c) The tests proposed in this paper exhibited non-trivial empirical power (i.e. power higher than the nominal level of the test) properties against a range of alternative models. These include only the error distribution is different from that in the null model, only the mean function is different from that in the null model, the conditional variance is different from that in the null model, and the models for $\{\mu_i, \Psi_i\}$ are the same as for the null model but the errors are not iid.

(d) For testing the specification of a GARCH model with a known distribution function, for example the standard normal, the KMe test performed better than the tests proposed in this paper for small samples. This was expected since KMe used the true population parameter value, instead of the estimator, thus avoiding an important source of variation. Our tests exhibited increasingly better performance as the sample size increased and outperformed the KMe in large samples.

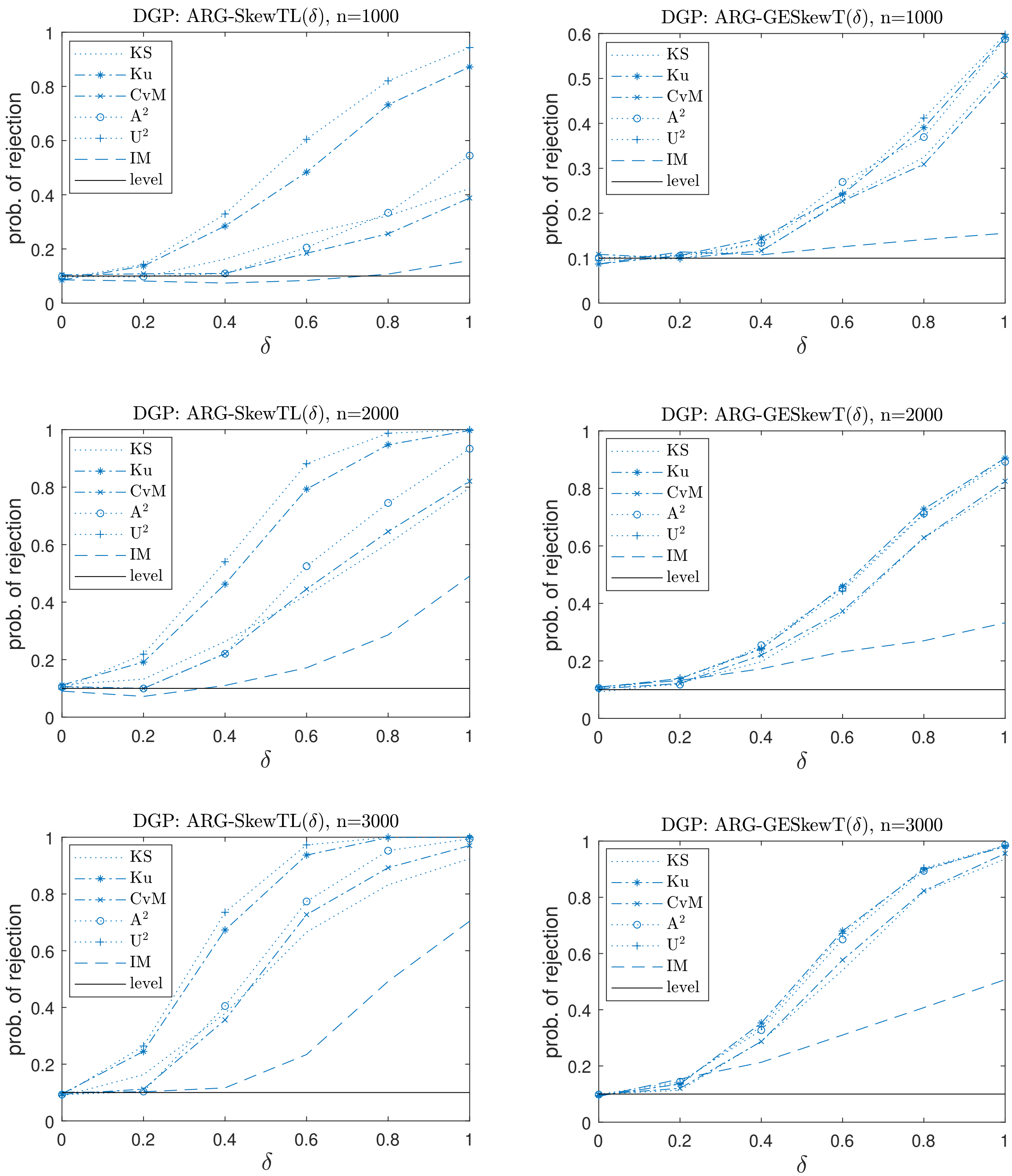
(e) In terms of empirical power, the Anderson–Darling test (A^2), the Kuiper's test (Ku) and the Watson's test (U^2) performed marginally better than the KS and the CvM tests.

Table 1: Empirical type I error rates and power (%) at 5% nominal level for testing 'H_0: AR(1)-GARCH(1,1) with error distribution $F_{\boldsymbol{\theta}}$'; the sample size $n = 1000$.

DGP	$F_{\boldsymbol{\theta}}$ (in H_0)	KS	Ku	CvM	A^2	U^2
		Type-I error rate(%)				
ARG-N	N	5.0	4.2	4.9	4.6	4.4
ARG-L	L	4.9	5.3	4.9	4.7	5.2
ARG-SEVD	SEVD	4.8	5.2	5.4	4.7	5.5
ARG-Std $t(5)$	Std $t(5)$	4.6	4.9	4.1	3.9	3.7
ARG-N	GED($\boldsymbol{\theta}$)	5.3	4.4	4.1	3.9	4.4
ARG-L	GED($\boldsymbol{\theta}$)	5.1	5.5	4.6	5.0	4.9
ARG-GED(1.5)	GED($\boldsymbol{\theta}$)	5.0	5.9	5.5	4.5	5.7
ARG-Skewed-$t(7, -0.1)$	Skewed-$t(v, \lambda)$	5.1	4.6	4.5	5.0	6.1
ARG-Std $t(5)$	Std $t(d)$	4.8	6.4	4.7	4.1	5.7
		Empirical power (%)				
EGARCH-N	N	99.9	99.9	99.9	99.9	99.9
TARG-SEVD	SEVD	99.3	99.2	99.8	99.6	99.3
EGARCH-N	GED($\boldsymbol{\theta}$)	99.9	99.9	99.9	99.9	99.7
NLMA-L	GED($\boldsymbol{\theta}$)	25.8	34.7	38.5	76.9	28.1
BIL-Skewed-$t(7, -0.1)$	Skewed-$t(v, \lambda)$	34.3	15.6	45.6	47.3	17.5
TARG-Std $t(5)$	Std $t(d)$	51.2	32.1	59.1	61.5	37.9
AR2G22-GED(1.5)	N	41.7	60.6	50.3	61.3	67.0
ARG$^{(a)}$-N	N	98.5	91.9	99.4	99.8	94.9
ARG$^{(a)}$-L	L	70.1	51.5	76.8	79.5	54.7
ARG$^{(a)}$-SEVD	SEVD	68.6	41.5	71.1	94.5	45.8
ARG$^{(a)}$-Std $t(5)$	Std $t(5)$	77.7	52.6	81.8	82.2	54.7
ARG$^{(a)}$-N	GED($\boldsymbol{\theta}$)	81.9	42.2	96.2	97.6	41.3
ARG$^{(a)}$-L	GED($\boldsymbol{\theta}$)	85.3	68.7	88.5	91.0	72.0
ARG$^{(a)}$-GED(1.5)	GED($\boldsymbol{\theta}$)	77.1	45.7	90.2	93.9	44.5
ARG$^{(a)}$-Skewed-$t(7, -0.1)$	Skewed-$t(v, \lambda)$	90.6	74.7	93.1	92.8	78.6
ARG$^{(a)}$-Std $t(5)$	Std $t(d)$	62.0	49.3	64.8	67.4	45.5

Note: (1) $F_{\boldsymbol{\theta}}$ is the parametric family of error distributions under H_0, N: Normal; L: Laplace; Std t: Standardized t; GED: Generalized error distribution; SEVD: Standardized extreme value distribution; Skewed-t: Skewed t. (2) DGPs are defined by models *M2–M8*.

Figure 1: Empirical power curves at the 10% level. For panels on the left, the null hypothesis is 'H_0: AR(1)-GARCH(1,1) with the error distribution being Skewed-$t(v, \lambda)$' and the DGP, denoted ARG-SkewTL(δ), is AR(1)-GARCH(1,1) with Skewed-t and Laplace mixture for which the pdf is $f_{SkTL}(x) = (1-\delta) f_{SkT}(x) + \delta f_L(x)$. For panels on the right, the null hypothesis is 'H_0: AR(1)-GARCH(1,1) model with the error distribution being $GED(\boldsymbol{\theta})$', and the DGP, denoted ARG-GESkewT(δ), is AR(1)-GARCH(1,1) with generalized-error and Skewed-t mixture distribution for which the pdf is $f_{GESkT}(x) = (1-\delta) f_{GE}(x) + \delta f_{SkT}(x)$.



8 An empirical illustration

To illustrate the testing procedure, we briefly discuss an example based on the energy price growth rate data studied by Bai and Lam (2019) and by Huo and Cho (2021). The variable of interest is the growth rate of the weekly energy price. More precisely, we consider the growth rates of the Baltic liquefied petroleum gas (BLPG) price, propane Argus Far East index (PAFEI), and propane CP swap price (CPS). The data were obtained from the energy price data provided by Bai and Lam (2019). Each of these samples contains 601 observations and spans the period from the second week of 2005 to the 35th week of 2016. Huo and Cho (2021) analyzed this data set by applying several diagnostic tests, and concluded that an AR-GARCH(1,1) model with a skewed t-distribution provides a good fit; see Table 4 in Huo and Cho (2021).

Instead of using the AR-GARCH(1,1), we apply the AR-AGARCH(1,1) model, because the stylized facts of the energy price growth rates are known to display the so-called leverage effects. More precisely, we test the goodness-of-fit of the model

$$X_i = \mu_i + \sqrt{\Psi_i}\varepsilon_i, \quad \mu_i = \phi_1 + \phi_2 X_{i-1}, \quad \Psi_i = \phi_3 + \phi_4 \big\{|\eta_{i-1}| - \gamma\eta_{i-1}\big\}^2 + \phi_5 \Psi_{i-1}, \quad \eta_i = X_i - \mu_i, \tag{15}$$

with skew-$t(v, \lambda)$ error distribution, where v and λ are the degrees of freedom and the skewness parameters, respectively (Hansen 1994). To this end, we apply the tests based on $T_1, \ldots, T_5$ defined in Section 2 together with the bootstrap information matrix [IM] test considered in Section 7. The bootstrap p-values for the tests are provided in Table 2.

In view of the results in Table 2, the AR(1)-AGARCH(1,1) model with the skew-$t(v, \lambda)$ error distribution cannot be rejected for both BLPG and CPS. It does not appear that this model is adequate for the variable PAFEI. Although the IM test provides a relatively large p-value of 0.162 for PAFEI, in view of the low empirical power of the IM test in the simulation results of the previous section, it is prudent not to rely solely on the IM test.

Table 2: The bootstrap p-values for the specification tests of AR(1)-AGARCH(1,1) with error distribution skew-$t(v, \lambda)$.

Variable	KS	Ku	CvM	A^2	U^2	IM
BLPG	0.494	0.268	0.454	0.200	0.182	0.205
PAFEI	0.031	0.020	0.069	0.047	0.025	0.162
CPS	0.236	0.281	0.265	0.435	0.404	0.487

SUPPLEMENTARY MATERIAL

The online Supplement to this paper provides the verification of the regularity conditions for GARCH(p,q). It also contains a general result on asymptotic uniform expansions for a class of weighted empirical processes. In addition, some simulation results are also provided.

ACKNOWLEDGEMENTS

We wish to express our warm appreciation to three anonymous reviewers, the associate editor, and the editor for their comments, suggestions, and constructive criticisms. We thank Peter Phillips and Peter Robinson for some helpful suggestions, and Jin Seo Cho for providing us the R-code for implementing the bootstrap version of the IM test. Silvapulle acknowledges support from the Australian Government through the Australian Research Council Discovery Project scheme. Indeewara Perera gratefully acknowledges support from the Department of Economics, University of Sheffield, for visiting Monash University in 2018/19 to complete part of this research. We also thank the Monash e-Science and Grid Engineering Lab for assistance with computing.

A APPENDIX: Main steps for the proofs of Theorems 1, 2, and 3

Lemma 1. *Suppose that* $Y_i \overset{e.a.s.}{\to} 0$ *as* $i \to \infty$. *Then there exist a nonnegative random variable* Z *and* $\gamma > 1$ *such that* $\|Y_i\| \le \gamma^{-i} Z$ $(i \in \mathbb{N})$. *Further,* $\max_{1\le i\le n} \|Y_i\| = O_P(1)$ *and* $\sum_{i=1}^n \|Y_i\| = O_p(1)$.

