



DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

STABILITY OF NONLINEAR AR-GARCH MODELS

Mika Meitz and Pentti Saikkonen

Number 328

May 2007

Manor Road Building, Oxford OX1 3UQ

Stability of nonlinear AR–GARCH models*

Mika Meitz
University of Oxford

Pentti Saikkonen
University of Helsinki

First version June 2006, revised April 2007

Abstract

This paper studies the stability of nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a nonlinear autoregression of order p (AR(p)) with the conditional variance specified as a nonlinear first order generalized autoregressive conditional heteroskedasticity (GARCH(1,1)) model. Conditions under which the model is stable in the sense that its Markov chain representation is geometrically ergodic are provided. This implies the existence of an initial distribution such that the process is strictly stationary and β -mixing. Conditions under which the stationary distribution has finite moments are also given. The results cover several nonlinear specifications recently proposed for both the conditional mean and conditional variance, and only require mild moment conditions.

*The first author's work was financially supported by the Jan Wallander's and Tom Hedelius' Foundation, Grant No. J03–41. The second author acknowledges financial support by the Yrjö Jahnsson Foundation and the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki. We thank the editor and an anonymous referee for helpful comments and suggestions. Material from this paper has been presented at the 61st European Meeting of the Econometric Society, Vienna, August 2006, Workshop 'Volatility Day', Stockholm, November 2006, Zeuthen Workshop, Copenhagen, December 2006, and in seminars at LSE, Université catholique de Louvain, Université Libre de Bruxelles, and University of Oxford. We thank the participants in these occasions for their comments. Address correspondence to: Mika Meitz, Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, United Kingdom; e-mail: mika.meitz@economics.ox.ac.uk; or to: Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FIN-00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

This paper is concerned with the stability of nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a nonlinear autoregression of order p ($\text{AR}(p)$) with the conditional variance specified as a nonlinear first order generalized autoregressive conditional heteroskedasticity ($\text{GARCH}(1,1)$) model. This time series model can be viewed as a Markov chain, and our study makes heavy use of the stability theory developed for Markov chains. We refer the reader to Meyn and Tweedie (1993) for a comprehensive account of the needed Markov chain theory.

The stability concept employed in the paper is that of geometric ergodicity, or more precisely, Q -geometric ergodicity as defined by Liebscher (2005). Geometric ergodicity is a useful property, for it implies the existence of an initial distribution which makes the Markov chain strictly stationary and β -mixing (or absolutely regular). The Q -geometric ergodicity is even more useful in that it implies that certain moments of the stationary distribution exist and, moreover, the β -mixing property also holds for a variety of nonstationary initial distributions. In this paper we give conditions under which the Markov chain associated with our AR-GARCH model is Q -geometrically ergodic and has moments of known order. An important consequence of these results is that usual limit theorems can be applied and, therefore, it becomes possible to develop a rigorous asymptotic estimation theory for these models.

Results similar to ours have previously been obtained for nonlinear homoskedastic autoregressions in Bhattacharya and Lee (1995), An and Huang (1996), An and Chen (1997), and Lee (1998) among many others. These results have been extended to allow for ARCH, but not GARCH, type conditional heteroskedasticity by Masry and Tjøstheim (1995), Cline and Pu (1998), Lu (1998), Cline and Pu (1999), Chen and Chen (2001), Hwang and Woo (2001), Lu and Jiang (2001), Cline and Pu (2004), Liebscher (2005), and Saikkonen (2007). For related results for pure GARCH models, see Carrasco and Chen (2002), Meitz and Saikkonen (2007), and the references therein. Stability of nonlinear autoregressions with GARCH type conditional heteroskedasticity has previously been studied by Liu, Li, and Li (1997), Ling (1999), and, quite recently, Cline (2006). Of these papers Liu, Li, and Li (1997) and Ling (1999) are confined to threshold AR-

GARCH models whereas Cline (2006) studies a very general nonlinear autoregression with GARCH type conditional heteroskedasticity. Cline (2006) obtains sharp results for geometric ergodicity but a difficulty with the application of these results is that the employed assumptions are quite general and appear difficult to verify. A threshold AR–GARCH model is the only example that is explicitly treated in his paper. Note also that in this example the conditional heteroskedasticity is driven by the observed series instead of the autoregressive errors, as assumed in this paper.

A major difficulty in establishing geometric ergodicity in the present context is to prove irreducibility of the relevant Markov chain, which is typically required as a first step in the proof of geometric ergodicity. This difficulty also appears in Cline (2006) and it presumably explains the small number of related previous results. It is also a major reason why we focus on first order GARCH models. Our approach is to apply results on nonlinear state space models given in Meyn and Tweedie (1993, Chapter 7). This approach requires rather stringent smoothness assumptions about the nonlinear functions used to specify the conditional mean and conditional variance and, consequently, we are not able to handle threshold type nonlinearities. However, we are still able to cover a number of nonlinearities recently considered in both theoretical and applied literature. Our results are also easy to apply and they only require mild moment conditions.

A convenient feature of the assumptions needed to obtain our results is that most of them restrict the conditional mean and conditional variance of the model separately. Only one of our assumptions is common to both the conditional mean and conditional variance and in several cases of interest this assumption can be straightforwardly checked. In such cases the verification of our assumptions reduces to separately checking the assumptions of a homoskedastic nonlinear autoregressive model and a pure GARCH model. As far as the conditional mean is concerned, our results apply to smooth variants of the functional-coefficient autoregressive model of Chen and Tsay (1993) which encompasses various well-known nonlinear autoregressive models such as the smooth transition autoregressive models (see Teräsvirta (1994), van Dijk, Teräsvirta, and Franses (2002), and the references therein). The conditional variance may be specified as the linear GARCH model of Bollerslev (1986) or even a GARCH model with a rather complicated smooth nonlinear

structure.

The rest of this paper is organized as follows. The model and the assumptions needed are introduced in Section 2. In Section 3 the main result of the paper is presented, and examples are provided in Section 4. Section 5 concludes. Proofs of all the results are given in an Appendix.

2 Model and Assumptions

Let y_t , $t = 1, 2, \dots$, be a real valued stochastic process generated by

$$y_t = f(y_{t-1}, \dots, y_{t-p}) + h_t^{1/2} \varepsilon_t, \quad (1)$$

where h_t is a positive function of y_s , $s < t$, and ε_t is a sequence of (continuous) i.i.d. $(0, 1)$ random variables such that ε_t is independent of $\{y_s, s < t\}$. The function f is supposed to be nonlinear so that equation (1) defines a nonlinear autoregression with conditionally heteroskedastic errors. We assume that h_t , the conditional variance of y_t , is generated by a (possibly) nonlinear GARCH(1,1) process driven by regression errors. Specifically,

$$h_t = g(u_{t-1}, h_{t-1}), \quad (2)$$

where g is a function to be described shortly and

$$u_t = y_t - f(y_{t-1}, \dots, y_{t-p}). \quad (3)$$

From the definition of u_t it is readily seen that $Z_t = [y_t \ \dots \ y_{t-p} \ h_t]' \stackrel{def}{=} [Y_t' \ h_t]'$ is a Markov chain on $\mathcal{Z} = \mathbb{R}^{p+1} \times \mathbb{R}_+$ (here and in what follows the notation $\mathbb{R}_+ = (0, \infty)$ is used). To make the Markov chain representation of Z_t explicit, set

$$h(Z_{t-1}) = g(y_{t-1} - f(y_{t-2}, \dots, y_{t-1-p}), h_{t-1}) \quad (4)$$

and observe that then we can write

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p} \\ h_t \end{bmatrix} = \begin{bmatrix} f(y_{t-1}, \dots, y_{t-p}) \\ y_{t-1} \\ \vdots \\ y_{t-p} \\ h_t(Z_{t-1}) \end{bmatrix} + \begin{bmatrix} h(Z_{t-1})^{1/2} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (5)$$

or, more briefly,

$$Z_t = F(Z_{t-1}, \varepsilon_t), \quad (6)$$

where the function $F : \mathcal{Z} \times \mathbb{R} \rightarrow \mathcal{Z}$ is defined in an obvious way.

We set $F_1 = F$ and, for $k \geq 1$, $F_{k+1}(z, e_1, \dots, e_{k+1}) = F(F_k(z, e_1, \dots, e_k), e_{k+1})$, where $z \in \mathbb{R}^{p+1}$ and $e_i \in \mathbb{R}$. Then, for any initial condition $Z_0 = z_0$ and any $k \geq 1$, $Z_k = F_k(z_0, \varepsilon_1, \dots, \varepsilon_k)$. Following Meyn and Tweedie (1993) we call $\{e_i\}$ a control sequence and $z_k = F_k(z_0, e_1, \dots, e_k)$ ($k = 1, 2, \dots$) the associated deterministic control model for the nonlinear state space model (6). Our analysis of the Markov chain Z_t makes use of this deterministic control model.

We make the following assumptions about the error term ε_t and the function f . We call a function smooth if its (partial) derivatives exist up to any order and are continuous.

Assumption 1 *The i.i.d. $(0, 1)$ random variables ε_t have a (Lebesgue) density which is positive and lower semicontinuous on \mathbb{R} . Furthermore, for some real $r > 0$, $E[|\varepsilon_t|^{2r}] < \infty$.*

Assumption 2 *The function f is of the form $f(x) = a(x)'x + b(x)$, $x \in \mathbb{R}^p$, where the functions $a : \mathbb{R}^p \rightarrow \mathbb{R}^p$ and $b : \mathbb{R}^p \rightarrow \mathbb{R}$ are bounded and smooth.*

Assumption 1 is mild and met in most applications where no bounds for the values of the considered process are assumed. Assumption 2 imposes a certain structure on the nonlinear function f which specifies the conditional expectation of the process. As mentioned in the Introduction, similar structures have previously appeared in the functional-coefficient autoregressive model of Chen and Tsay (1993) and its special cases such as smooth transition autoregressive models. For these models the required smoothness assumption is also satisfied. This, as well as Assumption 1, is needed to make use of the results for nonlinear state space models in Meyn and Tweedie (1993, Chapter 7).

