TÜSİAD-KOÇ UNIVERSITY ECONOMIC RESEARCH FORUM
WORKING PAPER SERIES

A NOTE ON THE GEOMETRIC ERGODICITY OF A NONLINEAR AR–ARCH MODEL

Mika Meitz
Pentti Saikkonen

Working Paper 1003
January 2010

TÜSİAD-KOÇ UNIVERSITY ECONOMIC RESEARCH FORUM
Rumeli Feneri Yolu 34450 Sarıyer/Istanbul

A note on the geometric ergodicity of a nonlinear AR–ARCH model*

Mika Meitz
Koç University

Pentti Saikkonen
University of Helsinki

First version January 2009, revised December 2009

Abstract

This note studies the geometric ergodicity of nonlinear autoregressive models with conditionally heteroskedastic errors. A nonlinear autoregression of order p (AR(p)) with the conditional variance specified as the conventional linear autoregressive conditional heteroskedasticity model of order q (ARCH(q)) is considered. Conditions under which the Markov chain representation of this nonlinear AR–ARCH model is geometrically ergodic and has moments of known order are provided. The obtained results complement those of Liebscher [Journal of Time Series Analysis, 26 (2005), 669–689] by showing how his approach based on the concept of the joint spectral radius of a set of matrices can be extended to establish geometric ergodicity in nonlinear autoregressions with conventional ARCH(q) errors.

*The second author acknowledges financial support from the Academy of Finland and the Okobank Group Research Foundation. We thank a co-editor and an anonymous referee for helpful comments and suggestions. Address correspondence to: Mika Meitz, Department of Economics, Koç University, Rumelifeneri Yolu, 34450 Sariyer, Istanbul, Turkey; e-mail: mmeitz@ku.edu.tr; 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 note is concerned with the geometric ergodicity of nonlinear autoregressive models with conditionally heteroskedastic errors. We consider a nonlinear autoregression of order p (AR(p)) with the conditional variance specified as the conventional linear autoregressive conditional heteroskedasticity model of order q (ARCH(q)). We give conditions under which the Markov chain associated with this nonlinear AR–ARCH model is geometrically ergodic (or, more precisely, Q–geometrically ergodic in a sense to be defined in Section 3) and has moments of known order. Our study makes heavy use of the stability theory developed for Markov chains, and we refer the reader to Meyn and Tweedie (1993) for a comprehensive account of the needed Markov chain theory.

Stability of conditionally heteroskedastic nonlinear autoregressions has previously been studied by several authors. Masry and Tjøstheim (1995), Lu (1998), Chen and Chen (2001), and Lu and Jiang (2001), among others, have provided sufficient conditions for geometric ergodicity in models similar to ours. In these papers, the proof of geometric ergodicity essentially assumes that (i) the conditional mean is dominated by a linear autoregression as the values of the observed process approach infinity and that (ii) this linear autoregression is stable in the sense of having a companion matrix whose spectral radius is less than one. However, such conditions may unnecessarily restrict the types of nonlinearity allowed and lead to overly restrictive regions of the parameter space ensuring ergodicity.

In a series of papers, Cline and Pu (1998, 1999, 2004) and Cline (2007) have used a different approach to establish geometric ergodicity in nonlinear conditionally heteroskedastic autoregressions. Based on the concept of the Lyapunov exponent they obtain conditions that often ensure geometric ergodicity in much larger regions of the parameter space than obtained in the abovementioned references. These conditions are sharp but the assumptions employed are quite general and appear difficult to verify.

In a recent paper, Liebscher (2005) takes yet another approach and employs the concept of the joint spectral radius of a set of matrices (to be defined in Section 2) to deduce geometric ergodicity in nonlinear conditionally heteroskedastic autoregressions. As his results show, this approach also ensures geometric ergodicity in larger regions of the parameter space than obtained by Masry and Tjøstheim (1995), Lu (1998), Chen and Chen (2001), and Lu and Jiang (2001). However, in the case of a general nonlinear autoregressive model, Liebscher's results only allow for limited forms of conditional heteroskedasticity. In particular, nonlinear autoregressions with conventional ARCH errors are ruled out.

The purpose of this note is to complement Liebscher's (2005) results and show how his approach based on the joint spectral radius can be extended to obtain sharpened conditions for geometric ergodicity in a nonlinear autoregression with errors following a standard linear ARCH(q) process. In earlier work (Meitz and Saikkonen, 2008) we obtained similar results when the errors of the autoregression follow a nonlinear first-order generalized ARCH (GARCH(1,1)) process. It should be noted, however, that these results do not directly extend to the higher-order ARCH(q) case, although the method of proof is similar.

