# Stability of Nonlinear Stochastic Recursions with Application to Nonlinear AR-GARCH Models

Daren B.H. Cline[1]
Department of Statistics
Texas A&M University
College Station TX 77843-3143

25 March 2007

**Abstract.**

We characterize the Lyapounov exponent and ergodicity of nonlinear stochastic recursion models, including nonlinear AR-GARCH models, in terms of an easily defined uniformly ergodic process. Properties of this latter process, known as the collapsed process, also determine the existence of moments for the stochastic recursion when it is stationary. As a result, both the the stability of a given model and the existence of its moments may be evaluated with relative ease.

The method of proof involves piggybacking a Foster-Lyapounov drift condition on certain characteristic behavior of the collapsed process.

AMS 2000 subject classification: primary 60G10, 60J05; secondary 62M10, 91B84.
key words: stochastic recursion, GARCH, ergodicity, stationary distribution
running title: Stability of Nonlinear Stochastic Recursions

[1]dcline@stat.tamu.edu

## 1. Introduction.

**1.1. Overview and Objectives.** Interest in the highly successful generalized autoregressive conditional heteroscedastic (GARCH) time series models has been, in recent years, extended to models with autoregression and threshold or other nonlinear behavior. We term such models generally as "nonlinear AR-GARCH type" time series. Like most time series models, they have a state space representation that is a stochastic recursion of the form

$$X_t = F(X_{t-1}, e_t), \tag{1.1}$$

with $F : \mathbb{R}^m \times \mathbb{E} \to \mathbb{X} \subset \mathbb{R}^m$ and $\{e_t\}$ is an iid sequence of random variables in some Euclidean space $\mathbb{E}$. The definitions for the state vector $X_t$, the state space $\mathbb{X}$ and the recursion function $F$ will depend on the time series model.

Nonlinear AR-GARCH time series, characteristically, exhibit stochastic volatility because the random errors get multiplied by something roughly proportional to the size of the previous state vector. Indeed, they typically can be expressed in a $m$-dimensional Markov (state space) representation of the form

$$X_t = B\left(\tfrac{X_{t-1}}{||X_{t-1}||}, e_t\right) ||X_{t-1}|| + C(X_{t-1}, e_t) \tag{1.2}$$

where $0 < ||B(x/||x||, u)|| \le \bar{b}(1 + |u|)$ and $||C(x, u)|| \le \bar{c}(x)(1 + |u|)$ for finite $\bar{b}$ and $\bar{c}(x) = o(||x||)$. (See examples in section 2.) The point to be made here is that the first term on the right is *homogeneous* in $X_{t-1}$, and it dominates when the process becomes very large.

We are interested whether such a time series is ergodic, and specifically whether it is stable and does not grow indefinitely. Accordingly, the geometric "drift" of the state process $\{X_t\}$, when large, is characterized in terms of the limiting parameter

$$\bar{\gamma} \stackrel{\text{def}}{=} \liminf_{n\to\infty} \limsup_{||x||\to\infty} \frac{1}{n} E\left(\log\left(\frac{1 + ||X_n||}{1 + ||X_0||}\right) \,\Big|\, X_0 = x\right). \tag{1.3}$$

We call this constant the Lyapounov exponent of the process because it describes the longterm change in the magnitude of $X_t$, at least when the initial state is large. Furthermore, it is the critical value for stability. Under irreducibility and aperiodicity assumptions, geometric ergodicity is a consequence of a negative Lyapounov exponent (Cline and Pu (1999)). Conversely, with similar assumptions, a positive Lyapounov exponent ensures that $\{X_t\}$ is transient (Cline and Pu (2001)).

Since $\bar{\gamma}$ is defined as a double limit, it seemingly is difficult to compute. Estimating it by simulation requires observing the time series during its most extreme (and most volatile) behavior. Our primary objective in this paper, therefore, is to give a precise value to $\bar{\gamma}$ in terms that may more easily be computed.

To see how this may be accomplished, observe that only the first term in (1.2) describes the behavior of $X_t$ when $||X_{t-1}||$ is large. To exploit this, we consider the related (though inherently non-ergodic) Markov process

$$X_t^* = B\left(\tfrac{X_{t-1}^*}{||X_{t-1}^*||}, e_t\right) ||X_{t-1}^*||. \tag{1.4}$$

This is the homogeneous form of (1.2). Let $\Theta = \{x \in \mathbb{X} : ||x|| = 1\}$ and define

$$w(\theta, u) = ||B(\theta, u)||, \quad \eta(\theta, u) = B(\theta, u)/||B(\theta, u)||, \quad \text{for } \theta \in \Theta,\ u \in \mathbb{E}.$$

The homogeneous process can be collapsed to $\Theta$:

$$\theta_t^* = X_t^*/||X_t^*|| = \eta(\theta_{t-1}^*, e_t). \tag{1.5}$$

Also, let $W_t^* = w(\theta_{t-1}^*, e_t)$. The "collapsed" process $\{\theta_t^*\}$ also is Markov and typically it is uniformly ergodic. Its behavior (and more specifically, the behavior of $W_t^*$) determines both the stability and the existence of moments of the original process $\{X_t\}$.

In fact, we will demonstrate (Theorem 3.3 below) that the Lyapounov exponent is equal to the stationary value of

$$\gamma \stackrel{\text{def}}{=} E\Big(\log(W_t^*)\Big).$$

One very practical consequence of this result is that it provides a straightforward method to evaluate the Lyapounov exponent of $\{X_t\}$ (given the model and parameter values) because an expectation for a uniformly ergodic process frequently is very easy to estimate.

Our second objective is to determine which moments are finite for the stationary distribution of $\{X_t\}$ when the process is ergodic. Again, the answer can be expressed precisely in terms of $\{(\theta_t^*, W_t^*)\}$. Simply stated, the $\zeta$-moment for $\{X_t\}$ exists when there is a bounded, positive function $\lambda(\theta)$ such that

$$\sup_{\theta_\Theta} E\Big(\tfrac{\lambda(\theta_1^*)}{\lambda(\theta)}(W_t^*)^\zeta \;\Big|\; \theta_0^* = \theta\Big) < 1.$$

The exact result is contained in the statements of Theorems 3.2 and 3.4.

This paper extends similar results for a threshold AR-ARCH model (Cline and Pu (2004)) and advances that work in several respects. First, the model is more general and includes models with GARCH-like behavior. This is notable because the state vector is more complex for GARCH type time series. In particular, it includes each new "volatility" value which depends completely on the previous state, thereby adding to the singularity of the stochastic recursion and complicating both proofs and assumptions. The new results also apply to regime switching models, threshold variable driven switching models and to models that are not strictly piecewise versions of GARCH models. The arguments have been disentangled with certain of the assumptions and have been clarified substantially. In addition, there is more detail and precision in the theorem statements than there was for the analogous AR-ARCH versions. Finally, we describe the behavior of the process when it is not stable and demonstrate the sharpness of our stability condition.

An ultimate goal, not undertaken here, is to investigate the tail properties of stationary distributions when they exist. We simply note that various authors have shown stationary distributions typically have regularly varying probability tails, at least for the special cases studied so far (Borkovec and Klüppelberg (2001), Basrak, Davis and Mikosch (2002), Klüppelberg and Pergamenchtchikov (2004)), Zhang and Tong (2004), de Saporta (2005) and Cline (2007b)).

**1.2. Motivation and Background.** A motivating factor for desiring a simple but precise condition for stability is that, with only a limited understanding, a time series analyst may make parameter assumptions that are based on the requirements for ordinary GARCH and AR-GARCH models. Although shown to be sufficient for some models (Lanne and Saikkonen (2005) and Lee and Shin (2005)), such assumptions can be unduly harsh. Like threshold autoregression models, the parameter spaces of stable nonlinear AR-(G)ARCH models can be unexpectedly large. See Cline (2007a) for a simple example. Similarly, while moment conditions have been ascertained for ordinary GARCH models (e.g., He and

Teräsvirta (1999a,b)), those conditions do not translate well to nonlinear models or even to GARCH models with autoregression terms.

Identifying conditions for ergodicity of the model (1.1) is a problem of longstanding interest, and determining when the process is (geometrically) ergodic and mixing implies useful statistical behavior. In this paper we establish geometric ergodicity conditions for (1.1) when

(i) the stochastic recursion is neither linear nor "close" to linear,

(ii) $F(\cdot, u)$ is discontinuous, albeit piecewise continuous,

(iii) $F(x, e_1)$ has a singular distribution with support on a low-dimensional manifold of $\mathbb{R}^m$ that, in addition, depends on $x$, and

(iv) the process exhibits stochastic volatility such that the random part of $F(x, e_1)$ is of the order $||x||$ for large $||x||$.

Each of the characteristics just described presents complications for proving accurate/sharp conditions of ergodicity. For example, (ii) and (iii) each rule out that (1.1) is a Feller process, and usually they imply great care is required just to establish irreducibility and aperiodicity. The stochastic volatility prevents comparisons to dynamical systems and the nonlinearity implies that conditions based on "linear bounds" may not be sharp.

Numerous results are available in more "ideal" scenarios. For example, Bougerol and Picard (1992a,b) showed that embedding pure GARCH time series into (linear) random coefficients models is very beneficial, and obtained necessary and sufficient conditions for ergodicity based on the work of Brandt (1986). (See also Vervaat (1979), Grincevičius (1981), and Goldie and Maller (2000).) For this approach to apply more generally, there must at least exist a matrix function $G(u)$ such that $E\left(\log_+\left(||F(x, e_1) - G(e_1)x||\right)\right)$ is bounded. The Lyapounov exponent turns out to be $\inf_{n\geq 1} \frac{1}{n} E\left(\log\left(\|\prod_{t=1}^n G(e_t)\|\right)\right)$. Since this is defined in terms of the matrix norm of a longterm product of random matrices, it is not necessarily easy to compute. (It coincides with (1.3), with little additional assumption.)

Likewise, a condition is known (in a very general setting) when $F(x, \cdot)$ is Lipschitz continuous (Barnsley and Elton (1988), Elton (1990), Diaconis and Freedman (1999), Alsmeyer and Fuh (2001) and Jarner and Tweedie (2001)). Again the condition for ergodicity is that the Lyapounov exponent is negative: $\inf_{n\geq 1} \frac{1}{n} E\left(\sup_{x,y\in\mathbb{X}} \log\left(\frac{d(X_n(x), X_n(y))}{d(x,y)}\right)\right) < 0$, where $d$ is a metric on the state space $\mathbb{X}$ and $X_n(x)$ is the value $X_n$ takes when $X_0 = x$. As this also often is difficult to compute, one may utilize a stronger condition such as $E\left(\left(\sup_{x,y\in\mathbb{X}} \frac{d(F(x,e_1), F(y,e_1))}{d(x,y)}\right)^{\zeta}\right) < 1$ for some $\zeta > 0$. Steinsaltz (1999) has an apparently weaker condition which posits the existence of a test function $\lambda(x)$ such that $\sup_{x\in\mathbb{X}} E\left(\frac{\lambda(F(x,e_1))}{\lambda(x)} \limsup_{y\to x} \frac{d(F(x,e_1), F(y,e_1))}{d(x,y)}\right) < 1$, but then the problem also requires finding an optimal test function. Chan and Tong (1985) also used a Lipschitz continuity assumption early in the study of nonlinear time series.

Additionally, if the time series is a nonlinear autoregression with (roughly) constant noise variance then dynamical systems analysis and other methods have yielded important results. The literature is quite rich and we simply mention that it primarily begins with the work of Tong and Lim (1980), Tong (1981, 1983), Chan and Tong (1985) and Tong (1990), and was aided greatly by the Foster-Lyapounov drift condition theory ultimately consolidated in Meyn and Tweedie (1993). The drift condition approach has the advantage that it also provides sufficient conditions for the existence of moments.

Establishing sharp conditions for the ergodicity of the models we discuss here, however, has proven to be more difficult. The nonlinear, discontinuous, singular and stochastically volatile behaviors all have thwarted attempts to capitalize on comparisons to simpler, and more easily analyzed, models. There have been many papers that identify sufficient conditions for ergodicity of these models. These include the works of Rabemanajara and Zakoian (1993), Li and Li (1996), Liu, Li and Li (1997), Lu (1998), Ling (1999), Hwang and Woo (2001), Lu and Jaing (2001), Lanne and Saikkonen (2005), Lee and Shin (2005) and Meitz and Saikonnen (2006). But the conditions imposed are often restrictive because they rely on some sort of linear bound. For example, a typical condition requires $||F(x,u) - G(x,u)x||$ to be bounded in $x$ for some matrix function $G(x,u)$ such that the elements of $G(x,u)$ are uniformly bounded in $x$ by $G^*(u)$, and $G^*(e_1)$ satisfies a stability condition such as those for ordinary GARCH models or random coefficients models. Alternatively, the condition may be difficult to compute because it requires accurately bounding a high order iteration of $F$.

Ultimately, regardless of the assumptions, the issue of stability lies with an accurate computation of the Lyapounov exponent for the process. Even if the process is an ordinary high order GARCH time series and fits the results described in the literature above, we need to show how to compute the Lyapounov exponent efficiently.

The remainder of this section will explain the assumptions we use. Examples of nonlinear AR-GARCH models are given in section 2. Section 3 contains our results, section 4 has the proofs and section 5 verifies that the assumptions will hold for typical threshold models.

**1.3. Assumptions.** We list here the various assumptions held throughout the paper (except in section 5 where they are verified for certain models).

To be general, we express our results in terms of the $m$-dimensional state process $\{X_t\}$ as given in (1.2). In practice, the specifics will depend on the particular time series model under consideration. However, we assume $B(\cdot, u)$ is piecewise continuous so that our results will be useful for threshold models. In section 2 we will illustrate the decomposition for several types of models.

Let $||\cdot||$ be the usual Euclidean norm on $\mathbb{R}^m$. For the sake of generality, $\{e_t\}$ will be an iid error sequence on some open set $\mathbb{E}$ contained in an Euclidean space. We let the norm for this be denoted $|\cdot|$ to distinguish it from the other. In many cases, $\mathbb{E}$ will in fact be $\mathbb{R}$. The state space, $\mathbb{X} \subset \mathbb{R}^m$, for $\{X_t\}$ will be a cone: $x \in \mathbb{X}$ implies $cx \in \mathbb{X}$ for all $c > 0$. Let $\mathbb{X}^* = \mathbb{X}\backslash\{0\}$ be the appropriate space for the homogeneous process $\{X_t^*\}$, defined in (1.4).

The first assumption below is standard and the second is easily verified for typical models. The third assumption, while standard, may require technical verification and the fourth would certainly have to be checked for specific models. We discuss the validity of these assumptions for threshold AR-GARCH models in section 5.

