Econometric Theory, **40**, 2024, 1422–1455.
doi:10.1017/S0266466623000142

INFERENCE ON GARCH-MIDAS MODELS WITHOUT ANY SMALL-ORDER MOMENT

CHRISTIAN FRANCQ
CREST–ENSAE and University of Lille

BAYE MATAR KANDJI
CREST–ENSAE

JEAN-MICHEL ZAKOIAN
CREST-ENSAE and University of Lille

In GARCH-mixed-data sampling models, the volatility is decomposed into the product of two factors which are often interpreted as "short-run" (high-frequency) and "long-run" (low-frequency) components. While two-component volatility models are widely used in applied works, some of their theoretical properties remain unexplored. We show that the strictly stationary solutions of such models do not admit any small-order finite moment, contrary to classical GARCH. It is shown that the strong consistency and the asymptotic normality of the quasi-maximum likelihood estimator hold despite the absence of moments. Tests for the presence of a long-run volatility relying on the asymptotic theory and a bootstrap procedure are proposed. Our results are illustrated via Monte Carlo experiments and real financial data.

1. INTRODUCTION

Despite their ability to capture a number of empirical characteristics of financial returns, the restrictive features of "one-factor" classical GARCH models are well known. The parameter β in a GARCH(1,1) has to be close to 1 to ensure high-volatility persistence, but this may induce undesirable restrictions on the marginal distribution of the returns. Moreover, parameters governing the short-run effect of shocks (α in the usual GARCH(1,1) parameterization, as in the equation of σ_t^2 in Model (1)) also impact the long-run response through the coefficients ($\alpha\beta^i$) of the asymptotic expansion of the volatility as a function of the past squared returns. This lack of flexibility, in particular, the necessity to disentangle short- and long-run impacts of shocks, has motivated the introduction of alternative volatility specifications in the econometric and finance literatures. Additive component GARCH models were introduced by Ding and Granger (1996) and Engle and

The authors are grateful to two anonymous referees, the Co-Editor, and the Editor for insightful comments, suggestions, and useful criticisms. The authors are also grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MLforRisk (ANR-21-CE26-0007). Address correspondence to Jean-Michel Zakoïan, CREST–ENSAE, Palaiseau, France; e-mail: zakoian@ensae.fr

Lee (1999) but, in recent years, multiplicative component GARCH processes have attracted more attention. In such models, called GARCH-mixed-data sampling (GARCH-MIDAS), the volatility is decomposed into the product of two factors which may receive different interpretations, generally in terms of "short-run" (high-frequency) and "long-run" (low-frequency) components. To cite just a few recent references, the reader is referred to Engle, Ghysels, and Sohn (2013), Wang and Ghysels (2015), Amado and Teräsvirta (2017), Conrad, Custovic, and Ghysels (2018), and Conrad and Engle (2021).

While GARCH-MIDAS volatility models are widely used in applied works, some of their theoretical properties remain unexplored. An exception is the paper by Wang and Ghysels (2015) who consider stationarity and ergodicity, as well as asymptotic theory for the quasi-maximum likelihood (QML) estimator, under assumptions, we will further discuss. In this paper, we consider three issues: first, the existence of small-order moments for the strictly stationary solution of the two-component volatility model; second, the consistency and asymptotic normality of the QML estimator; and third, testing the existence of a long-run volatility. The first two issues are closely related because all existing proofs of the consistency and asymptotic normality of QML estimators in standard GARCH models rely on the existence of small-order moments. The third issue was also considered by Conrad and Schienle (2020) who proposed a score-based test in a general multiplicative component model.

One characteristic of most commonly used GARCH-type models is that strict stationarity entails the existence of a small-order moment. Hence, even if stationary solutions (r_t) of standard GARCH models are generally characterized by heavy tails (a desirable property for the modeling of financial returns), a maximal moment exponent exists: for a sufficiently small power s (depending on both the volatility parameters and the innovations distribution), we have $\mathbb{E}|r_t|^s < \infty$. In a sense, this means that such one-factor volatility models are too constrained, as the conditions ensuring stability of the dynamics produce unexpected restrictions on the marginal distributions. By contrast, the models we consider in this paper have the surprising property of admitting strictly stationary solutions that do not have any power moment (unless a very restrictive condition is imposed on the errors distribution). This heaviness of the tails of the marginal distribution entails formidable statistical difficulties for proving the consistency and asymptotic normality of the QML estimator. Indeed, the existence of a small moment for the observed process is crucial to derive the asymptotic properties of the QMLE in most GARCH-type models (see, for instance, Francq and Zakoïan, 2019, Section 7.4). In particular, contrary to the standard GARCH case, the proof of the consistency cannot rely on the existence of a limiting QML criterion. To circumvent the absence of moments, we use a property of exponential control of the trajectories which will be detailed below.

The rest of the paper is organized as follows: In the next section, we study the existence of strictly stationary solutions to the GARCH-MIDAS volatility model and their moment properties. Section 3 considers the estimation by QML of the model parameters. In Section 4, we propose tests for the existence of a long-run volatility. Two approaches are considered to handle the problem of unidentified parameters under the null and bootstrap procedures are proposed. Numerical and empirical results are presented in Section 5. Section 6 concludes. Proofs are given in the Appendix.

2. MODEL AND AN UNEXPECTED PROPERTY OF THE STATIONARY SOLUTION

We study, in this article, a class of GARCH-MIDAS processes (r_t) defined by

$$\begin{cases} r_t = \tau_t \epsilon_t, \quad \tau_t^2 = 1 + a_0 \sum_{i=1}^{Q} \varphi_i(\boldsymbol{\vartheta}_0) RV_{t-i}, \\ \epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{cases} \tag{1}$$

where (η_t) is an i.i.d. sequence with $\mathbb{E}\eta_t^2 = 1$, $a_0 \geq 0$, $\omega_0 > 0$, $\alpha_0 \geq 0$, and $\beta_0 \geq 0$, $RV_t = \sum_{i=0}^{N-1} r_{t-i}^2$ is a rolling window realized volatility, Q and N are positive integers, and $\varphi_i(\boldsymbol{\vartheta}_0)$ are positive weights, depending on some d-variate parameter $\boldsymbol{\vartheta}_0$. For the specification of the functions φ_i used to smooth the realized volatilities Engle et al. (2013), under a slightly different parametrization,[1] suggest exponential weights

$$\varphi_i(\vartheta_0) = \frac{\vartheta_0^i}{\sum_{j=1}^{Q} \vartheta_0^i}, \quad \vartheta_0 \in (0, \infty), \tag{2}$$

or beta weights

$$\varphi_i(\vartheta_0) = \frac{\{1 - i/(Q+1)\}^{\vartheta_0 - 1}}{\sum_{j=1}^{Q} \{1 - j/(Q+1)\}^{\vartheta_0 - 1}}, \quad \vartheta_0 \in (0, \infty).^{2} \tag{3}$$

The standard GARCH(1,1) is obtained for $a_0 = 0$. For $a_0 > 0$, the volatility component τ_t^2 is often referred to as the *long-run volatility* (for large q), whereas the *short-run volatility* σ_t^2 is a function of the normalized (long-run detrended) squared returns r_{t-i}^2 / τ_{t-i}^2.

Model (1) can be written under the following form, which will be used throughout,

$$\begin{cases} r_t = \tau_t \epsilon_t, \quad \tau_t^2 = 1 + a_0 \sum_{i=1}^{q} \phi_i(\boldsymbol{\vartheta}_0) r_{t-i}^2, \\ \epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{cases} \tag{4}$$

where the $\phi_i(\boldsymbol{\vartheta}_0)$'s are nonnegative, with at least one strictly positive coefficient. Model (4) is the model we focus on, and is more general than the GARCH-

[1] Engle et al. (2013) considered a unit-variance GARCH(1,1) equation, $\sigma_t^2 = 1 - \alpha_0 - \beta_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2$ for the short-run volatility and introduced an intercept m in the equation of τ_t^2. This choice is guided by the necessity to identify short- and long-run volatilities. The alternative identifiability condition we adopt here is a unit intercept, $m = 1$, in the long-term volatility dynamics. This constraint is not restrictive, whereas imposing a unit-variance for the short-run volatility requires $\alpha_0 + \beta_0 < 1$, which is not necessary for strict stationarity. Note that Engle et al. (2013) also allow for an intercept in the equation of r_t.

[2] In these examples, the weight parameter ϑ_0 is scalar, and therefore is not shown in bold.

MIDAS for which we have $q = N + Q - 1$. Without loss of generality, assume that $\sum_{i=1}^{q} \phi_i(\boldsymbol{\vartheta}_0) = 1$.

Next, we turn to the existence of strictly stationary solutions to Model (4).

Let $\delta_t = \alpha_0 \eta_t^2 + \beta_0$. Under the assumption

A1 $\gamma := \mathbb{E} \log \delta_1 < 0$,

the GARCH(1,1) equation in (4) admits the strictly stationary, non anticipative and ergodic solution

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} \delta_{t-j} \right). \tag{5}$$

Note that **A1** is less restrictive than the condition $\alpha_0 + \beta_0 < 1$ used in Wang and Ghysels (2015).

It is known that, for $r > 0$,

$$\mathbb{E}(\sigma_t^{2r}) < \infty \text{ if and only if } \mathbb{E}\delta_1^r < 1. \tag{6}$$

Therefore σ_t, and thus ϵ_t, cannot admit moments of any order when δ_t is not almost surely bounded by 1, i.e., when

A2 $P(\delta_1 > 1) \neq 0$.

Indeed, for $\iota > 0$ such that $P(\delta_1 > 1 + \iota) > 0$, we have

$$\mathbb{E}\delta_1^r \geq (1+\iota)^r P(\delta_1 > 1 + \iota) \to \infty,$$

as $r \to \infty$. In particular, **A2** is satisfied when η_t^2 is not bounded and $\alpha_0 \neq 0$.[3] It follows from (6) that $\mathbb{E}(\sigma_t^{2r}) = \infty$ for r large enough. This is a well-known property dating back to Kesten (1973; see also Mikosch and Stărică, 2000).

Write (4) in matrix form as

$$\boldsymbol{r}_t = \boldsymbol{A}_t \boldsymbol{r}_{t-1} + \boldsymbol{b}_t, \tag{7}$$

where $\boldsymbol{r}_t = (r_t^2, \ldots, r_{t-q+1}^2)'$, $\boldsymbol{b}_t = (\epsilon_t^2, \boldsymbol{0}_{q-1}')'$, and $\boldsymbol{A}_t = \boldsymbol{A}(\epsilon_t)$ is a companion-like matrix:

$$\boldsymbol{A}_t = \begin{pmatrix} a_0 \phi_1(\boldsymbol{\vartheta}_0)\epsilon_t^2 & \ldots & a_0 \phi_{q-1}(\boldsymbol{\vartheta}_0)\epsilon_t^2 & a_0 \phi_q(\boldsymbol{\vartheta}_0)\epsilon_t^2 \\ 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Noting that, under **A1**, the sequence $(\boldsymbol{A}_t, \boldsymbol{b}_t)$ is strictly stationary and ergodic, equation (7) admits, by Brandt (1986, Thm. 1), the strictly stationary solution

$$\boldsymbol{r}_t = \boldsymbol{b}_t + \sum_{i=1}^{\infty} \left(\prod_{j=1}^{i} \boldsymbol{A}_{t+1-j} \right) \boldsymbol{b}_{t-i} \tag{8}$$

under the assumption

[3]This assumption is therefore very mild. Moreover, it can be verified in practice by estimating $P(\delta_1 > 1)$.

A3 $\gamma_A < 0, \quad$ where $\gamma_A = \lim_{k\to\infty} \frac{1}{k}\mathbb{E}\log \|\boldsymbol{A}_k\boldsymbol{A}_{k-1}\ldots\boldsymbol{A}_1\| < 0.$

The top-Lyapounov exponent γ_A involved in **A3** is well defined in $[-\infty,\infty)$ because $\mathbb{E}\log^+\|\boldsymbol{A}_t\| < \infty$, in view of (9). Wang and Ghysels (2015) obtained explicit conditions entailing **A3** for particular submodels. The next assumption guarantees that the long- and short-run volatilities τ_t and σ_t are not degenerate.

A4 $a_0 > 0$ and $\alpha_0 > 0$.

According to Lemma 2.3 in Berkes, Horváth, and Kokoszka (2003), the strictly stationary solution ϵ_t of the standard GARCH(1,1) equation satisfies

$$\mathbb{E}|\epsilon_t|^s < \infty \quad \text{for some } s > 0. \tag{9}$$

The following proposition shows that, surprisingly, this feature does not extend to the solution (r_t) of the GARCH-MIDAS model (4).

We start by proving the following lemma, of independent interest as it concerns the GARCH(1,1) process (ϵ_t).

LEMMA 1. *Assume **A1** and **A2**. For all integer $k \geq 2$, all real numbers $p_j > 0$ and integers i_j, $j = 1,\ldots k$, there exists $K \in (0,\infty]$ such that*

$$\mathbb{E}|\epsilon_{t-i_1}|^{p_1}|\epsilon_{t-i_1-i_2}|^{p_2}\ldots|\epsilon_{t-i_1-\cdots-i_k}|^{p_k} \geq K\mathbb{E}|\epsilon_1|^{p_1+\cdots+p_k}.$$

The right-hand side, and thus the left-hand side, of the inequality is infinite when $p_1+\cdots+p_k$ is large enough.

PROPOSITION 1. *Under **A1** and **A3**, there exists a strictly stationary and ergodic solution (r_t) to (4). If in addition **A2** and **A4** hold, this solution does not admit any moments, in the sense that*

$$\mathbb{E}|r_t|^s = \infty \quad \textit{for all } s > 0. \tag{10}$$

Note that Wang and Ghysels (2015) showed that $\mathbb{E}|r_t|^2 = \infty$, under slightly more restrictive assumptions on the distribution of η_t (see their Proposition 3.9).

Remark 1. Without Assumption **A2**, the GARCH-MIDAS process may admit moments at any order. Indeed, suppose that $\delta_1 \in [0,1]$ with probability 1. It follows that, for any $s > 0$, $\mathbb{E}|\delta_t|^s < 1$ using (5). Since $|\eta_t|$ is bounded when $\delta_t < 1$, both σ_t^2 and ϵ_t^2 admit finite moments at any order. If in addition ϵ_t^2 is bounded with probability 1 (which holds when $|\delta_1| < \overline{\delta} < 1$ with probability 1), let $\overline{\boldsymbol{A}}$ the upper bound of the matrices $\boldsymbol{A}_t$ componentwise. If the spectral radius of $\overline{\boldsymbol{A}}$ is less than 1, then Assumption **A3** is satisfied and, by (8), r_t^2 admits moments at any order.

Example 1 (Trajectory of a process without any finite moment). Figure 1 displays a simulated trajectory of the simplest version of Model (4), which we know, from Proposition 1 that it is a strictly stationary process without any finite moment. Other simulations have been carried out, but the absence of any finite moment is, to say the least, difficult to detect on the trajectories.

