Supplementary material for "Specification tests for GARCH processes with nuisance parameters on the boundary"

Giuseppe Cavaliere
Department of Economics, University of Exeter, UK and
Department of Economics, University of Bologna, Italy
and
Indeewara Perera
Department of Economics, The University of Sheffield, UK
and
Anders Rahbek*
Department of Economics, University of Copenhagen, Denmark

January 14, 2023

Abstract

This supplementary material provides the proofs of the theoretical results stated in the paper "Specification tests for GARCH processes with nuisance parameters on the boundary", as well as some auxiliary lemmas and additional simulation results.

*Corresponding author: Anders Rahbek; email: anders.rahbek@econ.ku.dk.

S.1 Some preliminary results

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

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

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

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

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

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

The next lemma shows that $\check{F}_n$ defined by $\check{F}_n(x) := n^{-1}\sum_{t=1}^n \mathbb{I}(\check{\varepsilon}_t \le x),\ x \in \mathbb{R}$, in (12) converges to F_0 with probability 1.

Lemma S.1. (a) *Suppose that Assumptions (A1)–(A4) and* H_0 *hold, and* $\boldsymbol{\phi}_0$ *is an interior point in* Φ. *Then,* $d_2(\check{F}_n, F_0) \overset{a.s.}{\to} 0$ *as* $n \to \infty$. (b) *Additionally, assume that Assumption (A6) is also satisfied, then* $d_2(\check{F}_n, F_0) \overset{a.s.}{\to} 0$ *as* $n \to \infty$*, irrespective of whether* $\boldsymbol{\phi}_0$ *is in the interior of* Φ.

Proof of Lemma S.1. Recall that $d_2(F_X, F_Y)$ is the Mallows metric for the distance between two probability distributions F_X and F_Y defined by $d_2(F_X, F_Y) = \inf\{\mathrm{E}|X - Y|^2\}^{1/2}$, where the infimum is over all square integrable random variables X and Y with marginal distributions F_X and F_Y.

Proof of Part (a): Let $H_n(x) := n^{-1}\sum_{t=1}^n \mathbb{I}(\varepsilon_t \le x)$, be the e.d.f. of the unobserved errors $\{\varepsilon_1, \ldots, \varepsilon_n\}$. From the triangular inequality

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

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

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

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

$$\{d_2(\check{F}_n, H_n)\}^2 = \inf\{\mathrm{E}|X - Y|^2\} \leq \mathrm{E}\{X^{(1)} - Y^{(1)}\}^2$$
$$= n^{-1}\sum_{t=1}^{n}(\varepsilon_t - \check{\varepsilon}_t)^2 = (n\hat{\sigma}_n^2)^{-1}\sum_{t=1}^{n}\{\hat{\sigma}_n\varepsilon_t - (\widehat{\varepsilon}_t - n^{-1}\sum_{j=1}^{n}\widehat{\varepsilon}_j)\}^2, \tag{S.1}$$

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

Since Assumptions (A1)–(A4) are satisfied, and $\boldsymbol{\phi}_0$ is an interior point of Φ, QMLE $\hat{\boldsymbol{\phi}}_n$ is asymptotically linear and satisfies (10), and hence $\hat{\boldsymbol{\phi}}_n \stackrel{a.s.}{\to} \boldsymbol{\phi}_0$ and $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$, and it follows that $\hat{\sigma}_n^2 \stackrel{a.s.}{\to} 1$. Hence, for some constant $K > 0$, (S.1) is bounded from above by

$$Kn^{-1}\sum_{t=1}^{n}(\widehat{\varepsilon}_t - \varepsilon_t)^2 + Kn^{-2}\left(\sum_{t=1}^{n}\varepsilon_t\right)^2 + M_n$$

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

Since $h_t(\boldsymbol{\phi}) > \omega_L > 0$ for all $\boldsymbol{\phi} \in \Phi$, we have that

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}(\widehat{\varepsilon}_t - \varepsilon_t)^2 &= n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\big[\{h_t(\boldsymbol{\phi}_0)\}^{1/2} - \{h_t(\hat{\boldsymbol{\phi}}_n)\}^{1/2}\big]^2/h_t(\hat{\boldsymbol{\phi}}_n) \\ &\leq \omega_L^{-1}n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\big[\{h_t(\boldsymbol{\phi}_0)\}^{1/2} - \{h_t(\hat{\boldsymbol{\phi}}_n)\}^{1/2}\big]^2. \end{aligned}$$

From Proposition S.1 below, we obtain that

$$\{h_t(\hat{\boldsymbol{\phi}}_n)\}^{1/2} - \{h_t(\boldsymbol{\phi}_0)\}^{1/2} = 2^{-1}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\dot{h}_t(\boldsymbol{\phi}_0)/\{h_t(\boldsymbol{\phi}_0)\}^{1/2} + o_p(n^{-1/2}). \tag{S.2}$$

Because $\hat{\boldsymbol{\phi}}_n \stackrel{a.s.}{\to} \boldsymbol{\phi}_0$ and $\|\mathrm{E}\varepsilon_t^2\dot{h}_t(\boldsymbol{\phi}_0)/\{h_t(\boldsymbol{\phi}_0)\}^{1/2}\| < \infty$, then by the Ergodic Theorem

$$n^{-1}\sum_{t=1}^{n}\varepsilon_t^2[\{h_t(\hat{\boldsymbol{\phi}}_n)\}^{1/2} - \{h_t(\boldsymbol{\phi}_0)\}^{1/2}] \stackrel{a.s.}{\to} 0 \text{ as } n \to \infty.$$

Hence $n^{-1}\sum_{t=1}^{n}(\widehat{\varepsilon}_t - \varepsilon_t)^2 \stackrel{a.s.}{\to} 0$. Since $\mathrm{E}\varepsilon_t = 0$, by the strong law of large numbers we also have $n^{-2}\left(\sum_{t=1}^{n}\varepsilon_t\right)^2 \stackrel{a.s.}{\to} 0$ as $n \to \infty$. Therefore, $d_2(\check{F}_n, H_n) \stackrel{a.s.}{\to} 0$ and hence $d_2(\check{F}_n, F_0) \stackrel{a.s.}{\to} 0$.

Proof of part (b): Under Assumptions (A1)–(A3), $\hat{\boldsymbol{\phi}}_n$ converges to $\boldsymbol{\phi}_0$ (a.s.), irrespective of whether $\boldsymbol{\phi}_0$ is in the interior of the parameter space (see Lemma 2). Since Assumptions (A4)–(A6) are also satisfied, we also have $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$ by Lemma 2. Furthermore, under Assumptions (A1)–(A6), from Proposition S.1 below, (S.2) continues to hold irrespective of whether $\boldsymbol{\phi}_0$ is in the interior of Φ. Since $\hat{\boldsymbol{\phi}}_n \stackrel{a.s.}{\to} \boldsymbol{\phi}_0$, and $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$, then it follows that $d_2(\check{F}_n, F_0) \stackrel{a.s.}{\to} 0$ by repeating the arguments of the proof of part (a). □

For $(\boldsymbol{\phi}, F) = (\boldsymbol{\phi}_0, F_0)$ the data generating model (19) in the main paper is equivalent to the DGP defined by (1)–(4). Usually, $\boldsymbol{\phi}_0$ and F_0 are unknown. Hence, in order to generate data from a model that mimics (1)–(4), one needs to replace $(\boldsymbol{\phi}_0, F_0)$ by some known $(\boldsymbol{\phi}_n, F_n)$ which is sufficiently close to $(\boldsymbol{\phi}_0, F_0)$. Let $(\boldsymbol{\phi}_n, F_n)$ be such a sequence in the product space $\bar{\Phi}^* \times \mathcal{F}$, such that $(\boldsymbol{\phi}_n, F_n) \to (\boldsymbol{\phi}_0^*, F_0^*)$ as $n \to \infty$, with $\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_0^*\| \to 0$ and $d_2(F_n, F_0^*) \to 0$ as $n \to \infty$. Note that, since $(\boldsymbol{\phi}_0^*, F_0^*) = \operatorname{plim}(\hat{\boldsymbol{\phi}}_n, \check{F}_n)$, we have $(\boldsymbol{\phi}_0^*, F_0^*) = (\boldsymbol{\phi}_0, F_0)$ under H_0, and $(\boldsymbol{\phi}_0^*, F_0^*)$ is the pseudo-true value under H_1.

In what follows, when the DGP (19) corresponds to $(\boldsymbol{\phi}_n, F_n)$ instead of using $h_t^{(\phi_n, F_n)}(\cdot)$ and $\tau_t^{(\phi_n, F_n)}(\cdot)$, we let the analogs of $h_t(\cdot)$ and $\tau_t(\cdot)$ be denoted by $h_{nt}(\cdot)$ and $\tau_{nt}(\cdot)$, respectively. Note that, under H_0, the probability laws of $h_t^{(\phi_0, F_0)}(\cdot)$ and $\tau_t^{(\phi_0, F_0)}(\cdot)$ are identical to those of $h_t(\cdot)$ and $\tau_t(\cdot)$, respectively.

Recall that $\mathcal{H}_t$ denotes the information available up to time t, $t \in \mathbb{Z}$.

Lemma S.2. *If $\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_0^*\| \to 0$ and $d_2(F_n, F_0^*) \to 0$ as $n \to \infty$ with $\mathrm{E}(|F_n{}^{-1}(U_t)|^{4+d}) < \infty$, then for every $y \in \bar{\mathbb{R}}$, we have that $\mathrm{E}\mathbb{I}(Y_1^{(\phi_n, F_n)} \leq y) \to \mathrm{E}\mathbb{I}(Y_1^{(\phi_0^*, F_0^*)} \leq y)$ as $n \to \infty$.*

Proof. Fix $y \in \bar{\mathbb{R}}$. Since $Y_t^{(\phi_n, F_n)} = h_{nt}(\boldsymbol{\phi}_n) F_n^{-1}(U_t)$, $F_n^{-1}(U_t) := \inf\{y \in \mathbb{R} : F_n(y) \geq U_t\}$, and $\{U_t, t \in \mathbb{Z}\}$ are i.i.d. uniform(0,1), $\mathrm{E}\mathbb{I}(Y_1^{(\phi_n, F_n)} \leq y) - \mathrm{E}\mathbb{I}(Y_1^{(\phi_0^*, F_0^*)} \leq y)$ can be written as

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{E}\left\{\mathbb{I}\left(F_n^{-1}(U_t) \leq y/h_{nt}(\boldsymbol{\phi}_n)\right) |\mathcal{H}_{t-1}\right\}\right] - \mathrm{E}\left[\mathrm{E}\left\{\mathbb{I}\left(F_0^{*-1}(U_t) \leq y/h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\right) |\mathcal{H}_{t-1}\right\}\right] \\
&= \mathrm{E}\left[F_n\{y/h_{nt}(\boldsymbol{\phi}_n)\} - F_0^*\{y/h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\}\right] = \mathrm{E}\Delta_{n1} + \mathrm{E}\Delta_{n2}, \quad \text{where}
\end{aligned}$$

$$\Delta_{n1} = F_n[y/h_{nt}(\boldsymbol{\phi}_n)] - F_n[y/h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)], \quad \Delta_{n2} = F_n[y/h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)] - F_0^*[y/h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)].$$

Hence, to complete the proof it is sufficient to show that $\mathrm{E}\Delta_{n1} = o(1)$ and $\mathrm{E}\Delta_{n2} = o(1)$.

The recursion relation (19) yields the following Volterra series expansion of $h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)$:

$$h_t^{(\phi_0^*, F_0^*)}(\boldsymbol{\phi}_0^*) = b_0^* + b_0^* \sum_{k=1}^{\infty} \sum_{-\infty < s_k < \ldots < s_1 < t}^{\infty} b_{t-s_1}^* b_{s_1-s_2}^* \ldots b_{s_{k-1}-s_k}^* [F_0^{*-1}(U_{s_1})]^2 \ldots [F_0^{*-1}(U_{s_k})]^2,$$

where $b_0^* = \omega_0^*/(1 - \sum_{j=1}^{p_2} \beta_{0j}^*)$ and the weights $b_j^*, j \geq 1$, are defined by the generating function $\mathcal{A}_{\phi_0^*}(z)/\mathcal{B}_{\phi_0^*}(z) = \sum_{j=1}^{\infty} b_j^* z^j$. Recall that $\boldsymbol{\phi}_0^* = (\omega_0^*, \alpha_{01}^*, \ldots, \alpha_{0p_1}^*, \beta_{01}^*, \ldots, \beta_{0p_2}^*)'$, $\mathcal{A}_{\phi_0^*}(z) = \sum_{j=1}^{p_1} \alpha_{0j}^* z^j$ and $\mathcal{B}_{\phi_0^*}(z) = 1 - \sum_{j=1}^{p_2} \beta_{0j}^* z^j$.