Proof. Let Ω_0 be a set of probability 1 such that, for some $\gamma > 1$, we have $\gamma^i \|Y_i(\omega)\| \to 0$ as $i \to \infty$ for every $\omega \in \Omega_0$. Let $\epsilon > 0$ and $\omega \in \Omega_0$ be given. Then, there exists $n_0(\epsilon, \omega)$ such that $\gamma^i \|Y_i(\omega)\| < \epsilon$ for all $i \ge n_0(\epsilon, \omega)$ and $\omega \in \Omega_0$. Let $Z(\omega) = \epsilon + \max_{1\le i\le n_0(\omega,\epsilon)} \gamma^i \|Y_i(\omega)\|$ $(\omega \in \Omega_0)$. Then, $\|Y_i(\omega)\| \le \gamma^{-i} Z(\omega)$ for $\omega \in \Omega_0$ and $i \in \mathbb{N}$. The other two parts follow easily. □

A.1 Proof of Theorem 1

Recall that $\{\varepsilon_i, i \in \mathbb{Z}\}$ are iid and $\widetilde{\varepsilon}_i = \widetilde{\eta}_i(\hat{\boldsymbol{\phi}})/\{\widetilde{\Psi}_i(\hat{\boldsymbol{\phi}})\}^{1/2}$ $(i = 1, \dots, n)$. Also recall the definitions of $F_n, W_n, \widetilde{F}_n$, and $\widetilde{W}_n$ from (6) and (7). In what follows, we write '$\sum$' for '$\sum_{i=1}^n$'. Let us introduce the following terms corresponding to $\{\widetilde{\varepsilon}_i, \widetilde{F}_n, \widetilde{W}_n\}$ by replacing $\{\widetilde{\eta}_i(\cdot), \widetilde{\Psi}_i(\cdot)\}$ with $\{\eta_i(\cdot), \Psi_i(\cdot)\}$:

$$\widehat{\varepsilon}_i = \eta_i(\hat{\boldsymbol{\phi}})/\{\Psi_i(\hat{\boldsymbol{\phi}})\}^{1/2}, \quad \widehat{F}_n(x) := n^{-1} \sum I(\widehat{\varepsilon}_i \le x), \quad \widehat{W}_n(x) := \sqrt{n}\{\widehat{F}_n(x) - F_{\hat{\boldsymbol{\theta}}}(x)\}. \qquad \text{(A.1)}$$

To prove Theorem 1, we show that (a) $\sup_x |\widetilde{W}_n(x) - \widehat{W}_n(x)| = o_p(1)$, and (b) $\widehat{W}_n(\cdot)$ converges weakly to the centered Gaussian process $W_0(\cdot)$ in Theorem 1. Then, the weak convergence of $\widetilde{W}_n(\cdot)$

and that of each of the test statistics would follow. Let

$$\mathcal{V}_i = \{\Psi_i^{1/2}(\hat{\phi}) - \Psi_i^{1/2}(\phi_0)\}\Psi_i^{-1/2}(\phi_0), \quad \mathcal{U}_i = \{\mu_i(\hat{\phi}) - \mu_i(\phi_0)\}\Psi_i^{-1/2}(\phi_0), \tag{A.2}$$

$$\widetilde{\mathcal{V}}_i = \{\widetilde{\Psi}_i^{1/2}(\hat{\phi}) - \Psi_i^{1/2}(\phi_0)\}\Psi_i^{-1/2}(\phi_0), \quad \widetilde{\mathcal{U}}_i = \{\widetilde{\mu}_i(\hat{\phi}) - \mu_i(\phi_0)\}\Psi_i^{-1/2}(\phi_0). \tag{A.3}$$

In order to show that $\sup_x |\widetilde{W}_n(x) - \widehat{W}_n(x)| = o_p(1)$ we use Theorem A1 of Escanciano (2010). To this end, in the next lemma, we first establish that four regularity conditions are satisfied.

Lemma 2. *Under Assumptions 1–7, the following four conditions are satisfied.*

(L1). *Let* $\mathcal{Z}_{b,c} := \{(z_1, z_2)^\top \in \mathbb{R}^2 : |z_1| + |z_2| \le b, z_1 \ge c-1\}$. *Then, for all* $0 < b < \infty$ *and* $0 < c < \infty$, $\sup_{x\in\mathbb{R},\,(z_1,z_2)^\top\in\mathcal{Z}_{b,c}} |x| f_{\boldsymbol{\theta}_0}(x + xz_1 + z_2) < \infty$.

(L2). *Let* $K > 0$, $0 < \delta < 1$ *be arbitrary but fixed. Let* $\Delta_K = \{\boldsymbol{s} \in \mathbb{R}^p : \|\boldsymbol{s}\| < K\}$. *Then, for every* $0 < c < 1$ *there exists* $n_0 \in \mathbb{N}$, *such that,* $P(\inf_{\boldsymbol{s}\in\Delta_K} \min_{1\le i\le n}\{\Psi_i(\phi_0 + n^{-1/2}\boldsymbol{s})/\Psi_i(\phi_0)\} > c^2) \ge 1-\delta$ *for all* $n \ge n_0$. *Further, there exists a constant* $c_0 > 0$ *such that, for all* n *sufficiently large,* $P(\inf_{\boldsymbol{s}\in\Delta_K} \min_{1\le i\le n}\{\widetilde{\Psi}_i(\phi_0 + n^{-1/2}\boldsymbol{s})/\Psi_i(\phi_0)\} > c_0^2) \ge 1-\delta$.

(L3). $\max_{1\le i\le n}\{|\widetilde{\mathcal{V}}_i| + |\widetilde{\mathcal{U}}_i|\} = O_p(1)$, $\max_{1\le i\le n}\{|\mathcal{V}_i| + |\mathcal{U}_i|\} = o_p(1)$.

(L4). $\forall\delta > 0, \exists c > 0$ *such that* $P(\min_{1\le i\le n} \widetilde{\mathcal{V}}_i > c-1) \ge 1-\delta$, *for all* n *sufficiently large.*

Proof. Proof of (L1) : This follows from the facts that $\sup_x f_{\boldsymbol{\theta}_0}(x)$ and $\sup_x |x f_{\boldsymbol{\theta}_0}(x)|$ are finite.
Proof of (L2) : To prove the first part of (L2), let $\xi_{i,n,s} = \{\Psi_i(\phi_0 + n^{-1/2}\boldsymbol{s})\}^{1/2}\{\Psi_i(\phi_0)\}^{-1/2}$. Write

$$\begin{aligned}\xi_{i,n,s} = \quad & 1 + \frac{\{\Psi_i(\phi_0 + n^{-1/2}\boldsymbol{s})\}^{1/2} - \{\Psi_i(\phi_0)\}^{1/2} - (n^{-1/2}\boldsymbol{s}/2)^\top\dot{\Psi}_i(\phi_0)\{\Psi_i(\phi_0)\}^{-1/2}}{\{\Psi_i(\phi_0)\}^{1/2}} \\ & + 2^{-1}(n^{-1/2}\boldsymbol{s})^\top\dot{\Psi}_i(\phi_0)/\Psi_i(\phi_0), \qquad (\boldsymbol{s} \in \Delta_K),\end{aligned}$$

by adding and subtracting the same terms. By Assumptions 4.2 and 5, $\xi_{i,n,s}$ converges to one, in probability, uniformly over $\boldsymbol{s} \in \Delta_K$ and $1 \le i \le n$. Hence, the first part of (L2) follows.

To prove the second part, write $d_{i,n,s} := \{\sqrt{\widetilde{\Psi}_i(\phi_0 + n^{-1/2}\boldsymbol{s})} - \sqrt{\Psi_i(\phi_0 + n^{-1/2}\boldsymbol{s})}\}/\sqrt{\Psi_i(\phi_0)}$, and $\widetilde{\xi}_{i,n,s} = \sqrt{\widetilde{\Psi}_i(\phi_0 + n^{-1/2}\boldsymbol{s})}/\sqrt{\Psi_i(\phi_0)}$. Then $\widetilde{\xi}_{i,n,s} = d_{i,n,s} + \xi_{i,n,s}$, $(\boldsymbol{s} \in \Delta_K)$, and

$$\inf_{\boldsymbol{s}\in\Delta_K} \min_{n^*\le i\le n} \widetilde{\xi}_{i,n,s} \ge \inf_{\boldsymbol{s}\in\Delta_K} \min_{n^*\le i\le n} d_{i,n,s} + \inf_{\boldsymbol{s}\in\Delta_K} \min_{n^*\le i\le n} \xi_{i,n,s},$$

for any given pair $\{n^*, n\}$ with $1 < n^* \le n$. Since $\Psi_i(\phi_0) > \alpha_L > 0$, by Assumption 2 and Lemma 2.3 of Straumann and Mikosch (2006), $\sup_{\boldsymbol{s}\in\Delta_K,\,n\in\mathbb{N}} |d_{i,n,s}| \overset{e.a.s.}{\to} 0$ as $i \to \infty$. Therefore, the second part of (L2) follows.

Proof of (L3) : Since $n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0) = O_p(1)$, by using Assumptions 3, 4, 5, we obtain $\max_{1\le i\le n}\left|\mathcal{V}_i\right| = o_p(1)$ and $\max_{1\le i\le n}|\mathcal{U}_i| = o_p(1)$. Since $\Psi_i(\boldsymbol{\phi}_0) > \alpha_L > 0$, for each $i \ge 1$, we have that

$$|\widetilde{\mathcal{V}}_i| \;\le\; \alpha_L^{-1/2}\left|\sqrt{\widetilde{\Psi}_i(\hat{\boldsymbol{\phi}})} - \sqrt{\Psi_i(\hat{\boldsymbol{\phi}})}\right| + |\mathcal{V}_i|. \tag{A.4}$$

From Assumption 2, $\sup_{\boldsymbol{\phi}\in\Phi}|\widetilde{\Psi}_i(\boldsymbol{\phi}) - \Psi_i(\hat{\boldsymbol{\phi}})| \overset{e.a.s.}{\to} 0$, and hence by using Lemmas 2.1 and 2.3 in Straumann and Mikosch (2006), we obtain that $\max_{1\le i\le n}\left|\sqrt{\widetilde{\Psi}_i(\hat{\boldsymbol{\phi}})} - \sqrt{\Psi_i(\hat{\boldsymbol{\phi}})}\right| = O_p(1)$. Since $\max_{1\le i\le n}\left|\mathcal{V}_i\right| = o_p(1)$ then it follows from (A.4) that $\max_{1\le i\le n}\{|\widetilde{\mathcal{V}}_i|\} = O_p(1)$. By using similar arguments, we also obtain that $\max_{1\le i\le n}|\widetilde{\mathcal{U}}_i| = O_p(1)$.

Proof of (L4) : Fix $0 < \delta < 1$. Since $n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0) = O_p(1)$, there exists a constant $K > 0$ such that $P(n^{1/2}|\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0| > K) < \delta/2$ for all n sufficiently large. Further, in view of Condition (L2), there exists a constant $c > 0$ such that $P(\inf_{\boldsymbol{s}\in\Delta_K}\min_{1\le i\le n}\{\sqrt{\widetilde{\Psi}_i(\boldsymbol{\phi}_0 + n^{-1/2}\boldsymbol{s})}/\sqrt{\Psi_i(\boldsymbol{\phi}_0)}\} \le c) < \delta/2$. Let $A_{n,K} = \{n^{1/2}|\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0| > K\}$. Then, for all n sufficiently large, $P\big(\min_{1\le i\le n}\widetilde{\mathcal{V}}_i \le c-1\big) \le P\big(\inf_{\boldsymbol{s}\in\Delta_K}\min_{1\le i\le n}\big\{\sqrt{\widetilde{\Psi}_i(\boldsymbol{\phi}_0 + n^{-1/2}\boldsymbol{s})}/\sqrt{\Psi_i(\boldsymbol{\phi}_0)}\big\} \le c\big) + P(A_{n,K}) \le \delta$. □

Proposition 2. *Under Assumptions 1–7, we have that* $\sup_{x\in\mathbb{R}}|\widetilde{W}_n(x) - \widehat{W}_n(x)| = o_p(1)$.

Proof. Let $\widetilde{K}_n(x) := n^{-1/2}\sum\big\{I[\varepsilon_i \le \widetilde{\varrho}_i(x)] - F_{\boldsymbol{\theta}_0}[\widetilde{\varrho}_i(x)]\big\}$, $x \in \mathbb{R}$, where $\widetilde{\varrho}_i(x) = x + x\widetilde{\mathcal{V}}_i + \widetilde{\mathcal{U}}_i$. Then $\widetilde{K}_n(x) - W_n(x)$ forms a sum of conditionally centered bounded random variables ($x \in \mathbb{R}$), and hence by (L1)–(L4) in Lemma 2, we have $\mathrm{Var}(\widetilde{K}_n(x) - W_n(x)) \le E\big[n^{-1}\sum|F_{\boldsymbol{\theta}_0}[\widetilde{\varrho}_i(x)] - F_{\boldsymbol{\theta}_0}(x)|\big] = o_p(1)$. By Theorem A1 of Escanciano (2010) and Lemma 2, we also obtain that the process $\widetilde{K}_n - W_n$ is asymptotically equicontinuous, and hence from the asymptotic tightness of $\widetilde{K}_n - W_n$, it follows that $\sup_{x\in\mathbb{R}}|\widetilde{K}_n(x) - W_n(x)| = o_p(1)$.