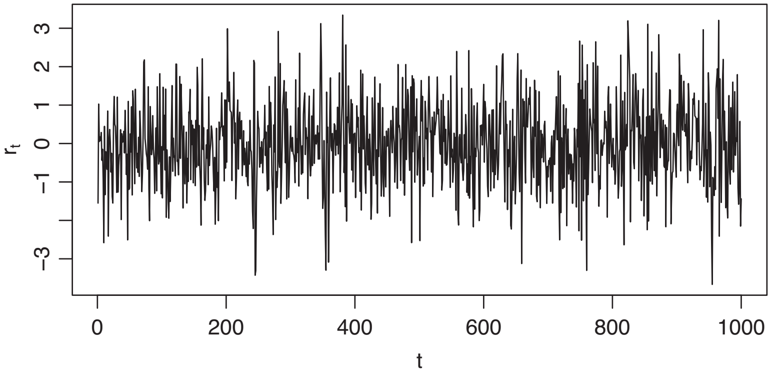

FIGURE 1. Simulation of $r_t = \sqrt{1+0.1r_{t-1}^2}\epsilon_t$ with $\epsilon_t = \sqrt{1+0.05\epsilon_{t-1}^2}\eta_t$, $\eta_t \sim \mathcal{N}(0,1)$.

3. QMLE WITHOUT MOMENT ASSUMPTION ON THE OBSERVED PROCESS

In this section, we study the estimation of the true parameter value $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0, a_0, \boldsymbol{\vartheta}_0')'$ in Model (4), assuming the functions ϕ_i are known and such that $\sum_{i=1}^q \phi_i(\cdot) = 1$. We start by introducing a consequence of the strict stationarity which will replace the existence of a small moment in the proof of the consistency and asymptotic normality (CAN) of the QMLE.

3.1. Exponential Control of the Trajectories

Wang and Ghysels (2015) studied the asymptotic distribution of the QMLE of the GARCH-MIDAS under the assumption that

$$\mathbb{E}|r_t|^s < \infty \text{ for some } s > 0. \tag{11}$$

This is a key assumption to show the CAN of the QMLE of GARCH (see Berkes, Horváth, and Kokoszka, 2003; Francq and Zakoïan, 2004). To the authors' knowledge, the consistency of the QMLE has never been shown without an assumption that implies (11). Proposition 1 however entails that (11) cannot be assumed in our framework.

To circumvent the failure of the small-order moment assumption, we will use the following lemma, which is a consequence of Kandji (2023).

LEMMA 2 (Kandji, 2023). *Under **A1** and **A3**, the strictly stationary solution of (4) satisfies*

$$\limsup_{k\to\infty} \frac{1}{k} \log r_{t+k}^2 \le 0, \qquad \limsup_{k\to\infty} \frac{1}{k} \log r_{t-k}^2 \le 0 \quad a.s., \tag{12}$$

for all $t \in \mathbb{Z}$.

This property can be interpreted as an exponential control of the trajectories. It is easy to see that (11) implies (12),[4] but the converse is false.[5]

Assume that the observations $r_1,\dots,r_n$ constitute a realization (of length n) of the two-factor GARCH process defined by (4), for the value $\boldsymbol{\theta}_0$ of the parameter. Let $\boldsymbol{\Theta}$ a compact subset of $(0,\infty)\times[0,\infty)^2\times[0,1)\times\mathbb{R}^d$ and assume $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}$. For initial values $r_0,\dots,r_{-q},\tilde{\sigma}_0^2$, and for $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, the conditional Gaussian quasi-likelihood is given by

$$\tilde{L}_n(\boldsymbol{\theta})=\tilde{L}_n(\boldsymbol{\theta};r_1,\dots,r_n)=\prod_{t=1}^{n}\frac{1}{\sqrt{2\pi\tilde{\tau}_t^2\tilde{\sigma}_t^2}}\exp\left(-\frac{r_t^2}{2\tilde{\tau}_t^2\tilde{\sigma}_t^2}\right),$$

where the $\tilde{\tau}_t^2$ and $\tilde{\sigma}_t^2$ are recursively defined, for $t\geq 1$, by

$$\tilde{\tau}_t^2=\tilde{\tau}_t^2(\boldsymbol{\theta})=1+a\sum_{i=1}^{q}\phi_i(\boldsymbol{\vartheta})r_{t-i}^2,$$

$$\tilde{\sigma}_t^2=\tilde{\sigma}_t^2(\boldsymbol{\theta})=\omega+\alpha\tilde{\epsilon}_{t-1}^2+\beta\tilde{\sigma}_{t-1}^2,\quad \tilde{\epsilon}_t^2=\frac{r_t^2}{\tilde{\tau}_t^2}.$$

A QMLE of $\boldsymbol{\theta}_0$ is defined as any measurable solution of

$$\widehat{\boldsymbol{\theta}}_n=\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\,\tilde{L}_n(\boldsymbol{\theta})=\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\,\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_t(\boldsymbol{\theta}):=\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\,\tilde{\mathbf{l}}_n(\boldsymbol{\theta}),$$

where $\tilde{\ell}_t(\boldsymbol{\theta})=\frac{r_t^2}{\tilde{\tau}_t^2\tilde{\sigma}_t^2}+\log\tilde{\tau}_t^2+\log\tilde{\sigma}_t^2$.

3.2. Asymptotic Properties of the QMLE

To establish the strong consistency of the QMLE, we need the following additional assumptions.

A5 The support of the law of η_t^2 contains three distinct points.

A6 $(\phi_i(\boldsymbol{\vartheta}))_{i=1\dots,q}=(\phi_i(\boldsymbol{\vartheta}_0))_{i=1\dots,q}\ \Rightarrow\ \boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0$.

A7 $\mathbb{E}\log\eta_t^2>-\infty$.

For the volatility of a standard GARCH to be nondegenerate, we know that the support of the law of η_t must contain three distinct points. To show the identifiability of the GARCH-MIDAS, we need the slightly stronger assumption **A5**. This is due to the fact that the volatility of this model can be written as a polynomial of order 2 in η_{t-1}^2 (instead of order 1 in the GARCH case), with coefficients belonging to the sigma-field generated by $\{\eta_u,u\leq t-2\}$. Assumption **A6** is another identifiability condition which is satisfied, in particular, for the

[4]See, for instance, Exercise 4.12 in Francq and Zakoïan (2019).

[5]Let a sequence (X_t) of identically distributed random variables such that $\mathbb{E}|X_t|<\infty$ but $\mathbb{E}X_t^2=\infty$. Then $r_t=e^{|X_t|/2}$ satisfies (12) because $k^{-1}\log r_{t+k}^2=k^{-1}|X_{t+k}|\to 0$ a.s. (see, for instance, Exercise 2.13 in Francq and Zakoïan, 2019). On the other hand, (11) is not satisfied because $\mathbb{E}|r_t|^s=\mathbb{E}e^{s|X_t|/2}\geq\frac{1}{2}\mathbb{E}(s|X_t|/2)^2=\infty$, for any $s>0$.

exponential weights $\phi_i(\vartheta) = \vartheta^i / \sum_{j=1}^{q} \vartheta^i$ (except when $q = 1$). The assumption is also satisfied for the beta weighting schemes (3) (with obvious change of notation). Assumption **A7**, precluding densities with too much mass around 0, is satisfied by most commonly used distributions. It is not required for the consistency of the standard GARCH (see Berkes, Horváth, and Kokoszka, 2003; Francq and Zakoïan, 2004), but it is introduced here to circumvent the absence of any moments (Proposition 1), which constitutes the major difficulty of the proof of the next consistency result.

THEOREM 1. *Under Assumptions **A1** and **A3**– **A7**, we have*

$$\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0, \quad a.s.\ as\ n \to \infty.$$

We now turn to the asymptotic normality. We introduce the following additional assumptions.

A8 $\boldsymbol{\theta}_0 \in \overset{\circ}{\boldsymbol{\Theta}}$, where $\overset{\circ}{\boldsymbol{\Theta}}$ denotes the interior of $\boldsymbol{\Theta}$.

A9 $\kappa_\eta := \mathbb{E}\eta_t^4 < \infty$.

Denote by $\nabla_{\boldsymbol{\theta}}$ (resp. $\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}$) the partial derivative operator (resp. the second-order derivative operator) with respect to $\boldsymbol{\theta}$ (resp. $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$). Similarly, we denote by ∇_{θ_i} the partial derivative with respect to any component θ_i of $\boldsymbol{\theta}$.

A10 The functions $\phi_i(\cdot)$, for $i = 1, \ldots, q$, admit continuous second-order derivatives and the matrix $\left[\nabla_{\boldsymbol{\vartheta}}\phi_1(\boldsymbol{\vartheta}_0), \ldots, \nabla_{\boldsymbol{\vartheta}}\phi_q(\boldsymbol{\vartheta}_0)\right]$ has full-row rank.

A11 For $i = 1, \ldots, q$, either $\phi_i(\cdot) = 0$ or $\phi_i(\boldsymbol{\vartheta}_0) \neq 0$.

Assumption **A8** and **A9** are also made in the standard GARCH case. Note that **A10** and **A11** are satisfied in the cases of exponential and beta weights.

The next result establishes the asymptotic normality of the QMLE. Let $V_t(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta})\tau_t^2(\boldsymbol{\vartheta})$.

THEOREM 2. *Under the Assumptions of Theorem 1 and **A8–A11**,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\kappa_\eta - 1)\boldsymbol{J}^{-1}),$$

where

$$\boldsymbol{J} := \mathbb{E}\left(\frac{1}{V_t^2(\boldsymbol{\theta}_0)} \nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}_0) \nabla'_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}_0)\right) \tag{13}$$

is a positive-definite matrix.

Despite the absence of moments for the return process (which complicates the proof), the form of the asymptotic variance is thus the same as in the standard GARCH model (with obviously a multiplicative component volatility in the definition of $\boldsymbol{J}$).

4. TESTING THE EXISTENCE OF A LONG-RUN VOLATILITY

To test the existence of a long-term volatility component, i.e., the null hypothesis $H_0 : a_0 = 0$, usual tests such as the Wald test may have nonstandard asymptotic distributions due to the presence of the unidentified parameter $\boldsymbol{\vartheta}$ under the null. Indeed, it is known that in similar situations (see, e.g., Figure 1 in Francq, Horvàth, and Zakoïan, 2010) the Wald, score, and likelihood-ratio (LR) test statistics do not follow the standard distributions[6] under the null. To solve the problem, we consider two approaches. First, we fix the unidentified parameter to some value $\boldsymbol{\vartheta}^*$. This gives rise to test procedures which have standard, χ^2 or chi-bar-square, asymptotic distributions under the null, but whose power properties depend on the arbitrary choice of $\boldsymbol{\vartheta}^*$. We thus consider a second approach consisting in estimating by QMLE all the parameters, including the unidentified parameter $\boldsymbol{\vartheta}$, and estimating the critical value of the resulting Wald test by a residual-based bootstrap procedure. Note that the identifiability problem is not present in the framework of Conrad and Schienle (2020), in which a score-based test, not requiring the bootstrap, is developed.

4.1. Fixing $\boldsymbol{\vartheta}$

The first approach relies on the auxiliary model

$$\begin{cases} r_t = \tau_t \epsilon_t, & \tau_t^2 = 1 + a_0 \sum_{i=1}^q \phi_i(\boldsymbol{\vartheta}^*) r_{t-i}^2, \\ \epsilon_t = \sigma_t \eta_t, & \sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{cases} \tag{14}$$

where $\boldsymbol{\vartheta}^*$ is given, and the unknown parameter is $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0, a_0)'$. Let $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(\boldsymbol{\vartheta}^*) = (\widehat{\omega}_n, \widehat{\alpha}_n, \widehat{\beta}_n, \widehat{a}_n)'$ be the QMLE of $\boldsymbol{\theta}_0$. Denote also by $\widehat{\boldsymbol{\theta}}_G = (\widehat{\omega}^c, \widehat{\alpha}^c, \widehat{\beta}^c)'$ the QMLE of a standard GARCH(1,1) model. In other words, $\widehat{\boldsymbol{\theta}}_n^c = (\widehat{\boldsymbol{\theta}}_G', 0)'$ is the QMLE of $\boldsymbol{\theta}_0$ under H_0. Let $\boldsymbol{e}_i$ be the ith column of the 4×4 identity matrix. Let also $\widehat{\eta}_t = r_t / \widetilde{V}_t^{1/2}(\widehat{\boldsymbol{\theta}}_n)$, where $\widetilde{V}_t(\boldsymbol{\theta}) = \widetilde{\sigma}_t^2(\boldsymbol{\theta}) \widetilde{\tau}_t^2(\boldsymbol{\theta})$, and

$$\widehat{\eta}_t^c = r_t / \widetilde{V}_t^{1/2}(\widehat{\boldsymbol{\theta}}_n^c) = r_t / \widetilde{\sigma}_t(\widehat{\boldsymbol{\theta}}_G),$$

$\widehat{\kappa}_n = n^{-1} \sum_{t=1}^n |\widehat{\eta}_t|^4$, and $\widehat{\kappa}_n^c = n^{-1} \sum_{t=1}^n \left|\widehat{\eta}_t^c\right|^4$. The Wald, score, and LR test statistics are defined, respectively, by

$$\mathrm{W}_n = \frac{n}{\widehat{\kappa}_n - 1} \frac{\widehat{a}_n^2}{\boldsymbol{e}_4' \widehat{\boldsymbol{J}}_n^{-1} \boldsymbol{e}_4}, \qquad \widehat{\boldsymbol{J}}_n = \frac{1}{n} \sum_{t=1}^n \frac{1}{\widetilde{V}_t^2} \nabla_{\boldsymbol{\theta}} \widetilde{V}_t \nabla_{\boldsymbol{\theta}}' \widetilde{V}_t(\widehat{\boldsymbol{\theta}}_n),$$

$$\mathrm{R}_n = \frac{n}{\widehat{\kappa}_n^c - 1} \nabla_{\boldsymbol{\theta}}' \tilde{\mathbf{l}}_n(\widehat{\boldsymbol{\theta}}_n^c) \left(\widehat{\boldsymbol{J}}_n^c\right)^{-1} \nabla_{\boldsymbol{\theta}} \tilde{\mathbf{l}}_n(\widehat{\boldsymbol{\theta}}_n^c), \quad \widehat{\boldsymbol{J}}_n^c = \frac{1}{n} \sum_{t=1}^n \frac{1}{\widetilde{V}_t^2} \nabla_{\boldsymbol{\theta}} \widetilde{V}_t \nabla_{\boldsymbol{\theta}}' \widetilde{V}_t(\widehat{\boldsymbol{\theta}}_n^c),$$

[6] χ^2 for the score, chi-bar-square for the Wald, and LR statistics due to the positivity constraints on the estimator of a_0.

and

$$\mathrm{L}_n = 2\frac{n}{\widehat{\kappa}_n - 1}\left\{\tilde{\mathbf{l}}_n(\widehat{\boldsymbol{\theta}}_n^c) - \tilde{\mathbf{l}}_n(\widehat{\boldsymbol{\theta}}_n)\right\}.$$

Denote by χ_1^2 the chi-square distribution with one degree of freedom, and the chi-bar-square distribution $\frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$ that is the equal-weighted mixture of the Dirac measure at 0 and the χ_1^2 distribution. The following proposition gives the asymptotic distributions of the previous test statistics under the null.

