Adv. Appl. Prob. **39,** 462–491 (2007)
Printed in Northern Ireland

STABILITY OF NONLINEAR STOCHASTIC RECURSIONS WITH APPLICATION TO NONLINEAR AR-GARCH MODELS

DAREN B. H. CLINE,* *Texas A&M University*

Abstract

We characterize the Lyapunov 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 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–Lyapunov drift condition on certain characteristic behavior of the collapsed process.

Keywords: Stochastic recursion; GARCH model; ergodicity; stationary distribution

2000 Mathematics Subject Classification: Primary 60G10; 60J05
Secondary 62M10; 91B84

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

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

where $F\colon \mathbb{R}^m \times \mathbb{E} \to \mathbb{X} \subset \mathbb{R}^m$ and $\{e_t\}$ is a sequence of independent, identically distributed (i.i.d.) 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 an m-dimensional Markov (state space) representation of the form

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

where $0 < ||B(x/||x||, u)|| \leq \bar{b}(1 + |u|)$ and $||C(x, u)|| \leq \bar{c}(x)(1 + |u|)$ for finite $\bar{b}$ and $\bar{c}(x) = o(||x||)$. (See the 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.

Received 16 October 2006; revision received 25 March 2007.
* Postal address: Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.
Email address: dcline@stat.tamu.edu

We are interested in whether such a time series is ergodic, and specifically in 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

$$\bar{\gamma} := \liminf_{n\to\infty} \limsup_{||x||\to\infty} \frac{1}{n} \mathrm{E}\bigg(\log\bigg(\frac{1+||X_n||}{1+||X_0||}\bigg) \;\bigg|\; X_0 = x\bigg). \tag{1.3}$$

We call this constant the Lyapunov exponent of the process because it describes the long-term 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 Lyapunov exponent (see Cline and Pu (1999)). Conversely, under similar assumptions, a positive Lyapunov exponent ensures that $\{X_t\}$ is transient (see Cline and Pu (2001)).

Since $\bar{\gamma}$ is defined as a double limit, it is seemingly difficult to compute. Estimating it by simulation requires observing the time series during periods of 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 nonergodic) Markov process

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

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

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

The homogeneous process can be collapsed to Θ:

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

Also, let $W_t^* = w(\theta_{t-1}^*, e_t)$. The 'collapsed' process, $\{\theta_t^*\}$, is also Markov and typically 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 (see Theorem 3.3) that the Lyapunov exponent is equal to the stationary value of

$$\gamma := \mathrm{E}(\log(W_t^*)).$$

One very practical consequence of this result is that it provides a straightforward method to evaluate the Lyapunov exponent of $\{X_t\}$ (given the model and parameter values), because an expectation of 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 ζ-moment for $\{X_t\}$ exists when there is a bounded, positive function $\lambda(\theta)$ such that

$$\sup_{\theta\in\Theta} \mathrm{E}\bigg(\frac{\lambda(\theta_1^*)}{\lambda(\theta)}(W_t^*)^\zeta \;\bigg|\; \theta_0^* = \theta\bigg) < 1 \quad \text{for } \zeta > 0.$$

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 (see 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 from 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 sought here, is to investigate the tail properties of stationary distributions when they exist. We simply note that various authors have shown that stationary distributions typically have regularly varying probability tails, at least for the special cases studied so far (see Borkovec and Klüppelberg (2001), Basrak *et al.* (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 (see Lanne and Saikkonen (2005) and Lee and Shin (2005)), such assumptions can be unduly restrictive. Like threshold autoregression models, the parameter spaces of stable nonlinear AR-ARCH and AR-GARCH models can be unexpectedly large. See Cline (2007a) for a simple example. Similarly, while moment conditions have been ascertained for ordinary GARCH models (see, e.g. He and Teräsvirta (1999a), (1999b)), those conditions do not translate well to nonlinear models or even to GARCH models with autoregression terms.

Identifying conditions for ergodicity of the model in (1.1) is a problem of longstanding interest, and determining when the process is (geometrically) ergodic and mixing is useful for statistical inference. In this paper we establish geometric ergodicity conditions for the model in (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 the possibility that (1.1) is a Feller process, and they usually imply that 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), (1992b) showed that embedding pure-GARCH time series into (linear) random-coefficient 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 $\mathrm{E}(\log_+(||F(x, e_1) - G(e_1)x||))$ is bounded. The Lyapunov exponent turns out to be $\inf_{n\geq 1}(1/n)\,\mathrm{E}(\log(\|\prod_{t=1}^{n} G(e_t)\|))$. Since this is defined in terms of the matrix norm of a long-term product of random matrices, it is not necessarily easy to compute. (It coincides with (1.3) with few additional assumptions.)

Likewise, a condition is known (in a very general setting) when $F(x, \cdot)$ is Lipschitz continuous (see 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 Lyapunov exponent be negative:

$$\inf_{n\geq 1} \frac{1}{n}\,\mathrm{E}\biggl(\sup_{x,y\in\mathbb{X}} \log\biggl(\frac{d(X_n(x), X_n(y))}{d(x, y)}\biggr)\biggr) < 0,$$

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 is often difficult to compute, we may utilize a stronger condition such as

$$\mathrm{E}\biggl(\biggl(\sup_{x,y\in\mathbb{X}} \frac{d(F(x, e_1), F(y, e_1))}{d(x, y)}\biggr)^{\zeta}\biggr) < 1 \quad \text{for some } \zeta > 0.$$

Steinsaltz (1999) has an apparently weaker condition which posits the existence of a test function $\lambda(x)$ such that

$$\sup_{x\in\mathbb{X}} \mathrm{E}\biggl(\frac{\lambda(F(x, e_1))}{\lambda(x)} \limsup_{y\to x} \frac{d(F(x, e_1), F(y, e_1))}{d(x, y)}\biggr) < 1,$$

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, for time series that are nonlinear autoregressions with (roughly) constant noise variances, dynamical systems analysis and other methods have yielded important results. The literature is quite rich and we simply mention that it primarily began 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–Lyapunov 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 have all thwarted attempts to capitalize on comparisons with 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 Zakoïan (1993), Li and Li (1996), Liu *et al.* (1997), Lu (1998), Ling (1999), Hwang and Woo (2001), Lu and Jaing (2001), Lanne and Saikkonen (2005), Lee and Shin (2005), and Meitz and Saikkonen (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 that $G^*(e_1)$ satisfies a stability condition such as those for ordinary GARCH models or random-coefficient 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 Lyapunov exponent for the process. Even if the process is an ordinary high-order GARCH time series and fits with the results described in the literature above, we need to show how to compute the Lyapunov exponent efficiently.

In the remainder of this section, we will discuss the assumptions we use. Examples of nonlinear AR-GARCH models are given in Section 2. In Section 3 we present the results, in Section 4 we detail the proofs, and in Section 5 we verify that the assumptions will hold for typical threshold models.

1.3. Assumptions

In this section we list the various assumptions applied throughout the paper (except in Section 5 where the assumptions are validated for typical threshold models).

For generality, 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 model.

Let $||\cdot||$ denote the usual Euclidean norm on $\mathbb{R}^m$. For the sake of generality, $\{e_t\}$ will be an i.i.d. error sequence on some open set $\mathbb{E}$ contained in a Euclidean space. We let the norm for this be denoted by $|\cdot|$, 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 $\mathbb{X}^* = \mathbb{X} \setminus \{0\}$ 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 1.1. *The error sequence* $\{e_t\}$ *is i.i.d. with* $\mathrm{E}(|e_t|^\beta) < \infty$, *for some* $\beta > 0$.

