

MIXING PROPERTIES OF A GENERAL CLASS OF GARCH(1,1) MODELS WITHOUT MOMENT ASSUMPTIONS ON THE OBSERVED PROCESS

CHRISTIAN FRANCO
Université Lille 3, GREMARS

JEAN-MICHEL ZAKOÏAN
Université Lille 3, GREMARS
and
CREST

We consider general, and possibly nonparametric, GARCH(1,1) processes. First we give conditions for the existence and the uniqueness of stationary ergodic solutions. Then we identify additional conditions for geometric ergodicity. These conditions consist of mild restrictions on the distribution of the latent independent process. No moment assumption is made on the generalized autoregressive conditionally heteroskedastic (GARCH) process. Applications to the asymptotic behavior of sample autocorrelations and to unit-root tests are proposed.

1. INTRODUCTION

With the increasing popularity of generalized autoregressive conditional heteroskedasticity (GARCH) modeling, there is also increased interest in general, even nonparametric, models and in moving away from the particular specification of the classical GARCH models, as introduced by Engle (1982) and Bollerslev (1986). In this paper, we assume that (ϵ_t) belongs to the general class of GARCH(1,1) processes, defined by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ h_t =: h(\sigma_t) = \omega(\eta_{t-1}) + a(\eta_{t-1})h_{t-1}, \end{cases} \quad (1)$$

where the sequence (η_t) is independent and identically distributed (i.i.d.). For statistical purposes the assumption that η_t has zero mean and unit variance is

This work was supported by INTAS (research project 03-51-3714). The authors gratefully acknowledge the quick and careful reading of the manuscript by Bruce Hansen and three referees. Their detailed comments led to a greatly improved presentation. Address correspondence to: Jean-Michel Zakoïan, CREST, MK1, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France; e-mail: zakoian@ensae.fr.

often required, but we do not need this assumption in this paper. The following assumptions are made on the functions ω , a , and h .

- (i) $\omega: \mathbb{R} \rightarrow \mathbb{R}^+$ is such that its restrictions to \mathbb{R}^+ and \mathbb{R}^- are either constant and strictly positive, or continuous and, respectively, strictly increasing and strictly decreasing;
- (ii) $a: \mathbb{R} \rightarrow \mathbb{R}^+$ is such that its restrictions to \mathbb{R}^+ and \mathbb{R}^- are continuous and, respectively, strictly increasing and strictly decreasing;
- (iii) $h: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is 1-1, onto and increasing.

The standard GARCH(1,1) model is obtained for $\omega(x) = \omega$, $h(x) = x^2$, and $a(x) = \alpha x^2 + \beta$ with $\alpha \geq 0, \beta \geq 0$. This specification has been found adequate for a number of financial series and is arguably the most popular volatility model. When h has a form inspired by the Box–Cox transformation, and for some particular specifications of the functions ω and a , we get the augmented GARCH introduced by Duan (1997). For $h(x) = x^\delta$ we get the class of GARCH(1,1) models defined by He and Teräsvirta (1999), which includes a variety of other first-order specifications.¹ See Ling and McAleer (2002) for strict stationarity and moment conditions for such models. Note that with the preceding assumptions, the volatility σ_t increases with the magnitude of positive “shocks” η_{t-1} as well as it increases with the magnitude of negative ones. Yet positive and negative shocks may have different impacts on the volatility.

Nelson (1990) and Bougerol and Picard (1992) showed in the standard GARCH case that if $E\eta_t = 0$ and $\text{Var}(\eta_t) = 1$,

$$E \log a(\eta_t) < 0 \tag{2}$$

is a necessary and sufficient condition for the existence of a unique strictly stationary and nonanticipative solution to model (1). A nonanticipative solution is a process (ϵ_t) such that ϵ_t is a measurable function of the variables η_{t-s} , $s \geq 0$. The extension to model (1) will be given subsequently. For statistical inference, however, strict stationarity is not a sufficient assumption, and it can be crucial to know when the stationary solution possesses mixing properties. Knowing that these properties hold may make it possible or easier to establish other properties such as central limit theorems (CLTs).

Mixing properties of classes of models including GARCH-type processes have been investigated by Ango Nze (1992, 1998), Lu (1996), Carrasco and Chen (2002), Rahbek, Hansen, and Dennis (2002), Lee and Shin (2004), Hwang and Kim (2004), and Meitz and Saikkonen (2004), among others. Unfortunately, when applied to standard GARCH processes, their results require moment assumptions that are much stronger than the strict stationarity assumption. Typically the condition $\alpha + \beta < 1$ is imposed for the standard GARCH(1,1), which amounts to restricting the class of strictly stationary solutions to those admitting a second-order moment. To our knowledge the most significant contribution, specifically devoted to the standard GARCH(p, q), is the dissertation by

Boussama (1998), which establishes strong mixing under conditions we will further discuss. However, the proof relies on heavy geometric algebra based upon the Mokkadem (1990) result for polynomial autoregressive processes. See also Kristensen (2006).

Our main contribution is to show that under (2), the β -mixing of the strictly stationary solution holds without any additional restriction on the function $a(\cdot)$. In particular we do not make any moment assumptions on the process (ϵ_t) . We provide simple sufficient conditions on the process η_t under which the strictly stationary solution to model (1) is β -mixing with exponential decay. We do not impose a continuous distribution for η_t , contrary to the preceding references dealing with mixing. This may have interest for financial applications because prices, and hence returns, are not observed continuously but are multiples of a monetary unit called the tick. A continuous distribution for the i.i.d. process would typically imply a continuous distribution for ϵ_t . On the other hand, dealing with the mixing properties of discrete-valued time-series models is in general a difficult task. For these reasons, and for the sake of generality, we will allow for both a discrete and a continuous part in the distribution of η_t . We rely on the results displayed in the book by Meyn and Tweedie (1996) so that the proof can be easily followed.

The fact that requirements for the existence of second-order moments can be ignored is particularly important for the statistical inference of GARCH(1,1) models. Indeed, recent references establish asymptotic normality of the maximum likelihood estimator essentially under the assumption (2) using the martingale theory.² In this framework, the asymptotic distribution of other statistics of interest (such as the autocorrelations of functions of ϵ_t) may be difficult, if not impossible, to derive using limit theorems for martingale differences. In such situations, the β -mixing property will be of invaluable help.

In the next section we give, for the reader's convenience, the Markov chain results we need. Section 3 is devoted to strict stationarity. In Section 4 we establish geometric ergodicity of the strictly stationary solution. Two statistical applications are proposed in Section 5.