Similarly, with $\boldsymbol{\phi}_n = (\omega_n, \alpha_{n1}, \ldots, \alpha_{np_1}, \beta_{n1}, \ldots, \beta_{np_2})'$, $b_0^{(n)} = \omega_n/(1 - \sum_{j=1}^{p_2} \beta_{nj})$ and $b_j^{(n)}, j \geq 1$ defined by the generating function $\mathcal{A}_{\phi_n}(z)/\mathcal{B}_{\phi_n}(z) = \sum_{j=1}^{\infty} b_j^{(n)} z^j$, one obtains

$$h_{nt}(\boldsymbol{\phi}_n) = b_0^{(n)} + b_0^{(n)} \sum_{k=1}^{\infty} \sum_{-\infty < s_k < \ldots < s_1 < t}^{\infty} b_{t-s_1}^{(n)} b_{s_1-s_2}^{(n)} \ldots b_{s_{k-1}-s_k}^{(n)} [F_n{}^{-1}(U_{s_1})]^2 \ldots [F_n{}^{-1}(U_{s_k})]^2.$$

Write $h_{nt}(\boldsymbol{\phi}_n) - h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*) = (b_0^{(n)} - b_0^*) + \sum_{k=1}^{\infty} \sum_{-\infty<s_k<\ldots<s_1<t}^{\infty} \{D_{n1}^* + D_{n2}^*\}$, where

$$\begin{aligned} D_{n1}^* &= (b_0^{(n)} b_{t-s_1}^{(n)} \ldots b_{s_{k-1}-s_k}^{(n)} - b_0^* b_{t-s_1}^* \ldots b_{s_{k-1}-s_k}^*)[F_n{}^{-1}(U_{s_1})]^2 \ldots [F_n{}^{-1}(U_{s_k})]^2, \\ D_{n2}^* &= b_0^* b_{t-s_1}^* \ldots b_{s_{k-1}-s_k}^* \{[F_n{}^{-1}(U_{s_1})]^2 \ldots [F_n{}^{-1}(U_{s_k})]^2 - [F_0^{*-1}(U_{s_1})]^2 \ldots [F_0^{*-1}(U_{s_k})]^2\}. \end{aligned}$$

Since $\boldsymbol{\phi}_n \to \boldsymbol{\phi}_0^*$, $\mathrm{E}(|F_n{}^{-1}(U_t)|^{4+d}) < \infty$ and the weights $b_j^*, b_j^{(n)}, j \geq 1$, are exponentially decaying, we have that $\sum_{k=1}^{\infty} \sum_{-\infty<s_k<\ldots<s_1<t}^{\infty} D_{n1}^* \stackrel{a.s.}{\to} 0$. Since $d_2(F_n, F_0^*) \to 0$ and $b_j^*, j \geq 1$, are exponentially decaying, it also follows that $\sum_{k=1}^{\infty} \sum_{-\infty<s_k<\ldots<s_1<t}^{\infty} D_{n2}^* \stackrel{a.s.}{\to} 0$. Therefore,

$$|h_{nt}(\boldsymbol{\phi}_n) - h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \leq |b_0^{(n)} - b_0^*| + \left|\sum_{k=1}^{\infty} \sum_{-\infty<s_k<\ldots<s_1<t}^{\infty} (D_{n1}^* + D_{n2}^*)\right| \stackrel{a.s.}{\to} 0.$$

In what follows $C > 0$ denotes a generic constant, which may take different values at different places. Since $\sup_{\zeta \in D} \|h_t^{(\zeta)}(\cdot)\|_{\bar{\Phi}^*} > \omega_L > 0$ and $|h_{nt}(\boldsymbol{\phi}_n) - h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \stackrel{a.s.}{\to} 0$ we obtain that $|\Delta_{n1}| \leq y \omega_L^{-2} C |h_{nt}(\boldsymbol{\phi}_n) - h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \stackrel{a.s.}{\to} 0$, and hence $\mathrm{E}\Delta_{n1} = o(1)$. Similarly, since $d_2(F_n, F_0^*) \to 0$, $\mathrm{E}|\Delta_{n2}| \leq C d_2(F_n, F_0^*) \to 0$, and hence $\mathrm{E}\Delta_{n2} = o(1)$. □

Recall that $\tau_{nt}(\boldsymbol{\phi}) := \dot{h}_{nt}(\boldsymbol{\phi}) / h_{nt}(\boldsymbol{\phi})$ where $h_{nt}(\boldsymbol{\phi}) = h_t^{(\boldsymbol{\phi}_n, F_n)}(\boldsymbol{\phi})$.

Lemma S.3. *For every nonrandom sequence* $\zeta_n := (\boldsymbol{\phi}_n, F_n) \in \bar{\Phi}^* \times \mathcal{F}$ *with* $\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_0^*\| \to 0$ *and* $d_2(F_n, F_0^*) \to 0$, *where* $\mathrm{E}(|F_n{}^{-1}(U_t)|^{4+d}) < \infty$, *we have that* $\mathrm{E}\|\tau_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\|^2 < \infty$ *and* $\mathrm{E}\|\ddot{h}_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\{h_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\}^{-1/2}\|^2 < \infty$. *Furthermore,* $\mathrm{E}[\tau_{n1}(\boldsymbol{\phi}_n)] \to \mathrm{E}[\tau_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)]$ *and* $\mathrm{E}[\ddot{h}_{n1}(\boldsymbol{\phi}_n)\{h_{n1}(\boldsymbol{\phi}_n)\}^{-1/2}] \to \mathrm{E}[\ddot{h}_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\{h_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\}^{-1/2}]$ *as* $n \to \infty$.

Proof. Since $h_{nt}(\boldsymbol{\phi}_n), h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*) > \omega_L > 0$,

$$\begin{aligned} &|\tau_{nt}(\boldsymbol{\phi}_n) - \tau_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \\ &\quad= \left| \dot{h}_{nt}(\boldsymbol{\phi}_n) \left(\frac{1}{h_{nt}(\boldsymbol{\phi}_n)} - \frac{1}{h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)} \right) + \frac{\dot{h}_{nt}(\boldsymbol{\phi}_n) - \dot{h}_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)}{h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)} \right| \\ &\quad\leq \|\tau_{n1}(\boldsymbol{\phi}_n)\| \omega_L^{-1} |h_{nt}(\boldsymbol{\phi}_n) - h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| + \omega_L^{-1} \|\dot{h}_{nt}(\boldsymbol{\phi}_n) - \dot{h}_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\|. \end{aligned} \tag{S.3}$$

From the proof of Lemma S.2, $|h_{nt}(\boldsymbol{\phi}_n) - h_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \stackrel{a.s.}{\to} 0$. By arguing as in the proof of Proposition S3 of Perera and Silvapulle (2022), $\|\dot{h}_{nt}(\boldsymbol{\phi}_n) - \dot{h}_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \stackrel{a.s.}{\to} 0$. Furthermore, from Propositions S3 and S4 of Perera and Silvapulle (2022), $\mathrm{E}\|\tau_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\|^2 < \infty$ and $\mathrm{E}\|\ddot{h}_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\{h_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\}^{-1/2}\|^2 < \infty$. Since $\mathrm{E}\|\tau_{n1}(\boldsymbol{\phi}_n)\|^2 < \infty$, by the above two convergence results and the inequality (S.3), we obtain that $|\tau_{nt}(\boldsymbol{\phi}_n) - \tau_t^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)| \stackrel{a.s.}{\to} 0$, and hence $\mathrm{E}[\tau_{n1}(\boldsymbol{\phi}_n)] \to \mathrm{E}[\tau_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)]$. By using similar arguments we also obtain that $\mathrm{E}[\ddot{h}_{n1}(\boldsymbol{\phi}_n)\{h_{n1}(\boldsymbol{\phi}_n)\}^{-1/2}] \to \mathrm{E}[\ddot{h}_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\{h_1^{(\boldsymbol{\phi}_0^*, F_0^*)}(\boldsymbol{\phi}_0^*)\}^{-1/2}]$. □

Lemma S.4. *Suppose that either (a) Assumptions (A1)–(A3) and* H_0 *are satisfied and* $\{Y_t\}_{t\in\mathbb{Z}}$ *is square integrable, or (b) Assumptions (B1) and (B2) are satisfied and* H_1 *holds. Then, for every* $\zeta = (\boldsymbol{\phi}, F) \in \bar{\Phi}^* \times \mathcal{F}$*, the model (19) has a unique stationary ergodic solution* $\{Y_t^{(\zeta)} : t \in \mathbb{Z}\}$ *with finite second moments.*

Proof. Let $\zeta = (\boldsymbol{\phi}, F) \in \bar{\Phi}^* \times \mathcal{F}$ be fixed and arbitrary. Since F has zero mean and unit variance, the condition $\sum_{i=1}^{p_1} \alpha_i + \sum_{j=1}^{p_2} \beta_j < 1$ is necessary and sufficient for the process $\{Y_t^{(\zeta)}; t \in \mathbb{Z}\}$ to be strictly stationary and have finite second moments with $\mathrm{E}(Y_t^{(\zeta)}) = 0$ and $\mathrm{E}[\{Y_t^{(\zeta)}\}^2] = \omega/(1 - \sum_{i=1}^{p_1} \alpha_i - \sum_{j=1}^{p_2} \beta_j)$; see, e.g., Nelson (1990); Chen and An (1998). □

Lemma S.5. *Suppose that the assumptions of Lemma S.4 are satisfied and* $\zeta_n := (\boldsymbol{\phi}_n, F_n) \to \zeta_0^* := (\boldsymbol{\phi}_0^*, F_0^*)$*, with* $\mathrm{E}(|F_n{}^{-1}(U_t)|^{4+d}) < \infty$*,* $\|\boldsymbol{\phi}_n - \boldsymbol{\phi}_0^*\| \to 0$ *and* $d_2(F_n, F_0^*) \to 0$*, where* $\zeta_n \in \bar{\Phi}^* \times \mathcal{F}$*. Then, for every constant* $C < \infty$*,*

$$\sup |h_{nt}^{1/2}(\boldsymbol{b}) - h_{nt}^{1/2}(\boldsymbol{a}) - 2^{-1}(\boldsymbol{b} - \boldsymbol{a})'\dot{h}_{nt}(\boldsymbol{a})h_{nt}^{-1/2}(\boldsymbol{a})|h_{nt}^{-1/2}(\boldsymbol{a}) = o_p(n^{-1/2}),$$

where the supremum is taken over $1 \le t \le n$ *and over* $\{(\boldsymbol{b}, \boldsymbol{a}) : \boldsymbol{b}, \boldsymbol{a} \in \bar{\Phi}^*, n^{1/2}\|\boldsymbol{b} - \boldsymbol{a}\| \le C\}$*.*

Proof of Lemma S.5. Let $D := \bar{\Phi}^* \times \mathcal{F}$, $p = p_1 + p_2 + 1$. Let $\wp_{nt}(\boldsymbol{\phi}) = \{h_{nt}(\boldsymbol{\phi})\}^{1/2}$ for $\boldsymbol{\phi} \in \bar{\Phi}^*$. Let $\Delta_{nt}(\boldsymbol{a}, \boldsymbol{b}) := h_{nt}^{1/2}(\boldsymbol{b}) - h_{nt}^{1/2}(\boldsymbol{a}) - 2^{-1}(\boldsymbol{b} - \boldsymbol{a})'\dot{h}_{nt}(\boldsymbol{a})h_{nt}^{-1/2}(\boldsymbol{a})$. Let $\boldsymbol{a}, \boldsymbol{b} \in \bar{\Phi}^*$ be fixed but arbitrary. Then, for each $n \in \mathbb{N}$, there exists $\boldsymbol{\delta}_{n1} \in \mathbb{R}^p$, such that $\boldsymbol{b} = \boldsymbol{a} + n^{-1/2}\boldsymbol{\delta}_{n1}$. Hence, $\Delta_{nt}(\boldsymbol{a}, \boldsymbol{b}) = \wp_{nt}(\boldsymbol{b}) - \wp_{nt}(\boldsymbol{a}) - n^{-1/2}\boldsymbol{\delta}_{n1}'\dot{\wp}_{nt}(\boldsymbol{a})$. Therefore, by the Mean Value Theorems for functions from $\mathbb{R}^p$ to $\mathbb{R}$, and $\mathbb{R}^p$ to $\mathbb{R}^p$, with right partial derivatives, for every $n \in \mathbb{N}$, there exist $\boldsymbol{\delta}_{n2}, \boldsymbol{\delta}_{n3} \in \mathbb{R}^p$ with $\|\boldsymbol{\delta}_{n3}\| \le \|\boldsymbol{\delta}_{n2}\| \le \|\boldsymbol{\delta}_{n1}\|$, such that