The errors are often assumed to have a positive bounded density as well, mainly for the purpose of establishing Assumption 1.3 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 an ARCH- or GARCH-type model. The second part of Assumption 1.2 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 Lyapunov exponent is positive.

Assumption 1.2. (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* $u \in \mathbb{E}$ *and* $\theta \in \Theta$, *and* $||C(x, u)|| \leq \bar{c}(x)(1 + |u|)$, *for all* $u \in \mathbb{E}$ *and* $x \in \mathbb{X}$.

(ii) *For some finite* K *and* $\alpha \in (0, 1]$, $\mathrm{P}(||B(x/||x||, e_1)||x|| + C(x, e_1)|| < \delta||x||) < K\delta^\alpha$, *for all* $x \in \mathbb{X}$, *such that* $||x|| > 1$, *and for 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 because $B(x, e_1)$ is typically discontinuous in x (e.g. for threshold models) and has a singular

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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, Chapters 4–6)) for definitions of the properties mentioned here.

Assumption 1.3. *$\{X_t\}$ and $\{X_t^*\}$ are each aperiodic ϕ-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 1.3 that bounded sets are small is necessary essentially because the critical behavior for stability of a time series is off bounded sets, that is, for large values. This requirement 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 ϕ-irreducible, then this requirement is equivalent to saying that $\{X_t\}$ and $\{X_t^*\}$ are T-chains (cf. Meyn and Tweedie (1993, Theorem 6.2.5)).

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 θ_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 1.3, 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 1.4. *There exists a set $\Theta_\#$, open in $\Theta = \{x \in \mathbb{X}\colon ||x|| = 1\}$, such that*

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

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

(iii) *for every $\varepsilon > 0$ there exists $n \geq 1$ and $L < \infty$ such that $\mathrm{P}(\tilde{\theta}_n \in \Theta_\# \mid X_0 = x) > 1 - \varepsilon$, 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 1.4 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 1.3, we have to deal with the closure of $\Theta_\# \times [0, \infty)$ one way or another.

2. Examples

2.1. The threshold AR-GARCH model

The 'GARCH' in AR-GARCH can be interpreted in two different ways. In this subsection and the next we present the simpler interpretation, called 'double autoregression' by Ling (2004). In Section 2.3 we discuss the alternative 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 that $C_1, \dots, C_N$ are disjoint cones partitioning $\mathbb{R}^p$ and having, say, affine boundaries (which are the thresholds). Let $y = (y_1, \dots, y_p) \in \mathbb{R}^p$ and $s = (s_1, \dots, s_q) \in [0, \infty)^q$, and define

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

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

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

where

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

The index I_t is called the regime index.

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

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

Lee and Shin (2005) have recently verified that an extension of the standard assumption for the 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

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

By explicitly identifying the Lyapunov 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, as otherwise σ_t is certain to converge to infinity.

The state vector for the model in (2.2) is

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

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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 σ_t to take the value of zero, 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, $d_{0i}(y, s) := (b_i^2(y) + c_i^2(s))^{1/2} - ((b_i^*(y))^2 + c_i^2(s))^{1/2}$ is bounded. Let $i(y) = \sum_{i=1}^N i\, \mathbf{1}_{y \in C_i}$. Therefore, we can express the model in the form of (1.2) with $x = (y, s)$,

$$\begin{aligned} B(x, u) &= B\left(\frac{x}{||x||}, u\right)||x|| \\ &= (a_{i(y)}^*(y) + ((b_{i(y)}^*(y))^2 + c_{i(y)}^2(s))^{1/2}u,\ y_1, \dots, y_{p-1}, \\ &\qquad ((b_{i(y)}^*(y))^2 + c_{i(y)}^2(s))^{1/2},\ s_1, \dots, s_{q-1}), \end{aligned} \tag{2.5}$$

and $C(x, u) = (a_{0,i(y)} + d_{0,i(y)}(y, s)u,\ 0, \dots, 0,\ d_{0,i(y)}(y, s),\ 0, \dots, 0)$. 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 $c_{1i} + \cdots + c_{qi} < 1$. As σ_t is completely determined by past values, verifying Assumption 1.3 for the threshold AR-GARCH model is much more involved than it would be for a threshold AR-ARCH model (that is, with $q = 0$).

2.2. The nonlinear AR-GARCH model

More generally, a nonlinear AR-GARCH model may satisfy

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

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 $y = (y_1, \dots, y_p)$ and $s = (s_1, \dots, s_q)$. 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

$$\begin{aligned} B(x, u) &= B\left(\frac{x}{||x||}, u\right)||x|| \\ &= (a^*(y) + ((b^*(y))^2 + c^2(s))^{1/2}u,\ y_1, \dots, y_{p-1}, \\ &\qquad ((b^*(y))^2 + c^2(s))^{1/2},\ s_1, \dots, s_{q-1}), \end{aligned}$$

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

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 $k \le p$ could have

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

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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

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

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. The 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

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

where a_i, b_i and c_i are as in Section 2.1 and $I_t = \sum_{i=1}^N i\, \mathbf{1}_{(\xi_{t-1},\ldots,\xi_{t-p})\in C_i}$. The state vector is now (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 before in Section 2.1, we may express (for $x = (y, z, s)$)

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

Our results apply equally well to models of this variety. The only difficulty might be the extra effort required to verify Assumption 1.3 and Assumption 1.4.

2.4. The 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 also fits into our framework. In the ordinary threshold model of Section 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 (2007) 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

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

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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

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

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

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

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. The Markov regime-switching model

Regime-switching models have traditionally 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 θ_t^* and Θ be defined in terms of (Y_t, S_t) just as before and let $\mathbb{I} = \{1, \dots, N\}$ 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 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 $h_j((\theta, i)) = \mathrm{P}(I_t = j \mid I_{t-1} = i)$ for the latter.

Francq 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 λ of the next section.

3. Main results

3.1. The collapsed process

We remind the reader that Assumptions 1.1–1.4 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 $\{X_t\}$. The proofs are given in Section 4.

Theorem 3.1. *$\{\theta_t^*\}$ and $\{(\theta_t^*, W_t^*)\}$ are uniformly ergodic. Furthermore, with β as in Assumption 1.1 and α as in Assumption 1.2(ii),*

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

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

$$\gamma = \int_{\Theta} \mathrm{E}(\log W_1^* \mid \theta_0^* = \theta)\pi(\mathrm{d}\theta) = \int_{\Theta} \mathrm{E}(\log w(\theta, e_1))\pi(\mathrm{d}\theta) \tag{3.1}$$

exists finite.

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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

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

(Meyn and Tweedie (1993, Theorem 17.4.2)). This solution, unfortunately, is generally not continuous on Θ. However, we can show that it is uniformly continuous on $\Theta_\#$ (hence piecewise continuous on Θ) and that will be sufficient.

Theorem 3.2. *Let $\Theta_\#$ be as in Assumption 1.4 and let π be the stationary distribution for $\{\theta_t^*\}$.*

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

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

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

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

The Poisson equation, (3.2), will be the basis for identifying the Lyapunov 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–Lyapunov drift conditions that are piggybacked on the collapsed process, specifically on (3.2) and (3.3). We begin by identifying the Lyapunov exponent for $\{X_t\}$. Again, the proofs are given in Section 4.

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

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

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), (1999b) and of Basrak *et al.* (2002).

Theorem 3.4. *Let γ be as in (3.1).*

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

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

$$\mathrm{E}(V(X_1) \mid X_0 = x) \le \rho_0 V(x)\, \mathbf{1}_{V(x) > M_0} + K_0\, \mathbf{1}_{V(x) \le M_0} \quad \textit{for all } x \in \mathbb{X}, \tag{3.5}$$

and

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

(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* ζ*-moment for every* $\zeta \in (0, \kappa)$.