PROPOSITION 2. *Assume **A1**, **A2**, **A3**, **A5**, **A7**, **A9** and that* $(\omega_0, \alpha_0, \beta_0)' \in \overset{\circ}{\boldsymbol{\Theta}}_G$, *where* $\overset{\circ}{\boldsymbol{\Theta}}_G$ *denotes the interior of the GARCH(1,1) parameter space* $\boldsymbol{\Theta}_G$, *a compact subset of* $(0,\infty)^2 \times [0,1)$. *Under* H_0, *we have* $\mathrm{W}_n \xrightarrow{\mathcal{L}} \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$, $\mathrm{R}_n \xrightarrow{\mathcal{L}} \chi_1^2$, *and* $\mathrm{L}_n \xrightarrow{\mathcal{L}} \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$ *as* $n \to \infty$.

We will see in the numerical section that the finite sample distributions of the test statistics are not always well approximated by their asymptotic laws. To solve the problem, we will approximate the test statistic distributions by means of a residual-based bootstrap procedure. Recent papers dealing with similar bootstrap inference procedures are Leucht, Kreiss, and Neumann (2015), Beutner, Heinemann, and Smeekes (2020), and Cavaliere, Nielsen, Pedersen, and Rahbek (2022).

Because the Wald test was found to be more powerful than the other tests in our Monte Carlo experiments, we present the resampling scheme and study its asymptotic behavior for the Wald-type statistic only. The algorithm is the following.

1. On the observations $r_1, \ldots, r_n$, compute the QMLE $\widehat{\boldsymbol{\theta}}_G = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})'$ of a GARCH(1,1) model and compute the standardized residuals (discarding the first n_0 values the alleviate the effect of the initial values) $\widehat{\eta}_t^0 = (\widehat{\eta}_t^c - m_n)/s_n$, for $t = n_0 + 1, \ldots, n$, where $\widehat{\eta}_t^c$, m_n, and s_n are, respectively, the nonstandardized GARCH residuals, their empirical mean and standard deviation. Denote by F_n the empirical distribution of these standardized residuals. Also compute the QMLE of the auxiliary GARCH-MIDAS model (14). Let $\widehat{a}_n$ be the estimator of the parameter a.
2. Simulate a trajectory of length n of a GARCH(1,1) model with parameter $\widehat{\boldsymbol{\theta}}_G$ and i.i.d. noise (η_t^*) with distribution F_n, compute the QMLE $\widehat{\boldsymbol{\theta}}_n^* = (\widehat{\omega}_n^*, \widehat{\alpha}_n^*, \widehat{\beta}_n^*, \widehat{a}_n^*)'$ of the GARCH-MIDAS model (14).
3. Repeat B times Step 2, and denote by $\widehat{a}_n^{*1}, \ldots, \widehat{a}_n^{*B}$ the bootstrap estimates of a. Approximate the p-value of the test $H_0 : a_0 = 0$ against $H_1 : a_0 > 0$ by $p_B^* = (1 + \#\{\widehat{a}_n^{*j} \geq \widehat{a}_n; j = 1, \ldots, B\})/(B+1)$.

To reduce the computational burden of bootstrap procedures, Kreiss et al. (2011) and Shimizu (2013) proposed to simulate the distribution of the (Q)MLE by using a Newton–Raphson-type iteration. This trick can not be used directly here because $\boldsymbol{\theta}_0$ belongs to the boundary of the parameter space under H_0, which implies that the Bahadur-type approximation

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \boldsymbol{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\eta_t^2 - 1\right) \frac{1}{V_t} \nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}_0) + o_P(1),$$

used for the Newton–Raphson iteration, is not valid when $a_0 = 0$. By the arguments of Francq and Zakoïan (2009), it can however be seen that in this case

$$\sqrt{n}\widehat{a}_n = \max\left\{\boldsymbol{e}_4' \boldsymbol{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\eta_t^2 - 1\right) \frac{1}{V_t} \nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}_0), 0\right\} + o(1) \quad \text{a.s.}$$

This suggests replacing $\widehat{a}_n^*$ in Step 2 by

$$\widehat{a}_n^* = \max\left\{\boldsymbol{e}_4' \left(\widehat{\boldsymbol{J}}_n^c\right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \left(\eta_t^{*2} - 1\right) \frac{1}{\widetilde{V}_t} \nabla_{\boldsymbol{\theta}} \widetilde{V}_t(\widehat{\boldsymbol{\theta}}_n^c), 0\right\}. \tag{15}$$

Since White (1982), it is known that the (Q)MLE of a misspecified model generally converges to some pseudo-true value. The resampling algorithm is valid in the following sense.

THEOREM 3. *Let the assumptions of Proposition 2 hold. Assume also that the distribution of η_t admits a bounded density with respect to the Lebesgue measure. Let $\widehat{a}_n^*$ defined by (15). Under H_0, for almost all realization (r_t), as $n \to \infty$ we have, given (r_t),*

$$\sqrt{n}\widehat{a}_n^* \xrightarrow{\mathcal{L}} N\mathbb{1}_{N\geq 0}, \quad N \sim \mathcal{N}\left(0, \sigma^2 := (\kappa - 1)\boldsymbol{e}_4 \boldsymbol{J}^{-1} \boldsymbol{e}_4\right), \tag{16}$$

and thus

$$\mathrm{W}_n^* := \frac{n}{\widehat{\kappa}_n - 1} \frac{\left(\widehat{a}_n^*\right)^2}{\boldsymbol{e}_4' \widehat{\boldsymbol{J}}_n^{-1} \boldsymbol{e}_4} \xrightarrow{\mathcal{L}} \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2.$$

Under $H_1 : a_0 > 0$, for almost all realization (r_t), if $\widehat{\boldsymbol{\theta}}_G$ converges to some pseudo-true value $\boldsymbol{\theta}_G \in \boldsymbol{\Theta}_G$ such that

$$\boldsymbol{J} := \mathbb{E} \frac{1}{V_t^2} \nabla_{\boldsymbol{\theta}} V_t \nabla_{\boldsymbol{\theta}} V_t' \begin{pmatrix} \boldsymbol{\theta}_G \\ 0 \end{pmatrix}$$

exists and is invertible and if $\widehat{a}_n \to a_0$ then $p^* \to 0$ as $n \to \infty$, where $p^* = \lim_{B\to\infty} p_B^*$ a.s.

The previous result thus shows that the distribution of $\widehat{a}_n^*$ (resp. W_n^*) given (r_t) well mimics the (unconditional) distribution of $\widehat{a}_n$ (resp. W_n) under H_0 when n is large. It is also expected that in finite samples the bootstrap distribution of $\sqrt{n\widehat{a}_n^*}$ better approaches the distribution of $\sqrt{n\widehat{a}_n}$ than its asymptotic distribution. The consistency of the bootstrap is also ensured as soon as $\liminf_{n\to\infty} \widehat{a}_n > 0$ and $\sqrt{n\widehat{a}_n^*} = O_P(1)$, which holds under the conditions of the theorem, but also under more general conditions.

4.2. Bootstrapping the Full Wald Test

The asymptotic properties of the test statistics defined in the previous section do not depend on the fixed value of the parameter $\boldsymbol{\vartheta}^*$ in (14). However, the illustrations presented in the numerical section show that the finite sample behavior of the tests depends on this parameter. In addition, there is no obvious choice of the parameter that one could recommend to the practitioner. When $\boldsymbol{\vartheta}$ is estimated by QMLE, together with the other parameters, the test statistics have nonstandard asymptotic distributions under the null, and the bootstrap techniques become particularly appealing. The resampling scheme is then modified as follows.

1. On the observations $r_1, \dots, r_n$, compute the GARCH(1,1) QMLE $\widehat{\boldsymbol{\theta}}_G = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})'$ and the standardized residuals $\widehat{\eta}_t^0 \sim F_n$, exactly as in the previous algorithm. Compute the QMLE of the GARCH-MIDAS model (4). Let $\widehat{a}_n$ be the estimator of the parameter a.
2. Simulate a trajectory of length n of a GARCH(1,1) model with parameter $\widehat{\boldsymbol{\theta}}_G$ and i.i.d. noise (η_t^*) with distribution F_n, compute the QMLE $\widehat{\boldsymbol{\theta}}_n^* = \left(\widehat{\omega}_n^*, \widehat{\alpha}_n^*, \widehat{\beta}_n^*, \widehat{a}_n^*, \widehat{\boldsymbol{\vartheta}}_n^{*'}\right)'$ of the GARCH-MIDAS model (4).
3. Repeat B times Step 2, and compute the bootstrap estimated p-value p_B^* exactly as in the previous algorithm.

Under H_0, the distribution of $\widehat{a}_n$ in the QML estimation of the full GARCH-MIDAS model (4) is an unknown function $G_n(\boldsymbol{\theta}_0, F)$ of the GARCH parameter $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0)'$ and the distribution F of the noise η_t. The previous residual bootstrap algorithm estimates $G_n(\boldsymbol{\theta}_0, F)$ by $G_n(\widehat{\boldsymbol{\theta}}_G, F_n)$. A formal justification, similar to that given in Theorem 3, would certainly rely on the strong consistency of $\widehat{\boldsymbol{\theta}}_G$ and on the consistency of F_n, in the sense of (A.16), and would require establishing a kind of continuity of $G_n(\cdot)$ and/or the asymptotic form of G_n as $n \to \infty$. To obtain the latter, techniques used to obtain the asymptotic distribution of sup-type test statistics, as in Hansen (1996), could be considered but the problem seems difficult because the parameter is on the boundary under the null hypothesis (see Andrews, 2001). To the best of our knowledge, there is no available result dealing with sup-tests when the parameter is on the boundary of the parameter set.

Note that the choice of B has little effect on the size and power of the test. Consider the test which rejects the null when $p_B^* \leq 5\%$. If $B = 19$ or $B = 99$, the size is exactly 5%. Note also that the bootstrap is a randomized procedure, in the sense that the statistical decision depends not only on the observations $r_1, \dots, r_n$, but also on the random bootstrap trials (for a formal definition, see, e.g., van der Vaart, 2000, p. 98). Taking a large value of B (we took $B = 999$ for the numerical illustrations of Section 5.2) has the advantage of reducing the test randomness. To assess the performance of the bootstrap test on Monte Carlo simulation experiments, the randomness of the procedure is not an issue. We thus follow the so-called "warp-speed" methodology of Giacomini, Politis, and White (2013) by computing $\widehat{a}_n$ on a large number K of Monte Carlo replications of a GARCH-MIDAS model (1). For each of the K Monte Carlo simulations, we generated $B = 1$ bootstrap simulation and computed the corresponding bootstrap statistic $\widehat{a}_n^*$. Let ξ_α^* be the α-quantile of the K values of $\widehat{a}_n^*$. The size (resp. power) of the bootstrap test of nominal level α is then approximated by the proportion of $\widehat{a}_n > \xi_{1-\alpha}^*$ over the K replications when $a_0 = 0$ (resp. $a_0 > 0$) in the simulated GARCH-MIDAS model.

5. NUMERICAL RESULTS

We first present the results of Monte Carlo experiments. Our objectives are twofold: (i) evaluating the effect of the absence of moments on the accuracy of the QMLE and (ii) assessing the performance of the QML in detecting and estimating the two volatility components. Then, we will present an application on real financial data.

5.1. Monte Carlo Experiments

The aim of our first Monte Carlo experiment is to study the effect of the absence or presence of marginal moments on the empirical accuracy of the QMLE. We simulated the simplest version of model (4) with $q = 1$, $\phi_i(\boldsymbol{\vartheta}) \equiv 1$ and parameter $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0, a_0)$ given in the column "True" of Table 1. For the first data generating process (DGP A), the noise η_t is $\mathcal{N}(0,1)$-distributed, so that **A2** is satisfied, and the DGP is stationary but does not admit any moment. For the second data generating process (DGP B), the noise η_t follows an equal-weighted mixture of $\mathcal{N}(m,1)$ and $\mathcal{N}(-m,1)$ distributions truncated on the interval $[-b,b]$, where m is chosen such that $\mathbb{E}\eta_t^2 = 1$ and $b = \sqrt{(1-\iota-\beta)/\alpha}$ with $0 < \iota < 1-\beta$. Since $a_t < 1-\iota$ a.s., we have $\epsilon_t^2 \leq b\omega/\iota$. If $\iota > ab\omega$, then $a\epsilon_t^2 < 1$, which entails that r_t is bounded. For DGP B, we took $\iota = 0.05$, so that $b = \sqrt{3}$, $0 < \iota < 1-\beta = 0.2$ and $\iota > ab\omega = 0.02\sqrt{3}$. This DGP thus admits moments of any order.

The number of replications of each simulation is $R = 1{,}000$, with sample sizes $n = 2{,}000$ and $n = 4{,}000$. The two DGPs have been estimated by QMLE. Table 1 displays the results of these Monte Carlo experiments. The columns "Min," "Q1," "Q2," "Q3," "Max," "Bias", and "RMSE" provide, respectively, the minimum, the first quartile, the median, the third quartile, the maximum, the bias, and the root mean square error (RMSE) of the R estimated values of the parameter. The column "MASE" refers to the estimated standard error based on the asymptotic theory. The ith Mean Asymptotic Standard Error (MASE) is defined as the empirical mean over the R replications of the estimated standard errors $\sqrt{\widehat{\boldsymbol{\Sigma}}(i,i)/n}$, where $\widehat{\boldsymbol{\Sigma}}$ is the empirical estimator of the asymptotic variance $\boldsymbol{\Sigma} = (\kappa_\eta - 1)\boldsymbol{J}^{-1}$ of the QMLE. As expected, bias and RMSE decrease when the sample size increases. The values of RMSE and MASE get closer as the sample size increases, which means that the empirical distribution of the estimator becomes closer to its asymptotic distribution. Unsurprisingly, the QMLE turns out to be more accurate when all moments exist (DGP B) than when there is no moment (DGP A), but the difference in accuracy is quite small.

