Stability of nonlinear AR–GARCH models*

Mika Meitz
Stockholm School of Economics

Pentti Saikkonen
University of Helsinki

SSE/EFI Working Paper Series in Economics and Finance No. 632

June 1, 2006

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.

*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. Address correspondence to: Mika Meitz, Department of Economic Statistics, Stockholm School of Economics, P. O. Box 6501, SE–113 83 Stockholm, Sweden; e-mail: mika.meitz@hhs.se; 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 (AR(p)) with the conditional variance specified as a nonlinear first order generalized autoregressive conditional heteroskedasticity (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), Lu (1998), Cline and Pu (1998), Cline and Pu (1999), Lu and Jiang (2001), Chen and Chen (2001), Saikkonen (2005), and Liebscher (2005). For related results for pure GARCH models, see Carrasco and Chen (2002), Meitz and Saikkonen (2005), and the references therein. To the best of our knowledge, this paper provides the first practically applicable stability results for nonlinear autoregressive models with GARCH errors.

A major difficulty in establishing geometric ergodicity in the present context is to prove irreducibility and aperiodicity of the relevant Markov chain, which is typically required as a first step in the proof of geometric ergodicity. This difficulty may actually explain the aforementioned lack of related previous results and 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 characterized by discontinuous functions (see, e.g., Tong (1990) and Chen and Tsay (1993) for models for conditional mean and Glosten, Jaganathan, and Runkle (1993), and Rabemananjara and Zakoïan (1993) for GARCH models). However, we are still able to cover a number of nonlinearities recently considered in both theoretical and applied literature.

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 quite often 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, $t = 1, 2, \ldots$, 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 (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,

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

where g is a function to be described shortly and

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

From the definition of u_t it is readily seen that $Z_t = [y_t \ \cdots \ y_{t-p} \ \ h_t]' \stackrel{def}{=} [Y_t' \ \ h_t]'$ 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

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

and observe that then we can write

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

or, more briefly,

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

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

We set $F_1 = F$ and, for $k \geq 1$, $F_{k+1}(z, e_1, \ldots, e_{k+1}) = F(F_k(z, e_1, \ldots, e_k), e_{k+1})$, where $z \in \mathbb{R}^{p+1}$ and $e_i \in \mathbb{R}$. Then, for any initial condition $Z_0 = z_0$ and any $k \geq 1$, $Z_k = F_k(z_0, \varepsilon_1, \ldots, \varepsilon_k)$. Following Meyn and Tweedie (1993) we call $\{e_i\}$ a control sequence and $z_k = F_k(z_0, e_1, \ldots, e_k)$ $(k = 1, 2, \ldots)$ 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 $r \geq 1$, $E\left[\varepsilon_t^{2r}\right] < \infty$.*

Assumption 2 *The function f is of the form*

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

where the functions $a : \mathbb{R}^p \to \mathbb{R}^p$ and $b : \mathbb{R}^p \to \mathbb{R}$ 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 this reason we also need similar smoothness assumptions for the function g used to model the conditional variance. These assumptions are satisfied by several well-known models but, as discussed in the introduction, they rule out threshold type nonlinearities.

For later purposes it will be convenient to introduce some notation. For any integer $p \geq 2$ and $x \in \mathbb{R}^p$ 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

$$A(x) = \bar{A}_{p+1}\left(\left[a(x)' \ \ 0\right]'\right) = \begin{bmatrix} a_1(x) & a_2(x) & \cdots & a_p(x) & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

With this notation the model takes the form

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

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 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.

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* $\lambda^p - \alpha_1 \lambda^{p-1} - \ldots - \alpha_p = 0$ *are inside the unit circle, where* $\alpha_j = \sup_{x\in\mathbb{R}^p} |a_j(x)|$ $(j = 1, \ldots, p)$.

As already indicated, Assumption 3 is needed to prove the geometric ergodicity of the Markov chain Z_t. A similar assumption based on the concept of joint spectral radius or Lemma 1(i) was recently used by Liebscher (2005) who established geometric ergodicity for various nonlinear autoregressive models. In these models conditional heteroskedasticity was also allowed but only of a limited nature. In particular, 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 $g : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ *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 $x \in \mathbb{R}_+$, $g(u, x) \to \infty$ *as* $u \to \infty$.

(c) There exists a real number $h^* \in \mathbb{R}_+$ *such that the sequence* h_k $(k = 1, 2, \ldots)$ *defined by* $h_k = g(0, h_{k-1})$, $k = 1, 2, \ldots$, *converges to* h^* *as* $k \to \infty$ *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 nonnegative real numbers a and c, and a Borel measurable function* $\varphi : \mathbb{R} \to \mathbb{R}_+$ *such that*

$$g(x^{1/2}\varepsilon_t, x) \le (a + \varphi(\varepsilon_t))\, x + c$$

for all $x \in \mathbb{R}_+$. *Furthermore,* $a + \varphi(0) < 1$ *and* $E[(a + \varphi(\varepsilon_t))^r] < 1$ *where the real number* $r \ge 1$ *is as in Assumption 1.*

As mentioned above, 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 (2005).

In Assumption 4(c) 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

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

(this follows from the contraction map principle, see for example 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. For instance, in some cases $g(u, x) \ge g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$ and it suffices that the function $g(0, x)$ is nondecreasing for $x \ge h^*$.

