J. Appl. Probab. **60,** 1501–1515 (2023)
*doi:*10.1017/jpr.2023.13

EXPONENTIAL CONTROL OF THE TRAJECTORIES OF ITERATED FUNCTION SYSTEMS WITH APPLICATION TO SEMI-STRONG GARCH *(P, Q)* MODELS

BAYE MATAR KANDJI ,* *CREST, ENSAE, Institut Polytechnique de Paris*

Abstract

We establish new results on the strictly stationary solution to an iterated function system. When the driving sequence is stationary and ergodic, though not independent, the strictly stationary solution may admit no moment but we show an exponential control of the trajectories. We exploit these results to prove, under mild conditions, the consistency of the quasi-maximum likelihood estimator of GARCH(p,q) models with non-independent innovations.

Keywords: Inference without moments; quasi-maximum likelihood; semi-strong GARCH; stochastic recurrence equation

2020 Mathematics Subject Classification: Primary 60H25
Secondary 62M10; 62P05

1. Introduction

Since [19], the theoretical properties of the stochastic recurrence equation (SRE) $X_t = A_tX_{t-1} + B_t$ has received much attention. This equation gathers a large class of classical econometric processes such as the GARCH and ARMA models, and their numerous variants. A sufficient condition of existence and uniqueness of a strictly stationary solution was proposed in [5] in the case where $(A_t, B_t)_t$ is stationary and ergodic. Under an irreducibility condition, [4] established that this condition is also necessary when the sequence (A_t, B_t) is independent and identically distributed (i.i.d.). The probabilistic properties of the stationary solution of SRE model in the i.i.d. case are well known. In the scalar case, [19] showed that $\mathbb{P}(\pm X_1 > x) \sim c_\pm x^{-a}$ as $x \to \infty$ for some positive constants $c_\pm$. A thorough study of SRE models, in particular their tail behavior, is presented in [6]. The SRE model is the affine-mapping-particular case of the so-called stochastic iterated function system (IFS) $X_t = \Psi(\boldsymbol{\theta}_t, X_{t-1})$. Most of the theoretical properties established for SRE models (stationary, tail properties) can be extended to IFS equations.

One important application of SREs in time series analysis is the study of the stationarity properties of GARCH processes. Assuming i.i.d. innovations, [3] deduced from [5] a necessary and sufficient condition for the existence of a unique stationary solution of a general GARCH(p, q) model. In recent years, the i.i.d. assumption on the innovations has often been replaced by a less restrictive conditional moment assumption (the model is then called 'semi-strong' GARCH). See [10] for the classical GARCH(p,q) model, and [12, 17] for GARCH-X

Received 14 July 2022; revision received 9 February 2023; accepted 10 February 2023.
* Postal address: 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France. Email: bayematar.kandji@ensae.fr

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

models. The GARCH-MIDAS models of [9] constitute another class of IFS models which are not driven by an i.i.d. sequence. Another example is given by GARCH-X models which are IFS driven by a (generally non-i.i.d.) sequence of innovations and covariates. This motivates studying IFS equations driven by non-i.i.d. innovations.

However, strict stationarity generally does not suffice for establishing the asymptotic properties of estimators, such as the quasi-maximum likelihood estimator (QMLE). To our knowledge, all existing works on the QML inference of IFS models *assume* the existence of a small-order moment of the observed process. Surprisingly, however, the strictly stationary solutions of IFS equations with non-i.i.d. innovations may not admit any finite moment.

The aim of this paper is to establish that the stationary trajectories of the IFS equations enjoy an exponential control property. We also show that this property is sufficient to establish the consistency of the QMLE of semi-strong GARCH models.

The rest of the paper is organized as follows. In Section 2 we present our main result, and Section 3 is devoted to its proof. Section 4 investigates the estimation of the semi-strong GARCH(p, q) model. Complementary proofs are displayed in the Appendices.

2. Stochastic IFS without moments

Let $(E, \mathcal{E})$ be a measurable space and (F, d) a complete and separable metric space (Polish space). Let $(\boldsymbol{\theta}_t)_{t\in\mathbb{Z}}$ be a stationary and ergodic process valued in E, and let $\Psi\colon E \times F \to F$ be a function such that $x \mapsto \Psi(\theta, x)$ is Lipschitz continuous for all $\theta \in E$. Let

$$\boldsymbol{\Lambda}_t = \Lambda(\boldsymbol{\Psi}_t) = \sup_{x_1, x_2 \in F, x_1 \neq x_2} \frac{d(\boldsymbol{\Psi}_t(x_1), \boldsymbol{\Psi}_t(x_2))}{d(x_1, x_2)},$$

where $\boldsymbol{\Psi}_t = \Psi(\boldsymbol{\theta}_t, \cdot)$. Let $\boldsymbol{\Lambda}_t^{(0)} = 1$ and $\boldsymbol{\Lambda}_t^{(r)} = \Lambda(\boldsymbol{\Psi}_t \circ \cdots \circ \boldsymbol{\Psi}_{t-r+1})$ for all $r > 0$.

Consider the IFS

$$X_t = \Psi(\boldsymbol{\theta}_t, X_{t-1}) = \boldsymbol{\Psi}_t(X_{t-1}) \qquad \text{for all } t \in \mathbb{Z}. \tag{1}$$

A solution (X_t) of (1) is said to be causal if, for every t, X_t is $\sigma(\boldsymbol{\theta}_k,\ k \le t)$-measurable.

Under a slightly different form, the following result has been established in [8, Theorem 3] and [2, Theorem 3.1]; see also [22, Theorem 2.8] and the review in [7].

Theorem 1. *Assume the following conditions hold: (i) there exists a constant* $c \in F$ *such that* $\mathbb{E} \ln^+ d(\boldsymbol{\Psi}_0(c), c) < \infty$*; (ii)* $\mathbb{E} \ln^+ \boldsymbol{\Lambda}_0 < \infty$*; and (iii)* $\lim_{r\to\infty} (1/r) \ln \boldsymbol{\Lambda}_0^{(r)} < 0$ *almost surely (a.s.). Then there exists a unique stationary (causal and ergodic) solution* $(X_t)_{t\in\mathbb{Z}}$ *to (1).*

Moreover,

$$\text{for all } t \in \mathbb{Z}, \quad d(X_t, c) \le \sum_{n=0}^{\infty} \boldsymbol{\Lambda}_t^{(n)} d(\boldsymbol{\Psi}_{t-n}(c), c) < \infty \quad a.s. \tag{2}$$

Note that $(\ln \boldsymbol{\Lambda}_0^{(r)})_{r\ge 1}$ is a sub-additive sequence. Therefore, by the sub-additive ergodic theorem of [20], the limit in assumption (iii) exists.

For the reader's convenience and because we have not been able to find (2) exactly under this form, we provide a proof of Theorem 1 in Appendix A.

Remark 1. If $(\boldsymbol{\theta}_t)$ is i.i.d., it is possible to prove in particular cases, including the affine mapping, that $d(X_1, c)$ has a power-law tail [6, Theorem 5.3.6]. More generally, it can be shown

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

that, under the conditions of Theorem 1, there exists $s > 0$ such that $\mathbb{E}d(X_1, c)^s < \infty$. This small moment property is often used in the statistical inference of IFS models, for example, to prove the consistency of GARCH models and their derivatives (see [1] for the GARCH model and [13] for the EGARCH and Log-GARCH models). If $(\boldsymbol{\theta}_t)$ is not i.i.d., the examples below show that the stationary solution may not admit any small-order moment.

