CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

WORKING PAPER SERIES

Inference on Multiplicative Component GARCH without any Small-Order Moment

Christian Francq, Baye Matar Kandji, Jean-Michel Zakoian

N°09/April 2022

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

Inference on multiplicative component GARCH without any small-order moment

Christian Francq*, Baye Matar Kandji† and Jean-Michel Zakoian‡

March 18, 2022

Abstract

In multiplicative component GARCH models, the volatility is decomposed into the product of two factors which often received interpretations in terms of "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.

JEL Classification: C12, C13, C22 and C58

Keywords: GARCH-MIDAS, Moments existence, QMLE, Residual Bootstrap, Tests on boundary parameters.

*CREST-ENSAE and University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-Mail: christian.francq@univ-lille3.fr

†CREST-ENSAE. E-Mail: bayematar.kandji@ensae.fr

‡Corresponding author: Jean-Michel Zakoïan, University of Lille and CREST-ENSAE, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France. E-mail: zakoian@ensae.fr.

1 Introduction

Despite their ability to capture 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) parametrization) 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 Lee (1999) but, in recent years, multiplicative component GARCH processes have attracted more attention. In such models, 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), Conrad and Engle (2020).

While two-component volatility models are widely used in applied works, some of their theoretical properties remain unexplored. In this paper, we consider two such properties: first, the existence of small-order moments for the strictly stationary solution of the two-component volatility model and, second, the asymptotic properties of the Quasi-Maximum Likelihood (QML) estimator. The 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.

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 modelling of financial returns), there exists a sufficiently small power s (depending on both the volatility parameters and the innovations distribution) such that $E|r_t|^s < \infty$. 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 heavyness of the tails of the marginal distribution entails formidable statistical difficulties for proving the consistency and asymptotic normality of the QML estimator.

The rest of the paper is organized as follows. In the next section we study the existence of strictly stationary solutions to the two-component 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 two-factor GARCH process (r_t) defined by

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

where (η_t) is an iid sequence with $E\eta_t^2 = 1$, the $\phi_i(\boldsymbol{\vartheta}_0)$'s are nonnegative coefficients depending on some d-variate parameter $\boldsymbol{\vartheta}_0$, $a_0 \geq 0$, $\omega_0 > 0$, $\alpha_0 \geq 0$ and $\beta_0 \geq 0$. Some coefficients $\phi_i(\boldsymbol{\vartheta}_0)$, but not all of them, are allowed to be zero. Without loss of generality assume that $\sum_{i=1}^q \phi_i(\boldsymbol{\vartheta}_0) = 1$. 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*, while 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.

Remark 1. In particular, Model (1) includes the class of GARCH-mixed-data sampling (GARCH-MIDAS) model proposed by Engle et al. (2013). In this paper, under a slightly different parametrization[1], the τ_t component is specified by smoothing realized volatilities, as

$$\tau_t^2 = m + a \sum_{i=1}^{Q} \varphi_i(\boldsymbol{\vartheta}) RV_{t-i}, \tag{2}$$

[1] Engle et al. (2013) considered a unit-variance GARCH(1,1) equation, $\sigma_t^2 = 1 - \alpha - \beta + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ for the short-run volatility. 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 $\alpha + \beta < 1$, which is not necessary for the strict stationarity. Note that Engle et al. (2013) allow for an intercept in the equation of r_t.

where $RV_t = \sum_{i=0}^{N-1} r_{t-i}^2$ is a rolling window realized volatility and the weights $\varphi_i(\boldsymbol{\vartheta})$'s are positive and sum to one. Engle et al. (2013) suggest exponential or Beta weighting schemes for the specification of the functions φ_i.

Next, we turn to the existence of strictly stationary solutions to Model (1). The problem was investigated by Wang and Ghysels (2015) in the GARCH-MIDAS case under the condition $\alpha_0 + \beta_0 < 1$.

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

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

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

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} \delta_{t-j} \right). \tag{3}$$

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

$$E(\sigma_t^{2r}) < \infty \text{ if and only if } E\delta_1^r < 1. \tag{4}$$

Note that 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

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

as $r \to \infty$. Note that **A2** is satisfied when η_t is not bounded and $\alpha_0 \neq 0$. It follows from (4) that $E(\sigma_t^{2r}) = \infty$ for r large enough.

Write (1) in matrix form as

$$\boldsymbol{r}_t = \boldsymbol{A}_t \boldsymbol{r}_{t-1} + \boldsymbol{b}_t, \tag{5}$$

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 (5) admits, by Brandt (Theorem 1, 1986), the strictly stationary solution

$$\boldsymbol{r}_t = \boldsymbol{b}_t + \sum_{i=1}^{\infty} \left(\prod_{j=1}^{i} \boldsymbol{A}_{t+1-j} \right) \boldsymbol{b}_{t-i} \tag{6}$$

under the assumption

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

Note that the top-Lyapounov exponent $\gamma_{\boldsymbol{A}}$ involved in **A3** is well defined in $[-\infty, \infty)$ because $E \log^+ \|\boldsymbol{A}_t\| < \infty$, in view of (7) below. Wang and Ghysels (2015) obtained explicit conditions entailing **A3** for particular sub-models. 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

$$E|\epsilon_t|^s < \infty \quad \text{for some } s > 0. \tag{7}$$

The following proposition shows that, surprisingly, this feature does not extend to the solution (r_t) of the multi-volatility model (1).

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

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

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

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

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

$$E|r_t|^s = \infty \quad \textit{for all } s > 0. \tag{8}$$

Note that, for a particular class of models of the form (1), Wang and Ghysels (2015) showed Proposition 1 for $s = 2$ (see their Proposition 3.9).

Remark 2. Without Assumption **A2**, the two factor GARCH process may admit moments at any order. Indeed, suppose that $\delta_1 \in [0,1]$ with probability 1. It follows that, for any $s > 0$, $E|\delta_t|^s < 1$ using (3). 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 one, then Assumption **A3** is satisfied and, by (6), r_t^2 admits moments at any order.

3 QMLE without any moment

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 (1), 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

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

This is a key assumption to show the CAN of the QMLE of GARCH (see Berkes, Horváth and Kokoszka (2003) and Francq and Zakoian (2004)). To the authors knowledge, the consistency of the QMLE has never been shown without an assumption that implies (9). Proposition 1 however entails that (9) cannot be assumed in our framework.