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 the proofs are given in Section 4.

2 Model

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

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

where h_t is a positive function of y_s, $s < t$, and ε_t is a sequence of independent and identically distributed 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. When $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = 1$ one can interpret $f(y_{t-1}, \ldots, y_{t-p})$ and h_t as the conditional mean and conditional variance of y_t, respectively. For convenience, we will use this standard terminology although neither $E[\varepsilon_t] = 0$ nor $E[\varepsilon_t^2] = 1$ (or even existence of these moments) is necessary for our results to hold.

The function f describing the conditional mean is supposed to be of the form

$$f(x) = a(x)' x + b(x), \quad x \in \mathbb{R}^p, \tag{2}$$

where the functions $a : \mathbb{R}^p \to \mathbb{R}^p$ and $b : \mathbb{R}^p \to \mathbb{R}$ are bounded and Borel measurable. This assumption restricts the nonlinearity permitted in the conditional expectation but still covers several popular cases. In particular, the general functional-coefficient autoregressive model of Chen and Tsay (1993) and its special cases such as threshold autoregressive models (see, e.g., Tong (1990)) and smooth transition autoregressive models (see, e.g., Teräsvirta (1994)) are included. The boundedness requirement imposed on the function b

is somewhat stronger than required in Theorem 3 of Liebscher (2005) where $b(x) = o(\|x\|)$ as $\|x\| \to \infty$ is only assumed. It seems difficult to allow for this extension in our context. The reason for this is that in our proof of geometric ergodicity, we are forced to rely on an m-step-ahead drift criterion instead of a more conventional one-step-ahead criterion (a more detailed explanation of the arising difficulties can be found at the end of the paper following the proof of Theorem 1).

We assume that the conditional variance h_t is generated by a standard ARCH process driven by regression errors. Specifically,

$$h_t = \omega + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2, \tag{3}$$

where $\omega > 0$, $\alpha_j \geq 0$ $(j = 1, \ldots, q)$, and

$$u_t = y_t - f(y_{t-1}, \ldots, y_{t-p}). \tag{4}$$

Clearly, h_{t+1} is a function of the random vector $Z_t = [y_t \ \cdots \ y_{t-p-q+1}]'$, and we express this as $h_t = h(Z_{t-1})$. This conventional ARCH model for the conditional variance was ruled out in Theorem 3 of Liebscher (2005) where only weaker forms of conditional heteroskedasticity satisfying $h^{1/2}(z) = o(\|z\|)$ as $\|z\| \to \infty$ were allowed for. On the other hand, his condition permits limited forms of nonlinearity ruled out in our model. (In his Theorem 4, Liebscher (2005) makes a milder assumption about the conditional variance which also covers our ARCH model (3), but this is made at the cost of considerably restricting the nonlinearity permitted in the conditional expectation.)

From the definition of u_t it is readily seen that $Z_t = [y_t \ \cdots \ y_{t-p-q+1}]'$ is a Markov chain on $\mathcal{Z} = \mathbb{R}^{p+q}$. To make the Markov chain representation of Z_t explicit observe that

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

or, more briefly,

$$Z_t = F(Z_{t-1}, \varepsilon_t), \quad t = 1, 2, \ldots, \tag{6}$$

where the function $F : \mathbb{R}^{p+q+1} \to \mathbb{R}^{p+q}$ is defined in an obvious way.

We now discuss assumptions on the error term ε_t, the conditional mean function f, and the conditional variance h_t that are used to prove our results. In what follows, we shall always assume that the process y_t is defined by (1) with the function f given by (2) and h_t given by (3) and (4). Our first assumption concerns the error term ε_t.

Assumption 1. *The independent and identically distributed random variables ε_t have a (Lebesgue) density which is bounded away from zero on compact subsets of $\mathbb{R}$. Furthermore, for some real $r > 0$, $E[|\varepsilon_t|^{2r}] < \infty$.*

The first part of Assumption 1 ensures that Z_t in (6) is an irreducible and aperiodic T–chain (see Meyn and Tweedie (1993) for the definitions of these concepts). As (2) and (3) are assumed, this can be seen as in Example 2.1 of Cline and Pu (1998). The latter part of Assumption 1 requires the error term to have a finite moment of some (small) order. This is needed to apply a drift criterion in the proof of Theorem 1 below, and it also ensures that y_t and h_t have finite moments of some (small) order. Note that this assumption is weaker than in some of the related previous work (see Masry and Tjøstheim (1995), Lu (1998), Chen and Chen (2001), Lu and Jiang (2001), and Liebscher (2005)) where at least existence of a finite expectation is assumed. On the other hand, it coincides with the assumption used in, for example, Cline and Pu (2004).