For any integer $p \geq 2$ and any $x \in \mathbb{R}^p$ we define the $p \times p$ matrix

$$\bar{A}_p(x) = \begin{bmatrix} x_1 & x_2 & \cdots & x_{p-1} & x_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Then, using the function $a(x)$ in Assumption 2, set $a(x) = [a_1(x) \cdots a_p(x)]'$ and define the $(p+1) \times (p+1)$ matrix $A(x) = \bar{A}_{p+1} \begin{pmatrix} [a(x)' \ 0]' \end{pmatrix}$. Now the model takes the form

$$\begin{aligned} Y_t &= A(S'Y_{t-1})Y_{t-1} + \iota b(S'Y_{t-1}) + \iota h(Z_{t-1})^{1/2} \varepsilon_t \\ h_t &= h(Z_{t-1}), \end{aligned}$$

where $\iota = [1 \ 0 \ \cdots \ 0]'$ $((p+1) \times 1)$ and $S = [I_p \ 0]'$ $((p+1) \times p)$. To be able to establish geometric ergodicity we need to restrict the matrix $A(x)$. A general way to do this is provided by the following assumption where $\mathcal{A}_* = \{A(x) : x \in \mathbb{R}^p\}$.

Assumption 3 *There exists a matrix norm $\|\cdot\|^*$ induced by a vector norm, also denoted by $\|\cdot\|^*$, such that $\|A\|^* \leq \rho$ for all $A \in \mathcal{A}_*$ and some $0 < \rho < 1$.*

To make Assumption 3 operational in practice, two concrete cases are considered. For the first one we need the concept of joint spectral radius defined for a set of bounded square matrices \mathcal{A} by

$$\rho(\mathcal{A}) = \limsup_{k \rightarrow \infty} \left(\sup_{A \in \mathcal{A}^k} \|A\| \right)^{1/k},$$

where $\mathcal{A}^k = \{A_1 A_2 \cdots A_k : A_i \in \mathcal{A}, i = 1, \dots, k\}$ and $\|\cdot\|$ can be any matrix norm (the value of $\rho(\mathcal{A})$ does not depend on the choice of this norm). If the set \mathcal{A} only contains a single matrix A then the joint spectral radius of \mathcal{A} coincides with $\rho(A)$, the spectral radius of A . Several useful results about the joint spectral radius are given in the recent paper by Liebscher (2005) where further references can also be found.

Sufficient conditions for Assumption 3 can now be given.

Lemma 1 *Either of the following conditions is sufficient for Assumption 3 to hold.*

(i) $\rho(\mathcal{A}_*) < 1$ or, equivalently, $\rho(\mathcal{A}_1) < 1$, where $\mathcal{A}_1 = \{A_1(x) : x \in \mathbb{R}^p\}$ with the $p \times p$ matrix $A_1(x)$ defined by deleting the last row and last column of $A(x)$.

(ii) $\sum_{j=1}^p \alpha_j < 1$ or, equivalently, the roots of the characteristic polynomial $\lambda^p - \alpha_1 \lambda^{p-1} - \cdots - \alpha_p = 0$ are inside the unit circle, where $\alpha_j = \sup_{x \in \mathbb{R}^p} |a_j(x)|$ ($j = 1, \dots, p$).

An assumption similar to that in Lemma 1(i) was used by Liebscher (2005) who established geometric ergodicity for various nonlinear autoregressive models. In these models conditional heteroskedasticity was also allowed but GARCH type or even ARCH type

conditional heteroskedasticity was ruled out. A practical difficulty with the application of Lemma 1(i) is that the computation of the joint spectral radius is very computer-intensive unless the dimension of the matrix $A(x)$ is reasonably small (for a discussion, see Liebscher (2005)). In practice one should therefore consider $\rho(\mathcal{A}_1)$ rather than $\rho(\mathcal{A}_*)$. This computational difficulty has also been a motivation for the second part of Lemma 1 which gives the condition used by Chen and Tsay (1993) to provide a sufficient condition for geometric ergodicity in their functional-coefficient autoregressive model. The main advantage of this latter condition is its simplicity, for Liebscher (2005, Section 7) shows by an example that the condition based on the joint spectral radius can provide a larger region in the parameter space ensuring geometric ergodicity than the condition given in Lemma 1(ii).

The following assumption contains conditions which restrict the dynamics of the conditional variance process.

Assumption 4

- (a) *The function $g : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is smooth and, for some $\underline{g} > 0$, $\inf_{(u,x) \in \mathbb{R} \times \mathbb{R}_+} g(u,x) = \underline{g}$.*
- (b) *For all $x \in \mathbb{R}_+$, $g(u,x) \rightarrow \infty$ as $u \rightarrow \infty$.*
- (c) *There exists an $h^* \in \mathbb{R}_+$ such that the sequence h_k ($k = 1, 2, \dots$) defined by $h_k = g(0, h_{k-1})$, $k = 1, 2, \dots$, converges to h^* as $k \rightarrow \infty$ for all $h_0 \in \mathbb{R}_+$. If $g(u,x) \geq h^*$ for all $u \in \mathbb{R}$ and all $x \geq h^*$ it suffices that this convergence holds for all $h_0 \geq h^*$.*
- (d) *There exist $a, c \in [0, \infty)$ and a Borel measurable function $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ such that $g(x^{1/2}\varepsilon_t, x) \leq (a + \varphi(\varepsilon_t))x + c$ for all $x \in \mathbb{R}_+$. Furthermore, $a + \varphi(0) < 1$.*

The smoothness condition in Assumption 4(a) is needed to make use of the results for nonlinear state space models in Meyn and Tweedie (1993, Chapter 7). The same is true for Assumption 4(c) which is a high level assumption. Sufficient conditions for this assumption are discussed below. The latter condition in Assumption 4(a) implies that the conditional variance h_t is bounded away from zero, a property shared by most GARCH models. Assumption (b) is technical and needed in the proofs. It is also satisfied by most

commonly used first order GARCH models. Assumption 4(d) supplements Assumption 3 in that it is needed to prove the geometric ergodicity of the Markov chain Z_t . Assumptions closely related to Assumption 4(d) have also been used by Lanne and Saikkonen (2005) and Meitz and Saikkonen (2007). Further restrictions on the real number a and the function $\varphi(\cdot)$ will be imposed below in Assumption 5.

Assumption 4(c) is used to establish irreducibility of the Markov chain Z_t which, as already noted, is the challenge in proving geometric ergodicity of nonlinear AR–GARCH models. In this assumption the existence of a fixed point h^* of the function $g(0, x)$ is assumed. A well-known sufficient condition which implies that a unique fixed point exists and can be found by the stated recursion is the Lipschitz condition

$$|g(0, x_1) - g(0, x_2)| \leq \kappa |x_1 - x_2| \text{ for some } 0 \leq \kappa < 1 \text{ and all } x_1, x_2 \in \mathbb{R}_+ \quad (7)$$

(this follows from the contraction map principle, see e.g. Simmons (1963, Appendix 1)). This condition applies to the standard (linear) GARCH(1,1) model and, more generally, when Assumption 4(d) holds with $g(0, x) = (a + \varphi(0))x + c$. However, when the function $g(0, x)$ is nonlinear the Lipschitz condition (7) may not hold or it can be difficult to verify. Then the second condition of Assumption 4(c) may be useful. Combined with the other parts of Assumption 4, this condition implies the convergence of the stated recursion to a maximal fixed point h^* .¹ An instance when this second condition is convenient is when $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$ and $g(0, x)$ is nondecreasing for $x \geq h^*$. Then, $h^* = g(0, h^*) \leq g(0, x) \leq g(u, x)$ for all $u \in \mathbb{R}$ and $x \geq h^*$, implying the second condition of Assumption 4(c) and, hence, the convergence of the required recursion.

Our next assumption further restricts the dynamics of the conditional variance process, and is critical in proving the geometric ergodicity of the Markov chain Z_t .

¹To see this, first note that by Assumption 4(d), $g(0, x) \leq (a + \varphi(0) + \epsilon)x$ for all x large enough and some $\epsilon > 0$ such that $a + \varphi(0) + \epsilon < 1$. From this and Assumption 4(a) it follows that the function $g(0, x)$ has a maximal fixed point h^* such that $g(0, h^*) = h^*$ and $g(0, x) < x$ for all $x > h^*$. This, together with the latter condition of Assumption 4(c), implies that for any initial value $h_0 > h^*$ the sequence h_k , $k \geq 0$, is nonincreasing and bounded from below by h^* . Therefore it converges to, say, h_* ($\geq h^*$) and, because $g(0, h_k) = h_{k+1}$, we also have $g(0, h_k) \rightarrow h_*$. On the other hand, by the continuity of $g(0, \cdot)$, $g(0, h_k) \rightarrow g(0, h_*)$. Thus we must have $g(0, h_*) = h_*$ and, since h^* is the maximal fixed point, $h_* = h^*$.

Assumption 5 *Let the real number $r > 0$ be as in Assumption 1 and a and $\varphi(\cdot)$ as in Assumption 4(d). Assume that either*

$$(a) \ E[(a + \varphi(\varepsilon_t))^r] < 1, \text{ or}$$

$$(b) \ E[\varphi(\varepsilon_t)^r] < \infty \text{ and } E[\log(a + \varphi(\varepsilon_t))] < 0.$$

The moment condition of Assumption 5(a) is convenient in the proofs and it also enables us to obtain explicit results about existence of moments. However, if one is only interested in proving geometric ergodicity condition (b) can also be employed. Under this condition there exists an $r_0 \in (0, r)$ such that $E[(a + \varphi(\varepsilon_t))^{r_0}] < 1$ (see Basrak, Davis, and Mikosch (2002, Remark 2.9) and Meitz and Saikkonen (2007, Lemma 3)). Note, however, that this result only guarantees the existence of such an r_0 , but the precise value of r_0 cannot be determined.