To circumvent the failure of the small-order moment assumption, we will use Theorem 2.2 in Kandji (2022), establishing that the strictly stationary solution of (1) satisfies

$$\limsup_{k\to\infty} \frac{1}{k} \log r_{t+k}^2 \leq 0, \quad \limsup_{k\to\infty} \frac{1}{k} \log r_{t-k}^2 \leq 0 \quad \text{a.s.} \tag{10}$$

for all $t \in \mathbb{Z}$. This property can be interpreted as an exponential control of the trajectories. It is easy to see that (9) implies (10),[2] but the converse is false.[3]

Assume that the observations $r_1, \dots, r_n$ constitute a realization (of length n) of the two-factor GARCH process defined by (1), 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

$$\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{L}_n(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\boldsymbol{\theta}) := \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}),$$

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 $E\log\eta_t^2 > -\infty$.

Assumption **A7**, precluding densities with too much mass around zero, 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 Zakoian (2004)) but it is introduced here to

[2]Indeed, by the strict stationarity of (r_t) we have, for any $\iota > 0$, under (9)

$$\sum_{k=1}^{\infty} P(k^{-1}\log|r_{t+k}| > \iota) \leq (s\iota)^{-1} E\log^+|r_t|^s \leq (s\iota)^{-1}E|r_t|^s < \infty,$$

(see for instance Exercise 2.13 in Francq and Zakoian (2019)). By the Borel-Cantelli lemma, this entails the first inequality in (10) and the second inequality is obtained similarly.

[3]Let a sequence (X_t) of identically distributed random variables such that $E|X_t| < \infty$ but $EX_t^2 = \infty$. Then $r_t = e^{|X_t|/2}$ satisfies (10) but not (9). Indeed, $k^{-1}\log r_{t+k}^2 = k^{-1}|X_{t+k}| \to 0$ a.s. On the other hand $E|r_t|^s = Ee^{s|X_t|/2} \geq \frac{1}{2}E(s|X_t|/2)^2 = \infty$, for any $s > 0$.

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***,* **A3**-**A7***, we have*

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

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 $\kappa_\eta := E\eta_t^4 < \infty$.

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 $i = 1, \dots, q$, admit continuous second-order derivatives and the matrix $[\nabla_{\boldsymbol{\vartheta}}\phi_1(\boldsymbol{\vartheta}_0), \dots, \nabla_{\boldsymbol{\vartheta}}\phi_q(\boldsymbol{\vartheta}_0)]$ has full-row rank.

A11 If $\phi_i(\boldsymbol{\vartheta}_0) = 0$ then there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that $\phi_i(\boldsymbol{\vartheta}) = 0$ for all $\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0) \cap \Theta$.

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. Recall that $V_t(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta})\tau_t^2(\boldsymbol{\vartheta})$.

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

$$\boldsymbol{J} := E\left(\frac{1}{V_t^2(\boldsymbol{\theta}_0)}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\nabla'_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right) \tag{11}$$

is a positive definite matrix.

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 non standard 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, Horvath and Zakoian, 2010) the Wald, score and

Likelihood-Ratio (LR) test statistics do not follow the standard χ^2 distribution 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 a test procedure which has a standard asymptotic distribution 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.

4.1 Fixing $\boldsymbol{\vartheta}$

The first approach relies on the auxiliary model

$$\begin{cases} r_t = \tau_t \epsilon_t, & \tau_t^2 = 1 + a_0 \sum_{i=1}^q \phi_i(\boldsymbol{\vartheta}^*) r_{t-i}^2, \\ \epsilon_t = \sigma_t \eta_t, & \sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{cases} \tag{12}$$

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

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

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

and

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

Denote by χ_1^2 the chi-square distribution with one degree of freedom, and by $\frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$ the equally 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 (2018), Cavaliere, Nielsen, Pedersen and Rahbek (2021).

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 $r_1, \dots, r_n$, compute the QMLE $\widehat{\boldsymbol{\theta}}_G = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})'$ of a GARCH(1,1) model and compute the standardized residuals $\widehat{\eta}_t^* = (\widehat{\eta}_t^c - m_n)/s_n$, for $t = n_0 + 1, \dots, n$, where $\widehat{\eta}_t^c$, m_n and s_n are respectively the non-standardized 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 (12). 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 iid 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 (12).

3. Repeat B times Step 2, and denote by $\widehat{a}_n^{*1}, \dots, \widehat{a}_n^{*B}$ the bootstrap estimates of a. Approximate the p-value of the test $H_0 : a_0 = 0$ against $H_1 : a_0 > 0$ by $p_B^* = (1 + \#\{\widehat{a}_n^{*j} \geq \widehat{a}_n; j = 1, \dots, B\})/(B + 1)$.

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 Zakoian (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

$$\widehat{a}_n^* = \max\left\{ \boldsymbol{e}_4' \left(\widehat{\boldsymbol{J}}_n^c\right)^{-1} \frac{1}{n}\sum_{t=1}^{n}\left(\eta_t^{*\,2}-1\right)\frac{1}{\widetilde{V}_t}\nabla_{\boldsymbol{\theta}}\widetilde{V}_t(\widehat{\boldsymbol{\theta}}_n^c), 0\right\}. \tag{13}$$

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 (13). Under H_0, for almost all realization (r_t), as $n \to \infty$ we have, given (r_t),*

$$\sqrt{n}\widehat{a}_n^* \xrightarrow{\mathcal{L}} N\mathbb{1}_{N\geq 0}, \quad N \sim \mathcal{N}\left(0, \sigma^2 := (\kappa-1)\boldsymbol{e}_4\boldsymbol{J}^{-1}\boldsymbol{e}_4\right), \tag{14}$$

and thus

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

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

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

exists and is invertible and if $\widehat{a}_n \to a_0 > 0$ 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 (12). 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 non standard asymptotic distributions under the null, and the bootstrap techniques become particularly appealing. The resampling scheme is then modified as follows.

1. On the observations $r_1, \ldots, r_n$, compute the GARCH(1,1) QMLE $\widehat{\boldsymbol{\theta}}_G = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})$ and the standardized residuals $\widehat{\eta}_t^* \sim F^*$, exactly as in the previous algorithm. Compute the QMLE of the GARCH-MIDAS model (1). 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 iid noise (η_t^*) with distribution F_n^*, compute the QMLE $\widehat{\boldsymbol{\theta}}_n^* = \left(\widehat{\omega}_n^*, \widehat{\alpha}_n^*, \widehat{\beta}_n^*, \widehat{a}_n^*, \widehat{\boldsymbol{\vartheta}}_n^{*\prime}\right)^{\prime}$ of the GARCH-MIDAS model (1).