The latter condition of Assumption 4(c), that is, $g(u, x) \ge h^*$ for all $u \in \mathbb{R}$ and all $x \ge h^*$, combined with the other conditions of this assumption implies the convergence of the stated recursion. This can be seen as follows. Note first that from Assumption 4(d) it follows that $g(0, x) \le (a + \varphi(0))\, x + c \le (a + \varphi(0) + \epsilon)\, x$ for all x large enough and some $\epsilon > 0$ such that $a + \varphi(0) + \epsilon < 1$. From this and Assumption 4(a) it is straightforward to check 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, $k \ge 0$, is nonincreasing and bounded from below by h^*. Therefore it converges to, say, h_* $(\ge h^*)$ and, because $g(0, h_k) = h_{k+1}$, we also have $g(0, h_k) \to h_*$. On the other hand, by the continuity of $g(0, \cdot)$, $g(0, h_k) \to g(0, h_*)$. Thus we must have $g(0, h_*) = h_*$ and, since h^* is the maximal fixed point, $h_* = h^*$.

Our final assumption concerns the deterministic control model $z_k = F_k(z_0, e_1, \ldots, e_k)$ $(k = 1, 2, \ldots)$ 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 5 *For each initial value* $z_0 \in \mathcal{Z}$, *there exists a control sequence* $e_1^{(0)}, \ldots, e_{p+2}^{(0)}$ *such that the* $(p+2) \times (p+2)$ *matrix*

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

is nonsingular.

By Proposition 7.1.4 of Meyn and Tweedie (1993), this assumption implies that the deterministic control model $z_k = F_k(z_0, e_1, \ldots, e_k)$ is forward accessible. This property is needed to apply the results obtained in Chapter 7 of Meyn and Tweedie (1993). Note that although Assumption 5 is sufficient for forward accessibility it is not necessary, as the aforementioned proposition of Meyn and Tweedie (1993) shows.

Although Assumption 5 may look difficult to verify in practice that is fortunately not the case for several commonly used models. To get an idea of the structure of the derivative matrix $\nabla F_{p+2}^{(0)}$, denote the components of the vector $F_{p+2}(z_0, e_1, \ldots, e_{p+2})$ briefly by $y_{p+2}, \ldots, y_2$ 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 $h_i^{1/2} = h(z_{i-1})^{1/2} > 0$ $(i = 2, \ldots, p+2)$ 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 5 to hold it suffices to find $e_1^{(0)}, \ldots, e_{p+2}^{(0)}$ such that, for all initial values, $\partial h_{p+2}/\partial e_1$ is nonzero and $\partial h_{p+2}/\partial e_2, \ldots, \partial h_{p+2}/\partial e_{p+1}$ are zero when evaluated at $\begin{bmatrix} e_1^{(0)} & \cdots & e_{p+2}^{(0)} \end{bmatrix}'$. 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 turns out to be different.

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 5, 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, Assumption 4 only concerns the GARCH model specified for the error term in (1) and restricts it by conditions which are very similar to previous counterparts used in pure GARCH(1,1) models. As for Assumption 5, it concerns both the conditional mean and conditional variance but, as the examples of Section 4 show, this assumption is often easy to check 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 introduced by Liebscher (2005). 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 $P(z, A)$ will be used.)

Definition 1 *(Liebscher (2005)) 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*

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

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), Saikkonen (2005), Lanne and Saikkonen (2005), and Meitz and Saikkonen (2005)). However, we have found it difficult to establish it in the present context. Therefore, we use the weaker Q–geometric ergodicity which, as discussed above, 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, as already discussed, its application assumes the smoothness conditions imposed in Assumptions 2 and 4(a). We have the following lemma.

Lemma 2 *If Assumptions 1–4 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 $A_N = \{z \in Z : \|y\|^{2r} \leq N, h^r(z) \leq N\}$ is small for any vector norm and for all positive r and N such that $\underline{g}^r < N$, where $\underline{g}$ is as in Assumption 4(a).*

Thus, Lemma 1 provides the necessary prerequisites for the application of the drift criterion. Note that Lemma 1 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 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 Liebscher (2005) with a function $Q^*(z) \geq 1 + \|y\|^{2r} + h^r(z)$.*

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, respectively (the latter moment result follows because $h_t = h(Z_{t-1})$). An important consequence of Theorem 1 is that usual limit theorems apply. As far as we know, there is no equivalent to this result in the previous literature on nonlinear autoregressions with GARCH errors.

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 2–5 of which Assumptions 2 and 3 concern the conditional mean of the model, that is, the function f, whereas Assumption 4 restricts the form of permitted conditional heteroskedasticity. As already indicated, Assumption 5 can often be checked without paying attention to the conditional mean. In such cases it is only necessary to check conditions imposed on the conditional mean and conditional variance separately.