Example 1. Let $\delta \in (0, 1)$ and let $(z_t)_{t\in\mathbb{Z}}$ be an i.i.d. non-negative real process with $\mathbb{E}z_t = \frac{1}{2}(1-\delta)$ and $\mathbb{E}z_t^2 = \infty$. The process $(\boldsymbol{\theta}_t)$ defined by $\boldsymbol{\theta}_t = \sum_{k=0}^{\infty} \delta^k z_{t-k}$ for all $t \in \mathbb{Z}$ satisfies $\mathbb{E}\boldsymbol{\theta}_t = \frac{1}{2}$ and is such that, for all $t \in \mathbb{Z}$, $x_t = 1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k} \boldsymbol{\theta}_{t-j+1}$ exists a.s. Moreover, (x_t) is the unique stationary solution of $x_t = \boldsymbol{\theta}_t x_{t-1} + 1$, $t \in \mathbb{Z}$. Note that $x_t \geq \prod_{j=1}^{k} \boldsymbol{\theta}_{t-j+1} \geq \delta^{k(k-1)/2}(z_{t-k+1})^k$ for all $k \in \mathbb{N}^*$. For all $s > 0$, we thus have $\mathbb{E}x_0^s \geq \mathbb{E}\delta^{sk(k-1)/2}(z_0)^{sk} = \infty$ for k such that $sk > 2$.

The previous example is simple, but probably a little artificial. We now give an example of commonly used econometric models, for which it was recently proven that the strictly stationary solution does not admit any finite moment.

Example 2. Consider the following GARCH-MIDAS model [9]:

$$\begin{cases} r_t = \sqrt{\boldsymbol{\tau}_t}\boldsymbol{\sigma}_t\boldsymbol{\eta}_t, \\ \boldsymbol{\tau}_t = a + br_{t-1}^2, \\ \boldsymbol{\sigma}_t^2 = 1 - \alpha - \beta + \alpha r_{t-1}^2/\boldsymbol{\tau}_t + \beta\boldsymbol{\sigma}_{t-1}^2, \end{cases}$$

where $(\boldsymbol{\eta}_t)_t$ is a zero-mean and unit-variance i.i.d. sequence, $\alpha > 0$, $\beta \geq 0$, $\alpha + \beta < 1$, $a > 0$, and $b > 0$. Noting that $\boldsymbol{\epsilon}_t := \boldsymbol{\sigma}_t\boldsymbol{\eta}_t$ is a GARCH process, we see that $(\boldsymbol{\tau}_t)$ follows the SRE $\boldsymbol{\tau}_t = a + br_{t-1}^2 = a + (b\boldsymbol{\epsilon}_{t-1}^2)\boldsymbol{\tau}_{t-1}$ driven by a non-i.i.d. sequence $\boldsymbol{\epsilon}_t$. It can be shown that, when $b \leq 1$, the process (r_t) is strictly stationary but, when $\boldsymbol{\eta}_0$ has unbounded support, then, for any $s > 0$, $E|r_t|^s = \infty$. See [11, Proposition 1] for the proof of the previous result.

We now state our main result, which provides a way to circumvent the non-existence of small-order moments for models such as those of Examples 1 and 2. Section 4 will be devoted to the statistical study of a class of econometric models where the existence of moments is not guaranteed.

Theorem 2. *Under the conditions of Theorem 1, for all $t \in \mathbb{Z}$,*

$$\text{(i)} \quad \limsup_{n\to\infty} (1/n) \ln d(X_{t+n}, c) \leq 0; \qquad \text{(ii)} \quad \limsup_{n\to\infty} (1/n) \ln d(X_{t-n}, c) \leq 0 \quad a.s.$$

Theorem 2 can be interpreted as an exponential control of the trajectory of the stationary solution. Note that the property $\mathbb{E} \ln^+ d(X_1, c) < \infty$ (a weaker condition than the existence of a small-order moment) implies the results of Theorem 2 (see Appendix B). However, the converse is false (see [23, Example (a)]).

As a consequence of the previous theorem, we obtain the following result. Its proof is provided in Appendix C.

Corollary 1. *Under the conditions of Theorem 2, almost surely, $\lim_{|n|\to\infty} (1/|n|) \ln^+ d(X_{t+n}, c)$ exists and is equal to 0; if $\mathbb{E} \ln^- d(X_1, c) < \infty$, then*

$$\lim_{|n|\to\infty} \frac{1}{|n|} \ln d(X_{t+n}, c) \text{ exists and is equal to } 0. \tag{3}$$

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

3. Proof of the main result

To show Theorem 2, we first define an SRE which bounds the distance between X_t and c. Note that, by [20],

$$\lim_{r\to\infty} \frac{1}{r} \ln \mathbf{\Lambda}_0^{(r)} = \inf_{r\in\mathbb{N}^*} \frac{1}{r}\mathbb{E} \ln \mathbf{\Lambda}_0^{(r)} = \lim_{r\to\infty} \frac{1}{r}\mathbb{E} \ln \mathbf{\Lambda}_0^{(r)} \quad \text{a.s.}, \tag{4}$$

so by Theorem 1(iii) there exists a positive integer r_0 such that $\mathbb{E} \ln \mathbf{\Lambda}_0^{(r_0)} < 0$. It can be shown that $\mathbb{E}[\ln ((\mathbf{\Lambda}_0^{(r_0)} + u))] \xrightarrow{u\downarrow 0} \mathbb{E} \ln \mathbf{\Lambda}_0^{(r_0)}$ [22, proof of Theorem 2.10]. Therefore, there exists $u_0 > 0$ such that $\ln (u_0) \le \gamma_0 := \mathbb{E}[\ln ((\mathbf{\Lambda}_0^{(r_0)} + u_0))] < 0$. We thus have, for all $v \in [\gamma_0, 0)$,

$$\mathbb{E}[\ln (\delta(v)(\mathbf{\Lambda}_0^{(r_0)} + u_0))] = v, \tag{5}$$

with $\delta(v) = \exp (v - \gamma_0) \ge 1$.

Now, for any integer $p \in [0, r_0 - 1]$, define $(a_{p,t}(v), b_{p,t})_{t\in\mathbb{Z}}$ by

$$a_{p,t}(v) = \delta(v)(\mathbf{\Lambda}_{r_0 t+p}^{(r_0)} + u_0), \qquad b_{p,t} = 1 + \sum_{k=0}^{r_0-1} \mathbf{\Lambda}_{r_0 t+p}^{(k)} d(\mathbf{\Psi}_{r_0 t+p-k}(c), c).$$

By Theorem 1(i) and (ii), and by the elementary inequality $\ln (\sum_{i=1}^n a_i) \le \ln n + \sum_{i=1}^n \ln^+ a_i$ for non-negative $\{a_i\}_{i=1}^n$, we have $\mathbb{E} \ln^+ a_{p,t}(v) < \infty$ and $\mathbb{E} \ln^+ b_{p,t}(v) < \infty$. Therefore, in view of (5), there exists a unique stationary solution $(z_{p,t}(v))_t$ to the equation

$$z_{p,t}(v) = a_{p,t}(v) z_{p,t-1}(v) + b_{p,t}. \tag{6}$$

Note that, by [5],

$$z_{p,t}(v) = \sum_{q=0}^{\infty} \left(\prod_{i=0}^{q-1} a_{p,t-i}(v) \right) b_{p,t-q}. \tag{7}$$

By iterating (6), we have

$$z_{p,t}(v) = \sum_{q=0}^{n} \left(\prod_{i=0}^{q-1} a_{p,t-i}(v) \right) b_{p,t-q} + \left(\prod_{i=0}^{n} a_{p,t-i}(v) \right) z_{p,t-(n+1)}(v), \quad \text{for all } n \ge 1. \tag{8}$$

By (7) and (8), $\big(\prod_{i=0}^n a_{p,t-i}(v)\big) z_{p,t-(n+1)}(v)$ is the remainder of a convergent series, and hence almost surely converges to 0. That is,

$$\left(\prod_{k=0}^{n-1} a_{p,t-k}(v) \right) z_{p,t-n}(v) \overset{n\to\infty}{\to} 0 \quad \text{a.s.} \tag{9}$$

We now give a technical lemma linking the processes (X_t) and $(z_{p,t}(v))_t$.