TABLE 1. Distribution of the QMLE over 1,000 replications

n		True	Min	Q1	Q2	Q3	Max	Bias	RMSE	MASE
DGP A satisfying **A2** (no moments)										
2,000	ω	0.2	0.023	0.146	0.221	0.343	1.391	0.076	0.206	0.760
	α	0.05	0.000	0.037	0.054	0.082	0.240	0.015	0.045	0.043
	β	0.8	0.000	0.676	0.781	0.849	0.978	−0.064	0.174	0.642
	a	0.1	0.000	0.061	0.089	0.115	0.236	−0.012	0.044	0.046
4,000	ω	0.2	0.008	0.153	0.210	0.283	0.901	0.037	0.139	0.112
	α	0.05	0.000	0.038	0.052	0.068	0.253	0.007	0.031	0.024
	β	0.8	0.212	0.730	0.790	0.841	0.991	−0.031	0.120	0.098
	a	0.1	0.000	0.076	0.096	0.115	0.193	−0.005	0.032	0.029
DGP B that does not satisfy **A2** (moments at any order)										
2,000	ω	0.2	0.008	0.149	0.227	0.340	1.030	0.073	0.197	0.310
	α	0.05	0.000	0.038	0.055	0.081	0.205	0.014	0.041	0.040
	β	0.8	0.161	0.680	0.774	0.846	0.992	−0.061	0.167	0.256
	a	0.1	0.000	0.067	0.091	0.112	0.187	−0.012	0.039	0.039
4,000	ω	0.2	0.020	0.154	0.208	0.280	1.010	0.034	0.130	0.107
	α	0.05	0.005	0.041	0.051	0.066	0.222	0.006	0.027	0.023
	β	0.8	0.160	0.731	0.791	0.838	0.975	−0.028	0.111	0.093
	a	0.1	0.000	0.081	0.097	0.113	0.181	−0.005	0.026	0.025

In a second set of Monte Carlo experiments, we assess the ability of our estimation approach to estimate and detect the presence of long-term volatility. We chose to estimate the GARCH-MIDAS specification of τ_t in (1), with beta weights given by (3). We thus simulated 1,000 trajectories of size $n = 4{,}000$ of Model (1) with $N = 22$, $Q = 250$, and $(\omega_0, \alpha_0, \beta_0, \vartheta_0, a_0) = (0.028, 0.115, 0.831, 2.067, 0.056)$.[7] For the distribution of η_t, we took a standardized Student distribution with $\nu = 5.41$ degrees of freedom.[8] The estimation results are presented in the top panel of Table 2. Interestingly, the parameter a_0 is estimated with a small bias, and its estimated standard deviation is on average very close to the observed RMSE. We have redone the estimation exercise on simulations of a standard GARCH (corresponding to Model (4) with $a_0 = 0$). The bottom panel of Table 2 shows that at least one half of the estimated values of a are exactly equal to 0. Unsurprisingly, the estimations of ϑ, whose true value is undefined when $a_0 = 0$, are erratic. Figure 2 displays

[7]These parameters are those estimated on the NASDAQ index considered in Section 5.2, with RVs computed over 1 month and 1 MIDAS lag year, on a set of historical data of size $n = 12{,}654$ (for our simulations, we consider the smallest sample size $n = 4{,}000$).

[8]The kurtosis thus corresponds to the empirical kurtosis of the residuals of the model fitted to the NASDAQ series.

Table 2. Distribution of the QMLE of a GARCH-MIDAS, when the DGP is a GARCH-MIDAS (first part) and when it is a standard GARCH (second part of the table)

	True	Min	Q1	Q2	Q3	Max	Bias	RMSE	MASE
ω	0.028	0.009	0.025	0.033	0.042	0.139	0.007	0.016	0.013
α	0.115	0.057	0.103	0.116	0.128	0.203	0.001	0.019	0.020
β	0.831	0.572	0.804	0.828	0.846	0.922	−0.008	0.037	0.033
ϑ	2.067	0.000	1.447	2.122	3.173	68.541	0.743	3.329	5.925
a	0.056	0.000	0.030	0.045	0.064	0.256	−0.005	0.033	0.032
ω	0.028	0.010	0.024	0.028	0.034	0.084	0.002	0.009	0.009
α	0.115	0.062	0.105	0.116	0.129	0.204	0.002	0.020	0.021
β	0.831	0.534	0.802	0.823	0.841	0.908	−0.012	0.038	0.033
ϑ	UD	0.000	2.067	2.067	2.067	4,650.425	10.653	207.455	199.696
a	0	0.000	0.000	0.000	0.013	0.116	0.010	0.020	0.022

Note: In the latter case, the parameter ϑ is undefined (UD).

a typical example of estimates of the short- and long-term volatilities of the two DGPs of Table 2. The distinction between the dynamics of the two DGPs is clear from the figure, and can be confirmed by a formal test of the null hypothesis $H_0 : a_0 = 0$. Figure 3 shows that the estimation of the volatilities is fortunately not too sensitive to the choice of the integers N and Q in (1). Finally, we estimated a (misspecified) standard GARCH(1,1) on simulations of a GARCH-MIDAS (with same parameters as in the first part of Table 2). Table 3 presents the estimation results. The columns "Mean" and "SD" stand for the mean and standard deviation of the estimates over the 1,000 replications. It can be noted that the estimated value of $\alpha+\beta$ is always very close to 1, a stylized fact that is often observed on real series. Over a small sub-period of a randomly chosen simulation, Figure 4 graphically compares the volatility estimates obtained by the correctly specified GARCH-MIDAS model with those obtained by the misspecified standard GARCH(1,1). Even if the volatility estimation of the standard GARCH is, as expected, dominated by the GARCH-MIDAS estimation, the difference is not huge. Table 4 confirms that the estimates obtained from the GARCH-MIDAS model are indeed better, but only slightly better, than those obtained from the GARCH model, as measured by the QLIK loss defined by

$$QLIK = \frac{1}{n}\sum_{t=r_0+1}^{n} \frac{V_t^2}{\widehat{V}_t^2} + \log \widehat{V}_t^2,$$

where V_t denotes the true volatility and $\widehat{V}_t$ denotes the estimated volatility (for the GARCH or the GARCH-MIDAS). We took $r_0 = 100$ to avoid the effect of the initial values required to compute the volatility estimates. The reader is

TABLE 3. Distribution of the QMLE of a GARCH(1,1) when the DGP is the GARCH-MIDAS of Table 2 (top panel)

	Min	Q1	Q2	Q3	Max	Mean	SD
ω	0.004	0.025	0.033	0.043	0.131	0.036	0.015
α	0.027	0.092	0.105	0.118	0.170	0.105	0.019
β	0.747	0.857	0.873	0.890	0.969	0.872	0.026
$\alpha+\beta$	0.876	0.968	0.979	0.989	1.015	0.977	0.017

TABLE 4. Distribution of the QLIK losses over 1,000 replications when the GARCH-MIDAS volatility is estimated by the GARCH-MIDAS model or by the GARCH model

Model	Min	Q1	Q2	Q3	Max	Mean	SD
MIDAS	0.589	0.956	1.101	1.265	3.285	1.133	0.263
GARCH	0.591	0.959	1.108	1.272	3.301	1.139	0.265

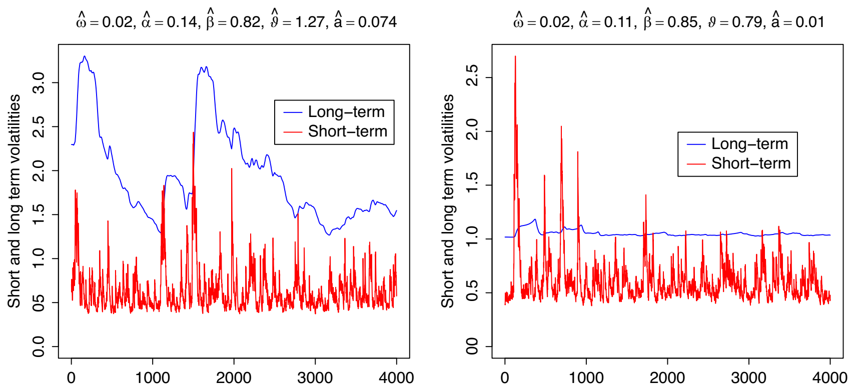

FIGURE 2. Examples of estimated short- and long-term volatilities when the GDP is a GARCH-MIDAS (left) or a standard GARCH (right), with $N = 22$ and $Q = 250$.

referred to Patton (2011) for arguments in favor of the QLIK loss to compare volatility forecasts/estimates. We did not use the MSE loss because we know from Proposition 1 that σ_t^2 does not admit any moment.

Table 5 gives the empirical relative frequency of rejection of the score, Wald, and LR tests of Section 4.1 for the null of no long-run volatility. The DGP is that used in Table 2, except that $a_0 = 0$ (under the null) or $a_0 \in \{0.01, 0.05\}$ (under the alternative). The number of replications is 1,000. Different values of $\vartheta \geq 1$ are used. With $\vartheta = 1$, all the RVs involved in (1) have the same weight; the larger ϑ, the higher the weights of the most recent RVs. It can be seen from this table that the three tests are conservative, but the size is better controlled with the score test.

TABLE 5. Empirical relative frequency of rejection of the null that there exists no long-run volatility (i.e., $a_0 = 0$) using the score, Wald, and LR tests with a fixed value of ϑ, for nominal levels varying from 0.1% to 20%

a_0	ϑ	Test	0.1%	1%	2%	3%	4%	5%	6%	7%	10%	20%
0	1	R_n	0	1	2	3.3	3.9	4.9	6.2	6.9	10.1	18.9
		W_n	0	0	0	0	0	0	0.2	0.4	1.6	9.7
		L_n	0	0.2	0.6	1.4	1.8	1.9	2	2.5	3.2	6.6
	2	R_n	0	0.8	1.5	3	4.1	4.9	5.4	6.6	9.2	19.2
		W_n	0	0	0	0	0.1	0.1	0.3	0.6	2.1	8.8
		L_n	0	0.2	0.6	1	1.8	2.2	2.6	2.7	3.4	6.1
	3	R_n	0	0.4	1.5	2.9	3.8	4.8	5.8	6.6	9.3	18.6
		W_n	0	0	0	0.1	0.1	0.4	0.7	1.2	3.1	10.1
		L_n	0	0.3	0.6	1.2	1.6	2.1	2.4	2.6	3.3	5.9
	9	R_n	0.3	0.8	1.5	2.1	3.2	3.6	4.1	5.1	7.5	15.9
		W_n	0	0	0.5	1.4	1.7	2.3	3.2	3.6	5.8	12.6
		L_n	0	0.6	1.1	1.8	2.2	2.4	2.7	3.2	3.9	6.7
0.01	1	R_n	0.6	2.5	3.8	4.8	5.3	6.8	7.7	8.6	11.5	20.8
		W_n	0	0	0.1	0.3	0.4	0.6	1.3	1.9	4.9	20.6
		L_n	0.3	1.9	2.6	3.5	4.3	5	5.6	6.5	8.9	14.6
	2	R_n	0.2	2.2	3.5	4.7	5.6	6.4	7.6	8.5	11.3	20.8
		W_n	0	0.1	0.2	0.3	0.6	1.1	2	3	7.3	22.7
		L_n	0.5	2.1	3.4	4.1	4.6	5.6	6.3	6.9	9.4	16.4
	3	R_n	0.4	2.2	3.1	3.9	4.7	5.4	6.2	7.2	10.1	20.3
		W_n	0	0.1	0.2	0.4	1.1	1.9	3	4.5	8.8	24.3
		L_n	0.3	2.2	3	4.1	4.9	5.4	6	6.6	9.5	15.7
	9	R_n	0.4	0.9	1.2	1.4	2.1	3.1	3.5	4	6	15
		W_n	0	0.4	1.2	2.8	4.1	5.5	7.1	8.3	12.2	25.2
		L_n	0.3	2	2.9	3.7	4.7	5.4	5.8	6.3	7.9	12.6
0.05	1	R_n	11.1	29.4	37.4	43.2	48.1	52.8	55.4	57.1	62.4	73.3
		W_n	0.1	1.3	4.6	9.5	17.2	25.1	34	42.1	61.2	84.4
		L_n	17.3	38.2	46.9	52.6	57.6	60.4	63.2	65.2	70.9	81
	2	R_n	9.2	24.2	32.6	38.6	41	43.8	47	50.3	55.2	65.6
		W_n	0	1.4	7.2	17.9	29.3	38.2	48.1	55.3	71.3	90.2
		L_n	24.4	46.8	54.9	60.6	65	67.7	70.2	71.8	76.7	85
	3	R_n	4.1	14.6	20.1	23.4	26.7	28.7	31	33.3	38.6	51.3
		W_n	0	2	11.8	25.9	38.1	47.7	55.2	61.9	74.4	90.6
		L_n	22.4	44.6	53.5	58.5	61.5	64.3	67.5	69.6	74.5	83.1
	9	R_n	0.6	2.6	3.4	4.9	6.1	6.9	8.2	9.5	12.8	23.2
		W_n	0	13.6	33.4	44.1	52.4	58.4	62	66.2	73.2	84.9
		L_n	8.8	21.6	27.7	32.7	36.5	39.9	43.3	45.1	51.5	62.3

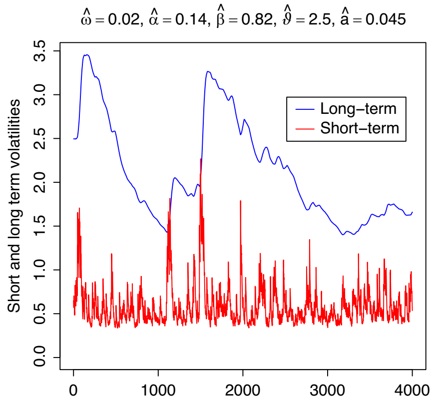

FIGURE 3. As Figure 2 (left), with $N = 44$ and $Q = 500$.

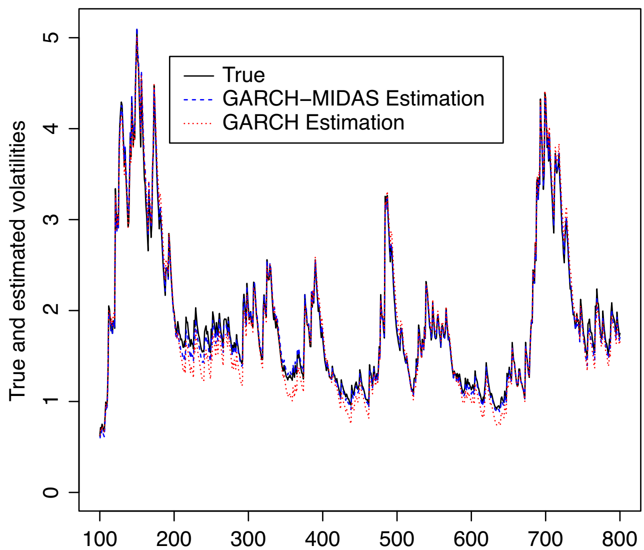

FIGURE 4. True and estimated volatility estimated by a GARCH-MIDAS and by a standard GARCH.

It can also be seen that the ranking of the three tests, in terms of power, vary a lot with a_0, the nominal level, and the parameter ϑ. Other numeric experiments, not presented here, show that the Wald test seems slightly more powerful than the two other tests when the sample size n is larger. The poor control of the error of the first kind, as well as the sensitivity to the choice of the fixed parameter ϑ, motivated us to consider the bootstrapped Wald test of Section 4.2. Table 6 shows

Table 6. Empirical relative frequency of rejection of the null that there exists no long-run volatility (i.e., $a_0 = 0$) using the bootstrapped version of the Wald test, for nominal levels varying from 0.1% to 20%

a_0	0.1%	1%	2%	3%	4%	5%	6%	7%	10%	20%
0	0.2	1.2	3.2	3.6	4.3	4.9	6.5	7.7	10.9	23.4
0.01	0.0	1.1	4.2	7.8	10.2	11.5	13.8	15.6	20.5	36.5
0.05	0.5	5.3	20.7	32.4	45.9	56.2	61.2	66.2	75.2	93.1

that this bootstrap test much better controls the error of the first kind, without degrading the power. Note that these empirical sizes and powers are obtained from the warp-speed methodology of Giacomini, Politis, and White (2013), as explained in Section 4.2, with $K = 1,000$.