To present our assumption restricting the conditional mean, set $a(x) = [a_1(x) \ \cdots \ a_p(x)]'$ ($x \in \mathbb{R}^p$) and define the companion matrix

$$A(x) = \begin{bmatrix} a_1(x) & \cdots & a_p(x) & 0 & \cdots & 0 & 0 \\ 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \qquad ((p+q) \times (p+q)).$$

By $A_1(x)$ we denote the matrix obtained by deleting the last q rows and columns from $A(x)$. Then, equation (5) can be expressed as

$$Z_t = A(S'Z_{t-1}) Z_{t-1} + \iota_{p+q} b(S'Z_{t-1}) + \iota_{p+q} h(Z_{t-1})^{1/2} \varepsilon_t, \tag{7}$$

where $S' = [I_p : 0]$ ($p \times (p+q)$) and $\iota_{p+q} = [1 \ 0 \ \cdots \ 0]'$ ($(p+q) \times 1$). Following Liebscher (2005) we restrict the matrix $A(x)$ by using the concept of the joint spectral radius of a (bounded) set of (square) matrices. To introduce this concept, let $\mathcal{A}$ be a set of bounded square matrices and $\mathcal{A}^k = \{A_1 A_2 \cdots A_k : A_i \in \mathcal{A},\ i = 1, \ldots, k\}$. Then the joint spectral radius of the set $\mathcal{A}$ is defined by

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

where $\|\cdot\|$ 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.

Now we can state our next assumption that restricts the conditional mean.

Assumption 2. $\rho(\mathcal{A}_1) < 1$, *where* $\mathcal{A}_1 = \{A_1(x) : x \in \mathbb{R}^p\}$.

It is straightforward to see that this assumption is equivalent to $\rho(\mathcal{A}_*) < 1$ where $\mathcal{A}_* = \{A(x) : x \in \mathbb{R}^p\}$ (see Lemma 1(i) of Meitz and Saikkonen (2008)). In the proofs of the paper we use the joint spectral radius $\rho(\mathcal{A}_*)$ but $\rho(\mathcal{A}_1)$ is more convenient in practice because, due to a smaller dimension, its value is easier to compute than that of $\rho(\mathcal{A}_*)$. Because $\mathcal{A}_*$ is a bounded set of matrices Assumption 2 implies that 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 $0 < \rho < 1$ (see Theorem 1 of Liebscher (2005)).

To present our assumption restricting the conditional variance, define the vector $X_t = \begin{bmatrix} h_t \; u_{t-1}^2 & \cdots & u_{t-q+1}^2 \end{bmatrix}'$. For $t = 1$, X_t is determined by the initial values of the Markov chain Z_t, that is, $X_1 = \begin{bmatrix} h(Z_0) & u_0^2 & \cdots & u_{-q+1}^2 \end{bmatrix}'$ where $u_0^2, \ldots, u_{-q+1}^2$ depend on $Z_0 = [y_0 \ \cdots \ y_{-p-q+1}]'$ (see (4)). For larger values of t, we have

$$\begin{bmatrix} h_t \\ u_{t-1}^2 \\ \vdots \\ u_{t-q+1}^2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \varepsilon_{t-1}^2 & \alpha_2 & \cdots & \alpha_{q-1} & \alpha_q \\ \varepsilon_{t-1}^2 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} h_{t-1} \\ u_{t-2}^2 \\ \vdots \\ u_{t-q}^2 \end{bmatrix} + \begin{bmatrix} \omega \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad t = 2, 3, \ldots,$$

or

$$X_t = \Lambda_{t-1} X_{t-1} + c, \quad t = 2, 3, \ldots \tag{8}$$

with the initial value X_1 as described above and Λ_t being a sequence of independent and identically distributed matrices. Because X_t is a function of Z_{t-1} we can write $X_t = G(Z_{t-1})$.

Our last assumption restricts the conditional variance process via the matrices Λ_t.