4.1 Conditional mean

First consider the conditional mean. A very general specification only assumes that the function f has the general structure imposed in Assumption 2. In this case, general sufficient conditions for Assumption 3 are obtained from Lemma 1. This approach is relevant for the general functional-coefficient autoregressive model of Chen and Tsay (1993). For more concrete examples, we have to be more specific about the function a in Assumption 2. For instance, suppose that

$$f\left(y_{t-1},\ldots,y_{t-p}\right)=\phi_0+\psi_0 G(y_{t-1},\ldots,y_{t-p})+\sum_{j=1}^{p}\left(\phi_j+\psi_j G(y_{t-1},\ldots,y_{t-p})\right)y_{t-j}, \qquad (9)$$

where $\phi_j,\psi_j\in\mathbb{R}$, $j=0,\ldots,p$, and G is a smooth function with range $[0,1]$. In this case Lemma 1(ii) shows that a sufficient condition for Assumption 3 is

$$\sum_{j=1}^{p}\max\left\{|\phi_j|,|\phi_j+\psi_j|\right\}<1, \qquad (10)$$

a condition previously obtained by Chen and Tsay (1993, Example 2) for a special choice of the function G.

To apply Lemma 1(i) to the specification (9), define $A_1=\bar{A}_p((\phi_1,\ldots,\phi_p)')$ and $A_2=\bar{A}_p((\phi_1+\psi_1,\ldots,\phi_p+\psi_p)')$. From Theorem 1 and Proposition 5 of Liebscher (2005) we can then conclude that a sufficient condition for Lemma 1(i), and hence, Assumption 3 is that the joint spectral radius of the set of two matrices $\{A_1,A_2\}$ is smaller than one, or that

$$\rho\left(\{A_1,A_2\}\right)<1. \qquad (11)$$

Liebscher (2005, Section 7) provides a numerical example of this condition with $p=2$ and $G(y_{t-1},\ldots,y_{t-p})=\exp\left(-\gamma y_{t-1}^2\right)$ $(\gamma>0)$. This choice of the function G corresponds to the exponential autoregressive (EXPAR) model introduced by Haggan and Ozaki (1981) and also studied by Tong (1990, p. 108). In this numerical example condition (11) holds but the simpler condition (10) is violated. Clearly, the same conclusion is obtained even if the general function G is assumed.

The EXPAR model discussed above is closely related to the exponential smooth transition autoregressive (ESTAR) model which, along with other smooth transition autoregressive models, have been considered by Teräsvirta (1994) and van Dijk, Teräsvirta, and Franses (2002) amongst others. In the ESTAR case, (9) applies with $G(y_{t-1},\ldots,y_{t-p})=1-\exp\left(-\gamma(y_{t-d}-c)^2\right)$ whereas the logistic smooth transition autoregressive (LSTAR) specification is given by $G(y_{t-1},\ldots,y_{t-p})=\left(1+\exp\left(-\gamma(y_{t-d}-c)\right)\right)^{-1}$ $(\gamma>0,\ c\in\mathbb{R},\ 1\le d\le p)$. A generalization of the latter is obtained by

$$G(y_{t-1},\ldots,y_{t-p})=\left(1+\exp\left(-\gamma\prod_{j=1}^{k}(y_{t-d}-c_j)\right)\right)^{-1},$$

where γ and d are as above and $c_1<\cdots<c_k$. Of course, conditions (10) and (11) which apply to the general specification (9) also apply to all these special cases. It may also be noted that, although our smoothness assumption rules out the possibility that G is an indicator function, an approximating smooth counterpart such as a logistic function is allowed.

4.2 Conditional variance: GARCH

Now consider the conditional variance. 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 (2005). In one variant of this model the dynamics of the conditional variance process are governed by

$$h_t = g\left(u_{t-1}, h_{t-1}\right) = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + \alpha^* G(u_{t-1}) u_{t-1}^2, \tag{12}$$

where G is a smooth 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 Assumption 4 for this model is straightforward. Assumptions 4(a) and 4(b) clearly hold with the lower bound in the former given by $\underline{g} = \omega$. For Assumption 4(d) we choose $a = \beta$, $\varphi(\varepsilon_t) = \max\{\alpha, \alpha + \alpha^*\}\varepsilon_t^2$, and $c = \omega$, giving the moment condition $E[\left(\beta + \max\{\alpha, \alpha + \alpha^*\}\varepsilon_t^2\right)^r] < 1$ which reduces to $\beta + \max\{\alpha, \alpha + \alpha^*\} < 1$ when $r = 1$. Finally, note that $g(0, x) = \omega + \beta x$ so that, because $\beta < 1$, the Lipschitz condition (7), and hence Assumption 4(c), is satisfied.

We shall now demonstrate that Assumption 5 holds for the GARCH model (12). Following the discussion in Section 2 we consider the last row of the derivative matrix $\nabla F_{p+2}^{(0)}$. The needed derivatives can be straightforwardly obtained from equation (12) and, unless otherwise stated, all derivatives below are evaluated at $e_2^{(0)} = \cdots = e_{p+2}^{(0)} = 0$. First note that $\partial h_i / \partial e_j = \beta \partial h_{i-1} / \partial e_j$, $i = 3, \ldots, p+2$, $j = 1, \ldots, i-2$, and $\partial h_i / \partial e_{i-1} = 0$, $i = 3, \ldots, p+2$, and thus the last row of the matrix $\nabla F_{p+2}^{(0)}$ becomes $\left[\ \partial h_{p+2}/\partial e_1 \quad 0 \quad \cdots \quad 0\ \right]$. 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

$$\begin{aligned} \partial h_{p+2}/\partial e_1 &= 2\beta^p h(z_0) e_1 \left(\alpha + \alpha^* G\left(h(z_0)^{1/2} e_1\right)\right) \\ &\quad + \beta^p \alpha^* h(z_0)^{3/2} G'\left(h(z_0)^{1/2} e_1\right) e_1^2. \end{aligned}$$