(ii) *If* $\gamma > 0$ *and there exists an irreducibility measure* ϕ *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* $\rho < 1$*,*

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

If $\{x \in \mathbb{X}: V(x) \le M\}$ is bounded in $\mathbb{R}^m$ for all finite M then (3.5) is known (under Assumption 1.3) to imply V-uniform ergodicity. That is, for some $\rho < 1$ and $K < \infty$,

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

where Π is the stationary distribution (cf. Meyn and Tweedie (1993, Theorem 16.0.1)). This, in turn, implies geometric ergodicity and geometric mixing properties (Meyn and Tweedie (1993, Theorem 16.1.5)). It also implies $\int_{\mathbb{X}} V(x)\Pi(\mathrm{d}x) < \infty$ 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 ζ-moment does not exist for $\zeta > \kappa$. A more precise description of the probability tails for Π would establish this.

3.3. Evaluating the stability conditions

Geometric ergodicity is a consequence of a negative Lyapunov exponent γ, which in turn is a parameter of the collapsed process. As suggested by (3.1),

$$\frac{1}{n}\sum_{t=1}^{n} \log W_t^* \to \gamma$$

almost surely, as $n \to \infty$. Thus, γ 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 will be very reliable.

Alternatively, γ may be determined as a consequence of solving (3.2) numerically. For example, suppose e_1 has density f. Define $\bar{\nu}_0(\theta) = q_0(\theta) = \int_{\mathbb{E}} \log w(\theta, u) f(u)\,\mathrm{d}u$, and let μ be a probability measure on Θ. Then an iterative procedure, apparent from the proof of Theorem 3.2(i), is to compute numerically

$$m_n = \int_\Theta \bar{\nu}_n(\theta)\mu(\mathrm{d}\theta), \qquad \nu_n(\theta) = \bar{\nu}_n(\theta) - m_n,$$

and

$$\bar{\nu}_{n+1}(\theta) = q_0(\theta) + \int_{\mathbb{E}} \nu_n(\eta(\theta, u)) f(u)\,\mathrm{d}u,$$

and then to estimate γ using either

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

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 3.1. *Define $\bar{\gamma}_n$, $\underline{\gamma}_n$ and ν_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 our purpose (say, in modeling a time series) is merely to determine stability then we do not need to solve (3.2) or evaluate γ precisely, as the following corollary shows.

Corollary 3.2. *If there exists a bounded function $\nu\colon \Theta \to \mathbb{R}$ such that*

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

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

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

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

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

4. Proofs

4.1. Analyzing the collapsed process

In this subsection we verify the properties of $\{\theta_t^*\}$ that we require. We begin 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 1.3, $\{X_t^*\}$ is aperiodic and ϕ-irreducible on $\mathbb{X}^*$. It easily follows that $\{\theta_t^*\}$ is aperiodic and ϕ-irreducible on Θ. Since Θ is bounded and bounded away from zero, it is small for $\{X_t^*\}$, again by Assumption 1.3. Thus, there exist $n_1 \geq 1$ and a nontrivial measure μ such that

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

That is,

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

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

From Assumption 1.2(i),

$$\sup_{\theta\in\Theta} \mathrm{E}((W_t^*)^\beta \mid \theta_{t-1}^* = \theta) = \sup_{\theta\in\Theta} \mathrm{E}(||B(\theta, e_1)||^\beta) < \bar{b}^\beta\, \mathrm{E}((1 + |e_1|)^\beta) < \infty. \tag{4.1}$$

From Assumption 1.2(i) and (ii), on the other hand,

$$\begin{aligned} \sup_{\theta\in\Theta} \mathrm{P}(W_t^* < \delta \mid \theta_{t-1}^* = \theta) &= \sup_{\theta\in\Theta} \mathrm{P}(||B(\theta, e_1)|| < \delta) \\ &\le \sup_{\theta\in\Theta} \lim_{||x||\to\infty} \mathrm{P}\Big(||B(\theta, e_1) + \frac{C(\theta||x||, e_1)}{||x||}|| \le \delta\Big) \\ &\le K\delta^\alpha, \end{aligned} \tag{4.2}$$

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

$$\begin{aligned} \sup_{\theta\in\Theta} \mathrm{E}((W_t^*)^\zeta \mid \theta_{t-1}^* = \theta) &\le 1 + \int_1^\infty \sup_{\theta\in\Theta} \mathrm{P}(W_t^* < u^{1/\zeta} \mid \theta_{t-1}^* = \theta)\, \mathrm{d}u \\ &\le 1 + K \int_1^\infty u^{\alpha/\zeta}\, \mathrm{d}u < \infty. \end{aligned} \tag{4.3}$$

It follows from (4.1) and (4.3) that $\{\log w(\theta, e_1)\}_{\theta\in\Theta}$ is uniformly integrable. Hence, γ exists and is finite.

Before we continue, the following lemma will be used to ensure uniform continuity of the functions ν and λ on $\Theta_\#$.

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

Proof. The boundedness of $\bar{q}(\theta)$ follows from Theorem 3.1. Suppose $\varepsilon > 0$. Since $\{q(\eta(\theta, e_1))(w(\theta, e_1))^\zeta\}_{\theta\in\Theta}$ is uniformly integrable, by Theorem 3.1, and

$$q(\eta(\theta, u))(w(\theta, u))^\zeta \to \infty$$

requires that $|u| \to \infty$ (cf. Assumption 1.2(i)), we may choose $M < \infty$ such that

$$\mathrm{E}(|q(\theta_1^*)|(W_1^*)^\zeta\, \mathbf{1}_{|e_1|>M} \mid \theta_0^* = \theta) = \mathrm{E}(|q(\eta(\theta, e_1))|(w(\theta, e_1))^\zeta\, \mathbf{1}_{|e_1|>M}) < \frac{\varepsilon}{4}. \tag{4.4}$$

On the other hand, $\{B(\theta, u)\}_{|u|\le M}$ is equicontinuous on $\Theta_\#$ by Assumption 1.4(i). This implies $\{q(\eta(\theta, u))(w(\theta, u))^\zeta\}_{|u|\le M}$ is likewise equicontinuous on $\Theta_\#$. Hence, there exists sufficiently small δ such that $|\theta - \theta'| < \delta$, where $\theta,\ \theta' \in \Theta_\#$, implies

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

Thus,

$$\begin{aligned} &|\,\mathrm{E}(q(\theta_1^*)(W_1^*)^\zeta\, \mathbf{1}_{|e_1|\le M} \mid \theta_0^* = \theta) - \mathrm{E}(q(\theta_1^*)(W_1^*)^\zeta\, \mathbf{1}_{|e_1|\le M} \mid \theta_0^* = \theta')| \\ &\qquad \le \mathrm{E}(|q(\eta(\theta, e_1))(w(\theta, e_1))^\zeta - q(\eta(\theta', e_1))(w(\theta', e_1))^\zeta|\, \mathbf{1}_{|e_1|\le M}) < \frac{\varepsilon}{2}. \end{aligned} \tag{4.5}$$

Combining (4.4) and (4.5), we conclude that $|\theta - \theta'| < \delta$, where $\theta,\ \theta' \in \Theta_\#$, implies $|\bar{q}(\theta) - \bar{q}(\theta')| < \varepsilon$. This demonstrates that $\bar{q}(\theta)$ is uniformly continuous on $\Theta_\#$.