Assumption 3. *There exists an induced matrix norm* $\|\cdot\|^\bullet$ *such that* $E[\|\Lambda_t\|^{\bullet r}] < 1$, *where* $r > 0$ *is as in Assumption 1.*

This assumption is formulated in a way which is convenient in the proofs but, in general, is not easy to check in practice. However, the usual conditions for covariance stationarity of an ARCH(q) model, $\sum_{j=1}^{q} \alpha_j < 1$ and $E[\varepsilon_t^2] = 1$, imply Assumption 3 with $r = 1$. This can be seen by using arguments similar to those in the proof of Lemma 1(ii) of Meitz and Saikkonen (2008); for details, see Section 4. Moreover, it can be shown that Assumption 3 holds with some (unknown) $r > 0$ if $E[\ln \|\Lambda_t\|^{\bullet}] < 0$ and $E[\|\Lambda_t\|^{\bullet s}] < \infty$ for some $s > 0$ (see Remark 2.9 of Basrak, Davis, and Mikosch (2002)). On the other hand, if Assumption 3 holds, then $E[\ln \|\Lambda_t\|^{\bullet}] < 0$ as can be seen using Jensen's inequality.

It is of interest to compare our assumptions on conditional heteroskedasticity with those in Liebscher's (2005) Theorem 3 where ARCH models are not allowed. One difference is that we explicitly assume the process for conditional heteroskedasticity to be driven by regression errors whereas Liebscher (2005) is more general in this respect. This difference is reflected in our method of proof which differs from that used by Liebscher (2005). In our proof, the structure of the conventional ARCH(q) process and equation (8) combined with Assumption 3 make it possible to establish geometric ergodicity.

3 Result

We will show that the Markov chain Z_t defined in (6) is Q–geometrically ergodic. This type of geometric ergodicity was defined and employed by Liebscher (2005) and further applied by Meitz and Saikkonen (2008). For convenience, we repeat the definition here in the form given in the latter paper. 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 $\mathcal{Z}$.

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

$$\sup_{v:|v|\leq 1} \left| \int_{\mathcal{Z}} P^n(z, dw)\, v(w) - \int_{\mathcal{Z}} \pi(dw) v(w) \right| \leq (a + bQ(z))\, \varrho^n \quad \text{for all } z \in \mathcal{Z} \text{ and all } n \geq 1. \tag{9}$$

Q–geometric ergodicity implies the existence of an initial value Z_0 which makes Z_t a stationary process such that $Q(Z_t)$ has finite expectation (for this and other implications of Q–geometric ergodicity, see Liebscher (2005) and Meitz and Saikkonen (2008)). Furthermore, for any initial value with a distribution such that $Q(Z_0)$ has finite expectation, Z_t is β–mixing (absolutely regular), implying that usual limit theorems hold.

Now we can state our main result.

Theorem 1. *Suppose that Assumptions 1, 2, and 3 hold, and let $\|\cdot\|_{(p+q)}$ and $\|\cdot\|_{(q)}$ be any vector norms on $\mathbb{R}^{p+q}$ and $\mathbb{R}^q$, respectively. Then the Markov chain Z_t on $\mathcal{Z}$ is Q^*–geometrically ergodic in the sense of Definition 1 with a function $Q^*(z) \geq 1 + \|z\|_{(p+q)}^{2r} + \|G(z)\|_{(q)}^r$.*

Thus, Theorem 1 shows that the Markov chain Z_t is Q^*–geometrically ergodic with a function $Q^*(\cdot)$ such that the stationary distribution of Z_t has moments of order $2r$. Moreover, as seen in the proof of the theorem, $h^r(z) \leq C \|G(z)\|^{\bullet r}$, $C < \infty$, so that we can also conclude that in the stationary case the conditional variance process h_t has moments of order r.

Theorem 1 also demonstrates how Liebscher's (2005) approach based on the joint spectral radius can be used to prove Q–geometric ergodicity in a nonlinear autoregressive model with conventional ARCH errors. Thus, we are able to extend the scope of Liebscher's (2005) Theorem 3. This is achieved at the cost of only a moderate strengthening of the nonlinearity in the conditional expectation and by ruling out only very weak forms of nonlinearity in the conditional variance. On the other hand, our Theorem 1 also applies in the case when only moments of some small order exist. Compared with Liebscher's (2005) Theorem 4, our assumptions on the conditional variance are only moderately more stringent, although the nonlinearity we can permit in the conditional expectation is considerably stronger. It may also be noted that to prove his Theorem 4 Liebscher (2005) does not need the concept of joint spectral radius because the ordinary spectral radius works as well.