For the standard linear GARCH model $\alpha^* = 0$ so that, since $\alpha > 0$ and $\beta > 0$ is assumed, $\partial h_{p+2}/\partial e_1 = 2\beta^p \alpha h(z_0) e_1$ is nonzero for any $e_1 \neq 0$ whereas $\partial h_{p+2}/\partial e_2, \ldots, \partial h_{p+2}/\partial e_{p+1}$ are zero. Thus, as discussed in Section 2, Assumption 5 holds. The same conclusion is obtained even if $\alpha^* \neq 0$. In this case $\partial h_{p+2}/\partial e_1 \neq 0$ 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 5.

In an alternative smooth transition GARCH model, suggested by Lanne and Saikkonen (2005), the conditional heteroskedasticity is specified as

$$h_t = g\left(u_{t-1}, h_{t-1}\right) = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + (\omega^* + \beta^* h_{t-1}) G(h_{t-1}), \tag{13}$$

where G is again a smooth function with range $[0,1]$ and the parameters satisfy $\omega > 0$, $\alpha > 0$, $\beta > 0$, $\omega^* \geq 0$, and $\beta^* \geq 0$. Again, the validity of Assumptions 4(a) and (b) is clear with $\underline{g} = \omega$ in the former. For Assumption 4(d), we may choose $a = \beta + \beta^*$, $\varphi(\varepsilon_t) = \alpha\varepsilon_t^2$, and $c = \omega + \omega^*$. This gives the condition $E[\left(\beta + \beta^* + \alpha\varepsilon_t^2\right)^r] < 1$ or $\beta + \beta^* + \alpha < 1$ when $r = 1$. To verify Assumption 4(c), the Lipschitz condition (7) appears inconvenient, and the latter part of this assumption becomes useful. For this model, $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$. Thus, if we assume that 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). That Assumption 5 holds will be discussed below in the context of a related model.

In the preceding model one can also consider the case $\omega > 0$, $\omega^* > 0$, $\alpha > 0$, and $\beta = \beta^* = 0$. This special case was applied by Lanne and Saikkonen (2005) with the function G (strictly) increasing. As above, one can check that Assumption 4 holds for this specification with the moment condition in part (d) given by $E[\alpha^r \varepsilon_t^{2r}] < 1$ or $\alpha < 1$ when $r = 1$. More generally, the assumption that the (increasing) function G is bounded can be relaxed by requiring that $G(x) = o(x)$ as $x \to \infty$. It suffices to discuss Assumptions 4(c) and (d). Regarding the latter, one can write

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

where $\mathbf{1}(\cdot)$ signifies the indicator function. Choosing M large enough the last expression can be bounded from above by $\left(\epsilon + \alpha\varepsilon_t^2\right) x + c$ where $0 < \epsilon < 1$ is so small that $E\left[\left(\epsilon + \alpha\varepsilon_t^2\right)^r\right] < 1$ holds whenever $E[\alpha^r \varepsilon_t^{2r}] < 1$. Thus, Assumption 4(d) holds. Since $g(u, x) \geq g(0, x)$ for all $(u, x) \in \mathbb{R} \times \mathbb{R}_+$ also Assumption 4(c) holds.

Now consider Assumption 5 in the preceding model. In the same way as in the standard GARCH model discussed in the context of model (12), it is straightforward to check that, when evaluated at $e_2^{(0)} = \cdots = e_{p+2}^{(0)} = 0$, $\partial h_i / \partial e_j = \omega^* G'(h_{i-1}) \partial h_{i-1} / \partial e_j$, $i = 3, \ldots, p+2$, $j = 1, \ldots, i-2$, and $\partial h_i / \partial e_{i-1} = 2\alpha h_{i-1} e_{i-1}$, $i = 3, \ldots, p+2$. 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 $e_2^{(0)} = \cdots = e_{p+2}^{(0)} = 0$ becomes nonzero while $\partial h_{p+2} / \partial e_2 = \cdots = \partial h_{p+2} / \partial e_{p+1} = 0$. Hence Assumption 5 holds. The same reasoning applies to model (13), for it suffices to consider the (increasing) function $\beta x + (\omega^* + \beta^* x) G(x)$ in place of $\omega^* G(x)$.