2. BASIC MARKOV CHAIN THEORY

This section is drawn from the papers by Tjøstheim (1990) and Basrak, Davis, and Mikosch (2002) and the book by Meyn and Tweedie (1996). All the random variables considered in this paper are defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\{X_t, t \geq 0\}$ be a homogeneous Markov chain on (E, \mathcal{E}) where $E \subset \mathbb{R}^d$ and \mathcal{E} is the Borel σ -field on E . We denote the probability of moving from x to the set B in t steps by

$$P^t(x, B) = \mathbb{P}(X_t \in B | X_0 = x), \quad x \in E, \quad B \in \mathcal{E}.$$

The Markov chain (X_t) is ϕ -irreducible if, for some nontrivial measure ϕ on (E, \mathcal{E}) ,

$$\forall B \in \mathcal{E}, \phi(B) > 0 \Rightarrow \forall x \in E, \exists t > 0, P^t(x, B) > 0.$$

If (X_t) is ϕ -irreducible, there exists a maximal irreducibility measure M (see Meyn and Tweedie, 1996, Prop. 4.2.2), and we set $\mathcal{E}^+ = \{B \in \mathcal{E} \mid M(B) > 0\}$. We call the chain positive recurrent if

$$\limsup_{t \rightarrow \infty} P^t(x, B) > 0, \quad \forall x \in E, \quad \forall B \in \mathcal{E}^+.$$

For a ϕ -irreducible Markov chain, positive recurrence is equivalent (see Meyn and Tweedie, 1996, Thm. 18.2.2) to the existence of a (unique) invariant probability measure, that is, a probability π such that

$$\forall B \in \mathcal{E}, \pi(B) = \int P(x, B) \pi(dx).$$

Let $\|\cdot\|$ denote the total variation norm. The Markov chain (X_t) is said to be geometrically ergodic if there exists a $\rho, \rho \in (0, 1)$, such that

$$\forall x \in E, \rho^{-t} \|P^t(x, \cdot) - \pi\| \rightarrow 0 \quad \text{as } t \rightarrow +\infty. \tag{3}$$

Recall that for a stationary process, the strong (α -) mixing coefficients are defined by

$$\alpha_X(k) = \sup_{f, g} |\text{Cov}(f(\dots, X_{-1}, X_0), g(X_k, X_{k+1}, \dots))| \tag{4}$$

$$= \sup_{A, B} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, \tag{5}$$

where the first supremum is taken over the set of measurable functions f and g such that $|f| \leq 1, |g| \leq 1$, and the second supremum is taken over the sets $A \in \sigma(X_s, s \leq 0)$ and $B \in \sigma(X_s, s \geq k)$, whereas the β -mixing coefficients are defined by

$$\beta_X(k) = E \sup_{B \in \sigma(X_s, s \geq k)} |\mathbb{P}(B | \sigma(X_s, s \leq 0)) - \mathbb{P}(B)| \tag{6}$$

$$= \frac{1}{2} \sup \sum_{i=1}^I \sum_{j=1}^J |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|, \tag{7}$$

where in the last equality the sup is taken over all pairs of partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of Ω such that $A_i \in \sigma(X_s, s \leq 0)$ for each i and $B_j \in \sigma(X_s, s \geq k)$ for each j . The process is called α -mixing (resp. β -mixing) if $\lim_{k \rightarrow \infty} \alpha_X(k) = 0$ (resp. $\lim_{k \rightarrow \infty} \beta_X(k) = 0$). We have $\alpha_X(k) \leq \beta_X(k)$, so that β -mixing implies α -mixing. If $Y = (Y_t)$ is a process such that $Y_t = f(X_t, \dots, X_{t-r})$ for some measurable function f and some integer $r \geq 0$, then $\sigma(Y_t, t \leq s) \subset \sigma(X_t, t \leq s)$ and $\sigma(Y_t, t \geq s) \subset \sigma(X_{t-r}, t \geq s)$. Thus

$$\alpha_Y(k) \leq \alpha_X(k - r) \quad \text{and} \quad \beta_Y(k) \leq \beta_X(k - r) \quad \forall k \geq r. \tag{8}$$

Note that for a stationary Markov process we have $\alpha_X(k) = \sup_{f,g} |\text{Cov}(f(X_0), g(X_k))|$, where f and g are as in the previous definition (see Bradley, 1986). One consequence of the geometric ergodicity is that the Markov chain (X_t) is β -mixing, and hence *strongly mixing, with geometric rate*. Indeed, Davydov (1973) showed that for an ergodic Markov chain (X_t) with invariant probability measure π ,

$$\beta_X(k) = \int \|P^k(x, \cdot) - \pi\| \pi(dx).$$

Thus $\beta_X(k) = O(\rho^k)$ if (3) holds.

To state the following criterion for the geometric ergodicity of a Markov chain, we need the idea of a Feller chain. We call (X_t) a *Feller Markov chain* (or weak Feller chain) if the function

$$E(g(X_t) | X_{t-1} = x), \quad x \in E$$

is continuous for every bounded and continuous function g on E .

THEOREM 1 (Feigin and Tweedie, 1985, Thm. 1). *Assume that*

- (i) (X_t) is a Feller Markov chain;
- (ii) (X_t) is ϕ -irreducible for some measure ϕ on (E, \mathcal{E}) ;
- (iii) there exists a compact set $C \subset E$ such that $\phi(C) > 0$ and a nonnegative continuous function (test function) $V: E \rightarrow \mathbb{R}$ such that

$$V(x) \geq 1, \quad \forall x \in C \tag{9}$$

and for some $c > 0$

$$E[V(X_t) | X_{t-1} = x] \leq (1 - c)V(x), \quad \forall x \notin C. \tag{10}$$

Then (X_t) is geometrically ergodic.

3. STRICT STATIONARITY

The assumption that the sequence (η_t) is i.i.d. with $E \log^+ \{a(\eta_t)\} < \infty$, where $\log^+ x = \max(\log x, 0)$ for $x \geq 0$, is maintained throughout. Our first result can be deduced from results established by Bougerol and Picard (1992) for generalized autoregressive vector equations. But we prefer to give a simple self-contained proof of this theorem.

THEOREM 2. *If (2) holds, then the series*

$$h_t = \omega(\eta_{t-1}) + \sum_{i=1}^{\infty} a(\eta_{t-1}) \dots a(\eta_{t-i}) \omega(\eta_{t-i-1}) \tag{11}$$

converges almost surely (a.s.), and the process (ϵ_t) , defined by $\epsilon_t = h^{-1}(h_t)\eta_t$, is a strictly stationary solution of (1). This solution is unique, nonanticipative, and ergodic.

If (2) does not hold and $P[\eta_t = 0] \neq 1$, there exists no strictly stationary solution to model (1).

Proof. First note that $\gamma =: E \log a(\eta_t)$ exists in $[-\infty, +\infty)$. Now let

$$h_t = \lim_{N \rightarrow \infty} h_t(N) \quad \text{a.s.} \quad \text{where}$$

$$h_t(N) = \omega(\eta_{t-1}) + \sum_{i=1}^N a(\eta_{t-1}) \dots a(\eta_{t-i}) \omega(\eta_{t-i-1}), \tag{12}$$

the limit being well defined in $\bar{\mathbb{R}}^+$ in view of the positivity of the summands. Because $h_t(N) = \omega(\eta_{t-1}) + a(\eta_{t-1})h_{t-1}(N - 1)$ for all N , we have, letting N go to infinity, $h_t = \omega(\eta_{t-1}) + a(\eta_{t-1})h_{t-1}$. It remains to show that h_t is a.s. finite. Let $\omega^*(x) = \omega(x) + \tau \mathbb{1}_{\omega(x)=0}$ for some constant $\tau > 0$ and $x \in \mathbb{R}$. Let $h_t^*(N)$ be obtained by replacing ω by ω^* in $h_t(N)$ and denote by h_t^* its a.s. limit. We have

$$\begin{aligned} & [a(\eta_{t-1}) \dots a(\eta_{t-n}) \omega^*(\eta_{t-n-1})]^{1/n} \\ &= \exp \left[\frac{1}{n} \sum_{i=1}^n \log \{a(\eta_{t-i})\} + \frac{1}{n} \log \{\omega^*(\eta_{t-n-1})\} \right] \rightarrow e^\gamma \quad \text{a.s.} \end{aligned} \tag{13}$$

as $n \rightarrow \infty$, by the strong law of large numbers applied to the i.i.d. sequence $(\log \{a(\eta_t)\})$. It follows from the Cauchy rule that for any t , the sequence $\{h_t^*(N), N \geq 1\}$ converges a.s. in \mathbb{R} . Because $h_t \leq h_t^*$ we thus have $h_t < \infty$ a.s. As a function of an i.i.d. sequence, the limit h_t is thus strictly stationary and ergodic, in which case so is ϵ_t .