4 Proofs

Proof that $\sum_{j=1}^q \alpha_j < 1$ and $E[\varepsilon_t^2] = 1$ imply the validity of Assumption 3 with $r = 1$. First note that the assumption $\sum_{j=1}^q \alpha_j < 1$ is equivalent to $\rho(\Lambda) < 1$, where

$$\Lambda \overset{def}{=} E[\Lambda_t] = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{q-1} & \alpha_q \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Now, as in the proof of Lemma 1(ii) of Meitz and Saikkonen (2008), we can find a $q \times 1$ vector κ with positive components such that the components of the row vector

$\nu' = \kappa'(I_q - \Lambda)$ are positive and, furthermore, $0 < \underline{\nu}/\overline{\kappa} < 1$ where $\underline{\nu}$ and $\overline{\kappa}$ are the smallest and largest components of ν and κ, respectively. Define the vector norm $\|\cdot\|^{\bullet}$ in $\mathbb{R}^q$ by

$$\|x\|^{\bullet} = \sum_{j=1}^{q} \kappa_j |x_j| = \kappa' |x|, \text{ where } \quad |x| = [|x_1| \ \cdots \ |x_q|]',$$

and consider the random matrix Λ_t and any $x \in \mathbb{R}^q$, $x \neq 0$. With probability one,

$$\|\Lambda_t x\|^{\bullet} \leq \kappa_1 \alpha_1 \varepsilon_t^2 |x_1| + \kappa_1 \sum_{j=2}^{q} \alpha_j |x_j| + \kappa_2 \varepsilon_t^2 |x_1| + \sum_{j=3}^{q} \kappa_j |x_{j-1}|,$$

and hence

$$E[\|\Lambda_t x\|^{\bullet}] \leq \kappa_1 \sum_{j=1}^{q} \alpha_j |x_j| + \sum_{j=2}^{q} \kappa_j |x_{j-1}| = \kappa' |x| \left(1 - \frac{\nu' |x|}{\kappa' |x|}\right) \leq \|x\|^{\bullet} (1 - \underline{\nu}/\overline{\kappa}).$$

Because $0 < 1 - \underline{\nu}/\overline{\kappa} < 1$, this shows that Assumption 3 holds with the matrix norm induced by $\|\cdot\|^{\bullet}$ and $r = 1$. ■

Proof of Theorem 1. First note that, as discussed after Assumption 1, Z_t is an irreducible and aperiodic T–chain. Let $\|\cdot\|_{(p+q)}$ and $\|\cdot\|_{(q)}$ be any vector norms on $\mathbb{R}^{p+q}$ and $\mathbb{R}^q$, respectively, 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 the discussion following Assumption 2). Let $\|\cdot\|^{\bullet}$ be an induced matrix norm satisfying Assumption 3. By the equivalence of all vector norms in finite-dimensional vector spaces, there exist finite $C_1, C_2 > 0$ such that $\|z\|_{(p+q)} \leq C_1^{1/2r} \|z\|^*$ for all $z \in \mathbb{R}^{p+q}$ and $\|x\|_{(q)} \leq C_2^{1/r} \|x\|^{\bullet}$ for all $x \in \mathbb{R}^q$ (see e.g. Horn and Johnson (1985, Sec. 5.4)). Denote $V_*(z) = 1 + C_1 \|z\|^{*2r} + C_2 \|G(z)\|^{\bullet r}$. As in Lemma 5 of Meitz and Saikkonen (2008) the idea is to examine the conditional expectation $E[V_*(Z_t) \mid Z_{t-m} = z]$ and demonstrate that condition (19.15) of Meyn and Tweedie (1993) holds for the function $V_*(z)$ (with the choice $n(z) \equiv m$) after which an application of Lemma 6 of Meitz and Saikkonen (2008) shows that Z_t is V_*–geometrically ergodic in the sense of Definition 1.

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

$$\left(\sum_{i=1}^{n} x_i\right)^r \leq \Delta_{r,n} \sum_{i=1}^{n} x_i^r \tag{10}$$

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

Using (8) and repeated substitution one obtains, for $m, k \geq 1$,

$$G(Z_{t-m+k}) = \prod_{j=1}^{k} \Lambda_{t-m+j} G(Z_{t-m}) + \left(I_q + \sum_{j=0}^{k-2} \prod_{i=0}^{j} \Lambda_{t-m+k-i}\right) c,$$

which in conjunction with usual properties of vector and matrix norms and (10) gives

$$\Delta_{r,k+1}^{-1}\left\|G\left(Z_{t-m+k}\right)\right\|^{\bullet r}\leq\prod_{j=1}^{k}\left\|\Lambda_{t-m+j}\right\|^{\bullet r}\left\|G(Z_{t-m})\right\|^{\bullet r}+\left(\left\|I_q\right\|^{\bullet r}+\sum_{j=0}^{k-2}\prod_{i=0}^{j}\left\|\Lambda_{t-m+k-i}\right\|^{\bullet r}\right)\left\|c\right\|^{\bullet r}.$$