Lemma 1. *For all $v \in [\gamma_0, 0)$, $0 \le p \le r_0 - 1$, and $t \in \mathbb{Z}$,*

$$d(X_{r_0 t+p}, c) \le z_{p,t}(v) \quad a.s. \tag{10}$$

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

Proof of Lemma 1. For any integer n, let q and m denote the quotient and remainder of the Euclidean division of n by r_0: $n = qr_0 + m$. By sub-multiplicativity we have

$$\mathbf{\Lambda}_t^{(n)} \leq \left(\prod_{i=0}^{q-1} \mathbf{\Lambda}_{t-ir_0}^{(r_0)}\right) \mathbf{\Lambda}_{t-qr_0}^{(m)}, \qquad \prod_{i=0}^{-1} \mathbf{\Lambda}_{t-ir_0}^{(r_0)} = 1.$$

For all $q \in \mathbb{N}$, we then obtain

$$\sum_{n=qr_0}^{(q+1)r_0-1} \mathbf{\Lambda}_t^{(n)} d(\mathbf{\Psi}_{t-n}(c), c) \leq \left(\prod_{i=0}^{q-1} \mathbf{\Lambda}_{t-ir_0}^{(r_0)}\right) \sum_{m=0}^{r_0-1} \mathbf{\Lambda}_{t-qr_0}^{(m)} d(\mathbf{\Psi}_{t-qr_0-m}(c), c).$$

It follows that

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbf{\Lambda}_t^{(n)} d(\mathbf{\Psi}_{t-n}(c), c) &= \sum_{q=0}^{\infty} \sum_{n=qr_0}^{(q+1)r_0-1} \mathbf{\Lambda}_t^{(n)} d(\mathbf{\Psi}_{t-n}(c), c) \\ &\leq \sum_{q=0}^{\infty} \left(\prod_{i=0}^{q-1} \mathbf{\Lambda}_{t-ir_0}^{(r_0)}\right) \sum_{m=0}^{r_0-1} \mathbf{\Lambda}_{t-qr_0}^{(m)} d(\mathbf{\Psi}_{t-qr_0-m}(c), c).\end{aligned}$$

Since $\delta(v) \geq 1$ and $u_0 > 0$, we obtain

$$\left(\prod_{i=0}^{q-1} a_{p,t-i}(v)\right) b_{p,t-q} \geq \left(\prod_{i=0}^{q-1} \mathbf{\Lambda}_{(r_0t+p)-ir_0}^{(r_0)}\right) \sum_{m=0}^{r_0-1} \mathbf{\Lambda}_{(r_0t+p)-qr_0}^{(m)} d(\mathbf{\Psi}_{(r_0t+p)-qr_0-m}(c), c).$$

In view of the last two inequalities, together with (7) and (2), we have

$$z_{p,t}(v) \geq \sum_{n=0}^{\infty} \mathbf{\Lambda}_{r_0t+p}^{(n)} d(\mathbf{\Psi}_{r_0t+p-n}(c), c) \geq d(X_{r_0t+p}, c),$$

which proves (10). □

Let **Aff** denote the set of affine maps from $\mathbb{R}$ into $\mathbb{R}$. An element $f_{a,b}$ of **Aff** can be written as $f_{a,b}(x) = ax + b$, $x \in \mathbb{R}$, where $(a, b) \in \mathbb{R}^2$.

Lemma 2. *Let us define a function Φ from* **Aff** *to $\mathbb{R}_+$ by $\Phi(f_{a,b}) = |a| + |b|$.*

(i) *For any x with $|x| \geq 1$, $|f_{a,b}(x)| \leq \Phi(f_{a,b})|x|$.*

(ii) *If $|d| \geq 1$ then $\Phi(f_{a,b} \circ f_{c,d}) \leq \Phi(f_{a,b})\Phi(f_{c,d})$.*

Since Lemma 2 is elementary, its proof is skipped. Note that Φ is the 1-norm in the vector space of affine maps.

Lemma 3. *For all $p \in \{0, \ldots, r_0 - 1\}$ and $t \in \mathbb{Z}$, letting $Q_p(t) = r_0t + p$,*

$$\text{(i)} \quad \limsup_{n\to\infty} \frac{1}{n} \ln d(X_{Q_p(t+n)}, c) \leq 0; \qquad \text{(ii)} \limsup_{n\to\infty} \frac{1}{n} \ln d(X_{Q_p(t-n)}, c) \leq 0 \quad \text{a.s.}$$

Lemma 3 distinguishes between cases (i) and (ii) because their proofs are different.

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

Proof of Lemma 3. We start by proving (i). Let f_t be the random affine map defined by $f_t(x) = a_{p,t}(v)x + b_{p,t}$ for all $x \in \mathbb{R}$. Define also the maps $\boldsymbol{\gamma}_{t,n} = f_t \circ f_{t-1} \cdots \circ f_{t-n+1}$ and $\boldsymbol{\zeta}_{t,n} = f_{t+n} \circ f_{t+n-1} \cdots \circ f_{t+1}$ for all $(t, n) \in \mathbb{Z} \times \mathbb{N}^*$. Note that

$$\boldsymbol{\zeta}_{t,n} = \boldsymbol{\gamma}_{t+n,n}, \qquad z_{p,t}(v) = \boldsymbol{\gamma}_{t,n}(z_{p,t-n}(v)), \qquad z_{p,t+n}(v) = \boldsymbol{\zeta}_{t,n}(z_{p,t}(v)) \quad \text{a.s.} \tag{11}$$

Since $b_{p,t} \geq 1$, by Lemma 2(ii),

$$(u_{t,n})_n := (\ln \Phi(\boldsymbol{\gamma}_{t,n}))_n, \qquad (w_{t,n})_n := (\ln \Phi(\boldsymbol{\zeta}_{t,n}))_n \tag{12}$$

are sub-additive sequences. By arguments already used, we have $\mathbb{E}|\ln \Phi(\boldsymbol{\gamma}_{t,1})| = \mathbb{E}|\ln \Phi(\boldsymbol{\zeta}_{t,1})| = \mathbb{E}|\ln \Phi(f_t)| < \infty$. In view of (11) and Lemma 2(i),

$$\limsup_{n\to\infty} \frac{1}{n} \ln z_{p,t+n}(v) \leq \limsup_{n\to\infty} \frac{1}{n} w_{t,n} + \limsup_{n\to\infty} \frac{1}{n} \ln z_{p,t}(v) \quad \text{a.s.}$$