To prove uniqueness, let (\tilde{h}_t) be another strictly stationary solution process of (1). Suppose $\mathbb{P}(h_t \neq \tilde{h}_t) > 0$ for some t . Iterating the second equation in (1) we have $|h_t - \tilde{h}_t| = a(\eta_{t-1}) \dots a(\eta_{t-n}) |h_{t-n-1} - \tilde{h}_{t-n-1}|$. From the strong law of large numbers and (2), we have $a(\eta_{t-1}) \dots a(\eta_{t-n}) \rightarrow 0$ with probability 1 as $n \rightarrow \infty$. Thus $\mathbb{P}(|h_{t-n-1} - \tilde{h}_{t-n-1}| \rightarrow \infty) > 0$, which entails $|h_{t-n-1}| \rightarrow \infty$ or $|\tilde{h}_{t-n-1}| \rightarrow \infty$ with nonzero probability. This is not possible because the sequences $(h_t)_t$ and $(\tilde{h}_t)_t$ are stationary. Therefore $h_t = \tilde{h}_t$ for any t , a.s.

To prove the necessary part, suppose there exists a strictly stationary solution (h_t) of (1). We have for $n > 0$,

$$h_0 \geq \omega(\eta_{-1}) + \sum_{i=1}^n a(\eta_{-1}) \dots a(\eta_{-i}) \omega(\eta_{-i-1}),$$

from which we deduce that $a(\eta_{-1}) \dots a(\eta_{-n})\omega(\eta_{-n-1})$ converges to zero, a.s., when $n \rightarrow \infty$, or, equivalently, that

$$\sum_{i=1}^n \log a(\eta_i) + \log \omega(\eta_{-n-1}) \rightarrow -\infty \text{ a.s. as } n \rightarrow \infty. \tag{14}$$

First suppose $E \log\{a(\eta_i)\} > 0$. Then by the strong law of large numbers, $\sum_{i=1}^n \log a(\eta_i) \rightarrow +\infty$ a.s., and it is necessary for (14) to hold that $\log \omega(\eta_{-n-1}) \rightarrow -\infty$ a.s. This convergence implies $\omega(\eta_0) = 0$ a.s., which is precluded because η_t is not identically equal to zero. Now suppose $E \log\{a(\eta_i)\} = 0$. By the Chung–Fuchs theorem (see, e.g., Chow and Teicher, 1997) we have $\overline{\lim} \sum_{i=1}^n \log a(\eta_i) = +\infty$ with probability 1 and, using the elementary Lemma 1, which follows, the convergence (14) entails $\log \omega(\eta_{-n-1}) \rightarrow -\infty$ in probability. Thus, we are led to a contradiction as in the previous case. Thus, the assumption that a strictly stationary solution exists when $E \log\{a(\eta_i)\} \geq 0$ entails a contradiction. ■

LEMMA 1. *If (X_n) and (Y_n) are two independent sequences of random variables such that $X_n + Y_n \rightarrow -\infty$ and $X_n \not\rightarrow -\infty$ in probability, then $Y_n \rightarrow -\infty$ in probability.*

Remark 1. It can be seen from the proof that a solution (h_t) , as given by (12), always exists in \mathbb{R}^+ but that when (2) does not hold and $\mathbb{P}(\eta_t = 0) \neq 1$, this solution satisfies $\mathbb{P}[h_t = +\infty] > 0$ and $\mathbb{P}[|\epsilon_t| = +\infty] > 0$. See Klüppelberg, Lindner, and Maller (2004) for more detailed results in the standard GARCH(1,1) case.

4. GEOMETRIC ERGODICITY

To prove geometric ergodicity we require additional assumptions on the i.i.d. process (η_t) , essentially to ensure that the transition kernel has a Lebesgue component.

Assumption A. The distribution \mathbb{P}_η of the variable η_t is a mixture of an absolutely continuous component with respect to the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and Dirac masses at some points $\mu_i \in \mathbb{R}$, $i = 1 \dots N$. With standard notation we then have

$$d\mathbb{P}_\eta = \sum_{i=1}^N p_i d\delta_{\mu_i} + (1 - p) f d\lambda, \quad p_i \geq 0, \quad \sum_{i=1}^N p_i = p \in [0, 1), \tag{15}$$

where f is a density of the continuous component. Let $\eta_+^0 = \inf\{\eta | \eta > 0, f(\eta) > 0\}$ and $\eta_-^0 = \sup\{\eta | \eta < 0, f(\eta) > 0\}$, when these sets are nonempty, and assume that

$$(\eta_-^0 - \tau, \eta_-^0) \cup (\eta_+^0, \eta_+^0 + \tau) \subset \{f > 0\}$$

for some $\tau > 0$. By convention $(\eta_-^0 - \tau, \eta_-^0) = \emptyset$ (resp. $(\eta_+^0, \eta_+^0 + \tau) = \emptyset$) when η_-^0 (resp. η_+^0) is not defined. Finally, $E\{\omega(\eta_t)^r\} < \infty$ and $E\{a(\eta_t)^r\} < \infty$ for some $r > 0$.

Remark 2. The standard case where \mathbb{P}_η is absolutely continuous with respect to the Lebesgue measure is obtained by taking $p = 0$. Note however that the case $p = 1$, that is, when the law of η_t has no continuous component, is excluded. In such a case, criteria based on topological properties of the chain fail to prove ergodic properties (see a similar example in Meyn and Tweedie, 1996, p. 127). This does not mean that the process is not geometrically ergodic in those situations: for example, in the standard GARCH(1,1) case, if $\eta_t^2 = 1$ with probability 1, then the strictly stationary solution process is an independent white noise, which is obviously geometrically ergodic.

The main result of this paper is as follows.

THEOREM 3. *Under Assumption A and if the strict stationarity condition (2) holds, then the strictly stationary and nonanticipative solution (ϵ_t) of the GARCH(1,1) model (1) is β -mixing with exponential decay.*

Remark 3. The proof of this theorem relies on showing that (2) entails the geometric ergodicity of (h_t) . Moreover, geometric ergodicity implies strict stationarity. Under Assumption A, which entails $\mathbb{P}[\eta_t = 0] \neq 1$, condition (2) is therefore necessary and sufficient for the existence of a geometrically ergodic solution (h_t) and also for the existence of a strictly stationary and geometrically β -mixing solution (ϵ_t) .

Remark 4. To our knowledge, existing results on mixing conditions for non-standard GARCH(1,1) processes (see references in the introduction) are demanding in terms of moment assumptions. For instance, in Carrasco and Chen (2002), the mixing properties are obtained for various GARCH(1,1) models under moment conditions on the process (ϵ_t) (see their Table 1). By contrast, we find that the strictly stationary solution is β -mixing without any moment restriction.

Remark 5. When applied to standard GARCH(1,1) models, this theorem is also more general than those already established. In Boussama (1998), the geometric ergodicity of standard GARCH models is proved under the assumption that η_t has an absolutely continuous distribution with respect to the Lebesgue measure (i.e., $p = 0$ in our framework), with a positive density in a neighborhood of zero. In this case Assumption A holds with $\eta_-^0 = \eta_+^0 = 0$. Note however that our assumption allows us to handle more general cases, where the density is null on a neighborhood of zero or where the distribution of η_t does not admit a density with respect to the Lebesgue measure.