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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 $q_0(\theta) = \mathrm{E}(\log w(\theta, e_1))$, which is bounded on Θ and uniformly continuous on $\Theta_\#$ (similar to the proof of the previous lemma). Let $q_t(\theta) = \mathrm{E}(q_0(\theta_t^*) \mid \theta_0^* = \theta)$. By the Markov property,

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

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

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

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

(cf. Meyn and Tweedie (1993, Theorem 16.2.1)). It follows (see, e.g. Meyn and Tweedie (1993, Theorem 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 $\varepsilon > 0$ and choose n_1 such that $4Kr^{n_1}/(1-\delta) \leq \varepsilon$. Then

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

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

$$\left| \sum_{t=0}^{n_1}(q_t(\theta) - \gamma) - \sum_{t=0}^{n_1}(q_t(\theta') - \gamma) \right| < \frac{\varepsilon}{2},$$

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

(ii) Now suppose $\gamma < 0$ and choose $\varepsilon \in (0, -\gamma)$ with ν as above. By Theorem 3.1, we may choose $M < \infty$ such that

$$\sup_{\theta \in \Theta} \frac{1}{\beta} \mathrm{E}((\exp(\beta(\nu(\theta_1^*) - \nu(\theta)))(W_1^*)^\beta - 1)\, \mathbf{1}_{|\log W_1^*| > M} \;\mid \theta_0^* = \theta) < \frac{\varepsilon}{3}. \tag{4.6}$$

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

$$\limsup_{\zeta \downarrow 0}\sup_{\theta \in \Theta} \mathrm{E}\left(\left| \frac{\exp(\zeta\nu(\theta_1^*) - \zeta\nu(\theta))(W_1^*)^\zeta - 1}{\zeta} - (\nu(\theta_1^*) - \nu(\theta) + \log W_1^*) \right| \times \mathbf{1}_{|\log W_1^*| \leq M} \;\mid \theta_0^* = \theta \right) = 0. \tag{4.7}$$

By (3.2), (4.6), and (4.7),

$$\sup_{\theta \in \Theta} \frac{1}{\zeta} \mathrm{E}(\exp(\zeta\nu(\theta_1^*) - \zeta\nu(\theta))(W_1^*)^\zeta - 1 \mid \theta_0^* = \theta) < \gamma + \varepsilon < 0,$$

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

Now define κ 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 ζ, ρ, and λ, and suppose $s \in (0, \zeta)$. Jensen's inequality implies

$$\mathrm{E}((\lambda(\theta_1^*))^{s/\zeta}(W_1^*)^s \mid \theta_0^* = \theta) \le (\mathrm{E}(\lambda(\theta_1^*)(W_1^*)^\zeta \mid \theta_0^* = \theta))^{s/\zeta} \le (\rho\lambda(\theta))^{s/\zeta}$$

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

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

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

$$\begin{aligned}
\mathrm{E}\biggl(\lambda(\theta_t^*)\prod_{k=1}^{t}(W_k^*)^\zeta \biggm| \theta_0^* = \theta\biggr) &= \mathrm{E}\biggl(\mathrm{E}(\lambda(\theta_t^*)(W_t^*)^\zeta \mid \theta_{t-1}^*)\prod_{k=1}^{t-1}(W_k^*)^\zeta \biggm| \theta_0^* = \theta\biggr) \\
&\le \rho\,\mathrm{E}\biggl(\lambda(\theta_{t-1}^*)\prod_{k=1}^{t-1}(W_k^*)^\zeta \biggm| \theta_0^* = \theta\biggr) \\
&\le \cdots \le \rho^t\lambda(\theta) \quad \text{for all } \theta \in \Theta \text{ and } t \ge 1.
\end{aligned} \tag{4.8}$$

Again suppose $\varepsilon \in (0, 1 - \rho)$. As λ is bounded and bounded away from zero, it follows from (4.8) that, for large enough n_2,

$$\sup_{\theta\in\Theta}\biggl(\mathrm{E}\biggl(\prod_{k=1}^{n_2}(W_k^*)^\zeta \biggm| \theta_0^* = \theta\biggr)\biggr)^{1/n_2} < \rho + \varepsilon < 1. \tag{4.9}$$

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

$$\inf_{\theta\in\Theta}\bar{q}_1(\theta) \ge \inf_{\theta\in\Theta}(\mathrm{E}((w(\theta, e_1))^{-\alpha/2}))^{-2\zeta/\alpha} > 0, \tag{4.10}$$

by Jensen's inequality and Theorem 3.1. We define

$$\begin{aligned}
\bar{q}_t(\theta) &= \mathrm{E}\biggl(\prod_{k=1}^{t}(W_k^*)^\zeta \biggm| \theta_0^* = \theta\biggr) \\
&= \mathrm{E}\biggl(\mathrm{E}\biggl(\prod_{k=1}^{t}(W_k^*)^\zeta \biggm| \theta_1^*,\ W_1^*\biggr) \biggm| \theta_0^* = \theta\biggr) \\
&= \mathrm{E}\biggl(\mathrm{E}\biggl(\prod_{k=2}^{t}(W_k^*)^\zeta \biggm| \theta_1^*\biggr)(W_1^*)^\zeta \biggm| \theta_0^* = \theta\biggr) \\
&= \mathrm{E}(\bar{q}_{t-1}(\theta_1^*)(W_1^*)^\zeta \mid \theta_0^* = \theta).
\end{aligned} \tag{4.11}$$

By Lemma 4.1, (4.10), and induction, each $\bar{q}_t(\theta)$ is bounded, bounded away from zero, 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 is likewise uniformly continuous on $\Theta_{\#}$. Finally, by Hölder's inequality and (4.11),

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

for all $\theta \in \Theta$, with the final inequality following from (4.9), thus completing the proof.

4.2. Showing stability

In this subsection we verify that γ is the Lyapunov exponent and, thus, that $\gamma < 0$ 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 is greatly improved and clarified, and handles the distinctive features of a GARCH-type model (as opposed to ARCH-type). We begin with a couple of lemmas.

Lemma 4.2. *Let β be as in Assumption 1.1 and α as in Assumption 1.2. Then*

$$\sup_{x \in \mathbb{X}} \mathrm{E}\bigg(\bigg(\frac{1 + ||X_1||}{1 + ||x||}\bigg)^{\zeta} \,\bigg|\, X_0 = x\bigg) < \infty \quad \text{for all } \zeta \in (-\alpha, \beta]. \tag{4.12}$$

Consequently, $\bar{\gamma}$, as defined in (1.3), is finite. Furthermore, for each $\varepsilon > 0$, there exists $L < \infty$ such that

$$\mathrm{P}\bigg(\frac{||X_{t_1}||}{L} \le ||X_{t_2}|| \le L||X_{t_1}||, 0 \le t_1 < t_2 \le n \,\bigg|\, X_0 = x\bigg) > 1 - n\varepsilon, \tag{4.13}$$

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

Proof. For any $\zeta \in (0, \beta]$, (4.12) is easy to confirm. From Assumption 1.2(i) we know that

$$\bar{M} := \sup_{x \in \mathbb{X}} \sup_{u \in \mathbb{E}} \frac{1 + ||B(x/||x||, u)||x|| + C(x, u)||}{(1 + ||x||)(1 + |u|)} < \infty. \tag{4.14}$$

Then

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

and (4.12) follows. Now, suppose $-\alpha < \zeta < 0$ and let $\delta \in (0, 1)$ be arbitrary. We have

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

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

when $||x|| > 1$, by Assumption 1.2(ii). In the same way that (4.2) implies (4.3), this suffices to show (4.12).