ASSUMPTION A.1 *The error sequence $\{e_t\}$ is iid with $E(|e_t|^\beta) < \infty$ for some $\beta > 0$.*

The errors are often assumed to have a positive bounded density as well, mainly for the purpose of establishing Assumption A.3 below.

The first part of the next assumption ensures that the stochastic volatility component of $\{X_t\}$ is roughly proportional to the magnitude of the process. That is, the process is something like a (G)ARCH type model. The second part of Assumption A.2 guarantees the existence of a negative moment for $W_t^*$. This could be relaxed slightly for the ergodicity proof but we use it when we verify that $\{X_t\}$ is transient if the Lyapounov exponent is positive.

ASSUMPTION A.2

*(i) There exist* $\underline{b}_1 > 0$, $\underline{b}_2 \geq 0$, $\bar{b} < \infty$ *and* $\bar{c}(x) = o(||x||)$, *as* $||x|| \to \infty$, *such that* $\max(\underline{b}_1|u| - \underline{b}_2, 0) \leq ||B(\theta, u)|| \leq \bar{b}(1 + |u|)$ *for all* $u \in \mathbb{E}$, $\theta \in \Theta$, *and* $||C(x, u)|| \leq \bar{c}(x)(1 + |u|)$ *for all* $u \in \mathbb{E}$, $x \in \mathbb{X}$.

*(ii) For some finite* $K$ *and* $\alpha \in (0, 1]$, $P\Big(||B(x/||x||, e_1)||x|| + C(x, e_1)|| < \delta||x||\Big) < K\delta^\alpha$ *for all* $x \in \mathbb{X}$ *such that* $||x|| > 1$, *all* $\delta \in (0, 1]$.

The next assumption is a requirement for ergodicity of a Markov chain in the time series setting. It can be easy or difficult to verify for specific models. In our case, the difficulty arises since $B(x, e_1)$ typically is discontinuous in $x$ (e.g., for threshold models) and has a singular multivariate distribution. We provide a common approach in Theorem 5.1 and use it to look at threshold AR-GARCH models. The reader may wish to consult a text on discrete time Markov processes (such as Meyn and Tweedie (1993)) for definitions of the properties mentioned here.

ASSUMPTION A.3 $\{X_t\}$ *and* $\{X_t^*\}$ *are each aperiodic* $\phi$*-irreducible Markov chains, on* $\mathbb{X}$ *and* $\mathbb{X}^*$, *respectively. Furthermore, bounded subsets of* $\mathbb{X}$ *are small for* $\{X_t\}$ *and subsets of* $\mathbb{X}^*$ *that are bounded and bounded away from* $\{0\}$ *are small for* $\{X_t^*\}$.

The requirement in Assumption A.3 that bounded sets are small is needed essentially because the critical behavior for stability of a time series is off bounded sets, that is, for large values. It is at the heart of the drift condition argument we will use. It is a continuity assumption but, unlike the Feller assumption, it is reasonable for most threshold models. If $\mathbb{X}$ is closed in $\mathbb{R}^m$ and the chains are aperiodic and $\phi$-irreducible then this requirement is equivalent to saying that $\{X_t\}$ and $\{X_t^*\}$ are $T$-chains (cf. Meyn and Tweedie (1993)).

It will be useful to study the directional component of the process, $\tilde{\theta}_t = X_t/||X_t||$, in addition to the collapsed process defined in (1.5). Also, let $\tilde{\theta}_1^* = \eta(X_0/||X_0||, e_1)$ which is the value of $\theta_1^*$ if $\theta_0^* = \tilde{\theta}_0 = X_0/||X_0||$. When $||X_0||$ is large, we would expect $\tilde{\theta}_1$ and $\tilde{\theta}_1^*$ to behave similarly. However, they must avoid the thresholds (points of discontinuity) and because $B(x, e_1)$ is singular this may require even further restriction. Thus, the purpose of the next assumption is to control the discontinuous behavior of the process. In particular, we expect $B(\cdot, u)$ to be piecewise uniformly continuous for each $u$ with its discontinuities on thresholds that do not attract the process. In combination with Assumption A.3, this imparts enough regularity for the homogeneous process to properly mimic the state space model. See section 5 for specifics in the case of a threshold AR-GARCH model.

ASSUMPTION A.4 *There exists a set* $\Theta_\#$, *open in* $\Theta = \{x \in \mathbb{X} : ||x|| = 1\}$, *such that*

*(i)* $\{B(\cdot, u)\}_{|u| \leq M}$ *is equicontinuous on* $\Theta_\#$ *for all finite* $M$. *That is, for each* $\epsilon > 0$ *and* $M < \infty$, *there is* $\delta > 0$ *such that* $|\theta - \theta'| < \delta$, $\theta, \theta' \in \Theta_\#$ *implies* $|B(\theta, u) - B(\theta', u)| < \epsilon$ *for all* $|u| \leq M$.

*Additionally,*

*(ii) for each* $\epsilon > 0$ *there exists* $L < \infty$ *such that* $P\Big(\tilde{\theta}_1 \in \Theta_\#, \tilde{\theta}_1^* \in \Theta_\# \;\Big|\; X_0 = x\Big) > 1 - \epsilon$ *for all* $x \in \mathbb{X}$ *with* $x/||x|| \in \Theta_\#$ *and* $||x|| > L$,

*and*

*(iii) for every $\epsilon > 0$ there exists $n \geq 1$ and $L < \infty$ such that $P\Big(\tilde{\theta}_n \in \Theta_\# \;\Big|\; X_0 = x\Big) > 1 - \epsilon$ for all $x \in \mathbb{X}$ with $||x|| > L$.*

The equicontinuity assumption above is used to prove uniform continuity of certain conditional expectations. Part (ii) indicates that $\tilde{\theta}_1$ and $\tilde{\theta}_1^*$ are similarly restricted to the "nice" set $\Theta_\#$ while part (iii) provides that $\Theta_\#$ will eventually be reached.

Assumption A.4 allows for less regularity on bounded sets. In some cases, $\{X_t\}$ may in fact be weak Feller when restricted to $\Theta_\# \times [0, \infty)$ but such a restriction does not immediately simplify the problem: in order to establish Assumption A.3, we have to deal with the closure of $\Theta_\# \times [0, \infty)$ one way or another.

## 2. Examples.

**2.1. Threshold AR-GARCH Model.** The "GARCH" in AR-GARCH can be interpreted two different ways. In this subsection and the next, we present the simpler interpretation also called "double autoregression" by Ling (2004). In subsection 2.3 we discuss the other interpretation, which is a nonlinear autoregression with GARCH errors and which (ironically) includes the "double threshold ARCH" model investigated by Li and Li (1996) and Ling (1999).

Essentially, a threshold AR-GARCH model is defined like a GARCH model with autoregression but with different parameter values on different subsets ("regimes") of the state space. To describe a typical threshold model explicitly, first suppose $C_1, \ldots, C_N$ are disjoint cones partitioning $\mathbb{R}^p$ and having, say, affine boundaries (which are the thresholds). Let $y = (y_1, \ldots, y_p) \in \mathbb{R}^p$, $s = (s_1, \ldots, s_q) \in [0, \infty)^q$ and define

$$\begin{aligned} a_i(y) &= a_{0i} + a_{1i}y_1 + \cdots + a_{pi}y_p, \\ b_i(y) &= \Big(b_{0i} + b_{1i}y_1^2 + \cdots + b_{pi}y_p^2\Big)^{1/2}, \\ \text{and} \quad c_i(s) &= \Big(c_{1i}s_1^2 + \cdots + c_{qi}s_q^2\Big)^{1/2}, \quad i = 1, \ldots, N, \end{aligned} \tag{2.1}$$

with each $b_{ki} \geq 0$ and each $c_{ki} \geq 0$. The threshold AR-GARCH$(p, q)$ model is expressed by

$$\begin{aligned} \xi_t &= a_{I_t}(\xi_{t-1}, \ldots, \xi_{t-p}) + \sigma_t e_t, \\ \sigma_t &= \Big(b_{I_t}^2(\xi_{t-1}, \ldots, \xi_{t-p}) + c_{I_t}^2(\sigma_{t-1}, \ldots, \sigma_{t-q})\Big)^{1/2}, \end{aligned} \tag{2.2}$$

where

$$I_t = \sum_{i=1}^{N} i 1_{(\xi_{t-1}, \ldots, \xi_{t-p}) \in C_i}. \tag{2.3}$$

The index $I_t$ is called the regime index.

In the simplest scenario, there are just two regimes and the threshold variable is $\xi_{t-k}$, $k \leq p$, meaning the regime is chosen with a delay. The regimes (cones) for this scenario are

$$C_1 = \{y \in \mathbb{R}^p : y_k \leq 0\} \quad \text{and} \quad C_2 = \{y \in \mathbb{R}^p : y_k > 0\}.$$

Lee and Shin (2005) have recently verified that extending the standard assumption for stability of a GARCH model suffices for models including those of the form we have just described, when each $a_i(y) = 0$. Specifically, their condition is

$$\sum_{k=1}^{p} \max_{1 \leq i \leq N} b_{ki} E(e_1^2) + \sum_{k=1}^{q} \max_{1 \leq i \leq N} c_{ki} < 1. \tag{2.4}$$

By explicitly identifying the Lyapounov exponent of the model, we will obtain a much weaker, and sharper, sufficient condition. Lee and Shin (2005) also show that (2.4) ensures a finite second moment for the stationary distribution. Again, we can weaken the condition. Note that $c_{1i} + \cdots + c_{qi} < 1$ is necessary for at least one $i$, and possibly all, since otherwise $\sigma_t$ is certain to converge to infinity.

The state vector for the model (2.2) is

$$X_t = (Y_t, S_t) \quad \text{with} \quad Y_t = (\xi_t, \ldots, \xi_{t-p+1}) \quad \text{and} \quad S_t = (\sigma_t, \ldots, \sigma_{t-q+1}),$$

and $m = p + q$. Additionally, the state space for $\{X_t\}$ is $\mathbb{X} = \mathbb{R}^p \times [0, \infty)^q$. For topological reasons, we allow $\sigma_t$ to take the value 0 but this is a transitory state.

Now let $a_i^*(y) = a_i(y) - a_{0i}$ and $b_i^*(y) = (b_i^2(y) - b_{0i})^{1/2}$. Then both are homogeneous: $a_i^*(y) = a_i^*(y/||y||)||y||$ and $b_i^*(y) = b_i^*(y/||y||)||y||$. Also, $d_{0i}(y,s) \stackrel{\text{def}}{=} (b_i^2(y) + c_i^2(s))^{1/2} - ((b_i^*(y))^2 + c_i^2(s))^{1/2}$ is bounded. Let $i(y) = \sum_{i=1}^N i 1_{y \in C_i}$. We therefore can express the model in the form (1.2) with $x = (y, s)$,

$$B(x,u) = B(x/||x||, u)||x|| = \Big(a_{i(y)}^*(y) + ((b_{i(y)}^*(y))^2 + c_{i(y)}^2(s)))^{1/2} u, y_1, \ldots, y_{p-1}, \\ ((b_{i(y)}^*(y))^2 + c_{i(y)}^2(s)))^{1/2}, s_1, \ldots, s_{q-1}\Big) \tag{2.5}$$

and $C(x,u) = (a_{0,i(y)} + d_{0,i(y)}(y,s)u, 0, \ldots, 0, d_{0,i(y)}(y,s), 0, \ldots, 0)$. Only the leading components of $B(x, e_1)$ and $C(x, e_1)$ are random.

In section 5, we will verify that this model satisfies our assumptions when the coefficients $b_{ki}$ and $c_{ki}$ are positive and $c_{1i} + \cdots + c_{qi} < 1$. Because $\sigma_t$ is completely determined by past values, verifying Assumption A.3 for the threshold AR-GARCH is much more involved than it would be for a threshold AR-ARCH model (that is, with $q = 0$).

**2.2. Nonlinear AR-GARCH Model.** More generally, a nonlinear AR-GARCH model may satisfy

$$\begin{aligned} \xi_t &= a(\xi_{t-1}, \ldots, \xi_{t-p}) + \sigma_t e_t, \\ \sigma_t &= \Big(b^2(\xi_{t-1}, \ldots, \xi_{t-p}) + c^2(\sigma_{t-1}, \ldots, \sigma_{t-q})\Big)^{1/2}, \end{aligned} \tag{2.6}$$

assuming only that

$$a(y) = a_0(y) + a^*(y) \quad \text{and} \quad b(y) = b_0(y) + b^*(y),$$

where $a_0$ and $b_0$ are bounded functions, $b$, $b^*$ and $c$ are positive, $a^*$ and $b^*$ are homogeneous, as well as piecewise continuous, and $c$ is assumed to be homogeneous and piecewise continuous itself. Let $x = (y, s)$ with $y = (y_1, \ldots, y_p)$ and $s = (s_1, \ldots, s_q)$. The state vector and state space are as in the previous subsection. Define

$$d_0(x) = (b^2(x) + c^2(s))^{1/2} - ((b^*(y))^2 + c^2(s))^{1/2}$$

and note that $|d_0(x)| \le |b_0(y)|$. Then the model satisfies (1.2) with

$$B(x,u) = B(x/||x||, u)||x|| = \Big(a^*(y) + ((b^*(y))^2 + c^2(s)))^{1/2} u, y_1, \ldots, y_{p-1}, \\ ((b^*(y))^2 + c^2(s)))^{1/2}, s_1, \ldots, s_{q-1}\Big)$$

and $C(x,u) = (a_0(y) + d_0(x)u, 0, \ldots, 0, d_0(x), 0, \ldots, 0)$.

One advantage of such a representation is that it allows for threshold-like models with smooth transitions. For example, a smooth transition model with delay lag $k \leq p$ could have

$$\begin{aligned} b(y) = &\left(b_{01} + b_{11}y_1^2 + \cdots + b_{p1}y_p^2\right)^{1/2} G(y_k/||y||) \\ &+ \left(b_{02} + b_{12}y_1^2 + \cdots + b_{p2}y_p^2\right)^{1/2} (1 - G(y_k/||y||)) \end{aligned}$$

with $G$ a continuous probability distribution on $[-1, 1]$. Then

$$\begin{aligned} b^*(y) = &\left(b_{11}y_1^2 + \cdots + b_{p1}y_p^2\right)^{1/2} G(y_k/||y||) \\ &+ \left(b_{12}y_1^2 + \cdots + b_{p2}y_p^2\right)^{1/2} (1 - G(y_k/||y||)) \end{aligned}$$

and $b_0(y) = b(y) - b^*(y)$ which is bounded. The functions $a(y)$ and $c(s)$ could have similar representations. See Lanne and Saikkonen (2005) for a discussion on models of this type (without the AR component). They also impose quite strong conditions for stability.