3. Repeat B times Step 2, and compute the bootstrap estimated p-value p_B^* exactly as in the previous algorithm.

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 $r_1, \ldots, r_n$, but also on the random bootstrap trials (for a formal definition, see *e.g.* Page 98 in van der Vaart, 2000). 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)-(2). 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.

n		True	Min	Q1	Q2	Q3	Max	Bias	RMSE	MASE
				DGP A satisfying **A2** (no moments)						
2000	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
4000	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)						
2000	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
4000	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

Table 1: Distribution of the QMLE over 1000 replications

5.1 Monte Carlo experiments

The aim of our first Monte Carlo experiments 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 (1) 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 equally 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 $E\eta_t^2 = 1$ 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 $R = 1000$, with sample sizes $n = 2000$ and $n = 4000$. 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 i-th Mean Asymptotic Standard Error (MASE) is defined as the empirical mean over the R replications of the estimated standard errors $\sqrt{\widehat{\mathbf{\Sigma}}(i,i)/n}$, where $\widehat{\mathbf{\Sigma}}$ is the empirical estimator of the asymptotic variance $\mathbf{\Sigma} = (\kappa_\eta - 1)\boldsymbol{J}^{-1}$ 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.

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 (2) with $m = 1$, with Beta weights

$$\varphi_i(\vartheta) = \frac{\{1 - i/(Q+1)\}^{\vartheta-1}}{\sum_{j=1}^{Q}\{1 - j/(Q+1)\}^{\vartheta-1}}$$

where $\vartheta \in (0, \infty)$. We thus simulated 1000 trajectories of size $n = 12654$ of Model (1)-(2) with $m = 1$, $N = 22$, $Q = 250$ and $(\omega_0, \alpha_0, \beta_0, \vartheta_0, a_0) = (0.028, 0.115, 0.831, 2.067, 0.056)$.[4] For the distribution of η_t we took a standardized Student distribution with $\nu = 5.41$ degrees of freedom[5]. The estimation results are presented in the top panel of Table 2. Interestingly, the parameter a_0 is very accurately estimated. Its estimated value over the 1000 replications is always positive, and its estimated standard deviation is in average very close to the observed RMSE. We have redone the estimation exercise on simulations of a standard GARCH (corresponding to Model (1) 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 zero. Unsurprisingly, the estimations of ϑ, whose true value is undefined when $a_0 = 0$, are erratic. Figure 1 displays 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 2 shows that the estimation of the volatilities is fortunately not too sensitive

[4]These parameters are those estimated on the NASDAQ index considered in Section 5.2, with RVs computed over one month and one MIDAS lag year.

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

to the choice of the integers N and Q in (2). 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 1000 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 3 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

$$QLIK = \frac{1}{n} \sum_{t=r_0+1}^{n} \frac{\sigma_t^2}{\hat{\sigma}_t^2} + \log \hat{\sigma}_t^2,$$

where we took $r_0 = 100$ to avoid the effect of the initial values required to compute the volatility estimates. The reader is referred to Patton (2011) for arguments in favour 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.

	True	Min	Q1	Q2	Q3	Max	Bias	RMSE	MASE
ω	0.028	0.015	0.025	0.029	0.034	0.063	0.002	0.007	0.006
α	0.115	0.081	0.107	0.116	0.123	0.160	0.001	0.012	0.012
β	0.831	0.736	0.817	0.829	0.840	0.884	-0.003	0.019	0.018
ϑ	2.067	0.517	1.716	2.140	2.587	13.598	0.196	0.936	0.752
a	0.056	0.015	0.042	0.053	0.066	0.183	-0.001	0.019	0.018
ω	0.028	0.017	0.025	0.028	0.031	0.045	0.000	0.004	0.005
α	0.115	0.084	0.108	0.116	0.124	0.162	0.001	0.012	0.012
β	0.831	0.731	0.816	0.826	0.837	0.874	-0.005	0.019	0.018
ϑ	UD	0.000	2.067	2.067	2.067	118.580	0.993	6.193	104.949
a	0	0.000	0.000	0.000	0.009	0.061	0.005	0.010	0.010

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

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

	Min	Q1	Q2	Q3	Max	Mean	SD
ω	0.007	0.025	0.030	0.035	0.060	0.030	0.008
α	0.065	0.096	0.104	0.112	0.147	0.104	0.012
β	0.832	0.868	0.878	0.887	0.929	0.878	0.014
$\alpha+\beta$	0.943	0.977	0.983	0.988	1.012	0.982	0.009

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

Model	Min	Q1	Q2	Q3	Max	Mean	SD
MIDAS	0.783	1.030	1.115	1.212	3.479	1.139	0.179
GARCH	0.787	1.035	1.122	1.219	3.497	1.146	0.181

Table 4: Distribution of the QLIK losses over 1000 replications when the GARCH-MIDAS volatility is estimated by the GARCH-MIDAS model or by the GARCH model.

$a_0 = 0$ (under the null) or $a_0 \in \{0.01, 0.05\}$ (under the alternative). The number of replications is 1000. Different values of $\vartheta \geq 1$ are used. With $\vartheta = 1$ all the RVs involved in (2) 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 size of the Wald test, and especially that of the LR test, is not well controlled. It can also be seen that the Wald test seems to be slightly more powerful than the two other tests, and that the score test has a low power when $\vartheta = 9$. The poor control of the error of 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 that this bootstrap test much better controls the error of 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 $K = 1000$.

5.2 Application to stock indices

We estimated the GARCH-MIDAS model (1)-(2) with exponential weights on the daily returns of the CAC 40, DAX, NASDAQ and Hang Seng indices, from 1990-03-01 to 2021-04-08. Table 7 displays the estimated coefficients when $N = 65$ (corresponding to RVs over a quarter) and $Q = 1000$ (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 one month and one 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