By Assumption 3, $E[\|\Lambda_t\|^{\bullet r}]<1$ and we denote this expectation by δ. Furthermore, denote $d=\left(\|I_q\|^{\bullet r}+\delta/(1-\delta)\right)\|c\|^{\bullet r}$. By the independence of the Λ_t's,

$$\begin{aligned}\Delta_{r,k+1}^{-1}E\left[\left\|G\left(Z_{t-m+k}\right)\right\|^{\bullet r}\mid Z_{t-m}=z\right] &\leq \|G(z)\|^{\bullet r}\delta^k+\left(\|I_q\|^{\bullet r}+\sum_{j=0}^{k-2}\delta^{j+1}\right)\|c\|^{\bullet r}\\ &\leq \|G(z)\|^{\bullet r}\delta^k+d. \qquad (11)\end{aligned}$$

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

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

Denote $\iota_q=[1\ 0\ \cdots\ 0]'$ $(q\times 1)$ and $\|\iota_q\|^{\bullet r}=\iota_q^{\bullet}$. Then, as $h^r(Z_{t-m+k})=\left|\iota_q' G\left(Z_{t-m+k}\right)\right|^r$, we also have

$$\begin{aligned}E\left[h^r(Z_{t-m+k})\mid Z_{t-m}=z\right] &\leq \|\iota_q\|^{\bullet r}E\left[\left\|G\left(Z_{t-m+k}\right)\right\|^{\bullet r}\mid Z_{t-m}=z\right]\\ &\leq \iota_q^{\bullet}\Delta_{r,m}\|G(z)\|^{\bullet r}\delta^k+\iota_q^{\bullet}d'. \qquad (12)\end{aligned}$$

Now consider Z_t which we wish to express in terms of past values of the process Z_t until $t-m$. Repeated substitution in equation (7), usual properties of vector and matrix norms, and an application of (10) yield (cf. the proof of Lemma 5 of Meitz and Saikkonen (2008), the paragraph following inequality (15) with Y_t therein replaced by Z_t)

$$\begin{aligned}\Delta_{2r,2m+1}^{-1}\|Z_t\|^{*2r}\leq&\prod_{j=0}^{m-1}\|A(S'Z_{t-1-j})\|^{*2r}\|Z_{t-m}\|^{*2r}+\|\iota_{p+q}b\left(S'Z_{t-1}\right)\|^{*2r}\\ &+\sum_{j=0}^{m-2}\prod_{i=0}^{j}\|A\left(S'Z_{t-1-i}\right)\|^{*2r}\|\iota_{p+q}b\left(S'Z_{t-2-j}\right)\|^{*2r}+\|\iota_{p+q}h\left(Z_{t-1}\right)^{1/2}\varepsilon_t\|^{*2r}\\ &+\sum_{j=0}^{m-2}\prod_{i=0}^{j}\|A\left(S'Z_{t-1-i}\right)\|^{*2r}\|\iota_{p+q}h\left(Z_{t-2-j}\right)^{1/2}\varepsilon_{t-1-j}\|^{*2r}. \qquad (13)\end{aligned}$$

Denote $\|\iota_{p+q}\|^{*2r}=\iota_{p+q}^{*}$ and note that $\|A\left(\cdot\right)\|^{*2r}\leq\rho^{2r}$, $\|\iota_{p+q}b\left(\cdot\right)\|^{*2r}\leq\iota_{p+q}^{*}B$ for some finite B (because $b\left(\cdot\right)$ is bounded), $\|\iota_{p+q}h\left(\cdot\right)^{1/2}\varepsilon_t\|^{*2r}\leq\iota_{p+q}^{*}h^r\left(\cdot\right)|\varepsilon_t|^{2r}$, and $E[|\varepsilon_t|^{2r}]\stackrel{def}{=}$