**2.3. Nonlinear Autoregression with GARCH Errors.** It is perhaps more common to define a model for which the conditional heteroscedasticity is an autoregression on past squared values of the error terms, rather than on past squared values of the time series. That is, the *error terms* are a (possibly nonlinear) GARCH process. This is the type of model discussed by Li and Li (1996), Ling (1999) and Meitz and Saikkonen (2006), and it requires a longer state vector than the models of the preceding subsections.

The threshold version is defined by

$$\begin{aligned} \xi_t &= a_{I_t}(\xi_{t-1}, \ldots, \xi_{t-p}) + \zeta_t, \\ \zeta_t &= \sigma_t e_t, \\ \sigma_t &= \left(b_{I_t}^2(\zeta_{t-1}, \ldots, \zeta_{t-r}) + c_{I_t}^2(\sigma_{t-1}, \ldots, \sigma_{t-q})\right)^{1/2}, \end{aligned}$$

where $a_i$, $b_i$ and $c_i$ are as in subsection 2.1 and $I_t = \sum_{i=1}^{N} i1_{(\xi_{t-1},\ldots,\xi_{t-p})\in C_i}$. The state vector now is $(Y_t, Z_t, S_t)$ with $Y_t$ and $S_t$ as before and $Z_t = (\zeta_t, \ldots, \zeta_{t-r+1})$. Letting $a_i^*$ and $b_i^*$ also be defined as they were in subsection 2.1, we may express (for $x = (y, z, s)$)

$$\begin{aligned} B(x, u) &= B(x/||x||, u)||x|| \\ &= \Big(a_{i(y)}^*(y) + ((b_{i(y)}^*(z))^2 + c_{i(y)}^2(s)))^{1/2}u, y_1, \ldots, y_{p-1}, \\ &\quad ((b_{i(y)}^*(z))^2 + c_{i(y)}^2(s)))^{1/2}u, z_1, \ldots, z_{r-1}, ((b_{i(y)}^*(z))^2 + c_{i(y)}^2(s)))^{1/2}, s_1, \ldots, s_{q-1}\Big). \end{aligned}$$

Our results apply equally well to models of this variety. The only difficulty might be the the extra effort required to verify Assumptions A.3 and A.4.

**2.4. Variable Driven Switching Model.** Wu and Chen (2007) recently introduced a variable driven switching model that combines the advantages of random regime switching with those of thresholds. Such a model fits into our framework as well. In the ordinary threshold model of subsection 2.1, the "threshold variable" (one or more linear combinations of $(\xi_{t-1}, \ldots, \xi_{t-p})$) determines the regime variable $(I_t)$ for the next observation. In a switching model, the regime is randomly determined. The model that Wu and Chen proposed uses the threshold variable to drive the random switching.

Specifically, in our context, suppose $h_1(\theta), \ldots, h_N(\theta)$ are positive functions of $\theta \in \Theta$. Recall that we defined $\tilde{\theta}_t = X_t/||X_t||$. The regimes are randomly chosen so that

$$P\Big(I_t = i \;\Big|\; \tilde{\theta}_{t-1} = \theta\Big) = \frac{h_i(\theta)}{\sum_{j=1}^N h_j(\theta)}.$$

Even more precisely, we augment the errors with an iid uniform(0,1) sequence $\{v_t\}$, independent of $\{e_t\}$, and define

$$q_i(\theta) = \frac{\sum_{j\le i} h_j(\theta)}{\sum_{j=1}^N h_j(\theta)}, \quad i = 0, \ldots, N.$$

Then, in (2.3), the equation for $I_t$ is replaced with

$$I_t = i(\tilde{\theta}_{t-1}, v_t) \stackrel{\text{def}}{=} \sum_{i=1}^N i 1_{q_{i-1}(\tilde{\theta}_{t-1}) < v_t \le q_i(\tilde{\theta}_{t-1})},$$

while (2.1) and (2.2) stay the same. The state vector is as before, and the representations for $B$ and $C$ remain much the same except that now the index function also depends on $v_t$ and the random error is $(e_t, v_t)$.

**2.5. Markov Regime Switching Model.** Regime switching models traditionally have involved a finite state Markov chain that dictates the regime. Letting $\{I_t\}$ be this chain, the model is further defined by (2.1) and (2.2). Now, however, the state vector must be $X_t = (Y_t, S_t, I_t)$ with the corresponding change in the state space. Let $\theta_t^*$ and $\Theta$ be defined in terms of $(Y_t, S_t)$ just as before and let $\mathbb{I} = \{1, \ldots, N\}$ be the set of regime indices. Then the collapsed model is now $\{(\theta_t^*, I_t)\}$, a Markov chain on $\Theta \times \mathbb{I}$. The results and proofs of the following sections are easily modified to accommodate this change. In fact, this really is just a special case of both the threshold and the variable driven models, with cones $C_i = \Theta \times \{i\}$ for the former and link functions $h_j((\theta, i)) = P(I_t = j \mid I_{t-1} = i)$ for the latter.

Franq and Zakoïan (2005) investigate even integer order moment properties of this model, without the AR term $a_i(y)$, obtaining necessary and sufficient conditions that agree with ours. Their conditions are expressed in terms of expectations of Kronecker products of random matrices but do not require an implicitly determined (test) function such as the $\lambda$ of the next section.

## 3. Main Results.

**3.1. The Collapsed Process.** We remind the reader that Assumptions A.1–A.4 are in effect for all the results. The collapsed process $\{\theta_t^*\}$, defined in (1.5), is evidently Markov. In this subsection we consider the behaviors it exhibits that are necessary for describing the stability results of the next subsection. The first result is preliminary and establishes ergodicity of $\{\theta_t^*\}$ and existence of moments of $W_t^* = w(\theta_{t-1}^*, e_t)$ while the second identifies the specific behavior we will use to verify stability of $\{X_t\}$. Proofs are in section 4.

THEOREM 3.1 *$\{\theta_t^*\}$ and $\{(\theta_t^*, W_t^*)\}$ are uniformly ergodic. Furthermore, with $\beta$ as in Assumption A.1 and $\alpha$ as in Assumption A.2(ii),*

$$\sup_{\theta\in\Theta} E\Big((W_t^*)^\zeta \;\Big|\; \theta_{t-1}^* = \theta\Big) < \infty \quad \textit{for any } \zeta \in (-\alpha, \beta].$$

*In particular, if $\pi$ is the stationary distribution for $\{\theta_t^*\}$ then*

$$\gamma = \int_\Theta E\Big(\log W_1^* \;\Big|\; \theta_0^* = \theta\Big)\,\pi(d\theta) = \int_\Theta E\Big(\log w(\theta, e_1)\Big)\,\pi(d\theta) \quad \textit{exists finite.} \tag{3.1}$$

In view of (3.1) and the uniform ergodicity of the collapsed process, there must exist a bounded solution $\nu$ to the Poisson equation

$$E\Big(\nu(\theta_1^*) - \nu(\theta) + \log w(\theta, e_1) \;\Big|\; \theta_0^* = \theta\Big) = \gamma \tag{3.2}$$

(Meyn and Tweedie (1993, Thm. 17.4.2)). This solution, unfortunately, is generally not continuous on $\Theta$. But we can show that it is uniformly continuous on $\Theta_\#$ (hence piecewise continuous on $\Theta$) and that will suffice.

THEOREM 3.2 *Let $\Theta_\#$ be as in Assumption A.4 and let $\pi$ be the stationary distribution for $\{\theta_t^*\}$.*

*(i) There exists a bounded, measurable function $\nu : \Theta \to \mathbb{R}$, uniformly continuous on $\Theta_\#$, such that (3.2) holds for all $\theta \in \Theta$.*

*(ii) Let $\Lambda$ be the class of measurable, positive functions on $\Theta$ that are bounded and bounded away from 0. If $\gamma < 0$ then there exists a unique $\kappa \in (0, \infty]$ such that for all $\zeta \in (0, \kappa)$*

$$E\Big(\lambda(\theta_1^*)(W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta\Big) \le \rho\lambda(\theta) \quad \textit{for all } \theta \in \Theta \textit{ and some } \rho \in (0,1), \lambda \in \Lambda, \tag{3.3}$$

*but not for any $\zeta > \kappa$. Furthermore, such $\lambda$ may be chosen to be uniformly continuous on $\Theta_\#$ (possibly by increasing the value of $\rho$).*

The Poisson equation (3.2) will be the basis for identifying the Lyapounov exponent of $\{X_t\}$ and (3.3) will be the basis for constructing the proof of ergodicity and the existence of moments. We now turn to these questions.

**3.2. Geometric Ergodicity.** Our approach is to verify Foster-Lyapounov drift conditions that are piggybacked on the collapsed process, specifically on (3.2) and (3.3). The first order of business is to identify the Lyapounov exponent for $\{X_t\}$. Again, the proofs are in section 4.

THEOREM 3.3 *Let $\gamma$ be as in (3.1). The Lyapounov exponent $\bar\gamma$ for $\{X_t\}$ is $\gamma$. Indeed,*

$$\lim_{n\to\infty} \lim_{||x||\to\infty} \Big| \frac{1}{n} E\Big(\log\Big(\frac{1 + ||X_n||}{1 + ||X_0||}\Big) \;\Big|\; X_0 = x\Big) - \gamma \Big| = 0. \tag{3.4}$$

We now turn to the primary result, a sharp condition for ergodicity. In addition to establishing the condition, we also determine which moments exist under stationarity. As indicated below, these two objectives can be attained simultaneously. Other approaches have been used to determine existence of moments, primarily for the basic GARCH model. These include the methods of He and Teräsvirta (1999a,b) and of Basrak, Davis and Mikosch (2002).

THEOREM 3.4 *Let $\gamma$ be as in (3.1).*

*(i) If $\gamma < 0$ then $\{X_t\}$ is geometrically ergodic. Furthermore, let $\kappa$ be as in Theorem 3.2(ii). For every $\zeta \in (0, \kappa)$ there exists a Foster-Lyapounov test function $V : \mathbb{X} \to \mathbb{R}_+$ satisfying*

*(a) there exist $\rho_0 < 1$ and finite $M_0$, $K_0$ such that*

$$E\Big(V(X_1) \;\Big|\; X_0 = x\Big) \le \rho_0 V(x) 1_{V(x)>M_0} + K_0 1_{V(x)\le M_0}, \quad \text{for all } x \in \mathbb{X}, \tag{3.5}$$

*and*

*(b) there exist finite, positive $d_1, d_2$ such that*

$$d_1(1 + ||x||)^\zeta \le V(x) \le d_2(1 + ||x||)^\zeta. \tag{3.6}$$

*Consequently, the stationary distribution for $\{X_t\}$ has finite $\zeta$-moment for every $\zeta \in (0, \kappa)$.*

*(ii) If $\gamma > 0$ and there exists an irreducibility measure $\phi$ for $\{X_t\}$ such that $\phi(\{x \in \mathbb{X} : ||x|| > M\}) > 0$ for all $M < \infty$, then $\{X_t\}$ is transient and, for some $\rho < 1$,*

$$P\Big(\lim_{n\to\infty} \rho^n ||X_n|| = \infty \;\Big|\; X_0 = x\Big) = 1, \quad \text{for all } x \in \mathbb{X}.$$

If $\{x \in \mathbb{X} : V(x) \le M\}$ is bounded in $\mathbb{R}^m$ for all finite $M$ then condition (3.5) is known (under Assumption A.3) to imply $V$-uniform ergodicity. That is, for some $\rho < 1$ and $K < \infty$,

$$\sup_{A\in\mathcal{B}(\mathbb{X})} |P(X_n \in A | X_0 = x) - \Pi(A)| \le K\rho^n V(x) \quad \text{for all } x \in \mathbb{X},\ n \ge 1,$$

where $\Pi$ is the stationary distribution (cf. Meyn and Tweedie (1993, Thm. 16.0.1)). This, in turn, implies geometric ergodicity and geometric mixing properties (Meyn and Tweedie (1993, Thm. 16.1.5)). It also implies $\int_{\mathbb{X}} V(x)\,\Pi(dx) < \infty$ which, in conjunction with (3.6), is what determines the existence of moments.

It is easy to conjecture, although we do not show it here, that the $\zeta$-moment does not exist for $\zeta > \kappa$. A more precise description of the probability tails for $\Pi$ would establish this.

**3.3. Evaluating the Stability Conditions.** Geometric ergodicity is a consequence of a negative Lyapounov exponent $\gamma$ which in turn is a parameter of the collapsed process. As suggested by (3.1),

$$\frac{1}{n}\sum_{t=1}^{n} \log W_t^* \to \gamma \quad \text{almost surely, as } n \to \infty.$$

Thus $\gamma$ may be estimated simply by simulating the collapsed process and obtaining the sample average of $\log W_t^*$. Since $\{(\theta_t^*, W_t^*)\}$ is uniformly ergodic, such an estimator will be easy to generate and very reliable.

Alternatively, $\gamma$ may be determined as a consequence of numerically solving (3.2). For example, suppose $e_1$ has density $f$. Define $\bar{\nu}_0(\theta) = q_0(\theta) = \int_{\mathbb{E}} \log w(\theta, u)\, f(u)\, du$ and let $\mu$

be a probability measure on $\Theta$. Then an iterative procedure, apparent from the proof of Theorem 3.2(i), is to compute numerically

$$m_n = \int_\Theta \bar{\nu}_n(\theta)\,\mu(d\theta), \quad \nu_n(\theta) = \bar{\nu}_n(\theta) - m_n \quad \text{and} \quad \bar{\nu}_{n+1}(\theta) = q_0(\theta) + \int_{\mathbb{E}} \nu_n(\eta(\theta,u))\,f(u)\,du,$$

and then to estimate $\gamma$ by either of

$$\bar{\gamma}_n = \sup_{\theta\in\Theta}(\bar{\nu}_n(\theta) - \nu_{n-1}(\theta)) \quad \text{or} \quad \underline{\gamma}_n = \inf_{\theta\in\Theta}(\bar{\nu}_n(\theta) - \nu_{n-1}(\theta)).$$

The purpose for computing and subtracting $m_n$ is to provide numerical stability. See Cline (2007a) for a discussion of this approach. We state the result as follows.

COROLLARY 3.5 *Define $\bar{\gamma}_n$, $\underline{\gamma}_n$ and $\nu_n$ as above. Then $\bar{\gamma}_n \downarrow \gamma$ and $\underline{\gamma}_n \uparrow \gamma$ as $n \to \infty$ and $\nu_n(\theta)$ converges uniformly to a solution of (3.2) that is uniformly continuous on $\Theta_\#$.*

In fact, if one's purpose (say, in modeling a time series) is merely to determine stability then one does not need to solve (3.2) or evaluate $\gamma$ precisely, as the following corollary shows.