Before proving the theorem, we start by establishing geometric ergodicity of (h_t) .

LEMMA 2. Under the assumptions of Theorem 3, the strictly stationary and nonanticipative solution (h_t) of model (1) is geometrically ergodic.

Proof. By the second equation in model (1), (h_t) is obviously a homogeneous Markov chain on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$. The proof consists in checking the three conditions of Theorem 1.

Step (i): Feller property. For any bounded and continuous function g on \mathbb{R}^+ , the function

$$E[g(h_t) | h_{t-1} = x] = E[g\{\omega(\eta_t) + a(\eta_t)x\}]$$

is continuous in x over \mathbb{R}^+ , by the Lebesgue dominated convergence theorem, which shows that the Markov chain (h_t) is Feller.

Step (ii): Irreducibility. Let $\tau' \in (0, \tau)$ be small enough so that the set $D_{\tau'} =: (\eta_-^0 - \tau', \eta_-^0) \cup (\eta_+^0, \eta_+^0 + \tau')$ does not contain any mass μ_i .

(a) First assume that $\mathbb{P}[a(\eta_t) = 0] > 0$. Note that $\{a(x) = 0\} \subset \{x = 0\}$ in view of the assumptions made on the function a . Set $H(x, y) = \omega(y) + a(y)\omega(x)$ and remark that H is strictly increasing in $|y|$ for fixed x . Let

$$I = H(0, D_{\tau'}) = (H(0, \eta_-^0), H(0, \eta_-^0 - \tau')) \cup (H(0, \eta_+^0), H(0, \eta_+^0 + \tau'))$$

and denote by λ_I the restriction of the Lebesgue measure to the (nonempty) set I . For $h_0 = x \in \mathbb{R}^+$, we have

$$h_2 = H(\eta_0, \eta_1) + a(\eta_1)a(\eta_0)x.$$

We have, for any Borel set B ,

$$\begin{aligned} \mathbb{P}(h_2 \in B) &\geq \mathbb{P}(h_2 \in B | a(\eta_0) = 0, \eta_1 \in D_{\tau'}) \mathbb{P}(a(\eta_0) = 0) \mathbb{P}(\eta_1 \in D_{\tau'}) \\ &= \mathbb{P}(H(0, \eta_1) \in B | \eta_1 \in D_{\tau'}) \mathbb{P}(a(\eta_0) = 0) \mathbb{P}(\eta_1 \in D_{\tau'}). \end{aligned}$$

Note that, conditional on the event $\{\eta_1 \in D_{\tau'}\}$, the variable $H(0, \eta_1)$ admits a density, f_H , that is positive over I . Thus

$$\lambda_I(B) > 0 \Rightarrow \mathbb{P}(h_2 \in B) \geq \mathbb{P}(a(\eta_0) = 0) \mathbb{P}(\eta_1 \in D_{\tau'}) \int_{I \cap B} f_H d\lambda > 0,$$

which proves that the Markov chain (h_t) is λ_I -irreducible.

(b) Now suppose $\mathbb{P}[a(\eta_t) = 0] = 0$. For ease of presentation we shall assume that

$$p_i > 0, \quad \forall i = 1, \dots, N, \quad \text{and} \quad \prod_{i=1}^N a(\mu_i) > 0. \tag{16}$$

The case where the distribution of η is absolutely continuous with respect to λ , that is, $p = 0$ in (15), can be handled by a straightforward adaptation of what follows.

We have, in view of (15),

$$\begin{aligned}
 E \log \{a(\eta_t)\} &= \sum_{i=1}^N p_i \log a(\mu_i) + (1-p) \int_{(-\infty, \eta_-^0] \cup [\eta_+^0, +\infty)} \log a(x) f(x) d\lambda(x) \\
 &\geq \sum_{i=1}^N p_i \log a(\mu_i) + q \log a(\eta_-^0) + r \log a(\eta_+^0),
 \end{aligned}
 \tag{17}$$

where

$$q = (1-p) \int_{(-\infty, \eta_-^0]} f(x) d\lambda(x), \quad r = (1-p) \int_{[\eta_+^0, +\infty)} f(x) d\lambda(x).$$

By convention take $a(\eta_-^0) = 1$ if $\{\eta \mid \eta < 0, f(\eta) > 0\} = \emptyset$ and $a(\eta_+^0) = 1$ if $\{\eta \mid \eta > 0, f(\eta) > 0\} = \emptyset$. Inequality (17) follows from the fact that a is monotonous over the positive and negative real semilines. It follows that, under (2),

$$a(\eta_-^0)^q a(\eta_+^0)^r \prod_{i=1}^N a(\mu_i)^{p_i} < 1.$$

Hence there exist some integers n_0, m_0 , and n_i , for $i = 1, \dots, N$, such that

$$\rho =: a(\eta_-^0)^{n_0} a(\eta_+^0)^{m_0} \prod_{i=1}^N a(\mu_i)^{n_i} < 1.
 \tag{18}$$

By continuity of a , it is not restrictive to assume that τ' is small enough so that

$$\rho_1 =: a(\eta_-^0 - \tau')^{n_0} a(\eta_+^0 + \tau')^{m_0} \prod_{i=1}^N a(\mu_i)^{n_i} < 1.
 \tag{19}$$

Now let $h_0 = x \in \mathbb{R}^+$. We have, for all $t > 0$,

$$h_t = \omega(\eta_{t-1}) + \sum_{i=1}^{t-1} a(\eta_{t-1}) \dots a(\eta_{t-i}) \omega(\eta_{t-i-1}) + a(\eta_{t-1}) \dots a(\eta_0) x.
 \tag{20}$$

The inequalities (18) and (19) will allow us to control the products of this sum, provided that we constrain the i.i.d. process to visit some states with appropriate frequencies. To this aim we introduce, for $K = 1, 2, \dots$, the event

$$\begin{aligned}
 A_K &= \bigcap_{k=0}^{K-1} \bigcap_{s=0}^{n_0-1} \{\eta_{s+kn} \in (\eta_-^0 - \tau', \eta_-^0)\} \bigcap_{s=n_0}^{n_0+m_0-1} \{\eta_{s+kn} \in (\eta_+^0, \eta_+^0 + \tau')\} \\
 &\quad \times \bigcap_{s=n_0+m_0}^{n_0+m_0+n_1-1} \{\eta_{s+kn} = \mu_1\} \dots \bigcap_{s=n-n_N}^{n-1} \{\eta_{s+kn} = \mu_N\},
 \end{aligned}$$

where $n = n_0 + m_0 + \sum_{i=1}^N n_i$. When η_+^0 or η_-^0 is not defined, the corresponding terms can be withdrawn from the definition of A_K . Thus we have

$$\mathbb{P}(A_K) > 0. \tag{21}$$

Denote $y_k = (a(\eta_{s+kn}))_{0 \leq s \leq n_0+m_0-1} \in (\mathbb{R}^+)^{n_0+m_0}$, for $k = 0, \dots, K - 1$. The conditional distribution given A_K of the vector

$$Y_K = (y_0, \dots, y_{K-1})$$

has a density with respect to the Lebesgue measure on $(\mathbb{R}^+)^{K(n_0+m_0)}$.