Remark 1. *To show the tightness of* $(\widetilde{K}_n - W_n)$*, the standard techniques are insufficient as explained in Escanciano (2010) (see pages 749, 750, and Appendix).*

Next, let $K_n(x) := n^{-1/2}\sum\big\{I[\varepsilon_i \le \varrho_i(x)] - F_{\boldsymbol{\theta}_0}[\varrho_i(x)]\big\}$, $x \in \mathbb{R}$, where $\varrho_i(x) = x + x\mathcal{V}_i + \mathcal{U}_i$. Since $\max_{1\le i\le n}|\mathcal{V}_i| = o_p(1)$ and $\max_{1\le i\le n}|\mathcal{U}_i| = o_p(1)$, by using Theorem S1 in the Supplementary Material we obtain that $\sup_{x\in\mathbb{R}}|K_n(x) - W_n(x)| = o_p(1)$. Hence, $\sup_{x\in\mathbb{R}}|\widetilde{K}_n(x) - K_n(x)| = o_p(1)$. Since $\widetilde{W}_n(x) - \widehat{W}_n(x) = n^{-1/2}\sum I[\varepsilon_i \le \widetilde{\varrho}_i(x)] - n^{-1/2}\sum I[\varepsilon_i \le \varrho_i(x)]$, it follows that

$$\sup_{x\in\mathbb{R}}\left|\widetilde{W}_n(x) - \widehat{W}_n(x) - n^{-1/2}\sum\big\{F_{\boldsymbol{\theta}_0}[\widetilde{\varrho}_i(x)] - F_{\boldsymbol{\theta}_0}[\varrho_i(x)]\big\}\right| = \sup_{x\in\mathbb{R}}|\widetilde{K}_n(x) - K_n(x)| = o_p(1). \tag{A.5}$$

Next, fix $\delta > 0$. From (L1)–(L4), there exist constants $K > 0$ and $c > 0$ such that the event $B^n_{K,c} := \{(\widetilde{\mathcal{V}}_i, \widetilde{\mathcal{U}}_i)^\top \in \mathcal{Z}_{K,c}, (\mathcal{V}_i, \mathcal{U}_i)^\top \in \mathcal{Z}_{K,c}, 1 \le i \le n\}$, satisfies $P(B^n_{K,c}) > 1-\delta$ for all n sufficiently large, and on the event $B^n_{K,c}$, we have

$$\sup_{x\in\mathbb{R}}\left|\sum F_{\boldsymbol{\theta}_0}[\widetilde{\varrho}_i(x)] - F_{\boldsymbol{\theta}_0}[\widetilde{\varrho}_i(x)]\right| \le A_1 n^{-1/2}\sum\left|\widetilde{\mathcal{V}}_i - \mathcal{V}_i\right| + \sup_{x\in\mathbb{R}} f_{\boldsymbol{\theta}_0}(x) n^{-1/2}\sum\left|\widetilde{\mathcal{U}}_i - \mathcal{U}_i\right|, \tag{A.6}$$

where $A_1 = \sup_{x\in\mathbb{R},\,(z_1,z_2)^\top\in\mathcal{Z}_{K,\,c}} |x| f_{\boldsymbol{\theta}_0}(x+xz_1+z_2)$. From Condition (L1), $A_1 < \infty$, and $\sup_{x\in\mathbb{R}} f_{\boldsymbol{\theta}_0}(x)$ is finite by Assumption 7.4. Since $\Psi_i(\boldsymbol{\phi}_0) > \alpha_L > 0$, $|\{\widetilde{\Psi}_i(\hat{\boldsymbol{\phi}})\}^{1/2} - \{\Psi_i(\hat{\boldsymbol{\phi}})\}^{1/2}| \overset{e.a.s.}{\to} 0$, and $|\widetilde{\mu}_i(\hat{\boldsymbol{\phi}}) - \mu_i(\hat{\boldsymbol{\phi}})| \overset{e.a.s.}{\to} 0$ as $i \to \infty$ (see Assumptions 1.3 and 2), we also obtain that $|\widetilde{\mathcal{V}}_i - \mathcal{V}_i| \overset{e.a.s.}{\to} 0$ and $|\widetilde{\mathcal{U}}_i - \mathcal{U}_i| \overset{e.a.s.}{\to} 0$ as $i \to \infty$. Therefore, an application of Lemma 2.1 of Straumann and Mikosch (2006), or Lemma 1, yields that $\sum\{|\widetilde{\mathcal{V}}_i - \mathcal{V}_i| + |\widetilde{\mathcal{U}}_i - \mathcal{U}_i|\} = O_p(1)$. Since δ is arbitrary, it follows from (A.5) and (A.6) that $\sup_{x\in\mathbb{R}} |\widetilde{W}_n(x) - \widehat{W}_n(x)| = o_p(1)$. □

Recall that Assumption 3.2 provides an asymptotic representation for $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ as a sum of stationary terms up to $o_p(1)$. The next proposition gives a similar representation for $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$.

Proposition 3. *Suppose that Assumptions 1–7 are satisfied. Then,* $\hat{\theta}_j - \theta_{0j} = n^{-1}\sum h_{\boldsymbol{\theta}_0 j}(\varepsilon_i) - (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)^\top \mathcal{Q}_{nj} + o_p(n^{-1/2})$, *where* $\mathcal{Q}_{nj} = n^{-1}\sum[\varphi_i(\boldsymbol{\phi}_0) + \tau_i(\boldsymbol{\phi}_0)\varepsilon_i/2] h'_{\boldsymbol{\theta}_0 j}(\varepsilon_i)$, $(j = 1, 2, \ldots, q)$.

Proof. By the usual one-term Taylor series expansion of the first derivative of the objective function $Q_n(\boldsymbol{\theta})$, we obtain $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{-1/2}\sum h_{\boldsymbol{\theta}_0}(\widetilde{\varepsilon}_i) + o_p(1)$, where, $h_{\boldsymbol{\theta}}(\varepsilon) := -E[\ddot{\hbar}(\boldsymbol{\theta};\varepsilon_0)]^{-1}\dot{\hbar}(\boldsymbol{\theta};\varepsilon)$. Again, by a Taylor series expansion of $h_{\boldsymbol{\theta}_0}(\widetilde{\varepsilon}_i)$ around $\widetilde{\varepsilon}_i = \varepsilon_i$, we obtain $n^{-1/2}\sum h_{\boldsymbol{\theta}_0 j}(\widetilde{\varepsilon}_i) = n^{-1/2}\sum \left[h_{\boldsymbol{\theta}_0 j}(\varepsilon_i) + (\widetilde{\varepsilon}_i - \varepsilon_i)h'_{\boldsymbol{\theta}_0 j}(\varepsilon_i)\right] + o_p(1)$. Then, approximate $\widetilde{\varepsilon}_i$ on the right hand side by $\widehat{\varepsilon}_i$ and obtain that $n^{-1/2}\sum(\widetilde{\varepsilon}_i - \varepsilon_i)h'_{\boldsymbol{\theta}_0 j}(\varepsilon_i) = n^{-1/2}\sum(\widehat{\varepsilon}_i - \varepsilon_i)h'_{\boldsymbol{\theta}_0 j}(\varepsilon_i) + o_p(1)$. Since $\widehat{\varepsilon}_i = d_i(\hat{\boldsymbol{\phi}})$ and $\varepsilon_i = d_i(\boldsymbol{\phi}_0)$, where $d_i(\boldsymbol{\phi}) = \eta_i(\boldsymbol{\phi})/\Psi_i(\boldsymbol{\phi})$, using one-term Taylor expansions and Assumptions 4 and 5, we obtain that $\{\widehat{\varepsilon}_i(\hat{\boldsymbol{\phi}}) - \varepsilon_i(\boldsymbol{\phi}_0)\}$ is asymptotically linear in $(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$. Hence, the proof follows by substituting the expansion for $(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ in Assumption 3.2. □

Proof of Theorem 1. First, we show that $\widehat{W}_n(\cdot) \Rightarrow W_0(\cdot)$. To this end, we show that $\widehat{W}_n(\cdot)$ is tight and its finite dimensional distributions converge. Then, we use Proposition 2 to deduce that $\widetilde{W}_n(\cdot) \Rightarrow W_0(\cdot)$ and use the continuous mapping theorem to conclude that the null distributions of the test statistics converge. Let $B_{1n}(x) = n^{1/2}\{F_n(x) - F_{\boldsymbol{\theta}_0}(x)\}$, $B_{2n}(x) = n^{1/2}\{F_{\hat{\boldsymbol{\theta}}}(x) - F_{\boldsymbol{\theta}_0}(x)\}$, and $B_{3n}(x) = n^{1/2}\{\widehat{F}_n(x) - F_n(x)\}$, $x \in \mathbb{R}$. Then, $\widehat{W}_n(x) = B_{1n}(x) - B_{2n}(x) + B_{3n}(x)$, $x \in \mathbb{R}$.

We show, in turn, that B_{1n}, B_{2n}, and B_{3n} are tight, and hence so is $\widehat{W}_n$. The empirical process B_{1n} is tight since it is based on iid variables. Next, let $\mathcal{Q}_i = h'_{\boldsymbol{\theta}_0}(\varepsilon_i)\{\varphi_i(\boldsymbol{\phi}_0) + 2^{-1}\varepsilon_i\tau_i(\boldsymbol{\phi}_0)\}^\top$ and $\xi_i(\boldsymbol{\phi}_0, \varepsilon_i) = A(\boldsymbol{\phi}_0)\{\tau_i(\boldsymbol{\phi}_0)(1 - {\varepsilon_i}^2) - 2\varepsilon_i\varphi_i(\boldsymbol{\phi}_0)\}$. By Proposition 3 and Assumption 3.2 we have

$$\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 = n^{-1}\sum \xi_i(\boldsymbol{\phi}_0, \varepsilon_i) + o_p(n^{-1/2}), \quad \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = n^{-1}\sum h_{\boldsymbol{\theta}_0}(\varepsilon_i) - \mathcal{Q}_i(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) + o_p(n^{-1/2}). \tag{A.7}$$

By expanding $F_{\hat{\boldsymbol{\theta}}}(x)$ about $\boldsymbol{\theta}_0$, and substituting the foregoing representation for $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, we obtain

$$\sup_{x\in\mathbb{R}} \left|B_{2n}(x) - n^{-1/2}\sum\{h_{\boldsymbol{\theta}_0}(\varepsilon_i) - \mathcal{C}_i(\boldsymbol{\phi}_0, \varepsilon_i)\}^\top \dot{F}_{\boldsymbol{\theta}_0}(x)\right| = o_p(1), \tag{A.8}$$

where $\mathcal{C}_i(\boldsymbol{\phi}_0,\varepsilon_i)=\{n^{-1}\sum_{j=1}^n \mathcal{Q}_j\}\xi_i(\boldsymbol{\phi}_0,\varepsilon_i)$. Next, to obtain an expansion for B_{3n}, note that

$$B_{3n}(x)=n^{-1/2}\sum I(\varepsilon_i\le x+x\mathcal{V}_i+\mathcal{U}_i)-n^{-1/2}\sum I(\varepsilon_i\le x),\quad x\in\mathbb{R}.$$

Using Assumptions 1, 3, 4 and 5, and invoking Lemma 2, one obtains that $\max_{1\le i\le n}|\mathcal{V}_i|=o_p(1)$ and $\max_{1\le i\le n}|\mathcal{U}_i|=o_p(1)$. Hence, by Theorem S1 in the Supplementary Material, we obtain

$$\sup_{x\in\mathbb{R}}\left|B_{3n}(x)-n^{1/2}\{F_{\boldsymbol{\theta}_0}(x+x\mathcal{V}_i+\mathcal{U}_i)-F_{\boldsymbol{\theta}_0}(x)\}\right|=o_p(1).$$