COROLLARY 3.6 *If there exists a bounded function $\nu : \Theta \to \mathbb{R}$ such that*

$$\sup_{\theta\in\Theta} E\left(\nu(\theta_1^*) - \nu(\theta) + \log w(\theta, e_1) \mid \theta_0^* = \theta\right) < 0 \tag{3.7}$$

*then $\{X_t\}$ is geometrically ergodic.*

Analogous to the last corollary, the expression in (3.3) may be used to determine the existence of a $\zeta$-moment for the stationary distribution of $\{X_t\}$. See Theorem 3.4(i). (As (3.3) also implies that (3.7) holds with $\nu(\theta) = \frac{1}{\zeta}\log\lambda(\theta)$, by Jensen's inequality, it simultaneously verifies the existence of the stationary distribution.) To determine whether (3.3) holds for a particular choice of $\zeta$, one may iteratively compute

$$\begin{aligned}\tilde{\lambda}_n(\theta) &= \int_{\mathbb{E}} \lambda_{n-1}(\eta(\theta,u))(w(\theta,u))^\zeta\,f(u)\,du, \quad \rho_n = \sup_{\theta\in\Theta}\Big(\tilde{\lambda}_n(\theta)/\lambda_{n-1}(\theta)\Big),\\ \lambda_n(\theta) &= \tilde{\lambda}_n(\theta)/\int_\Theta \tilde{\lambda}_n(\tilde{\theta})\,\mu(d\tilde{\theta}),\end{aligned}$$

with $\lambda_0(\theta) \equiv 1$ and where $\rho_n < 1$ indicates the condition is satisfied.

Unfortunately, the moment condition *cannot* be determined by evaluating $E((W_1^*)^\zeta)$ under the stationary distribution of the collapsed process even though $E((W_1^*)^\zeta) < 1$ does imply that $\gamma < 0$ (also by Jensen's inequality).

## 4. Proofs.

**4.1. Analyzing the Collapsed Process.** Here we verify the needed properties of $\{\theta_t^*\}$. We start by proving uniform ergodicity of the collapsed chain. Recall, from section 1.1, that $\theta_t^* = \eta(\theta_{t-1}^*, e_t)$ and $W_t^* = w(\theta_{t-1}^*, e_t)$, where $w(\theta,u) = ||B(\theta,u)||$ and $\eta(\theta,u) = B(\theta,u)/||B(\theta,u)||$.

PROOF OF THEOREM 3.1 By Assumption A.3, $\{X_t^*\}$ is aperiodic and $\phi$-irreducible on $\mathbb{X}^*$. It easily follows that $\{\theta_t^*\}$ is aperiodic and $\phi$-irreducible on $\Theta$. Since $\Theta$ is bounded and

bounded away from 0, it is small for $\{X_t^*\}$, again by Assumption A.3. Thus, there exist $n_1 \geq 1$ and a non-trivial measure $\mu$ such that

$$P\Big(X_{n_1}^* \in A \;\Big|\; X_0^* = x\Big) \geq \mu(A) \quad \text{for all } x \in \Theta,\ A \in \mathcal{B}(\mathbb{X}).$$

That is,

$$\begin{aligned} P\Big(\theta_{n_1}^* \in A \;\Big|\; \theta_0^* = \theta\Big) &= P\Big(X_{n_1}^* \in A \times \mathbb{R}_+ \;\Big|\; X_0^* = \theta\Big) \\ &\geq \mu(A \times \mathbb{R}_+), \quad \text{for all } \theta \in \Theta,\ A \in \mathcal{B}(\Theta). \end{aligned} \tag{4.1}$$

Hence $\Theta$ is small for $\{\theta_t^*\}$. By Meyn and Tweedie (1993, Thm. 16.0.2)), $\{\theta_t^*\}$ is uniformly ergodic. Since $W_t^* = w(\theta_{t-1}^*, e_t)$ and $e_t$ is independent of $\theta_{t-1}^*$, it follows trivially that $\{(\theta_t^*, W_t^*)\}$ also is uniformly ergodic.

From Assumption A.2(i),

$$\sup_{\theta\in\Theta} E\Big((W_t^*)^\beta \;\Big|\; \theta_{t-1}^* = \theta\Big) = \sup_{\theta\in\Theta} E\Big(||B(\theta, e_1)||^\beta\Big) < \bar{b}^\beta E((1 + |e_1|)^\beta) < \infty. \tag{4.2}$$

From Assumption A.2(i,ii), on the other hand,

$$\begin{aligned} \sup_{\theta\in\Theta} P\Big(W_t^* < \delta \;\Big|\; \theta_{t-1}^* = \theta\Big) &= \sup_{\theta\in\Theta} P\Big(||B(\theta, e_1)|| < \delta\Big) \\ &\leq \sup_{\theta\in\Theta} \lim_{||x||\to\infty} P\Big(||B(\theta, e_1) + \tfrac{C(\theta||x||, e_1)}{||x||}|| \leq \delta\Big) \\ &\leq K\delta^\alpha, \end{aligned} \tag{4.3}$$

if $0 < \delta \leq 1$. Therefore, for any $\zeta \in (-\alpha, 0)$,

$$\begin{aligned} \sup_{\theta\in\Theta} E\Big((W_t^*)^\zeta \;\Big|\; \theta_{t-1}^* = \theta\Big) &\leq 1 + \int_1^\infty \sup_{\theta\in\Theta} P\Big(W_t^* < u^{1/\zeta} \;\Big|\; \theta_{t-1}^* = \theta\Big)\, du \\ &\leq 1 + K\int_1^\infty u^{\alpha/\zeta}\, du < \infty. \end{aligned} \tag{4.4}$$

It follows from (4.2) and (4.4) that $\{\log w(\theta, e_1)\}_{\theta\in\Theta}$ is uniformly integrable. Hence $\gamma$ exists and is finite. □

Before we continue, we have a lemma that will be used to ensure uniform continuity of the functions $\nu$ and $\lambda$ on $\Theta_\#$ .

LEMMA 4.1 *Suppose $q : \Theta \to \mathbb{R}$ is bounded on $\Theta$ and uniformly continuous on $\Theta_\#$. Then, for each $\zeta \in [0, \beta)$, $\bar{q}(\theta) = E\left(q(\theta_1^*)(W_1^*)^\zeta \mid \theta_0^* = \theta\right)$ is bounded on $\Theta$ and uniformly continuous on $\Theta_\#$.*

PROOF The boundedness of $\bar{q}(\theta)$ is obvious by Theorem 3.1.

Suppose $\epsilon > 0$. Since $\{q(\eta(\theta, e_1))(w(\theta, e_1))^\zeta\}_{\theta\in\Theta}$ is uniformly integrable, again by Theorem 3.1, and $q(\eta(\theta, u))(w(\theta, u))^\zeta \to \infty$ requires $|u| \to \infty$ (cf. Assumption A.2(i)), we may choose $M < \infty$ so that

$$\begin{aligned} E\left(|q(\theta_1^*)|(W_1^*)^\zeta 1_{|e_1|>M} \mid \theta_0^* = \theta\right) &= E\Big(|q(\eta(\theta, e_1))|(w(\theta, e_1))^\zeta 1_{|e_1|>M}\Big) \\ &< \epsilon/4. \end{aligned} \tag{4.5}$$

On the other hand, $\{B(\theta,u)\}_{|u|\le M}$ is equicontinuous on $\Theta_\#$ by Assumption A.4(i). This implies $\{q(\eta(\theta,u))(w(\theta,u))^\zeta\}_{|u|\le M}$ likewise is equicontinuous on $\Theta_\#$. Hence, there exists sufficiently small $\delta$ such that $|\theta-\theta'|<\delta$, $\theta,\theta'\in\Theta_\#$ implies

$$|q(\eta(\theta,u))(w(\theta,u))^\zeta - q(\eta(\theta',u))(w(\theta',u))^\zeta| < \epsilon/2 \quad \text{for all } |u|\le M.$$

Thus,

$$\begin{aligned}&\left|E\left(q(\theta_1^*)(W_1^*)^\zeta 1_{|e_1|\le M} \mid \theta_0^*=\theta\right) - E\left(q(\theta_1^*)(W_1^*)^\zeta 1_{|e_1|\le M} \mid \theta_0^*=\theta'\right)\right| \\ &\quad\le E\left(\left|q(\eta(\theta,e_1))(w(\theta,e_1))^\zeta - q(\eta(\theta',e_1))(w(\theta',e_1))^\zeta\right| 1_{|e_1|\le M}\right) \\ &\quad< \epsilon/2. \end{aligned} \tag{4.6}$$

Combining (4.5) and (4.6), we conclude that $|\theta-\theta'|<\delta$, $\theta,\theta'\in\Theta_\#$ implies $|\bar q(\theta)-\bar q(\theta')|<\epsilon$. This demonstrates that $\bar q(\theta)$ is uniformly continuous on $\Theta_\#$. □

We now establish the characteristic behavior of the collapsed process that is used to describe our conditions for the stability of the stochastic recursion process.

Proof of Theorem 3.2 (i) Define $q_0(\theta)=E(\log w(\theta,e_1))$, which is bounded on $\Theta$ and uniformly continuous on $\Theta_\#$ (similar to the proof of the previous lemma). Let $q_t(\theta)=E(q_0(\theta_t^*)\mid\theta_0^*=\theta)$. By the Markov property,

$$E(q_0(\theta_t^*)\mid\theta_0^*=\theta) = E(E(q_0(\theta_t^*)\mid\theta_1^*)\mid\theta_0^*=\theta) = E(q_{t-1}(\theta_1^*)\mid\theta_0^*=\theta)$$

and so, by Lemma 4.1 and induction, $q_t(\theta)$ is bounded on $\Theta$ and uniformly continuous on $\Theta_\#$ for each $t\ge 1$.

By (3.1) and the uniform ergodicity of $\{\theta_t^*\}$, there exists $r<1$ and $K<\infty$ such that

$$|q_t(\theta)-\gamma| = |E(q(\theta_t^*)\mid\theta_0^*=\theta)-\gamma| < Kr^t \quad \text{for all } t\ge 1 \text{ and all } \theta\in\Theta$$

(cf. Meyn and Tweedie (1993, Thm. 16.2.1)). It follows (see, e.g., Meyn and Tweedie (1993, Thm. 17.4.2) that $\nu(\theta)=\sum_{t=0}^\infty (q_t(\theta)-\gamma)$ is a bounded solution to the Poisson equation (3.2).

Now, fix $\epsilon>0$ and choose $n_1$ such that $\frac{4Kr^{n_1}}{1-\delta}\le\epsilon$. Then

$$\sum_{t=n_1+1}^\infty |q_t(\theta)-\gamma| < \epsilon/4 \quad \text{for all } \theta\in\Theta,$$

while $\sum_{t=0}^{n_1}(q_t(\theta)-\gamma)$ is uniformly continuous on $\Theta_\#$ by the comments above. So there exists $\delta>0$ such that $|\theta-\theta'|<\delta$, $\theta,\theta'\in\Theta_\#$ implies

$$\left|\sum_{t=0}^{n_1}(q_t(\theta)-\gamma) - \sum_{t=0}^{n_1}(q_t(\theta)-\gamma)\right| < \epsilon/2,$$

and hence $|\nu(\theta)-\nu(\theta')|<\epsilon$. Therefore, $\nu$ is uniformly continuous on $\Theta_\#$.

(ii) Now suppose $\gamma<0$ and choose $\epsilon\in(0,-\gamma)$ and $\nu$ as above. By Theorem 3.1, we may choose $M<\infty$ so that

$$\sup_{\theta\in\Theta}\frac{1}{\beta}E\left((e^{\beta(\nu(\theta_1^*)-\nu(\theta))}(W_1^*)^\beta - 1)1_{|\log W_1^*|>M} \;\middle|\; \theta_0^*=\theta\right) < \epsilon/3. \tag{4.7}$$

Observe that $\frac{1}{\zeta}(e^{\zeta v}-1) \downarrow v$, as $\zeta \downarrow 0$, uniformly on compact sets. Thus, since $\nu$ also is bounded,

$$\lim_{\zeta \downarrow 0} \sup_{\theta \in \Theta} E\Big( \Big| \frac{e^{\zeta\nu(\theta_1^*)-\zeta\nu(\theta)}(W_1^*)^\zeta - 1}{\zeta} - (\nu(\theta_1^*) - \nu(\theta) + \log W_1^*) \Big| 1_{|\log W_1^*| \le M} \;\Big|\; \theta_0^* = \theta \Big) = 0. \quad (4.8)$$

By (3.2), (4.7) and (4.8),

$$\sup_{\theta \in \Theta} \frac{1}{\zeta} E\Big( e^{\zeta\nu(\theta_1^*)-\zeta\nu(\theta)}(W_1^*)^\zeta - 1 \;\Big|\; \theta_0^* = \theta \Big) < \gamma + \epsilon < 0$$

for $\zeta$ small enough. Therefore, for some $\zeta > 0$, (3.3) holds with $\lambda(\theta) = e^{\zeta\nu(\theta)}$, $\rho = 1+\zeta(\gamma+\epsilon)$.

Now define $\kappa$ to be the supremum of all $\zeta > 0$ such that (3.3) holds for some $\lambda \in \Lambda$ and $\rho \in (0,1)$. Choose any such $\zeta$, $\rho$ and $\lambda$, and suppose $s \in (0,\zeta)$. Jensen's inequality implies

$$E\Big( (\lambda(\theta_1^*))^{s/\zeta}(W_1^*)^s \;\Big|\; \theta_0^* = \theta \Big) \le \Big( E\Big( \lambda(\theta_1^*)(W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) \Big)^{s/\zeta} \le (\rho\lambda(\theta))^{s/\zeta},$$

for all $\theta \in \Theta$. Thus, we may conclude there exists $\lambda \in \Lambda$ and $\rho \in (0,1)$ (both depending on $\zeta$) such that (3.3) holds, for every $\zeta \in (0,\kappa)$.

It remains to be shown that we can choose $\lambda$ to be uniformly continuous on $\Theta_\#$.