Additionally, let

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

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

$$\begin{aligned}\frac{1}{n}\,\mathrm{E}\biggl(\biggl|\log\biggl(\frac{1+||X_n||}{1+||x||}\biggr)\biggr| \;\Big|\; X_0 = x\biggr) &\le \frac{1}{n}\,\mathrm{E}\biggl(\max_{0\le t_1<t_2\le n}\biggl|\log\biggl(\frac{1+||X_{t_2}||}{1+||X_{t_1}||}\biggr)\biggr| \;\Big|\; X_0 = x\biggr)\\ &\le \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\biggl(\biggl|\log\biggl(\frac{1+||X_t||}{1+||X_{t-1}||}\biggr)\biggr| \;\Big|\; X_0 = x\biggr)\\ &\le M_1\\ &< \infty, \end{aligned} \tag{4.15}$$

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

Also, if $||x|| > L \ge 2\mathrm{e}^{M_1/\varepsilon}$, then Markov's inequality and (4.15) show that

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

verifying (4.13).

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 $\bar{\varepsilon} > 0$, there exist $n \ge 1$ and a positive function $V_1(x)$ satisfying*

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

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 $\bar{\varepsilon} > 0$ and set

$$M_1 = \sup_{x\in\mathbb{X}} \mathrm{E}\biggl(\biggl|\log\frac{1+||B(x/||x||, e_1)||x|| + C(x, e_1)||}{1+||x||}\biggr|\biggr),$$

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

and $K_1 \geq \max(2\sup_{\theta\in\Theta}|\nu(\theta)|, 1)$. Fix $\varepsilon = \bar{\varepsilon}/(5+2M_1)$ and choose ν according to Theorem 3.2(i). According to Assumption 1.4(ii) and (iii), choose $n \geq 1$ and $L_0 \geq 1$ such that

$$\sup_{||x||>L_0,\, x/||x||\in\Theta_\#} \mathrm{P}(\tilde{\theta}_{1,x} \notin \Theta_\# \text{ or } \tilde{\theta}^*_{1,x} \notin \Theta_\#) < \frac{\varepsilon}{K_1}, \tag{4.16}$$

and

$$\sup_{||x||>L_0} \mathrm{P}(\tilde{\theta}_n \notin \Theta_\# \mid X_0 = x) < \frac{\varepsilon}{K_1}. \tag{4.17}$$

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

$$\sup_{||x||>L_1} \mathrm{P}\biggl(||X_n|| \leq \frac{||x||}{L_1} \biggm| X_0 = x\biggr) < \frac{\varepsilon}{K_1}. \tag{4.18}$$

Define

$$\tilde{\theta}_{1,x} = \frac{B(x/||x||, e_1)||x|| + C(x, e_1)}{||B(x/||x||, e_1)||x|| + C(x, e_1)||} \quad \text{and} \quad \tilde{\theta}^*_{1,x} = \frac{B(x/||x||, e_1)}{||B(x/||x||, e_1)||},$$

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 1.2(i), and since ν is uniformly continuous on $\Theta_\#$, it follows that

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

Thus, by (4.16) there exists $L_2 \geq L_1$ such that

$$\begin{aligned}
&\sup_{||x||>L_2,\, x/||x||\in\Theta_\#} \mathrm{E}(|\nu(\tilde{\theta}_{1,x}) - \nu(\tilde{\theta}^*_{1,x})|) \\
&\qquad \leq \sup_{||x||>L_2, x/||x||\in\Theta_\#} \mathrm{E}(|\nu(\tilde{\theta}_{1,x}) - \nu(\tilde{\theta}^*_{1,x})|\,\mathbf{1}_{\tilde{\theta}_{1,x}\in\Theta_\#,\tilde{\theta}^*_{1,x}\in\Theta_\#}) \\
&\qquad\quad + \sup_{||x||>L_2, x/||x||\in\Theta_\#} K_1\,\mathrm{P}(\tilde{\theta}_{1,x} \notin \Theta_\# \text{ or } \tilde{\theta}^*_{1,x} \notin \Theta_\#) \\
&\qquad < 2\varepsilon.
\end{aligned} \tag{4.19}$$

Analogous to the definition of $\tilde{\theta}^*_1$ prior to Assumption 1.4, let $\tilde{\theta}^*_{n+1} = \eta(X_n/||X_n||, e_{n+1})$. Conditioning on X_n, (4.19) gives

$$\begin{aligned}
&\sup_{x\in\mathbb{X}} \mathrm{E}(|\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}^*_{n+1})|\,\mathbf{1}_{||X_n||>L_2,\tilde{\theta}_n\in\Theta_\#} \mid X_0 = x) \\
&\qquad = \sup_{x\in\mathbb{X}} \mathrm{E}(\mathrm{E}(|\nu(\tilde{\theta}_{1,X_n}) - \nu(\tilde{\theta}^*_{1,X_n})| \mid X_n)\,\mathbf{1}_{||X_n||>L_2,\tilde{\theta}_n\in\Theta_\#} \mid X_0 = x) \\
&\qquad < 2\varepsilon.
\end{aligned} \tag{4.20}$$

Then, by combining (4.20) with (4.17) and (4.18), we obtain

$$\begin{aligned}
&\sup_{||x||>L_2^2} \mathrm{E}(|\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}^*_{n+1})| \mid X_0 = x) \\
&\qquad \leq \sup_{||x||>L_2^2} \mathrm{E}(|\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}^*_{n+1})|\,\mathbf{1}_{||X_n||>L_2,\tilde{\theta}_n\in\Theta_\#} \mid X_0 = x) \\
&\qquad\quad + \sup_{||x||>L_2^2} K_1\,\mathrm{P}\biggl(\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \leq \frac{||x||}{L_2} \biggm| X_0 = x\biggr) \\
&\qquad < 4\varepsilon.
\end{aligned} \tag{4.21}$$

We also have

$$\frac{1 + ||B(x/||x||, e_1)||x|| + C(x, e_1)||}{1 + ||x||} - w\left(\frac{x}{||x||}, e_1\right) \to 0$$

almost surely, as $||x|| \to \infty$, so that by uniform integrability (cf. Theorem 3.1 and Lemma 4.2), we obtain

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

Thus, for some $L_3 \geq L_2$,

$$\sup_{||x|| > L_3} \mathrm{E}\left(\left|\log \frac{1 + ||B(x/||x||, e_1)||x|| + C(x, e_1)||}{1 + ||x||} - \log w\left(\frac{x}{||x||}, e_1\right)\right|\right) < \varepsilon. \tag{4.22}$$

Recall the definition of M_1 and observe that $\sup_{\theta \in \Theta} \mathrm{E}(\log w(\theta, e_1)) \leq M_1$. Using (4.18) and (4.22), condition on X_n to obtain

$$\begin{aligned}
&\sup_{||x|| > L_3^2} \mathrm{E}\left(\left|\log \frac{1 + ||X_{n+1}||}{1 + ||X_n||} - \log w(\tilde{\theta}_n, e_{n+1})\right| \;\middle|\; X_0 = x\right) \\
&\qquad \leq \sup_{||x|| > L_3^2} \mathrm{E}\left(\left|\log \frac{1 + ||X_{n+1}||}{1 + ||X_n||} - \log w(\tilde{\theta}_n, e_{n+1})\right| \mathbf{1}_{||X_n|| > L_3} \;\middle|\; X_0 = x\right) \\
&\qquad\quad + \sup_{||x|| > L_3^2} 2M_1 \, \mathrm{P}(||X_n|| \leq L_3 \mid X_0 = x) \\
&\qquad < (2M_1 + 1)\varepsilon
\end{aligned} \tag{4.23}$$

(since $K_1 \geq 1$).