Therefore, from the special case of Lemma 6 below with $\boldsymbol{\omega}_n=\boldsymbol{\omega}_0=(\boldsymbol{\phi}_0^\top,\boldsymbol{\theta}_0^\top)^\top$, we have that $\sup_{x\in\mathbb{R}}|B_{3n}(x)-B_0(x)|=o_p(1)$ where $B_0(x)=n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0)^\top E[(x/2)\tau_1(\boldsymbol{\phi}_0)+\varphi_1(\boldsymbol{\phi}_0)]f_{\boldsymbol{\theta}_0}(x)$. Therefore, by substituting the expansion (A.7) for $n^{1/2}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0)$ in $B_0(x)$, we obtain that

$$\sup_{x\in\mathbb{R}}|B_{3n}(x)-f_{\boldsymbol{\theta}_0}(x)E[\frac{x}{2}\tau_1(\boldsymbol{\phi}_0)+\varphi_1(\boldsymbol{\phi}_0)]^\top n^{-1/2}\sum\xi_i(\boldsymbol{\phi}_0,\varepsilon_i)|=o_p(1).\tag{A.9}$$

Next, let $g_i(t)=a_i(t)-b_i(t)+c_i(t)$, where $a_i(t)=I(\varepsilon_i\le F_{\boldsymbol{\theta}_0}^{-1}(t))-t$, $b_i(t)=\{h_{\boldsymbol{\theta}_0}(\varepsilon_i)-\mathcal{C}_i(\boldsymbol{\phi}_0,\varepsilon_i)\}^\top\dot{F}_{\boldsymbol{\theta}_0}(F_{\boldsymbol{\theta}_0}^{-1}(t))$, $c_i(t)=f_{\boldsymbol{\theta}_0}(F_{\boldsymbol{\theta}_0}^{-1}(t))E[2^{-1}F_{\boldsymbol{\theta}_0}^{-1}(t)\tau_1(\boldsymbol{\phi}_0)+\varphi_1(\boldsymbol{\phi}_0)]^\top\xi_i(\boldsymbol{\phi}_0,\varepsilon_i)$, and $G_n(t)=n^{-1/2}\sum_{i=1}^n g_i(t)$. Since $\widehat{W}_n(x)=B_{1n}(x)-B_{2n}(x)+B_{3n}(x)$, from (A.8) and (A.9), it follows that

$$\sup_{t\in[0,1]}|\sqrt{n}\{\widehat{F}_n(F_{\boldsymbol{\theta}_0}^{-1}(t))-F_{\hat{\boldsymbol{\theta}}}(F_{\boldsymbol{\theta}_0}^{-1}(t))\}-G_n(t)|=o_p(1).\tag{A.10}$$

Furthermore, we also obtain that $cov[G_n\circ F_{\boldsymbol{\theta}_0}(x),G_n\circ F_{\boldsymbol{\theta}_0}(y)]$ converges, and with $G_0(\cdot)$ as in Theorem 1, $\mathrm{cov}\{G_n(s),G_n(t)\}=\mathrm{cov}\{G_0(s),G_0(t)\}+o(1)$, $s,t\in[0,1]$.

An application of Theorem 18.3 in Billingsley (1999) yields that the finite dimensional distributions of G_n converge to those of $W_0\circ F_{\boldsymbol{\theta}_0}^{-1}$, where W_0 is the centered Gaussian process in Theorem 1. By Markov's inequality, Proposition 3, and using Assumptions 7 and 5, one can verify that each of $n^{-1/2}\sum a_i(t), n^{-1/2}\sum b_i(t)$ and $n^{-1/2}\sum c_i(t)$ is asymptotically equi-continuous. Further, from the convergence of finite dimensional distributions of G_n to those of $W_0\circ F_{\boldsymbol{\theta}_0}^{-1}$, it follows that $G_n\overset{w}{\Rightarrow}W_0\circ F_{\boldsymbol{\theta}_0}^{-1}$ in $D[0,1]$. Then, in view of (A.10) and Proposition 2, we obtain that $\widetilde{W}_n\overset{w}{\Rightarrow}W_0$ in $D[-\infty,\infty]$. Theorem 1 follows by an application of the continuous mapping theorem. □

To show that the test statistic T_j ($j=1,\ldots,5$) converges in distribution, we showed that $\widetilde{W}_n=n^{1/2}(\widetilde{F}_n-F_{\hat{\boldsymbol{\theta}}})$ converges weakly. Since $\widetilde{W}_n$ is an empirical process based on the empirical distribution function $\widetilde{F}_n$ of the residuals $\{\widetilde{\varepsilon}_1,\ldots,\widetilde{\varepsilon}_n\}$ that have a complex dependence pattern, it is difficult to prove weak convergence of $\widetilde{W}_n$ by showing its tightness directly. Therefore, we showed that $\sup_x|\widetilde{W}_n(x)-\widehat{W}_n(x)|=o_p(1)$ and that $\widehat{W}_n$ converges weakly. To indicate the new techniques used for the proof of the former, note that $\widetilde{W}_n(x)-\widehat{W}_n(x)=n^{-1/2}\sum I[\varepsilon_i\le\widetilde{\varrho}_i(x)]-n^{-1/2}\sum I[\varepsilon_i\le$

$\varrho_i(x)]$, where $\varrho_i(x) = x + x\mathcal{V}_i + \mathcal{U}_i$ and $\widetilde{\varrho}_i(x) = x + x\widetilde{\mathcal{V}}_i + \widetilde{\mathcal{U}}_i$. Intuitively, we expect the influence of initial conditions to vanish in the limit and therefore $|\widetilde{\mathcal{U}}_i - \mathcal{U}_i|$ and $|\widetilde{\mathcal{V}}_i - \mathcal{V}_i|$, and hence $|\widetilde{\varrho}_i(x) - \varrho_i(x)|$ $(i = 1, \ldots, n)$ to be small enough to ensure that $\sup_x |\widetilde{W}_n(x) - \widehat{W}_n(x)| = o_p(1)$. While the intuition turns out to be correct, we used some new techniques based on stochastic recurrence equation and $\stackrel{e.a.s.}{\to}$ rate of convergence to provide a complete answer. To this end, we used Theorem A1 of Escanciano (2010); for a discussion of the relevance of this theorem to the context of our result, see pages 749, 750, and the introductory remarks to the Appendix therein. To apply this theorem, we showed that $\max_{1\le i\le n} |\widetilde{\mathcal{U}}_i| = O_p(1)$, $\max_{1\le i\le n} |\widetilde{\mathcal{V}}_i| = O_p(1)$, and $\sum_{i=1}^n |\widetilde{\mathcal{U}}_i - \mathcal{U}_i| + |\widetilde{\mathcal{V}}_i - \mathcal{V}_i| = O_p(1)$; to this end, we used Assumption (2.1), which is a high level regularity condition involving $\stackrel{e.a.s.}{\to}$ rate of convergence. In the Supplementary Material, we show that Assumption (2.1) is satisfied by GARCH(p, q) and note that the method of verification therein is more general and that it could be applied to other GARCH type models for verifying Assumption (2.1). These verifications depend on solutions $\widetilde{\mathcal{V}}_i$ and $\widetilde{\mathcal{U}}_i$ of the stochastic recurrence equations defining the GARCH models satisfying $|\widetilde{\mathcal{U}}_i - \mathcal{U}_i| \stackrel{e.a.s.}{\to} 0$ and $|\widetilde{\mathcal{V}}_i - \mathcal{V}_i| \stackrel{e.a.s.}{\to} 0$. Thus, this paper uses $\stackrel{e.a.s.}{\to}$ rate of convergence in a novel way to prove the weak convergence of $\widetilde{W}_n$ and hence that of T_j $(j = 1, \ldots, 5)$.

A.2 Proof of Theorem 2

In this section, we assume, without further comments, that Conditions (M1)–(M4) are satisfied.

A.2.1 Some results when the true parameter is $\boldsymbol{\omega}_n$, and $\boldsymbol{\omega}_n \to \boldsymbol{\omega}_0$ as $n \to \infty$

Let Ω_0 denote a set of probability one such that $\hat{\boldsymbol{\omega}} := (\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ converges to $\boldsymbol{\omega}_0$ along every sample path in Ω_0. To establish the consistency of the bootstrap test, in probability, we restrict attention to a sample path in Ω_0. If $\hat{\boldsymbol{\omega}}$ is known to converge to $\boldsymbol{\omega}_0$ only in probability, we may work with subsequences along which $\hat{\boldsymbol{\omega}} \stackrel{a.s}{\to} \boldsymbol{\omega}_0$, and then deduce the bootstrap consistency, in probability. In what follows, we adopt this standard argument without repeating the details.

First, we introduce a notation. For a given sample path in Ω_0, let $\boldsymbol{\omega}_n$ denote $\hat{\boldsymbol{\omega}}$. Then $\boldsymbol{\omega}_n$ is not stochastic, and $\boldsymbol{\omega}_n \to \boldsymbol{\omega}_0$ as $n \to \infty$. For the bootstrap algorithm in Section 2.1, we present the samples in the form of a triangular array with the nth row corresponding to sample size n. Therefore, the sample in the nth row is generated from the model $\mathcal{M}$ with true parameter value $\boldsymbol{\omega}_n$, and one needs to use user supplied starting values at time zero to obtain $\{\widetilde{\Psi}_i(\boldsymbol{\phi}; \boldsymbol{\omega}_n), \widetilde{\mu}_i(\boldsymbol{\phi}; \boldsymbol{\omega}_n), , \widetilde{\eta}_i(\boldsymbol{\phi}; \boldsymbol{\omega}_n)\}$.

Additional notation: We adopt the following simpler notation to avoid having to show the true value $\boldsymbol{\omega}_n$ repeatedly: we write $\widetilde{\Psi}_{ni}(\boldsymbol{\phi})$ for $\widetilde{\Psi}_i(\boldsymbol{\phi}; \boldsymbol{\omega}_n)$, where the n in the double suffix of $\widetilde{\Psi}_{ni}$ indicates that the process is generated at the true parameter value $\boldsymbol{\omega}_n$, and $\widetilde{\cdot}$ indicates "conditional on user supplied initial values at time zero". More generally, we write Y_{ni} for $Y_i(\boldsymbol{\omega}_n)$ and $\widetilde{Y}_{ni}(\boldsymbol{\phi})$ for $\widetilde{Y}_i(\boldsymbol{\phi};\boldsymbol{\omega}_n)$. For example, $X_{ni}=X_i(\boldsymbol{\omega}_n)$, $\varPsi_{ni}(\cdot)=\varPsi_i(\cdot;\boldsymbol{\omega}_n)$, $\mu_{ni}(\cdot)=\mu_i(\cdot;\boldsymbol{\omega}_n)$, $\eta_{ni}(\cdot)=[X_{ni}-\mu_{ni}(\cdot)]$, $\widetilde{\mu}_{ni}(\cdot)=\widetilde{\mu}_i(\cdot;\boldsymbol{\omega}_n)$, $\widetilde{\varphi}_{ni}(\cdot)=\widetilde{\varphi}_i(\cdot;\boldsymbol{\omega}_n)$, $\widetilde{\tau}_{ni}(\cdot)=\widetilde{\tau}_i(\cdot;\boldsymbol{\omega}_n)$, $\widetilde{\eta}_{ni}(\cdot)=[X_{ni}-\widetilde{\mu}_{ni}(\cdot)]$. Similarly, we define

$$\widetilde{L}_{nn}(\boldsymbol{\phi})=\sum\widetilde{\ell}_{ni}(\boldsymbol{\phi}),\quad \widetilde{\ell}_{ni}(\boldsymbol{\phi})=\log\widetilde{\varPsi}_{ni}(\boldsymbol{\phi})+\frac{[\widetilde{\eta}_{ni}(\boldsymbol{\phi})]^2}{\widetilde{\varPsi}_{ni}(\boldsymbol{\phi})},\quad \hat{\boldsymbol{\phi}}_{nn}=\arg\min_{\boldsymbol{\phi}\in\Phi}\widetilde{L}_{nn}(\boldsymbol{\phi}),$$

and $\widetilde{\varepsilon}_{ni}=\widetilde{\eta}_{ni}(\hat{\boldsymbol{\phi}}_{nn})\{\widetilde{\varPsi}_{ni}(\hat{\boldsymbol{\phi}}_{nn})\}^{-1/2}$. Thus, $\hat{\boldsymbol{\phi}}_{nn}$ is the QML estimator based on the observable quasi log-likelihood function $-\widetilde{L}_{nn}(\boldsymbol{\phi})$ obtained with user supplied starting values at time zero, when the process is generated at $\boldsymbol{\omega}_n$. Similarly, the analog of $\hat{\boldsymbol{\theta}}$ for this setup, denoted by $\hat{\boldsymbol{\theta}}_{nn}$, is the estimator based on $\{\widetilde{\varepsilon}_{n1},\ldots,\widetilde{\varepsilon}_{nn}\}$. We assume, without loss of generality, that $\varepsilon_{ni}=F^{-1}_{\boldsymbol{\theta}_n}(U_i)$, $i\in\mathbb{Z}$, where $\{U_i\}_{i\in\mathbb{Z}}$ are iid uniform (0,1) random variables.