Assume $\lambda \in \Lambda$, $\rho < 1$ satisfy (3.3). By the Markov property,

$$\begin{aligned} E\Big( \lambda(\theta_t^*) \prod_{k=1}^t (W_k^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) &= E\Big( E\Big( \lambda(\theta_t^*)(W_t^*)^\zeta \;\Big|\; \theta_{t-1}^* \Big) \prod_{k=1}^{t-1} (W_k^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) \\ &\le \rho E\Big( \lambda(\theta_{t-1}^*) \prod_{k=1}^{t-1} (W_k^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) \\ &\le \cdots \le \rho^t \lambda(\theta), \quad \text{for all } \theta \in \Theta,\ t \ge 1. \end{aligned} \quad (4.9)$$

Again suppose $\epsilon \in (0, 1-\rho)$. Since $\lambda$ is bounded and bounded away from 0, it follows from (4.9) that, for large enough $n_2$,

$$\sup_{\theta \in \Theta} \left( E\Big( \prod_{k=1}^{n_2} (W_k^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) \right)^{1/n_2} < \rho + \epsilon < 1. \quad (4.10)$$

Next, let $\bar{q}_1(\theta) = E\left( (w(\theta, e_1))^\zeta \right)$, which is bounded. By Lemma 4.1, $\bar{q}_1$ is uniformly continuous on $\Theta_\#$. Also,

$$\inf_{\theta \in \Theta} \bar{q}_1(\theta) \ge \inf_{\theta \in \Theta} \left( E\left( (w(\theta, e_1))^{-\alpha/2} \right) \right)^{-2\zeta/\alpha} > 0, \quad (4.11)$$

by Jensen's inequality and Theorem 3.1. We define

$$\begin{aligned} \bar{q}_t(\theta) &= E\Big( \prod_{k=1}^t (W_k^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) = E\Big( E\Big( \prod_{k=1}^t (W_k^*)^\zeta \;\Big|\; \theta_1^*, W_1^* \Big) \;\Big|\; \theta_0^* = \theta \Big) \\ &= E\Big( E\Big( \prod_{k=2}^t (W_k^*)^\zeta \;\Big|\; \theta_1^* \Big) (W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big) \\ &= E\Big( \bar{q}_{t-1}(\theta_1^*)(W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta \Big). \end{aligned} \quad (4.12)$$

By Lemma 4.1, (4.11) and induction, each $\bar{q}_t(\theta)$ is bounded, bounded away from 0 and is uniformly continuous on $\Theta_\#$.

We now define $\bar{\lambda}(\theta) = \prod_{t=1}^{n_2-1}(\bar{q}_t(\theta))^{1/n_2}$. Then $\bar{\lambda} \in \Lambda$ and it likewise is uniformly continuous on $\Theta_\#$. Finally, by Hölder's inequality and (4.12),

$$\begin{aligned}
E\Big(\bar{\lambda}(\theta_1^*)(W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta\Big) &= E\Big(\prod_{t=1}^{n_2-1}\Big(\bar{q}_t(\theta_1^*)(W_1^*)^\zeta\Big)^{1/n_2}\Big((W_1^*)^\zeta\Big)^{1/n_2} \;\Big|\; \theta_0^* = \theta\Big) \\
&\le \prod_{t=1}^{n_2-1}\Big(E\Big(\bar{q}_t(\theta_1^*)(W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta\Big)\Big)^{1/n_2}\Big(E\Big((W_1^*)^\zeta \;\Big|\; \theta_0^* = \theta\Big)\Big)^{1/n_2} \\
&= \prod_{t=1}^{n_2-1}\Big(\bar{q}_{t+1}(\theta)\Big)^{1/n_2}\Big(\bar{q}_1(\theta)\Big)^{1/n_2} \\
&= \Big(\bar{q}_{n_2}(\theta)\Big)^{1/n_2}\prod_{t=1}^{n_2-1}\Big(\bar{q}_t(\theta)\Big)^{1/n_2} < (\rho+\epsilon)\bar{\lambda}(\theta),
\end{aligned}$$

for all $\theta \in \Theta$, with the final inequality by (4.10), thus completing the proof. □

**4.2. Showing Stability.** In this subsection we verify that $\gamma$ is the Lyapounov exponent and thus $\gamma < 0$ implies geometric ergodicity for $\{X_t\}$. We also prove that the stationary distribution, when it exists, has moments of all positive orders $\zeta < \kappa$. The method is similar to the "piggyback" proofs in Cline and Pu (2004) but greatly improved and clarified, as well as handling the distinctive features of a GARCH-type model (as opposed to ARCH-type). We start with a couple of lemmas.

LEMMA 4.2 *Let $\beta$ be as in Assumption A.1 and $\alpha$ as in Assumption A.2. Then*

$$\sup_{x\in\mathbb{X}} E\left(\left(\frac{1+||X_1||}{1+||x||}\right)^\zeta \;\Big|\; X_0 = x\right) < \infty \quad \text{for all } \zeta \in (-\alpha, \beta]. \tag{4.13}$$

*Consequently, $\bar{\gamma}$, as defined in (1.3), is finite.*

*Furthermore, for each $\epsilon > 0$, there exists $L < \infty$ such that*

$$P\Big(||X_{t_1}||/L \le ||X_{t_2}|| \le L||X_{t_1}||,\ 0 \le t_1 < t_2 \le n \;\Big|\; X_0 = x\Big) > 1 - n\epsilon \tag{4.14}$$

*for all $x \in \mathbb{X}$ such that $||x|| > L$ and all $n \ge 1$.*

PROOF For any $\zeta \in (0, \beta]$, the conclusion (4.13) is easy. From Assumption A.2(i) we know that

$$\bar{M} \stackrel{\text{def}}{=} \sup_{x\in\mathbb{X}}\sup_{u\in\mathbb{E}} \frac{1+||B(x/||x||,u)||x|| + C(x,u)||}{(1+||x||)(1+|u|)} < \infty. \tag{4.15}$$

Then

$$E\Big((1+||X_1||)^\zeta \;\Big|\; X_0 = x\Big) \le \bar{M}^\zeta(1+||x||)^\zeta E((1+|e_1|)^\zeta)$$

and (4.13) follows. Now suppose $-\alpha < \zeta < 0$ and let $\delta \in (0,1)$ be arbitrary. We have

$$\begin{aligned}
P\Big(1+||X_1|| < \delta(1+||x||) \;\Big|\; X_0 = x\Big) &\le P\Big(||B(x/||x||, e_1)||x|| + C(x,e_1)|| < \delta||x||\Big) \\
&< K\delta^\alpha,
\end{aligned}$$

when $||x|| > 1$, by Assumption A.2(ii). In the same way that (4.3) implies (4.4), this suffices to show (4.13).

Additionally, let

$$M_1 = \sup_{x\in\mathbb{X}} E\Big(\Big|\log\Big(\frac{1+||X_1||}{1+||x||}\Big)\Big| \;\Big|\; X_0 = x\Big) \le (\log \bar{M}) E(\log(1+|e_1|)),$$

which is finite. By conditioning on $X_{t-1}$,

$$\begin{aligned}\frac{1}{n}E\Big(\Big|\log\Big(\frac{1+||X_n||}{1+||x||}\Big)\Big| \;\Big|\; X_0 = x\Big) &\le \frac{1}{n}E\Big(\max_{0\le t_1<t_2\le n}\Big|\log\Big(\frac{1+||X_{t_2}||}{1+||X_{t_1}||}\Big)\Big| \;\Big|\; X_0 = x\Big)\\ &\le \frac{1}{n}\sum_{t=1}^{n} E\Big(\Big|\log\Big(\frac{1+||X_t||}{1+||X_{t-1}||}\Big)\Big| \;\Big|\; X_0 = x\Big)\\ &\le M_1 < \infty, \end{aligned} \tag{4.16}$$

for all $n \ge 1$ and all $x \in \mathbb{X}$. This shows that $|\bar{\gamma}| < \infty$.

Also, if $||x|| > L \ge 2e^{M_1/\epsilon}$ then Markov's inequality and (4.16) show

$$\begin{aligned} &P\Big(||X_{t_1}||/L \le ||X_{t_2}|| \le L||X_{t_1}||,\ 0 \le t_1 < t_2 \le n \;\Big|\; X_0 = x\Big)\\ &\quad\ge P\Big(2/L \le \frac{1+||X_{t_2}||}{1+||X_{t_1}||} \le L/2,\ 0 \le t_1 < t_2 \le n \;\Big|\; X_0 = x\Big)\\ &\quad\ge 1 - \frac{1}{\log(L/2)} E\Big(\max_{0\le t_1<t_2\le n}\Big|\log\Big(\frac{1+||X_{t_2}||}{1+||X_{t_1}||}\Big)\Big| \;\Big|\; X_0 = x\Big)\\ &\quad\ge 1 - \frac{nM_1}{\log(L/2)} \ge 1 - n\epsilon, \end{aligned}$$

verifying (4.14). □

The next lemma is where we piggyback a drift condition for $\{X_t\}$ on the Poisson equation for $\{\theta_t^*\}$.

LEMMA 4.3 *For each $\bar{\epsilon} > 0$ there exist $n \ge 1$ and a positive function $V_1(x)$ satisfying*

$$\limsup_{||x||\to\infty} \Big|E\Big(\log\frac{V_1(X_{n+1})}{V_1(X_n)} \;\Big|\; X_0 = x\Big) - \gamma\Big| < \bar{\epsilon}.$$

*Furthermore, for some positive $d_1$ and finite $d_2$, $V_1$ may be chosen to satisfy*

$$d_1(1+||x||) \le V_1(x) \le d_2(1+||x||).$$

PROOF Fix $\bar{\epsilon} > 0$ and set $M_1 = \sup_{x\in\mathbb{X}} E\Big(\Big|\log\frac{1+||B(x/||x||,e_1)||x||+C(x,e_1)||}{1+||x||}\Big|\Big)$ and $K_1 \ge \max(2\sup_{\theta\in\Theta}|\nu(\theta)|, 1)$. Fix $\epsilon = \frac{\bar{\epsilon}}{5+2M_1}$ and choose $\nu$ according to Theorem 3.2(i). According to Assumption A.4(ii,iii), choose $n \ge 1$ and $L_0 \ge 1$ such that

$$\sup_{||x||>L_0,\, x/||x||\in\Theta_\#} P\Big(\tilde{\theta}_{1,x} \notin \Theta_\# \text{ or } \tilde{\theta}^*_{1,x} \notin \Theta_\#\Big)\Big) < \epsilon/K_1 \tag{4.17}$$

and

$$\sup_{||x||>L_0} P\Big(\tilde{\theta}_n \notin \Theta_\# \;\Big|\; X_0 = x\Big) < \epsilon/K_1. \tag{4.18}$$

With $n$ fixed, Lemma 4.2 allows us to choose $L_1 \ge L_0$ to satisfy

$$\sup_{||x||>L_1} P\Big(||X_n|| \le ||x||/L_1 \;\Big|\; X_0 = x\Big) < \epsilon/K_1. \tag{4.19}$$

Define $\tilde{\theta}_{1,x} = \frac{B(x/||x||,e_1)||x||+C(x,e_1)}{||B(x/||x||,e_1)||x||+C(x,e_1)||}$ and $\tilde{\theta}^*_{1,x} = \frac{B(x/||x||,e_1)}{||B(x/||x||,e_1)||}$, that is, $\tilde{\theta}_1$ and $\tilde{\theta}^*_1$ as functions of $x$, respectively. Since $||\tilde{\theta}_{1,x} - \tilde{\theta}^*_{1,x}|| \to 0$ almost surely as $||x|| \to \infty$, by Assumption A.2(i), and since $\nu$ is uniformly continuous on $\Theta_\#$, it follows that

$$\limsup_{||x||\to\infty} E\Big(|\nu(\tilde{\theta}_{1,x}) - \nu(\tilde{\theta}^*_{1,x})|\, 1_{\tilde{\theta}_{1,x}\in\Theta_\#,\, \tilde{\theta}^*_{1,x}\in\Theta_\#}\Big) = 0.$$

Thus, by (4.17) there exists $L_2 \geq L_1$ such that

$$\begin{aligned}
&\sup_{||x||>L_2,\, x/||x||\in\Theta_\#} E\Big(|\nu(\tilde{\theta}_{1,x}) - \nu(\tilde{\theta}^*_{1,x})|\Big) \\
&\quad\leq \sup_{||x||>L_2,\, x/||x||\in\Theta_\#} E\Big(|\nu(\tilde{\theta}_{1,x}) - \nu(\tilde{\theta}^*_{1,x})|\, 1_{\tilde{\theta}_{1,x}\in\Theta_\#,\, \tilde{\theta}^*_{1,x}\in\theta_\#}\Big) \\
&\qquad + \sup_{||x||>L_2,\, x/||x||\in\Theta_\#} K_1 P\Big(\tilde{\theta}_{1,x} \notin \Theta_\# \text{ or } \tilde{\theta}^*_{1,x} \notin \Theta_\#\Big) \\
&\quad< 2\epsilon,
\end{aligned} \tag{4.20}$$

Analogous to the definition of $\tilde{\theta}^*_1$ prior to Assumption A.4, let $\tilde{\theta}^*_{n+1} = \eta(X_n/||X_n||, e_{n+1})$. Conditioning on $X_n$, (4.20) gives

$$\begin{aligned}
&\sup_{x\in\mathbb{X}} E\Big(|\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}^*_{n+1})|\, 1_{||X_n||>L_2,\, \tilde{\theta}_n\in\Theta_\#} \;\Big|\; X_0 = x\Big) \\
&\quad= \sup_{x\in\mathbb{X}} E\Big(E\Big(|\nu(\tilde{\theta}_{1,X_n}) - \nu(\tilde{\theta}^*_{1,X_n})| \;\Big|\; X_n\Big) 1_{||X_n||>L_2,\, \tilde{\theta}_n\in\Theta_\#} \;\Big|\; X_0 = x\Big) \\
&\quad< 2\epsilon.
\end{aligned} \tag{4.21}$$

Then, by combining (4.21) with (4.18) and (4.19),

$$\begin{aligned}
&\sup_{||x||>L_2^2} E\Big(|\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}^*_{n+1})| \;\Big|\; X_0 = x\Big) \\
&\quad\leq \sup_{||x||>L_2^2} E\Big(|\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}^*_{n+1})|\, 1_{||X_n||>L_2,\, \tilde{\theta}_n\in\Theta_\#} \;\Big|\; X_0 = x\Big) \\
&\qquad + \sup_{||x||>L_2^2} K_1 P\Big(\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \leq ||x||/L_2 \;\Big|\; X_0 = x\Big) \\
&\quad< 4\epsilon.
\end{aligned} \tag{4.22}$$

We also have $\frac{1+||B(x/||x||,e_1)||x||+C(x,e_1)||}{1+||x||} - w(x/||x||, e_1) \to 0$ almost surely, as $||x|| \to \infty$, so that by uniform integrability (cf. Theorem 3.1 and Lemma 4.2),

$$\limsup_{||x||\to\infty} E\Big(\Big|\log \frac{1+||B(x/||x||,e_1)||x||+C(x,e_1)||}{1+||x||} - \log w(x/||x||, e_1)\Big|\Big) = 0.$$

Thus, for some $L_3 \geq L_2$,

$$\sup_{||x||>L_3} E\Big(\Big|\log \frac{1+||B(x/||x||,e_1)||x||+C(x,e_1)||}{1+||x||} - \log w(x/||x||, e_1)\Big|\Big) < \epsilon. \tag{4.23}$$

Recall the definition of $M_1$ and observe that $\sup_{\theta\in\Theta} E(\log w(\theta, e_1)) \leq M_1$ as well. Using (4.19) and (4.23), condition on $X_n$ to obtain

$$\sup_{||x||>L_3^2} E\Big(\Big|\log \frac{1+||X_{n+1}||}{1+||X_n||} - \log w(\tilde{\theta}_n, e_{n+1})\Big| \;\Big|\; X_0 = x\Big)$$

$$\leq \sup_{||x||>L_3^2} E\Big(\Big|\log\frac{1+||X_{n+1}||}{1+||X_n||} - \log w(\tilde\theta_n, e_{n+1})\Big| 1_{||X_n||>L_3} \;\Big|\; X_0 = x\Big)$$
$$+ \sup_{||x||>L_3^2} 2M_1 P\Big(||X_n|| \leq L_3 \;\Big|\; X_0 = x\Big)$$
$$< (2M_1+1)\epsilon \tag{4.24}$$

(since $K_1 \geq 1$).