Because $z_{p,t}(v)$ does not depend on n, we have $\limsup_{n\to\infty} (1/n) \ln z_{p,t}(v) = 0$ a.s. Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \ln z_{p,t+n}(v) \leq \limsup_{n\to\infty} \frac{1}{n} w_{t,n} \quad \text{a.s.} \tag{13}$$

Since, for any $n \in \mathbb{N}^*$, $u_{t,n}$ and $w_{t,n}$ have the same law, by (12) and Kingman's sub-additive ergodic theorem,

$$\limsup_{n\to\infty} \frac{1}{n} w_{t,n} = \limsup_{n\to\infty} \frac{1}{n} \mathbb{E} u_{t,n} = \limsup_{n\to\infty} \frac{1}{n} u_{t,n} \quad \text{a.s.} \tag{14}$$

On the other hand, in view of (8), we have, by the positivity of the coefficients,

$$\Phi(\boldsymbol{\gamma}_{t,n+1}) = \sum_{q=0}^{n} \left(\prod_{i=0}^{q-1} a_{p,t-i}(v) \right) b_{p,t-q} + \left(\prod_{i=0}^{n} a_{p,t-i}(v) \right) \overset{n\to\infty}{\to} z_{p,t}(v) \quad \text{a.s.}$$

Therefore, $\lim_{n\to\infty} u_{t,n} = \ln z_{p,t}(v)$ a.s., which entails

$$\limsup_{n\to\infty} \frac{1}{n} u_{t,n} = 0 \quad \text{a.s.} \tag{15}$$

By (13), (14), and (15), we get $\limsup_{n\to\infty} (1/n) \ln z_{p,t+n}(v) \leq 0$ a.s., which implies, by (10), part (i) of the lemma.

For (ii), by (10), (9), (5), and the ergodic theorem, we have

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \ln d(X_{Q_p(t-n)}, c) &\leq \limsup_{n\to\infty} \frac{1}{n} \ln z_{p,t-n}(v) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \ln \left(\prod_{i=0}^{n-1} a_{p,t-i}(v) \right) z_{p,t-n}(v) \\
&\quad - \liminf_{n\to\infty} \frac{1}{n} \ln \left(\prod_{i=0}^{n-1} a_{p,t-i}(v) \right) \leq -v \quad \text{a.s.}
\end{aligned}$$

for all $v \in [\gamma_0, 0)$. Letting $v \to 0^-$, we get the result. □

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

We are now ready to prove Theorem 2.

Proof of Theorem 2. For all $t \in \mathbb{Z}$, let $t' \in \mathbb{Z}$ and p', $0 \le p' \le r_0 - 1$, be such that $t = r_0 t' + p'$. Note that $\{t + k, k \in \mathbb{N}\} \subset \bigcup_{0 \le p \le r_0 - 1} \{r_0(t' + k) + p, k \in \mathbb{N}\}$. This and Lemma 3(i) imply that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \ln d(X_{t+n}, c) &\le \max_{0 \le p \le r_0 - 1} \left(\limsup_{n\to\infty} \frac{1}{Q_p(t'+n)} \ln d(X_{Q_p(t'+n)}, c) \right) \\ &\le C \max_{0 \le p \le r_0 - 1} \left(\limsup_{n\to\infty} \frac{1}{n} \ln d(X_{Q_p(t'+n)}, c) \right) \le 0\end{aligned}$$

for

$$C = \max_{0 \le p \le r_0 - 1} \left(\sup_{n \ge 0} \frac{n}{Q_p(t'+n)} \right),$$

which establishes (i). Part (ii) follows from similar arguments. □

4. Inference for semi-strong GARCH(*p*, *q*)

Consider the GARCH(p, q) model

$$\epsilon_t = \sqrt{h_t}\eta_t, \qquad h_t = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}, \quad \text{for all } t \in \mathbb{Z}, \tag{16}$$

where $\omega_0 > 0$, $\alpha_{0i} \geqslant 0$ ($i = 1, \ldots, q$), and $\beta_{0j} \geqslant 0$ ($j = 1, \ldots, p$). When (η_t) is i.i.d., the model in (16) is a standard strong GARCH, for which the statistical inference has been thoroughly studied. In particular, [1, 14] studied the QMLE under the stationarity of (ϵ_t), and [18] explored the asymptotic behavior of the QMLE in the explosive case. In the stationary framework, [10] proved the consistency and asymptotic normality of the QMLE without i.i.d.-ness for (η_t), but had to assume that $E|\epsilon_t|^s < \infty$ for some small $s > 0$. The aim of this section is to relax this extra moment assumption.

4.1. Property of the strictly stationary solution

Let

$$A_t = \begin{pmatrix} \alpha_{01}\eta_t^2 & \cdots & \alpha_{0q}\eta_t^2 & \beta_{01}\eta_t^2 & \cdots & \beta_{0p}\eta_t^2 \\ & I_{q-1} & & & 0_{(q-1)\times p} & \\ \alpha_{01} & \cdots & \alpha_{0q} & \beta_{01} & \cdots & \beta_{0p} \\ & 0_{(p-1)\times q} & & & I_{p-1} & \end{pmatrix}, \qquad b_t = \begin{pmatrix} \omega_0 \eta_t^2 \\ 0_{q-1} \\ \omega_0 \\ 0_{p-1} \end{pmatrix}$$

with standard notation.

The model in (16) is a special case of (1) using $\theta_t = (A_t, b_t)$, $X_t = (\epsilon_t^2, \ldots, \epsilon_{t-q+1}^2, h_t^2, \ldots, h_{t-p+1}^2)'$, $\Psi(\theta, x) = Ax + b$, and $d(x, y) = \|x - y\|$ for any norm $\|\cdot\|$ on $\mathbb{R}^{p+q}$. Note that $\mathbf{\Lambda}_t^{(r)} = \|A_t A_{t-1} \ldots A_{t-r+1}\|$.

In what follows, we do not assume that (η_t) is i.i.d., we only assume that it is stationary and ergodic. If $\mathbb{E} \ln^+ \eta_1^2 < \infty$, Theorem 1 applies with $c = 0_{p+q}$. Therefore, in view of (4), there exists a unique non-anticipative strictly stationary solution (ϵ_t) to model (16) if

$$\gamma(\mathbf{A}) := \inf_{r \in \mathbb{N}^*} \frac{1}{r} \mathbb{E}(\ln \|A_0 A_{-1} \ldots A_{-r+1}\|) = \lim_{r\to\infty} \frac{1}{r} \ln \|A_0 A_{-1} \ldots A_{-r+1}\| < 0 \quad \text{a.s.}$$

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

By Theorem 2, it follows that the strictly stationary solution of (16) satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \ln \boldsymbol{\epsilon}_{t+n}^2 \le 0, \qquad \limsup_{n\to\infty} \frac{1}{n} \ln \boldsymbol{\epsilon}_{t-n}^2 \le 0 \quad \text{a.s.} \tag{17}$$

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

In the GARCH(1,1) case, it is easy to check that $\gamma(\mathbf{A}) = \mathbb{E} \ln(\alpha_{01}\boldsymbol{\eta}_t^2 + \beta_{01})$. For general GARCH($p$,$q$) of the form (16), it seems impossible to compute $\gamma(\mathbf{A})$ explicitly. This issue has been discussed in several papers, e.g. [3, p. 117] and [6, pp. 148, 149]. Both papers recommend estimation by computer simulation.