For our bootstrap test to be consistent, we essentially need an extension of Theorem 1 to hold for the foregoing sampling scheme in a triangular array with true parameter value $\boldsymbol{\omega}_n$ converging to $\boldsymbol{\omega}_0$. To prove such results in the triangular array setting, we first prove the corresponding result in the double array setting, and then show that the difference between the corresponding quantities from the two arrays, converges to zero. The results for the double array are easier to show since each row is stationary. This is a general approach adopted throughout.

Let P_n denote the probability law of the DGP when the true parameter is $\boldsymbol{\omega}_n$, with O_{p_n}, o_{p_n}, and E_n denoting the usual stochastic orders of magnitude and expectation, respectively, with respect to [w.r.t.] P_n. Let $K_{\boldsymbol{\phi}}(\subset\Phi)$ and $K_{\boldsymbol{\theta}}(\subset\Theta)$ be the compact neighborhoods of $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$, respectively, in Condition M. Let n_0 be a positive integer such that $\boldsymbol{\omega}_n\in K_{\boldsymbol{\omega}}:=K_{\boldsymbol{\phi}}\times K_{\boldsymbol{\theta}}$ for all $n\geq n_0$.

For a continuous matrix-valued function H on a compact set $\Lambda\subset\mathbb{R}^r$, define the norm $\|\cdot\|_\Lambda$ by $\|H\|_\Lambda=\sup_{s\in\Lambda}\|H(s)\|$, where $\|\cdot\|$ is a *consistent matrix norm*; if H is real valued then $\|H\|_\Lambda=\sup_{s\in\Lambda}|H(s)|$; if H is vector valued then $\|H\|_\Lambda=\sup_{s\in\Lambda}\|H(s)\|$ where $\|\cdot\|$ is the Euclidean norm.

By Assumption 2.1, we have $\sup_{\boldsymbol{\phi}\in\Phi}|\widetilde{\varPsi}_i(\boldsymbol{\phi})-\varPsi_i(\boldsymbol{\phi})|\overset{e.a.s.}{\to}0$ and $\sup_{\boldsymbol{\phi}\in\Phi}|\widetilde{\mu}_i(\boldsymbol{\phi})-\mu_i(\boldsymbol{\phi})|\overset{e.a.s.}{\to}0$ as $i\to\infty$. This condition is used in the proof of the consistency of the QMLE $\hat{\boldsymbol{\phi}}$. Hence, we expect that a more general version of Assumption 2.1 is likely to be required when the true parameter value is $\boldsymbol{\omega}_n$. The next lemma, which follows from Condition (M1), formalizes this requirement.

Lemma 3. (a). $\sup_{n\geq n_0}\|\widetilde{Y}_{ni}(\cdot)-Y_{ni}(\cdot)\|_\Phi\overset{e.a.s.}{\to}0$ *as* $i\to\infty$, *for* $Y_{ni}=\mu_{ni},\,\varPsi_{ni}$.
[For example, for $Y_{ni}=\mu_{ni}$, the statement says that $\sup_{n\geq n_0}\|\widetilde{\mu}_{ni}(\cdot)-\mu_{ni}(\cdot)\|_\Phi\overset{e.a.s.}{\to}0$ as $i\to\infty$.]
(b). $\sup_{n\geq n_0}\|\widetilde{Y}_{ni}(\cdot)-Y_{ni}(\cdot)\|_{K_{\boldsymbol{\phi}}}\overset{e.a.s.}{\to}0$ *as* $i\to\infty$, *for* $Y_{ni}=\dot{\mu}_{ni},\,\dot{\varPsi}_{ni}$.

Recall that Assumption 5 states that $\max_{1\leq i\leq n}\|\tau_i(\boldsymbol{\phi}_0;\boldsymbol{\omega}_0)\|$ and $\max_{1\leq i\leq n}\|\varphi_i(\boldsymbol{\phi}_0;\boldsymbol{\omega}_0)\|$ are $o_p(n^{1/2})$. Since this condition was used in the proof of Theorem 1, we expect that we are likely to require that it holds when the true parameter value is $\boldsymbol{\omega}_n$ as well. The required condition is given in part(b) of Lemma 5. It follows by Chebyshev inequality. The moment condition to apply the Chebyshev inequality is given in the next Lemma; it follows easily from Condition (M2).

Lemma 4. *There exists $d > 0$ such that, $\sup_{n\geq n_0} E_n\|\tau_{n0}(\boldsymbol{\phi}_n)\|^{2+d}$, $\sup_{n\geq n_0} E_n\|\varphi_{n0}(\boldsymbol{\phi}_n)\|^{4+d}$, $\sup_{n\geq n_0} E_n|\mu_{n0}(\boldsymbol{\phi}_n)|^{4+d}$, and $\sup_{n\geq n_0} E_n|\Psi_{n0}(\boldsymbol{\phi}_n)|^{2+d}$ are finite.*

Next, consider Assumptions 3 and 4. The crux of Assumption 3 used in the proof of Theorem 1 is that $\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)$ is asymptotically proportional to $n^{-1/2}\sum \dot{\ell}_i(\boldsymbol{\phi}_0;\boldsymbol{\omega}_0)$, and $\{\dot{\ell}_i(\boldsymbol{\phi}_0;\boldsymbol{\omega}_0)\}$ is a stationary process. Consequently, when we expand a function of the QMLE $\hat{\boldsymbol{\phi}}$ about the true value $\boldsymbol{\phi}_0$, the expansion lends itself to apply LLN/CLT. Part (a) of the next lemma states an extension of this result to the above triangular array setting.

Similarly, we need a suitable extension of Assumption 4. This assumption essentially says the following: Let $\boldsymbol{s} \in \Phi$ and consider a neighbourhood of radius $n^{-1/2}$ centered at $\boldsymbol{s}$. Then, if the functions $\mu_i(\boldsymbol{t};\boldsymbol{\omega}_0)$ and $\{\Psi_i(\boldsymbol{t};\boldsymbol{\omega}_0)\}^{1/2}$ are expanded about $\boldsymbol{t} = \boldsymbol{s}$ using a one-term Taylor series, then the leading term is linear in $(\boldsymbol{t} - \boldsymbol{s})$, and the remainder term converges to zero in probability, uniformly over $i \in \{1,\ldots,n\}$ and uniformly over $\boldsymbol{s} \in \Phi$. Part (c) in the next lemma says that these conditions are satisfied in the triangular array setting as well. The proof of the next lemma follows a familiar route using the mean value theorem and Chebyshev inequality; the proof is omitted.

Lemma 5. (a). *For all n sufficiently large, the matrices $E_n[\ddot{\ell}_{n0}(\boldsymbol{\phi}_n)]$ and $E_n[\dot{\ell}_{n0}(\boldsymbol{\phi}_n)\dot{\ell}_{n0}(\boldsymbol{\phi}_n)^\top]$ are nonsingular, $\|\hat{\boldsymbol{\phi}}_{nn} - \boldsymbol{\phi}_n\| = o_{p_n}(1)$ and $\|\sqrt{n}(\hat{\boldsymbol{\phi}}_{nn} - \boldsymbol{\phi}_n) - [n^{-1/2}A_n(\boldsymbol{\phi}_n)\sum \dot{\ell}_{ni}(\boldsymbol{\phi}_n)]\| = o_{p_n}(1)$, where $A_n(\boldsymbol{\phi}) = E_n[-\ddot{\ell}_{n0}(\boldsymbol{\phi})]^{-1}$ and $\hat{\boldsymbol{\phi}}_{nn} = \arg\min_{\boldsymbol{\phi}\in\Phi} L_{nn}(\boldsymbol{\phi})$.*

(b). $\max_{1\leq i\leq n} \|\tau_{ni}(\cdot)\|_{K_{\boldsymbol{\phi}}} = o_{p_n}(n^{1/2})$, $\max_{1\leq i\leq n} \|\varphi_{ni}(\cdot)\|_{K_{\boldsymbol{\phi}}} = o_{p_n}(n^{1/2})$.

(c). *Let $\Lambda_0(\subset K_{\boldsymbol{\phi}})$ be the closure of an open neighborhood of $\boldsymbol{\phi}_0$. Then, for every constant $C > 0$:*

(i) $\sup \sqrt{n} \mid \mu_{ni}(\boldsymbol{b}) - \mu_{ni}(\boldsymbol{a}) - (\boldsymbol{b} - \boldsymbol{a})^\top \dot{\mu}_{ni}(\boldsymbol{a}) \mid \{\Psi_{ni}(\boldsymbol{a})\}^{-1/2} = o_{p_n}(1)$,

(ii) $\sup \sqrt{n} \mid \Delta_{ni}(\boldsymbol{a},\boldsymbol{b}) \mid \{\Psi_{ni}(\boldsymbol{a})\}^{-1/2} = o_{p_n}(1)$, *where*

the supremum is taken over $1 \leq i \leq n$ and $\{(\boldsymbol{a},\boldsymbol{b}) : \boldsymbol{a},\boldsymbol{b} \in \Lambda_0, \sqrt{n}\|\boldsymbol{b} - \boldsymbol{a}\| \leq C\}$, with

$\Delta_{ni}(\boldsymbol{a},\boldsymbol{b}) = [\Psi_{ni}(\boldsymbol{b})]^{1/2} - [\Psi_{ni}(\boldsymbol{a})]^{1/2} - 2^{-1}(\boldsymbol{b} - \boldsymbol{a})^\top \dot{\Psi}_{ni}(\boldsymbol{a})[\Psi_{ni}(\boldsymbol{a})]^{-1/2}$.

Let $\quad B_n(x,\boldsymbol{\phi}) := n^{1/2}(\boldsymbol{\phi} - \boldsymbol{\phi}_n)^\top[(x/2)E_n\{\tau_{n1}(\boldsymbol{\phi}_n)\} + E_n\{\varphi_{n1}(\boldsymbol{\phi}_n)\}]f_{\boldsymbol{\theta}_n}(x)$,

$$u_{ni}(\boldsymbol{\phi}) := \frac{\mu_{ni}(\boldsymbol{\phi}) - \mu_{ni}(\boldsymbol{\phi}_n)}{\{\Psi_{ni}(\boldsymbol{\phi}_n)\}^{1/2}} \quad\text{and}\quad v_{ni}(\boldsymbol{\phi}) := \frac{\{\Psi_{ni}(\boldsymbol{\phi})\}^{1/2} - \{\Psi_{ni}(\boldsymbol{\phi}_n)\}^{1/2}}{\{\Psi_{ni}(\boldsymbol{\phi}_n)\}^{1/2}}.$$

In the proofs provided below, the expression $\sum F_{\boldsymbol{\theta}_n}\{x + xv_{ni}(\boldsymbol{\phi}) + u_{ni}(\boldsymbol{\phi})\}$ arises in the asymptotic arguments. The next lemma provides a simple asymptotic representation for this expression.

Lemma 6. $\sup_{x,\boldsymbol{\phi},K} |n^{-1/2} \sum\{F_{\boldsymbol{\theta}_n}(x + xv_{ni}(\boldsymbol{\phi}) + u_{ni}(\boldsymbol{\phi})) - F_{\boldsymbol{\theta}_n}(x)\} - B_n(x, \boldsymbol{\phi})| = o_{p_n}(1)$, *for any* $0 < K < \infty$, *where* $\sup_{x,\boldsymbol{\phi},K}$ *denotes the supremum over all* $x \in \mathbb{R}$ *and* $\{\boldsymbol{\phi} \in \Phi : \sqrt{n}\|\boldsymbol{\phi} - \boldsymbol{\phi}_n\| \leq K\}$.

Proof. For brevity, write (u_{ni}, v_{ni}) for $\big(u_{ni}(\boldsymbol{\phi}), v_{ni}(\boldsymbol{\phi})\big)$. Let $a > 0$ be as in Assumption 7.3. Then, for every $x \in \mathbb{R}$, there exists an $u_x \in \mathbb{R}$ with $|u_x| < a$ such that, for all large enough n, we have that $F_{\boldsymbol{\theta}_n}(x + xv_{ni} + u_{ni}) - F_{\boldsymbol{\theta}_n}(x) = P_{ni}(x) + Q_{ni}(x)$, where $P_{ni}(x) = (xv_{ni} + u_{ni})f_{\boldsymbol{\theta}_n}(x)$ and $Q_{ni}(x) = 2^{-1}(xv_{ni} + u_{ni})^2 f'_{\boldsymbol{\theta}_n}\{x(1 + u_x)\}$. Select $n_1 \in \mathbb{N}$ large enough to have $\boldsymbol{\theta}_n \in K_{\boldsymbol{\theta}}$ for all $n \geq n_1$. From Condition (M3)(a), we obtain $\sup_{n \geq n_1,\, x \in \mathbb{R},\, |u|<a}(x^2 + 1)|f'_{\boldsymbol{\theta}_n}\{x(1 + u)\}| < \infty$. Hence, using Lemmas 5(b) and 5(c), Condition (M3)(a), and Assumption 7.3, we obtain that $\sup_{x,\boldsymbol{\phi},K} |\sum Q_{ni}(x)| = O(\sup_{\boldsymbol{\phi},K} \sum[v_{ni} + u_{ni}]^2) = o_{p_n}(n^{1/2})$. Therefore, the proof follows. □

A.2.2 Asymptotic negligibility of the effect of initial values on bootstrap validity

It is well-known that the effect of initial values for the asymptotic normality of the QMLE for GARCH(p,q) becomes asymptotically negligible (for example, see Berkes *et al.* 2003, and pages 172 and 177 in Francq and Zakoïan 2010). By a more general version of essentially the same idea, the effect of the initial values on the bootstrap test in this paper also become asymptotically negligible. We do not provide a detailed rigorous proof, but a brief outline is provided below.