Conditioning on $X_n$ one last time, now with (3.2),

$$\sup_{x\in\mathbb{X}} \Big| E\Big(\nu(\tilde\theta^*_{n+1}) - \nu(\tilde\theta_n) + \log w(\tilde\theta_n, e_{n+1}) \;\Big|\; X_0 = x\Big) - \gamma\Big| = 0. \tag{4.25}$$

Combining (4.22), (4.24) and (4.25), we arrive at

$$\sup_{||x||>L_3^2} \Big| E\Big(\nu(\tilde\theta_{n+1}) - \nu(\tilde\theta_n) + \log\frac{1+||X_{n+1}||}{1+||X_n||} \;\Big|\; X_0 = x\Big) - \gamma\Big| < (5+2M_1)\epsilon.$$
$$= \bar\epsilon, \tag{4.26}$$

and the result is obtained with $V_1(x) = e^{\nu(x/||x||)}(1+||x||)$. □

We now can identify the Lyapounov exponent.

Proof of Theorem 3.3 We start with (4.26) and the constants chosen in its proof. Note that $n$ is fixed and depends on $\bar\epsilon$. Set $\bar L_0 = L_3^2$. By Lemma 4.2, there exists $\bar L_t \geq \bar L_{t-1}$ such that

$$\sup_{||x||>\bar L_t} P\Big(||X_t|| \leq L_3^2 \;\Big|\; X_0 = x\Big) \leq \tfrac{\bar\epsilon}{K_1+2M_1+|\gamma|}, \quad \text{for } t \geq 1. \tag{4.27}$$

Hence, using (4.26) and conditioning on $X_t$ in the manner of the previous proof,

$$\sup_{||x||>\bar L_t} \Big| E\Big(\Big(\nu(\tilde\theta_{n+t+1}) - \nu(\tilde\theta_{n+t}) + \log\frac{1+||X_{n+t+1}||}{1+||X_{n+t}||} - \gamma\Big) 1_{||X_t||>L_3^2} \;\Big|\; X_0 = x\Big)\Big| < \bar\epsilon, \tag{4.28}$$

$t \geq 1$, whereas by (4.27)

$$\sup_{||x||>\bar L_t} \Big| E\Big(\Big(\nu(\tilde\theta_{n+t+1}) - \nu(\tilde\theta_{n+t}) + \log\frac{1+||X_{n+t+1}||}{1+||X_{n+t}||} - \gamma\Big) 1_{||X_t||\leq L_3^2} \;\Big|\; X_0 = x\Big)\Big|$$
$$\leq (K_1 + 2M_1 + |\gamma|) \sup_{||x||>\bar L_t} P\Big(||X_t|| \leq L_3^2 \;\Big|\; X_0 = x\Big) < \bar\epsilon. \tag{4.29}$$

Combining (4.28) and (4.29) and summing over $t = 0, \ldots, k-1$, we obtain

$$\sup_{||x||>\bar L_{k-1}} \Big| E\Big(\nu(\tilde\theta_{n+k}) - \nu(\tilde\theta_n) + \log\frac{1+||X_{n+k}||}{1+||X_n||} \;\Big|\; X_0 = x\Big) - k\gamma\Big| < 2k\bar\epsilon, \tag{4.30}$$

In other words, (4.30) declares

$$\limsup_{||x||\to\infty} \Big| \frac{1}{k} E\Big(\nu(\tilde\theta_{n+k}) - \nu(\tilde\theta_n) + \log\frac{1+||X_{n+k}||}{1+||X_n||} \;\Big|\; X_0 = x\Big) - \gamma\Big| < 2\bar\epsilon,$$

for all $k \geq 1$. Since $\nu$ is bounded and $E\Big(\log\frac{1+||X_n||}{1+||X_0||} \;\Big|\; X_0 = x\Big)$ is bounded (as $n$ is fixed), we conclude

$$\limsup_{k\to\infty} \limsup_{||x||\to\infty} \Big| \frac{1}{k} E\Big(\log\frac{1+||X_{n+k}||}{1+||X_0||} \;\Big|\; X_0 = x\Big) - \gamma\Big| \leq 2\bar\epsilon.$$

Finally, $\bar{\epsilon}$ may be chosen arbitrarily and thus the conclusion (3.4) follows. □

Now we prove the principal result for stability and the existence of moments by piggybacking a drift condition on (3.3).

PROOF OF THEOREM 3.4 (i) We may suppose $\zeta \in (0, \kappa)$, $\lambda \in \Lambda$ and $\rho \in (0,1)$ satisfy (3.3), and $\lambda$ is uniformly continuous on $\Theta_\#$. It is well known that (3.5) and (3.6) imply geometric ergodicity and finite $\zeta$-moment of the stationary distribution, under Assumption A.3. So our proof will be to validate (3.5) and (3.6). To this end, let $\bar{\epsilon} \in (0, 1-\rho)$ and set $\rho_0 = \rho + \bar{\epsilon}$. Also, choose $K_2$ so that $1/K_2 < \lambda(\theta) < K_2$ for all $\theta$.

Note that (3.3) and Assumption A.2(i) imply $E(|e_1|^\beta) < \infty$ for all $\beta < \kappa$. Indeed, we may assume $\beta > \zeta$ in Theorem 3.1 and Lemma 4.2. Let $M_3 = \bar{M}^\zeta E((1+|e_1|)^\zeta)$, where $\bar{M}$ is defined in (4.15). Then

$$\sup_{x \in \mathbb{X}} E\Big((w(x/||x||, e_1))^\zeta\Big) \le \sup_{x \in \mathbb{X}} E\left(\left(\frac{1+||X_1||}{1+||x||}\right)^\zeta \;\Big|\; X_0 = x\right) \le M_3. \tag{4.31}$$

Now let $\epsilon = \frac{1}{32}\left(\frac{\bar{\epsilon}}{K_2^2 M_3}\right)^2$. According to Assumption A.4(ii,iii), fix $n$ and $L_0$ so that

$$\sup_{||x||>L_0,\, x/||x|| \in \Theta_\#} P\Big(\tilde{\theta}_1 \notin \Theta_\# \text{ or } \tilde{\theta}_1^* \notin \Theta_\# \;\Big|\; X_0 = x\Big) < \frac{\bar{\epsilon}}{8K_2^2 M_3} \tag{4.32}$$

and

$$\sup_{||x||>L_0} P\Big(\tilde{\theta}_n \notin \Theta_\# \;\Big|\; X_0 = x\Big) < \epsilon. \tag{4.33}$$

Next, select $L_1 \ge L_0$ so that (by Lemma 4.2)

$$\sup_{||x||>L_1} P\Big(||X_n|| \le ||x||/L_1 \;\Big|\; X_0 = x\Big) < \epsilon. \tag{4.34}$$

Then, using (4.31) and (4.32),

$$\begin{aligned}
&\sup_{||x||>L_1,\, x/||x|| \in \Theta_\#} E\Big(\Big|\frac{\lambda(\tilde{\theta}_1) - \lambda(\tilde{\theta}_1^*)}{\lambda(x/||x||)}\Big| \, (w(x/||x||, e_1))^\zeta 1_{\tilde{\theta}_1 \notin \Theta_\# \text{ or } \tilde{\theta}_1^* \notin \Theta_\#} \;\Big|\; X_0 = x\Big) \\
&\quad \le 2K_2^2 M_3 \sup_{||x||>L_1,\, x/||x|| \in \Theta_\#} P\Big(\tilde{\theta}_1 \notin \Theta_\# \text{ or } \tilde{\theta}_1^* \notin \Theta_\# \;\Big|\; X_0 = x\Big) \\
&\quad < \bar{\epsilon}/4.
\end{aligned} \tag{4.35}$$

Recall that, by Assumption A.2(i), $\frac{1+||B(x/||x||,e_1)||x||+C(x,e_1)||}{1+||x||} - w(x/||x||, e_1) \to 0$ almost surely, as $||x|| \to \infty$. By dominated convergence and the uniform integrability implied by Theorem 3.1 and Lemma 4.2, we get

$$\limsup_{||x||\to\infty} E\Big(\frac{\lambda(\tilde{\theta}_1)}{\lambda(x/||x||)}\Big|\frac{(1+||X_1||)^\zeta}{(1+||x||)^\zeta} - (w(x/||x||, e_1))^\zeta\Big| \;\Big|\; X_0 = x\Big) = 0. \tag{4.36}$$

Using the uniform continuity of $\lambda$ on $\Theta_\#$, on the other hand, and uniform integrability again,

$$\limsup_{||x||\to\infty} E\Big(\Big|\frac{\lambda(\tilde{\theta}_1) - \lambda(\tilde{\theta}_1^*)}{\lambda(x/||x||)}\Big| \, (w(x/||x||, e_1))^\zeta 1_{\tilde{\theta}_1 \in \Theta_\#, \tilde{\theta}_1^* \in \Theta_\#} \;\Big|\; X_0 = x\Big) = 0. \tag{4.37}$$

Therefore, from (3.3) and (4.35)–(4.37), there exists $L_2 \geq L_1$ such that

$$\sup_{||x||>L_2,\, x/||x|| \in \Theta_\#} E\Big( \frac{\lambda(\tilde{\theta}_1)}{\lambda(x/||x||)} \frac{(1+||X_1||)^\zeta}{(1+||x||)^\zeta} \;\Big|\; X_0 = x \Big) < \rho + \bar{\epsilon}/2. \tag{4.38}$$

Let $V_2(x) = \lambda(x/||x||)(1+||x||)^\zeta$. Thus, from (4.38), we have

$$E\Big(V_2(X_1) \;\Big|\; X_0 = x\Big) < (\rho + \bar{\epsilon}/2) V_2(x), \quad \text{for all } ||x|| > L_2,\, x/||x| \in \Theta_\#. \tag{4.39}$$

Observe that, by conditioning on $X_n$ and applying (4.31),

$$E\Big(V_2(X_{n+1}) \;\Big|\; X_0 = x\Big) \leq K_2^2 M_3 E\Big(V_2(X_n) \;\Big|\; X_0 = x\Big). \tag{4.40}$$

Now we fix $\beta \in (\zeta, 2\zeta)$ such that

$$\Big(E\Big((V_2(X_{n+1}))^{\beta/\zeta} \;\Big|\; X_0 = x\Big)\Big)^{\zeta/\beta} \leq 2E\Big(V_2(X_{n+1}) \;\Big|\; X_0 = x\Big). \tag{4.41}$$

For $||x|| > L_2^2$, (4.33), (4.34), (4.40), (4.41) and the choice of $\epsilon$ and $\beta$ yield

$$\begin{aligned}
&E\Big(V_2(X_{n+1}) 1_{\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \leq L_1} \;\Big|\; X_0 = x\Big) \\
&\quad \leq \Big(E\Big((V_2(X_{n+1}))^{\beta/\zeta} \;\Big|\; X_0 = x\Big)\Big)^{\zeta/\beta} \Big(P\Big(\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \leq L_1 \;\Big|\; X_0 = x\Big)\Big)^{1-\zeta/\beta} \\
&\quad \leq 2E\Big(V_2(X_{n+1}) \;\Big|\; X_0 = x\Big)(2\epsilon)^{1-\zeta/\beta} \\
&\quad < 2K_2^2 M_3 (2\epsilon)^{1/2} E\Big(V_2(X_n) \;\Big|\; X_0 = x\Big) \\
&\quad \leq (\bar{\epsilon}/2) E\Big(V_2(X_n) \;\Big|\; X_0 = x\Big).
\end{aligned} \tag{4.42}$$

Conditioning on $X_n$ once again, (4.39) and (4.42) yield, for $||x|| > L_2^2$,

$$\begin{aligned}
E\Big(V_2(X_{n+1}) \;\Big|\; X_0 = x\Big) &\leq E\Big((\rho + \bar{\epsilon}/2) V_2(X_n) + V_2(X_{n+1}) 1_{\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \leq L_1} \;\Big|\; X_0 = x\Big) \\
&< (\rho + \bar{\epsilon}) E\Big(V_2(X_n) \mid X_0 = x\Big) = \rho_0 E\Big(V_2(X_n) \;\Big|\; X_0 = x\Big).
\end{aligned} \tag{4.43}$$

The desired test function is $V(x) = E\Big(V_2(X_n) \;\Big|\; X_0 = x\Big)$ which, by Lemma 4.2 applied iteratively and the fact $\lambda$ is bounded and bounded away from 0, clearly satisfies (3.6). Consequently, we may select $M_0$ such that $V(x) > M_0$ implies $||x|| > L_2^2$, and then (3.6) confirms also

$$K_0 \stackrel{\text{def}}{=} \sup_{V(x) \leq M_0} E\Big(V(X_1) \;\Big|\; X_0 = x\Big) = \sup_{V(x) \leq M_0} E\Big(V_2(X_{n+1}) \;\Big|\; X_0 = x\Big) < \infty.$$

With (4.43), this completes the verification that (3.5) holds as well.

(ii) A positive Lyapounov exponent implies $\{X_t\}$ is transient by Theorem 3.3, Lemma 4.2 and Cline and Pu (2001, Thm. 2.2).

Indeed, from the arguments of Cline and Pu (2001), it follows that, for some $\rho < 1$, $\zeta \in (0,1)$, $K < \infty$ and $L < \infty$,

$$E\Big(\Big(\frac{1+||X_0||}{1+||X_n||}\Big)^\zeta \;\Big|\; X_0 = x\Big) < K\rho^n, \quad \text{for all } n \geq 1 \text{ and } ||x|| > L.$$

Hence, by Markov's inequality,

$$\begin{aligned} P\Big(\rho^n||X_n|| \le M \;\Big|\; X_0 = x\Big) &\le P\Big(\Big(\frac{1+||X_0||}{1+||X_n||}\Big) \ge \rho^n L/M \;\Big|\; X_0 = x\Big) \\ &\le (\rho^n L/M)^{-\zeta}\, E\Big(\Big(\frac{1+||X_0||}{1+||X_n||}\Big)^{\zeta} \;\Big|\; X_0 = x\Big) \\ &\le K(M/L)^{\zeta}\rho^{(1-\zeta)n}, \end{aligned} \tag{4.44}$$

for every $M \ge L$ and all $||x|| > L$.