$\gamma_{2r} < \infty$. Thus,

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

where the last inequality makes use of (12) and the fact that $E\left[h^r(Z_{t-m}) \mid Z_{t-m} = z\right] = h^r(z) = \left|\iota_q' G(z)\right|^r \leq \iota_q^{\bullet} \|G(z)\|^{\bullet r}$. Defining $\phi = \max\{\rho^{2r}, \delta\} < 1$ and $\phi' = \frac{1}{1-\phi}$ we get

$$
\begin{aligned}
& \Delta_{2r,2m+1}^{-1} E\left[\|Z_t\|^{*2r} | Z_{t-m} = z\right] \\
\leq\ & \phi^m \|z\|^{*2r} + \iota_{p+q}^{*} B \left(1 + \sum_{j=0}^{m-2} \phi^{j+1}\right) + \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} \left(\Delta_{r,m} \phi^{m-1} \|G(z)\|^{\bullet r} + d'\right) \\
& + \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} \left(\sum_{j=0}^{m-3} \phi^{j+1} \left(\Delta_{r,m} \phi^{m-2-j} \|G(z)\|^{\bullet r} + d'\right) + \phi^{m-1} \|G(z)\|^{\bullet r}\right) \\
\leq\ & \phi^m \|z\|^{*2r} + \iota_{p+q}^{*} B \phi' + \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} \left(\Delta_{r,m} \phi^{m-1} \|G(z)\|^{\bullet r} + d'\right) \\
& + \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} \left(\sum_{j=0}^{m-3} \phi^{m-1} \Delta_{r,m} \|G(z)\|^{\bullet r} + \sum_{j=0}^{m-3} \phi^{j+1} d' + \Delta_{r,m} \phi^{m-1} \|G(z)\|^{\bullet r}\right) \\
\leq\ & \phi^m \|z\|^{*2r} + m \cdot \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} \Delta_{r,m} \phi^{m-1} \|G(z)\|^{\bullet r} + \iota_{p+q}^{*} \phi' (B + \iota_q^{\bullet} \gamma_{2r} d'). \qquad (14)
\end{aligned}
$$

Combining the inequalities (11) (with $k = m$) and (14) yields

$$
\begin{aligned}
& E\left[V_*(Z_t) \mid Z_{t-m} = z\right] \\
=\ & E\left[1 + C_1 \|Z_t\|^{*2r} + C_2 \|G(Z_t)\|^{\bullet r} \mid Z_{t-m} = z\right] \\
\leq\ & 1 + C_1 \Delta_{2r,2m+1} \left(\phi^m \|z\|^{*2r} + \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} m \Delta_{r,m} \phi^{m-1} \|G(z)\|^{\bullet r} + \iota_{p+q}^{*} \phi' (B + \iota_q^{\bullet} \gamma_{2r} d')\right) \\
& + C_2 \Delta_{r,m+1} \left(\|G(z)\|^{\bullet r} \delta^m + d\right) \\
=\ & 1 + C_1 \left[\Delta_{2r,2m+1} \phi^m\right] \|z\|^{*2r} + C_2 \left[C_1 C_2^{-1} \iota_{p+q}^{*} \iota_q^{\bullet} \gamma_{2r} \Delta_{2r,2m+1} m \Delta_{r,m} \phi^{m-1} + \Delta_{r,m+1} \delta^m\right] \|G(z)\|^{\bullet r} \\
& + \left\{C_1 \Delta_{2r,2m+1} \iota_{p+q}^{*} \phi' (B + \iota_q^{\bullet} \gamma_{2r} d') + C_2 \Delta_{r,m+1} d\right\}. \qquad (15)
\end{aligned}
$$

Because $0 < \delta \leq \phi < 1$, it follows from the definitions that we can choose an m large enough so that both of the expressions in square brackets in (15) are smaller than some

$\lambda < 1$. The expression in curly brackets in (15) is clearly finite, and thus for some $L < \infty$

$$E\left[V_*(Z_t) \mid Z_{t-m} = z\right] \leq \lambda \left(1 + C_1 \|z\|^{*2r} + C_2 \left\|G(z)\right\|^{\bullet r}\right) + L. \tag{16}$$

What remains to be examined is the behavior of (16) on and off a small set. To this end, write the right-hand-side of (16) as

$$\lambda^{1/2} \left(1 + C_1 \|z\|^{*2r} + C_2 \left\|G(z)\right\|^{\bullet r}\right) \cdot \lambda^{1/2} \left(1 + \frac{L}{\lambda \left(1 + C_1 \|z\|^{*2r} + C_2 \left\|G(z)\right\|^{\bullet r}\right)}\right). \tag{17}$$