Now we wish to show that

conditional on A_K and $(h_0 = x)$, the variable h_{K_n} has a density with respect to the Lebesgue measure on \mathbb{R}^+ . (22)

First suppose that the function ω is constant over \mathbb{R}^+ and \mathbb{R}^- . For any integer ℓ and any vector $u = (u_1, \dots, u_\ell)$ denote $u^{\leftarrow} = (u_1, \dots, u_{\ell-1})$ and $u^+ = u_\ell$. By convention let $\prod_{j=k}^n u_j = 1$ for $n < k$. Let $\underline{n} = \sum_{i=1}^N n_i$. We have, given A_K ,

$$\begin{aligned} h_{K_n} &= \sum_{j=0}^{\underline{n}} \omega(\eta_{K_n-j-1}) \prod_{\ell=1}^j a(\eta_{K_n-\ell}) \\ &\quad + \left\{ \prod_{\ell=1}^{\underline{n}} a(\eta_{K_n-\ell}) \right\} \\ &\quad \times \left\{ \sum_{j=\underline{n}+1}^{K_n-1} \omega(\eta_{K_n-j-1}) \prod_{\ell=\underline{n}+2}^j a(\eta_{K_n-\ell}) + \prod_{\ell=\underline{n}+2}^{K_n} a(\eta_{K_n-\ell})x \right\} a(\eta_{K_n-\underline{n}-1}) \\ &=: \varpi + \left\{ \prod_{i=1}^N a(\mu_i)^{n_i} \right\} S(Y_K^{\leftarrow}) Y_K^+, \end{aligned}$$

where $\varpi = \varpi(\mu_1, \dots, \mu_N)$ is a constant and $S(Y_K^{\leftarrow})$ is an a.s. positive random variable. In view of (16) it follows that given A_K the mapping

$$Y_K = (Y_K^{\leftarrow}, Y_K^+) \rightarrow Z_K =: (Y_K^{\leftarrow}, h_{K_n})$$

is a C^1 diffeomorphism between open sets of $(\mathbb{R}^+)^{K(n_0+m_0)}$. Indeed, the determinant of the Jacobian matrix of this mapping is given by $\{\prod_{i=1}^N a(\mu_i)^{n_i}\} S(Y_K^{\leftarrow})$ which is a.s. positive. Therefore the distribution of Z_K conditional on A_K has a density with respect to the Lebesgue measure on $(\mathbb{R}^+)^{K(n_0+m_0)}$. Consequently (22) holds.

Now suppose that ω is nonconstant over \mathbb{R}^+ . Let $\tilde{y}_k = (a(\eta_{s+kn}), \omega(\eta_{s+kn}))_{0 \leq s \leq n_0+m_0-1}$ and let $\tilde{Y}_K = (\tilde{y}_0, \dots, \tilde{y}_{K-1})$. We have

$$\begin{aligned} h_{Kn} &= \sum_{j=0}^{n-1} \omega(\eta_{Kn-j-1}) \prod_{\ell=1}^j a(\eta_{Kn-\ell}) + \omega(\eta_{Kn-n-1}) \prod_{\ell=1}^n a(\eta_{Kn-\ell}) \\ &\quad + \left\{ \prod_{\ell=1}^{n+1} a(\eta_{Kn-\ell}) \right\} \\ &\quad \times \left\{ \sum_{j=n+1}^{Kn-1} \omega(\eta_{Kn-j-1}) \prod_{\ell=n+2}^j a(\eta_{Kn-\ell}) + \prod_{\ell=n+2}^{Kn} a(\eta_{Kn-\ell})x \right\} \\ &=: \tilde{\omega} + \left\{ \prod_{i=1}^N a(\mu_i)^{n_i} \right\} \tilde{Y}_K^+ + T(\tilde{Y}_K^{\leftarrow}), \end{aligned}$$

where $\tilde{\omega}$ is a constant and $T(\tilde{Y}_K^{\leftarrow})$ is a random variable. The conclusion follows from the same argument as before, noting that the mapping

$$Y_K = (\tilde{Y}_K^{\leftarrow}, \tilde{Y}_K^+) \rightarrow \tilde{Z}_K = (\tilde{Y}_K^{\leftarrow}, h_{Kn})$$

is a C^1 diffeomorphism between open sets of $(\mathbb{R}^+)^{2K(n_0+m_0)}$. The case where ω is nonconstant over \mathbb{R}^- can be handled similarly.

To determine the support I_K of the conditional distribution of h_{Kn} , first note that, for $t = Kn$, the last term on the right-hand side of (20) writes, conditional on A_K ,

$$a(\eta_{Kn-1}) \dots a(\eta_0)x = \left\{ \prod_{i=1}^N a(\mu_i)^{n_i} \right\} \prod_{k=0}^{K-1} \prod_{s=0}^{n_0+m_0-1} a(\eta_{s+kn})x$$

and therefore belongs to the set $[\rho^Kx, \rho_1^Kx]$, in view of the assumptions made on the function a . Products of the form $a(\eta_{Kn-1}) \dots a(\eta_{Kn-kn})$, for $k = 1, \dots, K - 1$, can be handled similarly. To deal with the other products in (20) we introduce the notation

$$\begin{aligned} M(y, z) &= \sum_{j=1}^{n_0} a(y)^{n_0-j} a(z)^{m_0} \prod_{i=1}^N a(\mu_i)^{n_i} \omega(y) + \sum_{j=1}^{m_0} a(z)^{m_0-j} \prod_{i=1}^N a(\mu_i)^{n_i} \omega(z) \\ &\quad + \sum_{j=1}^{n_1} a(\mu_1)^{n_1-j} \prod_{i=2}^N a(\mu_i)^{n_i} \omega(\mu_1) + \dots + \sum_{j=1}^{n_N} a(\mu_N)^{n_N-j} \omega(\mu_N). \end{aligned}$$

If A_K holds true, using the assumptions (i) and (ii) on the functions a and ω , we have

$$h_{Kn} \in I_K =: \left[M(\eta_-^0, \eta_+^0) \sum_{i=0}^{K-1} \rho^i + \rho^Kx, \quad M(\eta_-^0 - \tau', \eta_+^0 + \tau') \sum_{i=0}^{K-1} \rho_1^i + \rho_1^Kx \right],$$

and thus, for K sufficiently large, in view of (18) and (19),

$$h_{Kn} \in I = \left[\frac{M(\eta_-^0, \eta_+^0)}{1 - \rho}, \frac{M(\eta_-^0 - \tau', \eta_+^0 + \tau')}{1 - \rho_1} \right].$$

Indeed, I is the closure of the limit of the sets I_K when K tends to infinity. Because the lower and upper bounds of I_K are reached, by the intermediate values theorem and in view of (22), we conclude that

conditional on A_K and $(h_0 = x)$, the density of the variable h_{Kn} is positive over I_K . (23)

Denoting by λ_I the restriction of the Lebesgue measure to the set I , it follows that λ_I is an irreducibility measure because, for any set $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} \lambda_I(B) > 0 &\Rightarrow \exists K, \quad \lambda(B \cap I_K) > 0 \\ &\Rightarrow \exists t = Kn, \quad \mathbb{P}[h_t \in B] \geq \mathbb{P}[h_t \in B | A_K] \mathbb{P}(A_K) > 0, \end{aligned}$$

by (21) and (23).

Step (iii): First note that the assumptions $E \log a(\eta_t) < 0$ and $E\{a(\eta_t)^r\} < \infty$ for $r > 0$ imply the existence of a number $s \in (0, 1)$ such that

$$\rho_2 =: E\{a(\eta_{t-1})^s\} < 1$$

(see Nelson, 1990; Berkes, Horváth, and Kokoszka, 2003, Lem. 2.3).