Since $\{x : ||x|| \le L\}$ is small the irreducibility assumption ensures there exist $n_0 \ge 1$ and $\delta > 0$ such that $P\Big(||X_{n_0}|| > L \;\Big|\; X_0 = x\Big) \ge \delta$ for $||x|| \le L$. It is then easy to show by induction and the Markov property that, for any $k \ge 1$,

$$P\Big(||X_n|| \le L,\, n \le kn_0 \;\Big|\; X_0 = x\Big) \le P\Big(||X_{jn_0}|| \le L,\, j \le k \;\Big|\; X_0 = x\Big) \le (1-\delta)^k.$$

That is, $P\Big(||X_t|| \le L,\, t \le n \;\Big|\; X_0 = x\Big) \le (1-\delta)^{n/n_0 - 1}$ for all $n$. Combining this with (4.44), we have

$$\begin{aligned} &P\Big(\rho^n||X_n|| \le M \;\Big|\; X_0 = x\Big) \\ &\quad\le P\Big(||X_t|| \le L,\, t \le n/2 \;\Big|\; X_0 = x\Big) + \sum_{t=1}^{\lfloor n/2 \rfloor} E\Big(P\Big(\rho^n||X_n|| \le M \;\Big|\; X_t\Big) 1_{||X_t|| \le L} \;\Big|\; X_0 = x\Big) \\ &\quad\le (1-\delta)^{n/2n_0 - 2} + \frac{n}{2} K(M/L)^{\zeta}\rho^{(1-\zeta)(n/2-1)}., \end{aligned}$$

for all $x \in \mathbb{X}$, $M \ge L$. This and Borel-Cantelli thus imply $P\Big(\rho^n||X_n|| \to \infty \;\Big|\; X_0 = x\Big) = 1$ for all $x$. □

Proof of Corollary 3.5 Define $q_t(\theta)$ as in the proof of Theorem 3.2(i). It is then easy to show that

$$\nu_n(\theta) = \sum_{t=0}^{n} \Big( q_t(\theta) - \int_\Theta q_t(\theta)\, \mu(d\theta) \Big) = \sum_{t=0}^{n} (q_t(\theta) - \gamma) - \int_\Theta \sum_{t=0}^{n} (q_t(\theta) - \gamma)\, \mu(d\theta),$$

which converges uniformly, as $n \to \infty$, by the theorem's proof. From that proof it also is clear that the limit will be uniformly continuous on $\Theta_\#$.

Furthermore,

$$\bar\nu_n(\theta) - \nu_{n-1}(\theta) = E(\nu_{n-1}(\theta_1^*) - \nu_{n-1}(\theta) + \log W_1^* \mid \theta_0^* = \theta) \to \gamma, \quad \text{uniformly.}$$

This shows that the limit solves (3.2) and that both $\bar\gamma_n$ and $\underline{\gamma}_n$ converge to $\gamma$. Additionally,

$$\begin{aligned} \bar\gamma_n &= \sup_{\theta\in\Theta} (\bar\nu_n(\theta) - \nu_{n-1}(\theta)) = \sup_{\theta\in\Theta} q_n(\theta) \\ &= \sup_{\theta\in\Theta} \int_{\mathbb{E}} q_{n-1}(\eta(\theta,u))\, f(u)\, du \le \sup_{\theta\in\Theta} q_{n-1}(\theta) = \bar\gamma_{n-1}. \end{aligned}$$

Therefore $\bar\gamma_n \downarrow \gamma$ and, in like manner, $\underline{\gamma}_n \uparrow \gamma$. □

Proof of Corollary 3.6 Let $\pi$ be the stationary distribution for $\{\theta_t^*\}$. Then (3.7) implies

$$\gamma = \int_\Theta E(\log w(\theta, e_1))\, \pi(d\theta) = \int_\Theta E\left(\nu(\theta_1^*) - \nu(\theta) + \log w(\theta, e_1) \mid \theta_0^* = \theta\right) \pi(d\theta) < 0,$$

and geometric ergodicity follows by Theorem 3.4(i). □

**5. Verifying Assumptions.**

Here we sketch out verification of the regularity Assumptions A.2–A.4 for the threshold AR-GARCH model as defined in subsection 2.1. In particular, we will assume Assumption A.1 and the following.

(i) $e_t$ has a density $f$ which is bounded and locally bounded away from 0 on $\mathbb{R}$,

(ii) $b_{ki} > 0$, $k = 0, \ldots, p$, $i = 1, \ldots, N$,

(iii) $c_{1i} + \cdots + c_{qi} < 1$ and $c_{ki} > 0$, $k = 1, \ldots, q$, $i = 1, \ldots, N$, and

(iv) the thresholds (boundaries of $C_1, \ldots, C_N$) are affine.

These undoubtedly are not necessary assumptions and one may wish to weaken them slightly, especially (ii). But to do so would make the following arguments more intricate than they already are. Recall also that $\mathbb{X} = \mathbb{R}^p \times [0, \infty)^q$.

Throughout this section, we let $x = (y, s)$ where $y = (y_1, \ldots, y_p)$ and $s = (s_1, \ldots, s_q)$. We write $b(y) = b_{i(y)}(y)$, $b^*(y) = (b^2(y) - b_{0,i(y)})^{1/2}$ and $c(s, y) = c_{i(y)}(s)$. Set

$$\delta_0 = \inf_{||(y,s)||=1} ((b^*(y))^2 + c^2(s, y))^{1/2},$$

which is positive under (ii) and (iii) above.

Assumption A.2(i) clearly holds. Recall the expressions for $B(x, u)$ and $C(x, u)$ in section 2.1. Since (note $b(y) \geq b^*(y)$)

$$||B(x/||x||, e_1)||x|| + C(x, e_1)|| \geq ((b^*(y))^2 + c^2(s, y))^{1/2} \geq \delta_0 ||x||,$$

Assumption A.2(ii) is trivially true with $K = 1/\delta_0$ and $\alpha = 1$.

Since $\mathbb{X}$ is closed, Assumption A.3 holds if $\{X_t\}$ is an aperiodic $\phi$-irreducible $T$-chain on $\mathbb{X}$ and $\{X_t^*\}$ is an aperiodic $\phi$-irreducible $T$-chain on $\mathbb{X}^* = \mathbb{X} \backslash \{0\}$. (See Meyn and Tweedie (1993).) To verify these, we rely on the following result.

THEOREM 5.1 *Let $\{X_t\}$ be a Markov process on $\mathbb{X} \subset \mathbb{R}^m$ with transition kernel $T$ defined by $T(x, A) = P(X_1 \in A \mid X_0 = x)$ and suppose the following three conditions hold.*

*(i) For some $k \geq 1$, $T^k(x, \cdot)$ is absolutely continuous for all $x \in \mathbb{X}$ and hence for each $n \geq k$, $T^n(x, \cdot)$ is absolutely continuous with some density $g_n(x, \cdot)$.*

*(ii) For each $x \in \mathbb{X}$ there exists $n \geq k$ satisfying: there exists open $A \subset \mathbb{X}$ and $\delta > 0$ such that*

$$\inf_{\hat{x}: ||\hat{x} - x|| < \delta} g_n(\hat{x}, \tilde{x}) > 0 \quad \text{for } \tilde{x} \in A.$$

*(iii) There is $\tilde{x} \in \mathbb{X}$ satisfying: for each $x \in \mathbb{X}$ and $\delta > 0$, there exists $n \geq k$ such that*

$$P\Big(||X_n - \tilde{x}|| < \delta \;\Big|\; X_0 = x\Big) > 0 \quad \text{and} \quad P\Big(||X_{n+1} - \tilde{x}|| < \delta \;\Big|\; X_0 = x\Big) > 0.$$

*Then $\{X_t\}$ is an aperiodic $\phi$-irreducible $T$-chain.*

PROOF The first two conditions imply that $\{X_t\}$ is a $T$-chain by the argument in Meyn and Tweedie (1993, Prop. 6.2.4). The first inequality of the third condition states that $\tilde{x}$ is "reachable" which, combined with the above, implies $\{X_t\}$ is $\phi$-irreducible by Meyn and Tweedie (1993, Prop. 6.2.1). The irreducibility measure $\phi$ is absolutely continuous with respect to Lebesgure measure. The additional inequality of the third condition then implies aperiodicity. □

THEOREM 5.2 *Assume $e_t$ has a density $f$ which is locally bounded and locally bounded away from 0 on $\mathbb{R}$ and suppose $b_{0i} > 0$, $b_{pi} > 0$, $c_{qi} > 0$ and $c_{1i} + \cdots + c_{qi} < 1$ for all $i = 1, \ldots, N$. Then the threshold AR-GARCH(p, q) model of section 2.1 satisfies Assumption A.3.*

PROOF This is a sketch as the algebra involved can get tedious. We focus here on the basic process $\{(Y_t, S_t)\}$ but the argument is valid for the homogeneous process also. The main distinction is in showing that $\{X_t^*\}$ is a $T$-chain on $\mathbb{X}^*$ rather than on $\mathbb{X}$.

Let $\underline{b}_0 = \min_{i=1,\ldots,N} b_{0i}$ which is positive, and $\bar{b}_0 = \max_{i=1,\ldots,N} \max(\max_{j=1,\ldots,p} b_{ji}, 1)$. Suppose $Y_0 = y = (y_1, \ldots, y_p)$ and $S_0 = s = (s_1, \ldots, s_q)$. Then $0 < \underline{b}_0 \leq \sigma_t^2 \leq \bar{b}_0^t(1 + ||(y, s)||^2)$ for all $t \geq 1$. (For the homogeneous process, we instead let $d_0 = \min_{i=1,\ldots,N} \min_{j=1,\ldots,q} c_{ji} \in (0, 1)$. Then $\sigma_t^2 \geq d_0^t \max(s_1^2, \ldots, s_q^2) > 0$, for all $t \geq 1$.) Let $\sigma_1(y, s) = ((b(y))^2 + c^2(s, y))^{1/2}$. It is clear, then, that the conditional distribution of $\xi_1$, given $Y_0 = y$ and $S_0 = s$, has density

$$\frac{1}{\sigma_1(y, s)} f\left(\frac{u - a(y)}{\sigma_1(y, s)}\right),$$

which is both locally bounded and locally bounded away from 0.

Indeed, one may show iteratively that the conditional distribution of $(\xi_n, \ldots, \xi_1)$, given $Y_0 = y$ and $S_0 = s$, has a density $h_n((y, s), \cdot)$ which is both locally bounded and locally bounded away from 0, and if $\epsilon_t \in (0, 1)$, $t \geq 1$, then

$$\begin{aligned} \underline{h}_n((y, s), z_1, \ldots, z_n) \prod_{t=1}^{n} \epsilon_t \leq P\Big(\bigcap_{t \leq n} \{|\xi_t - z_t| < \epsilon_t\} \;\Big|\; Y_0 = y, S_0 = s\Big) \\ \leq \bar{h}_n((y, s), z_1, \ldots, z_n) \prod_{t=1}^{n} \epsilon_t, \end{aligned} \tag{5.1}$$

for some $\underline{h}_n$ and $\bar{h}_n$, both locally bounded and locally bounded away from 0 in $(z_1, \ldots, z_n, y, s)$.

We first discuss a condition to ensure the existence of a $k$-step transition density for some $k \geq p + q$. Throughout, we assume we are conditioning on $Y_0 = y$ and $S_0 = s$. The regime indices $I_1, \ldots, I_k$ are determined by $(\xi_{k-1}, \ldots, \xi_1, y)$. It follows that $(\sigma_k^2, \ldots, \sigma_1^2)$ is a linear transformation of $(\xi_{k-1}^2, \ldots, \xi_1^2)$, depending on $I_1, \ldots, I_k$ and $(y, s)$. We show by the analysis below that $(\xi_k^2, \ldots, \xi_{k-p+1}^2, \sigma_k^2, \ldots, \sigma_{k-q+1}^2)$ is a linear transformation of $(\xi_k^2, \ldots, \xi_1^2)$, again depending on $I_1, \ldots, I_k$ and $(y, s)$.

To be specific, let $Y_k^2 = (\xi_k^2, \ldots, \xi_{k-p+1}^2)$, $\tilde{Y}_{k-p}^2 = (\xi_{k-p}^2, \ldots, \xi_1^2)$, $S_k^2 = (\sigma_k^2, \ldots, \sigma_{k-q+1}^2)$ and $\tilde{S}_{k-q}^2 = (\sigma_{k-q}^2, \ldots, \sigma_1^2)$. Also let $y^2 = (y_1^2, \ldots, y_p^2) = Y_0^2$ and $s^2 = (s_1^2, \ldots, s_p^2) = S_0^2$. (All these are to be treated as column vectors.) We can define column vectors $B_{01}$ and $B_{02}$ and matrices $B_{11}$, $B_{12}$, $B_{22}$, $C_{11}$, $C_{12}$, $C_{22}$, $D_1$ and $D_2$, with elements determined by the parameters and the values of $I_1, \ldots, I_k$, so that

$$\begin{aligned} \begin{pmatrix} S_k^2 \\ \tilde{S}_{k-q}^2 \end{pmatrix} = \begin{pmatrix} B_{01} \\ B_{02} \end{pmatrix} + \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} Y_k^2 \\ \tilde{Y}_{k-p}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ D_1 \end{pmatrix} y^2 \\ + \begin{pmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{pmatrix} \begin{pmatrix} S_k^2 \\ \tilde{S}_{k-q}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ D_2 \end{pmatrix} s^2. \end{aligned}$$

Note that $C_{11}$ (which is $q \times q$) and $C_{22}$ (which is $k - q \times k - q$) are upper triangular matrices with 0's for the diagonal elements. Thus, $I - C_{11}$ and $I - C_{22}$ are invertible upper

triangular matrices. Let $F_1 = (I - C_{11})^{-1} B_{11}$ and $F_2 = (I - C_{11})^{-1}(B_{12} + C_{12}(I - C_{22})^{-1} B_{22})$. Therefore, it may be seen that

$$\begin{pmatrix} Y_k^2 \\ S_k^2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ F_1 & F_2 \end{pmatrix} \begin{pmatrix} Y_k^2 \\ \tilde{Y}_{k-p}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ G \end{pmatrix},$$

where $G$ is linear in $(y^2, s^2)$. Again, all this is conditional on $I_1, \ldots, I_k$.