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

where

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

Since $\sup_{\zeta \in D} \|h_t^{(\zeta)}(\cdot)\|_{\bar{\Phi}^*} > \omega_L > 0$, for any given constant $C > 0$, w.p. 1,

$$\max_{1 \le t \le n} \sup_{\boldsymbol{a}, \boldsymbol{b} \in \bar{\Phi}^*, n^{1/2}\|\boldsymbol{b} - \boldsymbol{a}\| \le C} n^{1/2}|\Delta_{nt}(\boldsymbol{a}, \boldsymbol{b})|h_{nt}^{-1/2}(\boldsymbol{a}) \le \omega_L^{-1/2} n^{-1/2} C^2 \max_{1 \le t \le n} \|\ddot{\wp}_{nt}\|_{\bar{\Phi}^*}.$$

Since $n^{-1/2}\max_{1 \le t \le n} \|\ddot{\wp}_{nt}\|_{\bar{\Phi}^*} = o_p(1)$, see, e.g. proof of Lemma 4 in Perera and Silvapulle (2021), it follows that $\max_{1 \le t \le n} \sup_{\boldsymbol{a}, \boldsymbol{b} \in \bar{\Phi}^*, n^{1/2}\|\boldsymbol{b} - \boldsymbol{a}\| \le C} |\Delta_{nt}(\boldsymbol{a}, \boldsymbol{b})|h_{nt}^{-1/2}(\boldsymbol{a}) = o_p(n^{-1/2})$. □

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

Lemma S.6. *Let $\{(a_i, b_i); 1 \leq i \leq n\}$ be i.i.d. square-integrable bivariate random vectors with* $\mathrm{E}(a_i) = \mathrm{E}(b_i) = 0$, $1 \leq i \leq n$. *Then we have that* $\mathrm{E}\{(\sum_{t=1}^{n} a_i)^2(\sum_{j=1}^{n} b_j)^2\} \leq n\mathrm{E}(a_1^2 b_1^2) + 3n(n-1)\mathrm{E}(a_1^2)\mathrm{E}(b_1^2)$.

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

Proposition S.1. *Suppose that the assumptions of Lemma 1 hold. Then, for every $K < \infty$,* $n^{1/2} \sup \mid h_t^{1/2}(\boldsymbol{b}) - h_t^{1/2}(\boldsymbol{a}) - 2^{-1}(\boldsymbol{b}-\boldsymbol{a})'\dot{h}_t(\boldsymbol{a})h_t^{-1/2}(\boldsymbol{a}) \mid h_t^{-1/2}(\boldsymbol{\phi}_0) = o_p(1)$, *where the supremum is taken over $1 \leq t \leq n$ and over $\{(\boldsymbol{b}, \boldsymbol{a}) : \boldsymbol{b}, \boldsymbol{a} \in \varPhi, n^{1/2}\|\boldsymbol{b}-\boldsymbol{a}\| \leq K\}$.*

Proof. The proof follows as a special case of the proof of Lemma S.5. □

S.2 Initialization effect

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

If data are generated from (19) for $t \geq -m$, conditional on a vector of starting values $\varsigma_0 = (y_0, \cdots, y_{1-q}, s_0, \cdots, s_{1-p})'$, then we use the superscript "(m, ζ)" instead of "(ζ)", where $\zeta = (\boldsymbol{\phi}, F)$. For example, $h_t^{(m,\zeta)}(\cdot)$ and $\tau_t^{(m,\zeta)}(\cdot)$ are the analogues of $h_t^{(\zeta)}(\cdot)$ and $\tau_t^{(\zeta)}(\cdot)$, respectively, when the data generating model obeys (19) for $t \geq -m$, conditional on the starting values ς_0.

Lemma S.7. *Suppose that the assumptions of Lemma S.4 are satisfied. Then, there exists a compact set $K_1 \subseteq \bar{\Phi}^*$, which contains $\boldsymbol{\phi}_0^*$, such that* $\sup_{\zeta \in K} \|h_t^{(m,\zeta)} - h_t^{(\zeta)}\|_{K_1} \overset{e.a.s.}{\to} 0$ *and* $\sup_{\zeta \in K} \|\dot{h}_t^{(m,\zeta)} - \dot{h}_t^{(\zeta)}\|_{K_1} \overset{e.a.s.}{\to} 0$ *as $t \to \infty$, where $K = K_1 \times \mathcal{F}$.*

Proof. The proof follows from the verifications of Conditions (M1) and (M2) in Perera and Silvapulle (2022) for the GARCH(p_1, p_2) setup, with the set $K_{\boldsymbol{\theta}}$ being replaced by $\mathcal{F}$. Sine everything follows after suitable modifications of the arguments already developed in Perera and Silvapulle (2022), we omit the details. □

Conditional on $(Y_1, \ldots, Y_n)$, and using the notation introduced above, we have

$$\{Y_t^*, h_t^*(\boldsymbol{\phi}), \tau_t^*(\boldsymbol{\phi})\} \equiv \{Y_t^{(0,\zeta^*)}, h_t^{(0,\zeta^*)}(\boldsymbol{\phi}), \tau_t^{(0,\zeta^*)}(\boldsymbol{\phi})\}, \quad \boldsymbol{\phi} \in \Phi, t \in \mathbb{N}, \tag{S.4}$$

where $\zeta^* = (\hat{\boldsymbol{\phi}}_n, \check{F}_n)$ for the bootstrap method in Section 3.2, and $\zeta^* = (\hat{\boldsymbol{\phi}}_n^\dagger, \check{F}_n)$ for the bootstrap method in Section 4.2 with $\hat{\boldsymbol{\phi}}_n^\dagger$ given by (22). Similarly, let the bootstrap process generated by (19), conditional on $(Y_1, \ldots, Y_n)$, without any initialization, be defined as

$$\{Y_t^{*(\infty)}, h_t^{*(\infty)}(\boldsymbol{\phi}), \tau_t^{*(\infty)}(\boldsymbol{\phi})\} = \{Y_t^{(\zeta^*)}, h_t^{(\zeta^*)}(\boldsymbol{\phi}), \tau_t^{(\zeta^*)}(\boldsymbol{\phi})\}, \ \boldsymbol{\phi} \in \Phi, t \in \mathbb{Z}, \tag{S.5}$$

where $\zeta^* = (\hat{\boldsymbol{\phi}}_n, \check{F}_n)$ and $\zeta^* = (\hat{\boldsymbol{\phi}}_n^\dagger, \check{F}_n)$ for the bootstraps in Section 3.2 and Section 4.2, respectively. This notation is similar to the one used in Section 3.3 in Perera and Silvapulle (2018) to define a similar hypothetical bootstrap process. Thus, $\{Y_t^{*(\infty)}, h_t^{*(\infty)}(\boldsymbol{\phi}), \tau_t^{*(\infty)}(\boldsymbol{\phi})\}$ represents a hypothetical (non-operational) bootstrap process generated by (19), without any initialization. Further, let

$$\hat{\boldsymbol{\phi}}_n^{*(\infty)} = \arg\min_{\boldsymbol{\phi} \in \Phi} \sum_{t=1}^n \ell_t^{*(\infty)}(\boldsymbol{\phi}), \quad \ell_t^{*(\infty)}(\boldsymbol{\phi}) = \log h_t^{*(\infty)}(\boldsymbol{\phi}) + \frac{\{Y_t^{*(\infty)}\}^2}{h_t^{*(\infty)}(\boldsymbol{\phi})},$$

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

Since no initialization is used in the bootstrap data generation in (S.5), the marked empirical process $\mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}_n^{*(\infty)})$, unlike $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*)$, is not subject to any initialization error.

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

Lemma S.8. *Suppose that the assumptions of Theorem 2 are satisfied with $\boldsymbol{\phi}_0$ being an interior point in Φ, or the assumptions of Theorem 3 are satisfied. Then, conditional on $(Y_1, \ldots, Y_n)$, $\sup_{y \in \mathbb{R}} | \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}_n^{*(\infty)}) | = o_p^*(1)$, in probability.*

Proof. Fix a sample path along which $\hat{\boldsymbol{\phi}}_n \to \boldsymbol{\phi}_0$ and $d_2(\check{F}_n, F_0) \to 0$ (this approach is similar to the one used in the proof of Lemma 7 in Perera and Silvapulle (2022)). Assume without loss of generality that $Y_t^* \leq Y_t^{*(\infty)}$. Then, we have

$$\begin{aligned} & \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}_n^{*(\infty)}) \\ = \;& n^{-1/2} \sum_{t=1}^n \left(\frac{\{Y_t^*\}^2}{h_t^*(\hat{\boldsymbol{\phi}}_n^*)} - \frac{\{Y_t^{*(\infty)}\}^2}{h_t^{*(\infty)}(\hat{\boldsymbol{\phi}}_n^{*(\infty)})} \right) \mathbb{I}(Y_{t-1}^{*(\infty)} \leq y) \\ & + n^{-1/2} \sum_{t=1}^n \left(\frac{\{Y_t^*\}^2}{h_t^*(\hat{\boldsymbol{\phi}}_n^*)} - 1 \right) \mathbb{I}(Y_{t-1}^* \leq y < Y_{t-1}^{*(\infty)}) \\ = \;& \mathbf{I} + \mathbf{II}, \quad \text{say.} \end{aligned} \tag{S.6}$$

Since Lemma S.7 shows that $|h_t^*(\hat{\boldsymbol{\phi}}_n) - h_t^{*(\infty)}(\hat{\boldsymbol{\phi}}_n)| \stackrel{e.a.s.}{\to} 0$ as $t \to \infty$, from Lemma 2.1 in Straumann and Mikosch (2006), $|\{h_{t-1}^*(\hat{\boldsymbol{\phi}}_n)\}^{1/2} - \{h_{t-1}^{*(\infty)}(\hat{\boldsymbol{\phi}}_n)\}^{1/2}| \stackrel{e.a.s.}{\to} 0$, as $t \to \infty$, and hence from Lemma 2.3 in Straumann and Mikosch (2006), it follows that

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

and hence the sum $\mathbf{II}$ in (S.6) is of order $o_p^*(1)$, uniformly in $y \in \mathbb{R}$, along the fixed sample path.

The first sum in (S.6) is bounded as

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

Since $|\{Y_t^*\}^2 - \{Y_t^{*(\infty)}\}^2| = |\varepsilon_t^*|^2|h_t^*(\hat{\boldsymbol{\phi}}_n) - h_t^{*(\infty)}(\hat{\boldsymbol{\phi}}_n)| \stackrel{e.a.s.}{\to} 0$, as $t \to \infty$, $\sum_{t=1}^n |\{Y_t^*\}^2 - \{Y_t^{*(\infty)}\}^2|$ converges a.s. as $n \to \infty$, by Lemma 2.1 in Straumann and Mikosch (2006), and hence sum $\mathbf{I}_B$ is of order $o_p^*(1)$ along the fixed sample path.

From the proof of Lemma 8 in Perera and Silvapulle (2017), $n^{1/2}(\hat{\boldsymbol{\phi}}_n^* - \hat{\boldsymbol{\phi}}_n^{*(\infty)}) = o_p^*(1)$. By Lemma S.7, $\sup_{\boldsymbol{\phi}\in K_1} |h_t^*(\boldsymbol{\phi}) - h_t^{*(\infty)}(\boldsymbol{\phi})| \stackrel{e.a.s.}{\to} 0$ for some compact set $K_1 \subseteq \bar{\Phi}^*$. Hence, Lemma 2.3 in Straumann and Mikosch (2006) yields that sum $\mathbf{I}_A$ is also of order $o_p^*(1)$, along the fixed sample path. Since $\hat{\boldsymbol{\phi}}_n \stackrel{a.s.}{\to} \boldsymbol{\phi}_0$ and $d_2(\check{F}_n, F_0) \stackrel{a.s.}{\to} 0$, as $n \to \infty$, it follows that $\sup_{y\in\mathbb{R}} | \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}_n^{*(\infty)}) |= o_p^*(1)$, in probability. □

The next lemma shows that, conditional on $(Y_1, \ldots, Y_n)$, for the bootstrap method in Section 4.2, $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*)$ and $\mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}_n^{*(\infty)})$ are uniformly close, in probability.