Let the test function defined by $V(x) = 1 + x^s$, let $0 < c < 1 - \rho_2$, and let the compact set

$$C = \{x \in \mathbb{R}^+; \quad \omega_s + c + (\rho_2 - 1 + c)x^s \geq 0\},$$

where $\omega_s = E\{\omega(\eta_{t-1})^s\}$ and s is chosen small enough so that $\omega_s < \infty$. We have, for $x \notin C$, using the elementary inequality $(a + b)^s \leq a^s + b^s$ for $a, b \geq 0$ and $s \in [0, 1]$,

$$\begin{aligned} E[V(h_t) | h_{t-1} = x] &\leq 1 + \omega_s + \rho_2 x^s \\ &< (1 - c)V(x), \end{aligned}$$

which proves (10). Moreover (9) holds true.

It remains to check that $\phi(C) > 0$ where $\phi = \lambda_I$ is the irreducibility measure obtained previously. Given the shape of the intervals I and C , it is clear that

$$\begin{aligned} \phi(C) > 0 &\Leftrightarrow \frac{M(\eta_-^0, \eta_+^0)}{1 - \rho} \in \dot{C} \quad (\text{interior of } C) \\ &\Leftrightarrow \omega_s + c + (\rho_2 - 1 + c) \left\{ \frac{M(\eta_-^0, \eta_+^0)}{1 - \rho} \right\}^s > 0. \end{aligned}$$

If c is chosen close enough to $\rho_2 - 1$ the latter inequality will be verified. For such a c , the compact set C meets the assumptions of Theorem 1. It follows that the Markov chain (h_t) is geometrically ergodic. ■

Proof of Theorem 3. We will show that the process (ϵ_t) inherits the mixing property established for (h_t) . We first show that the process $Y_t = (h_t, \eta_t)'$ has the mixing property. It is clear that (Y_t) is a Markov chain on $\mathbb{R}^+ \times \mathbb{R}$ endowed with its Borel σ -field. Moreover (Y_t) is strictly stationary as a measurable function of $\eta_t, \eta_{t-1}, \dots$. By independence between h_t and η_t we can denote by $\mathbb{P}_Y = \mathbb{P}_h \otimes \mathbb{P}_\eta$ the stationary distribution of Y_t , where \mathbb{P}_h is that of h_t and \mathbb{P}_η that of η_t . Denote by $\tilde{P}^t(y, \cdot)$ the transition probabilities of the Markov chain (Y_t) . We have, for $y = (y_1, y_2) \in \mathbb{R}^+ \times \mathbb{R}$, $B_1 \in \mathcal{B}(\mathbb{R}^+), B_2 \in \mathcal{B}(\mathbb{R})$, and $t > 0$,

$$\begin{aligned} \tilde{P}^t(y, B_1 \times B_2) &= \mathbb{P}(h_t \in B_1, \eta_t \in B_2 | h_0 = y_1, \eta_0 = y_2) \\ &= \mathbb{P}_\eta(B_2) \mathbb{P}(h_t \in B_1 | h_0 = y_1, \eta_0 = y_2) \\ &= \mathbb{P}_\eta(B_2) \mathbb{P}(h_t \in B_1 | h_1 = \omega(y_2) + a(y_2)y_1) \\ &= \mathbb{P}_\eta(B_2) P^{t-1}(\omega(y_2) + a(y_2)y_1, B_1). \end{aligned}$$

Therefore, because \mathbb{P}_η is a probability measure,

$$\|\tilde{P}^t(y, \cdot) - \mathbb{P}_Y(\cdot)\| = \|\tilde{P}^{t-1}(\omega(y_2) + a(y_2)y_1, \cdot) - \mathbb{P}_h(\cdot)\|.$$

The right-hand-side term converging to 0 at exponential rate, by the geometric ergodicity of (h_t) , we can deduce that (Y_t) is geometrically ergodic and thus geometrically β -mixing. Because $\epsilon_t = h^{-1}(h_t)\eta_t$ is a measurable function of Y_t , we can conclude in view of (8) that the process (ϵ_t) is geometrically β -mixing. ■

Remark 6. The theorem could be straightforwardly extended to the case where $N = \infty$, provided that $a(\mu_i) < 1$ for a finite number, say, N_0 , of indexes i . Indeed, in this case the inequality (17) continues to hold with N replaced by N_0 .

Remark 7. In Pham (1986), irreducibility is established for a class of models that is very similar to our model for (h_t) . However we cannot use Pham’s results because he assumes a continuous distribution for the i.i.d. process with a positive density in a neighborhood of 0, and more importantly, he requires $\omega(0) = 0$ (with our notations), which does not hold, in particular, for the standard GARCH model.

Examples

1. Consider the standard ARCH(1) model and assume that η_t has a mass at zero, with $E\eta_t^2 = 1$. Thus (2) is met because

$$\begin{aligned} E \log a(\eta_t) &= E \log \alpha \eta_t^2 = (-\infty) \times \mathbb{P}(\eta_t = 0) \\ &\quad + E(\log \alpha \eta_t^2 | \eta_t \neq 0) \times \mathbb{P}(\eta_t \neq 0) = -\infty. \end{aligned}$$

It follows that geometric ergodicity holds, under Assumption A, without any restriction on the parameter α .

2. Consider the threshold ARCH(1) model, introduced by Zakoian (1994), where $\omega(x) = \omega > 0$, $h(x) = x$, and $a(x) = \max(0, -x)\alpha_- + \max(0, x)\alpha_+$, with $\alpha_- > 0$, $\alpha_+ > 0$. Let $\mathbb{P}(\eta_t < 0) = p_-$ and $\mathbb{P}(\eta_t > 0) = p_+$. We have

$$E \log a(\eta_t) = \log(\alpha_-)p_- + \log(\alpha_+)p_+ + E \log|\eta_t|.$$

By Remark 3, under Assumption A, a necessary and sufficient condition for the existence of a stationary and geometrically ergodic solution is

$$\alpha_-^{p_-} \alpha_+^{p_+} < \exp\{-E \log|\eta_t|\}.$$

5. STATISTICAL APPLICATIONS

The time-series literature abounds in statistical results requiring mixing assumptions. Consequently, Theorem 3 has numerous direct applications. We only give two of them. The first application concerns the asymptotic distribution of sample autocorrelations and is directly inspired from Romano and Thombs (1996). The second application shows that the standard Dickey–Fuller unit-root tests remain asymptotically valid when the error term follows the GARCH(1,1) model that we consider in this paper.

5.1. Behavior of the Sample Autocorrelations

For a time series $\epsilon_1, \dots, \epsilon_n$ the identification stage of a model of the form (1) may involve the use of many statistics. Traditional estimators of the population autocorrelations of the squares are given by

$$\hat{\rho}(\ell) = \frac{\hat{\gamma}(\ell)}{\hat{\gamma}(0)}, \quad \text{where for } \ell \geq 0, \quad \hat{\gamma}(\ell) = \frac{1}{n} \sum_{t=1}^{n-\ell} \overline{\epsilon_t^2 \epsilon_{t+\ell}^2}$$

and $\overline{X_t}$ stands for the mean-corrected observations of a time series ($X_t, 1 \leq t \leq n$). Such statistics are often used to get an insight into the fourth-order structure of the process (ϵ_t). For the model of Ding, Granger, and Engle (1993), based on a Box–Cox power transformation of the conditional standard deviation process, the squares can be replaced by powers δ of the $|\epsilon_t|$. For nonparametric GARCH models, such as model (1), general transformations of the data lead to statistics of the form

$$\hat{\rho}_g(\ell) = \frac{\hat{\gamma}_g(\ell)}{\hat{\gamma}_g(0)}, \quad \text{where for } \ell \geq 0, \quad \hat{\gamma}_g(\ell) = \frac{1}{n} \sum_{t=1}^{n-\ell} \overline{g(\epsilon_t) g(\epsilon_{t+\ell})},$$

for some measurable function g . The asymptotic distributions of such statistics are easily deduced from Theorem 3.