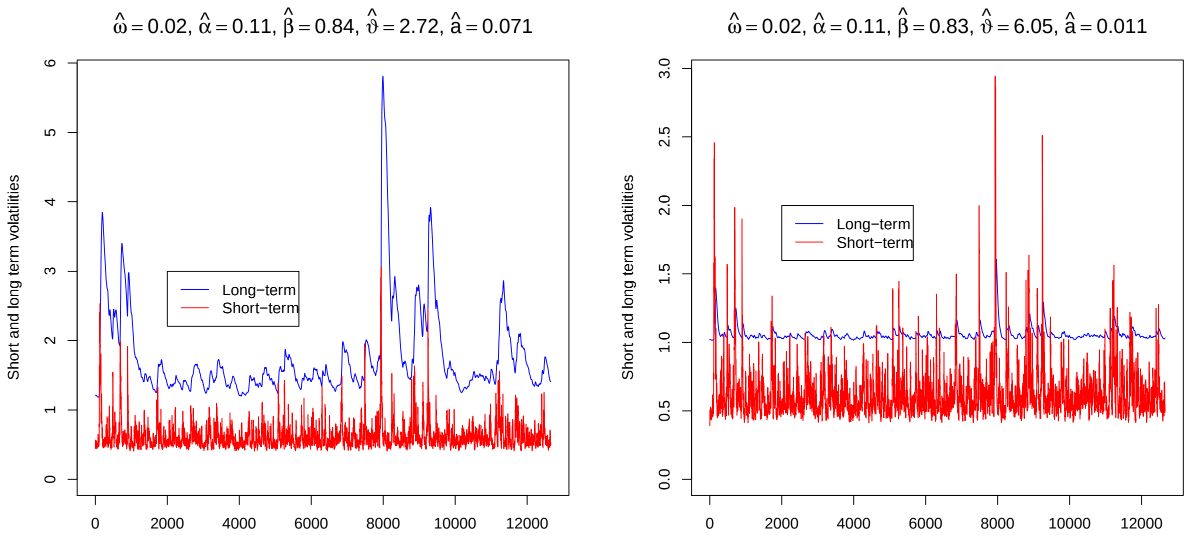

Figure 1: Examples of estimated short and long term volatilities when the GDP is a GARCH-MIDAS (left figure) or a standard GARCH (right figure)

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 4 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 zero and a continuous distribution on $(0, \infty)$, has been estimated by a Kernel density estimator—using the reflection method for boundary correction. Figure 5 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 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 pandemic.

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. We also established the asymptotic properties of the QML estimator and proposed tests of the

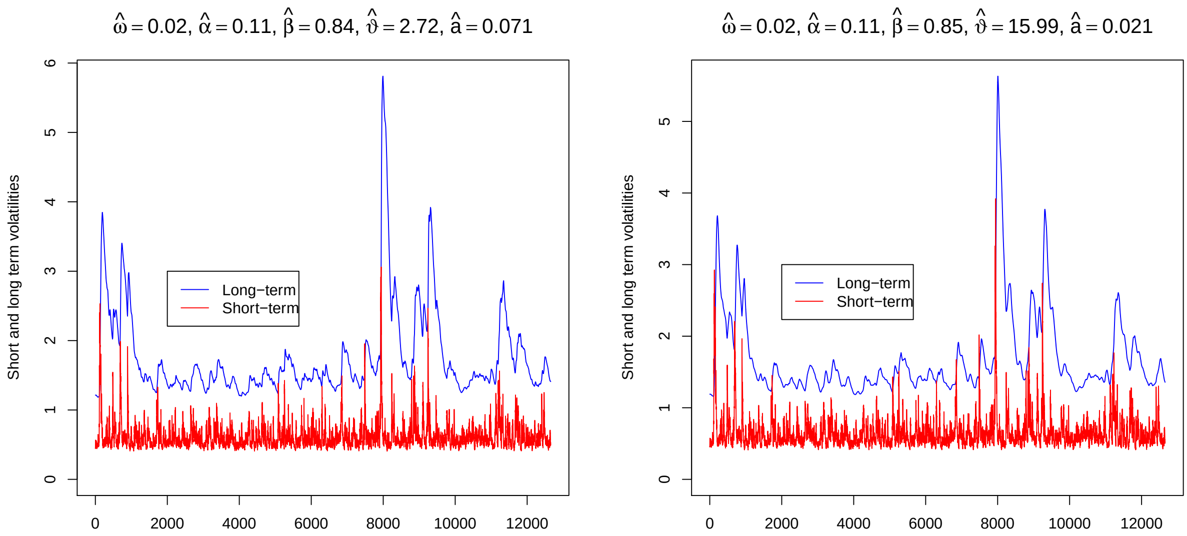

Figure 2: Estimated short and long term volatilities of the GARCH-MIDAS model with $N = 22$ and $Q = 250$ (left figure) and with $N = 65$ and $Q = 1000$ (right figure)

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.

References

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

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

[3] Beutner, E., Heinemann, A. and Smeekes, S. (2018) A residual bootstrap for conditional Value-at-Risk. *Preprint arXiv:1808.09125.*

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

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

[6] Cavaliere, G., Nielsen, H.B., Pedersen R.S. and Rahbek, A. (2021) Bootstrap inference on the boundary of the parameter space with application to conditional volatility models. *Journal of Econometrics*, in press.

[7] Conrad, C. and Engle, R. (2020) Modelling Volatility Cycles: the $(\mathrm{MF})^2$ GARCH model. *Working document available at SSRN.*

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

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

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

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

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

[13] Engle, R. F. and Lee, G. (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*. Oxford, UK: Oxford University Press, 475–497.

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

[15] Francq, C. and Zakoïan, J-M. (2004) Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH. *Bernoulli* 10, 605–637.

[16] Francq, C. and Zakoïan, J-M. (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.

[17] Francq, C. and Zakoïan, J-M. (2019) *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons, Second Edition.

[18] Francq, C. and Zakoïan, J-M. (2021) Testing the existence of moments for GARCH processes. *Journal of Econometrics*, in press.

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

[20] Kandji, B.M. (2022) Iterated Function Systems driven by non independent sequences: structure and inference. CREST Working Papers Series 2022-03. https://crest.science/wp-content/uploads/2022/01/2022-03.pdf

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

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

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

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

[25] van der Vaart, A. W. (2000) *Asymptotic statistics.* Cambridge university press.

[26] Wang, F. and Ghysels, E. (2015) Econometric Analysis of Volatility Component Models. *Econometric Theory* 31, 362–393.

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

Appendix: proofs

Proof of Lemma 1

Let $\mathcal{F}_t$ be the sigma-field generated by $\{\eta_u, u \leq t\}$. Let $\mu_p = E|\eta_1|^p$ 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 zero with probability one.

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

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

By succesive applications of this inequality, it follows that

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

Iterating the argument, we obtain the result with

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

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

Proof of Proposition 1

The strictly stationary solution is obtained from (3) and (6), 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 $E|\epsilon_t|^{ks} = \infty$. 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

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

by the Cauchy rule and the second inequality in (10).