Lemma S.9. *Suppose that either assumptions of Theorem 4 or Theorem 5 are satisfied. Then, conditional on $\{Y_1, \ldots, Y_n\}$, $\sup_{y\in\mathbb{R}} | \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*) - \mathcal{U}_n^{*(\infty)}(y, \hat{\boldsymbol{\phi}}_n^{*(\infty)}) |= o_p^*(1)$, in probability.*

Proof. The proof follows from arguing as for the proof of Lemma S.8 with $\hat{\boldsymbol{\phi}}_n$ replaced by $\hat{\boldsymbol{\phi}}_n^\dagger$. □

Lemmas S.8 and S.9 show that the effect of initialization in the bootstrap data generation is asymptotically negligible. Hence, in the next section we only focus on $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}_n^*)$.

S.3 Main proofs

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

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

$$
\begin{aligned}
&\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n) - \mathcal{U}_n(y, \boldsymbol{\phi}_0) \\
&= n^{-1/2} \sum_{t=1}^{n} \{Y_t^2/h_t(\hat{\boldsymbol{\phi}}_n) - Y_t^2/h_t(\boldsymbol{\phi}_0))\} \mathbb{I}(Y_{t-1} \le y) \\
&= n^{-1/2} \sum_{t=1}^{n} \varepsilon_t^2 \{h_t(\boldsymbol{\phi}_0)/h_t(\hat{\boldsymbol{\phi}}_n) - 1\} \mathbb{I}(Y_{t-1} \le y) \\
&= -n^{-1/2} \sum_{t=1}^{n} \varepsilon_t^2 \left(\frac{h_t(\hat{\boldsymbol{\phi}}_n) - h_t(\boldsymbol{\phi}_0)}{h_t(\boldsymbol{\phi}_0)} \right) \mathbb{I}(Y_{t-1} \le y) \\
&\quad + n^{-1/2} \sum_{t=1}^{n} \varepsilon_t^2 \left(h_t(\hat{\boldsymbol{\phi}}_n) - h_t(\boldsymbol{\phi}_0) \right) \left(\frac{1}{h_t(\boldsymbol{\phi}_0)} - \frac{1}{h_t(\hat{\boldsymbol{\phi}}_n)} \right) \mathbb{I}(Y_{t-1} \le y).
\end{aligned} \tag{S.7}
$$

Since $\mathrm{E}(\varepsilon_0^2) = 1$ and $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$, by applying Proposition S.1 and the Ergodic Theorem to the expansion (S.7), we obtain that, uniformly in $y \in \mathbb{R}$,

$$
\begin{aligned}
\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n) &= \mathcal{U}_n(y, \boldsymbol{\phi}_0) - n^{-1} \sum_{t=1}^{n} \{\varepsilon_t^2 n^{1/2} (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)' \tau_t(\boldsymbol{\phi}_0)\} \mathbb{I}(Y_{t-1} \le y) + o_p(1) \\
&= \mathcal{U}_n(y, \boldsymbol{\phi}_0) - n^{1/2} (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)' \mathrm{E}[\tau_1(\boldsymbol{\phi}_0) \mathbb{I}(Y_0 \le y)] + o_p(1).
\end{aligned}
$$

□

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

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

$$
\begin{aligned}
\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n) &= \mathcal{U}_n(y, \boldsymbol{\phi}_0) \\
&\quad - \mathrm{E}[\tau_1'(\boldsymbol{\phi}_0) \mathbb{I}(Y_0 \le y)] \Sigma_n^{-1}(\boldsymbol{\phi}_0) n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t^2 - 1) \tau_t(\boldsymbol{\phi}_0) + o_p(1).
\end{aligned} \tag{S.8}
$$

By the Ergodic Theorem (e.g., Theorem 2.5.2 of Giraitis *et al.*, 2012) $\Sigma_n(\boldsymbol{\phi}_0) \stackrel{a.s.}{\to} \Sigma(\boldsymbol{\phi}_0)$, as

$n \to \infty$. Hence, by using the above asymptotic uniform expansion of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}}_n)$, we derive

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

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

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

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

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

$$\begin{aligned}
a_t &= \{\varepsilon_t^2 - 1\}\mathbb{I}(G^{-1}(u_1) < Y_{t-1} \le G^{-1}(u)), \\
b_t &= \{\varepsilon_i^2 - 1\}\mathbb{I}(G^{-1}(u) < Y_{t-1} \le G^{-1}(u_2)).
\end{aligned}$$

Note that $\mathrm{E}(a_t) = \mathrm{E}(b_t) = 0$ and $a_t b_t = 0$. Further, $\bar{\mathcal{U}}_n(u, \boldsymbol{\phi}_0) - \bar{\mathcal{U}}_n(u_1, \boldsymbol{\phi}_0) = n^{-1/2}\sum_{t=1}^n a_t$ and $\bar{\mathcal{U}}_n(u_2, \boldsymbol{\phi}_0) - \bar{\mathcal{U}}_n(u, \boldsymbol{\phi}_0) = n^{-1/2}\sum_{j=1}^n b_j$. Therefore, from Lemma S.6, we obtain that

$$\begin{aligned}
&\mathrm{E}[\{\bar{\mathcal{U}}_n(u, \boldsymbol{\phi}_0) - \bar{\mathcal{U}}_n(u_1, \boldsymbol{\phi}_0)\}\{\bar{\mathcal{U}}_n(u_2, \boldsymbol{\phi}_0) - \bar{\mathcal{U}}_n(u, \boldsymbol{\phi}_0)\}] \\
&= n^{-2}\mathrm{E}\{(\sum_{t=1}^{n} a_t)^2(\sum_{j=1}^{n} b_j)^2\} \le [n\mathrm{E}(a_1^2 b_1^2) + 3n(n-1)\mathrm{E}(a_1^2)\mathrm{E}(b_1^2)]/n^2 \\
&= 3n(n-1)n^{-2}[\mathrm{E}\{\varepsilon_1^2 - 1\}^2]^2(u - u_1)(u_2 - u) \\
&\le 3(\kappa_\varepsilon - 1)^2(u_2 - u_1)^2.
\end{aligned}$$

Since u_1 and u_2 are arbitrary, then it follows that $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ is $\mathcal{C}$-tight, e.g. by Theorem 15.7 of Billingsley (1968). Since G is continuous, the last term in (S.8),

$$\mathrm{E}[\tau_1'(\boldsymbol{\phi}_0)\mathbb{I}(Y_0 \le y)]c\Sigma_n^{-1}(\boldsymbol{\phi}_0)n^{-1/2}\textstyle\sum_{t=1}^n \tau_t(\boldsymbol{\phi}_0)\varrho(\varepsilon_t),$$

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

We next state the proof of Theorem 2. In view of Lemmas S.8 and S.9, and the continuous mapping theorem, the effect of initialization in the bootstrap data generation is asymptotically negligible. Therefore, in the next proof, and in the sequel, we do not distinguish between $\{Y_t^{*(\infty)}, h_t^{*(\infty)}(\boldsymbol{\phi}), \tau_t^{*(\infty)}(\boldsymbol{\phi})\}$ in (S.5), and the operational bootstrap process $\{Y_t^*, h_t^*(\boldsymbol{\phi}), \tau_t^*(\boldsymbol{\phi})\}$ in (S.4).

Note that, since $(\boldsymbol{\phi}_0^*, F_0^*) = \text{plim}(\hat{\boldsymbol{\phi}}_n, \check{F}_n)$, we have $(\boldsymbol{\phi}_0^*, F_0^*) = (\boldsymbol{\phi}_0, F_0)$ under H_0, and $(\boldsymbol{\phi}_0^*, F_0^*)$ is the pseudo-true value under H_1.

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

$$\begin{aligned}\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^*) &= \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n) - \frac{1}{n}\sum_{t=1}^n (\varepsilon_t^*)^2 n^{1/2}(\hat{\boldsymbol{\phi}}_n^* - \hat{\boldsymbol{\phi}}_n)'\tau_t^*(\hat{\boldsymbol{\phi}}_n)\mathbb{I}(Y_{t-1}^* \le y) + o_p^*(1), \\ &= \mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n) - \frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\boldsymbol{\phi}}_n^* - \hat{\boldsymbol{\phi}}_n)'\tau_t^*(\hat{\boldsymbol{\phi}}_n)\mathbb{I}(Y_{t-1}^* \le y) + o_p^*(1), \qquad \text{(S.9)}\end{aligned}$$

in probability. Since plim $\hat{\boldsymbol{\phi}}_n = \boldsymbol{\phi}_0$, and $\boldsymbol{\phi}_0$ is an interior point in Φ, one obtains

$$n^{1/2}(\hat{\boldsymbol{\phi}}_n^* - \hat{\boldsymbol{\phi}}_n) = -\Sigma_n^{*-1}(\hat{\boldsymbol{\phi}}_n) n^{-1/2}\sum_{t=1}^n (1 - \varepsilon_t^{*2})\tau_t^*(\hat{\boldsymbol{\phi}}_n) + o_p^*(1), \qquad \text{(S.10)}$$

$$\Sigma_n^*(\boldsymbol{\phi}) := n^{-1}\sum_{t=1}^n \tau_t^*(\boldsymbol{\phi})\tau_t^*(\boldsymbol{\phi})', \quad \tau_t^*(\boldsymbol{\phi}) := \dot{h}_t^*(\boldsymbol{\phi})/h_t^*(\boldsymbol{\phi}), \quad \boldsymbol{\phi} \in \Phi;$$

this follows from the proof of Proposition 3.2 in Hidalgo and Zaffaroni (2007), pp. 850-851. More precisely, $H_T^*(\widehat{\theta})$ and $T^{1/2}h_T^*$ in Hidalgo and Zaffaroni (2007) are the same as $2^{-1}n^{-1}\sum_{t=1}^n (\partial^2/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi})\ell_t^*(\hat{\boldsymbol{\phi}}_n)$ and $2^{-1}n^{-1/2}\sum_{t=1}^n (1 - \varepsilon_t^{*2})\tau_t^*(\hat{\boldsymbol{\phi}}_n)$, respectively, in our notation. Note that $\ell_t^*(\boldsymbol{\phi}) = \log h_t^*(\boldsymbol{\phi}) + [Y_t^{*2}/h_t^*(\boldsymbol{\phi})]$. By a routine application of Slutzky's theorem $\Sigma_n^{*-1}(\hat{\boldsymbol{\phi}}_n) - [n^{-1}\sum_{t=1}^n (\partial^2/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi})\ell_t^*(\hat{\boldsymbol{\phi}}_n)]^{-1} = o_p^*(1)$, and hence the arguments in Hidalgo and Zaffaroni (2007) leading to $T^{1/2}(\hat{\theta}^* - \hat{\theta}) = -H_T^{*-1}(\hat{\theta})T^{1/2}h_T^* + o_p^*(1)$ imply (S.10).