Conditioning on X_n one last time, now using (3.2), yields

$$\sup_{x \in \mathbb{X}} |\, \mathrm{E}(\nu(\tilde{\theta}^*_{n+1}) - \nu(\tilde{\theta}_n) + \log w(\tilde{\theta}_n, e_{n+1}) \mid X_0 = x) - \gamma \,| = 0. \tag{4.24}$$

Combining (4.21), (4.23), and (4.24), we arrive at

$$\begin{aligned}
\sup_{||x|| > L_3^2} \left| \mathrm{E}\left(\nu(\tilde{\theta}_{n+1}) - \nu(\tilde{\theta}_n) + \log \frac{1 + ||X_{n+1}||}{1 + ||X_n||} \;\middle|\; X_0 = x\right) - \gamma \right| &< (5 + 2M_1)\varepsilon \\
&= \bar{\varepsilon},
\end{aligned} \tag{4.25}$$

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

We can now identify the Lyapunov exponent.

Proof of Theorem 3.3. We begin with (4.25) and the constants chosen in its proof. Note that n is fixed and depends on $\bar{\varepsilon}$. Set $\bar{L}_0 = L_3^2$. By Lemma 4.2, there exists $\bar{L}_t \geq \bar{L}_{t-1}$ such that

$$\sup_{||x|| > \bar{L}_t} \mathrm{P}(||X_t|| \leq L_3^2 \mid X_0 = x) \leq \frac{\bar{\varepsilon}}{K_1 + 2M_1 + |\gamma|} \quad \text{for } t \geq 1. \tag{4.26}$$

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

Hence, using (4.25) and conditioning on X_t in the manner of the previous proof yields

$$\sup_{||x||>\bar{L}_t}\left|\mathrm{E}\left(\left(\nu(\tilde{\theta}_{n+t+1})-\nu(\tilde{\theta}_{n+t})+\log\frac{1+||X_{n+t+1}||}{1+||X_{n+t}||}-\gamma\right)\mathbf{1}_{||X_t||>L_3^2}\;\Big|\;X_0=x\right)\right|<\bar{\varepsilon}, \tag{4.27}$$

$t \geq 1$, whereas by (4.26)

$$\begin{aligned}
&\sup_{||x||>\bar{L}_t}\left|\mathrm{E}\left(\left(\nu(\tilde{\theta}_{n+t+1})-\nu(\tilde{\theta}_{n+t})+\log\frac{1+||X_{n+t+1}||}{1+||X_{n+t}||}-\gamma\right)\mathbf{1}_{||X_t||\leq L_3^2}\;\Big|\;X_0=x\right)\right|\\
&\qquad\leq (K_1+2M_1+|\gamma|)\sup_{||x||>\bar{L}_t}\mathrm{P}(||X_t||\leq L_3^2\mid X_0=x)\\
&\qquad<\bar{\varepsilon}.
\end{aligned} \tag{4.28}$$

Combining (4.27) and (4.28) and summing over $t = 0, \ldots, k-1$, we obtain

$$\sup_{||x||>\bar{L}_{k-1}}\left|\mathrm{E}\left(\nu(\tilde{\theta}_{n+k})-\nu(\tilde{\theta}_n)+\log\frac{1+||X_{n+k}||}{1+||X_n||}\;\Big|\;X_0=x\right)-k\gamma\right|<2k\bar{\varepsilon}. \tag{4.29}$$

In other words, (4.29) declares

$$\limsup_{||x||\to\infty}\left|\frac{1}{k}\,\mathrm{E}\left(\nu(\tilde{\theta}_{n+k})-\nu(\tilde{\theta}_n)+\log\frac{1+||X_{n+k}||}{1+||X_n||}\;\Big|\;X_0=x\right)-\gamma\right|<2\bar{\varepsilon}$$

for all $k \geq 1$. As ν is bounded and

$$\mathrm{E}\left(\log\frac{1+||X_n||}{1+||X_0||}\;\Big|\;X_0=x\right)$$

is bounded (as n is fixed), we conclude that

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

Finally, $\bar{\varepsilon}$ may be chosen arbitrarily and thus (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 λ is uniformly continuous on $\Theta_\#$. It is well known that (3.5) and (3.6) imply geometric ergodicity and finite ζ-moment of the stationary distribution, under Assumption 1.3. Thus, in our proof we will validate (3.5) and (3.6). To this end, let $\bar{\varepsilon} \in (0, 1-\rho)$ and set $\rho_0 = \rho + \bar{\varepsilon}$. Also, choose K_2 such that $1/K_2 < \lambda(\theta) < K_2$ for all θ.

Note that (3.3) and Assumption 1.2(i) imply $\mathrm{E}(|e_1|^\beta) < \infty$, for all $\beta < \kappa$. Indeed, we may assume $\beta > \zeta$ in Theorem 3.1 and Lemma 4.2. Let $M_3 = \bar{M}^\zeta\,\mathrm{E}((1+|e_1|)^\zeta)$, where $\bar{M}$ is defined in (4.14). Then

$$\sup_{x\in\mathbb{X}}\mathrm{E}\left(\left(w\left(\frac{x}{||x||},e_1\right)\right)^\zeta\right)\leq\sup_{x\in\mathbb{X}}\mathrm{E}\left(\left(\frac{1+||X_1||}{1+||x||}\right)^\zeta\;\Big|\;X_0=x\right)\leq M_3. \tag{4.30}$$

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

Now let $\varepsilon = \frac{1}{32}(\bar{\varepsilon}/K_2^2 M_3)^2$. According to Assumption 1.4(ii) and (iii), fix n and L_0 such that

$$\sup_{||x||>L_0,\, x/||x||\in\Theta_\#} \mathrm{P}(\tilde{\theta}_1 \notin \Theta_\# \text{ or } \tilde{\theta}_1^* \notin \Theta_\# \mid X_0 = x) < \frac{\bar{\varepsilon}}{8K_2^2 M_3}, \tag{4.31}$$

and

$$\sup_{||x||>L_0} \mathrm{P}(\tilde{\theta}_n \notin \Theta_\# \mid X_0 = x) < \varepsilon. \tag{4.32}$$

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

$$\sup_{||x||>L_1} \mathrm{P}\biggl(||X_n|| \le \frac{||x||}{L_1} \biggm| X_0 = x\biggr) < \varepsilon. \tag{4.33}$$

Then, using (4.30) and (4.31),

$$\begin{aligned}
&\sup_{||x||>L_1,\, x/||x||\in\Theta_\#} \mathrm{E}\biggl(\biggl|\frac{\lambda(\tilde{\theta}_1)-\lambda(\tilde{\theta}_1^*)}{\lambda(x/||x||)}\biggr|\biggl(w\biggl(\frac{x}{||x||}, e_1\biggr)\biggr)^{\zeta} \mathbf{1}_{\tilde{\theta}_1\notin\Theta_\# \text{ or } \tilde{\theta}_1^*\notin\Theta_\#} \biggm| X_0 = x\biggr) \\
&\quad \le 2K_2^2 M_3 \sup_{||x||>L_1,\, x/||x||\in\Theta_\#} \mathrm{P}(\tilde{\theta}_1 \notin \Theta_\# \text{ or } \tilde{\theta}_1^* \notin \Theta_\# \mid X_0 = x) \\
&\quad < \frac{\bar{\varepsilon}}{4}.
\end{aligned} \tag{4.34}$$

Recall that, by Assumption 1.2(i),

$$\frac{1 + ||B(x/||x||, e_1)||x|| + C(x, e_1)||}{1 + ||x||} - w\biggl(\frac{x}{||x||}, e_1\biggr) \to 0$$

almost surely, as $||x|| \to \infty$. By dominated convergence and the uniform integrability implied by Theorem 3.1 and Lemma 4.2, we obtain

$$\limsup_{||x||\to\infty} \mathrm{E}\biggl(\frac{\lambda(\tilde{\theta}_1)}{\lambda(x/||x||)}\biggl|\frac{(1+||X_1||)^{\zeta}}{(1+||x||)^{\zeta}} - \biggl(w\biggl(\frac{x}{||x||}, e_1\biggr)\biggr)^{\zeta}\biggr| \biggm| X_0 = x\biggr) = 0. \tag{4.35}$$