However, contrary to the standard GARCH case, the limiting criterion $E\ell_t(\boldsymbol{\theta})$ 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 Zakoian (2019) Section 7.4), do not involve a limiting criterion :

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

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

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

$iv)$ 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

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

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

$$|\sigma_t^2 - \tilde{\sigma}_t^2| \leq \beta|\sigma_{t-1}^2 - \tilde{\sigma}_{t-1}^2| \leq \beta^{t-q}|\sigma_q^2 - \tilde{\sigma}_q^2|. \tag{15}$$

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

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

where $\rho = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \beta < 1$ and K is a fixed (independent of n) random variable. By the first inequality in (10), we have

$$\limsup_{k\to\infty} \frac{1}{k}\log \rho^k r_{t+k}^2 < 0, \text{a.s.}$$

from which it follows that $\rho^t(1 + r_t^2) \to 0$, a.s. as $t \to \infty$. By Cesàro's lemma the conclusion follows.

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

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

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

$$\frac{\tau_{t-1}^2(\boldsymbol{\theta})}{\tau_{t-1}^2(\boldsymbol{\theta}_0)} = \frac{a\alpha\phi_1(\boldsymbol{\vartheta})}{a_0\alpha_0\phi_1(\boldsymbol{\vartheta}_0)} := c. \tag{16}$$

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 (16) 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

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

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

$$\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0) \geq g(W_t(\boldsymbol{\theta}), \eta_t^2)\mathbb{1}_{W_t(\boldsymbol{\theta})\in A_K} - \left\{g(W_t(\boldsymbol{\theta}), \eta_t^2)\right\}^- \mathbb{1}_{W_t(\boldsymbol{\theta})\in A_K^c}. \tag{17}$$

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

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

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,

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

To deal with the expectation of the second term in the r.h.s. of (17) we use the fact that for $y > 0$, $g(x, y) \geq g(1/y, y)$. It follows that

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

because, by **A7**, $E\left[\left\{g(1/\eta_t^2, \eta_t^2)\right\}^-\right] < \infty$. 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)*

$$\begin{aligned}
&\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0)\right) \\
&\geq \liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right) - 2\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{\mathbf{l}}_n(\boldsymbol{\theta}) - \mathbf{l}_n(\boldsymbol{\theta})| \\
&= \liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right).
\end{aligned} \tag{18}$$

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

$$\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^* \in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*) - \mathbf{l}_n(\boldsymbol{\theta}_0)\right) \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^n \inf_{\boldsymbol{\theta}^* \in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0). \tag{19}$$

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

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

Therefore $E\left(\inf_{\boldsymbol{\theta}^* \in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*) - \ell_t(\boldsymbol{\theta}_0)\right)$ exists in $\mathbb{R}\cup\{+\infty\}$, and the ergodic theorem applies (see Francq and Zakoian (2019), Exercises 7.3 and 7.4). From (19) we obtain

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

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

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

which by *i)* is strictly positive. In view of (18), 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, $\inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta}_0)\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) \leq \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0), \quad a.s.$ Moreover, by compactness of $\boldsymbol{\Theta}$, there exists a finite subcover of the form $V(\boldsymbol{\theta}_0), V(\boldsymbol{\theta}_1), \dots, V(\boldsymbol{\theta}_M)$. By *iv)*, for $i = 1, \dots, M$, 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 $n \geq \max_{i=1,\dots,M}(n_i)$,

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

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

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

$$E \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left|\frac{V_t(\boldsymbol{\theta}_0)}{V_t}\right|^{2p_1} < \infty, \quad E \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left\|\frac{1}{V_t^2}\nabla_{\boldsymbol{\theta}}V_t\nabla'_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta})\right\|^{q_1} < \infty, \tag{20}$$

$$E \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left|\frac{V_t(\boldsymbol{\theta}_0)}{V_t}\right|^{p_2} < \infty, \quad E \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left\|\frac{1}{V_t}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}V_t(\boldsymbol{\theta})\right\|^{q_2} < \infty, \tag{21}$$

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 $\boldsymbol{\theta}_0$),

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

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

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

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

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

The inequality

$$\frac{\tau_t^2(\boldsymbol{\theta}_0)}{\tau_t^2(\boldsymbol{\theta})} \leq 1 + \frac{a_0}{a}\sum_{i\in I}\frac{\phi_i(\boldsymbol{\vartheta}_0)}{\phi_i(\boldsymbol{\vartheta})} \quad \forall \boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0), \tag{22}$$

A11 and (7) entail $E\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} |\epsilon_t(\boldsymbol{\theta})|^s < \infty$. 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,

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

The existence of the second expectation in (20) 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 (22) and already given arguments, there exist a generic $K \in (0,\infty)$ such that, for $s_0 \in (0,1)$,

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

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 (20) and (21) follows, for all values of p_1 and p_2.

Turning to second-order derivatives, we have

$$\frac{1}{V_t}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}V_t = \frac{1}{\sigma_t^2}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\sigma_t^2 + \frac{1}{\tau_t^2}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}'}\tau_t^2 + \frac{1}{V_t}\nabla_{\boldsymbol{\theta}}\tau_t^2\nabla_{\boldsymbol{\theta}'}\sigma_t^2 + \frac{1}{V_t}\nabla_{\boldsymbol{\theta}}\sigma_t^2\nabla_{\boldsymbol{\theta}'}\tau_t^2. \tag{23}$$

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 (23), allows to show the existence of the second expectation in (21) 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

$$\begin{aligned} 0 =& \boldsymbol{x}'\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0) \\ =& \sigma_t^2\boldsymbol{x}'\left\{(\nabla_{\boldsymbol{\theta}}a_0)\sum_{i=1}^q\phi_i(\boldsymbol{\vartheta}_0)r_{t-i}^2 + a_0\sum_{i=1}^q\nabla_{\boldsymbol{\theta}}\phi_i(\boldsymbol{\vartheta}_0)r_{t-i}^2\right\} \\ &+ \tau_t^2\boldsymbol{x}'\left\{\nabla_{\boldsymbol{\theta}}\omega_0+\epsilon_{t-1}^2\nabla_{\boldsymbol{\theta}}\alpha_0-\alpha_0\epsilon_{t-1}^2\nabla_{\boldsymbol{\theta}}\log\tau_{t-1}^2+\sigma_{t-1}^2\nabla_{\boldsymbol{\theta}}\beta_0+\beta_0\nabla_{\boldsymbol{\theta}}\sigma_{t-1}^2\right\} \\ :=& e_{t-1}\eta_{t-1}^4+f_{t-1}\eta_{t-1}^2+g_{t-1}, \quad a.s. \end{aligned} \tag{24}$$