Now let $k = p + 2q$. We want to show that $F_2$ (which is $q \times 2q$) has rank $q$. The rightmost $q$ columns of $B_{12}$ (which is $q \times 2q$) are 0's and, since $b_{pi} > 0$ and $c_{qi} > 0$ for all $i = 1, \ldots, N$, the rightmost $q$ columns of $B_{22}$ (which is $p + q \times 2q$) have rank $q$ and $C_{12}$ (which is $q \times p + q$) also has rank $q$. It follows that $F_2$ will have rank $q$ for any choice of the values of $I_1, \ldots, I_k$. This means for each $\Xi = (I_1, \ldots, I_k)$ there is a nonsingular transformation $H_\Xi$ and a vector $K_\Xi$ such that

$$\begin{pmatrix} Y_k^2 \\ S_k^2 \\ \tilde{Y}_q^2 \end{pmatrix} = H_\Xi \begin{pmatrix} Y_k^2 \\ \tilde{Y}_{k-p}^2 \end{pmatrix} + K_\Xi.$$

Now, we put $\mathcal{Z} = (Y_k, S_k, \tilde{Y}_q)$ and as a column vector and let $\mathcal{Z}^2$ denote the corresponding vector of squared terms. Let $v$ be a vector of $+1$'s and $-1$'s and set

$$\mathcal{Y}(\mathcal{Z}, v, \Xi) = v \cdot (H_\Xi^{-1}(\mathcal{Z}^2 - K_\Xi))^{1/2}$$

where the square root is taken componentwise and the multiplication by $v$ is also componentwise. Note that, if the regime indices match up correctly, $\mathcal{Y}(\mathcal{Z}, v, \Xi)$ is a possible value for $(Y_k, \tilde{Y}_{k-p})$. Also, we define $\mathcal{I}(Y_k, \tilde{Y}_{k-p})$ to be the regime indices $I_1, \ldots, I_k$ that are determined by $(Y_k, \tilde{Y}_{k-p})$. We can then explicitly give a density for $\mathcal{Z}$ (abusing the notation somewhat for clarity), namely

$$\sum_v \sum_{\Xi : \mathcal{I}(\mathcal{Y}(\mathcal{Z}, v, \Xi)) = \Xi} \det(H_\Xi^{-1}) h_k((y, s), \mathcal{Y}(\mathcal{Z}, v, \Xi)), \tag{5.2}$$

with the sum being taken over all possible combinations that give a legitimate set of regime indices. Integrating out the values for $\tilde{Y}_q$, we conclude that $(Y_k, S_k)$ has a conditional density $g_k((y, s), \cdot)$, given $(Y_0, S_0) = (y, s)$, (however messy) and condition (i) of Theorem 5.1 holds.

This density is not everywhere positive due to lower bounds on the values of the $\sigma_t$'s. However, from (5.1) and (5.2) we may deduce that for some open set $A \subset \mathbb{R}^p \times (0, \infty)^q$, possibly depending on $(y, s)$, $g_k((\tilde{y}, \tilde{s}), \tilde{x})$ is bounded away from 0 for $(\tilde{y}, \tilde{s})$ in a neighborhood of $(y, s)$, for each $\tilde{x}$ in $A$. This verifies condition (ii) of Theorem 5.1.

To show the existence of a reachable point, suppose $(1, \ldots, 1)$ is in the interior of cone $C_i$. (If it is not in the interior, the following can be modified for an arbitrary point in the interior of some cone.) Define $\bar{b}_i = b_{1i} + \cdots + b_{pi}$ and $\bar{c}_i = c_{1i} + \cdots + c_{qi}$ and let $\sigma_0^2 = \frac{b_{0i} + \bar{b}_i}{1 - \bar{c}_i}$. Suppose $\epsilon_t \in (0, 1)$, $t \geq 1$, and each is small enough that $|\xi_t - 1| < \epsilon_t$, implies $Y_t \in C_i$ for $t \geq p$. Then $|\sigma_t^2 - \sigma_0^2| < \delta_t$, $t = 1, \ldots, n-1$, further implies

$$|\sigma_n^2 - \sigma_0^2| < 3\bar{b}_i \max_{t=1,\ldots,p} \epsilon_{n-t} + \bar{c}_i \max_{t=1,\ldots,q} \delta_{n-t} \quad \text{for } t \geq \max(p, q).$$

Since $\bar{c}_i < 1$, choosing $\epsilon_t \to 0$ means we can also have $\delta_t \to 0$, whence it follows that $|\xi_t - 1| < \epsilon_t$ for all $t$ implies $X_n$ is in a neighborhood of $(1, \ldots, 1, \sigma_0, \ldots, \sigma_0)$ for large $n$. With (5.1), we conclude that for any $\epsilon > 0$

$$P\Big(\max_{t=1,\ldots,p} |\xi_{n-t} - 1| < \epsilon, \max_{t=1,\ldots,q} |\sigma_{n-t} - \sigma_0| < \epsilon \;\Big|\; Y_0 = y, S_0 = s\Big) > 0 \quad \text{for all } n \text{ large enough.}$$

This verifies that condition (iii) of Theorem 5.1 is satisfied. □

We finish by verifying the final assumption.

THEOREM 5.3 *Suppose the boundaries of $C_1, \ldots, C_N$ are contained in the union of affine planes*

$$H_j = \{y = (y_1, \ldots, y_p) : \sum_{i=1}^{p} h_{ji} y_i = 0\}, \quad j = 1, \ldots, \tilde{N}.$$

*Assume $e_t$ has a density $f$ which is bounded and suppose $b_{ki} > 0$, $k = 0, \ldots, p$, and $c_{ki} > 0$, $k = 1, \ldots, q$, for all $i = 1, \ldots, N$. Then the threshold AR-GARCH$(p, q)$ model of section 2.1 satisfies Assumption A.4.*

PROOF Let $n = \max(p, q)$ and suppose $\tilde{\theta}_0 = x/||x||$, $||x|| > 0$. The thresholds (in $\Theta$) depend only on the first $p$ components of $\theta$. For each $j = 1, \ldots, \tilde{N}$ define $k(j) = \min\{k \geq 1 : h_{jk} \neq 0\}$. There is no loss of generality to assume $h_{j,k(j)} = 1$. We denote $\theta = (\theta_1, \ldots, \theta_{p+q})$. For $m \leq k(j)$, define

$$\Delta_{jm}(\theta) = |h_{j,k(j)}\theta_m + \cdots + h_{jp}\theta_{p+m-k(j)}|,$$

so that the thresholds are given by $\Delta_{j,k(j)}(\theta) = 0$. We define $D_{jm} = \{\theta : \Delta_{jm}(\theta) > 0\}$ and $\Theta_\# = \bigcap_{j=1}^{\tilde{N}} \bigcap_{m=1}^{k(j)} D_{jm}$. Note that $\Theta_\#$ excludes the thresholds and is open and nonempty.

When $m = 1$,

$$\Delta_{j1}(\tilde{\theta}_1) = \Big|a(y) + (b^2(y) + c^2(s, y))^{1/2} e_1 + \sum_{i=k(j)+1}^{p} h_{ji} y_{i-k(j)}\Big| \, \frac{||x||}{||X_1||}$$

and $\Delta_{j1}(\tilde{\theta}_1^*)$ is similar with $a$ replaced by $a^*$ and $b$ replaced by $b^*$. Thus,

$$P(\Delta_{j1}(\tilde{\theta}_1) > 0 \mid X_0 = x) = P(\Delta_{j1}(\tilde{\theta}_1^*) > 0 \mid X_0 = x) = 1$$

for all $j$ and all $\theta \in \Theta$. Additionally, when $m > 1$, we have $\Delta_{jm}(\tilde{\theta}_1) = \Delta_{jm}(\tilde{\theta}_1^*) = \Delta_{j,m-1}(\theta)||x||/||X_1||$. Hence $\theta \in D_{j,m-1}$ implies each of $\tilde{\theta}_1 \in D_{jm}$ and $\tilde{\theta}_1^* \in D_{jm}$.

It follows from all the preceding, therefore, that $\theta \in \Theta_\#$ implies

$$P(\tilde{\theta}_1 \in \Theta_\# \mid X_0 = x) = P(\tilde{\theta}_1^* \in \Theta_\# \mid X_0 = x) = 1,$$

and $\theta \in \Theta$ implies $P(\tilde{\theta}_n \in \Theta_\# \mid X_0 = x) = 1$.

The above assumed $||x|| > 0$. If, however, this is not so then we still have $X_1 \neq 0$ with probability 1 and thus $P(\tilde{\theta}_{n+1} \in \Theta_\# \mid X_0 = x) = 1$.

Finally (see (2.5)), $\{B(\theta, u)\}_{|u| \leq M}$ clearly is equicontinuous on $\Theta_\#$, for any finite $M$. □

The above arguments can be modified to include the variable driven switching model of subsection 2.3, at least if the link functions $h_1, \ldots, h_N$ are sufficiently smooth (such as piecewise continuous on the regimes $C_1, \ldots, C_n$). In fact, by conditioning on the values of $\{v_t\}$, the argument in Theorem 5.2 can be used basically intact.

**References.**


Alsmeyer, G. and Fuh, C.-D. (2001). Limit theorems for iterated random functions by regenerative methods, *Stoch. Proc. Appl.* 96, 123–142.

Barnsley, M.F. and Elton, J.H. (1988). A new class of Markov processes for image encoding, *Adv. Appl. Probab.* 20, 14–32.

Basrak, B., Davis, R. and Mikosch, T. (2002). Regular variation of GARCH processes, *Stoch. Proc. Appl.* 99, 95–115.

Borkovec, M. and Klüppelberg, C. (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors, *Ann. Appl. Probab.* 11, 1220–1241.

Bougerol, P. and Picard, N. (1992a). Strict stationarity of generalized autoregressive processes, *Ann. Probab.* 20, 1714–1730.

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

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

Chan, K.-S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations, *Adv. Appl. Probab.* 17, 666–678.

Cline, D.B.H. (2007a). Evaluating the Lyapounov exponent and existence of moments for threshold AR-ARCH models, *J. Time Series Anal.* 28, 241–260.

Cline, D.B.H. (2007b). Regular variation of order 1 nonlinear AR-ARCH models, Statistics Dept., *Stoch. Proc. Appl.*, to appear.

Cline, D.B.H. and Pu, H.H. (1999). Geometric ergodicity of nonlinear time series, *Stat. Sinica* 9, 1103–1118.

Cline, D.B.H. and Pu, H.H. (2001). Geometric transience of nonlinear time series, *Stat. Sinica* 11, 273–287.

Cline, D.B.H. and Pu, H.H. (2004). Stability and the Lyaponouv exponent of threshold AR-ARCH models, *Ann. Appl. Probab.* 14, 1920–1949.

Diaconis, P. and Freedman, D. (1999). Iterated random functions, *SIAM Review* 41, 45–76.

Elton, J.H. (1990). A multiplicative ergodic theorem for Lipshitz maps, *Stoch. Proc. Appl.* 34, 39–47.

Franq, C. and Zakoïan, J.-M. (2005). The $L^2$-structures of standard and switching-regime GARCH models, *Stoch. Proc. Appl.* 115, 1577–1582.

Goldie, C.M. and Maller, R.A. (2000). Stability of perpetuities, *Ann Probab.* 28, 1195–1218.

Grincevičius, A.K. (1981). A random difference equation, *Lithuanian Math. J.* 20, 279–282.

He, C. and Teräsvirta, T. (1999a). Properties of moments of a family of GARCH processes, *J. Econom.* 92, 173–192.

He, C. and Teräsvirta, T. (1999b). Fourth moment structure of the GARCH$(p, q)$ process, *Econom. Theory* 15, 824–846.

Hwang, S.Y. and Woo, M.-J. (2001). Threshold ARCH(1) processes: asymptotic inference, *Stat. Prob. Letters* 53, 11–20.

Jarner, S.F. and Tweedie, R.L. (2001). Locally contractive iterated functions and stability of Markov chains, *J. Appl. Prob.* 38, 494–507.

Lanne, M. and Saikkonen, P. (2005). Nonlinear GARCH models for highly persistent volatility, *Econom. J.* 8, 251–276.

Lee, O. and Shin, D.W. (2005). On stationarity and $\beta$-mixing property of certain nonlinear GARCH$(p, q,)$ models, *Stat. Probab. Letters* 73, 25–35.

Li, C.W. and Li, W.K. (1996). On a double-threshold autoregressive heteroscedastic time series model, *J. Appl. Econom.* 11, 253–274.

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model, *J. Appl. Prob.* 36, 688–705.

Ling, S. (2004). Estimation and testing stationarity for double-autoregressive models, *J. Royal Stat. Soc.* B 66, 63–78.

Liu, J., Li, W.K. and Li, C.W. (1997). On a threshold autoregression with conditional heteroscedastic variances, *J. Stat. Plan. Inference* 62, 279–300.

Lu, Z. (1998b). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term, *Stat. Sinica* 8, 1205–1217.

Lu, Z. and Jiang, Z.(2001). $L_1$ geometric ergodicity of a multivariate nonlinear AR model with an ARCH term, *Stat. Prob. Letters* 51, 121–130.

Klüppelberg, C. and Pergamenchtchikov, P. (2004). The tail of the stationary distribution of a random coefficient AR($q$) model, *Ann. Appl. Probab.* 14, 971–1005.

Meitz, M. and Saikkonen, P. (2006). Stability of nonlinear AR-GARCH models, University of Helsinki, preprint.

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

Rabemanajara, R. and Zakoian, J.M. (1993). Threshold ARCH models and asymmetries in volatility, *J. Appl. Econometrics* 8, 31–49.

de Saporta, B. (2005). Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients, *Stoch. Proc. Appl.* 115, 1954–1978.

Steinsaltz, D. (1999). Locally contractive iterative functions, *Ann. Probab.* 27, 1952–1979.

Tong, H. (1981). A note on a Markov bilinear stochastic process in discrete time, *J. Time Series Anal.* 2, 279–284.

Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*, Notes in Statistics, No. 21, Springer Verlag (Heidelberg).

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

Tong, H. and Lim, K.S. (1980). Threshold autregression, limit cycles and cyclical data (with discussion), *J. Roy. Stat. Soc. (series B)* 42, 245–292.

Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables, *Adv. Appl. Prob.* 11, 750–783.

Wu, S. and Chen, R. (2007). Threshold variable selection and threshold variable driven switching autoregressive models, *Statistica Sinica* 17, 241–264.

Zhang, Z. and Tong, H. (2004). A note on stochastic difference equations and its application to GARCH models, *Chin. J. Appl. Probab.* 20, 259–269.