Our final assumption concerns the deterministic control model $z_k = F_k(z_0, e_1, \dots, e_k)$ ($k = 1, 2, \dots$) associated with the nonlinear state space model (6) and the concept of forward accessibility (for a definition, see Meyn and Tweedie (1993, p. 151)).

Assumption 6 *For each initial value $z_0 \in \mathcal{Z}$, there exists a control sequence $e_1^{(0)}, \dots, e_{p+2}^{(0)}$ such that the $(p+2) \times (p+2)$ matrix*

$$\nabla F_{p+2}^{(0)} = \left[\frac{\partial}{\partial e_1} F_{p+2} \left(z_0, e_1^{(0)}, \dots, e_{p+2}^{(0)} \right) : \dots : \frac{\partial}{\partial e_{p+2}} F_{p+2} \left(z_0, e_1^{(0)}, \dots, e_{p+2}^{(0)} \right) \right]$$

is nonsingular.

By Proposition 7.1.4 of Meyn and Tweedie (1993), this assumption implies that the deterministic control model $z_k = F_k(z_0, e_1, \dots, e_k)$ is forward accessible. This property is needed to apply the results obtained in Chapter 7 of Meyn and Tweedie (1993). Note that although Assumption 6 is sufficient for forward accessibility it is not necessary, as the aforementioned proposition of Meyn and Tweedie (1993) shows.

Combined with Assumption 4(c), Assumption 6 is the key factor in our proof of the irreducibility of the Markov chain Z_t . Although Assumption 6 may sometimes be difficult to verify this is not the case for several commonly used models, as the examples of Section 4 demonstrate. To get an idea of the structure of the derivative matrix $\nabla F_{p+2}^{(0)}$, denote

the components of the vector $F_{p+2}(z_0, e_1, \dots, e_{p+2})$ briefly by y_{p+2}, \dots, y_2 and h_{p+2} (cf. equations (5) and the subsequent discussion). Then it is straightforward to check that

$$\nabla F_{p+2} = \begin{bmatrix} \partial y_{p+2}/\partial e_1 & \partial y_{p+2}/\partial e_2 & \partial y_{p+2}/\partial e_3 & \cdots & \partial y_{p+2}/\partial e_{p+1} & h_{p+2}^{1/2} \\ \partial y_{p+1}/\partial e_1 & \partial y_{p+1}/\partial e_2 & \partial y_{p+1}/\partial e_3 & \cdots & h_{p+1}^{1/2} & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ \partial y_2/\partial e_1 & h_2^{1/2} & 0 & \cdots & 0 & 0 \\ \partial h_{p+2}/\partial e_1 & \partial h_{p+2}/\partial e_2 & \partial h_{p+2}/\partial e_3 & \cdots & \partial h_{p+2}/\partial e_{p+1} & 0 \end{bmatrix},$$

where $h_i^{1/2} = h(z_{i-1})^{1/2} > 0$ ($i = 2, \dots, p+2$) and the superscript has been suppressed from $\nabla F_{p+2}^{(0)}$ to indicate that the derivatives are evaluated at an arbitrary control sequence. Thus, for Assumption 6 to hold it suffices to find $e_1^{(0)}, \dots, e_{p+2}^{(0)}$ such that, for all initial values, $\partial h_{p+2}/\partial e_1$ is nonzero and $\partial h_{p+2}/\partial e_2, \dots, \partial h_{p+2}/\partial e_{p+1}$ are zero when evaluated at $[e_1^{(0)} \cdots e_{p+2}^{(0)}]'$. As will be seen in Section 4, this holds for the standard linear GARCH model and even for some nonlinear GARCH models without any further assumptions. However, for some models, including pure ARCH models, the situation is more difficult.

We close this section by noting that a convenient feature of the assumptions imposed on the conditional mean and conditional variance is that, except for Assumption 6, they are separate. Specifically, as Lemma 1 shows, Assumptions 2 and 3 restrict only the conditional mean in (1) and this is done in the same way as in previous models without conditional heteroskedasticity. On the other hand, Assumptions 4 and 5 only concern the GARCH model specified for the error term in (1) and restrict it by conditions which are very similar to previous counterparts used in pure GARCH(1,1) models. As for Assumption 6, it concerns both the conditional mean and conditional variance but, as the examples of Section 4 demonstrate, this assumption can often be checked by only considering the model specified for conditional heteroskedasticity.

3 Geometric Ergodicity

Under the assumptions stated in the previous section we are able to show that the Markov chain Z_t is geometrically ergodic. We use the Q -geometric ergodicity of a Markov chain

as defined in Liebscher (2005) (see also Chan (1989) and Meyn and Tweedie (1993, p. 356) for similar previous concepts). For convenience, we repeat the definition here in a slightly different, though equivalent, form. We use $P^n(z, A) = \Pr(Z_n \in A \mid Z_0 = z)$, $z \in \mathcal{Z}$, $A \in \mathcal{B}(\mathcal{Z})$, to signify the n -step transition probability measure of the Markov chain Z_t defined on $\mathcal{B}(\mathcal{Z})$, the Borel sets of \mathcal{Z} . (When $n = 1$ the notation $P(z, A)$ is used.)

Definition 1 *The Markov chain Z_t on \mathcal{Z} is Q -geometrically ergodic if there exists a function $Q : \mathcal{Z} \rightarrow [0, \infty]$, a probability measure π on $\mathcal{B}(\mathcal{Z})$, and constants $a > 0$, $b > 0$, and $0 < \varrho < 1$ such that $\int_{\mathcal{Z}} \pi(dw)Q(w) < \infty$ and*

$$\sup_{v: |v| \leq 1} \left| \int_{\mathcal{Z}} P^n(z, dw) v(w) - \int_{\mathcal{Z}} \pi(dw) v(w) \right| \leq (a + bQ(z)) \varrho^n \quad \text{for all } z \in \mathcal{Z} \text{ and all } n \geq 1. \quad (8)$$

Observing that the left hand side of (8) equals the total variation norm of the signed measure $P^n(z, \cdot) - \pi(\cdot)$ shows that our definition of Q -geometric ergodicity is equivalent to that in Liebscher (2005). Thus, geometric ergodicity entails that the n -step transition probability measure $P^n(z, \cdot)$ converges at a geometric rate to the probability measure $\pi(\cdot)$ with respect to the total variation norm for all $z \in \mathcal{Z}$. The probability measure π is often referred to as the stationary probability measure of Z_t . The reason is that geometric ergodicity implies stationarity of Z_t if the initial value Z_0 is distributed according to the probability measure π (see Meyn and Tweedie (1993, p. 230–231)). Another useful consequence of Q -geometric ergodicity is that it implies that the Markov chain Z_t is β -mixing for any initial value Z_0 with a distribution such that the expectation of $Q(Z_0)$ is finite (see Liebscher (2005)). Also, once Q -geometric ergodicity has been established the finiteness of the expectation $\int_{\mathcal{Z}} \pi(dw)Q(w)$ is automatically obtained. This fact can be used to show that the stationary distribution of Z_t has finite moments of some order.

Note that one should be careful with the term Q -geometric ergodicity because, except for the prefix Q , another similar concept is in use. This concept is defined by assuming $Q \geq 1$ and replacing the inequality $|v| \leq 1$ and the bound $a + bQ(z)$ in (8) by $|v| \leq Q$ and M_z , respectively (see Meyn and Tweedie (1993, p. 356)). This clearly results in a stronger convergence than assumed in (8). This stronger convergence has also been established in various nonlinear autoregressive models and GARCH models (see Meyn and Tweedie

(1993), Lanne and Saikkonen (2005), Meitz and Saikkonen (2007), and Saikkonen (2007)). However, we have found it difficult to establish it in the present context. Therefore, we use the weaker Q -geometric ergodicity which also provides us with useful results.

The standard method to establish Q -geometric ergodicity, as well as its aforementioned stronger counterpart, is based on the so called drift criterion (see Meyn and Tweedie (1993, Theorem 15.0.1) or Liebscher (2005)). Before the application of this criterion one needs to show that the considered Markov chain is irreducible and aperiodic. In many nonlinear autoregressions of the type (6) this can be done in a fairly straightforward way. That also applies to our model if the function g in (2) is independent of h_{t-1} so that the conditional heteroskedasticity is of pure ARCH type. Then the analysis can be reduced to that of the process Y_t which is a Markov chain and one can employ the ideas in Cline and Pu (1998) and Lu (1998) to show irreducibility and aperiodicity. However, when the function g also depends on h_{t-1} we have to consider the larger Markov chain Z_t in which the deterministic dependence of h_t on past values of the process y_t through the nonlinear function f makes the analysis complicated and the approach described in Cline and Pu (1998) and Lu (1998) gets difficult. Similar difficulties occur when one tries to establish the T -continuity of Z_t which, in conjunction with irreducibility and aperiodicity, implies that compact subsets of \mathcal{Z} are small, a fact also pertinent for the application of the drift criterion (see Theorems 6.2.5(ii) and 5.5.7 of Meyn and Tweedie (1993)).

Due to the aforementioned difficulties we establish the irreducibility, aperiodicity, and T -continuity of Z_t by using the approach described in Chapter 7 of Meyn and Tweedie (1993). This approach is based on the deterministic control model associated with the nonlinear state space model (6) and it gives us the following result.