4.2. QML estimator

Let $\{\boldsymbol{\epsilon}_t\}_{t=1}^n$ be a sample of size n of the unique non-anticipative strictly stationary solution of model (16). The vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{p+q+1})^\top = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)^\top$ belongs to a parameter space $\boldsymbol{\Theta} \subset]0, +\infty[\times [0, \infty[^{p+q}$. The true value of the parameter is unknown and is denoted by $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})^\top$. Conditionally on initial values $\boldsymbol{\epsilon}_0, \dots, \boldsymbol{\epsilon}_{1-q}, \tilde{\boldsymbol{\sigma}}_0^2, \dots, \tilde{\boldsymbol{\sigma}}_{1-p}^2$, the Gaussian quasi-likelihood is defined by

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}; \boldsymbol{\epsilon}_1 \dots, \boldsymbol{\epsilon}_n) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{\boldsymbol{\sigma}}_t^2}} \exp\left(-\frac{\boldsymbol{\epsilon}_t^2}{2\tilde{\boldsymbol{\sigma}}_t^2}\right),$$

where the $\tilde{\boldsymbol{\sigma}}_t^2$ are defined recursively, for $t \geqslant 1$, by

$$\tilde{\boldsymbol{\sigma}}_t^2 = \tilde{\boldsymbol{\sigma}}_t^2(\boldsymbol{\theta}) = \omega + \sum_{i=1}^q \alpha_i \boldsymbol{\epsilon}_{t-i}^2 + \sum_{j=1}^p \beta_j \tilde{\boldsymbol{\sigma}}_{t-j}^2.$$

For instance, the initial values can be chosen as

$$\boldsymbol{\epsilon}_0^2 = \dots = \boldsymbol{\epsilon}_{1-q}^2 = \tilde{\boldsymbol{\sigma}}_0^2 = \dots = \tilde{\boldsymbol{\sigma}}_{1-p}^2 = c, \tag{18}$$

with $c = \omega$ or $\boldsymbol{\epsilon}_1^2$. The standard estimator of the GARCH parameter $\boldsymbol{\theta}_0$ is the QMLE defined as any measurable solution $\hat{\boldsymbol{\theta}}_n$ of

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}), \tag{19}$$

where $\tilde{\mathbf{l}}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^n \tilde{\ell}_t$ and $\tilde{\ell}_t = \tilde{\ell}_t(\boldsymbol{\theta}) = (\boldsymbol{\epsilon}_t^2 / \tilde{\boldsymbol{\sigma}}_t^2) + \ln \tilde{\boldsymbol{\sigma}}_t^2$.

Let $\mathcal{A}_{\boldsymbol{\theta}}(z) = \sum_{i=1}^q \alpha_i z^i$ and $\mathcal{B}_{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^p \beta_j z^j$. It is not restrictive to assume that $q \ge 1$. By convention, $\mathcal{B}_{\boldsymbol{\theta}}(z) = 1$ if $p = 0$. Let $\mathcal{F}_{t-1}$ be the σ-field generated by $(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots)$. To show the strong consistency, we make the following assumptions.

Assumption 1. *$\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$ is compact.*

Assumption 2. *$\gamma(\mathbf{A}_0) < 0$ and, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\sum_{j=1}^p \beta_j < 1$.*

Assumption 3. *$(\boldsymbol{\eta}_t)$ is stationary and ergodic; $\boldsymbol{\eta}_t^2$ has a non-degenerate distribution with (i) $\mathbb{E}[\boldsymbol{\eta}_t^2 \mid \mathcal{F}_{t-1}] = 1$ a.s. and (ii) $\mathbb{E} \ln \boldsymbol{\eta}_t^2 > -\infty$.*

Assumption 4. *If $p > 0$, $\mathcal{A}_{\boldsymbol{\theta}_0}(z)$ and $\mathcal{B}_{\boldsymbol{\theta}_0}(z)$ have no common root, $\mathcal{A}_{\boldsymbol{\theta}_0}(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.*

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

Remark 2. Assumptions 1, 2, and 3 are standard (see [14] for comments on these assumptions). Assumption 3(i) is obviously less restrictive than the i.i.d. assumption with finite second-order moments. In Appendix D, we provide an explicit example of semi-strong GARCH based on a non-i.i.d. martingale difference innovation satisfying Assumption 3(i). This assumption was first used in [21] for the inference of GARCH models, and [10] established the consistency of the QMLE under this assumption, with a small-order moment condition on the observed process instead of our Assumption 3(ii). Note that the latter assumption precludes densities with too much mass around zero, but is satisfied by most commonly used distributions. It is also weaker than the regularity condition on the $\boldsymbol{\eta}_t$ law ($\lim_{t\to 0} t^{-\mu}\mathbb{P}\{\boldsymbol{\eta}_0^2 \leqslant t\} = 0$ for some $\mu > 0$) used in [1] (see Appendix E).

Assumption 2 implies that the roots of $\mathcal{B}_{\boldsymbol{\theta}}(z)$ are outside the unit disc. Therefore, by the second inequality of (17), we can define $(\boldsymbol{\sigma}_t^2) = \{\boldsymbol{\sigma}_t^2(\boldsymbol{\theta})\}$ as the (unique) strictly stationary, ergodic, and non-anticipative solution of

$$\boldsymbol{\sigma}_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \boldsymbol{\epsilon}_{t-i}^2 + \sum_{j=1}^{p} \beta_j \boldsymbol{\sigma}_{t-j}^2 \quad \text{for all } t; \tag{20}$$

see Appendix F.

Note that $\boldsymbol{\sigma}_t^2(\boldsymbol{\theta}_0) = h_t$. Let

$$\mathbf{l}_n(\boldsymbol{\theta}) = \mathbf{l}_n(\boldsymbol{\theta}; \boldsymbol{\epsilon}_n, \boldsymbol{\epsilon}_{n-1}\dots) = n^{-1}\sum_{t=1}^{n} \ell_t, \qquad \ell_t = \ell_t(\boldsymbol{\theta}) = \frac{\boldsymbol{\epsilon}_t^2}{\boldsymbol{\sigma}_t^2} + \ln \boldsymbol{\sigma}_t^2.$$

We are now able to establish the strong consistency of the QMLE.

Theorem 3. *Let $(\hat{\boldsymbol{\theta}}_n)$ be a sequence of QMLE satisfying (19), with any initial condition (18). Then, under Assumptions 1–4, $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$.*

Remark 3. [10] established the asymptotic normality of the QMLE under the assumption that a small-order moment exists. This moment condition is mainly used to justify the existence of the asymptotic covariance of the QMLE. To the best of our knowledge, the asymptotic normality has never been shown without a hypothesis that implies the existence of a small-order moment. In some cases, the asymptotic covariance matrix may not exist without a finite moment of sufficiently large order [15, Section 3.1]. Study of the asymptotic distribution of the semi-strong GARCH without any moment condition is left for future work.

Proof of Theorem 3. The proof relies on the following intermediate results.

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

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

(iii) If $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ then $\mathbb{E}\{\ell_1(\boldsymbol{\theta}) - \ell_1(\boldsymbol{\theta}_0)\} > 0$.

(iv) Any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ has a neighborhood $V(\boldsymbol{\theta})$ such that $\liminf_{n\to\infty} (\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0)) > 0$ a.s.