The following arguments are conditional on $X_1, \ldots, X_n$. To avoid ambiguities in the proofs, it would be better to make the notation for bootstrap quantities more precise. Let $\Psi_i^*(\cdot) = \Psi_i(\cdot; \hat{\boldsymbol{\omega}})$, $\mu_i^*(\cdot) = \mu_i(\cdot; \hat{\boldsymbol{\omega}})$, and $\eta_i^*(\cdot) = [X_i^* - \mu_i^*(\cdot)]$; recall that $X_i^* = X_i(\hat{\boldsymbol{\omega}})$. Define $\widehat{\varepsilon}_i^* = \eta_i^*(\hat{\boldsymbol{\phi}}^*)\{\Psi_i^*(\hat{\boldsymbol{\phi}}^*)\}^{-1/2}$. Recall that $\widetilde{F}_n^*(x) = n^{-1}\sum_{i=1}^n I(\widetilde{\varepsilon}_i^* \leq x)$, and $\widetilde{W}_n^*(x) = \sqrt{n}\big\{\widetilde{F}_n^*(x) - F_{\hat{\boldsymbol{\theta}}^*}(x)\big\}$, $(x \in \mathbb{R})$ are the bootstrap analogs of $\widetilde{F}_n$ and $\widetilde{W}_n$ in (7). Similarly, let $\widehat{F}_n^*(x) = n^{-1}\sum_{i=1}^n I(\widehat{\varepsilon}_i^* \leq x)$, and $\widehat{W}_n^*(x) = \sqrt{n}\big\{\widehat{F}_n^*(x) - F_{\hat{\boldsymbol{\theta}}^*}(x)\big\}$, $(x \in \mathbb{R})$, be the bootstrap analogs of $\widehat{F}_n$ and $\widehat{W}_n$, respectively.

It turns out that, $\sup_{x \in \mathbb{R}} |\widetilde{F}_n^*(x) - \widehat{F}_n^*(x)| = o_{p_n^*}(n^{-1/2})$, in probability, as $n \to \infty$. The main driver of the proof of this result is that the stochastic recurrence equation that defines model $\mathcal{M}$ is *contractive on average* as defined in Straumann and Mikosch (2006). Consequently, we obtain $\sup_{x \in \mathbb{R}} |\widetilde{W}_n^*(x) - \widehat{W}_n^*(x)| = o_{p_n^*}(1)$, in probability, as $n \to \infty$. Therefore, to prove Theorem 2, it suffices to show that $\widehat{W}_n^* \Rightarrow W_0$, in probability, as $n \to \infty$. This is shown in the next subsection.

A.2.3 Proof of $\widehat{W}_n^* \Rightarrow W_0$, in probability, as $n \to \infty$

To prove that the bootstrap testing procedure is asymptotically valid, we need to establish that the bootstrap empirical process $\widehat{W}_n^*$ converges weakly. Let

$$v_{ni}^* = \{\Psi_i^*(\hat{\boldsymbol{\phi}})\}^{-1/2}\{\Psi_i^*(\hat{\boldsymbol{\phi}}^*)\}^{1/2} - 1, \qquad u_{ni}^* = \{\mu_i^*(\hat{\boldsymbol{\phi}}^*) - \mu_i^*(\hat{\boldsymbol{\phi}})\}\{\Psi_i^*(\hat{\boldsymbol{\phi}})\}^{-1/2} \tag{A.11}$$

$$\widehat{\varepsilon}_i^* = \{X_i^* - \mu_i^*(\hat{\boldsymbol{\phi}}^*)\}\{\Psi_i^*(\hat{\boldsymbol{\phi}}^*)\}^{-1/2}, \qquad \varepsilon_i^* = \{X_i^* - \mu_i^*(\hat{\boldsymbol{\phi}})\}\{\Psi_i^*(\hat{\boldsymbol{\phi}})\}^{-1/2}. \tag{A.12}$$

Recall that $\widehat{F}_n^*(x) = n^{-1}\sum I(\widehat{\varepsilon}_i^* \le x)$, $F_n^*(x) = n^{-1}\sum I(\varepsilon_i^* \le x)$, and $\widehat{W}_n^* = n^{1/2}(\widehat{F}_n^* - F_{\hat{\boldsymbol{\theta}}^*})$. In view of the similarity of $\widehat{W}_n^*$ to $\widehat{W}_n$, we adopt a method similar to that in the proof of Theorem 1 for $\widehat{W}_n$ to derive the weak convergence of $\widehat{W}_n^*$. Thus, we first decompose $\widehat{W}_n^*$ into three components as for $\widehat{W}_n$ in the proof of Theorem 1: $\widehat{W}_n^* = n^{1/2}(F_n^* - F_{\hat{\boldsymbol{\theta}}}) - n^{1/2}(F_{\hat{\boldsymbol{\theta}}^*} - F_{\hat{\boldsymbol{\theta}}}) + n^{1/2}(\widehat{F}_n^* - F_n^*)$. The third term, $n^{1/2}(\widehat{F}_n^* - F_n^*)$, is the most challenging of the three as there are no standard techniques that can be applied to expand this term. Several related results on asymptotic uniform expansions of weighted empirical processes have been developed in Koul and Ossiander (1994), Koul (2002) and Koul and Ling (2006). However, these results are not applicable to the setup of our bootstrap data generation. Therefore, in the Supplementary Material of this paper, we develop a general result on asymptotic uniform expansions applicable to $n^{1/2}(\widehat{F}_n^* - F_n^*)$ by extending some uniform expansion results of Koul and Ling (2006) and Koul and Ossiander (1994). The next lemma uses this general result to obtain a tractable asymptotic uniform expansion for $n^{1/2}(\widehat{F}_n^* - F_n^*)$.

Lemma 7. *Let* $\tilde{\mathcal{U}}_n^*(x) = \sqrt{n}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}})^\top[(x/2)E_*\{\tau_1^*(\hat{\boldsymbol{\phi}})\} + E_*\{\varphi_1^*(\hat{\boldsymbol{\phi}})\}]f_{\hat{\boldsymbol{\theta}}}(x)$ *and* $\mathcal{U}_n^*(x) = \sqrt{n}\{\widehat{F}_n^*(x) - F_n^*(x)\}$. *Then,* $\sup_{x\in\mathbb{R}}|\mathcal{U}_n^*(x) - \tilde{\mathcal{U}}_n^*(x)| = o_{p_n^*}(1)$, *in probability.*

Proof. Note that $\widehat{F}_n^*(x) = n^{-1}\sum I(\varepsilon_i^* \le x + xv_{ni}^* + u_{ni}^*)$. Fix a sample path along which $\boldsymbol{\omega}_n = \hat{\boldsymbol{\omega}} \to \boldsymbol{\omega}_0$. By applying Lemma 5(b), with $(\varphi_{ni}, \tau_{ni}, \boldsymbol{\phi}_n) = (\varphi_i^*, \tau_i^*, \hat{\boldsymbol{\phi}})$, we obtain that

$$\max_{1\le i\le n} n^{-1/2}\|\tau_i^*(\hat{\boldsymbol{\phi}})\| = o_{p_n^*}(1) \quad \text{and} \quad \max_{1\le i\le n} n^{-1/2}\|\varphi_i^*(\hat{\boldsymbol{\phi}})\| = o_{p_n^*}(1). \tag{A.13}$$

Let $v_i^*(t) := \{\Psi_i^*(\hat{\boldsymbol{\phi}})\}^{-1/2}\{\Psi_i^*(\hat{\boldsymbol{\phi}} + n^{-1/2}t)\}^{1/2} - 1$, $u_i^*(t) := \{\mu_i^*(\hat{\boldsymbol{\phi}} + n^{-1/2}t) - \mu_i^*(\hat{\boldsymbol{\phi}})\}\{\Psi_i^*(\hat{\boldsymbol{\phi}})\}^{-1/2}$,

$$\mathcal{U}_n^*(x,t) := n^{-1/2}\sum\{I(\varepsilon_i^* \le x + xv_i^*(t) + u_i^*(t)) - I(\varepsilon_i^* \le x)\}, \quad x \in \mathbb{R},\ t \in \mathbb{R}^r.$$

In what follows we apply Theorem S1 stated in the Supplementary Material to show that, for every $0 < b < \infty$, $\sup_{x\in\mathbb{R},\,\|t\|<b}|\mathcal{U}_n^*(x,t) - n^{-1/2}\sum\{F_{\boldsymbol{\theta}_n}(x + xv_i^*(t) + u_i^*(t)) - F_{\boldsymbol{\theta}_n}(x)\}| = o_{p_n^*}(1)$. In order to apply Theorem S1 to $\mathcal{U}_n^*(x,t)$ we need to show that Conditions (D1)–(D5) in the Supplementary Material are satisfied, with $\gamma_{ni}^*(t) = 1, \rho_{ni}^*(t) = v_i^*(t)$ and $\xi_{ni}^*(t) = u_i^*(t)$. With $\gamma_{ni}^*(t) = 1$, the Conditions (D1) and (D5) are trivially satisfied. Next, we consider (D2), (D3) and (D4).

By applying Lemma 5(c), with $(\mu_{ni}, \Psi_{ni}, \boldsymbol{\phi}_n) = (\mu_i^*, \Psi_i^*, \hat{\boldsymbol{\phi}})$, we obtain that, for every $t \in \mathbb{R}^r$,

$$\max_{1\le i\le n} |v_i^*(t) - 2^{-1}n^{-1/2}t^\top \tau_i^*(\hat{\boldsymbol{\phi}})| = o_{p_n^*}(n^{-1/2}), \quad \max_{1\le i\le n} |u_i^*(t) - n^{-1/2}t^\top \varphi_i^*(\hat{\boldsymbol{\phi}})| = o_{p_n^*}(n^{-1/2}). \quad \text{(A.14)}$$

By (A.13) and (A.14), for each $t \in \mathbb{R}^r$, $\max_{1\le i\le n} |v_i^*(t)| = o_{p_n^*}(1)$ and $\max_{1\le i\le n} |u_i^*(t)| = o_{p_n^*}(1)$. Hence (D2) holds. Let $t \in \mathbb{R}^r$ be fixed but arbitrary. By the triangle inequality

$$\begin{aligned} n^{-1/2}\sum |v_i^*(t)| &\le n^{-1/2}\sum |v_i^*(t) - 2^{-1}n^{-1/2}t^\top \tau_i^*(\hat{\boldsymbol{\phi}})| + n^{-1/2}\sum |2^{-1}n^{-1/2}t^\top \tau_i^*(\hat{\boldsymbol{\phi}})| \\ &\le n^{1/2}\max_{1\le i\le n} |v_i^*(t) - 2^{-1}n^{-1/2}t^\top \tau_i^*(\hat{\boldsymbol{\phi}})| + 2^{-1}\|t\|n^{-1}\sum \|\tau_i^*(\hat{\boldsymbol{\phi}})\|. \quad \text{(A.15)} \end{aligned}$$

Since $n^{1/2}\max_{1\le i\le n} |v_i^*(t) - 2^{-1}n^{-1/2}t^\top \tau_i^*(\hat{\boldsymbol{\phi}})| = o_{p_n^*}(1)$ by (A.13) and $n^{-1}\sum \|\tau_i^*(\hat{\boldsymbol{\phi}})\| = O_{p_n^*}(1)$ by the Ergodic Theorem, it follows from (A.15) that $n^{-1/2}\sum |v_i^*(t)| = O_{p_n^*}(1)$. Similarly, we also obtain $n^{-1/2}\sum |u_i^*(t)| = O_{p_n^*}(1)$. Thus, $n^{-1/2}\sum\{|v_i^*(t)| + |u_i^*(t)|\} = O_{p_n^*}(1)$. Hence (D3) also holds.