Lemma 2 *If Assumptions 1–4 and 6 hold then the Markov chain Z_t on \mathcal{Z} is an irreducible and aperiodic T -chain and, hence, all compact subsets of \mathcal{Z} are small. Moreover, the set $A_N = \{z \in \mathcal{Z} : \|y\|^{2r} \leq N, h^r(z) \leq N\}$ is small for any vector norm and for all N such that $\underline{g}^r < N$, where \underline{g} is as in Assumption 4(a).*

Thus, Lemma 2 provides the necessary prerequisites for the application of the drift criterion. Note that Lemma 2 also shows that certain noncompact subsets of \mathcal{Z} are small.

Unlike in many previous cases this result greatly facilitates the application of the drift criterion. This part of the lemma is based on ideas used by Cline and Pu (1998, Theorem 2.5) who also discuss its usefulness.

The following theorem presents the main result of the paper. In the proof of this theorem we apply an m -step ahead drift criterion for a sufficiently large value of m (cf. Theorem 19.1.3 of Meyn and Tweedie (1993)). In most previous cases 1-step ahead versions of this criterion have sufficed, but in the present model the combination of the assumed nonlinear autoregressive structure both in the conditional mean and conditional variance seems to make the application of this more conventional approach difficult. Although the possibility to make use of the m -step ahead drift criterion in nonlinear autoregressions was already pointed out by Tjøstheim (1990) it seems that its previous applications have been rather rare and confined to cases where a 1-step ahead drift criterion would have worked without any difficulty. A new m -step ahead drift criterion for Q -geometric ergodicity (Lemma 6), potentially of independent interest, is proven in the Appendix.

Theorem 1 *Suppose that Assumptions 1–4, 5(a), and 6 hold, and let $\|\cdot\|$ be any vector norm. Then the Markov chain Z_t on \mathcal{Z} is Q^* -geometrically ergodic in the sense of Definition 1 with a function $Q^*(z) \geq 1 + \|y\|^{2r} + h^r(z)$. If Assumption 5(b) is assumed instead of 5(a), the same conclusion holds with an unknown $r_0 \in (0, r)$ in place of r .*

As discussed after Definition 1, Theorem 1 implies that, with appropriate initial distributions, the process (y_t, h_t) is β -mixing and that there exists a stationary initial distribution such that y_t and h_t have moments of orders $2r$ and r (or $2r_0$ and r_0), respectively. An important consequence of Theorem 1 is that usual limit theorems apply.

4 Examples

We shall now consider concrete examples to which Theorem 1 applies. According to what was said after Assumption 1, it suffices to discuss Assumptions 2–6 of which Assumptions 2 and 3 concern the conditional mean of the model, that is, the function f , whereas Assumptions 4 and 5 restrict the form of permitted conditional heteroskedasticity. As

already indicated, Assumption 6 can often be checked without paying attention to the conditional mean. This is also the case in the examples below so that we can discuss verifying conditions imposed on the conditional mean and conditional variance separately.

First consider the verification of Assumptions 2 and 3 which restrict the conditional mean. Using either part of Lemma 1 these assumptions can be verified for the general functional-coefficient autoregressive model of Chen and Tsay (1993) and the exponential autoregressive model considered in Liebscher (2005). To this end, suppose that

$$f(y_{t-1}, \dots, y_{t-p}) = \phi_0 + \psi_0 G(y_{t-1}, \dots, y_{t-p}) + \sum_{j=1}^p (\phi_j + \psi_j G(y_{t-1}, \dots, y_{t-p})) y_{t-j},$$

where $\phi_j, \psi_j \in \mathbb{R}$, $j = 0, \dots, p$, and G is a smooth function with range $[0, 1]$. This model also encompasses the smooth transition autoregressive models discussed by Teräsvirta (1994) and van Dijk, Teräsvirta, and Franses (2002) among others. By Lemma 1(ii), a sufficient condition for Assumption 3 is $\sum_{j=1}^p \max\{|\phi_j|, |\phi_j + \psi_j|\} < 1$, a condition obtained by Chen and Tsay (1993, Example 2) for a special choice of the function G . Lemma 1(i) provides the sufficient condition $\rho(\{A_1, A_2\}) < 1$, where $A_1 = \bar{A}_p([\phi_1 \dots \phi_p]')$ and $A_2 = \bar{A}_p([\phi_1 + \psi_1 \dots \phi_p + \psi_p]')$ (cf. Liebscher (2005, Theorem 1 and Proposition 5)). As in Liebscher (2005, Section 7) one can find an example with $p = 2$ in which this latter condition is strictly weaker than the one obtained from Lemma 1(ii). In the case of a linear $\text{AR}(p)$ model, the conditions of Lemma 1(i) and (ii) become $\rho(A_1) < 1$ and $\sum_{j=1}^p |\phi_j| < 1$, respectively. The former is necessary and sufficient for the existence of a strictly stationary causal solution a linear $\text{AR}(p)$ model whereas the latter is only sufficient.

Now consider verifying Assumptions 4–6 for which it suffices to consider the model for conditional heteroskedasticity. Although the conditions imposed on the function g in Assumption 4 rule out threshold GARCH models they apply to smooth transition GARCH models introduced in Hagerud (1996) and González-Rivera (1998), and further discussed in Lundbergh and Teräsvirta (2002), Lanne and Saikkonen (2005), and Meitz and Saikkonen (2007). In one variant of this model the dynamics of the conditional variance process are governed by

$$h_t = g(u_{t-1}, h_{t-1}) = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + \alpha^* G(u_{t-1}) u_{t-1}^2, \quad (9)$$

where G is a smooth strictly increasing function with range $[0, 1]$ and the parameters satisfy $\omega > 0$, $\alpha > 0$, $\beta > 0$, and $\alpha + \alpha^* > 0$. This model reduces to the linear GARCH model of Bollerslev (1986) when $\alpha^* = 0$. Again, the possibility that G is an indicator function is ruled out but an approximating smooth counterpart is allowed giving a smooth version of the GJR specification of Glosten, Jaganathan, and Runkle (1993). Checking the validity of Assumptions 4–6 for this model is straightforward. The following proposition shows the result.

Proposition 1 *Let the real number $r > 0$ be as in Assumption 1 and suppose that h_t is generated by (9). Then, if either (a) $E[(\beta + \max\{\alpha, \alpha + \alpha^*\}\varepsilon_t^2)^r] < 1$, or (b) $E[|\varepsilon_t|^{2r}] < \infty$ and $E[\log(\beta + \max\{\alpha, \alpha + \alpha^*\}\varepsilon_t^2)] < 0$, Assumptions 4–6 hold.*

In the special case $r = 1$ the moment condition of Proposition 1(a) reduces to $\beta + \max\{\alpha, \alpha + \alpha^*\} < 1$. In case of the standard linear GARCH model this becomes $\alpha + \beta < 1$, a condition that is necessary and sufficient for the second order stationarity of the process u_t and hence for that of y_t (see Bollerslev (1986)). The latter condition in part (b) reduces to $E[\log(\beta + \alpha\varepsilon_t^2)] < 0$ in case of the standard linear GARCH model. This condition agrees with the necessary and sufficient condition for the (strict) stationarity and geometric ergodicity of the conditional variance process obtained in Nelson (1990) and Francq and Zakoïan (2006), respectively.

Our second example is concerned with an alternative smooth transition GARCH model suggested by Lanne and Saikkonen (2005). In this case the conditional heteroskedasticity is specified as

$$h_t = g(u_{t-1}, h_{t-1}) = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + (\omega^* + \beta^* h_{t-1})G(h_{t-1}). \quad (10)$$

Here the parameters and the function G are as described in the following proposition which provides sufficient conditions for Assumptions 4–6 to hold.

Proposition 2 *Let the real number $r > 0$ be as in Assumption 1 and suppose that h_t is generated by (10) where G is a smooth strictly increasing nonnegative function. If either of the following conditions is satisfied, then Assumptions 4–6 hold.*

(i) *The function G has range $[0, 1]$, the parameters satisfy $\omega > 0$, $\alpha > 0$, $\beta > 0$,*

$\omega^* \geq 0$, and $\beta^* \geq 0$, and either (a) $E[(\beta + \beta^* + \alpha\varepsilon_t^2)^r] < 1$ or (b) $E[|\varepsilon_t|^{2r}] < \infty$ and $E[\log(\beta + \beta^* + \alpha\varepsilon_t^2)] < 0$.

(ii) The function G is either as in (i) or satisfies $G(x) = o(x)$ as $x \rightarrow \infty$, the parameters satisfy $\omega > 0$, $\alpha > 0$, $\omega^* > 0$, and $\beta = \beta^* = 0$, and either (a) $E[\alpha^r |\varepsilon_t|^{2r}] < 1$ or (b) $E[|\varepsilon_t|^{2r}] < \infty$ and $E[\log(\alpha\varepsilon_t^2)] < 0$.

In the special case $r = 1$ the moment conditions of Proposition 2(i.a) and (ii.a) reduce to $\beta + \beta^* + \alpha < 1$ and $\alpha < 1$, respectively. Again, one can use the alternative conditions stated in parts (i.b) and (ii.b). The specification considered in case (ii) was applied by Lanne and Saikkonen (2005) with the function G satisfying the conditions in (i).

The result of Proposition 2(ii) actually also holds when $\omega > 0$, $\alpha > 0$, and $\omega^* = \beta = \beta^* = 0$, in which case the conditional variance reduces to the pure ARCH model $h_t = \omega + \alpha u_{t-1}^2$. However, in this case the verification of Assumption 6 requires a different argument, and we refer to Meitz and Saikkonen (2006) for details.

5 Conclusion

In this paper we have studied a nonlinear $AR(p)$ model with conditionally heteroskedastic errors specified as a nonlinear GARCH(1,1) model. We gave conditions under which the Markov chain representation of the model is Q -geometrically ergodic in the sense of Definition 1 and, hence, β -mixing. Conditions for existence of moments of the stationary distribution were also obtained. Only mild moment conditions were required to obtain these results. Furthermore, a convenient feature of the assumptions needed is that they often restrict the conditional mean and conditional variance separately and can be readily verified.