To prove (i), note that [14, (4.7)] shows that, almost surely,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbf{l}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta})| \leqslant \left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{\omega^2}\right\} C n^{-1} \sum_{t=1}^{n} \rho^t \boldsymbol{\epsilon}_t^2 + \left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{\omega}\right\} C n^{-1} \sum_{t=1}^{n} \rho^t$$

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

for some constants $C>0$ and $0<\rho<1$ (independent of n); (i) thus follows by Cesàro's lemma, since the first inequality of (17) implies that $\rho^t\boldsymbol{\epsilon}_t^2\to 0$ a.s. as $t\to\infty$:

$$\limsup_{n\to\infty}\frac{1}{k}\ln\rho^k\boldsymbol{\epsilon}_{t+k}^2\le\ln\rho+\limsup_{n\to\infty}\frac{1}{k}\ln\boldsymbol{\epsilon}_{t+k}^2\le\ln\rho<0.$$

The proof of (ii) uses the same arguments as those of step (ii) in the proof of [14, Theorem 2.1].

Now let us turn to the proof of (iii). For strong GARCH models it is known that $\mathbb{E}\ell_1(\boldsymbol{\theta}_0)$ is finite. This may not be the case in our framework, so we give an alternative proof of (iii). We first establish the existence of $\mathbb{E}\{\ell_1(\boldsymbol{\theta})-\ell_1(\boldsymbol{\theta}_0)\}$. Let $W_t(\boldsymbol{\theta})=\sigma_t^2(\boldsymbol{\theta}_0)/\sigma_t^2(\boldsymbol{\theta})$ and, for $K>0$, $A_K=[K^{-1},K]$, write

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

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

$$\ell_t(\boldsymbol{\theta})-\ell_t(\boldsymbol{\theta}_0)\ge g(W_t(\boldsymbol{\theta}),\boldsymbol{\eta}_t^2)\mathbb{1}_{W_t(\boldsymbol{\theta})\in A_K}-\{g(W_t(\boldsymbol{\theta}),\boldsymbol{\eta}_t^2)\}^-\mathbb{1}_{W_t(\boldsymbol{\theta})\in A_K^c}. \tag{21}$$

Noting that $W_t(\boldsymbol{\theta})$ is $\mathcal{F}_{t-1}$-measurable and, by Assumption 3(i), $\mathbb{E}[g(W_t(\boldsymbol{\theta}),\boldsymbol{\eta}_t^2)\mid\mathcal{F}_{t-1}]=g(W_t(\boldsymbol{\theta}),1)$, the expectation of the first term on the right-hand side of (21) is well-defined and satisfies

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

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

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

To deal with the expectation of the second term on the right-hand side of (21), we use the fact that, for $y>0$, $g(x,y)\ge g(1/y,y)$. It follows that

$$-\mathbb{E}[\{g(W_t(\boldsymbol{\theta}),\boldsymbol{\eta}_t^2)\}^-\mathbb{1}_{W_t(\boldsymbol{\theta})\in A_K^c}]\ge-\mathbb{E}[\{g(1/\boldsymbol{\eta}_t^2,\boldsymbol{\eta}_t^2)\}^-\mathbb{1}_{W_t(\boldsymbol{\theta})\in A_K^c}]\to 0\quad\text{as }K\to\infty,$$

because, by Assumption 3(ii), $\mathbb{E}[\{g(1/\boldsymbol{\eta}_t^2,\boldsymbol{\eta}_t^2)\}^-]<\infty$ and thus the convergence holds by Lebesgue's dominated convergence theorem. This completes the proof of (iii).

Now we prove (iv). As for (iii), the possible non-existence of $\mathbb{E}\ell_1(\boldsymbol{\theta})$ requires a modification of the standard proof. For any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ we have

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

Hence, using (i),

$$\begin{aligned}&\liminf_{n\to\infty}\Big(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}}\tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*)-\tilde{\mathbf{l}}_n(\boldsymbol{\theta}_0)\Big)\\&\quad\ge\liminf_{n\to\infty}\Big(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}}\mathbf{l}_n(\boldsymbol{\theta}^*)-\mathbf{l}_n(\boldsymbol{\theta}_0)\Big)-2\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{\mathbf{l}}_n(\boldsymbol{\theta})-\mathbf{l}_n(\boldsymbol{\theta})|\\&\quad=\liminf_{n\to\infty}\Big(\inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})\cap\boldsymbol{\Theta}}\mathbf{l}_n(\boldsymbol{\theta}^*)-\mathbf{l}_n(\boldsymbol{\theta}_0)\Big).\end{aligned}\tag{22}$$

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

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

$$\liminf_{n\to\infty}\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \mathbf{l}_n(\boldsymbol{\theta}^*)-\mathbf{l}_n(\boldsymbol{\theta}_0)\right) \geq \liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*)-\ell_t(\boldsymbol{\theta}_0). \tag{23}$$

By arguments already given, under Assumption 3(ii),

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

Therefore, $\mathbb{E}(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}^*)-\ell_t(\boldsymbol{\theta}_0))$ exists in $\mathbb{R}\cup\{+\infty\}$, and the ergodic theorem applies [16, Exercises 7.3, 7.4]). From (23) we obtain

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

The latter term in 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

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

which, by (i), is strictly positive. In view of (22), the proof of (iv) is complete.

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

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

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

Appendix A. Proof of Theorem 1

Proof For all $t\in\mathbb{Z}$ and $n\in\mathbb{N}$, let

$$X_{t,n} = \Psi\left(\boldsymbol{\theta}_t, X_{t-1,n-1}\right) \tag{24}$$

with $X_{t,0}=c$. Note that $X_{t,n}=\psi_n(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}, \ldots, \boldsymbol{\theta}_{t-n+1})$ for some measurable function $\psi_n\colon (E^n, \mathcal{B}_{E^n}) \to (F, \mathcal{B}_F)$, with the usual notation. For all n, the sequence $(X_{t,n})_{t\in\mathbb{Z}}$ is thus stationary and ergodic. If, for all t, the limit $X_t=\lim_{n\to\infty}X_{t,n}$ exists a.s., then by taking the limit of both sides of (24), it can be seen that the process (X_t) is a solution of (1). When it exists, the limit is a measurable function of the form $X_t=\psi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}, \ldots)$ and is therefore stationary, ergodic, and causal. For the measurability of X_t, we can consider the $X_{t,n}$ as functions of $(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}, \ldots)$ and argue that, in a metric space, a limit of measurable functions is measurable. The existence of $\lim_{n\to\infty}X_{t,n}$ was proved in [8], which showed that, a.s., the sequence $(X_{t,n})_{n\in\mathbb{N}}$ is a Cauchy sequence in the complete space F.