On the other hand, using the uniform continuity of λ on $\Theta_\#$, and uniform integrability again, we obtain

$$\limsup_{||x||\to\infty} \mathrm{E}\biggl(\biggl|\frac{\lambda(\tilde{\theta}_1)-\lambda(\tilde{\theta}_1^*)}{\lambda(x/||x||)}\biggr|\biggl(w\biggl(\frac{x}{||x||}, e_1\biggr)\biggr)^{\zeta} \mathbf{1}_{\tilde{\theta}_1\in\Theta_\#,\tilde{\theta}_1^*\in\Theta_\#} \biggm| X_0 = x\biggr) = 0. \tag{4.36}$$

Therefore, from (3.3) and (4.34)–(4.36), there exists $L_2 \ge L_1$ such that

$$\sup_{||x||>L_2,\, x/||x||\in\Theta_\#} \mathrm{E}\biggl(\frac{\lambda(\tilde{\theta}_1)}{\lambda(x/||x||)}\,\frac{(1+||X_1||)^{\zeta}}{(1+||x||)^{\zeta}} \biggm| X_0 = x\biggr) < \rho + \frac{\bar{\varepsilon}}{2}. \tag{4.37}$$

Let $V_2(x) = \lambda(x/||x||)(1+||x||)^{\zeta}$. Thus, from (4.37), we have

$$\mathrm{E}(V_2(X_1) \mid X_0 = x) < \biggl(\rho + \frac{\bar{\varepsilon}}{2}\biggr)V_2(x) \quad \text{for all } ||x|| > L_2 \text{ and } \tfrac{x}{||x||} \in \Theta_\#. \tag{4.38}$$

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

Observe that, by conditioning on X_n and applying (4.30), we obtain

$$\mathrm{E}(V_2(X_{n+1}) \mid X_0 = x) \le K_2^2 M_3 \, \mathrm{E}(V_2(X_n) \mid X_0 = x). \tag{4.39}$$

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

$$(\mathrm{E}((V_2(X_{n+1}))^{\beta/\zeta} \mid X_0 = x))^{\zeta/\beta} \le 2\,\mathrm{E}(V_2(X_{n+1}) \mid X_0 = x). \tag{4.40}$$

For $||x|| > L_2^2$, (4.32), (4.33), (4.39), (4.40), and the choice of ε and β yield

$$\begin{aligned}
&\mathrm{E}(V_2(X_{n+1})\,\mathbf{1}_{\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \le L_1} \mid X_0 = x \\
&\quad \le (\mathrm{E}((V_2(X_{n+1}))^{\beta/\zeta} \mid X_0 = x))^{\zeta/\beta} (\mathrm{P}(\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \le L_1 \mid X_0 = x))^{1-\zeta/\beta} \\
&\quad \le 2\,\mathrm{E}(V_2(X_{n+1}) \mid X_0 = x)(2\varepsilon)^{1-\zeta/\beta} \\
&\quad < 2K_2^2 M_3 (2\varepsilon)^{1/2}\,\mathrm{E}(V_2(X_n) \mid X_0 = x) \\
&\quad \le \left(\frac{\bar{\varepsilon}}{2}\right) \mathrm{E}(V_2(X_n) \mid X_0 = x).
\end{aligned} \tag{4.41}$$

Conditioning on X_n once again, (4.38) and (4.41) yield, for $||x|| > L_2^2$,

$$\begin{aligned}
\mathrm{E}(V_2(X_{n+1}) \mid X_0 = x) &\le \mathrm{E}\left(\left(\rho + \frac{\bar{\varepsilon}}{2}\right) V_2(X_n) + V_2(X_{n+1})\,\mathbf{1}_{\tilde{\theta}_n \notin \Theta_\# \text{ or } ||X_n|| \le L_1} \,\middle|\, X_0 = x\right) \\
&< (\rho + \bar{\varepsilon})\,\mathrm{E}(V_2(X_n) \mid X_0 = x) \\
&= \rho_0\,\mathrm{E}(V_2(X_n) \mid X_0 = x).
\end{aligned} \tag{4.42}$$

The desired test function is $V(x) = \mathrm{E}(V_2(X_n) \mid X_0 = x)$ which, by applying Lemma 4.2 iteratively and by the fact that λ is bounded and bounded away from zero, 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) also confirms that

$$K_0 := \sup_{V(x) \le M_0} \mathrm{E}(V(X_1) \mid X_0 = x) = \sup_{V(x) \le M_0} \mathrm{E}(V_2(X_{n+1}) \mid X_0 = x) < \infty.$$

With (4.42), this completes the verification that (3.5) holds as well.

(ii) A positive Lyapunov exponent implies $\{X_t\}$ is transient by Theorem 3.3, Lemma 4.2, and Cline and Pu (2001, Theorem 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 $L < \infty$,

$$\mathrm{E}\left(\left(\frac{1 + ||X_0||}{1 + ||X_n||}\right)^{\zeta} \,\middle|\, X_0 = x\right) < K\rho^n \quad \text{for all } n \ge 1 \text{ and } ||x|| > L.$$

Hence, by Markov's inequality,

$$\begin{aligned}
\mathrm{P}(\rho^n ||X_n|| \le M \mid X_0 = x) &\le \mathrm{P}\left(\left(\frac{1 + ||X_0||}{1 + ||X_n||}\right) \ge \rho^n \frac{L}{M} \,\middle|\, X_0 = x\right) \\
&\le \left(\rho^n \frac{L}{M}\right)^{-\zeta} \mathrm{E}\left(\left(\frac{1 + ||X_0||}{1 + ||X_n||}\right)^{\zeta} \,\middle|\, X_0 = x\right) \\
&\le K\left(\frac{M}{L}\right)^{\zeta} \rho^{(1-\zeta)n}
\end{aligned} \tag{4.43}$$

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

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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

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

That is, $\mathrm{P}(||X_t|| \le L,\ t \le n \mid X_0 = x) \le (1-\delta)^{n/n_0 - 1}$, for all n. Combining this with (4.43), we have

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

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

Proof of Corollary 3.1. Define $q_t(\theta)$ as in the proof of Theorem 3.2(i). It is then easy to show that

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

which converges uniformly, as $n \to \infty$, by the proof of Theorem 3.2(i). It is also clear from the proof of Theorem 3.2(i) that the limit will be uniformly continuous on $\Theta_{\#}$. Furthermore,

$$\bar{\nu}_n(\theta) - \nu_{n-1}(\theta) = \mathrm{E}(\nu_{n-1}(\theta_1^*) - \nu_{n-1}(\theta) + \log W_1^* \mid \theta_0^* = \theta) \to \gamma$$

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

$$\begin{aligned}
\bar{\gamma}_n &= \sup_{\theta \in \Theta} (\bar{\nu}_n(\theta) - \nu_{n-1}(\theta)) = \sup_{\theta \in \Theta} q_n(\theta)\\
&= \sup_{\theta \in \Theta} \int_{\mathbb{E}} q_{n-1}(\eta(\theta, u)) f(u)\, \mathrm{d}u \le \sup_{\theta \in \Theta} q_{n-1}(\theta) = \bar{\gamma}_{n-1}.
\end{aligned}$$

Therefore, $\bar{\gamma}_n \downarrow \gamma$ and, similarly, $\underline{\gamma}_n \uparrow \gamma$.