We shall show below that the set $A_N = \{z \in \mathcal{Z} : \|z\|^{*2r} \leq N, \|G(z)\|^{\bullet r} \leq N\}$ is small for any N so large that A_N is nonempty (see (3)). Off this set either $\|z\|^{*2r} > N$ or $\|G(z)\|^{\bullet r} > N$, and the ratio in (17) can clearly be made arbitrarily small by choosing N large enough. Therefore for a large enough N

$$\lambda^{1/2} \left(1 + \frac{L}{\lambda \left(1 + C_1 \|z\|^{*2r} + C_2 \left\|G(z)\right\|^{\bullet r}\right)}\right) < 1$$

and hence

$$E\left[V_*(Z_t) \mid Z_{t-m} = z\right] \leq \lambda^{1/2} \left(1 + C_1 \|z\|^{*2r} + C_2 \left\|G(z)\right\|^{\bullet r}\right)$$

off the set A_N. On the other hand, the right hand side of (16) is clearly bounded on the set A_N. Therefore, condition (19.15) of Meyn and Tweedie (1993) is satisfied and it only remains to be shown that the set A_N is small.

To show that the set A_N is small we can use arguments similar to those in Lemma 4 of Meitz and Saikkonen (2008). We present the details for completeness. Using (7) and (10),

$$\Delta_{2r,3}^{-1} E\left[\|Z_t\|^{*2r} \mid Z_{t-1} = z\right] \leq \left\|A\left(S'z\right)\right\|^{*2r} \|z\|^{*2r} + \iota_{p+q}^{*} \left|b\left(S'z\right)\right|^{2r} + \iota_{p+q}^{*} h\left(z\right)^{r} E\left[|\varepsilon_t|^{2r}\right],$$

where the majorant side is bounded on the set A_N (recall that $A(\cdot)$ and $b(\cdot)$ are bounded and $h(z) \leq \|\iota_q\|^{\bullet} \|G(z)\|^{\bullet}$). Therefore we can find an $M_N < \infty$ such that

$$\sup_{z \in A_N} E\left[\|Z_t\|^{*2r} \mid Z_{t-1} = z\right] < M_N^{2r}. \tag{18}$$

Now define the compact set $B_N = \{z \in \mathcal{Z} : \|z\|^{*} \leq M_N\}$. Because Z_t is an irreducible and aperiodic T–chain this set is small and

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

Here the first inequality is Markov's and the second one is due to (18). That the set A_N is small can now be concluded from Proposition 5.2.4 of Meyn and Tweedie (1993). ■

We now briefly discuss why we assume the function $b(\cdot)$ in equation (2) to be bounded instead of the often used weaker condition $b(x) = o(\|x\|)$ as $\|x\| \to \infty$. The difficulty in using this weaker condition in our proof arises from the fact that we are forced to rely on an m-step-ahead drift criterion (instead of a more conventional one-step-ahead criterion) to prove geometric ergodicity. This leads us to examine the conditional expectation $E[V_*(Z_t) \mid Z_{t-m} = z]$ on and off the small set A_N that restricts the values of z $(= Z_{t-m})$. In inequality (13) we obtain an upper bound for the term $\|Z_t\|^{*2r}$, which forms part of the function $V_*(Z_t)$. The upper bound contains the terms $\|\iota_{p+q} b(S'Z_{t-j})\|^{*2r}$, $j = 1, ..., m-1$, and the difficulty is how to control the conditional expectations of these terms when the conditioning only restricts the values of z $(= Z_{t-m})$ but not those of $Z_{t-1}, \ldots, Z_{t-m+1}$. Our solution is to restrict the function $b(\cdot)$ uniformly over its domain by requiring it to be bounded. Note that if the use of a standard one-step-ahead drift criterion had sufficed in our proof, this problem would not have arised at all.

References

Basrak, B., R. A. Davis, and T. Mikosch (2002): "Regular variation of GARCH processes," *Stochastic Processes and their Applications*, 99, 95–115.

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. (2007): "Evaluating the Lyapounov exponent and existence of moments for threshold AR–ARCH models," *Journal of Time Series Analysis*, 28, 241–260.

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.

Horn, R. A., and C. R. Johnson (1985): *Matrix Analysis*. Cambridge University Press, Cambridge.

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.

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.

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 (2008): "Stability of nonlinear AR–GARCH models," *Journal of Time Series Analysis*, 29, 453–475.

Meyn, S. P., and R. L. Tweedie (1993): *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

Teräsvirta, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208–218.

Tong, H. (1990): *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.