COROLLARY 1. Assume that $E|g(\epsilon_t)|^{4+\nu} < \infty$ for some $\nu > 0$. Then, under the assumptions of Theorem 3, and for any fixed $h \geq 0$, the vectors

$$\sqrt{n} \left\{ \begin{pmatrix} \hat{\gamma}_g(0) \\ \vdots \\ \hat{\gamma}_g(h) \end{pmatrix} - \begin{pmatrix} \gamma_g(0) \\ \vdots \\ \gamma_g(h) \end{pmatrix} \right\} \quad \text{and} \quad \sqrt{n} \left\{ \begin{pmatrix} \hat{\rho}_g(1) \\ \vdots \\ \hat{\rho}_g(h) \end{pmatrix} - \begin{pmatrix} \rho_g(1) \\ \vdots \\ \rho_g(h) \end{pmatrix} \right\},$$

where $\gamma_g(\ell) = \text{Cov}\{g(\epsilon_t), g(\epsilon_{t-\ell})\}$ and $\rho_g(\ell) = \gamma_g(\ell)/\gamma_g(0)$, are asymptotically normally distributed.

Proof. The proof is similar to that of Theorems 3.1 and 3.2 in Romano and Thombs (1996). Theorem 3 shows that $(Y_t) =: (g(\epsilon_t) - Eg(\epsilon_t))$ is geometrically β - (and α -) mixing. Let $\hat{\gamma}_g^*(\ell) = n^{-1} \sum_{t=1}^{n-|\ell|} Y_t Y_{t+|\ell|}$. The Wold–Cramer device and the CLT for strongly mixing processes, given in Ibragimov (1962) and Herrndorf (1984), show that $n^{1/2}\{\hat{\gamma}_g^*(0) - \gamma_g(0), \dots, \hat{\gamma}_g^*(h) - \gamma_g(h)\}'$ is asymptotically normally distributed with covariance matrix Σ given by

$$\Sigma(i, j) = \lim_{n \rightarrow \infty} n \text{Cov}\{\hat{\gamma}_g^*(i), \hat{\gamma}_g^*(j)\} = \sum_{\ell=-\infty}^{\infty} \text{Cov}(Y_t Y_{t+i}, Y_{t+\ell} Y_{t+\ell+j}).$$

The absolute convergence of the last sum follows from standard covariance inequalities for mixing processes. To show the asymptotic normality of the vector involving the $\hat{\gamma}_g(i)$, $i = 0, \dots, h$, it remains to show that $\sqrt{n}\{\hat{\gamma}_g^*(i) - \hat{\gamma}_g(i)\} = o_p(1)$. This can be proved by the arguments given in the proof of Proposition 7.3.7 of Brockwell and Davis (1991). The vector of the sample autocorrelations, because it is a differentiable function of the sample autocovariances vector, is also asymptotically normally distributed. ■

Remark 8. For $g(x) = x$, the result can be deduced from the Lindeberg CLT for martingale differences. However, for autocovariances of general transformations of ϵ_t , it may be difficult, if not impossible, to rely on asymptotic theorems for martingales. In such cases, mixing results offer an alternative.

5.2. Unit-Root Tests for Autoregressive (AR) Models with GARCH Errors

Many financial series, such as (logarithms of) stock-market indices, are suspected to behave roughly like random walks with conditionally heteroskedastic increments. For such series, one could consider a model of the form

$$X_0 = 0, \quad X_t = \phi X_{t-1} + \epsilon_t, \quad t = 1, 2, \dots, \tag{24}$$

where (ϵ_t) belongs to the general class of GARCH(1,1) models (1). Given the consequences of the random walk hypothesis, especially in terms of persistence of the economic shocks, it is important to consider tests for the unit-root

hypothesis $H_0: \phi = 1$ against the stationarity assumption $H_1: \phi \in (-1, 1)$. Let $\hat{\phi}_n =: \sum_{t=2}^n X_t X_{t-1} (\sum_{t=2}^n X_{t-1}^2)^{-1}$ be the least squares estimator (LSE) of ϕ and let $\{W(t), t \in [0, 1]\}$ be a standard Brownian motion. The following corollary demonstrates the asymptotic validity of the standard Dickey–Fuller tests, in our heteroskedastic framework.

COROLLARY 2. *Suppose that (X_t) satisfies (24) where (ϵ_t) is a general GARCH(1,1) process satisfying the assumptions of Theorem 3. Then if $E|\epsilon_t|^{2+\nu} < \infty$ for some $\nu > 0$,*

$$n(\hat{\phi}_n - 1) \Rightarrow \frac{(\frac{1}{2})\{W^2(1) - 1\}}{\int_0^1 W^2(t) dt} \quad \text{under } H_0: \phi = 1, \tag{25}$$

and if $E|\epsilon_t|^{4+\nu} < \infty$ for some $\nu > 0$,

$$\sqrt{n}(\hat{\phi}_n - \phi) \Rightarrow \sqrt{(EX_t^2)^{-2} E\epsilon_t^2 X_{t-1}^2} W(1) \quad \text{under } H_1: \phi \in (-1, 1). \tag{26}$$

Proof. Note that $\lim_{n \rightarrow \infty} \text{Var}\{n^{-1/2} \sum_{t=1}^n \epsilon_t\} = E\epsilon_t^2 > 0$. In view of Theorem 3, the weak convergence (25) is deduced from Phillips (1987), and (26) can be deduced, for instance, from Francq and Zakoian (1998). ■

Remark 9. For the standard GARCH(1,1) errors, Ling, Li, and McAleer (2003) derived the asymptotic distribution in (25) under the second moment condition, namely, $\alpha + \beta < 1$.

Remark 10. Consider model (1) with $h_t = \sigma_t^2$. Assume that η_t has a symmetric distribution, that ω and a are even functions, and that the following moments exist, for $i = 1, 2$,

$$\begin{aligned} \omega_i &= \frac{E\{\omega(\eta_t)^i \eta_t^2\}}{E\{\omega(\eta_t)^i\}}, & a_i &= E\{a(\eta_t)^i\}, \\ c &= \frac{E\{\omega(\eta_t)a(\eta_t)\}}{E\{\omega(\eta_t)\}}, & \tilde{a} &= E\{a(\eta_t)\eta_t^2\}, \end{aligned}$$

with $a_2 < 1$. Then tedious computations show that the asymptotic variance of the LSE is, under H_1 ,

$$\begin{aligned} \frac{E\epsilon_t^2 X_{t-1}^2}{(EX_t^2)^2} &= \frac{(1 - \phi^2)^2 (1 - a_1)^2}{\omega_1^2} \sum_{i=0}^{\infty} \phi^{2i} E\epsilon_t^2 \epsilon_{t-i-1}^2 \\ &= \frac{(1 - a_1)(1 - \phi^2)\{\phi^2(1 - a_2) + (1 - \phi^2)\}[\omega_1(1 - a_2) + \tilde{a}\{\omega_2(1 - a_1) + 2c\}]}{(1 - a_2)(1 - a_1\phi^2)}. \end{aligned}$$

It is interesting to note that in the unit-root case the asymptotic distribution of the LSE is the same with i.i.d. or GARCH errors, whereas it depends on the

GARCH parameters in the stationary case. Thus the Dickey–Fuller test statistic can still be used when the errors satisfy a model of the form (1). A similar finding was obtained by Rahbek et al. (2002) in a multivariate framework. They showed that the trace test for the cointegration rank remains valid when the standard i.i.d. Gaussian errors are replaced by ARCH-type innovations, with appropriate moment conditions.