5.2. Application to Stock Indices

We estimated the GARCH-MIDAS model (1) with exponential weights on the daily returns of the CAC 40, DAX, NASDAQ, and Hang Seng indices, from March 1, 1990 to April 8, 2021. Table 7 displays the estimated coefficients when $N = 65$ (corresponding to RVs over a quarter) and $Q = 1,000$ (corresponding to 4 MIDAS lag years). These values were advocated by Engle, Ghysels, and Sohn (2013). We checked that the short- and long-term volatilities are not much modified with other choices of these parameters (in particular with biannual rolling window RV, i.e., $N = 125$, and 2 MIDAS lag years, or with $N = 22$ and $Q = 250$, i.e., RVs over 1 month and 1 MIDAS lag year). The last column of Table 7 displays the estimated p-values of the bootstrap Wald test of Section 4.2 (with $B = 999$). The most noticeable output of that table is that these p-values are small and the estimated value of a is always clearly significant, except perhaps for the HSI series, showing the existence of time-varying long-term volatilities. Figure 5 confirms that the GARCH-MIDAS parameter estimate $\widehat{a}_n$ is well on the right of its estimated distribution under the null $H_0 : a = 0$. The latter distribution, which is a mixture of a Dirac mass at 0 and a continuous distribution on $(0,\infty)$, has been estimated by a Kernel density estimator (using the reflection method for boundary correction). Figure 6 displays the estimated short- and long-term volatilities. The most striking feature of this figure is that long-term volatility varies strongly, but as expected slowly, over time. The volatilities of the CAC and DAX indices are surprisingly similar, with in particular a strong increase in long-term volatility after the 2008 crisis and the recent COVID-19 crisis. The Nasdaq behaves similarly in the most recent period, but reacted much more to the 2001 recession. The HSI behaves quite differently, with an increase in long-term volatility after the Asian Crisis of 1997 and after the Global Financial Crisis of 2008, but with little response to the COVID-19 pandemic.

Table 7. GARCH-MIDAS fitted on stock returns

	ω	α	β	ϑ	a	p-value
CAC	0.031 0.007	0.110 0.011	0.846 0.017	16.308 6.656	0.013 0.005	0.003
DAX	0.027 0.008	0.095 0.012	0.867 0.018	11.724 5.729	0.012 0.005	0.010
NASDAQ	0.026 0.005	0.113 0.011	0.840 0.015	10.813 3.227	0.017 0.005	0.001
HSI	0.034 0.009	0.080 0.011	0.884 0.016	11.316 5.889	0.008 0.003	0.031

Note: The estimated standard deviations are displayed in small font, under the estimated values of the coefficients. The last column gives the bootstrap estimated p-value of the Wald test of $H_0 : a = 0$.

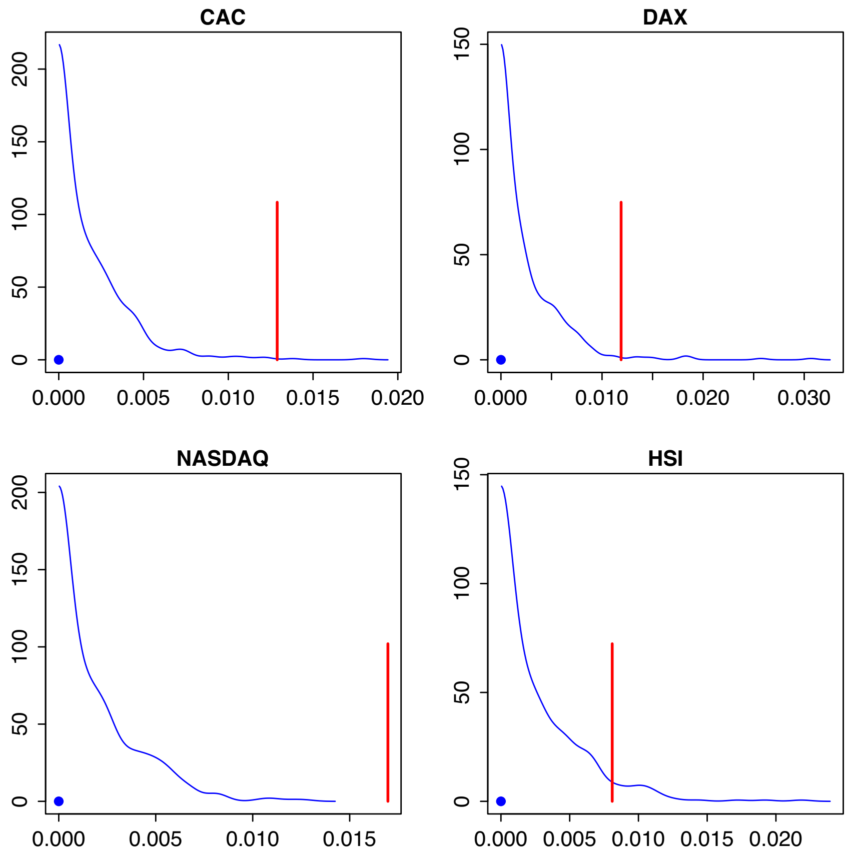

Figure 5. Bootstrap estimate of the distribution of $\widehat{a}_n$ when $a = 0$ (in blue) and observed value of $\widehat{a}_n$ (red vertical line).

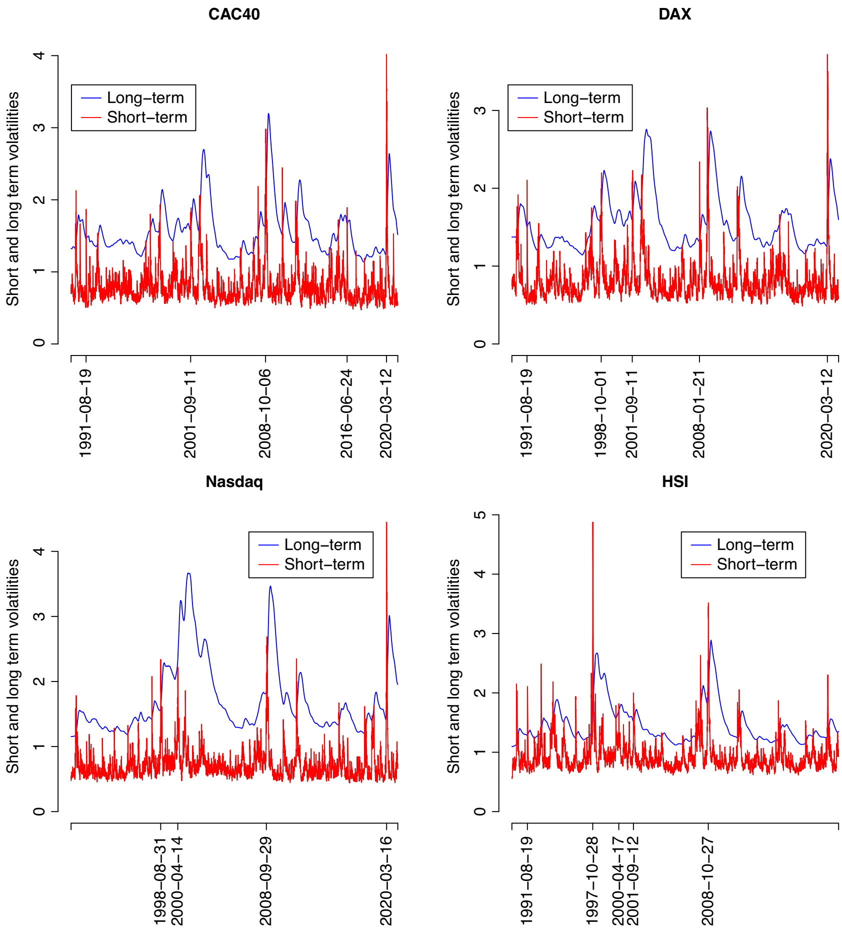

FIGURE 6. GARCH-MIDAS short- and long-term volatilities for four stock indices from March 1, 1990 to April 8, 2021.

6. CONCLUSION

In this article, we studied a class of models enabling long- and short-run volatilities. We showed that strictly stationary solutions are so heavy tailed that not even a small power moment exists. The main theoretical novelty with respect to the literature on GARCH estimation comes from showing that strong consistency and asymptotic normality of the QMLE hold despite the absence of moments. We also proposed tests of the existence of a long-run volatility component. Our numerical applications illustrated the ability of the QML to distinguish and accurately estimate the two components in finite sample, but also confirmed that a misspecified GARCH model can deliver reliable estimates of volatility. Other specifications of the long-run variance could be considered in further work, in particular, those including exogenous variables (such as macroeconomic factors) in the dynamics of τ_t, as in Conrad and Loch (2015), or Conrad and Schienle (2020) among many others. Recently, Ibragimov, Kim, and Skrobotov (2023) proposed a robust inference method for predictive regression models in which the error term can follow a two-factor volatility model, but their long-run component is nonstationary, in contrast with the GARCH-MIDAS studied in this paper.

GARCH-MIDAS is a complex model and several difficult questions remain open. In particular, does stationarity entail the existence of log-moments? It seems difficult to conjecture the result. On the one hand, it can be shown that (12) is equivalent to the existence of a finite log-moment when (r_t^2) is i.i.d. and bounded away from 0.[9] On the other hand, Tanny (1974) provided an example of stationary and ergodic sequence (r_t) where (12) is true and the log-moment is infinite. At least from a theoretical point of view, it would be interesting to know if this is also the case for GARCH-MIDAS processes. Other interesting questions concern the practical implications of the absence of moments. Starting from the general principle that it is better for a model to share the same characteristics as the data to which it applies, the question is whether financial returns (or other real-time series) are devoid of a finite moment. It is too difficult a problem to solve here. It seems from our experiments that the existence of moments might not be detectable from the trajectories (see the graph in Example 1). Another interesting question raised by a referee is the behavior of sample autocorrelations in the absence of theoretical autocorrelations. From Davis and Resnick (1986), the empirical ACF of an AR(1) with heavy-tailed i.i.d. innovations is known to converge to the AR coefficient. More recently, Ibragimov, Pedersen, and Skrobotov (2021) derived the asymptotic distribution of empirical autocorrelations of powers of absolute returns under heavy-tailed assumptions. Do these results hold true when innovations follow the GARCH-MIDAS model? The numerical experiments we have done lead us to believe that the convergence holds but proving the result is beyond the scope of this paper.

A. APPENDIX: Proofs

Proof of Lemma 1

Let $\mathcal{F}_t$ be the sigma-field generated by $\{\eta_u, u \leq t\}$. Let $\mu_p = \mathbb{E}|\eta_1|^p$ for any $p > 0$. Note that $\mu_{p_i} \in (0, \infty]$ because $\mu_2 = 1$ implies that $|\eta_1|$ can not be equal to 0 with probability 1. Without loss of generality, assume $i_2 \geq 1$. We can also assume $\mu_{p_1} < \infty$, otherwise the result is trivial.

Since $\sigma_t \geq \alpha_0^{1/2}|\epsilon_{t-1}|$, for all positive random variable $X_{t-2} \in \mathcal{F}_{t-2}$, we have

$$\mathbb{E}|\epsilon_{t-1}|^{p_1} X_{t-2} = \mu_{p_1} \mathbb{E}\sigma_{t-1}^{p_1} X_{t-2} \geq \mu_{p_1} \alpha_0^{\frac{p_1}{2}} \mathbb{E}|\epsilon_{t-2}|^{p_1} X_{t-2}.$$

[9]Using (i)–(iii) in Tanny (1974)

By successive applications of this inequality, it follows that

$$\begin{aligned}&\mathbb{E}|\epsilon_{t-i_1}|^{p_1}|\epsilon_{t-i_1-i_2}|^{p_2}\dots|\epsilon_{t-i_1-\dots-i_k}|^{p_k}\\ &\geq \left(\mu_{p_1}\alpha_0^{\frac{p_1}{2}}\right)^{i_2}\mathbb{E}|\epsilon_{t-i_1-i_2}|^{p_1+p_2}|\epsilon_{t-i_1-i_2-i_3}|^{p_3}\dots|\epsilon_{t-i_1-\dots-i_k}|^{p_k}.\end{aligned}$$

Iterating the argument, we obtain the result with

$$K=\left(\mu_{p_1}\alpha_0^{\frac{p_1}{2}}\right)^{i_2}\left(\mu_{p_1+p_2}\alpha_0^{\frac{p_1+p_2}{2}}\right)^{i_3}\cdots\left(\mu_{p_1+\dots+p_{k-1}}\alpha_0^{\frac{p_1+\dots+p_{k-1}}{2}}\right)^{i_k}.$$

Under **A2**, $\mathbb{E}|\epsilon_t|^{2r}=\infty$ for r large enough and the conclusion follows.

Proof of Lemma 2

First note that $\boldsymbol{X}_t=\boldsymbol{r}_t$ satisfies Stochastic Iterated Function Systems (2.1) in Kandji (2023), with $\theta_t=(\boldsymbol{A}_t,\boldsymbol{b}_t)$. Note that the assumptions of Theorem 2.1 in Kandji (2023) are satisfied: (i) is satisfied because GARCH possess small order moments, (ii) is satisfied with Λ_t the operator norm of $\boldsymbol{A}_t$, and (iii) is satisfied under **A3**. The lemma is thus a consequence of Theorem 2.2 in Kandji (2023), since $\log r_{t+k}^2\leq\frac{1}{2}\log d(\boldsymbol{r}_{t+k},c)$ with $c=0$ and d the euclidean norm.

Proof of Proposition 1

The strictly stationary solution is obtained from (5) and (8), by taking r_t equal to the square-root of the first component of $\boldsymbol{r}_t$ multiplied by the sign of η_t. Now, let i_0 such that $\phi_0=\phi_{i_0}(\boldsymbol{\vartheta}_0)>0$. We have

$$\tau_t^2\geq 1+a_0\phi_0\epsilon_{t-i_0}^2\tau_{t-i_0}^2=1+a_0\phi_0\epsilon_{t-i_0}^2+a_0^2\phi_0^2\epsilon_{t-i_0}^2\epsilon_{t-2i_0}^2+\cdots.$$

We thus have $|r_t|^s\geq(a_0\phi_0)^{ks/2}|\epsilon_t|^s|\epsilon_{t-i_0}|^s\cdots|\epsilon_{t-ki_0}|^s$ for any $s>0$ and any $k\geq 1$. By **A2**, for any $s>0$, there exists $k\geq 1$ such that $\mathbb{E}|\epsilon_t|^{ks}=\infty$. The conclusion follows from Lemma 1.