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,

$$\begin{aligned} 0=e_{t-1} =& \alpha_0V_{t-1}\sigma_{t-1}^2\boldsymbol{x}'\left\{\phi_1(\boldsymbol{\vartheta}_0)\nabla_{\boldsymbol{\theta}}a_0+a_0\nabla_{\boldsymbol{\theta}}\phi_1(\boldsymbol{\vartheta}_0)\right\} \\ &+ a_0\phi_1(\boldsymbol{\vartheta}_0)V_{t-1}\sigma_{t-1}^2\boldsymbol{x}'\left\{\nabla_{\boldsymbol{\theta}}\alpha_0-\alpha_0\nabla_{\boldsymbol{\theta}}\log\tau_{t-1}^2\right\}, \end{aligned}$$

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 (24), 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

$$\begin{aligned}
&\nabla_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}} \tilde{\ell}_t(\boldsymbol{\theta}) \\
=& \frac{r_t^2}{V_t \tilde{V}_t}(V_t - \tilde{V}_t)\nabla_{\boldsymbol{\theta}} \log V_t + \left(1 - \frac{r_t^2}{\tilde{V}_t}\right)(\nabla_{\boldsymbol{\theta}} \log V_t - \nabla_{\boldsymbol{\theta}} \log \tilde{V}_t).
\end{aligned} \tag{25}$$

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 (15), this entails, for t large enough,

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

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

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

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

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

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 (25) that

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

From the proof of a), we have

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

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

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

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

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

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

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

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

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

The proof is standard and uses the same arguments as those of Theorem 2 and Proposition 2 in Francq and Zakoian (2009).

Proof of Theorem 3

By the arguments used to show c) and d) in the Proof of Theorem 2, it can be seen that

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

with $\boldsymbol{x}_{t,n}=\left(\eta_t^{*2}-1\right)\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)$. To establish (14), by the Wold-Cramer device, it thus suffices to show that for any $\boldsymbol{\lambda}\neq\mathbf{0}\in\mathbb{R}^4$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,(\kappa_\eta-1)\boldsymbol{\lambda}'\boldsymbol{J}\boldsymbol{\lambda}\right). \tag{26}$$

Note that, conditional on (r_t), 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

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}\left(\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\right)\to(\kappa_\eta-1)\boldsymbol{\lambda}'\boldsymbol{J}\boldsymbol{\lambda}>0 \quad \text{as } n\to\infty, \tag{27}$$

and for all $\varepsilon>0$

$$\frac{1}{n}\sum_{t=1}^{n}E\left(\left\{\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\right\}^2\mathbb{1}_{\{|\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}|\geq\sqrt{n}\varepsilon\}}\right)\to 0 \quad \text{as } n\to\infty. \tag{28}$$

In Lemma A.1 in Francq and Zakoian (2021), 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) we have

$$\int h(x)F_n^*(dx)\to\int h(x)F(dx) \text{ as } n\to\infty. \tag{29}$$

Given (r_t), for t fixed we then have

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

from which (27) follows.

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

$$\begin{aligned}&E\left\{\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}\right\}^2\mathbb{1}_{\{|\boldsymbol{\lambda}'\boldsymbol{x}_{t,n}|\geq\sqrt{n}\varepsilon\}}\\&=\left\{\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right\}^2E\left|\eta_t^{*2}-1\right|^2\mathbb{1}_{\left\{\left|\eta_t^{*2}-1\right|\geq\frac{\sqrt{n}\varepsilon}{\left|\boldsymbol{\lambda}'\frac{1}{V_t}\nabla_{\boldsymbol{\theta}}V_t(\boldsymbol{\theta}_0)\right|}\right\}}.\end{aligned} \tag{30}$$

For any $A>0$ there exists n_A such that if $n>n_A$ then the expectation in the right-hand side of (30) is bounded by

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

By (29), this term tends as $n\to\infty$ to

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

which is arbitrarily small when A is sufficiently large. We then obtain (28) by the Cesàro Mean Theorem. The convergence (14) 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 Zakoian (2021) shows that (29) 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 $n\to\infty$.

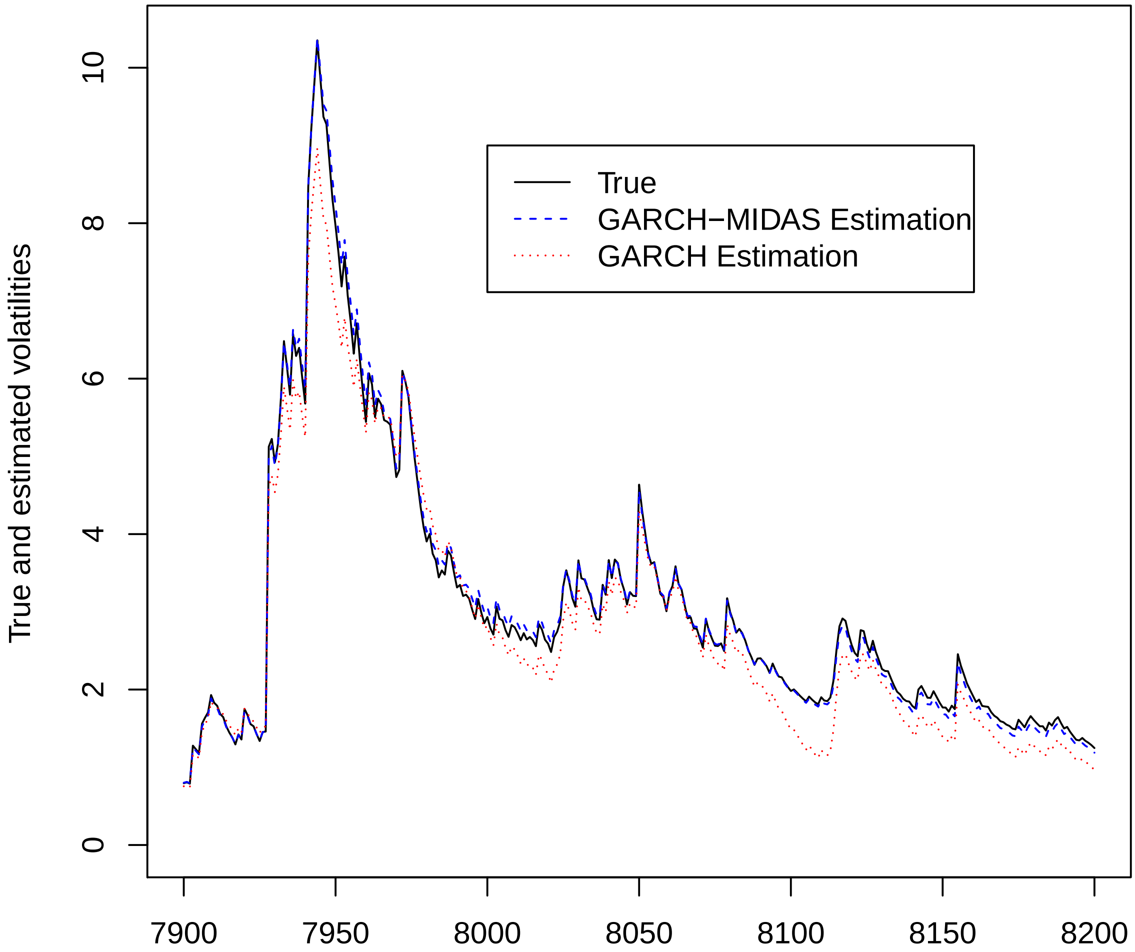

Figure 3: True and estimated volatility estimated by a GARCH-MIDAS and by a standard GARCH