Remark 11. Corollaries 1 and 2 are just given for illustrative purposes and can be straightforwardly extended. In particular, Corollary 2 could include more general models with augmented variables and/or an intercept and/or a deterministic trend, as in Phillips and Perron (1988). Similar convergences could also be stated for t -statistics.

NOTES

1. Namely, the absolute value GARCH model (Taylor, 1986; Schwert, 1989) for $\omega(x) = \omega$, $h(x) = x$, and $a(x) = \alpha|x| + \beta$; the threshold GARCH model of Zakoïan (1994) for $\omega(x) = \omega$, $h(x) = x$, and $a(x) = \alpha_- \max(0, -x) + \alpha_+ \max(0, x) + \beta$; the Glosten, Jagannathan, and Runkle (1993) model for $\omega(x) = \omega$, $h(x) = x^2$, and $a(x) = \alpha_- \{\max(0, -x)\}^2 + \alpha_+ \{\max(0, x)\}^2 + \beta$; the asymmetric power GARCH model of Ding et al. (1993) for $\omega(x) = \omega$, $h(x) = x^\delta$, and $a(x) = \alpha(|x| - \gamma x)^\delta + \beta$; a moving average GARCH process, inspired by the moving average conditionally heteroskedastic (MACH) model of Yang and Bewley (1995), for $h(x) = x^2$, $a(x) = a$, and, for instance, $\omega(x) = \omega_1 + \omega_2 x^2$; the sign-switching autoregressive conditional heteroskedasticity (ARCH) model of Fornari and Mele (1997) for $h(x) = x^2$, $\omega(x) = \omega_1 + \omega_2 \text{sign}(x)$, and $a(x) = \alpha x^2 + \beta$. See also Bühlmann and McNeil (2000) and Yang and Tschernig (2006) for recent references on nonparametric GARCH(1,1) modeling.

2. See, e.g., Lee and Hansen (1994), Lumsdaine (1996), Berkes, Horváth, and Kokoszka (2003), and Franco and Zakoïan (2004). See also Jensen and Rahbek (2004) for an extension to the non-stationary case.

REFERENCES

- Ango Nze, P. (1992) Critères d'ergodicité de quelques modèles à représentation markovienne. *Comptes Rendus de l'Académie des Sciences de Paris* 315, 1301–1304.
- Ango Nze, P. (1998) Critères d'ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation markovienne. *Comptes Rendus de l'Académie des Sciences de Paris* 326, 371–376.
- Basrak, B., R.A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.
- Berkes, I., L. Horváth, & P.S. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–227.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bougerol, P. & N. Picard (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.
- Boussama, F. (1998) Ergodicité, mélange et estimation dans les modèles GARCH. Ph.D. dissertation, Paris-7 University.
- Bradley, R.C. (1986) Basic properties of strong mixing conditions. In E. Eberlein & M.S. Taqqu (eds.), *Dependence in Probability and Statistics, A Survey of Recent Results*, pp. 165–192. Birkhäuser.
- Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*. Springer-Verlag.

- Bühlmann, P. & A.J. McNeil (2000) Nonparametric GARCH Models. Manuscript, ETH, Zürich.
- Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.
- Chow, Y.S. & H. Teicher (1997) *Probability Theory*, 3rd ed. Springer-Verlag.
- Davydov, Y. (1973) Mixing conditions for Markov chains. *Theory of Probability and Its Applications* 18, 313–328.
- Ding, Z., C. Granger, & R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Duan, J.-C. (1997) Augmented GARCH(p, q) process and its diffusion limit. *Journal of Econometrics* 79, 97–127.
- Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.
- Feigin, P.D. & R.L. Tweedie (1985) Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* 6, 1–14.
- Fornari, F. & A. Mele (1997) Sign- and volatility-switching ARCH models: Theory and applications to international stock markets. *Journal of Applied Econometrics* 12, 49–65.
- Franco, C. & J.-M. Zakoïan (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.
- Franco, C. & J.-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Glosten, L.R., R. Jagannathan, & D. Runkle (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.
- He, C. & T. Teräsvirta (1999) Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.
- Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12, 141–153.
- Hwang, S.Y. & T.Y. Kim (2004) Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure. *Stochastic Processes and Their Applications* 110, 295–314.
- Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349–382.
- Jensen, S.T. & A. Rahbek (2004) Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* 72, 641–646.
- Klüppelberg, C., A. Lindner, & R. Maller (2004) A continuous time GARCH process driven by a Lévy process: Stationarity and second order behaviour. *Journal of Applied Probability* 41, 601–622.
- Kristensen, D. (2006) Geometric ergodicity of a class of Markov chains with applications to time series models. Paper presented at the Econometric Society 2005 World Congress.
- Lee, O. & D.W. Shin (2004) Strict stationarity and mixing properties of asymmetric power GARCH models allowing a signed volatility. *Economics Letters* 84, 167–173.
- Lee, S.-W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.
- Ling, S., W.K. Li, & M. McAleer (2003) Estimation and testing for unit root processes with GARCH(1,1) errors: Theory and Monte Carlo evidence. *Econometric Reviews* 22, 179–202.
- Ling, S. & M. McAleer (2002) Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* 106, 109–117.
- Lu, Z. (1996) A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model. *Statistics and Probability Letters* 30, 305–311.
- Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

- Meitz, M. & P. Saikkonen (2004) Ergodicity, Mixing, and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models. Paper presented at the Econometric Society 2005 World Congress.
- Meyn, S.P. & R.L. Tweedie (1996) *Markov Chains and Stochastic Stability*, 3rd ed., Springer-Verlag.
- Mokkadem, A. (1990) Propriétés de mélange des processus autorégressifs polynomiaux. *Annales de l'Institut Henri Poincaré* 26, 219–260.
- Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.
- Pham, D.T. (1986) The mixing property of bilinear and generalised random coefficients autoregressive models. *Stochastic Processes and Their Applications* 23, 291–300.
- Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.
- Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
- Rahbek, A., E. Hansen, & J. Dennis (2002) ARCH Innovations and Their Impact on Cointegration Rank Testing. Working paper, University of Copenhagen.
- Romano, J.L. & L.A. Thombs (1996) Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* 91, 590–600.
- Schwert, G.W. (1989) Why does stock market volatility change over time? *Journal of Finance* 45, 1129–1155.
- Taylor, S. (1986) *Modelling Financial Time Series*. Wiley.
- Tjøstheim, D. (1990) Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
- Yang, L. & R. Tschernig (2006) Nonparametric Estimation of Generalized Impulse Response Functions. Unpublished document, Michigan State University.
- Yang, M. & R. Bewley (1995) Moving average conditional heteroskedastic processes. *Economics Letters* 49, 367–372.
- Zakoïan, J.-M. (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.