4.3 Conditional variance: ARCH

In the preceding cases the verification of Assumption 5 required that the considered GARCH models do not reduce to pure ARCH models. We shall now demonstrate that Assumption 5 can also be verified when the conditional heteroskedasticity is modeled by a pure ARCH model. Because in many fields of application pure ARCH models are seldom adequate and, as discussed in the introduction, stability results for them are already available we shall only consider the standard ARCH model $h_t = \omega + \alpha u_{t-1}^2$ $(\alpha > 0)$ and, for simplicity, assume that $p = 1$. It suffices to show that the matrix

$$D_1 = \begin{bmatrix} \partial y_2 / \partial e_1 & \partial y_2 / \partial e_2 \\ \partial h_3 / \partial e_1 & \partial h_3 / \partial e_2 \end{bmatrix}$$

is nonsingular when the partial derivatives are evaluated at suitable values of e_1 and e_2. By straightforward derivation,

$$\begin{aligned} &\partial y_2 / \partial e_1 = h(z_0)^{1/2} f'\left(f(y_0) + h(z_0)^{1/2} e_1\right) + h(z_0)\, h_2^{-1/2} \alpha e_1 e_2 \\ &\partial y_2 / \partial e_2 = h_2^{1/2} \\ &\partial h_3 / \partial e_1 = 2\alpha^2 h(z_0) e_1 e_2^2 \\ &\partial h_3 / \partial e_2 = 2\alpha h_2 e_2. \end{aligned}$$

Assuming that e_2 is nonzero, we can transform D_1 to

$$D_2 = \begin{bmatrix} h_2^{1/2} h(z_0)^{1/2} f'\left(f(y_0) + h(z_0)^{1/2} e_1\right) & 0 \\ \alpha h(z_0) e_1 e_2 & 1 \end{bmatrix},$$

a matrix whose rank equals that of D_1 (multiply the first row of D_1 by $h_2^{1/2}$, divide the second row and second column by $2\alpha e_2$ and h_2, respectively, and substract the second row from the first one). Thus, assuming that the function f is not constant, we can find an e_1 such that the element of D_2 in the upper left hand corner is nonzero. Then D_2, or equivalently, D_1 is nonsingular, and because this can be done for each initial value z_0, Assumption 5 holds.

5 Conclusion

In this paper we have studied a nonlinear autoregressive model of order p 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 Liebscher (2005) and, hence, β–mixing. Conditions for existence of moments of the stationary distribution were also obtained. The assumptions needed to obtain these results are convenient because in most cases they restrict the conditional mean and conditional variance separately. To the best of our knowledge, these are the first practically applicable stability results for nonlinear autoregressions with GARCH errors. They are of importance as they open up the way for the development of rigorous asymptotic estimation theory for these models.

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. It would be of interest to consider alternative approaches in which such smoothness assumptions would not be required. For simplicity, we also focused on the leading case of GARCH(1,1) errors and left the extension to the general GARCH(r,s) case for future work.

Appendix: Proofs

Proof of Lemma 1. 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 Theorem 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).

To justify the second part of the lemma, use the notation introduced in Section 2 and denote $A = \bar{A}_{p+1}((\alpha_1, \ldots, \alpha_p, 0)')$. By direct calculation, the characteristic polynomial of A is (up to a factor ± 1) $\lambda(\lambda^p - \alpha_1\lambda^{p-1} - \ldots - \alpha_p)$. To see the equivalence of the two conditions, denote $f_1(\lambda) = \lambda^p - \alpha_1\lambda^{p-1} - \ldots - \alpha_p$, $f_2(\lambda) = \lambda^p - f_1(\lambda)$, and first suppose that $\sum_{j=1}^p \alpha_j < 1$. Now for $|\lambda| \geq 1$, $|f_1(\lambda)| \geq |\lambda|^p - |f_2(\lambda)| \geq |\lambda|^p \,(1 - \sum_{j=1}^p \alpha_j) > 0$, and hence the roots are inside the unit circle. On the other hand, if $\sum_{j=1}^p \alpha_j \geq 1$, then $f_1(1) \leq 0$ while $f_1(\lambda) \to +\infty$ as $\lambda \to +\infty$ and hence there is a root on or outside the unit circle. Thus under either condition $\rho(A) < 1$ (cf. Chen and Tsay (1993, Proof of Theorem 1.1)).

Now, 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, $0 < \underline{\nu}/\overline{\kappa} < 1$ where $\underline{\nu}$ and $\overline{\kappa}$ are the smallest and largest components of ν and κ, respectively. Next define the vector norm $\|\cdot\|^*$ in $\mathbb{R}^{p+1}$ by $\|y\|^* = \sum_{j=1}^{p+1} \kappa_j\,|y_j| = \kappa'\,|y|$ where $|y| = [|y_1| \ \ldots \ |y_{p+1}|]'$. For arbitrary $A = A(x) \in \mathcal{A}_*$ and $y \in \mathbb{R}^{p+1}$, $y \neq 0$, we have

$$\begin{aligned} \|A(x)y\|^* &= \kappa'\,|A(x)y| \\ &\leq \kappa_1 \sum_{j=1}^p \alpha_j\,|y_j| + \sum_{j=1}^p \kappa_{j+1}\,|y_j| \\ &= \kappa' A\,|y| \\ &= \kappa'\,|y| - \kappa'(I_{p+1} - A)\,|y| \\ &= \kappa'\,|y| - \nu'\,|y| \\ &= \kappa'\,|y| \left(1 - \frac{\nu'\,|y|}{\kappa'\,|y|}\right) \\ &\leq \|y\|^* \left(1 - \frac{\underline{\nu}}{\overline{\kappa}}\right), \end{aligned}$$

where $0 < 1 - \underline{\nu}/\overline{\kappa} < 1$. 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 Meyn and Tweedie (1993) (note that under our assumptions, the conditions (NSS1)–(NSS3) in Meyn and Tweedie (1993, 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 Meyn and Tweedie (1993), respectively). As discussed in Section 2, forward accessibility follows from Assumption 5 and Proposition 7.1.4 of Meyn and Tweedie (1993). The existence of a globally attracting state is shown below in Lemma 3. Thus, from Propositions 7.1.5 and 7.2.5(i), and