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

By iterating (24) we have $X_{t,n} = \mathbf{\Psi}_t \circ \cdots \circ \mathbf{\Psi}_{t-n+1}(c)$. It follows that

$$d\left(X_{t,n}, X_{t,n-1}\right) \le \mathbf{\Lambda}_t^{(n-1)} d(\mathbf{\Psi}_{t-n+1}(c), c).$$

For $n < m$, we thus have

$$\begin{aligned} d(X_{t,m}, X_{t,n}) &\le \sum_{k=0}^{m-n-1} d(X_{t,m-k}, X_{t,m-k-1}) \\ &\le \sum_{k=0}^{m-n-1} \mathbf{\Lambda}_t^{(m-k-1)} d(\mathbf{\Psi}_{t-m+k+1}(c), c) \le \sum_{j=n}^{\infty} \mathbf{\Lambda}_t^{(j)} d(\mathbf{\Psi}_{t-j}(c), c). \end{aligned} \tag{25}$$

Note that

$$\limsup_{j\to\infty} \ln\left(\mathbf{\Lambda}_t^{(j)} d(\mathbf{\Psi}_{t-j}(c), c)\right)^{1/j} = \limsup_{j\to\infty} \frac{1}{j}\left(\ln \mathbf{\Lambda}_t^{(j)} + \ln d(\mathbf{\Psi}_{t-j}(c), c)\right) < 0$$

under (i) and (ii), by using Kingman's sub-additive ergodic theorem [20] and [16, Exercise 4.12]. We conclude, from the Cauchy criterion for the convergence of series with positive terms, that $\sum_{j=1}^{\infty} \mathbf{\Lambda}_t^{(j)} d(\mathbf{\Psi}_{t-j}(c), c)$ is a.s. finite under (i) and (ii). It follows that $(X_{t,n})_{n\in\mathbb{N}}$ is a.s. a Cauchy sequence in F. The existence of a stationary and ergodic solution to (1) follows.

Assume that there exists another stationary process (X_t^*) such that $X_t^* = \mathbf{\Psi}_t(X_{t-1}^*)$. For all $N \ge 0$,

$$d(X_t, X_t^*) \le \mathbf{\Lambda}_t^{(N+1)} d(X_{t-N}, X_{t-N}^*). \tag{26}$$

Since $\mathbf{\Lambda}_t^{(N+1)} \to 0$ a.s. as $N \to \infty$, and $d(X_{t-N}, X_{t-N}^*) = O_P(1)$ by stationarity, the right-hand side of (26) tends to zero in probability. Since the left-hand side does not depend on N, we have $\mathbb{P}(d(X_t, X_t^*) > \epsilon) = 0$ for all $\epsilon > 0$, and thus $\mathbb{P}(X_t = X_t^*) = 1$, which establishes the uniqueness. In view of (25), we have $d(X_t, c) \le \sum_{j=0}^{\infty} \mathbf{\Lambda}_t^{(j)} d(\mathbf{\Psi}_{t-j}(c), c)$ and (2) follows. □

Appendix B. Proof of the comment following Theorem 2

For all $\epsilon > 0$, since $\mathbb{P}(\ln d(X_1, c) > \epsilon) = \mathbb{P}(\ln^+ d(X_1, c) > \epsilon)$,

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(n^{-1} \ln d(X_{t+n}, c) > \epsilon) &= \sum_{n=1}^{\infty} \mathbb{P}(n^{-1} \ln^+ d(X_1, c) > \epsilon) \\ &\le \int_0^{\infty} \mathbb{P}(t^{-1} \ln^+ d(X_1, c) > \epsilon)\, \mathrm{d}t \\ &= \int_0^{\infty} \mathbb{P}(\epsilon^{-1} \ln^+ d(X_1, c) > t)\, \mathrm{d}t \\ &= \epsilon^{-1} \mathbb{E} \ln^+ d(X_1, c) < \infty. \end{aligned}$$

It follows by the Borel–Cantelli lemma that $\limsup n^{-1} \ln d(X_{t+n}, c) \le 0$ a.s. The second result is obtained by the same arguments.

Appendix C. Proof of Corollary 1

We have, for all $n \ge 1$, $\sup_{k\ge n} \max(0, \ln d(X_{t+k}, c)) = \max(0, \sup_{k\ge n} \ln d(X_{t+k}, c))$. It follows that

$$\limsup_n \frac{1}{n} \ln^+ d(X_{t+n}, c) = \max\left(0, \limsup_n \frac{1}{n} \ln d(X_{t+n}, c)\right) = 0 \quad \text{a.s.}$$

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

Since, in addition, $\ln^+ d(X_{t+n}, c)$ is non-negative, $\lim_{n\to\infty} (1/n) \ln^+ d(X_{t+n}, c)$ exists and is equal to 0 a.s. We get $\lim_{n\to\infty} (1/n) \ln^+ d(X_{t-n}, c)$ by the same arguments, which gives the first part of the corollary.

For (3), we have $\ln d(X_{t+n}, c) = \ln^+ d(X_{t+n}, c) - \ln^- d(X_{t+n}, c)$. Since $(1/|n|) \ln^+ d(X_{t-n}, c)$ converges a.s. to 0 and $(1/|n|) \ln^- d(X_{t-n}, c)$ also converges a.s. to 0 as $|n| \to \infty$ [16, Exercise 2.13], $(1/|n|) \ln d(X_{t+n}, c)$ converges a.s. to 0 as $|n| \to \infty$.

Appendix D. Construction of a semi-strong GARCH

We first define a non-i.i.d. martingale difference process. Consider a sequence $(x_t)_{t\in\mathbb{Z}}$ of i.i.d. random variables with standard normal distribution. Since, for all $z \in \mathbb{R}_+$, $x_t\sqrt{2z} - z \sim \mathcal{N}(-z, 2z)$, using the moment-generating function of the Gaussian distribution, we have

$$\mathbb{E}[\exp(x_t\sqrt{2z} - z)] = 1. \tag{27}$$

If (z_t) is a positive process, independent of (x_t), we also have $\mathbb{E}\eta_t^2 = 1$, where $\eta_t^2 = \exp(x_t\sqrt{2z_t} - z_t)$. This is the case if, for instance, z_t follows a causal AR(1) model of the form $z_t = \phi z_{t-1} + u_t$ with $\phi \in (0, 1)$ and u_t i.i.d. with positive variance. It is easy to see that $\mathrm{Cov}(z_1, z_0) \neq 0$, and thus

$$\begin{aligned}\mathrm{Cov}\{\ln(\eta_1^2), \ln(\eta_0^2)\} &= 2\mathbb{E}\{x_1\sqrt{z_1}x_0\sqrt{z_0}\} - \mathbb{E}\{x_1\sqrt{2z_1}z_0\} \\ &\quad - \mathbb{E}\{z_1x_0\sqrt{2z_0}\} + \mathbb{E}\{z_1z_0\} - \mathbb{E}z_1\mathbb{E}z_0 \\ &= \mathrm{Cov}\{z_1, z_0\} \neq 0.\end{aligned}$$

It follows that (η_t^2) is not i.i.d. We now define (η_t). Let (r_t) be an i.i.d. sequence of Rademacher variables (uniform distribution on $\{-1, 1\}$), independent of the two sequences (x_t) and (u_t). We define (η_t) by $\eta_t = r_t\sqrt{\eta_t^2}$.