Since $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n) = n^{-1/2}\sum_{t=1}^n (\varepsilon_t^* - 1)^2\mathbb{I}(Y_{t-1}^* \le y)$, by using Lemma S.2, for every $x, y \in \mathbb{R}$, with $x \wedge y := \min(x, y)$, we derive that

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

Further, by substituting the above expansion for $n^{1/2}(\hat{\boldsymbol{\phi}}_n^* - \hat{\boldsymbol{\phi}}_n)$ in (S.9), and arguing as for Theorem 1 in a triangular array, we also obtain that

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

where

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

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

$$
T_1^* = \mathrm{KS}^* \xrightarrow{d^*} \sup_y \left|\mathcal{U}_0(y)\right|, \quad T_2^* = \mathrm{CvM}^* \xrightarrow{d^*} \int \mathcal{U}_0^2(y)dG(y),
$$

in probability, and hence Part 2 also holds for the KS and CvM functional forms. □

Proof of Theorem 3. If some components of $\boldsymbol{\phi}_0^*$ lie on the boundary of the parameter space; i.e., $\phi_{0i}^* = 0$ for some $i = 2, \ldots, p_1 + p_2 + 1$, then the proof follows from arguing as in the proof of Theorem 5. Hence, here we only consider the case $\boldsymbol{\phi}_0^*$ is in the interior of Φ. Since Assumptions (B1) and (B2) hold, and $(\hat{\boldsymbol{\phi}}_n, \check{F}_n) \xrightarrow{p} (\boldsymbol{\phi}_0^*, F_0^*)$, by applying Lemmas S.2 and S.3 and arguing as in the proof of Theorem 2, conditional on $\{Y_1, \ldots, Y_n\}$, the process $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}_n^*)$ converges weakly to the centred Gaussian process $\mathcal{U}_0^\dagger(\cdot)$ specified by

$$
\begin{aligned}
Cov\{\mathcal{U}_0^\dagger(x), \mathcal{U}_0^\dagger(y)\} &= \mathrm{E}(\{F_0^{*-1}(U_t)\}^2 - 1)^2\mathbb{I}(Y_{t-1} \le x \wedge y) \\
&\quad +J'(x, \boldsymbol{\phi}_0^*)\mathrm{E}[V_t(\boldsymbol{\phi}_0)V_t'(\boldsymbol{\phi}_0^*)]J'(y, \boldsymbol{\phi}_0^*) \\
&\quad -J'(x, \boldsymbol{\phi}_0^*)\mathrm{E}[(\{F_0^{*-1}(U_t)\}^2 - 1)V_t(\boldsymbol{\phi}_0^*)\mathbb{I}(Y_{t-1} \le y)] \\
&\quad -J'(y, \boldsymbol{\phi}_0^*)\mathrm{E}[(\{F_0^{*-1}(U_t)\}^2 - 1)V_i(\boldsymbol{\phi}_0^*)\mathbb{I}(Y_{t-1} \le x)],
\end{aligned}
$$

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

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

Therefore, it suffices to show that $n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n)| = O_p(1)$, for some $y \in \mathbb{R}$. Under the assumptions of Theorem 3 there exists a $y \in \mathbb{R}$ satisfying $\mathrm{E}[\{h_1/h_1(\boldsymbol{\phi}_0^*) - 1\}\mathbb{I}(Y_0 \le y)] \neq 0$

under H_1. Fix such a y. Under H_1, $\varepsilon_t = Y_t/\sqrt{h_t}$, where $h_t = \mathrm{E}[Y_t^2 \mid \mathcal{H}_{t-1}]$, $t \in \mathbb{Z}$. Therefore,

$$\begin{aligned} n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n)| \;\; &\leq \;\; \Big|n^{-1}\sum_{t=1}^{n} \varepsilon_t^2 h_t\big[h_t^{-1}(\hat{\boldsymbol{\phi}}_n) - h_t^{-1}(\boldsymbol{\phi}_0^*)\big]\mathbb{I}(Y_{t-1} \leq y)\Big| \\ &\quad + \Big|n^{-1}\sum_{t=1}^{n} \big\{\varepsilon_t^2\big(h_t/h_t(\boldsymbol{\phi}_0^*)\big) - 1\big\}\mathbb{I}(Y_{t-1} \leq y)\Big|. \end{aligned} \tag{S.11}$$

Since Assumptions (B1) and (B2) are satisfied and H_1 holds, by applying Lemma S.5, on the set $n^{1/2}\|\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*\| \leq K$, the first term on the right hand side of (S.11), is bounded by

$$\omega_L^{-1} n^{-1}\sum_{t=1}^{n} Y_t^2\big|(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*)'\tau_t(\boldsymbol{\phi}_0^*)\big| + o_p(n^{-1/2}).$$

By the Ergodic Theorem, and because $\hat{\boldsymbol{\phi}}_n \xrightarrow{p} \boldsymbol{\phi}_0^*$, $n^{-1}\sum_{t=1}^{n} Y_t^2|(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*)'\tau_t(\boldsymbol{\phi}_0^*)| = o_p(1)$. Hence, the first term on the right hand side of (S.11) is of order $o_p(1)$. Since $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*) = O_p(1)$, then by using an extended Glivenko-Cantelli type argument, we obtain that

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

Since $\mathrm{E}[\{h_1/h_1(\boldsymbol{\phi}_0^*) - 1\}\mathbb{I}(Y_0 \leq y)] \neq 0$, then the proof follows. □

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

$$\sup_{y\in\mathbb{R}} | \,\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n) - \mathcal{U}_n(y, \boldsymbol{\phi}_0) + n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'J(y, \boldsymbol{\phi}_0)\, |= o_p(1), \tag{S.12}$$

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

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

$$n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = Z_n + o_p(1), \tag{S.13}$$

where $Z_n := -\Sigma_n^{-1}(\boldsymbol{\phi}_0)n^{-1/2}\sum_{t=1}^{n}(1 - \varepsilon_t^2)\tau_t(\boldsymbol{\phi}_0)$ and $\Sigma_n(\boldsymbol{\phi}) := n^{-1}\sum_{t=1}^{n}\tau_t(\boldsymbol{\phi})\tau_t(\boldsymbol{\phi})'$, and hence, with the asymptotic uniform expansion of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}}_n)$ in (S.12), it follows as in the proof of Theorem 1 that $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}}_n)$ converges weakly to $\mathcal{U}_0(\cdot)$ in $\mathcal{D}(\mathbb{R})$, where $\mathcal{U}_0$ is the centred Gaussian process given in Theorem 1.

In the method of bootstrap data generation in Section 4.2 the transformed estimator $\hat{\boldsymbol{\phi}}_n^\dagger$ plays the role of the true parameter $\boldsymbol{\phi}_0$; recall that $\hat{\boldsymbol{\phi}}_n^\dagger = (\hat{\phi}_{n1}^\dagger, \ldots, \hat{\phi}_{n(1+p_1+p_2)}^\dagger)'$ where

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

Next, we consider the validity of the bootstrap tests for the case some components of $\boldsymbol{\phi}_0$ lie on the boundary of the parameter space; i.e., $\phi_{0i} = 0$ for some $i = 2, \ldots, p_1 + p_2 + 1$. Since $\boldsymbol{\phi}_0$ is not an interior point, in this case, the limiting behaviour of $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)$ is not linear as in (S.13), and as in the proof of Theorem 2 of Francq and Zakoian (2007),

$$n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = \lambda^{\Lambda}_n + o_p(1), \quad \lambda^{\Lambda}_n := \arg\inf_{\lambda\in\Lambda}(Z_n - \lambda)'\Sigma_n(\boldsymbol{\phi}_0)(Z_n - \lambda). \tag{S.14}$$

The vector λ^{Λ}_n is the orthogonal projection of Z_n on the convex set Λ for the inner product $\langle x, y\rangle := x'\Sigma_n(\boldsymbol{\phi}_0)y$, and it is characterized by $\lambda^{\Lambda}_n \in \Lambda$, $\langle Z_n - \lambda^{\Lambda}_n, \lambda^{\Lambda}_n - \lambda\rangle \geq 0$, $\forall\lambda \in \Lambda$; see e.g. Lemma 1.1 in Zarantonello (1971). Thus, by arguing as in the proof of Theorem 2 in Francq and Zakoian (2007) we obtain that $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \xrightarrow{d} \lambda^{\Lambda} := \arg\inf_{\lambda\in\Lambda}(\lambda - Z)'\Sigma(\boldsymbol{\phi}_0)(\lambda - Z)$. Therefore, for a continuous G, the term $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'J(y, \boldsymbol{\phi}_0)$ in (S.12) is asymptotically $\mathcal{C}$-tight, as is $\mathcal{U}_n(\cdot, \boldsymbol{\phi}_0)$ by Lemma 1 and Theorem 1.

Next, consider the bootstrap data generation as in Section 4.2. Since $\hat{\boldsymbol{\phi}}^{\dagger}_n$ converges to $\boldsymbol{\phi}_0$, a.s., by a triangular array version of the proof of Lemma 1 replacing $\hat{\boldsymbol{\phi}}_n$ and $\boldsymbol{\phi}_0$ by $\hat{\boldsymbol{\phi}}^{*}_n$ and $\hat{\boldsymbol{\phi}}^{\dagger}_n$, respectively, we obtain that conditional on $\{Y_1, \ldots, Y_n\}$, uniformly in $y \in \mathbb{R}$,

$$\mathcal{U}^{*}_n(y, \hat{\boldsymbol{\phi}}^{*}_n) = \mathcal{U}^{*}_n(y, \hat{\boldsymbol{\phi}}^{\dagger}_n) - n^{1/2}(\hat{\boldsymbol{\phi}}^{*}_n - \hat{\boldsymbol{\phi}}^{\dagger}_n)'J^{*}(y, \hat{\boldsymbol{\phi}}^{\dagger}_n) + o^{*}_p(1), \tag{S.15}$$

in probability, where

$$J^{*}(y, \boldsymbol{\phi}) = \mathrm{E}^{*}[\tau^{*}_1(\boldsymbol{\phi})\mathbb{I}(Y^{*}_0 \leq y)], \quad \tau^{*}_t(\boldsymbol{\phi}) := \frac{(\partial/\partial\boldsymbol{\phi})h^{*}_t(\boldsymbol{\phi})}{h^{*}_t(\boldsymbol{\phi})}, \quad \boldsymbol{\phi} \in \Phi.$$

Since $\mathcal{U}^{*}_n(y, \hat{\boldsymbol{\phi}}^{\dagger}_n) = n^{-1/2}\sum_{t=1}^{n}(\varepsilon^{*}_t - 1)^2\mathbb{I}(Y^{*}_{t-1} \leq y)$, by using Lemma S.2, for every $x, y \in \mathbb{R}$, with $x \wedge y := \min(x, y)$, we obtain that

$$\begin{aligned}
\mathrm{cov}^{*}\{\mathcal{U}^{*}_n(x, \hat{\boldsymbol{\phi}}^{\dagger}_n), \mathcal{U}^{*}_n(y, \hat{\boldsymbol{\phi}}^{\dagger}_n)\} &= n^{-1}\sum_{t=1}^{n}\mathrm{E}^{*}(\varepsilon^{*}_t - 1)^2\mathbb{I}(Y^{*}_{t-1} \leq x \wedge y) \\
&= (\kappa_\varepsilon - 1)\mathrm{E}\mathbb{I}(Y_{t-1} \leq x \wedge y) + o_p(1) \\
&= \mathrm{cov}\{\mathcal{U}_n(x, \boldsymbol{\phi}_0), \mathcal{U}_n(y, \boldsymbol{\phi}_0)\} + o_p(1).
\end{aligned} \tag{S.16}$$

Further, it follows from Lemma S.2 that $J^*(y,\boldsymbol{\phi}) = \mathrm{E}[\tau_1(\boldsymbol{\phi})\mathbb{I}(Y_0 \le y)] + o_p(1)$.

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

$$n^{1/2}(\hat{\boldsymbol{\phi}}_n^* - \hat{\boldsymbol{\phi}}_n^\dagger) = \lambda_n^\Lambda + o_p^*(1), \text{ in probability.} \tag{S.17}$$

In order to obtain (S.17) we consider a triangular array version of the proof of (S.14).

In the proof of (S.14) in Francq and Zakoian (2007), first λ_n^Λ is represented as the orthogonal projection of Z_n on the convex set Λ, for the inner product $\langle x, y\rangle := x'\Sigma_n(\boldsymbol{\phi}_0)y$, and then this projection is approximated by that of Z_n on the set $n^{1/2}(\Phi - \boldsymbol{\phi}_0)$. Since Φ contains the hypercube $\bar{\Phi} = [\omega_L, \omega_U] \times [0, \epsilon]^{p_1+p_2}$ which includes $\boldsymbol{\phi}_0$, see Assumption (A1), the set $n^{1/2}(\Phi - \boldsymbol{\phi}_0)$ increases to Λ as $n \to \infty$. This plays a key role in the proof of (S.14). Recall that,

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

where $\Lambda_1 = \mathbb{R}$, and for each $i = 2, \dots, p_1 + p_2 + 1$, denoting $\boldsymbol{\phi}_0 = (\phi_{01}, \dots, \phi_{0(1+p_1+p_2)})'$, $\Lambda_i = \mathbb{R}$ if $\phi_{0i} \ne 0$ and $\Lambda_i = [0, \infty)$ if $\phi_{0i} = 0$. In order to extend the proof of (S.14) to the triangular array setup of the bootstrap data generation, we need to replace $\boldsymbol{\phi}_0$ by the bootstrap true parameter $\hat{\boldsymbol{\phi}}_n^\dagger$. However, to ensure that $n^{1/2}(\Phi - \hat{\boldsymbol{\phi}}_n^\dagger)$ increases to Λ, we need to show that $\hat{\boldsymbol{\phi}}_n^\dagger$ satisfies two important conditions. First, to allow the set $\bar{\Phi}^*$ in (18), which contains $\boldsymbol{\phi}_0$, to also contain $\hat{\boldsymbol{\phi}}_n^\dagger$ with probability converging to one, we need to have

$$\hat{\boldsymbol{\phi}}_n^\dagger \to \boldsymbol{\phi}_0 \text{ in probability as } n \to \infty. \tag{S.18}$$

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

$$n^{1/2}(\hat{\phi}_{ni}^\dagger - \phi_{0i}) = \begin{cases} o_p(1), & \text{if } \phi_{0i} = 0 \\ O_p(1), & \text{if } \phi_{0i} > 0 \end{cases}, \quad i = 1, 2, \dots, 1 + p_1 + p_2. \tag{S.19}$$

Since at least one component of $\boldsymbol{\phi}_0$ is zero, the rate of consistency (S.19) ensures that $n^{1/2}(\Phi - \hat{\boldsymbol{\phi}}_n^\dagger) = n^{1/2}(\Phi - \boldsymbol{\phi}_0) - n^{1/2}(\hat{\boldsymbol{\phi}}_n^\dagger - \boldsymbol{\phi}_0)$ converges to Λ in probability.