Due to the approach taken to obtain the results of the paper, rather stringent smoothness assumptions on the permitted nonlinearity were needed, and hence threshold type nonlinear models could not be covered. However, we were still able to provide stability results that cover a number of nonlinearities considered in the recent literature. Our results are of importance as they open up the way for the development of rigorous asymptotic estimation theory for these models.

Appendix: Proofs

In this Appendix, we refer to Meyn and Tweedie (1993) as ‘MT’. We also denote the indicator function of the event A with $\mathbf{1}_A$.

Proof of Lemma 1. (i) First note that \mathcal{A}_* and \mathcal{A}_1 are both bounded sets of matrices. That $\rho(\mathcal{A}_*) < 1$ implies Assumption 3 follows from Thm 1 of Liebscher (2005). To see that $\rho(\mathcal{A}_1) < 1$ is equivalent to this condition, notice that

$$A(x_1)A(x_2) \cdots A(x_k) = \begin{bmatrix} A_1(x_1)A_1(x_2) \cdots A_1(x_k) & 0 \\ \iota'_p A_1(x_2)A_1(x_3) \cdots A_1(x_k) & 0 \end{bmatrix},$$

where $\iota_p = [0 \cdots 0 \ 1]'$ ($p \times 1$). Thus, the stated equivalence can be established by choosing the norm in the definition of the joint spectral radius as the maximum of absolute row sums (the matrix norm induced by the l_∞ -norm).

(ii) Using the notation of Section 2, denote $A = \bar{A}_{p+1}([\alpha_1 \cdots \alpha_p \ 0]')$ and notice that the characteristic polynomial of A is (up to a factor ± 1) $\lambda(\lambda^p - \alpha_1 \lambda^{p-1} - \cdots - \alpha_p)$. The equivalence of the two conditions can be seen as in Ling (1999, Lemma 2.3). Thus, under either condition $\rho(A) < 1$ (cf. Chen and Tsay (1993, proof of Thm 1.1)).

Next, using the same argument as in Ling and McAleer (2003, proof of Lemma A.2) we can find a $(p+1) \times 1$ vector κ with positive components such that the components of the row vector $\nu' = \kappa'(I_{p+1} - A)$ are positive and, furthermore, $0 < \underline{\nu}/\bar{\kappa} < 1$ where $\underline{\nu}$ and $\bar{\kappa}$ are the smallest and largest components of ν and κ , respectively. Now define the vector norm $\|\cdot\|^*$ in \mathbb{R}^{p+1} by $\|y\|^* = \sum_{j=1}^{p+1} \kappa_j |y_j| = \kappa' |y|$ where $|y| = [|y_1| \cdots |y_{p+1}|]'$. For arbitrary $A = A(x) \in \mathcal{A}_*$ and $y \in \mathbb{R}^{p+1}$, $y \neq 0$, we have $\|A(x)y\|^* \leq \kappa_1 \sum_{j=1}^p \alpha_j |y_j| + \sum_{j=1}^p \kappa_{j+1} |y_j| = \kappa' |y| \left(1 - \frac{\nu'_j |y_j|}{\kappa'_j |y|}\right) \leq \|y\|^* (1 - \underline{\nu}/\bar{\kappa})$, where $0 < 1 - \underline{\nu}/\bar{\kappa} < 1$. This shows that the matrix norm induced by $\|\cdot\|^*$ satisfies Assumption 3. ■

Proof of Lemma 2. We consider Z_t as a nonlinear state space model and use the results in Chapter 7 of MT (note that under our assumptions, the conditions (NSS1)–(NSS3) in MT, pp. 32 and 156, are satisfied). For this we need to show that the deterministic control model associated with Z_t is forward accessible and attains a globally attracting state (for definitions of these concepts, see pp. 155 and 160 of MT, respectively). As discussed in Section 2, forward accessibility follows from Assumption 6 and Propn 7.1.4 of MT. The

existence of a globally attracting state is shown below in Lemma 3. Thus, from Props 7.1.5 and 7.2.5(i), and Thm 7.2.6 of MT we can conclude that the Markov chain Z_t is an irreducible T -chain. Aperiodicity is obtained from Thms 7.3.3 and 7.3.5(ii) of MT (see also the proof of Propn 7.4.1) because any cycle of the associated control model must contain the globally attracting state (in Lemma 3 we also show that there exists a control sequence such that the deterministic control model converges to the globally attracting state, and thus the period in Thm 7.3.3 of MT necessarily equals one). That every compact set is small now follows from Thms 6.2.5(ii) and 5.5.7 of MT. Finally, in Lemma 4 below it is shown that the set A_N is also small.

Thus, the proof of Lemma 2 is completed by the following two lemmas. ■

Lemma 3 *Under Assumptions 1–4 the deterministic control model associated with the Markov chain Z_t attains a globally attracting state.*

Proof. For a $z^* \in \mathcal{Z}$ to be a globally attracting state for the associated deterministic control model it suffices to establish that, for any initial value $z_0 \in \mathcal{Z}$, there exists a control sequence e_t such that z_t converges to z^* as $t \rightarrow \infty$ (see MT, p. 160). First suppose that the convergence in Assumption 4(c) holds for all $h_0 \in \mathbb{R}_+$ so that for every $z_0 \in \mathcal{Z}$, $h_t \rightarrow h^*$ as $t \rightarrow \infty$.

By Assumption 3 there exist an induced matrix norm $\|\cdot\|^*$ and a $\rho \in (0, 1)$ such that $\|A\|^* \leq \rho$ for all $A \in \mathcal{A}_*$. As in Assumption 3 we also use $\|\cdot\|^*$ for the vector norm corresponding to the matrix norm $\|\cdot\|^*$. Because $b(\cdot)$ is bounded, we can find a $c \in \mathbb{R}_+$ such that $\|\iota b(x)\|^* \leq c/2$ for all $x \in \mathbb{R}^p$. Define the compact set $K = \{y \in \mathbb{R}^{p+1} : \|y\|^* \leq c/(1 - \rho)\}$ and note that the mapping $y \mapsto A(S'y)y + \iota b(S'y)$ ($y \in \mathbb{R}^{p+1}$) is continuous. Furthermore, when $y \in K$, the range of this mapping is contained in K because, for $y \in K$,

$$\begin{aligned} \|A(S'y)y + \iota b(S'y)\|^* &\leq \|A(S'y)\|^* \|y\|^* + \|\iota b(S'y)\|^* \\ &\leq \rho \|y\|^* + c/2 \\ &\leq \rho c/(1 - \rho) + c/2 \\ &\leq c/(1 - \rho). \end{aligned}$$

Thus, it follows from Schauder's fixed point theorem (see e.g. Simmons (1963, Appendix 1)) that there exists a state $y^* \in K$ such that $y^* = A(S'y^*)y^* + \iota b(S'y^*)$.

We shall now demonstrate that, from any $z_0 \in \mathcal{Z}$, the associated control model can reach a state z_{p+1}^* whose first $p+1$ components are y_1^*, \dots, y_{p+1}^* , the components of the vector y^* , and the last component is h_{p+1} . Let $z_0 = [y_0' \ h_0']' \in \mathcal{Z}$ where $y_0 = [y_{0,1} \ \dots \ y_{0,p+1}]'$. From the first step of the associated control model one then obtains

$$\begin{aligned} y_1 &= [a(S'y_0)' : 0] y_0 + b(S'y_0) + h(z_0)^{1/2} e_1 \\ h_1 &= h(z_0) \end{aligned}$$

and with $e_1 = h(z_0)^{-1/2} (y_{p+1}^* - [a(S'y_0)' : 0] y_0 - b(S'y_0))$ we get $y_1 = y_{p+1}^*$. Next, setting $\bar{y}_1^* = [y_{p+1}^* \ y_{0,1} \ \dots \ y_{0,p}]'$ and $z_1^* = [\bar{y}_1^{*'} \ h_1']'$ the second step of the associated control model gives

$$\begin{aligned} y_2 &= [a(S'\bar{y}_1^*)' : 0] \bar{y}_1^* + b(S'\bar{y}_1^*) + h(z_1^*)^{1/2} e_2 \\ h_2 &= h(z_1^*), \end{aligned}$$

which with $e_2 = h(z_1^*)^{-1/2} (y_p^* - [a(S'\bar{y}_1^*)' : 0] \bar{y}_1^* - b(S'\bar{y}_1^*))$ yields $y_2 = y_p^*$. The next step is to set $\bar{y}_2^* = [y_p^* \ y_{p+1}^* \ y_{0,1} \ \dots \ y_{0,p-1}]'$ and $z_2^* = [\bar{y}_2^{*'} \ h_2']'$ and choose $e_3 = h(z_2^*)^{-1/2} \times (y_{p-1}^* - [a(S'\bar{y}_2^*)' : 0] \bar{y}_2^* - b(S'\bar{y}_2^*))$. This gives $y_3 = y_{p-1}^*$ and $z_3^* = [\bar{y}_3^{*'} \ h_3']'$ defined in an obvious way. Continuing in this way we reach the state $z_{p+1}^* = [y_1^* \ \dots \ y_{p+1}^* \ h(z_p^*)']' = [y^{*'} \ h_{p+1}^*']'$ in $p+1$ steps.

Next form z_t^* with $e_t = 0$, $t = p+2, p+3, \dots$. Because $y^* = A(S'y^*)y^* + \iota b(S'y^*)$ the first $p+1$ components of z_t^* will be the components of y^* for all $t \geq p+2$. Thus, $z_t^* = [y^{*'} \ h_t^*']'$ ($t \geq p+2$) where the last component satisfies $h_t^* = g(0, h_{t-1}^*)$ for $t \geq p+3$. Because Assumption 4(c) implies that $h_t^* \rightarrow h^*$ as $t \rightarrow \infty$ we can conclude that $z^* = [y^{*'} \ h^*']'$ is a globally attracting state for the associated control model.