By Lemma 5(c), (A.13), (A.14) and the Ergodic Theorem, for every constant $\delta > 0$,

$$\begin{aligned} & n^{-1/2}\sum \left\{ \sup_{\|t-s\|<\delta} |v_i^*(t) - v_i^*(s)| + \sup_{\|t-s\|<\delta} |u_i^*(t) - u_i^*(s)| \right\} \\ &\le n^{-1/2}\sum \left\{ |2^{-1}n^{-1/2}\delta^\top \tau_i^*(\hat{\boldsymbol{\phi}})| + |n^{-1/2}\delta^\top \varphi_i^*(\hat{\boldsymbol{\phi}})| \right\} + o_{p_n^*}(1) \\ &\le \delta n^{-1}\sum \left\{ 2^{-1}\|\tau_i^*(\hat{\boldsymbol{\phi}})\| + \|\varphi_i^*(\hat{\boldsymbol{\phi}})\| \right\} + o_{p_n^*}(1) = \delta O_{p_n^*}(1). \end{aligned}$$

Hence, for every $\epsilon > 0$, we can select a sufficiently small $\delta > 0$ and an $n_0 \in \mathbb{N}$ such that, $\forall n > n_0$,

$$P\left(n^{-1/2}\sum \left\{ \sup_{\|t-s\|<\delta} |v_i^*(t) - v_i^*(s)| + \sup_{\|t-s\|<\delta} |u_i^*(t) - u_i^*(s)| \right\} \le \epsilon \right) > 1 - \epsilon.$$

Therefore, Condition (D4) also holds. Hence, by applying Theorem S1 we obtain that, for every $0 < b < \infty$, $\sup_{x\in\mathbb{R}, \|t\|<b} |\mathcal{U}_n^*(x,t) - n^{-1/2}\sum\{F_{\boldsymbol{\theta}_n}(x + xv_i^*(t) + u_i^*(t)) - F_{\boldsymbol{\theta}_n}(x)\}| = o_{p_n^*}(1)$. From Lemma 5(a) and the Ergodic Theorem, we have that $n^{1/2}(\hat{\boldsymbol{\phi}}^* - \hat{\boldsymbol{\phi}}) = O_{p_n^*}(1)$, and hence it also follows that $\sup_{x\in\mathbb{R}} |\mathcal{U}_n^*(x) - n^{-1/2}\sum\{F_{\boldsymbol{\theta}_n}(x + xv_{ni}^* + u_{ni}^*) - F_{\boldsymbol{\theta}_n}(x)\}| = o_{p_n^*}(1)$. This result together with an application of Lemma 6 complete the proof. □

Proposition 4. *$\widehat{W}_n^* \circ F_{\hat{\boldsymbol{\theta}}}^{-1}(\cdot) = n^{1/2}[\widehat{F}_n^*\{F_{\hat{\boldsymbol{\theta}}}^{-1}(\cdot)\} - F_{\hat{\boldsymbol{\theta}}^*}\{F_{\hat{\boldsymbol{\theta}}}^{-1}(\cdot)\}] \overset{w}{\Rightarrow} G_0(\cdot)$, in probability, where $G_0(\cdot)$ is as in Theorem 1, with $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$ denoting a pseudo-true value if H_0 is not satisfied.*

Proof. As mentioned earlier, we fix a sample path along which $\boldsymbol{\omega}_n = \hat{\boldsymbol{\omega}} \to \boldsymbol{\omega}_0$, and partition the process $\widehat{W}_n^*(\cdot) = n^{1/2}\{\widehat{F}_n^*(\cdot) - F_{\hat{\boldsymbol{\theta}}^*}(\cdot)\}$ as

$$n^{1/2}\{\widehat{F}_n^*(\cdot) - F_{\hat{\boldsymbol{\theta}}^*}(\cdot)\} = n^{1/2}\{F_n^*(\cdot) - F_{\hat{\boldsymbol{\theta}}}(\cdot)\} - n^{1/2}\{F_{\hat{\boldsymbol{\theta}}^*}(\cdot) - F_{\hat{\boldsymbol{\theta}}}(\cdot)\} + n^{1/2}\{\widehat{F}_n^*(\cdot) - F_n^*(\cdot)\}. \quad \text{(A.16)}$$

We consider the three processes on the righthand side of (A.16) separately. The first process $n^{1/2}\{F_n^*(x) - F_{\hat{\boldsymbol\theta}}(x)\}$ in (A.16) is equal to $n^{-1/2}\sum\{I(\varepsilon_i^* \le x) - F_{\hat{\boldsymbol\theta}}(x)\}$; it can be handled as in the classical Donsker theorem because it is an empirical process of iid terms.

To study the term $n^{1/2}\{F_{\hat{\boldsymbol\theta}^*}(x) - F_{\hat{\boldsymbol\theta}}(x)\}$ in (A.16), let $\xi_i^*(\hat{\boldsymbol\phi}, \varepsilon_i^*) := \hat{A}(\hat{\boldsymbol\phi})[\tau_i^*(\hat{\boldsymbol\phi})(1-\varepsilon_i^{*2}) - 2\varepsilon_i^*\varphi_i^*(\hat{\boldsymbol\phi})]$ where $\hat{A}(\boldsymbol\phi) = E_*[-\ddot{\ell}_0^*(\boldsymbol\phi)]^{-1}$. Further, let $\mathcal{Q}_i^* := h_{\hat{\boldsymbol\theta}}'(\varepsilon_i^*)[\varphi_i^*(\hat{\boldsymbol\phi}) + 2^{-1}\varepsilon_i^*\tau_i^*(\hat{\boldsymbol\phi})]^\top$. Then, from Condition (M3) and Lemma 5(a), we have that

$$\hat{\boldsymbol\theta}^* - \hat{\boldsymbol\theta} = \frac{1}{n}\sum h_{\hat{\boldsymbol\theta}}(\varepsilon_i^*) - \mathcal{Q}_i^*(\hat{\boldsymbol\phi}^* - \hat{\boldsymbol\phi}) + o_{p_n^*}(n^{-1/2}),\ \hat{\boldsymbol\phi}^* - \hat{\boldsymbol\phi} = \frac{1}{n}\sum \xi_i^*(\hat{\boldsymbol\phi}, \varepsilon_i^*) + o_{p_n^*}(n^{-1/2}). \quad \text{(A.17)}$$

To study the asymptotic behaviour of $n^{1/2}\{F_{\hat{\boldsymbol\theta}^*}(x) - F_{\hat{\boldsymbol\theta}}(x)\}$, let us expand $F_{\hat{\boldsymbol\theta}^*}(x)$ about $\hat{\boldsymbol\theta}$ using Condition (M3), Lemma 5(a), and (A.17) and obtain

$$\sup_{x\in\mathbb{R}}\left|n^{1/2}\{F_{\hat{\boldsymbol\theta}^*}(x) - F_{\hat{\boldsymbol\theta}}(x)\} - n^{-1/2}\sum\{h_{\hat{\boldsymbol\theta}}(\varepsilon_i^*) - \mathcal{C}_{ni}^*(\hat{\boldsymbol\phi}, \varepsilon_i^*)\}^\top \dot{F}_{\hat{\boldsymbol\theta}}(x)\right| = o_{p_n^*}(1),$$

where $\mathcal{C}_{ni}^*(\hat{\boldsymbol\phi}, \varepsilon_i^*) = \{n^{-1}\sum_{j=1}^n \mathcal{Q}_j^*\}\xi_i^*(\hat{\boldsymbol\phi}, \varepsilon_i^*)$. By (A.17) and Lemma 7,

$$\sup_{x\in\mathbb{R}}|n^{1/2}\{\widehat{F}_n^*(x) - F_n^*(x)\} - f_{\hat{\boldsymbol\theta}}(x)E_*[(x/2)\tau_1^*(\hat{\boldsymbol\phi}) + \varphi_1^*(\hat{\boldsymbol\phi})]^\top n^{-1/2}\sum \xi_i^*(\hat{\boldsymbol\phi}, \varepsilon_i^*)| = o_{p_n^*}(1).$$

Let $G_n^*(t) = n^{-1/2}\sum_{i=1}^n\{a_{ni}(t) - b_{ni}(t) + c_{ni}(t)\}$, where $a_{ni}(t) = I(\varepsilon_i^* \le F_{\hat{\boldsymbol\theta}}^{-1}(t)) - t$, $b_{ni}(t) = \{h_{\hat{\boldsymbol\theta}}(\varepsilon_i^*) - \mathcal{C}_{ni}^*(\hat{\boldsymbol\phi}, \varepsilon_i^*)\}^\top \dot{F}_{\hat{\boldsymbol\theta}}(F_{\hat{\boldsymbol\theta}}^{-1}(t))$, and $c_{ni}(t) = f_{\hat{\boldsymbol\theta}}(F_{\hat{\boldsymbol\theta}}^{-1}(t))E_*[2^{-1}F_{\hat{\boldsymbol\theta}}^{-1}(t)\tau_1^*(\hat{\boldsymbol\phi}) + \varphi_1^*(\hat{\boldsymbol\phi})]^\top \xi_i^*(\hat{\boldsymbol\phi}, \varepsilon_i^*)$. Then, $\widehat{W}_n^* \circ F_{\hat{\boldsymbol\theta}}^{-1}(t) = \sqrt{n}\{\widehat{F}_n^*(F_{\hat{\boldsymbol\theta}}^{-1}(t)) - F_{\hat{\boldsymbol\theta}^*}(F_{\hat{\boldsymbol\theta}}^{-1}(t))\} = G_n^*(t) + o_{p_n^*}(1)$, uniformly in $t \in [0,1]$.

From a martingale central limit theorem (for example, Theorem 18.1 in Billingsley 1999), the finite dimensional distributions of G_n^* converge weakly, in probability, as $n \to \infty$ to those of G_0. By applying Markov's inequality and using Condition (M3) and Lemma 5, one obtains that each of $n^{-1/2}\sum a_{ni}(t), n^{-1/2}\sum b_{ni}(t)$ and $n^{-1/2}\sum c_{ni}(t)$ is asymptotically stochastically equi-continuous. Hence, $G_n^* \overset{w}{\Rightarrow} G_0$, in probability. Therefore, $\widehat{W}_n^* \circ F_{\hat{\boldsymbol\theta}}^{-1} \overset{w}{\Rightarrow} G_0$, in probability. □

A.3 Proof of Theorem 3

Let $P_0^{(n)}$ and $P_1^{(n)}$ denote the joint distributions of the random variables $\{X_1, \ldots, X_n\}$ under H_0 and H_{1n}, respectively. First, we show that $P_1^{(n)}$ is contiguous w.r.t. $P_0^{(n)}$. To this end, let $f_{(n)}$, $f_{\boldsymbol\theta_0}$, and $\tilde{f}$ denote the densities of $F_{(n)}$, $F_{\boldsymbol\theta_0}$, and $\widetilde{F}$, respectively. Let $\delta > 0$ be fixed but arbitrary. Since $F_{(n)} = (1 - n^{-1/2}\delta)F_{\boldsymbol\theta_0} + n^{-1/2}\delta\widetilde{F}$, we have $f_{(n)} = f_{\boldsymbol\theta_0} + n^{-1/2}\delta(\tilde{f} - f_{\boldsymbol\theta_0})$. Hence, it follows that

$$\int\{f_{(n)}^{1/2}(x) - f_{\boldsymbol\theta_0}^{1/2}(x) - (n^{-1/2}\delta/2)\xi(x)f_{\boldsymbol\theta_0}^{1/2}(x)\}^2 dx = o(n^{-1}), \quad \text{(A.18)}$$

where $\xi(x) = \{\tilde{f}(x) - f_{\boldsymbol\theta_0}(x)\}f_{\boldsymbol\theta_0}^{-1}(x)$. Further, for any sequence of real numbers $\{c_n : n \in \mathbb{N}\}$, with $f_{c_n} = f_{\boldsymbol\theta_0} + c_n(\tilde{f} - f_{\boldsymbol\theta_0})$, $|c_n| < 1$, where $c_n \to 0$ as $n \to \infty$, we have that $\int\{f_{(n)}^{1/2}(x) - f_{\boldsymbol\theta_0}^{1/2}(x) -$

$c_n\xi(x)f_{\boldsymbol{\theta}_0}^{1/2}(x)\}^2dx = o(c_n^2)$. The log likelihood ratio of $P_1^{(n)}$ to $P_0^{(n)}$ is $\lambda_n = \sum \log\{f_{(n)}(\varepsilon_i)/f_{\boldsymbol{\theta}_0}(\varepsilon_i)\}$. Let $\sigma^2 = \int\{\tilde{f}(x) - f_{\boldsymbol{\theta}_0}(x)\}^2 f_{\boldsymbol{\theta}_0}^{-1}(x)dx$. Then, under $P_0^{(n)}$, one obtains that $\lambda_n = \delta n^{-1/2}\sum\xi(\varepsilon_i) - 2^{-1}\delta^2\sigma^2 + o_p(1)$, by proceeding as in the proof of Theorem 7.2 in van der Vaart (1998). Hence, by a central limit theorem, λ_n converges in distribution, under $P_0^{(n)}$, to a random variable V that is distributed as $N(-2^{-1}\delta^2\sigma^2, \delta^2\sigma^2)$. Therefore, by Le Cam's first lemma (see van der Vaart and Wellner, 1996, Theorem 3.10.2), it follows that $P_1^{(n)}$ is contiguous w.r.t. $P_0^{(n)}$.