Theorem 7.2.6 of Meyn and Tweedie (1993) we can conclude that the Markov chain Z_t is an irreducible T–chain. Aperiodicity is obtained from Theorems 7.3.3 and 7.3.5(ii) of the same reference (see also the proof of Proposition 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 Theorem 7.3.3 of ibid. necessarily equals one). That every compact set is small now follows from Theorems 6.2.5(ii) and 5.5.7 of Meyn and Tweedie (1993). 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 $t \to \infty$ (see Meyn and Tweedie (1993, p. 160)). First suppose that the convergence in Assumption 4(c) holds for all $h_0 \in \mathbb{R}_+$ so that for every $z_0 \in \mathcal{Z}$, $h_t \to h^*$ as $t \to \infty$.

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

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

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 $z_0 \in \mathcal{Z}$, it is possible for the associated control model to reach a state z^* whose first $p+1$ components are $y_1^*, \ldots, y_{p+1}^*$, the components of the vector y^*, and the last component is h^*. Let $z_0 = [y_0' \;\; h_0]' \in \mathcal{Z}$ where $y_0 = [y_{0,1} \;\cdots\; y_{0,p+1}]'$. From the first step of the associated control model one then obtains

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

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

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

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

Next form z_t^* with $e_t = 0$, $t = p+2, p+3, \ldots$. 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, $z_t^* = [y^{*\prime}\ \ h_t^*]'$ $(t \geq p+2)$ where the last component satisfies $h_t^* = g(0, h_{t-1}^*)$ for $t \geq p+3$. Because Assumption 4(c) implies that $h_t^* \to h^*$ as $t \to \infty$ we can conclude that z^* 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 $e_2, \ldots, e_{p+2}$ to reach a state whose first $p+1$ components are $y_1^*, \ldots, y_{p+1}^*$, 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 $k = 4, \ldots, p+2$. Thus, after $p+2$ steps we are in a state $z_{p+2}^* = \left[y^{*\prime}\ \ h_{p+2}\right]'$ and we continue by forming z_t^* with $e_t = 0$, $t = p+3, p+4, \ldots$. 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 $t \to \infty$. Thus we have again shown that a globally attracting state exists for the associated control model. ■

Lemma 4 *Under Assumptions 1–4 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 positive r and 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

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

and, since the functions F_0 and h are bounded on the set A_N, we can find an $M_N < \infty$ such that

$$\sup_{z \in A_N} E\left[\|Z_t\| \mid Z_{t-1} = z\right] < M_N. \tag{14}$$

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 noted above. We have

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

Here the first inequality is justified by the fact that, for all z, $\Pr(Z_t \notin B_N \mid Z_{t-1} = z) = \Pr\left(\|Z_t\| > M_N \text{ or } h(z) < \underline{g} \mid Z_{t-1} = z\right)$ but $h(z) < \underline{g}$ is impossible by Assumption 4(a). The second inequality is Markov's and the third one is due to (14). That the set A_N is small can now be concluded from Proposition 5.2.4 of Meyn and Tweedie (1993). ■

Proof of Theorem 1. 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 with properties described in Assumption 3, i.e., a norm that satisfies $\|A\|^* \leq \rho$ for all $A \in \mathcal{A}_*$ and with $\rho \in (0,1)$. Since all vector norms are equivalent in finite-dimensional real (or complex) vector spaces, there exists a finite positive constant C such that $\|y\| \leq C^{1/2r}\|y\|^*$ for all $y \in \mathbb{R}^{p+1}$ (see e.g. Horn and Johnson (1985, Section 5.4)). Denote $V_*(z) = 1 + C\|y\|^{*2r} + h^r(z)$. In Lemma 5 the conditional expectation $E[V_*(Z_t) \mid Z_{t-m} = z]$ is examined and it is demonstrated that it satisfies an m–step ahead drift criterion (for a large m chosen in the proof of the lemma). More precisely, in this lemma it is shown that condition (19.15) of Meyn and Tweedie (1993) holds for the function $V_*(z)$ (with the choice $n(z) \equiv m$). Finally an application of our Lemma 6 below establishes that Z_t is V_*–geometrically ergodic in the sense of Liebscher (2005).

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*

$$E[V_*(Z_t) \mid Z_{t-m} = z] \leq \lambda^{1/2}\left(1 + C\|y\|^{*2r} + h^r(z) + \mathbf{1}_K(z)\right). \tag{15}$$

In other words, the drift condition (19.15) of Theorem 19.1.3 of Meyn and Tweedie (1993) holds (with the choice $n(z) \equiv m$*).*

Proof. First note that, by Hölder's inequality,

$$\left(\sum_{i=1}^{n} x_i\right)^r \leq \sum_{i=1}^{n} x_i^r \cdot n^{r-1} \tag{16}$$

for any positive x_i, $1 \leq i \leq n$, $n \in \mathbb{Z}_+$, and $r > 1$ (and this trivially holds also for $r = 1$).