Proof of Theorem 1

Let

$$\mathbf{l}_n(\boldsymbol{\theta})=\frac{1}{n}\sum_{t=1}^n\ell_t(\boldsymbol{\theta}),\qquad \ell_t(\boldsymbol{\theta})=\frac{r_t^2}{\tau_t^2\sigma_t^2}+\log\sigma_t^2+\log\tau_t^2,$$

where $\tau_t^2=\tau_t^2(\boldsymbol{\theta})=1+a\sum_{i=1}^q\phi_i(\boldsymbol{\vartheta})r_{t-i}^2$ and $\sigma_t^2=\sigma_t^2(\boldsymbol{\theta})=\omega+\alpha\epsilon_{t-1}^2(\boldsymbol{\theta})+\beta\sigma_{t-1}^2$, with $\epsilon_t^2(\boldsymbol{\theta})=r_t^2/\tau_t^2$. Note that σ_t^2 is well defined because

$$\sum_{k=0}^\infty\beta^k\epsilon_{t-k-1}^2(\boldsymbol{\theta})\leq\sum_{k=0}^\infty\beta^k r_{t-k-1}^2<\infty,\ \ a.s.$$

Since $\limsup_{k\to\infty} \frac{1}{k}\log\left(\beta^k r_{t-k-1}^2\right) \le \log\beta < 0$ by the second inequality in (12), the convergence of the latter infinite sum follows from the Cauchy rule.

However, contrary to the standard GARCH case, the limiting criterion $\mathbb{E}\ell_t(\boldsymbol{\theta})$ might not be defined, even at $\boldsymbol{\theta}_0$, because if **A2** holds the observed process has no moment.

The proof therefore relies on the following intermediate results which, contrary to the standard GARCH case (see, for instance, Francq and Zakoïan, 2019, Sect. 7.4), do not involve a limiting criterion:

(i) $\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbf{l}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta})| = 0, \quad a.s.,$

(ii) if $\sigma_t^2(\boldsymbol{\theta})\tau_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta}_0)\tau_t^2(\boldsymbol{\theta}_0)$ a.s., then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

(iii) if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then $\mathbb{E}\{\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)\} > 0$,

(iv) any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ has a neighborhood $V(\boldsymbol{\theta})$ such that

$$\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0)\right) > 0 \quad a.s.$$

We first show *(i)*. We have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbf{l}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta})|$$
$$\le \frac{1}{n}\sum_{t=1}^n \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{\left|\log\left(\frac{\sigma_t^2}{\tilde{\sigma}_t^2}\right)\right| + r_t^2\frac{|\sigma_t^2 - \tilde{\sigma}_t^2|}{\tilde{\tau}_t^2\sigma_t^2\tilde{\sigma}_t^2} + \left|\log\left(\frac{\tau_t^2}{\tilde{\tau}_t^2}\right)\right| + r_t^2\frac{|\tau_t^2 - \tilde{\tau}_t^2|}{\tilde{\tau}_t^2\tau_t^2\sigma_t^2}\right\}.$$

Noting that $\tau_t^2 = \tilde{\tau}_t^2$ for $t > q$, the last two terms asymptotically vanish and we have, for $t > q$,

$$|\sigma_t^2 - \tilde{\sigma}_t^2| \le \beta|\sigma_{t-1}^2 - \tilde{\sigma}_{t-1}^2| \le \beta^{t-q}|\sigma_q^2 - \tilde{\sigma}_q^2|. \quad \textbf{(A.1)}$$

Using the inequality $|\log(x/y)| \le |x-y|/(x\vee y)$ for $x, y > 0$, we deduce

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbf{l}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta})| \le \frac{K}{n}\sum_{t=1}^n \rho^t(1 + r_t^2),$$

where $\rho = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\beta < 1$ and, in view of (A.1), K is $\mathcal{F}_q$-measurable random variable. By the first inequality in (12), we have

$$\limsup_{t\to\infty}\frac{1}{t}\log\rho^t r_t^2 = \log\rho + \limsup_{t\to\infty}\frac{1}{t}\log r_t^2 \le \log\rho < 0, \quad \text{a.s.}$$

from which it follows that $\rho^t r_t^2 \to 0$, and then $\rho^t(1 + r_t^2) \to 0$, a.s. as $t \to \infty$. By Cesàro's lemma $\frac{1}{n}\sum_{t=1}^n \rho^t(1 + r_t^2) \to 0$, a.s. Because K is fixed (independent of n), the conclusion follows.

Next, we turn to *(ii)*. Letting $V_t(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta})\tau_t^2(\boldsymbol{\theta})$, we have

$$V_t(\boldsymbol{\theta}) = \left\{\omega + \alpha\frac{V_{t-1}(\boldsymbol{\theta}_0)}{\tau_{t-1}^2(\boldsymbol{\theta})}\eta_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta})\right\}\left\{1 + a\phi_1(\boldsymbol{\vartheta})V_{t-1}(\boldsymbol{\theta}_0)\eta_{t-1}^2 + a\sum_{i=2}^q \phi_i(\boldsymbol{\vartheta})r_{t-i}^2\right\}$$
$$:= b_{t-1}(\boldsymbol{\theta})\eta_{t-1}^4 + c_{t-1}(\boldsymbol{\theta})\eta_{t-1}^2 + d_{t-1}(\boldsymbol{\theta}),$$

where $b_{t-1}(\boldsymbol{\theta}), c_{t-1}(\boldsymbol{\theta}), d_{t-1}(\boldsymbol{\theta}) \in \mathcal{F}_{t-2}$. By Assumption **A5**, $V_t(\boldsymbol{\theta}) = V_t(\boldsymbol{\theta}_0)$ entails $b_{t-1}(\boldsymbol{\theta}) = b_{t-1}(\boldsymbol{\theta}_0), c_{t-1}(\boldsymbol{\theta}) = c_{t-1}(\boldsymbol{\theta}_0)$, and $d_{t-1}(\boldsymbol{\theta}) = d_{t-1}(\boldsymbol{\theta}_0)$. First, consider the case $\phi_1(\boldsymbol{\vartheta}_0) \neq 0$. The equality $b_{t-1}(\boldsymbol{\theta}) = b_{t-1}(\boldsymbol{\theta}_0)$ then implies

$$\frac{\tau_{t-1}^2(\boldsymbol{\theta})}{\tau_{t-1}^2(\boldsymbol{\theta}_0)} = \frac{a\alpha\phi_1(\boldsymbol{\vartheta})}{a_0\alpha_0\phi_1(\boldsymbol{\vartheta}_0)} := c. \tag{A.2}$$

Now $\tau_{t-1}^2(\boldsymbol{\theta}) = c\tau_{t-1}^2(\boldsymbol{\theta}_0)$ writes

$$\sum_{i=1}^{q}\{a\phi_i(\boldsymbol{\vartheta}) - ca_0\phi_i(\boldsymbol{\vartheta}_0)\}V_{t-i}(\boldsymbol{\theta}_0)\eta_{t-i}^2 = c-1$$

which, because $V_{t-i}(\boldsymbol{\theta}_0) > 0$ and by already given arguments, entails $a\phi_i(\boldsymbol{\vartheta}) = a_0\phi_i(\boldsymbol{\vartheta}_0)$, for $i = 1, \ldots, q$ and $c = 1$. Because the $\phi_i(\cdot)$'s sum up to 1, we deduce $a = a_0$ and then, by Assumptions **A4** and **A6**, $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. By (A.2), we also have $\alpha = \alpha_0$. In view of $c_{t-1}(\boldsymbol{\theta}) = c_{t-1}(\boldsymbol{\theta}_0)$, we obtain $\omega = \omega_0$. In view of $d_{t-1}(\boldsymbol{\theta}) = d_{t-1}(\boldsymbol{\theta}_0)$, we get $\beta\sigma_{t-1}^2(\boldsymbol{\theta}) = \beta_0\sigma_{t-1}^2(\boldsymbol{\theta}_0)$ from which we deduce $\beta = \beta_0$ by already given arguments. Now consider the case $\phi_1(\boldsymbol{\vartheta}_0) = 0$. The equality $b_{t-1}(\boldsymbol{\theta}) = b_{t-1}(\boldsymbol{\theta}_0)$ then implies $\phi_1(\boldsymbol{\vartheta}) = 0$, and $c_{t-1}(\boldsymbol{\theta}) = c_{t-1}(\boldsymbol{\theta}_0)$ in turn implies $\tau_{t-1}^2(\boldsymbol{\theta}_0) = c\tau_{t-1}^2(\boldsymbol{\theta})$ with $c = \alpha_0/\alpha$, which allows us to conclude by the previous arguments. Step *(ii)* is thus established.

Turning to *(iii)*, let $W_t(\boldsymbol{\theta}) = V_t(\boldsymbol{\theta}_0)/V_t(\boldsymbol{\theta})$ and, for $K > 0$, $A_K = [K^{-1}, K]$, write

$$\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0) = g(W_t(\boldsymbol{\theta}), \eta_t^2)\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K} + g(W_t(\boldsymbol{\theta}), \eta_t^2)\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K^c},$$

where, for $x > 0, y \geq 0$, $g(x,y) = -\log x + y(x-1)$. Introducing the negative part $x^- = \max(-x, 0)$ of any real number x, we thus have

$$\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0) \geq g(W_t(\boldsymbol{\theta}), \eta_t^2)\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K} - \left\{g(W_t(\boldsymbol{\theta}), \eta_t^2)\right\}^- \mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K^c}. \tag{A.3}$$

The expectation of the first term in the r.h.s. is well defined and satisfies

$$\mathbb{E}[g(W_t(\boldsymbol{\theta}), \eta_t^2)\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K}] = \mathbb{E}[g(W_t(\boldsymbol{\theta}), 1)\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K}] \geq 0,$$

since $g(x,1) \geq 0$ for any $x \geq 0$, with equality only if $x = 1$. By *(ii)*, we have that $W_t(\boldsymbol{\theta}) = 1$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. We thus have, by Beppo Levi's theorem,

$$\begin{aligned}\lim_{K\to\infty} \mathbb{E}[g(W_t(\boldsymbol{\theta}), \eta_t^2)\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K}] &= \mathbb{E}[g(W_t(\boldsymbol{\theta}), 1)\lim_{K\to\infty}\mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K}] \\ &= \mathbb{E}[g(W_t(\boldsymbol{\theta}), 1)] > 0 \quad \text{for} \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.\end{aligned}$$

To deal with the expectation of the second term in the r.h.s. of (A.3), we use the fact that for $y > 0$, $g(x,y) \geq g(1/y, y)$. It follows that

$$\begin{aligned}-\mathbb{E}\left[\left\{g(W_t(\boldsymbol{\theta}), \eta_t^2)\right\}^- \mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K^c}\right] &\geq -\mathbb{E}\left[\left\{g(1/\eta_t^2, \eta_t^2)\right\}^- \mathbf{1}_{W_t(\boldsymbol{\theta})\in A_K^c}\right] \\ &= -\mathbb{E}\left[\left\{g(1/\eta_t^2, \eta_t^2)\right\}^-\right] P[W_t(\boldsymbol{\theta}) \in A_K^c] \\ &\to 0 \quad \text{as } K \to \infty,\end{aligned}$$

because, by **A7**, $\mathbb{E}\left[\left\{g(1/\eta_t^2, \eta_t^2)\right\}^-\right] < \infty$. This completes the proof of Step *(iii)*.

Now we prove *(iv)*. For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have

$$\tilde{\mathbf{l}}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0) \geq \mathbf{l}_n(\boldsymbol{\theta}) - \mathbf{l}_n(\boldsymbol{\theta}_0) - |\tilde{\mathbf{l}}_n(\boldsymbol{\theta}) - \mathbf{l}_n(\boldsymbol{\theta})| - |\tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0) - \mathbf{l}_n(\boldsymbol{\theta}_0)|.$$

Hence, using *(i)*,

$$\begin{aligned}
&\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0)\right)\\
&\geq \liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right) - 2\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{\mathbf{l}}_n(\boldsymbol{\theta}) - \mathbf{l}_n(\boldsymbol{\theta})|\\
&= \liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right).
\end{aligned} \quad \textbf{(A.4)}$$

For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any positive integer k, let $V_k(\boldsymbol{\theta})$ the open ball of center $\boldsymbol{\theta}$ and radius $1/k$. We have

$$\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right) \geq \liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0). \quad \textbf{(A.5)}$$

By arguments already given, under **A7**,

$$\mathbb{E}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0)\right)^- \leq \mathbb{E}\left(g(1/\eta_t^2, \eta_t^2))\right)^- < \infty.$$

Therefore, $\mathbb{E}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0)\right)$ exists in $\mathbb{R}\cup\{+\infty\}$, and the ergodic theorem applies (see Francq and Zakoïan, 2019, Exercises 7.3 and 7.4). From (A.5), we obtain

$$\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right) \geq \mathbb{E}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0)\right).$$

The latter term into parentheses converges to $\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)$ as $k \to \infty$, and, by standard arguments using the positive and negative parts of $\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0)$, we have that

$$\lim_{k\to\infty}\mathbb{E}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0)\right) = \mathbb{E}\{\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)\},$$

which by *(i)* is strictly positive. In view of (A.4), the proof of *(iv)* is complete.

Now we complete the proof of the theorem. The set $\boldsymbol{\Theta}$ is covered by the union of an arbitrary neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and, for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, by neighborhoods $V(\boldsymbol{\theta})$ satisfying *(iv)*. Obviously, $\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta}_0)\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) \leq \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0)$, *a.s.* Moreover, by compactness of $\boldsymbol{\Theta}$, there exists a finite subcover of the form $V(\boldsymbol{\theta}_0), V(\boldsymbol{\theta}_1), \ldots, V(\boldsymbol{\theta}_M)$. By *(iv)*, for $i = 1, \ldots, M$, there exists n_i such that for $n \geq n_i$,

$$\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta}_i)\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) > \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0), \quad a.s.$$

Thus, for $n \geq \max_{i=1,\ldots,M}(n_i)$,

$$\inf_{\boldsymbol{\theta}^*\in\bigcup_{i=1,\ldots,M} V(\boldsymbol{\theta}_i)\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) > \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0), \quad a.s.$$

from which we deduce that $\widehat{\boldsymbol{\theta}}_n$ belongs to $V(\boldsymbol{\theta}_0)$ for sufficiently large n.