In the proof of Theorem 1, we constructed a process $G_n(\cdot)$, such that $G_n(t) = n^{-1/2}\sum g_i(t)$, where $g_i(\cdot)$ forms a martingale difference sequence. By proceeding as in the proof of Theorem 1, under $P_0^{(n)}$, $G_n(t)$ converges weakly to a centered Gaussian process with the same covariance kernel as $G_0(\cdot)$, and $\sup_{t\in[0,1]}|\widetilde{W}_n(F_{\boldsymbol{\theta}_0}^{-1}(t)) - G_n(t)| = o_p(1)$. Further, under $P_0^{(n)}$, $\mathrm{cov}[G_n(t), \lambda_n] = E[G_n(t)\lambda_n] = \delta m_a(t) + o(1)$, where $m_a(t) = m_{a1}(t) - m_{a2}(t)$, $m_{a1}(t) = [\widetilde{F}\{F_{\boldsymbol{\theta}_0}^{-1}(t)\} - t]$, and $m_{a2}(t) = [\int h_{\boldsymbol{\theta}_0}(\varepsilon)\ d\widetilde{F}(\varepsilon)]^\top \dot{F}_{\boldsymbol{\theta}_0}\{F_{\boldsymbol{\theta}_0}^{-1}(t)\}$. Therefore, by Le Cam's third lemma (see van der Vaart and Wellner, 1996, Theorem 3.10.7), under H_{1n}, $\widetilde{W}_n \circ F_{\boldsymbol{\theta}_0}^{-1}(\cdot) \overset{w}{\Rightarrow} \widetilde{G}(\cdot)$ in $D[0,1]$ where $\widetilde{G}(\cdot) = \delta m_a(\cdot) + W_0 \circ F_{\boldsymbol{\theta}_0}^{-1}(\cdot)$. Thus, the first part of Theorem 3 holds with $m(\cdot) = m_a \circ F_{\boldsymbol{\theta}_0}(\cdot)$.

Recall that $C[0,1]$ denotes the set of continuous functions on $[0,1]$. Let $\|h\|_\infty = \sup_{0\le t\le 1}|h(t)|$ and $\|h\|_2 = \{\int_0^1 h^2(t)\ dt\}^{1/2}$ $(h \in C[0,1])$. Then, $C[0,1]$ equipped with any one of these norms, or any equivalent norms, is a separable Banach space (Kreyszig 1978, pages 61, 62, 180).

For the test to have local power $m_a(\cdot)$ must be non-zero. To verify that $m_a(\cdot) \neq 0$, let us suppose that $m_a(\cdot) = 0$. Then $m_{a1}(t) = m_{a2}(t)$ for every $t \in [0,1]$, and hence $F_{(n)} = [1 + n^{-1/2}\delta\boldsymbol{b}^\top \dot{F}_{\boldsymbol{\theta}_0}]$, where $\boldsymbol{b} = \int h_{\boldsymbol{\theta}_0}(\varepsilon)\ d\widetilde{F}(\varepsilon)$. Since $F_{(n)}(-\infty) = 0$, we have $n^{-1/2}\delta\boldsymbol{b}^\top\dot{F}_{\boldsymbol{\theta}_0}(-\infty) = -1$ which is a contradiction. It follows from the assumptions in Theorem 3 that $m_a(\cdot)$ is continuous on $[0,1]$ and $m_a(0) = 0 = m_a(1)$; the same holds for $-m_a(\cdot)$ as well. Therefore, $m_a(\cdot)$ and $-m_a(\cdot)$ are nonzero and lie in the *support* of G_0, which we denote by $\mathrm{supp}(G_0)$; let us recall that $h \in \mathrm{supp}(G_0)$ if the G_0 measure of $\{y \in C[0,1] : \|y - h\| < \epsilon\}$ is positive for every $\epsilon > 0$.

Let $\|\cdot\|$ denote a norm on $C[0,1]$ and suppose that $C[0,1]$ equipped with $\|\cdot\|$ is a separable Banach space. Let ν denote the Gaussian probability measure on $\sigma\{C[0,1]\}$ generated by G_0. Further, let $c > 0$ be given and let B denote the ball $\{g \in C[0,1] : \|g\|_\infty < c\}$. Then, $\nu(B) = P[G_0 \in B] = P[\|G_0\|_\infty < c]$, and $\nu(B+\delta(-m_a)) = P[G_0 \in B - \delta m_a] = P[G_0 + \delta m_a \in B] = P[\|G_0 + \delta m_a\|_\infty < c]$. Therefore, by Lewandowski *et al.* (1995), $P[\|G_0 + \delta m_a\|_\infty > c]$ strictly increases from $P[\|G_0\|_\infty > c]$ to one as δ increases from zero to ∞. In what follows, we use this result with $\|\cdot\|$ corresponding to each of the test statistics $T_1, \ldots, T_5$ in turn.

Kolmogorov–Smirnov type test [T_1]: The asymptotic distribution of KS is $\|\delta m_a(\cdot) + G_0(\cdot)\|_\infty$, where G_0 is Gaussian. Let $c > 0$ be given. Then, $P[T_1 > c \mid H_{1n}] \to P[\|\delta m_a(\cdot) + G_0(\cdot)\|_\infty > c]$ as $n \to \infty$. Therefore, the Kolmogorov–Smirnov type test based on T_1 is asymptotically unbiased, and its asymptotic power against H_{1n}, strictly increases with δ.

Kuiper type test [T_3]: For a given $h \in C[0,1]$, let $\|h\|_{Ku} = \sup_{x\in\mathbb{R}} h(x) + \sup_{x\in\mathbb{R}}\{-h(x)\}$. Then $\|\cdot\|_\infty$ and $\|\cdot\|_{Ku}$ are equivalent since $\|h\|_\infty \le \|h\|_{Ku} \le 2\|h\|_\infty$. Therefore, $C[0,1]$ equipped with $\|\cdot\|_{Ku}$ is a separable Banach space. By repeating the foregoing arguments for the Kolmogorov–Smirnov type test, we conclude that T_3 is also asymptotically unbiased, and its asymptotic power against H_{1n} strictly increases with δ.

Cramér–von Mises Type test [T_2]: The proof for this follows similarly with $\|\cdot\|_\infty$ replaced by $\|\cdot\|_2$. The proof for the Anderson–Darling test [A_n^2] is very similar and hence omitted.

Watson's type test [T_5]: Let $X_0(\cdot) = G_0(\cdot) - \int_0^1 G_0(t)\,dt$ and $m_a^*(\cdot) = m_a(\cdot) - \int_0^1 m_a(t)\,dt$. It is easily verified that X_0 is Gaussian, $m_a^* \ne 0$, and m_a^* lies in the support of X_0 since m_a lies in the support of G_0. Further, $P[T_5 > c \mid H_{1n}] \to P[\|\delta m_a^* + X_0\|_\infty > c]$ as $n \to \infty$. Therefore, again by arguments similar to those in the previous paragraphs, the Watson's type test also has the desired local asymptotic properties stated in Theorem 3.

Next, to prove the second part, note that by proceeding as in the proof of Theorem 2, one obtains that, under H_{1n}, $T_j^* \xrightarrow{d^*} \mathfrak{g}_j\{G_0\}$, in probability $(j = 1, \ldots, 5)$. From the first part of this theorem and an application of the continuous mapping theorem $T_j \xrightarrow{d} \mathfrak{g}_j\{\widetilde{G}\}$ under H_{1n}, $(j = 1, \ldots, 5)$. Recall that G_0 is obtained by substituting $\delta = 0$ in $\widetilde{G}(\cdot) = \delta m_a(\cdot) + W_0 \circ F_{\boldsymbol{\theta}_0}^{-1}(\cdot)$. Hence, the foregoing local asymptotic properties of the tests also hold for their bootstrap implementations.

References

Andreou, E. and Werker, B. J. M. (2015). Residual-based rank specification tests for AR-GARCH type models. *J. Econometrics*, **185**(2), 305–331.

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

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

Bai, X. and Lam, J. S. L. (2019). A copula-garch approach for analyzing dynamic conditional dependency structure between liquefied petroleum gas freight rate, product price arbitrage and crude oil price. *Energy Economics*, **78**, 412–427.

Bera, A. K. and Jarque, C. M. (1981). Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals: Monte carlo evidence. *Econom. Lett.*, **7**, 313–318.

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

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

Bierens, H. J. and Wang, L. (2017). Weighted simulated integrated conditional moment tests for parametric conditional distributions of stationary time series processes. *Econometric Rev.*, **36**(1-3), 103–135.

Billingsley, P. (1999). *Convergence of probability measures.* John Wiley & Sons Inc., New York, second edition.

Chen, B. and Hong, Y. (2014). A unified approach to validating univariate and multivariate conditional distribution models in time series. *J. Econometrics*, **178**(part 1), 22–44.

Chen, S. X. and Gao, J. (2011). Simultaneous specification testing of mean and variance structures in nonlinear time series regression. *Econometric Theory*, **27**(4), 792–843.

Chen, S. X., Härdle, W., and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J. R. Stat. Soc. Ser. B*, **65**(3), 663–678.

Chen, Y.-T. (2012). A simple approach to standardized-residuals-based higher-moment tests. *Journal of Empirical Finance*, **19**(4), 427–453.

Cho, J. and White, H. (2014). Testing the equality of two positive-definite matrices with application to information matrix testing. In Y. Chang, T. Fomby, and J. Y. Park, editors, *Advances in econometrics: essays in honor of Peter CB Phillips*, volume 33, pages 491–556. Emerald Group Publishing Limited, West Yorkshire.

Conrad, C. and Mammen, E. (2016). Asymptotics for parametric garch-in-mean models. *J. Econometrics*, **194**(2), 319–329.

Corradi, V. and Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *J. Econometrics*, **133**(2), 779–806.

Delgado, M. A. and Stute, W. (2008). Distribution-free specification tests of conditional models. *J. Econometrics*, **143**(1), 37–55.

Dette, H., Pardo-Fernández, J. C., and Van Keilegom, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scand. J. Stat.*, **36**(4), 782–799.

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. *Ann. Statist.*, **46**(3), 1167–1196.

Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, **10**(4), 605–637.

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

Francq, C., Lepage, G., and Zakoï an, J.-M. (2011). Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *J. Econometrics*, **165**(2), 246–257.

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.

Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, **71**(1), 285–317.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, **35**(3), 705–30.

Horváth, L., Kokoszka, P., and Teyssière, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *J. Stat. Comput. Simul.*, **74**(7), 469–485.

Huo, L. and Cho, J. S. (2021). Testing for the sandwich-form covariance matrix of the quasi-maximum likelihood estimator. *Test*, **30**, 293–317.

Inglot, T. and Stawiarski, B. (2005). Data-driven score test of fit for conditional distribution in the GARCH$(1,1)$ model. *Probab. Math. Statist.*, **25**(2, Acta Univ. Wratislav. No. 2784), 331–362.

Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.*, **32**(3), 995–1034.

Koul, H. L. (2002). *Weighted empirical processes in dynamic nonlinear models*, volume 166 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

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

Koul, H. L. and Mimoto, N. (2012). A goodness-of-fit test for GARCH innovation density. *Metrika*, **75**(1), 127–149.

Koul, H. L. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.*, **22**(1), 540–562.

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.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. John Wiley and Sons, New York.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer Series in Statistics. Springer-Verlag, New York.

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH$(1,1)$ quasi-maximum likelihood estimator. *Econometric Theory*, **10**(1), 29–52.

Leucht, A., Kreiss, J.-P., and Neumann, M. H. (2015). A model specification test for GARCH$(1,1)$ processes. *Scand. J. Stat.*, **42**(4), 1167–1193.

Lewandowski, M., Ryznar, M., and Żak, T. (1995). Anderson inequality is strict for gaussian and stable measures. *Proceedings of the American Mathematical Society*, **123**, 3875–3880.

Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *J. Econometrics*, **140**(2), 849–873.

Ling, S. and Tong, H. (2011). Score based goodness-of-fit tests for time series. *Statist. Sinica*, **21**, 1807–1829.

Neumann, M. H. and Paparoditis, E. (2008). Goodness-of-fit tests for markovian time series models: Central limit theory and bootstrap approximations. *Bernoulli*, **14**(1), 14–46.

Perera, I. and Koul, H. L. (2017). Fitting a two phase threshold multiplicative error model. *J. 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. (2021). Bootstrap based probability forecasting in multiplicative error models. *J. Econometrics*, **221**, 1–24.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained statistical inference*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Stawiarski, B. (2009). Score test of fit for composite hypothesis in the GARCH(1,1) model. *J. Statist. Plann. Inference*, **139**(2), 593–616.

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

Troster, V. and Wied, D. (2021). A specification test for dynamic conditional distribution models with function-valued parameters. *Econometric Rev.*, **40**(2), 109–127.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York. With Applications to Statistics.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**(1), 1–25.