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 $k \geq 1$, and interpreting that an empty summation equals zero,

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

Using (16) we obtain

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

By Assumption 4(d), $E[c_t^r] < 1$ and we denote this expectation by δ.

Next note that (trivially) $E\left[h^r(Z_{t-m}) \mid Z_{t-m} = z\right] = h^r(z)$. Furthermore, using the notation $d = c^r/(1-\delta)$ and the independence of the c_t's,

$$\begin{aligned}(k+1)^{1-r} E\left[h^r(Z_{t-m+k}) \mid Z_{t-m} = z\right] &\leq h^r(z)\delta^k + c^r(1 + \sum_{j=0}^{k-2}\delta^{j+1}) \\ &\leq h^r(z)\delta^k + d.\end{aligned} \tag{17}$$

In particular, for $k = 1, \ldots, m-1$,

$$E\left[h^r(Z_{t-m+k}) \mid Z_{t-m} = z\right] \leq (k+1)^{r-1}\left(h^r(z)\delta^k + d\right) \leq m^{r-1}h^r(z)\delta^k + d', \tag{18}$$

where $d' = m^{r-1}d$.

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 and usual properties of vector and matrix norms yield

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

In the summation above there are $2m+1$ terms, and hence using (16)

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

Denote $\|\iota\|^{*2r} = \iota^*$ and note that $\|A\left(\cdot\right)\|^{*2r} \leq \rho^{2r}$, $\|\iota b\left(\cdot\right)\|^{*2r} \leq \iota^* B$ for some finite B (because

$b(\cdot)$ is bounded), $\|\iota h(\cdot)^{1/2}\varepsilon_t\|^{*2r} \le \iota^* h^r(\cdot)\varepsilon_t^{2r}$, and $E[\varepsilon_t^{2r}] \overset{def}{=} \gamma_{2r} < \infty$. Thus,

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

where the last inequality makes use of (18) and the fact that $E\left[h^r(Z_{t-m}) \mid Z_{t-m} = z\right] = h^r(z)$. Defining $\phi = \max\{\rho^{2r}, \delta\} < 1$ and $\phi' = \frac{1}{1-\phi}$ we get

$$\begin{aligned}
& (2m+1)^{1-2r} E\left[\|Y_t\|^{*2r} \,|Z_{t-m} = z\right] \\
\le\ & \phi^m \|y\|^{*2r} + \iota^* B \left(1 + \sum_{j=0}^{m-2} \phi^{j+1} \right) + \iota^* \gamma_{2r} \left(m^{r-1} \phi^{m-1} h^r(z) + d' \right) \\
& + \iota^* \gamma_{2r} \left(\sum_{j=0}^{m-3} \phi^{j+1} \left(m^{r-1} \phi^{m-2-j} h^r(z) + d' \right) + \phi^{m-1} h^r(z) \right) \\
\le\ & \phi^m \|y\|^{*2r} + \iota^* B \phi' + \iota^* \gamma_{2r} \left(m^{r-1} \phi^{m-1} h^r(z) + d' \right) \\
& + \iota^* \gamma_{2r} \left(\sum_{j=0}^{m-3} \phi^{m-1} m^{r-1} h^r(z) + \sum_{j=0}^{m-3} \phi^{j+1} d' + \phi^{m-1} h^r(z) \right) \\
\le\ & \phi^m \|y\|^{*2r} + \iota^* B \phi' + m \cdot \iota^* \gamma_{2r} m^{r-1} \phi^{m-1} h^r(z) + \iota^* \gamma_{2r} d' \phi' \\
=\ & \phi^m \|y\|^{*2r} + \iota^* \gamma_{2r} m^r \phi^{m-1} h^r(z) + \iota^* \phi' (B + \gamma_{2r} d'). \qquad (19)
\end{aligned}$$

Combining the inequalities (17) (with $k = m$) and (19) yields

$$\begin{aligned}
& E\left[V_*(Z_t) \mid Z_{t-m} = z\right] \\
=\ & E\left[1 + C\|Y_t\|^{*2r} + h^r(Z_t) \mid Z_{t-m} = z\right] \\
\le\ & 1 + C(2m+1)^{2r-1} \left(\phi^m \|y\|^{*2r} + \iota^* \gamma_{2r} m^r \phi^{m-1} h^r(z) + \iota^* \phi' (B + \gamma_{2r} d') \right) \\
& + (m+1)^{r-1} \left(h^r(z) \delta^m + d \right) \\
=\ & 1 + C\left[(2m+1)^{2r-1} \phi^m\right] \|y\|^{*2r} + \left[C(2m+1)^{2r-1} \iota^* \gamma_{2r} m^r \phi^{m-1} + (m+1)^{r-1} \delta^m\right] h^r(z) \\
& + \left\{ C(2m+1)^{2r-1} \iota^* \phi' (B + \gamma_{2r} d') + (m+1)^{r-1} d \right\}. \qquad (20)
\end{aligned}$$