Now suppose that the convergence in Assumption 4(c) holds for all $h_0 \geq h^*$. By Assumption 4(b) we can first choose an e_1 such that $h_2 = g(h_1^{1/2} e_1, h_1) > h^*$. As seen above, we can next choose e_2, \dots, e_{p+2} to reach a state whose first $p+1$ components are y_1^*, \dots, y_{p+1}^* , the components of the vector y^* . This can be done regardless of the initial value z_0 . Because $h_2 > h^*$, the relevant part of Assumption 4(c) implies $h_3 =$

$g(h_2^{1/2}e_2, h_2) \geq h^*$ and similarly $h_k \geq h^*$ for $k = 4, \dots, p+2$. Thus, after $p+2$ steps we are in a state $z_{p+2}^* = [y^{*'} \ h_{p+2}]'$ and we continue by forming z_t^* with $e_t = 0, t = p+3, p+4, \dots$. Then the first $p+1$ components of z_t^* will not change and, because $h_{p+2} \geq h^*$, the last one tends to h^* as $t \rightarrow \infty$. Thus a globally attracting state exists. ■

Lemma 4 *Under Assumptions 1–4 and 6 the set $A_N = \{z \in \mathcal{Z} : \|y\|^{2r} \leq N, h^r(z) \leq N\}$ is small for any vector norm and for all N such that $\underline{g}^r < N$.*

Proof. Writing equation (5) as $Z_t = F_0(Z_{t-1}) + \iota h(Z_{t-1})^{1/2} \varepsilon_t$ we have

$$\begin{aligned} E[\|Z_t\|^{2r} \mid Z_{t-1} = z] &= E\left[\left\|F_0(z) + \iota h(z)^{1/2} \varepsilon_t\right\|^{2r}\right] \\ &\leq \Delta_{2r} \|F_0(z)\|^{2r} + \Delta_{2r} \|\iota\|^{2r} h(z)^r E[|\varepsilon_t|^{2r}], \end{aligned}$$

where $\Delta_{2r} = 1$, when $2r \leq 1$ and $\Delta_{2r} = 2^{2r-1}$, when $2r \geq 1$ (see Davidson (1994, p. 140)). Since the functions F_0 and h are bounded on A_N , we can find an $M_N < \infty$ such that

$$\sup_{z \in A_N} E[\|Z_t\|^{2r} \mid Z_{t-1} = z] < M_N^{2r}. \quad (11)$$

Now define the set $B_N = \{z \in \mathcal{Z} : \|z\| \leq M_N, h \geq \underline{g}\}$ (where h is the last component of z). This set is small because it is compact, as already noted. Furthermore,

$$\begin{aligned} \inf_{z \in A_N} \Pr(Z_t \in B_N \mid Z_{t-1} = z) &= 1 - \sup_{z \in A_N} \Pr(Z_t \notin B_N \mid Z_{t-1} = z) \\ &\geq 1 - \sup_{z \in A_N} \Pr(\|Z_t\| \geq M_N \mid Z_{t-1} = z) \\ &\geq 1 - \sup_{z \in A_N} E[\|Z_t\|^{2r} \mid Z_{t-1} = z] / M_N^{2r} \\ &> 0. \end{aligned}$$

Here the first inequality is justified by the fact that, for all z , $\Pr(Z_t \notin B_N \mid Z_{t-1} = z) = \Pr(\|Z_t\| > M_N \text{ or } h(z) < \underline{g} \mid Z_{t-1} = z)$ but $h(z) < \underline{g}$ is impossible by Assumption 4(a). The second inequality is Markov's and the third one is due to (11). That the set A_N is small can now be concluded from Propn 5.2.4 of MT. ■

Proof of Theorem 1. We present the proof by assuming Assumption 5(a). If Assumption 5(b) is assumed instead, the proof remains valid with an unknown $r_0 \in (0, r)$ in place of r (in this case r_0 replaces r also in Lemmas 2, 4, and 5). First note that, by Lemma

2, Z_t is irreducible and aperiodic and the set A_N is small. Let $\|\cdot\|$ be any vector norm, and let $\|\cdot\|^*$ be an induced matrix norm that satisfies $\|A\|^* \leq \rho$ for all $A \in \mathcal{A}_*$ and with $\rho \in (0, 1)$ (see Assumption 3). By the equivalence of all vector norms in finite-dimensional vector spaces, there exists a finite $C > 0$ such that $\|y\| \leq C^{1/2r} \|y\|^*$ for all $y \in \mathbb{R}^{p+1}$ (see e.g. Horn and Johnson (1985, Sec. 5.4)). Denote $V_*(z) = 1 + C\|y\|^{*2r} + h^r(z)$. In Lemma 5 we demonstrate that the conditional expectation $E[V_*(Z_t) \mid Z_{t-m} = z]$ satisfies the m -step ahead drift criterion given in condition (19.15) of MT (with the choice $n(z) \equiv m$ and for a large m to be chosen in the proof of the lemma). An application of our Lemma 6 below then establishes that Z_t is V_* -geometrically ergodic in the sense of Definition 1.

Thus, the following two lemmas complete the proof of Theorem 1. ■

Lemma 5 *Suppose the assumptions of Theorem 1 are satisfied and define the function $V_*(z) = 1 + C\|y\|^{*2r} + h^r(z)$. Then, there exist a small set K , a positive integer m , and positive real numbers $\lambda < 1$ and b such that*

$$E[V_*(Z_t) \mid Z_{t-m} = z] \leq \lambda^{1/2} (1 + C\|y\|^{*2r} + h^r(z) + b\mathbf{1}_{\{z \in K\}}). \quad (12)$$

In other words, the drift condition (19.15) of Thm 19.1.3 of MT holds (with the choice $n(z) \equiv m$).

Proof. First note that, for any positive x_i , $1 \leq i \leq n$, $n \in \mathbb{Z}_+$, and $r > 0$,

$$\left(\sum_{i=1}^n x_i \right)^r \leq \Delta_{r,n} \sum_{i=1}^n x_i^r \quad (13)$$

where $\Delta_{r,n} = \max \{1, n^{r-1}\}$ (see Davidson (1994, p. 140)).

To analyze the conditional expectation in the lemma we first consider the quantity $h(Z_{t-1})$. From equations (1), (4), and (5) we obtain $h(Z_{t-m+1}) = g(h^{1/2}(Z_{t-m})\varepsilon_{t-m+1}, h(Z_{t-m}))$, and, by using Assumption 4(d) with the notation $c_{t-1} = a + \varphi(\varepsilon_{t-1})$, this quantity can be bounded from above with $h(Z_{t-m})c_{t-m+1} + c$. Therefore we have, for $k \geq 1$,

$$h(Z_{t-m+k}) \leq \prod_{j=1}^k c_{t-m+j} \cdot h(Z_{t-m}) + c \left(1 + \sum_{j=0}^{k-2} \prod_{i=0}^j c_{t-m+k-i} \right).$$

Using (13) we obtain

$$\Delta_{r,k+1}^{-1} h^r(Z_{t-m+k}) \leq \prod_{j=1}^k c_{t-m+j}^r \cdot h^r(Z_{t-m}) + c^r \left(1 + \sum_{j=0}^{k-2} \prod_{i=0}^j c_{t-m+k-i}^r \right).$$

By Assumption 5(a), $E[c_t^r] < 1$ and we denote this expectation by δ . Furthermore, denote $d = c^r/(1 - \delta)$. By the independence of the c_t 's,

$$\begin{aligned} \Delta_{r,k+1}^{-1} E[h^r(Z_{t-m+k}) \mid Z_{t-m} = z] &\leq h^r(z)\delta^k + c^r(1 + \sum_{j=0}^{k-2} \delta^{j+1}) \\ &\leq h^r(z)\delta^k + d. \end{aligned} \quad (14)$$

In particular, setting $d' = \Delta_{r,m}d$ we have, for $k = 1, \dots, m-1$

$$\begin{aligned} E[h^r(Z_{t-m+k}) \mid Z_{t-m} = z] &\leq \Delta_{r,k+1}(h^r(z)\delta^k + d) \\ &\leq \Delta_{r,m}h^r(z)\delta^k + d'. \end{aligned} \quad (15)$$

Now consider Y_t which we wish to express in terms of past values of the process Z_t until $t-m$. Recall that $\|\cdot\|^*$ and ρ are as in Assumption 3. Repeated substitution, usual properties of vector and matrix norms, and an application of (13) yield

$$\begin{aligned} \Delta_{2r,2m+1}^{-1} \|Y_t\|^{*2r} &\leq \prod_{j=0}^{m-1} \|A(S'Y_{t-1-j})\|^{*2r} \|Y_{t-m}\|^{*2r} + \|\iota b(S'Y_{t-1})\|^{*2r} \\ &\quad + \sum_{j=0}^{m-2} \prod_{i=0}^j \|A(S'Y_{t-1-i})\|^{*2r} \|\iota b(S'Y_{t-2-j})\|^{*2r} + \|\iota h(Z_{t-1})^{1/2} \varepsilon_t\|^{*2r} \\ &\quad + \sum_{j=0}^{m-2} \prod_{i=0}^j \|A(S'Y_{t-1-i})\|^{*2r} \|\iota h(Z_{t-2-j})^{1/2} \varepsilon_{t-1-j}\|^{*2r}. \end{aligned}$$