a_0	ϑ	Test	0.1%	1%	2%	3%	4%	5%	6%	7%	10%	20%
0	1	R_n	0.2	0.9	1.6	2.4	3.3	4.2	4.8	6.3	8.5	17.6
		W_n	0	0.2	0.2	0.4	0.7	1.2	1.4	2.1	4.1	13.2
		L_n	0.2	0.2	0.8	0.9	1.4	1.9	2.3	2.4	3.7	6.8
	2	R_n	0.1	1.1	2.1	2.9	3.7	4.4	5.2	5.8	8.3	18.4
		W_n	0	0.1	0.6	0.8	1.3	1.4	1.9	2.7	4.8	14.5
		L_n	0.2	0.9	1.2	1.5	1.7	2	2.5	2.7	3.8	8
	3	R_n	0	1.2	2.2	2.5	3.5	4.3	4.9	6.1	8.9	19.1
		W_n	0	0.1	0.7	1.1	1.5	1.8	2.6	3.5	5.6	14
		L_n	0.1	1.1	1.4	1.7	1.8	2.1	2.2	2.5	4.2	8.8
	9	R_n	0.2	1	1.3	2.1	2.5	3.1	4.1	5.2	8.4	18.1
		W_n	0	0.3	0.8	2	2.6	3.3	4.3	4.9	7.6	16.5
		L_n	0.1	0.5	0.7	1.2	1.7	2	2.4	2.6	4	8.7
0.01	1	R_n	2.1	7.3	10.2	12.6	15.3	16.9	19.2	21.3	24.2	36.7
		W_n	0.1	0.7	2.8	5.5	7.9	11.8	14.8	18.4	26	46.1
		L_n	1.2	5.7	8	10.6	13.2	15.3	17.5	19	23.2	34.7
	2	R_n	2.9	8	11.4	15.2	17.6	19.7	21.7	23.6	27.5	40.7
		W_n	0.1	2.1	5.5	8.9	12.8	17	21	23.6	31.4	52.5
		L_n	2.4	7.7	11.4	14.5	17.4	19.9	22.4	23.6	27	39.9
	3	R_n	2.3	6.8	10.7	13.1	15	16.5	18.5	20.3	24.6	36.8
		W_n	0.1	2.9	6.8	11.2	15	19	21.3	24.4	32	52.1
		L_n	2.5	7.8	12	14.7	16.8	19	21	22.8	25.9	38.4
	9	R_n	0.3	1.4	2.7	3.4	3.8	4.4	5.6	6.5	8.8	17.2
		W_n	0.2	4.6	9.2	13.1	16.2	18.1	21.3	23	27.5	41.1
		L_n	1.5	4.3	6.6	9	10.7	12.2	13.8	15.7	19.2	27
0.05	1	R_n	65.9	78.8	83	84.4	85.7	86.3	86.8	87.3	88.9	91.5
		W_n	19	80	90.8	94.6	96.5	97.6	98	98.6	99.3	100
		L_n	85.2	94.2	96.8	97.6	98.1	98.2	98.4	98.7	99.2	100
	2	R_n	51.3	66.3	69.6	72.2	73.1	74.6	75.6	76.1	77.7	81.7
		W_n	36.1	88.7	95.1	97.3	98.4	98.8	99.3	99.5	99.8	100
		L_n	91.1	97.1	98	98.4	99.1	99.5	99.5	99.5	99.6	99.9
	3	R_n	23.3	38.7	44.4	48.7	51.2	53.8	55.1	57.4	60.6	66.5
		W_n	50.9	92	96	98	98.7	99.1	99.5	99.5	99.8	100
		L_n	90.2	96.3	97.2	98.3	98.8	99.4	99.4	99.4	99.6	99.8
	9	R_n	1.4	4.2	6.4	8.4	11.3	13.3	14.9	16.8	23.2	37.8
		W_n	67.9	91.3	94.7	96.3	96.8	97.5	97.9	98.2	98.6	99
		L_n	56.6	74.8	80.5	83.9	85.4	86.7	87.4	88.4	90	92.5

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	0.1%	1%	2%	3%	4%	5%	6%	7%	10%	20%
0	0.0	0.5	1.4	2.2	2.8	4.2	5.1	6.4	7.6	16.6
0.01	0.0	2.8	7.6	10.8	13.0	16.0	18.5	23.2	32.1	56.9
0.05	18.5	82.1	96.5	97.9	98.7	99.5	99.5	99.6	99.9	100.0

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%.

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

Table 7: GARCH-MIDAS fitted on stock returns. The estimated standard deviations are displayed in small, under the estimated values of the coefficients. The last column gives the bootstrap estimated p-value of the Wald test of $H_0 : a = 0$.