Since $0 < \delta \le \phi < 1$, we can clearly choose an m large enough so that both of the expressions in square brackets in (20) are smaller than some $\lambda < 1$. The expression in curly brackets in (20) is clearly finite, and thus for some $L < \infty$ we have

$$E\left[V_*(Z_t) \mid Z_{t-m} = z\right] \le \lambda\left(1 + C\|y\|^{*2r} + h^r(z)\right) + L. \tag{21}$$

What remains to be examined is the behaviour of (21) on and off a small set. To this end, write the right-hand-side of (21) as

$$\lambda^{1/2}\left(1 + C\|y\|^{*2r} + h^r(z)\right) \cdot \lambda^{1/2}\left(1 + \frac{L}{\lambda\left(1 + C\|y\|^{*2r} + h^r(z)\right)}\right). \tag{22}$$

By Lemma 4 the set $A_N = \{z \in \mathcal{Z} : \|y\|^{*2r} \le N, h^r(z) \le N\}$ is small. Off this set either $\|y\|^{*2r} > N$ or $h^r(z) > N$, and the ratio in (22) can clearly be made arbitrarily small by choosing N large enough. Therefore for a large enough N

$$\lambda^{1/2}\left(1 + \frac{L}{\lambda\left(1 + C\|y\|^{*2r} + h^r(z)\right)}\right) < 1$$

and hence

$$E\left[V_*(Z_t) \mid Z_{t-m} = z\right] \le \lambda^{1/2}\left(1 + C\|y\|^{*2r} + h^r(z)\right)$$

off the set A_N. On the other hand, the right hand side of (21) is clearly bounded on the set A_N. Therefore, condition (15) 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 $V : \mathcal{X} \to [1, \infty)$ bounded on K, and positive constants $\lambda < 1$ and $b < \infty$*

$$E[V(X_t) \mid X_{t-m} = x] \le \lambda^m \left(V(x) + b\mathbf{1}_K(x)\right) \tag{23}$$

for all $x \in \mathcal{X}$. Then X_t is V–geometrically ergodic in the sense of Liebscher (2005).

Proof. If $m = 1$ then X_t is V–geometrically ergodic in the sense of Meyn and Tweedie (1993) by their Theorem 15.0.1, and hence the stated weaker form of geometric ergodicity also follows. Suppose now that $m > 1$. It immediately follows from Theorem 19.1.3 of Meyn and Tweedie (1993) that X_t is geometrically ergodic and for some $\varrho < 1$ and $R < \infty$

$$\|P_X^n(x, \cdot) - \pi_X(\cdot)\| \le \varrho^n R V(x),$$

where $\|\cdot\|$ signifies the total variation norm, and $P_X^n(x, \cdot)$ and $\pi_X(\cdot)$ are the n–step transition probability measure and stationary measure of X_t, respectively. What remains to be proven is that the expectation $\int_{\mathcal{X}} \pi_X(dy)V(y)$ is finite. To this end, we will first establish that X_{tm}, the m–skeleton of X_t, is V–geometrically ergodic in the sense of Meyn and Tweedie (1993). By Proposition 5.4.5(iii) of Meyn and Tweedie (1993) and the assumptions of the lemma, the m–skeleton is irreducible and aperiodic, and satisfies the drift criterion (23) where the set K is small for the original chain X_t (but not necessarily for the m–skeleton). To establish a drift criterion with a set that is small for the m–skeleton, first choose a $\bar{\lambda}$ such that $\lambda^m < \bar{\lambda} < 1$. By Lemma 14.2.8 of Meyn and Tweedie (1993) we can find a set K_m which is small for the m–skeleton and such that

$$1_K(x) \le \sum_{i=0}^{m-1} \int_{\mathcal{X}} P_X^i(x, dy)\, \mathbf{1}_K(y) \le m\mathbf{1}_{K_m}(x) + (\bar{\lambda} - \lambda^m)/\lambda^m b.$$

Therefore

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

because $1 \leq V(x)$. Thus the m–skeleton satisfies a drift criterion with a set K_m which is small for the m–skeleton. Therefore by Theorem 15.0.1 of Meyn and Tweedie (1993) the m–skeleton is V–geometrically ergodic in the sense of Meyn and Tweedie (1993).

To complete the proof, note that by Theorem 10.4.5 of Meyn and Tweedie (1993) the stationary distributions of the m–skeleton of X_t and X_t itself are the same and, by the V–geometric ergodicity of the m–skeleton, the expectation $\int_{\mathcal{X}} \pi_X(dy)V(y)$ is finite. ■

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.

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.

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., 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.

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 Nonlinear 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.

Haggan, V., and T. Ozaki (1981): "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, 68, 189–196.

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

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 pth-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., and M. McAleer (2003): "Asymptotic theory for a vector ARMA–GARCH model," *Econometric Theory*, 19, 280–310.

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 (2005): "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.

Rabemananjara, R., and J.-M. Zakoïan (1993): "Threshold ARCH models and asymmetries in volatility," *Journal of Applied Econometrics*, 8, 31–49.

Saikkonen, P. (2005): "Stability results for nonlinear error correction models, " *Journal of Econometrics*, 127, 69–81.

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.

Tong, H. (1990): *Non-linear time series: A dynamical system approach.* Oxford University Press, London.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002): "Smooth transition autoregressive models — A survey of recent developments," *Econometric Reviews*, 21, 1–47.