Denote $\|\iota\|^{*2r} = \iota^*$ and note that $\|A(\cdot)\|^{*2r} \leq \rho^{2r}$, $\|\iota b(\cdot)\|^{*2r} \leq \iota^* B$ for some finite B (since $b(\cdot)$ is bounded), $\|\iota h(\cdot)^{1/2} \varepsilon_t\|^{*2r} \leq \iota^* h^r(\cdot) |\varepsilon_t|^{2r}$, and $E[|\varepsilon_t|^{2r}] \stackrel{\text{def}}{=} \gamma_{2r} < \infty$. Thus,

$$\begin{aligned} &\Delta_{2r,2m+1}^{-1} E[\|Y_t\|^{*2r} \mid Z_{t-m} = z] \\ &\leq \left(\prod_{j=0}^{m-1} \rho^{2r} \right) \|y\|^{*2r} + \iota^* B + \sum_{j=0}^{m-2} \left(\prod_{i=0}^j \rho^{2r} \right) \iota^* B + \iota^* E[h^r(Z_{t-1}) \mid Z_{t-m} = z] \gamma_{2r} \\ &\quad + \sum_{j=0}^{m-2} \left(\prod_{i=0}^j \rho^{2r} \right) \iota^* E[h^r(Z_{t-2-j}) \mid Z_{t-m} = z] \gamma_{2r} \\ &\leq \rho^{2rm} \|y\|^{*2r} + \iota^* B \left(1 + \sum_{j=0}^{m-2} \rho^{2r(j+1)} \right) + \iota^* \gamma_{2r} (\Delta_{r,m} \delta^{m-1} h^r(z) + d') \\ &\quad + \iota^* \gamma_{2r} \left(\sum_{j=0}^{m-3} \rho^{2r(j+1)} (\Delta_{r,m} \delta^{m-2-j} h^r(z) + d') + \rho^{2r(m-1)} h^r(z) \right), \end{aligned}$$

where the last inequality makes use of (15) and the fact that $E[h^r(Z_{t-m}) \mid Z_{t-m} = z] = h^r(z)$. Defining $\phi = \max\{\rho^{2r}, \delta\} < 1$ and $\phi' = \frac{1}{1-\phi}$ we get

$$\begin{aligned}
& \Delta_{2r,2m+1}^{-1} E[\|Y_t\|^{*2r} \mid Z_{t-m} = z] \\
& \leq \phi^m \|y\|^{*2r} + \iota^* B \left(1 + \sum_{j=0}^{m-2} \phi^{j+1}\right) + \iota^* \gamma_{2r} (\Delta_{r,m} \phi^{m-1} h^r(z) + d') \\
& \quad + \iota^* \gamma_{2r} \left(\sum_{j=0}^{m-3} \phi^{j+1} (\Delta_{r,m} \phi^{m-2-j} h^r(z) + d') + \phi^{m-1} h^r(z)\right) \\
& \leq \phi^m \|y\|^{*2r} + \iota^* B \phi' + \iota^* \gamma_{2r} (\Delta_{r,m} \phi^{m-1} h^r(z) + d') \\
& \quad + \iota^* \gamma_{2r} \left(\sum_{j=0}^{m-3} \phi^{m-1} \Delta_{r,m} h^r(z) + \sum_{j=0}^{m-3} \phi^{j+1} d' + \Delta_{r,m} \phi^{m-1} h^r(z)\right) \\
& \leq \phi^m \|y\|^{*2r} + m \cdot \iota^* \gamma_{2r} \Delta_{r,m} \phi^{m-1} h^r(z) + \iota^* \phi' (B + \gamma_{2r} d'). \tag{16}
\end{aligned}$$

Combining the inequalities (14) (with $k = m$) and (16) yields

$$\begin{aligned}
& E[V_*(Z_t) \mid Z_{t-m} = z] \\
& = E[1 + C\|Y_t\|^{*2r} + h^r(Z_t) \mid Z_{t-m} = z] \\
& \leq 1 + C\Delta_{2r,2m+1} (\phi^m \|y\|^{*2r} + \iota^* \gamma_{2r} m \Delta_{r,m} \phi^{m-1} h^r(z) + \iota^* \phi' (B + \gamma_{2r} d')) \\
& \quad + \Delta_{r,m+1} (h^r(z) \delta^m + d) \\
& = 1 + C[\Delta_{2r,2m+1} \phi^m] \|y\|^{*2r} + [C\iota^* \gamma_{2r} \Delta_{2r,2m+1} m \Delta_{r,m} \phi^{m-1} + \Delta_{r,m+1} \delta^m] h^r(z) \\
& \quad + \{C\Delta_{2r,2m+1} \iota^* \phi' (B + \gamma_{2r} d') + \Delta_{r,m+1} d\}. \tag{17}
\end{aligned}$$

Since $0 < \delta \leq \phi < 1$, it follows from the definitions that we can choose an m large enough so that both of the expressions in square brackets in (17) are smaller than some $\lambda < 1$. The expression in curly brackets in (17) is clearly finite, and thus for some $L < \infty$

$$\begin{aligned}
& E[V_*(Z_t) \mid Z_{t-m} = z] \\
& \leq \lambda (1 + C\|y\|^{*2r} + h^r(z)) + L \\
& = \lambda^{1/2} (1 + C\|y\|^{*2r} + h^r(z)) \cdot \lambda^{1/2} \left(1 + \frac{L}{\lambda (1 + C\|y\|^{*2r} + h^r(z))}\right). \tag{18}
\end{aligned}$$

What remains to be examined is the behaviour of the last expression on and off a small set. By Lemma 4 the set $A_N = \{z \in \mathcal{Z} : \|y\|^{*2r} \leq N, h^r(z) \leq N\}$ is small. Off this set either $\|y\|^{*2r} > N$ or $h^r(z) > N$, and the ratio in the last expression of (18) can clearly

be made arbitrarily small by choosing N large enough. Therefore, for a large enough N , $E[V_*(Z_t) \mid Z_{t-m} = z] \leq \lambda^{1/2} (1 + C\|y\|^{*2r} + h^r(z))$ off the set A_N . Finally, on the set A_N the last expression of (18) is clearly bounded, and therefore condition (12) is satisfied. ■

Lemma 6 *Let X_t be an irreducible and aperiodic Markov chain on a state space \mathcal{X} , and let m be a positive integer. Suppose that for a small set K , a function $V : \mathcal{X} \rightarrow [1, \infty)$ bounded on K , and positive constants $\lambda < 1$ and $b < \infty$*

$$E[V(X_t) \mid X_{t-m} = x] \leq \lambda^m (V(x) + b\mathbf{1}_{\{x \in K\}}) \quad (19)$$

for all $x \in \mathcal{X}$. Then X_t is V -geometrically ergodic in the sense of Definition 1.

Proof. If $m = 1$ then X_t is V -geometrically ergodic in the sense of MT by their Thm 15.0.1, and hence the stated weaker form of geometric ergodicity also follows. Suppose now that $m > 1$. By Thm 19.1.3 of MT, X_t is geometrically ergodic and $\|P_X^n(x, \cdot) - \pi_X(\cdot)\| \leq \varrho^n R V(x)$ for some $\varrho < 1$ and $R < \infty$ ($\|\cdot\|$ signifies the total variation norm, and $P_X^n(x, \cdot)$ and $\pi_X(\cdot)$ the n -step transition probability measure and stationary measure of X_t , respectively). What remains to be proven is that $\int_{\mathcal{X}} \pi_X(dy) V(y)$ is finite. To this end, consider X_{tm} , the m -skeleton of X_t , which by Propn 5.4.5(iii) of MT is irreducible and aperiodic. The m -skeleton satisfies (19), but the set K is not necessarily small for X_{tm} . However, by Lemma 14.2.8 of MT, for any $\bar{\lambda}$ such that $\lambda^m < \bar{\lambda} < 1$, we can find a set K_m which is small for the m -skeleton and such that $\mathbf{1}_{\{x \in K\}} \leq \sum_{i=0}^{m-1} \int_{\mathcal{X}} P_X^i(x, dy) \mathbf{1}_{\{y \in K\}} \leq m\mathbf{1}_{\{x \in K_m\}} + (\bar{\lambda} - \lambda^m)/\lambda^m b$. Therefore

$$\begin{aligned} E[V(X_t) \mid X_{t-m} = x] &\leq \lambda^m V(x) + \lambda^m b [m\mathbf{1}_{\{x \in K_m\}} + (\bar{\lambda} - \lambda^m)/\lambda^m b] \\ &\leq \bar{\lambda} V(x) + \lambda^m b m \mathbf{1}_{\{x \in K_m\}}, \end{aligned}$$

and thus the m -skeleton satisfies a drift criterion with the set K_m . Now, by Thm 15.0.1 of MT, the m -skeleton is V -geometrically ergodic in the sense of MT. Finally, by Thm 10.4.5 of MT, the stationary distributions of X_{tm} and X_t are the same and, by the V -geometric ergodicity of the m -skeleton, $\int_{\mathcal{X}} \pi_X(dy) V(y)$ is finite. ■

Proof of Proposition 1. Assumptions 4(a) and 4(b) clearly hold. For Assumption 4(d) we choose $a = \beta$, $\varphi(\varepsilon_t) = \max\{\alpha, \alpha + \alpha^*\} \varepsilon_t^2$, and $c = \omega$. Since $g(0, x) = \omega + \beta x$

and $\beta < 1$, the Lipschitz condition (7), and hence Assumption 4(c), is satisfied. The assumptions made now guarantee the moment conditions of Assumption 5.