Proof of Corollary 3.2. Let π be the stationary distribution for $\{\theta_t^*\}$. Then (3.7) implies

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

and geometric ergodicity follows from Theorem 3.4(i).

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

5. Verifying assumptions

In this section we outline verifications for the regularity Assumptions 1.2–1.4 for the threshold AR-GARCH model defined in Section 2.1. In particular, we will assume Assumption 1.1 and that

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

(ii) $b_{ki} > 0$, for $k = 0, \dots, p$ and $i = 1, \dots, N$;

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

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

These, undoubtedly, are not necessary assumptions and one may wish to weaken them slightly, especially (ii). However, 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 $y = (y_1, \dots, y_p)$ and $s = (s_1, \dots, s_q)$. 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 1.2(i) clearly holds. Recall the expressions for $B(x, u)$ and $C(x, u)$ in Section 2.1. As (note $b(y) \ge b^*(y)$)

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

Assumption 1.2(ii) is trivially true with $K = 1/\delta_0$ and $\alpha = 1$.

Since $\mathbb{X}$ is closed, Assumption 1.3 holds if $\{X_t\}$ is an aperiodic ϕ-irreducible T-chain on $\mathbb{X}$ and $\{X_t^*\}$ is an aperiodic ϕ-irreducible T-chain on $\mathbb{X}^* = \mathbb{X} \backslash \{0\}$. (See Meyn and Tweedie (1993, Theorem 6.2.5).) 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 $T(x, A) = \mathrm{P}(X_1 \in A \mid X_0 = x)$ and suppose that the following three conditions hold.*

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

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

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

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

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

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

Proof. The first two conditions imply that $\{X_t\}$ is a T-chain by the argument in Meyn and Tweedie (1993, Proposition 6.2.4). The first inequality in the third condition states that $\tilde{x}$ is 'reachable' which, combined with the above, implies that $\{X_t\}$ is ϕ-irreducible by Meyn and

Tweedie (1993, Proposition 6.2.1). The irreducibility measure ϕ is absolutely continuous with respect to the Lebesgue measure. The second inequality in 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 that $b_{0i} > 0$, $b_{pi} > 0$, $c_{qi} > 0$, 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 1.3.*

Proof. We consider only a sketch of the proof here as the algebra involved can get tedious. We focus on the basic process $\{(Y_t, S_t)\}$, but the argument is also valid for the homogeneous process. 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 that $Y_0 = y = (y_1, \ldots, y_p)$ and $S_0 = s = (s_1, \ldots, s_q)$. Then, $0 < \underline{b}_0 \le \sigma_t^2 \le \bar{b}_0^t(1+||(y, s)||^2)$, for all $t \ge 1$. (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, $\sigma_t^2 \ge d_0^t \max(s_1^2, \ldots, s_q^2) > 0$, for all $t \ge 1$.) Let $\sigma_1(y, s) = ((b(y))^2 + c^2(s, y))^{1/2}$. It is clear, then, that the conditional distribution of ξ_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 zero.

Indeed, we 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 zero, and if $\varepsilon_t \in (0, 1)$ and $t \ge 1$, then

$$\begin{aligned} \underline{h}_n((y, s), z_1, \ldots, z_n) \prod_{t=1}^{n} \varepsilon_t &\le \mathrm{P}\left(\bigcap_{t \le n} \{|\xi_t - z_t| < \varepsilon_t\} \,\middle|\, Y_0 = y,\ S_0 = s\right) \\ &\le \bar{h}_n((y, s), z_1, \ldots, z_n) \prod_{t=1}^{n} \varepsilon_t \end{aligned} \tag{5.1}$$

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 $k \ge p + q$. 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

$$\begin{aligned} Y_k^2 &= (\xi_k^2, \ldots, \xi_{k-p+1}^2), \\ \tilde{Y}_{k-p}^2 &= (\xi_{k-p}^2, \ldots, \xi_1^2), \\ S_k^2 &= (\sigma_k^2, \ldots, \sigma_{k-q+1}^2), \quad \text{and} \\ \tilde{S}_{k-q}^2 &= (\sigma_{k-q}^2, \ldots, \sigma_1^2). \end{aligned}$$

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},

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

C_{12}, C_{22}, D_1, and D_2, with elements determined by the parameters and the values of $I_1, \dots, I_k$, so that

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

Note that C_{11} (which is $q \times q$) and C_{22} (which is $(k-q) \times (k-q)$) are upper triangular matrices with 0s 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 $I_1, \dots, I_k$.

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 0s and, since $b_{pi} > 0$ and $c_{qi} > 0$, for all $i = 1, \dots, N$, 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 $I_1, \dots, I_k$. This means for each $\Xi = (I_1, \dots, I_k)$ there is a nonsingular transformation H_Ξ and a vector K_Ξ 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 -1s 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 $I_1, \dots, I_k$ 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 that $(Y_0, S_0) = (y, s)$, (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 σ_ts. 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 zero 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 $(1, \dots, 1)$ 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 $\sigma_0^2 = (b_{0i} + \bar{b}_i)/(1 - \bar{c}_i)$. Suppose that $\varepsilon_t \in (0, 1)$, for $t \geq 1$, and each is small enough that $|\xi_t - 1| < \varepsilon_t$ implies $Y_t \in C_i$, for $t \geq p$. Then $|\sigma_t^2 - \sigma_0^2| < \delta_t$, for $t = 1, \dots, n-1$, further implies

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

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

$$\mathrm{P}\biggl(\max_{t=1,\dots,p} |\xi_{n-t} - 1| < \varepsilon, \max_{t=1,\dots,q} |\sigma_{n-t} - \sigma_0| < \varepsilon \biggm| Y_0 = y, S_0 = s\biggr) > 0,$$

for all n 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 $C_1, \dots, C_N$ are contained in the union of affine planes*

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

Assume that e_t has a density f which is bounded and suppose that $b_{ki} > 0$, for $k = 0, \dots, p$, and $c_{ki} > 0$, for $k = 1, \dots, q$, for all $i = 1, \dots, N$. Then the threshold AR-GARCH(p, q) model of Section 2.1 satisfies Assumption 1.4.

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

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

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

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

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

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

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

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

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

The above arguments can be modified to include the variable-driven switching model of Section 2.3, at least if the link functions $h_1, \dots, h_N$ are sufficiently smooth (such as piecewise continuous on the regimes $C_1, \dots, C_n$). 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. Process. Appl.* **96,** 123–142.

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

Basrak, B., Davis, R. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stoch. Process. 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. Prob.* **11,** 1220–1241.

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

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

Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Prob.* **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. Prob.* **17,** 666–678.

Cline, D. B. H. (2007a). Evaluating the Lyapunov 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. *Stoch. Process. Appl.* **117,** 840–861.

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

Cline, D. B. H. and Pu, H. H. (2001). Geometric transience of nonlinear time series. *Statist. 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. Prob.* **14,** 1920–1949.

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. Process. Appl.* **115,** 1954–1978.

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

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

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

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

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

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

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

Hwang, S. Y. and Woo, M.-J. (2001). Threshold ARCH(1) processes: asymptotic inference. *Statist. Prob. Lett.* **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.

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

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 β-mixing property of certain nonlinear GARCH$(p, q,)$ models. *Statist. Prob. Lett.* **73,** 25–35.

https://doi.org/10.1239/aap/1183667619 Published online by Cambridge University Press

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. R. Statist. Soc. B* **66,** 63–78.

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

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

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

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

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

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

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

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

Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis* (Notes in Stat. **21**). Springer, Heidelberg.

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

Tong, H. and Lim, K. S. (1980). Threshold autregression, limit cycles and cyclical data (with discussion). *J. R. Statist. Soc. Ser. 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. *Statist. Sinica* **17,** 241–264.

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