Proof of Theorem 2

The proof relies on the following steps. There exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that:

(a) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\nabla'_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|<\infty,\quad \mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_t(\boldsymbol{\theta})\right\|<\infty,$

(b) $\boldsymbol{J}$ is invertible and $\sqrt{n}\nabla_{\boldsymbol{\theta}}\mathbf{l}_n(\boldsymbol{\theta}_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,(\kappa_\eta-1)\boldsymbol{J})$,

(c) $\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|n^{-1/2}\sum_{t=1}^{n}\left\{\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})-\nabla_{\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\theta})\right\}\right\|\to 0$ in probability as $n\to\infty$,

$\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|n^{-1}\sum_{t=1}^{n}\left\{\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_t(\boldsymbol{\theta})-\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\theta})\right\}\right\|\to 0$ in probability as $n\to\infty$,

(d) $n^{-1}\sum_{t=1}^{n}\nabla^2_{\theta_i\theta_j}\ell_t(\boldsymbol{\theta}^*)\to\boldsymbol{J}(i,j)$ a.s. for any $\boldsymbol{\theta}^*$ between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

We have

$$\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})=\left(1-\frac{V_t(\boldsymbol{\theta}_0)\eta_t^2}{V_t}\right)\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t,$$
$$\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_t(\boldsymbol{\theta})=\left(1-\frac{V_t(\boldsymbol{\theta}_0)\eta_t^2}{V_t}\right)\frac{1}{V_t}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}V_t(\boldsymbol{\theta})+\left(2\frac{V_t(\boldsymbol{\theta}_0)\eta_t^2}{V_t}-1\right)\frac{1}{V_t^2}\nabla_{\boldsymbol{\theta}}V_t\nabla'_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}).$$

To establish (a), by the Hölder inequality, it thus suffices to show

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left|\frac{V_t(\boldsymbol{\theta}_0)}{V_t}\right|^{2p_1}<\infty,\quad \mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\frac{1}{V_t^2}\nabla_{\boldsymbol{\theta}}V_t\nabla'_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta})\right\|^{q_1}<\infty, \tag{A.6}$$

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left|\frac{V_t(\boldsymbol{\theta}_0)}{V_t}\right|^{p_2}<\infty,\quad \mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\frac{1}{V_t}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}V_t(\boldsymbol{\theta})\right\|^{q_2}<\infty, \tag{A.7}$$

for some conjugate numbers $p_i,q_i>1$ such that $p_i^{-1}+q_i^{-1}=1$, with $i=1,2$. We have $\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta})=\frac{1}{\tau_t^2(\boldsymbol{\theta})}\nabla_{\boldsymbol{\theta}}\tau_t^2(\boldsymbol{\theta})+\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\nabla_{\boldsymbol{\theta}}\sigma_t^2(\boldsymbol{\theta})$ and, omitting the dependence with respect to $\boldsymbol{\theta}$, for $a,\alpha>0$ and $\beta\in(0,1)$ (which holds in a neighborhood of $\boldsymbol{\theta}_0$),

$$|\tau_t^{-2}\nabla_a\tau_t^2|\le 1/a,\quad |\sigma_t^{-2}\nabla_\alpha\sigma_t^2|\le\frac{1}{\alpha},\quad |\sigma_t^{-2}\nabla_\omega\sigma_t^2|\le 1/\{\omega(1-\beta)\},$$
$$|\sigma_t^{-2}\nabla_a\sigma_t^2|\le\sigma_t^{-2}\alpha\sum_{k\ge 0}\beta^k\epsilon_{t-k-1}^2|\tau_{t-k-1}^{-2}\nabla_a\tau_{t-k-1}^2|\le\frac{1}{a}.$$

Let I be the set of the indices $i\in\{1,\ldots,q\}$ such that $\phi_i(\boldsymbol{\vartheta}_0)>0$. Using **A11** and the continuity of $\phi_i(\cdot)$, I is also the set of the indices $i\in\{1,\ldots,q\}$ such that $\phi_i(\boldsymbol{\vartheta})>0$ for $\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)$, when $\mathcal{V}(\boldsymbol{\theta}_0)$ is small enough. We thus obtain for $\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)$

$$\|\tau_t^{-2}\nabla_{\boldsymbol{\vartheta}}\tau_t^2\| \le \sum_{i\in I}\|\nabla_{\boldsymbol{\vartheta}}\log\phi_i(\boldsymbol{\vartheta})\|,$$

$$\|\sigma_t^{-2}\nabla_{\boldsymbol{\vartheta}}\sigma_t^2\| \le \sigma_t^{-2}\alpha\sum_{k\ge 0}\beta^k\epsilon_{t-k-1}^2\|\tau_{t-k-1}^{-2}\nabla_{\boldsymbol{\vartheta}}\tau_{t-k-1}^2\| \le \sum_{i\in I}\|\nabla_{\boldsymbol{\vartheta}}\log\phi_i(\boldsymbol{\vartheta})\|.$$

Moreover, for all $s_0\in(0,1)$, using $x/(1+x)\le x^{s_0}$ when $x\ge 0$,

$$\begin{aligned}|\sigma_t^{-2}\nabla_\beta\sigma_t^2| &= \sigma_t^{-2}\sum_{k\ge 0}(k+1)\beta^k(\omega+\alpha\epsilon_{t-k-2}^2)\\ &\le \frac{1}{(1-\beta)^2}+\frac{1}{\beta}\sum_{k\ge 0}(k+1)\frac{\alpha\beta^{k+1}\epsilon_{t-k-2}^2}{\omega+\alpha\beta^{k+1}\epsilon_{t-k-2}^2}\\ &\le \frac{1}{(1-\beta)^2}+\frac{1}{\beta}\sum_{k\ge 0}(k+1)\left(\frac{\alpha\beta^{k+1}\epsilon_{t-k-2}^2}{\omega}\right)^{s_0}.\end{aligned}$$

The inequality

$$\frac{\tau_t^2(\boldsymbol{\theta}_0)}{\tau_t^2(\boldsymbol{\theta})} \le 1+\frac{a_0}{a}\sum_{i\in I}\frac{\phi_i(\boldsymbol{\vartheta}_0)}{\phi_i(\boldsymbol{\vartheta})}\quad \forall\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0), \tag{A.8}$$

A11 and (9) entail $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}|\epsilon_t(\boldsymbol{\theta})|^s<\infty$. It follows that there exist $K\in(0,\infty)$ and $\rho\in(0,1)$ such that, for all $q_1>1$ and s_0 small enough,

$$\left\|\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left|\sigma_t^{-2}\nabla_\beta\sigma_t^2\right|\right\|_{2q_1} \le K+K\sum_{k\ge 0}k\rho^k\left\|\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left|\epsilon_{t-k-2}(\boldsymbol{\theta})\right|^{2s_0}\right\|_{2q_1}<\infty.$$

The existence of the second expectation in (A.6) follows.

Let $\iota>0$ and $\mathcal{V}(\boldsymbol{\theta}_0)$ such that $\beta_0/\beta<1+\iota$. For all $\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)$, using (A.8) and already given arguments, there exist a generic $K\in(0,\infty)$ such that, for $s_0\in(0,1)$,

$$\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \le K+K\sum_{i=0}^{\infty}\frac{\beta_0^i\frac{r_{t-i-1}^2}{\tau_{t-i-1}(\boldsymbol{\theta}_0)}}{\omega+\alpha\beta^i\frac{r_{t-i-1}^2}{\tau_{t-i-1}(\boldsymbol{\theta})}} \le K+K\sum_{i=0}^{\infty}(1+\iota)^i\beta^{is_0}\epsilon_{t-i-1}^{2s_0}(\boldsymbol{\theta}).$$

By choosing ι such that $\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}(1+\iota)\beta^{s_0}<1$ and s_0 sufficiently small, the expectation of the supremum over $\mathcal{V}(\boldsymbol{\theta}_0)$ of the last sum is finite. The existence of the first expectations in (A.6) and (A.7) follows, for all values of p_1 and p_2.

Turning to second-order derivatives, we have

$$\frac{1}{V_t}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}V_t=\frac{1}{\sigma_t^2}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\sigma_t^2+\frac{1}{\tau_t^2}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\tau_t^2+\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}\tau_t^2\nabla_{\boldsymbol{\theta}'}\sigma_t^2+\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}\sigma_t^2\nabla_{\boldsymbol{\theta}'}\tau_t^2. \tag{A.9}$$

The matrix $\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\tau_t^2$ has the form

$$\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\tau_t^2=\begin{pmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \sum_{i=1}^q\nabla_{\boldsymbol{\vartheta}}\phi_i(\boldsymbol{\vartheta})r_{t-i}^2\\ \mathbf{0} & \sum_{i=1}^q\nabla_{\boldsymbol{\vartheta}'}\phi_i(\boldsymbol{\vartheta})r_{t-i}^2 & a\sum_{i=1}^q\nabla^2_{\boldsymbol{\vartheta}\boldsymbol{\vartheta}'}\phi_i(\boldsymbol{\vartheta})r_{t-i}^2\end{pmatrix}.$$

Hence, by **A11** and already used arguments, $\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\tau_t^{-2}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\tau_t^2\right\|$ is bounded by a constant when $\mathcal{V}(\boldsymbol{\theta}_0)$ is sufficiently small. We similarly show that $\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\sigma_t^{-2}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\sigma_t^2\right\|$ admits moments of any order, which, using the triangle and Cauchy–Schwarz inequalities in (A.9), allows to show the existence of the second expectation in (A.7) and to complete the proof of (a).

Now we turn to (b). Suppose there exists a vector $\boldsymbol{x} = (x_1, x_2, x_3, x_4, \boldsymbol{x}_5')' \in \mathbb{R}^{d+4}$ such that $\boldsymbol{x}'\boldsymbol{J}\boldsymbol{x} = 0$. Then, in view of $\nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}_0) = \sigma_t^2(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}}\tau_t^2(\boldsymbol{\theta}_0) + \tau_t^2(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}}\sigma_t^2(\boldsymbol{\theta}_0)$, we have

$$\begin{aligned}0 =& \boldsymbol{x}'\nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}_0)\\ =& \sigma_t^2\boldsymbol{x}'\left\{(\nabla_{\boldsymbol{\theta}} a_0)\sum_{i=1}^q \phi_i(\boldsymbol{\vartheta}_0) r_{t-i}^2 + a_0\sum_{i=1}^q \nabla_{\boldsymbol{\theta}}\phi_i(\boldsymbol{\vartheta}_0) r_{t-i}^2\right\}\\ &+ \tau_t^2\boldsymbol{x}'\left\{\nabla_{\boldsymbol{\theta}}\omega_0 + \epsilon_{t-1}^2\nabla_{\boldsymbol{\theta}}\alpha_0 - \alpha_0\epsilon_{t-1}^2\nabla_{\boldsymbol{\theta}}\log\tau_{t-1}^2 + \sigma_{t-1}^2\nabla_{\boldsymbol{\theta}}\beta_0 + \beta_0\nabla_{\boldsymbol{\theta}}\sigma_{t-1}^2\right\}\\ :=& e_{t-1}\eta_{t-1}^4 + f_{t-1}\eta_{t-1}^2 + g_{t-1}, \quad a.s.,\end{aligned} \tag{A.10}$$

where $e_{t-1}, f_{t-1}, g_{t-1} \in \mathcal{F}_{t-2}$. By Assumption **A5**, we must have $e_{t-1} = f_{t-1} = g_{t-1} = 0$, a.s. Therefore,

$$\begin{aligned}0 = e_{t-1} =& \alpha_0 V_{t-1}\sigma_{t-1}^2\boldsymbol{x}'\left\{\phi_1(\boldsymbol{\vartheta}_0)\nabla_{\boldsymbol{\theta}} a_0 + a_0\nabla_{\boldsymbol{\theta}}\phi_1(\boldsymbol{\vartheta}_0)\right\}\\ &+ a_0\phi_1(\boldsymbol{\vartheta}_0)V_{t-1}\sigma_{t-1}^2\boldsymbol{x}'\left\{\nabla_{\boldsymbol{\theta}}\alpha_0 - \alpha_0\nabla_{\boldsymbol{\theta}}\log\tau_{t-1}^2\right\},\end{aligned}$$

from which we deduce

$$\begin{aligned}& a_0\phi_1(\boldsymbol{\vartheta}_0)\alpha_0\boldsymbol{x}'\nabla_{\boldsymbol{\theta}}\log\tau_{t-1}^2\\ =& \alpha_0\boldsymbol{x}'\left\{\phi_1(\boldsymbol{\vartheta}_0)\nabla_{\boldsymbol{\theta}} a_0 + a_0\nabla_{\boldsymbol{\theta}}\phi_1(\boldsymbol{\vartheta}_0)\right\} + a_0\phi_1(\boldsymbol{\vartheta}_0)\boldsymbol{x}'\nabla_{\boldsymbol{\theta}}\alpha_0 := c.\end{aligned}$$

We thus have

$$a_0\phi_1(\boldsymbol{\vartheta}_0)\alpha_0\boldsymbol{x}'\nabla_{\boldsymbol{\theta}}\tau_{t-1}^2 = c\tau_{t-1}^2,$$

that is,

$$a_0\sum_{i=1}^q\left[\phi_1(\boldsymbol{\vartheta}_0)\alpha_0\boldsymbol{x}'\left\{a_0\nabla_{\boldsymbol{\theta}}\phi_i(\boldsymbol{\vartheta}_0) + \phi_i(\boldsymbol{\vartheta}_0)\nabla_{\boldsymbol{\theta}} a_0\right\} - c\phi_i(\boldsymbol{\vartheta}_0)\right] r_{t-i}^2 = c.$$

By **A5**, it can be shown that any equality of the form $\sum_{i=1}^\infty b_i r_{t-i}^2 = b_0$, where the b_i's are real constants, entails $b_i = 0$ for all $i \geq 0$. We thus have $c = 0$ and, since $a_0\alpha_0 > 0$,

$$\phi_1(\boldsymbol{\vartheta}_0)\left\{x_4\phi_i(\boldsymbol{\vartheta}_0) + a_0\boldsymbol{x}_5'\nabla_{\boldsymbol{\vartheta}}\phi_i(\boldsymbol{\vartheta}_0)\right\} = 0, \quad i = 1,\ldots,q.$$

First, suppose $\phi_1(\boldsymbol{\vartheta}_0) \neq 0$. Then, since $\sum_{i=1}^q \phi_i(\boldsymbol{\vartheta}_0) = 1$ and $\sum_{i=1}^q \nabla_{\boldsymbol{\vartheta}}\phi_i(\boldsymbol{\vartheta}_0) = \mathbf{0}$, we get $x_4 = 0$. Thus, $\boldsymbol{x}_5'\left[\nabla_{\boldsymbol{\vartheta}}\phi_1(\boldsymbol{\vartheta}_0), \ldots, \nabla_{\boldsymbol{\vartheta}}\phi_q(\boldsymbol{\vartheta}_0)\right] = 0$, which by **A10** entails $\boldsymbol{x}_5 = \mathbf{0}$. The definition of c thus implies $x_2 = 0$. Turning back to (A.10), we obtain

$$0 = x_1 + x_3\sigma_{t-1}^2 + \beta(x_1\nabla_\omega\sigma_{t-1}^2 + x_3\nabla_\beta\sigma_{t-1}^2) = x_3(1+\beta)\sigma_{t-1}^2 + y_{t-2},$$

where $y_{t-2} \in \mathcal{F}_{t-3}$. Using again **A5**, we deduce $x_3 = 0$ and finally $x_1 = 0$. We have shown that $\boldsymbol{x} = \mathbf{0}$ and the proof of the first part of *(b)* is now complete. We have

$$\sqrt{n}\nabla_{\boldsymbol{\theta}}\mathbf{l}_n(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(1-\eta_t^2)\nabla_{\boldsymbol{\theta}}\log V_t(\boldsymbol{\theta}_0).$$

The convergence in distribution follows from the central limit theorem for square integrable stationary and ergodic martingale differences (Billingsley, 1961).