The consistency of $\hat{\boldsymbol{\phi}}_n^\dagger$ follows from that of $\hat{\boldsymbol{\phi}}_n$, and hence (S.18) holds. Since c_n converges to 0 at a rate slower than $n^{-1/2}$, (S.19) follows by arguing as in the proof of Lemma 1 in Cavaliere *et al.* (2022). Hence, by a triangular array extension of the proof of Theorem 2 of Francq and Zakoian (2007), under Assumption (A5), we obtain that (S.17) holds; e.g., by arguing as in the proof of Proposition 3.2 in Hidalgo and Zaffaroni (2007); see also the discussion under Assumption E2 in Andrews (1997). Therefore, from (S.12)–(S.15), and the asymptotic tightness of $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}_n^{\dagger})$, we obtain that the conditional weak limit of $\mathcal{U}_n^*(\cdot, \hat{\boldsymbol{\phi}}_n^*)$ is the same as that of $\mathcal{U}_n(\cdot, \hat{\boldsymbol{\phi}}_n)$, in probability. Hence, the continuous mapping theorem yields that the bootstrap test based on T_j is asymptotically valid ($j = 1, 2$). □

Proof of Theorem 5. If $\boldsymbol{\phi}_0^*$ is an interior point, then the proof follows from Theorem 3. Therefore, we only consider the case some components of $\boldsymbol{\phi}_0^*$ lie on the boundary of the parameter space; i.e., $\phi_{0i}^* = 0$ for some $i = 2, \ldots, p_1 + p_2 + 1$. Since Assumptions (B1), (B2) and (A5) hold, with $(\boldsymbol{\phi}_0^*, F_0^*)$ being the pseudo-true value under H_1, by applying Lemmas S.2 and S.3 and arguing as in the proof of Theorem 4, for every $y \in \mathbb{R}$, we have that $\mathcal{U}_n^*(y, \hat{\boldsymbol{\phi}}_n^{\dagger}) = O_p^*(1)$, in probability. Therefore, it suffices to show that $n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n)| = O_p(1)$, for some $y \in \mathbb{R}$. Under the assumptions of the theorem there exists a y satisfying $\mathrm{E}[\{h_1/h_1(\boldsymbol{\phi}_0^*) - 1\}\mathbb{I}(Y_0 \le y)] \ne 0$. Fix such a y. Under H_1, $\varepsilon_t = Y_t/\sqrt{h_t}$, where $h_t = \mathrm{E}[Y_t^2 \mid \mathcal{H}_{t-1}]$. Hence, by using Assumptions (B1) and (B2), Lemmas S.2 and S.3, and arguing as in the proof of Lemma S.5, on the set $n^{1/2}\|\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*\| \le C$, where $0 < C < \infty$, we obtain

$$\begin{aligned} &\Big|n^{-1}\sum_{t=1}^{n} Y_t^2\big[h_t^{-1}(\hat{\boldsymbol{\phi}}_n) - h_t^{-1}(\boldsymbol{\phi}_0^*)\big]\mathbb{I}(Y_{t-1} \le y)\Big| \\ &\le \ \omega_L^{-1} n^{-1}\sum_{t=1}^{n} Y_t^2\big|(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*)'\tau_t(\boldsymbol{\phi}_0^*)\big| + o_p(n^{-1/2}). \end{aligned}$$

By the Ergodic Theorem, and because $\hat{\boldsymbol{\phi}}_n \xrightarrow{p} \boldsymbol{\phi}_0^*$, the sum in the above upper bound is $o_p(1)$, and hence $n^{-1/2}|\mathcal{U}_n(y, \hat{\boldsymbol{\phi}}_n)|$ is bounded from the above by $|n^{-1}\sum_{t=1}^{n}\{\varepsilon_t^2\big(h_t/h_t(\boldsymbol{\phi}_0^*)\big) - 1\}\mathbb{I}(Y_{t-1} \le y)|$ up to a term of order $o_p(1)$. Since $n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0^*) = O_p(1)$, then the proof follows by an extended Glivenko-Cantelli type argument as in the proof of Theorem 3. □

S.4 Additional simulation results

In this section we provide some simulation results to support the Monte Carlo simulation experiments discussed in Section 5 in the main paper. More specifically, in Section S.4.1 below we provide additional simulation results to complement the Monte Carlo simulation study discussed in Section 5.2.1 in the main paper, in order to analyze the empirical rejection probabilities of the tests under the null hypothesis. In Section S.4.2 we provide simulation results to evaluate the empirical power of the tests under the alternative hypothesis. In Section S.4.3 we discuss the choice of the shrinkage sequence $\{c_n\}$ on our bootstrap tests and compare with the choice of the length of the bootstrap samples for the 'm out of n' bootstrap.

S.4.1 Uniform size properties near and away from the boundary

In this section we provide more detailed simulation results for the Monte Carlo simulation study discussed in Section 5.2.1 in the main paper for evaluating the (uniform) finite sample size properties of the tests when some component of the true parameter is near and away from the boundary of the parameter space. The simulation results are presented in Table S1 and Figure S1 below. The DGPs and the null hypothesis are as described in Section 5.2.1 in the main paper.

S.4.2 Empirical power

In this subsection we investigate the empirical rejection probabilities of the null hypothesis H_0 specified by $h_t(\boldsymbol{\phi})$ in (24) in the main paper against several fixed and local alternatives.

First we investigate the empirical power of the tests against the following DGPs:

DGP C1 [GJR-GARCH(1,1)]: $h_t = 0.10 + 0.1Y_{t-1}^2 + 0.5h_{t-1} + 0.3Y_{t-1}^2\mathbb{I}(Y_{t-1} < 0)$,
DGP C2 [EGARCH(1,1)]: $\ln h_t = 0.1 + 0.4 \ln h_{t-1} + 0.2(|\varepsilon_{t-1}| - \mathrm{E}|\varepsilon_{t-1}|) - 0.2\varepsilon_{t-1}$,
DGP C3 [Threshold GARCH(1,1)]: $h_t = 0.10 + 0.1Y_{t-1}^2 + 0.5h_{t-1} + 0.3h_{t-1}\mathbb{I}(Y_{t-1} < 0)$.

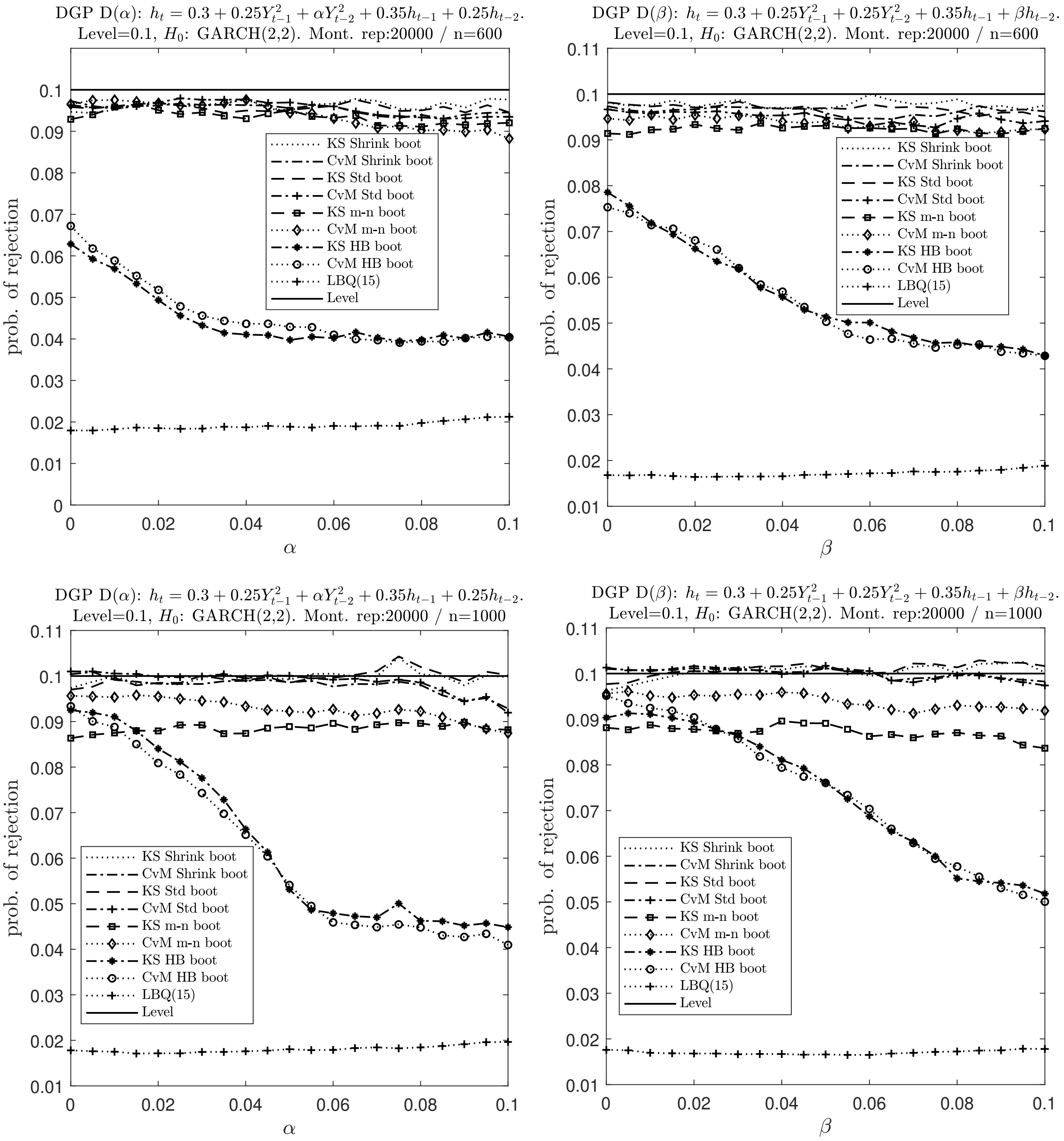
None of the DGPs C1–C3 is nested into the GARCH specification in (24). Hence, according to the theory results we expect that our bootstrap tests would reject H_0 as n diverges. The simulation results for the sample sizes $n = 500, 1000$ and 2000, at the nominal 10% level of significance, are presented in Table S2. In these simulations the shrinking based, standard and 'm out of n' bootstrap methods all perform very similarly and exhibit excellent empirical power properties. The empirical powers of the tests based on hybrid bootstrap approach are slightly lower than those based on the other three bootstrap methods, particularly for $n = 500$. However, the hybrid bootstrap approach catches up with the other bootstrap methods (in terms of empirical power of the tests) as the sample size increases to 2000. When comparing the relative performance of KS and CvM tests it can be seen that, for each bootstrap method and sample size, the CvM test exhibits better performance than the KS test when the DGP is either C1 or C2, and when the DGP is C3 the KS test performs better than the CvM test. The LBQ test does not exhibit any notable empirical power.

Table S1: Empirical rejection probabilities for testing the null hypothesis $\mathsf{H}_0: \ h_t = h_t(\boldsymbol{\phi}) = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$.