Let $(\mathcal{F}_t)$ be the canonical filtration of (η_t), i.e. $\mathcal{F}_t = \sigma(\eta_k, k \le t)$. Define a second filtration $\mathcal{H}_t = \sigma(r_k, x_{k+1}, u_{k+1}, k \le t)$. Since $\mathcal{F}_t \subset \mathcal{H}_t$ and r_t is independent of $\mathcal{H}_{t-1}$, we have

$$\mathbb{E}[\eta_t \mid \mathcal{F}_{t-1}] = \mathbb{E}\{\mathbb{E}[\eta_t \mid \mathcal{H}_{t-1}] \mid \mathcal{F}_{t-1}\} = \mathbb{E}\{\exp([x_t\sqrt{2z_t} - z_t]/2)\mathbb{E}[r_t \mid \mathcal{H}_{t-1}] \mid \mathcal{F}_{t-1}\} = 0.$$

Define a new filtration $\mathcal{I}_t = \sigma(r_k, x_k, u_{k+1}, k \le t)$. Since $\mathcal{F}_t \subset \mathcal{I}_t$, z_t is $\mathcal{I}_{t-1}$-measurable, and x_t is independent of $\mathcal{I}_{t-1}$, so by (27) we have

$$\mathbb{E}\left[\eta_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\{\mathbb{E}[\eta_t^2 \mid \mathcal{I}_{t-1}] \mid \mathcal{F}_{t-1}\} = \mathbb{E}\{\mathbb{E}[\exp(x_t\sqrt{2z_t} - z_t) \mid \mathcal{I}_{t-1}] \mid \mathcal{F}_{t-1}\} = 1.$$

We have thus shown the existence of a non-degenerate unit martingale difference sequence, that is, a stationary and ergodic sequence (η_t) satisfying the conditions

$$\mathbb{E}[\eta_t^2] < \infty, \qquad \mathbb{E}[\eta_t \mid \mathcal{F}_{t-1}] = 0, \qquad \mathbb{E}[\eta_t^2 \mid \mathcal{F}_{t-1}] = 1, \qquad (\eta_t^2) \text{ are not i.i.d.}$$

It is then easy to define a semi-strong GARCH with innovations (η_t).

Appendix E. Complement to Remark 2

Knowing that $\mathbb{E}(\ln^+(\eta_1^2)) < \infty$ by Assumption 3(i), to establish Assumption 3(ii) it is therefore sufficient to prove that $\mathbb{E}(\ln^-(\eta_1^2)) < \infty$. Using $\mathbb{E}(\ln^-(\eta_1^2)) = \int_0^\infty \mathbb{P}(\ln^+(1/\eta_1^2) \ge s)\,\mathrm{d}s = \int_0^\infty \mathbb{P}(\ln(1/\eta_1^2) \ge s)\,\mathrm{d}s = \int_0^\infty \mathbb{P}(1/\eta_1^2 \ge \exp(s))\,\mathrm{d}s = \int_0^\infty \mathbb{P}(\eta_1^2 \le \exp(-s))\,\mathrm{d}s$, we have, under the condition of [1], that $\mathbb{P}(\eta_1^2 \le \exp(-s)) = o(\exp(-\mu s))$ when $s \to \infty$, which gives the result.

https://doi.org/10.1017/jpr.2023.13 Published online by Cambridge University Press

Appendix F. Proof of the existence of a unique strictly stationary solution to (20)

Rewriting (20) in vector form as $\underline{\boldsymbol{\sigma}}_t^2 = \underline{c}_t + B\underline{\boldsymbol{\sigma}}_{t-1}^2$, where

$$\underline{\boldsymbol{\sigma}}_t^2 = \begin{pmatrix} \boldsymbol{\sigma}_t^2 \\ \boldsymbol{\sigma}_{t-1}^2 \\ \vdots \\ \boldsymbol{\sigma}_{t-p+1}^2 \end{pmatrix}, \qquad \underline{c}_t = \begin{pmatrix} \omega + \sum_{i=1}^{q} \alpha_i \boldsymbol{\epsilon}_{t-i}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix},$$

we have, by the second inequality of (17) that $\limsup_{n\to\infty} (1/n) \ln \|\underline{c}_n\| \le 0$. By Assumption 2, we deduce that

$$\limsup_{n\to\infty} \frac{1}{n} \ln \|B^n \underline{c}_{n-1}^2\| \le \limsup_{n\to\infty} \frac{1}{n} \ln \|B^n\| + \limsup_{n\to\infty} \frac{1}{n} \ln \|\underline{c}_n\| < 0.$$

From this, we deduce by the Cauchy rule that the series $\hat{\boldsymbol{\sigma}}_t^2 := \sum_{n=0}^{\infty} B^n \underline{c}_{t-n}^2$ converges almost surely. We note that $(\hat{\boldsymbol{\sigma}}_t^2)$ is a strictly stationary, ergodic, and non-anticipative solution of (20).

To show the uniqueness, assume that there exists another stationary process $(\underline{\boldsymbol{\sigma}}_t^2*)$ of (20). For all $n \ge 0$, we have $\|\underline{\boldsymbol{\sigma}}_t^2* - \hat{\boldsymbol{\sigma}}_t^2\| = \|B^n \underline{\boldsymbol{\sigma}}_{t-n}^2* - B^n \hat{\boldsymbol{\sigma}}_{t-n}^2\| \le \|B^n\| \|\underline{\boldsymbol{\sigma}}_{t-n}^2*\| + \|B^n\| \|\hat{\boldsymbol{\sigma}}_{t-n}^2\|$. Since $\|B^n\| \to 0$ a.s. as $n \to \infty$ and $\|\underline{\boldsymbol{\sigma}}_{t-n}^2*\|$ and $\|\hat{\boldsymbol{\sigma}}_{t-n}^2\|$ converge in law by stationarity, Slutsky's theorem entails that $\|\underline{\boldsymbol{\sigma}}_t^2* - \hat{\boldsymbol{\sigma}}_t^2\|$ converges in law to 0 as $n \to \infty$. Since $\|\underline{\boldsymbol{\sigma}}_t^2* - \hat{\boldsymbol{\sigma}}_t^2\|$ does not depend on n, we conclude that $\|\underline{\boldsymbol{\sigma}}_t^2* - \hat{\boldsymbol{\sigma}}_t^2\| = 0$ a.s.

Acknowledgements

I am most thankful to the Editor and to two referees for their constructive comments and suggestions. I also want to acknowledge Christian Francq and Jean-Michel Zakoan for their guidance and feedback.

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

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

[2] Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM J. Control Optimization* **31**, 942–959.

[3] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.

[4] Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *Ann. Prob.* **20**, 1714–1730.

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

[6] Buraczewski, D., Damek, E. and Mikosch, T. (2016). *Stochastic Models with Power-Law Tails*. Springer, New York.
[7] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 45–76.
[8] Elton, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stoch. Process. Appl.* **34**, 39–47.
[9] Engle, R. F., Ghysels, E. and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Rev. Econom. Statist.* **95**, 776–797.
[10] Escanciano, J. C. (2009). Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* **25**, 561–570.
[11] Francq, C., Kandji, B. M. and Zakoian, J.-M. (2023). Inference on multiplicative component GARCH without any small-order moment. Submitted.
[12] Francq, C. and Thieu, L. Q. (2019). QML Inference for volatility models with covariates. *Econometric Theory* **35**, 37–72.
[13] Francq, C., Wintenberger, O. and Zakoian, J.-M. (2018). Goodness-of-fit tests for Log-GARCH and EGARCH models. *Test* **27**, 27–51.
[14] Francq, C. and Zakoian, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605–637.
[15] Francq, C. and Zakoian, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stoch. Process. Appl.* **117**, 1265–1284.
[16] Francq, C. and Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley, Chichester.
[17] Han, H. and Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *J. Business Econom. Statist.* **32**, 416–429.
[18] Jensen, S. T. and Rahbek, A. (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203–1226.
[19] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.
[20] Kingman, J. F. C. (1973). Subadditive ergodic theory. *Ann. Prob.* **1**, 883–899.
[21] Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. *Econometric Theory* **10**, 29–52.
[22] Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* **34**, 2449–2495.
[23] Tanny, D. (1974). A zero–one law for stationary sequences. *Z. Wahrscheinlichkeitsth.* **30**, 139–148.