Now we turn to *(c)*. Note that

$$\begin{aligned}&\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\theta})\\ =&\frac{r_t^2}{V_t\tilde{V}_t}(V_t-\tilde{V}_t)\nabla_{\boldsymbol{\theta}}\log V_t + \left(1-\frac{r_t^2}{\tilde{V}_t}\right)(\nabla_{\boldsymbol{\theta}}\log V_t - \nabla_{\boldsymbol{\theta}}\log \tilde{V}_t).\end{aligned} \tag{A.11}$$

We have, for t large enough, $\nabla_{\boldsymbol{\theta}}\tau_t^2 = \nabla_{\boldsymbol{\theta}}\tilde{\tau}_t^2$. Moreover, $\tilde{\sigma}_t^2 = \omega + \alpha\tilde{\epsilon}_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2$, where $\tilde{\epsilon}_t = r_t/\tilde{\tau}_t$, thus

$$\nabla_{\boldsymbol{\theta}}\tilde{\sigma}_t^2 = \nabla_{\boldsymbol{\theta}}\omega + \tilde{\epsilon}_{t-1}^2\nabla_{\boldsymbol{\theta}}\alpha + \alpha\nabla_{\boldsymbol{\theta}}\tilde{\epsilon}_{t-1}^2 + \tilde{\sigma}_{t-1}^2\nabla_{\boldsymbol{\theta}}\beta + \beta\nabla_{\boldsymbol{\theta}}\tilde{\sigma}_{t-1}^2.$$

Therefore, for t large enough,

$$\nabla_{\boldsymbol{\theta}}\sigma_t^2 - \nabla_{\boldsymbol{\theta}}\tilde{\sigma}_t^2 = (\sigma_{t-1}^2 - \tilde{\sigma}_{t-1}^2)\nabla_{\boldsymbol{\theta}}\beta + \beta\{\nabla_{\boldsymbol{\theta}}\sigma_{t-1}^2 - \nabla_{\boldsymbol{\theta}}\tilde{\sigma}_{t-1}^2\}.$$

By (A.1), this entails, for t large enough,

$$\left\|\nabla_{\boldsymbol{\theta}}\sigma_t^2 - \nabla_{\boldsymbol{\theta}}\tilde{\sigma}_t^2\right\| \le Kt\beta^t,$$

and, given that $\tilde{\sigma}_t^2$ and σ_t^2 are uniformly bounded below, it is straightforward to deduce

$$\left\|\nabla_{\boldsymbol{\theta}}\log\sigma_t^2 - \nabla_{\boldsymbol{\theta}}\log\tilde{\sigma}_t^2\right\| \le K\beta^t\left\{t + \left\|\nabla_{\boldsymbol{\theta}}\log\sigma_t^2\right\|\right\}.$$

By $\nabla_{\boldsymbol{\theta}}\log V_t = \nabla_{\boldsymbol{\theta}}\log\sigma_t^2 + \nabla_{\boldsymbol{\theta}}\log\tau_t^2$, we also have

$$\left\|\nabla_{\boldsymbol{\theta}}\log V_t - \nabla_{\boldsymbol{\theta}}\log\tilde{V}_t\right\| \le K\beta^t\left\{t + \left\|\nabla_{\boldsymbol{\theta}}\log\sigma_t^2\right\|\right\}, \tag{A.12}$$

for large enough t. Noting that $V_t - \tilde{V}_t = (\sigma_t^2 - \tilde{\sigma}_t^2)\tau_t^2$ for large t, we deduce from (A.11) that

$$\left\|\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\theta})\right\| \le K\left\{1+\epsilon_t^2(\boldsymbol{\theta})\right\}\left\{t + \left\|\nabla_{\boldsymbol{\theta}}\log V_t\right\|\right\}\beta^t.$$

From the proof of (a), we have

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}|\epsilon_t(\boldsymbol{\theta})|^{4s_0} < \infty \text{ and } \mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\nabla_{\boldsymbol{\theta}}\log V_t\right\|^{2s_0} < \infty$$

for sufficiently small $s_0 \in (0,1)$. By the triangle and Hölder inequalities, for $K \in (0,\infty)$ and $\rho \in (0,1)$, we then have

$$\mathbb{E}\left(\sum_{t=1}^{\infty}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\theta})\right\|\right)^s \le K\sum_{t=1}^{\infty}(t^s+K)\rho^{ts} < \infty,$$

which entails that $\sum_{t=1}^{\infty}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\nabla_{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})-\nabla_{\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\theta})\right\|$ is finite almost surely. The convergence in the first part of (c) follows. The second convergence can be established along the same lines.

Turning to (d), we note that, by (a) and the ergodic theorem,

$$n^{-1}\sum_{t=1}^{n}\nabla^2_{\theta_i\theta_j}\ell_t(\boldsymbol{\theta}_0)\to\boldsymbol{J}(i,j)\quad\text{a.s. as } n\to\infty.$$

For all $\varepsilon>0$, by the same argument, the continuity of the second derivatives and the dominated convergence theorem, there exists a sufficiently small neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta})}\left|\nabla^2_{\theta_i\theta_j}\ell_t(\boldsymbol{\theta})-\nabla^2_{\theta_i\theta_j}\ell_t(\boldsymbol{\theta}_0)\right|\\ =&\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta})}\left|\nabla^2_{\theta_i\theta_j}\ell_t(\boldsymbol{\theta})-\nabla^2_{\theta_i\theta_j}\ell_t(\boldsymbol{\theta}_0)\right|\le\varepsilon.\end{aligned}$$

The point (d) is thus a consequence of the consistency of $\widehat{\boldsymbol{\theta}}_n$.

The proof of the theorem then follows from a Taylor expansion of the criterion around $\boldsymbol{\theta}_0$ and classical arguments.

Proof of Proposition 2

Conditional on (r_t), the bootstrap statistics a_n^* and W_n^* remain random because they depend on $\epsilon_t^*\sim F_n$. The proof is standard and uses the same arguments as those of Theorem 2 and Proposition 2 in Francq and Zakoïan (2009).

Proof of Theorem 3

Throughout the proof, we assume a fixed trajectory $(r_t)_{t\in\mathbb{Z}}$, belonging to a subset of events of probability 1 such that (A.12) holds uniformly in $\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)$ and $\widehat{\boldsymbol{J}}_n^c\to\boldsymbol{J}$ as $n\to\infty$. This sequence exists by the arguments used to show (c) and (d) in the proof of Theorem 2. We thus have

$$\left(\widehat{\boldsymbol{J}}_n^c\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\eta_t^{*2}-1\right)\frac{1}{\widetilde{V}_t}\nabla_{\boldsymbol{\theta}}\widetilde{V}_t(\widehat{\boldsymbol{\theta}}_n^c)=\boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{x}_{t,n}+o(1)$$

with $\boldsymbol{x}_{t,n}=\left(\eta_t^{*2}-1\right)\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)$. Conditional on (r_t), the previous quantity remains random because it depends on $\epsilon_t^*\sim F_n$. To establish (16), by the Cramér–Wold device, it thus suffices to show that, for any $\boldsymbol{\lambda}\neq\boldsymbol{0}\in\mathbb{R}^4$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,(\kappa_\eta-1)\boldsymbol{\lambda}'\boldsymbol{J}\boldsymbol{\lambda}\right).\qquad\textbf{(A.13)}$$

Note that, still conditioning by $(r_t)_{t\in\mathbb{Z}}$, for each n, the random variables $\boldsymbol{\lambda}'\boldsymbol{x}_{1,n},\boldsymbol{\lambda}'\boldsymbol{x}_{2,n},\ldots$ are independent and centered, with finite second-order moments. By the Lindeberg's CLT for triangular arrays of square integrable martingale increments, it remains to show that

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}\left(\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\right)\to(\kappa_\eta-1)\boldsymbol{\lambda}'\boldsymbol{J}\boldsymbol{\lambda}>0\quad\text{as } n\to\infty, \tag{A.14}$$

and, for all $\varepsilon>0$,

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\left\{\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\right\}^2\mathbb{1}_{\{|\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}|\geq\sqrt{n}\varepsilon\}}\right)\to 0\quad\text{as } n\to\infty. \tag{A.15}$$

In Lemma A.1 in Francq and Zakoïan (2022), it has been shown that, for standard GARCH, the distribution F_n of the standardized residuals tends to the (unconditional) distribution F of η_t. More precisely, for any almost everywhere continuous function h such that $|h(x|\leq ax^4+b$, where $a,b>0$, for almost all realization $(r_t)_{t\in\mathbb{Z}}$, we have

$$\int h(x)F_n(dx)\to\int h(x)F(dx)\text{ as } n\to\infty. \tag{A.16}$$

It can be assumed that $(r_t)_{t\in\mathbb{Z}}$ is such that (A.16) holds. Since $\eta_t^*\sim F_n$, given $(r_t)_{t\in\mathbb{Z}}$, under H_0, we have

$$\mathbb{E}\eta_t^*=0,\quad\mathbb{E}\eta_t^{*2}=1,\quad\text{and}\quad\mathbb{E}\eta_t^{*i}=\frac{1}{n-n_0}\sum_{k=n_0+1}^{n}\widehat{\eta}_k^{0i}\to\mathbb{E}\eta_1^i\quad\text{for } i\leq 4,$$

as $n\to\infty$. For t fixed, we then have

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\lambda}'\boldsymbol{x}_{t,n})&=\left\{\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right\}^2\left(\frac{1}{n-n_0}\sum_{k=n_0+1}^{n}\left(\widehat{\eta}_k^0\right)^4-1\right)\\&\to\left\{\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right\}^2(\kappa_\eta-1)\quad\text{as } n\to\infty,\end{aligned}$$

from which (A.14) follows.

Given (r_t), when $\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\neq 0$, we have

$$\begin{aligned}&\mathbb{E}\left\{\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\right\}^2\mathbb{1}_{\{|\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}|\geq\sqrt{n}\varepsilon\}}\\&=\left\{\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right\}^2\mathbb{E}\left\{\left|\eta_t^{*2}-1\right|^2\mathbb{1}_{\left(\left|\eta_t^{*2}-1\right|\geq\frac{\sqrt{n}\varepsilon}{\left|\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right|}\right)}\right\}.\end{aligned} \tag{A.17}$$

For any $A>0$, there exists n_A such that if $n>n_A$ then the expectation in the right-hand side of (A.17) is bounded by

$$\mathbb{E}\left|\eta_t^{*2}-1\right|^2\mathbb{1}_{\{|\eta_t^{*2}-1|\geq A\}}.$$

By (A.16), this term tends as $n\to\infty$ to

$$\int_{|x^2-1|\geq A}\left|x^2-1\right|^2F(dx),$$

which is arbitrarily small when A is sufficiently large. We then obtain (A.15) by the Cesàro Mean Theorem. The convergence (16) follows. The second convergence is obtained by noting that $\frac{1}{\sigma^2}N^2\mathbb{1}_{N\geq 0} \sim \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$. Under H_1 and the conditions given in the theorem, a careful examination of the proof of Lemma A.1 in Francq and Zakoïan (2022) shows that (A.16) holds if F denotes the marginal distribution of $r_t/\sigma_t(\boldsymbol{\theta}_G)$. It follows that

$$\left(\widehat{\boldsymbol{J}}_n^c\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\eta_t^{*2}-1\right)\frac{1}{\widetilde{V}_t}\nabla_{\boldsymbol{\theta}}\widetilde{V}_t(\widehat{\boldsymbol{\theta}}_n^c) = O_P(1),$$

and thus $\sqrt{n}\widehat{a}_n^* = O_P(1)$. Since $\sqrt{n}\widehat{a}_n \to \infty$ as $n \to \infty$, we have $P(\sqrt{n}\widehat{a}_n^* \geq \sqrt{n}\widehat{a}_n) \to 0$ as $n \to \infty$.

REFERENCES

Amado, C. & T. Teräsvirta (2017) Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Reviews* 36, 421–446.

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

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

Beutner, E., A. Heinemann, & S. Smeekes (2020) A residual bootstrap for conditional value-at-risk. Preprint, arXiv:1808.0912. https://EconPapers.repec.org/RePEc:arx:papers:1808.09125.

Billingsley, P. (1961) The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_nY_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–220.

Cavaliere, G., H.B. Nielsen, R.S. Pedersen, & A. Rahbek (2022) Bootstrap inference on the boundary of the parameter space with application to conditional volatility models. *Journal of Econometrics* 227, 241–263.

Conrad, C., A. Custovic, & E. Ghysels (2018) Long- and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management* 11, 23.

Conrad, C. & R. Engle (2021) Modelling volatility cycles: The (MF) ^{2} GARCH model. NYU Stern School of Business, Forthcoming.

Conrad, C. & K. Loch (2015) Anticipating long-term stock market volatility. *Journal of Applied Econometrics* 30, 1090–1114.

Conrad, C. & M. Schienle (2020) Testing for an omitted multiplicative long-term component in GARCH models. *Journal of Business & Economic Statistics* 38, 229–242.

Davis, R. & S. Resnick (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* 14, 533–558.

Ding, Z. & C. Granger (1996) Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics* 73, 185–215.

Engle, R.F., E. Ghysels, & B. Sohn (2013) Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics* 95, 776–797.

Engle, R.F. & G. Lee (1999) A long-run and short-run component model of stock return volatility. In R. Engle & H. White (eds), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, pp. 475–497. Oxford University Press.

Francq, C., L. Horvàth, & J.-M. Zakoïan (2010) Sup-tests for linearity in a general nonlinear AR (1) model. *Econometric Theory* 26, 965–993.

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

Francq, C. & J.-M. Zakoïan (2009) Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104, 313–324.

Francq, C. & J.-M. Zakoïan (2019) *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd Edition. Wiley.

Francq, C. & J.-M. Zakoïan (2022) Testing the existence of moments for GARCH processes. *Journal of Econometrics* 227, 47–64.

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

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Ibragimov, R., J. Kim, & A. Skrobotov (2023) New robust inference for predictive regressions. *Econometric Theory*. Preprint, arXiv:2006.01191v4.

Ibragimov, R., R.S. Pedersen, & A. Skrobotov (2021) New approaches to robust inference on market (non-)efficiency volatility clustering and non linear dependence. Preprint, arXiv:2006.01212v3.

Kandji, B.M. (2023) Iterated function systems driven by non independent sequences: Structure and inference. *Journal of Applied Probability*, forthcoming.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* 131, 207–248.

Kreiss, J.P., E. Paparoditis, & D.N. Politis (2011) On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics* 39, 2103–2130.

Leucht, A., J.P. Kreiss, & M.H. Neumann (2015) A model specification test for GARCH (1, 1) processes. *Scandinavian Journal of Statistics* 42, 1167–1193.

Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28, 1427–1451.

Patton, A.J. (2011) Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.

Shimizu, K. (2013) The bootstrap does not always work for heteroscedastic models. *Statistics & Risk Modeling* 30, 189–204.

Tanny, D. (1974) A zero-one law for stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 30, 139–148.

van der Vaart, A.W. (2000) *Asymptotic Statistics*. Cambridge University Press.

Wang, F. & E. Ghysels (2015) Econometric analysis of volatility component models. *Econometric Theory* 31, 362–393.

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