To see that Assumption 6 holds, we consider the last row of the derivative matrix $\nabla F_{p+2}^{(0)}$ (cf. the discussion in Section 2). The needed derivatives can be straightforwardly obtained from equation (9) and, unless otherwise stated, all derivatives below are evaluated at $e_2^{(0)} = \dots = e_{p+2}^{(0)} = 0$. First note that $\partial h_i / \partial e_j = \beta \partial h_{i-1} / \partial e_j$, $i = 3, \dots, p+2$, $j = 1, \dots, i-2$, and $\partial h_i / \partial e_{i-1} = 0$, $i = 3, \dots, p+2$, and thus the last row of the matrix $\nabla F_{p+2}^{(0)}$ becomes $[\partial h_{p+2} / \partial e_1 \quad 0 \quad \dots \quad 0]$. To obtain $\partial h_{p+2} / \partial e_1$ we calculate $\partial h_2 / \partial e_1$ (evaluated at an arbitrary e_1) and find that

$$\partial h_{p+2} / \partial e_1 = 2\beta^p h(z_0) e_1 (\alpha + \alpha^* G(h(z_0)^{1/2} e_1)) + \beta^p \alpha^* h(z_0)^{3/2} G'(h(z_0)^{1/2} e_1) e_1^2.$$

First suppose that $\alpha^* = 0$. Then, since $\alpha > 0$ and $\beta > 0$, $\partial h_{p+2} / \partial e_1 = 2\beta^p \alpha h(z_0) e_1$ is nonzero for any $e_1 \neq 0$ whereas $\partial h_{p+2} / \partial e_2, \dots, \partial h_{p+2} / \partial e_{p+1}$ are zero. Thus, as discussed in Section 2, Assumption 6 holds. The same conclusion is obtained even if $\alpha^* \neq 0$. Then $\partial h_{p+2} / \partial e_1 \neq 0$ may not hold for all $e_1 \neq 0$ without further assumptions on the derivative G' but it clearly holds for some $e_1 \neq 0$, which suffices for Assumption 6. ■

Proof of Proposition 2. (i) The validity of Assumptions 4(a) and (b) is again clear. For Assumption 4(d), we may choose $a = \beta + \beta^*$, $\varphi(\varepsilon_t) = \alpha \varepsilon_t^2$, and $c = \omega + \omega^*$. To verify Assumption 4(c), the latter part of this assumption is convenient. For this model, $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$. Thus, since the function $G(x)$ is nondecreasing the same is true for $g(0, x)$ ($x > 0$) and it follows that Assumption 4(c) holds (see the discussion after Assumption 4). Assumption 5 holds due to the assumptions made, and that Assumption 6 holds will be shown below in the proof of (ii).

(ii) If G is as in (i) the proof that Assumptions 4 and 5 hold is similar to that above. If G is increasing and $G(x) = o(x)$ as $x \rightarrow \infty$, it suffices to discuss Assumptions 4(c) and (d). The former holds again because $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$. Regarding the latter, we can write

$$\begin{aligned} g(x^{1/2} \varepsilon_t, x) &= \omega + \omega^* G(x) + \alpha x \varepsilon_t^2 \\ &= \left(\frac{\omega + \omega^* G(x)}{x} \mathbf{1}_{\{x > M\}} + \alpha \varepsilon_t^2 \right) x + (\omega + \omega^* G(x) \mathbf{1}_{\{x \leq M\}}). \end{aligned}$$

Choosing M large enough the last expression can be bounded from above by $(\epsilon + \alpha\epsilon_t^2)x + c$ where $0 < \epsilon < 1$ is so small that $E[(\epsilon + \alpha\epsilon_t^2)^r] < 1$ holds whenever $E[\alpha^r |\epsilon_t|^{2r}] < 1$. Thus, Assumptions 4(d) and 5 hold.

Now consider Assumption 6. In the same way as in the case $\alpha^* = 0$ in the proof of Proposition 1, it is straightforward to check that, when evaluated at $e_2^{(0)} = \dots = e_{p+2}^{(0)} = 0$, $\partial h_i / \partial e_j = \omega^* G'(h_{i-1}) \partial h_{i-1} / \partial e_j$, $i = 3, \dots, p+2$, $j = 1, \dots, i-2$, and $\partial h_i / \partial e_{i-1} = 2\alpha h_{i-1} e_{i-1}$, $i = 3, \dots, p+2$. Thus, since the function G is increasing we can choose $e_1 = e_1^{(0)} \neq 0$ such that $\partial h_{p+2} / \partial e_1$ evaluated at $e_1^{(0)}$ and $e_2^{(0)} = \dots = e_{p+2}^{(0)} = 0$ becomes nonzero while $\partial h_{p+2} / \partial e_2 = \dots = \partial h_{p+2} / \partial e_{p+1} = 0$. Hence Assumption 6 holds. The same reasoning applies to part (i), for it suffices to consider the (increasing) function $\beta x + (\omega^* + \beta^* x)G(x)$ in place of $\omega^* G(x)$. ■

References

- AN, H. Z., AND S. G. CHEN (1997): “A note on the ergodicity of non-linear autoregressive model,” *Statistics and Probability Letters*, 34, 365–372.
- AN, H. Z., AND F. C. HUANG (1996): “The geometrical ergodicity of nonlinear autoregressive models,” *Statistica Sinica*, 6, 943–956.
- BASRAK, B., R. A. DAVIS, AND T. MIKOSCH (2002): “Regular variation of GARCH processes,” *Stochastic Processes and their Applications*, 99, 95–115.
- BHATTACHARYA, R., AND C. LEE (1995): “On the geometric ergodicity of nonlinear autoregressive models,” *Statistics and Probability Letters*, 22, 311–315.
- BOLLERSLEV, T. (1986): “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.
- CARRASCO, M., AND X. CHEN (2002): “Mixing and moment properties of various GARCH and stochastic volatility models,” *Econometric Theory*, 18, 17–39.
- CHAN, K. S. (1989): “A note on the geometric ergodicity of a Markov chain,” *Advances in Applied Probability*, 21, 702–704.

- CHEN, G., AND M. CHEN (2001): “On a class of nonlinear AR(p) models with nonlinear ARCH errors,” *Australian and New Zealand Journal of Statistics*, 43, 445–454.
- CHEN, R., AND R. S. TSAY (1993): “Functional-coefficient autoregressive models,” *Journal of the American Statistical Association*, 88, 298–308.
- CLINE, D. B. H. (2006): “Stability of nonlinear stochastic recursions with application to nonlinear AR–GARCH models,” unpublished manuscript.
- CLINE, D. B. H., AND H. H. PU (1998): “Verifying irreducibility and continuity of a nonlinear time series,” *Statistics and Probability Letters*, 40, 139–148.
- CLINE, D. B. H., AND H. H. PU (1999): “Geometric ergodicity of nonlinear time series,” *Statistica Sinica*, 9, 1103–1118.
- CLINE, D. B. H., AND H. H. PU (2004): “Stability and the Lyapounov exponent of threshold AR–ARCH models,” *Annals of Applied Probability*, 14, 1920–1149.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.
- FRANCQ, C., AND J.-M. ZAKOÏAN (2006): “Mixing properties of a general class of GARCH(1,1) models without moment assumptions,” *Econometric Theory*, 22, 815–834.
- GLOSTEN, L. W., R. JAGANATHAN, AND D. E. 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.
- GONZÁLEZ-RIVERA, G. (1998): “Smooth-transition GARCH models,” *Studies in Non-linear Dynamics and Econometrics*, 3, 61–78.
- HAGERUD, G. E. (1996): “A smooth transition ARCH model for asset returns,” Working Paper Series in Economics and Finance No. 162, Stockholm School of Economics.
- HORN, R. A., AND C. R. JOHNSON (1985): *Matrix Analysis*. Cambridge University Press, Cambridge.

- HWANG, S. Y., AND M.-J. WOO (2001): “Threshold ARCH(1) processes: asymptotic inference,” *Statistics and Probability Letters*, 53, 11–20.
- LANNE, M. AND P. SAIKKONEN (2005): “Nonlinear GARCH models for highly persistent volatility,” *Econometrics Journal*, 8, 251–276.
- LEE, C. (1998): “Asymptotics of a class of p th-order nonlinear autoregressive processes,” *Statistics and Probability Letters*, 40, 171–177.
- LIEBSCHER, E. (2005): “Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes,” *Journal of Time Series Analysis*, 26, 669–689.
- LING, S. (1999): “On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model,” *Journal of Applied Probability*, 36, 688–705.
- LING, S., AND M. MCALEER (2003): “Asymptotic theory for a vector ARMA–GARCH model,” *Econometric Theory*, 19, 280–310.
- LIU, J., W. K. LI, AND C. W. LI (1997): “On a threshold autoregression with conditional heteroscedastic variances,” *Journal of Statistical Planning and Inference*, 62, 279–300.
- LU, Z. (1998): “On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term,” *Statistica Sinica*, 8, 1205–1217.
- LU, Z., AND Z. JIANG (2001): “ L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term,” *Statistics and Probability Letters*, 51, 121–130.
- LUNDBERGH, S., AND T. TERÄSVIRTA (2002): “Evaluating GARCH models,” *Journal of Econometrics*, 110, 417–435.
- MASRY, E., AND D. TJØSTHEIM (1995): “Nonparametric estimation and identification of nonlinear ARCH time series,” *Econometric Theory*, 11, 258–289.
- MEITZ, M., AND P. SAIKKONEN (2006): “Stability of nonlinear AR–GARCH models,” SSE/EFI Working Paper Series in Economics and Finance No. 632.

- MEITZ, M., AND P. SAIKKONEN (2007): “Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models,” SSE/EFI Working Paper Series in Economics and Finance No. 573.
- MEYN, S. P., AND R. L. TWEEDIE (1993): *Markov Chains and Stochastic Stability*. Springer-Verlag, London.
- NELSON, D. (1990): “Stationarity and persistence in the GARCH(1,1) model,” *Econometric Theory*, 6, 318–334.
- SAIKKONEN, P. (2007): “Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity, ” *Statistica Sinica*, 17, 221–239.
- SIMMONS, G. F. (1963): *Introduction to Topology and Modern Analysis*. McGraw-Hill, New York.
- TERÄSVIRTA, T. (1994): “Specification, estimation, and evaluation of smooth transition autoregressive models,” *Journal of the American Statistical Association*, 89, 208–218.
- TJØSTHEIM, D. (1990): “Non-linear time series and Markov chains,” *Advances in Applied Probability*, 22, 587–611.
- VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): “Smooth transition autoregressive models — A survey of recent developments,” *Econometric Reviews*, 21, 1–47.