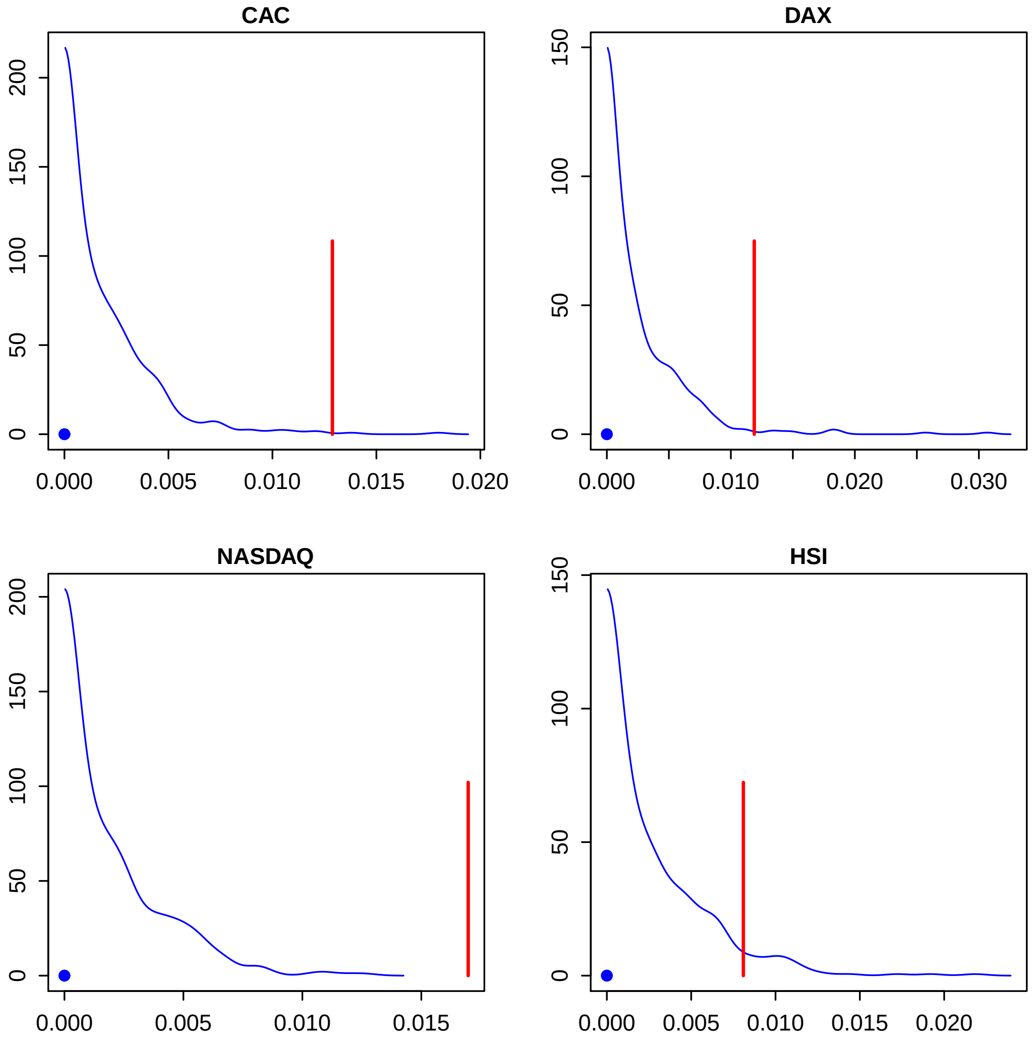

Figure 4: Bootstrap estimate of the distribution of $\widehat{a}_n$ when $a = 0$ (in blue) and observed value of $\widehat{a}_n$ (red vertical line).

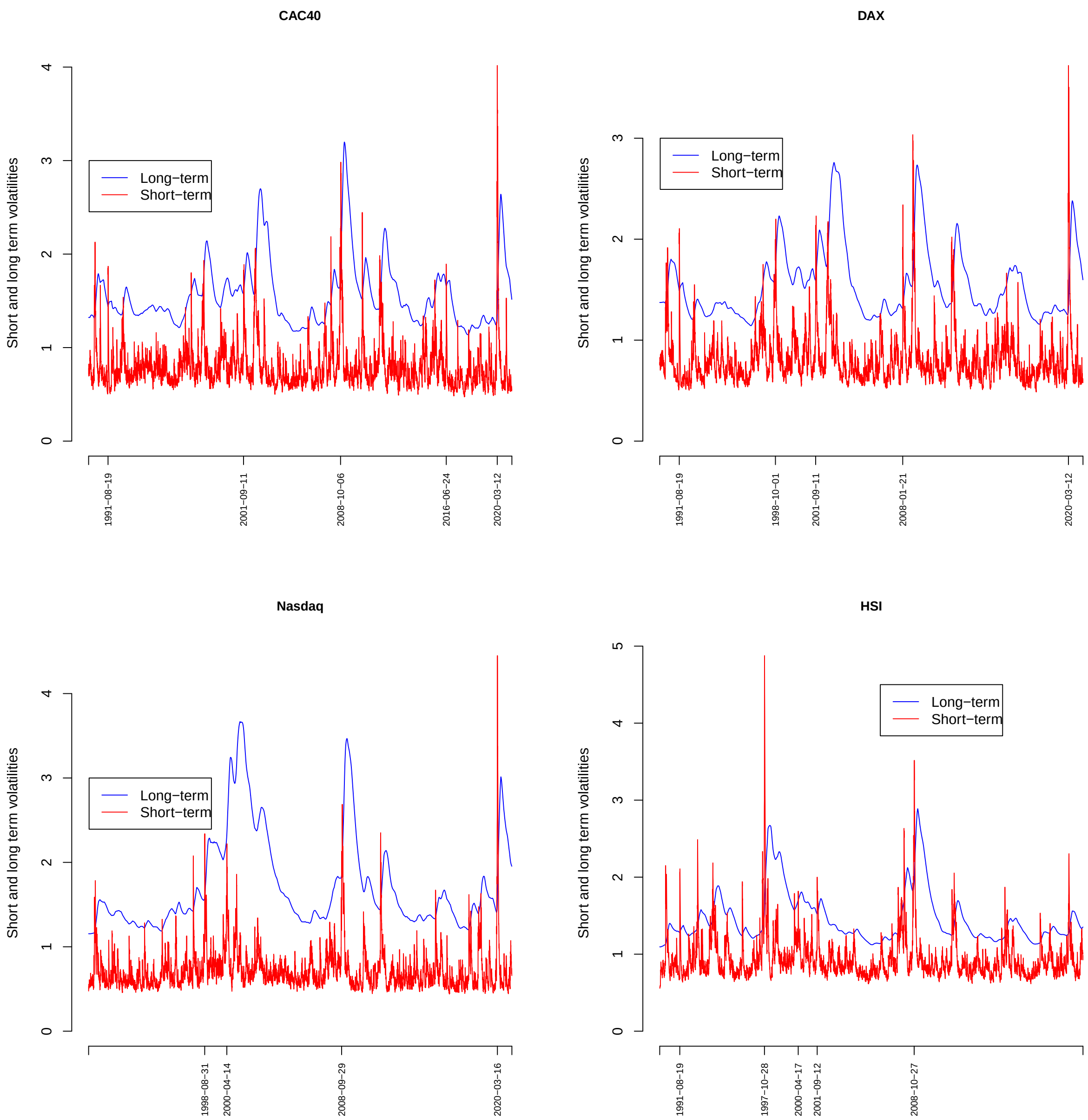

Figure 5: GARCH-MIDAS short and long term volatilities for four stock indices from 1990-03-01 to 2021-04-08.

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/