	Shrinking based bootstrap		Standard bootstrap		'm out of n' bootstrap		Hybrid bootstrap		
n	KS	CvM	KS	CvM	KS	CvM	KS	CvM	LBQ(15)
	DGP $\mathrm{D}_{n1} : (\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 1.4n^{-1/2}, 0.25, 0.35, 0.25)$								
100	0.111	0.106	0.106	0.095	0.137	0.111	0.031	0.032	0.026
200	0.097	0.099	0.090	0.089	0.109	0.100	0.032	0.035	0.023
400	0.098	0.101	0.094	0.095	0.104	0.098	0.037	0.039	0.018
600	0.102	0.105	0.098	0.100	0.100	0.103	0.040	0.045	0.016
800	0.105	0.106	0.102	0.102	0.099	0.102	0.044	0.044	0.015
1000	0.106	0.107	0.101	0.101	0.093	0.098	0.060	0.059	0.015
1200	0.101	0.105	0.098	0.096	0.095	0.100	0.073	0.069	0.013
1400	0.102	0.104	0.101	0.102	0.095	0.096	0.086	0.085	0.012
1600	0.105	0.109	0.102	0.104	0.098	0.103	0.092	0.094	0.013
1800	0.097	0.100	0.093	0.094	0.088	0.096	0.087	0.089	0.012
2000	0.107	0.109	0.104	0.104	0.090	0.098	0.099	0.100	0.014
	DGP $\mathrm{D}_{n2} : (\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.25, 0.25, 1.4n^{-1/2}, 0.35)$								
100	0.116	0.111	0.107	0.100	0.140	0.120	0.029	0.029	0.021
200	0.099	0.098	0.096	0.094	0.113	0.101	0.031	0.032	0.021
400	0.097	0.092	0.096	0.091	0.091	0.089	0.040	0.039	0.017
600	0.097	0.099	0.097	0.101	0.093	0.093	0.041	0.043	0.018
800	0.100	0.098	0.100	0.098	0.094	0.096	0.046	0.045	0.017
1000	0.102	0.101	0.104	0.102	0.091	0.094	0.054	0.055	0.017
1200	0.100	0.097	0.099	0.097	0.090	0.095	0.066	0.062	0.016
1400	0.098	0.099	0.099	0.099	0.093	0.096	0.074	0.076	0.016
1600	0.104	0.103	0.103	0.103	0.094	0.099	0.084	0.085	0.016
1800	0.095	0.094	0.095	0.094	0.085	0.094	0.083	0.082	0.015
2000	0.101	0.106	0.101	0.106	0.088	0.099	0.094	0.099	0.016

Notes: The true parameter is in the interior. The nominal level is 10%. The shrinking-based bootstrap uses $c_n = 1.6n^{-0.45}$, m-out-of-n bootstrap uses $m_n = 1.5n/\log(n)$, and for the hybrid bootstrap $\mathcal{C}_n = \{kc_n : k = 0.1, 0.2, \ldots, 0.9\}$.

Figure S1: Empirical size at the nominal 10% significance level for testing the null hypothesis '$\mathsf{H}_0 : h_t = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$'. The DGPs are $\mathrm{D}(\alpha)$: $Y_t = h_t^{1/2}\varepsilon_t$, $h_t = 0.3 + 0.25Y_{t-1}^2 + \alpha Y_{t-2}^2 + 0.35h_{t-1} + 0.25h_{t-2}$ and $\mathrm{D}(\beta)$: $Y_t = h_t^{1/2}\varepsilon_t$, $h_t = 0.3 + 0.25Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.35h_{t-1} + \beta h_{t-2}$.



In order to obtain further empirical power comparisons we also consider tests for the null specification $h_t(\boldsymbol{\phi})$ in (24) when the true DGP satisfies one of the following specifications:

$$\mathrm{M}_1(\delta): \quad Y_t = h_t^{1/2}\varepsilon_t, \quad h_t = 0.3 + 0.25Y_{t-1}^2 + 0.35h_{t-1} + \delta Y_{t-1}^2\mathbb{I}(Y_{t-1} \le 0),$$
$$\mathrm{M}_2(\delta): \quad Y_t = h_t^{1/2}\varepsilon_t, \quad h_t = 0.3 + 0.25Y_{t-1}^2 + 0.35h_{t-1} + \delta h_{t-1}\mathbb{I}(Y_{t-1} \le 0),$$

with the standard normal error distribution and with δ in the set $\{0, 0.04, 0.08, 0.12, ..., 0.4\}$. The DGPs $\mathrm{M}_1(\delta)$ and $\mathrm{M}_2(\delta)$ satisfy the null hypothesis of correct specification when $\delta = 0$, and both DGPs move away from the boundary into the alternative space as δ increases.

Figure S2 provides the empirical null rejection probabilities of the tests, at the nominal 10% level of significance, as functions of the parameter δ. The sample size is $n = 1000$ and the simulation is based on 20,000 Monte Carlo replications. In these simulations, when the DGPs move away from the boundary and move further into the alternative space, all four bootstrap approaches provide very similar results in terms of empirical power for both KS and CvM. The empirical null rejection probabilities of KS and CvM tests increase with δ, as expected. For the DGP $\mathrm{M}_1(\delta)$ the CvM test performs better than the KS test under each of the bootstrap methods, however, when the DGP is $\mathrm{M}_2(\delta)$ the KS test exhibits better performance than the CvM test. Note that the DGP $\mathrm{M}_1(\delta)$ moves into the alternative space in the direction of a misspecification of the ARCH parameters, and hence the simulation results seem to suggest that the CvM test is better suited than the KS test in detecting misspecifications of the ARCH parameters. Similarly, since the DGP $\mathrm{M}_2(\delta)$ moves in the direction of a misspecification of the GARCH parameters, it appears that the KS test is more suitable than the CvM test in detecting misspecifications of the GARCH parameters. The LBQ test does not exhibit any notable power.

We end this subsection by investigating the empirical power properties of the tests against the following two sequences of local alternative DGPs:

$$\mathrm{L}_{1n}: \quad Y_t = h_t^{1/2}\varepsilon_t, \quad h_t = 0.3 + 0.2Y_{t-1}^2 + 0.15Y_{t-2}^2 + 0.35h_{t-1} + 4n^{-1/2}Y_{t-1}^2\mathbb{I}(Y_{t-1} \le 0)$$
$$\mathrm{L}_{2n}: \quad Y_t = h_t^{1/2}\varepsilon_t, \quad h_t = 0.3 + 0.2Y_{t-1}^2 + 0.15Y_{t-2}^2 + 0.35h_{t-1} + 4n^{-1/2}h_{t-1}\mathbb{I}(Y_{t-1} \le 0),$$

with the standard normal error distribution. The tests are evaluated for testing the null hypothesis H_0 specified by $h_t(\boldsymbol{\phi})$ in (24). Thus, L_{1n} and L_{2n} both define sequences of local alternatives that approach the null model at the rate of $O_p(n^{-1/2})$.

Figure S3 presents the results on empirical power of the tests against local alternatives at 10% level of significance, for the DGPs L_{1n} and L_{2n}, based on 10,000 Monte Carlo

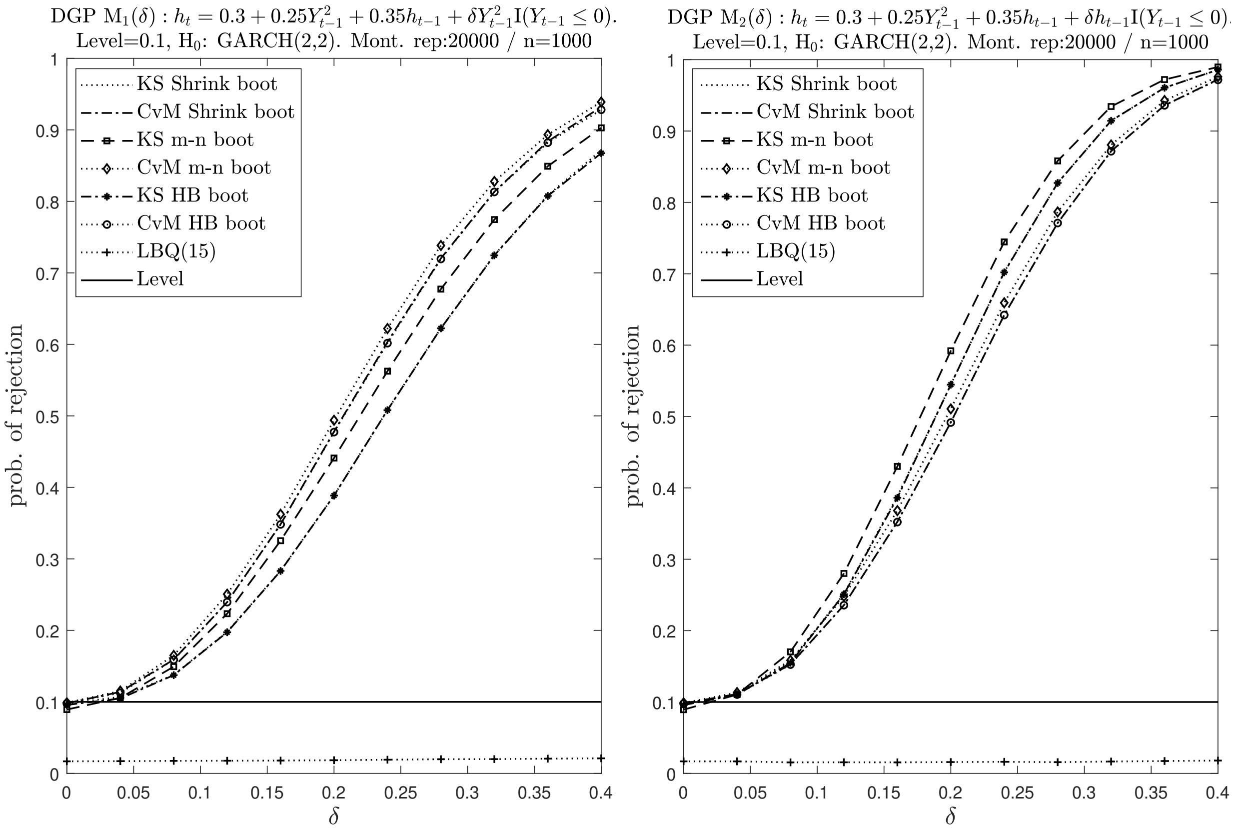
Table S2: Empirical power at 10% nominal level for testing $\mathsf{H}_0: \; h_t = h_t(\phi) = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$.

	Shrinking based bootstrap		Standard bootstrap		'm out of n' bootstrap		Hybrid bootstrap		
n	KS	CvM	KS	CvM	KS	CvM	KS	CvM	LBQ(15)
	DGP C1: $h_t = 0.10 + 0.1Y_{t-1}^2 + 0.5h_{t-1} + 0.3Y_{t-1}^2\mathbb{I}(Y_{t-1} < 0)$								
500	0.468	0.571	0.468	0.572	0.527	0.588	0.451	0.560	0.017
1000	0.776	0.855	0.775	0.856	0.811	0.865	0.775	0.855	0.019
2000	0.974	0.991	0.974	0.991	0.979	0.991	0.974	0.991	0.019
	DGP C2: $\ln h_t = 0.1 + 0.4 \ln h_{t-1} + 0.2(\lvert\varepsilon_{t-1}\rvert - \mathrm{E}\lvert\varepsilon_{t-1}\rvert) - 0.2\varepsilon_{t-1}$								
500	0.732	0.811	0.736	0.813	0.787	0.831	0.210	0.347	0.017
1000	0.963	0.983	0.964	0.983	0.970	0.984	0.951	0.979	0.018
2000	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.018
	DGP C3: $h_t = 0.10 + 0.1Y_{t-1}^2 + 0.5h_{t-1} + 0.3h_{t-1}\mathbb{I}(Y_{t-1} < 0)$								
500	0.546	0.521	0.553	0.522	0.605	0.545	0.385	0.346	0.016
1000	0.852	0.821	0.855	0.821	0.878	0.832	0.847	0.811	0.016
2000	0.992	0.983	0.992	0.983	0.993	0.984	0.992	0.983	0.015

Notes: The DGPs C1–C3 are under the alternative. The shrinking-based bootstrap uses $c_n = 1.6n^{-0.45}$, m-out-of-n bootstrap uses $m_n = 1.5n/\log(n)$, and for the hybrid bootstrap $\mathcal{C}_n = \{kc_n : k = 0.1, 0.2, \ldots, 0.9\}$.

replications (the results for standard bootstrap are not included in Figure S3 as they turn out to be similar to those for shrinking based bootstrap). In view of the simulation results all the bootstrap tests exhibit non-trivial empirical power against the two sequences of local alternatives L_{1n} and L_{2n}. The LBQ test does not exhibit any notable empirical power. The relative performance between KS and CvM tests, in terms of their empirical power against local alternatives, is similar in pattern to the previous setting on their empirical power against fixed alternatives. More specifically, for each bootstrap method, when the DGP is L_{1n} (with misspecified ARCH parameters) the CvM test exhibits better empirical power than the KS test, and when the DGP is L_{2n} (i.e. when the GARCH parameters are misspecified) the KS test performs better than the CvM test in terms of empirical power.

Figure S2: Empirical rejection probabilities of the tests at the nominal 10% significance level for testing the null hypothesis $\mathsf{H}_0: \; h_t = h_t(\boldsymbol{\phi}) = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$. The sample size $n = 1000$.

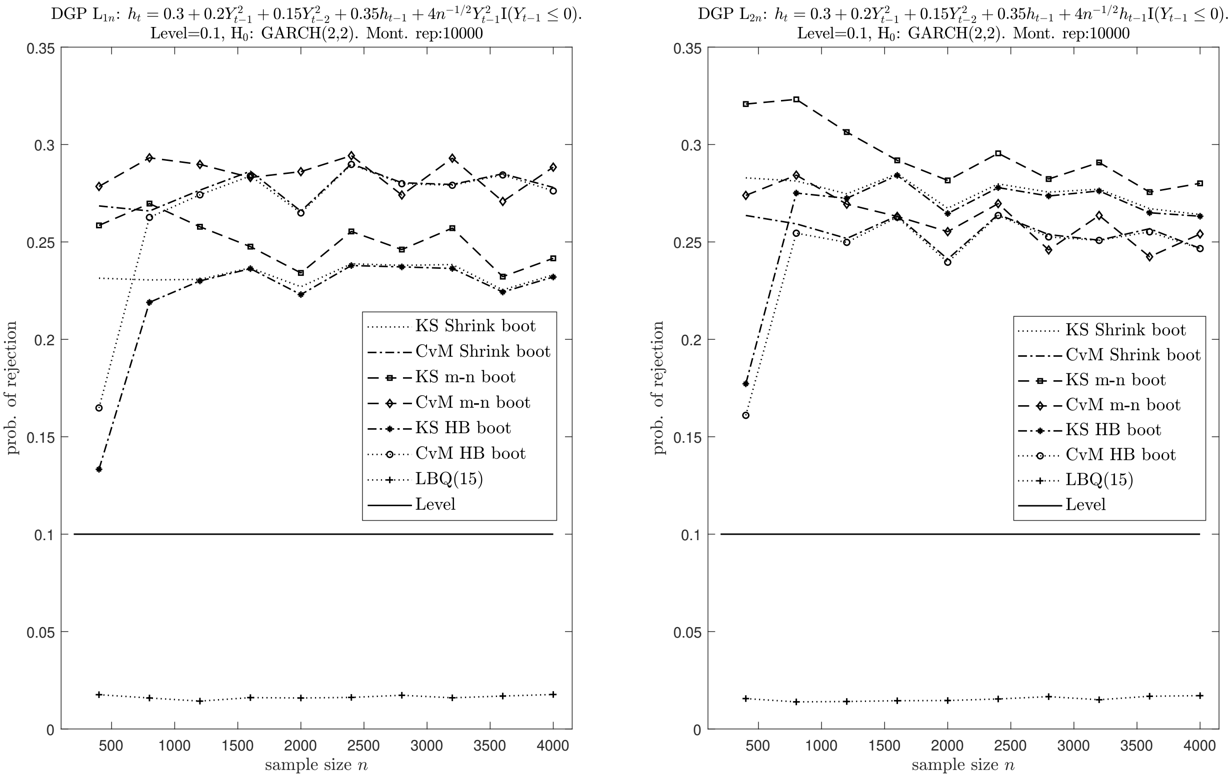


S.4.3 The choice of the shrinkage sequence

In order to investigate the effect of the choice of the shrinkage parameter c_n we evaluate the performance of the shrinking-based bootstrap tests while considering different choices for c_n. More specifically, we set $c_n := vn^{-\epsilon}$, with $\epsilon = 0.45$, and the tuning parameter v is chosen in the set $\mathcal{V} := \{0.2, 0.4, 0.8, 1.2, 1.6, 2.0\}$. Similarly, for the '$m$ out of n' bootstrap implementation, we set $m_n := cn/\log(n)$ with the tuning parameter c in the set $\mathcal{C} := \{1, 1.5, 2, 2.5, 3, 3.5\}$. Similar choices for c_n and m_n are also considered in Cavaliere *et al.* (2022). The results on empirical size and power presented above and in the main paper are based on $v = 1.6$ and $c = 1.5$.

Tables S3 and S4 provide the empirical null rejection probabilities, at the nominal 10% level of significance, for the shrinking-based and 'm out of n' bootstrap tests. The tests are evaluated when the parametric form $h_t(\boldsymbol{\phi})$ under H_0 is of the form (24), with the above

Figure S3: Empirical power against local alternatives at the nominal 10% significance level for testing '$\mathsf{H}_0 : h_t = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$', for the DGPs L_{1n} and L_{2n}. The simulation is based on 10,000 Monte Carlo replications.



choices of c_n and m_n. The DGPs D1–D5 are as defined in (25) in the main paper. The sample sizes $n = 100, 600$ and 2000 are considered. The simulation is based on 20,000 Monte Carlo replications. In these simulations, for the sample sizes $n = 100$ and $n = 600$, the tests based on the shrinking-based bootstrap show some sensitivity to the choice of the shrinkage parameter c_n. For example, when $n = 100$, the empirical null rejection rates of the CvM test for the DGP D2 varies between 9.5% and 11% as the tuning parameter v varies between 0.2 and 2.0. However, as the sample size increases to $n = 2000$, the shrinking-based bootstrap performs consistently well throughout with good empirical size properties for both KS and CvM tests and the test results do not vary significantly on the choice of the shrinkage parameter c_n. By comparison, the empirical rejection probabilities of the tests given by the 'm out of n' bootstrap exhibit significant sensitivity to the choice of the tuning parameter m_n, even when the sample size is as large as $n = 2000$, particularly when it comes to the KS test.

Table S3: Empirical size (%) of the KS test for shrinking-based and 'm out of n' bootstrap implementations with varying c_n and m_n parameters.

	Shrinking based boot ($c_n = vn^{-\epsilon}$)						'm out of n' boot ($m_n = cn/\log(n)$)					
	v						c					
n	0.2	0.4	0.8	1.2	1.6	2.0	1	1.5	2	2.5	3	3.5
	DGP D1: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.5, 0.45, 0, 0)$											
100	10.7	10.7	11.0	11.1	11.2	11.4	13.9	11.7	11.2	10.7	10.6	10.5
600	9.3	9.4	9.4	9.5	9.5	9.6	8.0	8.1	8.4	8.4	8.8	9.0
2000	9.5	9.5	9.5	9.5	9.5	9.5	7.6	7.8	8.6	8.5	8.8	8.7
	DGP D2: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.2, 0.3, 0.45, 0)$											
100	10.0	10.1	10.3	10.6	10.8	11.0	14.6	11.9	11.0	10.3	10.1	9.8
600	9.6	9.6	9.7	9.7	9.7	9.7	9.2	8.6	9.0	8.9	9.1	9.0
2000	10.0	10.0	9.9	10.0	10.0	10.1	8.1	8.6	8.9	9.2	9.2	9.6
	DGP D3: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0, 0.4, 0, 0.55)$											
100	10.8	10.8	10.9	10.8	10.9	10.9	18.1	14.7	13.3	11.9	11.4	10.8
600	10.5	10.5	10.5	10.5	10.5	10.5	11.1	10.2	10.0	10.2	10.2	10.2
2000	10.4	10.3	10.3	10.3	10.3	10.3	9.5	9.9	9.7	10.1	10.0	10.1
	DGP D4: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0, 0.4, 0.55, 0)$											
100	9.2	9.4	9.7	9.9	9.9	9.9	14.9	12.4	11.3	10.4	10.1	9.6
600	9.8	9.9	10.0	10.0	10.1	10.2	10.7	9.8	9.3	9.6	9.6	9.4
2000	10.1	10.1	10.2	10.3	10.5	10.5	8.9	9.4	9.7	9.8	9.7	9.6
	DGP D5: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.2, 0.25, 0.2, 0.3)$											
100	10.1	10.1	10.2	10.6	10.7	10.7	14.5	12.5	11.4	10.5	9.7	9.6
600	9.8	9.8	9.9	9.9	9.9	9.9	9.7	9.0	9.3	9.4	9.4	9.3
2000	10.1	10.1	10.1	10.2	10.2	10.2	8.6	9.2	9.2	9.4	9.1	9.3

Notes: The DGPs D1–D5 are of the form $Y_t = h_t^{1/2}\varepsilon_t$, $h_t = \omega_0 + \alpha_{01}Y_{t-1}^2 + \alpha_{02}Y_{t-2}^2 + \beta_{01}h_{t-1} + \beta_{02}h_{t-2}$. The null hypothesis is $\mathsf{H}_0: \ h_t = h_t(\boldsymbol{\phi}) = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$. The nominal level is 10%. The simulation is based on 20000 Monte Carlo replications.

Table S4: Empirical size (%) of the CvM test for shrinking-based and 'm out of n' bootstrap implementations with varying c_n and m_n parameters.

	Shrinking based boot ($c_n = vn^{-\epsilon}$)						'm out of n' boot ($m_n = cn/\log(n)$)					
	v						c					
n	0.2	0.4	0.8	1.2	1.6	2.0	1	1.5	2	2.5	3	3.5
	DGP D1: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.5, 0.45, 0, 0)$											
100	10.2	10.2	10.4	10.6	10.7	11.0	11.6	11.0	10.6	10.2	10.3	10.2
600	9.4	9.4	9.5	9.5	9.6	9.6	8.6	9.2	9.1	9.1	9.3	9.4
2000	9.3	9.3	9.5	9.6	9.6	9.6	8.8	8.6	8.8	8.8	8.6	8.8
	DGP D2: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.2, 0.3, 0.45, 0)$											
100	9.5	9.6	9.8	10.0	10.4	11.0	11.2	10.5	10.5	10.0	9.4	9.2
600	9.5	9.5	9.5	9.5	9.6	9.6	8.9	8.8	9.1	9.2	9.2	9.1
2000	9.6	9.6	9.6	9.6	9.6	9.7	8.8	9.0	9.1	9.3	9.3	9.4
	DGP D3: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0, 0.4, 0, 0.55)$											
100	10.6	10.6	10.7	10.7	10.7	10.7	13.9	12.8	12.0	11.7	11.1	10.6
600	10.4	10.5	10.6	10.6	10.6	10.6	10.4	10.3	10.0	10.4	10.2	10.3
2000	10.3	10.3	10.3	10.3	10.3	10.3	10.2	10.3	10.3	10.3	10.3	10.3
	DGP D4: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0, 0.4, 0.55, 0)$											
100	9.0	9.1	9.3	9.6	9.8	9.8	11.1	10.2	10.1	9.6	9.0	9.0
600	9.5	9.6	9.8	10.0	10.2	10.2	9.7	9.7	9.5	9.8	9.4	9.4
2000	10.1	10.2	10.3	10.4	10.4	10.4	9.3	9.7	10.0	10.1	9.9	10.0
	DGP D5: $(\omega_0, \alpha_{01}, \alpha_{02}, \beta_{01}, \beta_{02}) = (0.3, 0.2, 0.25, 0.2, 0.3)$											
100	9.9	9.8	10.1	10.4	10.6	10.7	11.6	11.0	10.8	10.2	9.7	9.6
600	9.1	9.1	9.1	9.2	9.3	9.4	9.1	9.1	9.1	9.4	9.3	9.0
2000	9.9	9.9	9.9	9.9	9.9	10.0	9.2	9.7	9.6	9.7	9.4	9.5

Notes: The DGPs D1–D5 are of the form $Y_t = h_t^{1/2}\varepsilon_t$, $h_t = \omega_0 + \alpha_{01}Y_{t-1}^2 + \alpha_{02}Y_{t-2}^2 + \beta_{01}h_{t-1} + \beta_{02}h_{t-2}$. The null hypothesis is $\mathsf{H}_0 : \ h_t = h_t(\boldsymbol{\phi}) = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$, for some $\omega > 0, \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$. The nominal level is 10%. The simulation is based on 20000 Monte Carlo replications.

References

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

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

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

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

Cavaliere, G., Nielsen, H. B., Pedersen, R. S., and Rahbek, A. (2022). Bootstrap Inference On The Boundary Of The Parameter Space With Application To Conditional Volatility Models. *Journal of Econometrics*, **227**(1), 241–263.

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

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

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

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

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

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

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. (2018). Specification tests for time series models with garch-type conditional variance. *Discussion paper: Available at SSRN: https://ssrn.com/abstract=3141822*.

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

Perera, I. and Silvapulle, M. J. (2022). Bootstrap specification tests for dynamic conditional distribution models. *Journal of Econometrics*, (in press).

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

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

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