Subgeometrically Ergodic Autoregressions with Autoregressive Conditional Heteroskedasticity

Meitz, Mika; Saikkonen, Pentti

2023-11-17
Cambridge University Press
http://hdl.handle.net/10138/589122
Meitz, M & Saikkonen, P 2023, 'Subgeometrically Ergodic Autoregressions with Autoregressive Conditional Heteroskedasticity', Econometric Theory. https://doi.org/10.1017/S026646662300035X

Subgeometrically ergodic autoregressions with autoregressive conditional heteroskedasticity*

Mika Meitz
University of Helsinki

Pentti Saikkonen
University of Helsinki

First version May 2022, revised April 2023

Abstract

In this paper, we consider subgeometric (specifically, polynomial) ergodicity of univariate nonlinear autoregressions with autoregressive conditional heteroskedasticity (ARCH). The notion of subgeometric ergodicity was introduced in the Markov chain literature in 1980s and it means that the transition probability measures converge to the stationary measure at a rate slower than geometric; this rate is also closely related to the convergence rate of β-mixing coefficients. While the existing literature on subgeometrically ergodic autoregressions assumes a homoskedastic error term, this paper provides an extension to the case of conditionally heteroskedastic ARCH-type errors, considerably widening the scope of potential applications. Specifically, we consider suitably defined higher-order nonlinear autoregressions with possibly nonlinear ARCH errors and show that they are, under appropriate conditions, subgeometrically ergodic at a polynomial rate. An empirical example using energy sector volatility index data illustrates the use of subgeometrically ergodic AR–ARCH models.

JEL classification: C22.

MSC2020 classifications: 60J05, 37A25.

Keywords: Nonlinear autoregressive model, autoregressive conditional heteroskedasticity, ARCH, subgeometric ergodicity, polynomial ergodicity, Markov chain, β-mixing.

*The authors thank the Academy of Finland (MM and PS), Foundation for the Advancement of Finnish Securities Markets (MM), and OP Group Research Foundation (MM) for financial support, and Co-Editor Robert Taylor and three anonymous referees for useful comments and suggestions. Contact addresses: Mika Meitz, Department of Economics, University of Helsinki, P. O. Box 17, FI–00014 University of Helsinki, Finland; e-mail: mika.meitz@helsinki.fi. Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FI–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

Let X_t $(t = 0, 1, 2, \ldots)$ be a Markov chain on the state space X and initialized from an X_0 following some initial distribution. If the n-step probability measures $P^n(x\,;\,\cdot) = \Pr(X_n \in \cdot \mid X_0 = x)$ converge in total variation norm $\|\cdot\|_{TV}$ to the stationary probability measure π at rate r^n (for some $r > 1$), that is,

$$\lim_{n\to\infty} r^n \|P^n(x\,;\,\cdot) - \pi(\cdot)\|_{TV} = 0, \qquad \pi \text{ a.e.}, \tag{1}$$

the Markov chain is said to be geometrically ergodic. When the convergence in (1) takes place at a suitably defined rate $r(n)$ slower than geometric, that is,

$$\lim_{n\to\infty} r(n) \|P^n(x\,;\,\cdot) - \pi(\cdot)\|_{TV} = 0, \qquad \pi \text{ a.e.}, \tag{2}$$

the Markov chain is called subgeometrically ergodic. Examples of common rates (where c denotes a positive constant) include geometric (or exponential) when $r(n) = e^{cn} = r^n$ $(r > 1)$, subexponential when $r(n) = e^{cn^\gamma}$ $(0 < \gamma < 1)$, polynomial when $r(n) = (1+n)^c$, and logarithmic when $r(n) = (1 + \ln(n))^c$. The authoritative and classic reference to Markov chain theory is the monograph of Meyn and Tweedie (2009), while an up-to-date treatment of subgeometric ergodicity can be found in Chapters 16 and 17 of Douc, Moulines, Priouret, and Soulier (2018).

To give some background, the notion of subgeometric ergodicity was introduced in the Markov chain literature in the 1980s when Nummelin and Tuominen (1983) and Tweedie (1983) obtained the first subgeometric ergodicity results for general state space Markov chains. Subsequent work by Tuominen and Tweedie (1994), Fort and Moulines (2000), Jarner and Roberts (2002), Fort and Moulines (2003), and Douc, Fort, Moulines, and Soulier (2004) lead to a formulation of a so-called drift condition to ensure subgeometric ergodicity, paralleling the use of a Foster-Lyapunov drift condition to establish geometric ergodicity (see, e.g., Meyn and Tweedie, 2009, Ch 15). Various topics in probability theory and statistics have also been considered under subgeometric assumptions; for instance, Douc, Guillin, and Moulines (2008) considered the central limit theorem and Berry-Esseen bounds, Atchadé and Fort (2010) the convergence of Markov chain Monte Carlo algorithms, Merlevède, Peligrad, and Rio (2011) a Bernstein-type inequality, and Meitz and Saikkonen (2021) the rate of β-mixing. In this paper we are interested in autoregressive time series models. Results regarding the subgeometric ergodicity of first-order autoregressions were obtained by Tuominen and Tweedie (1994), Veretennikov (2000), Fort and Moulines (2003), Douc et al. (2004), Klokov and Veretennikov (2004, 2005), and Klokov (2007), among others, whereas results for more general higher-order autoregressions were obtained by Meitz and Saikkonen (2022).

In this paper we consider subgeometric (specifically, polynomial) ergodicity of autoregressive models with autoregressive conditional heteroskedasticity (ARCH; Engle, 1982). The previous works on subgeometrically ergodic autoregressions listed above only considered the case of independent and identically distributed (IID) errors, and allowing for conditionally heteroskedastic errors considerably widens the scope of potential applications. This is particularly important in applications using economic and financial time series data. In the subgeometrically ergodic AR–ARCH models we consider, the conditional mean is similar to the (homoskedastic) AR

models already considered in Meitz and Saikkonen (2022). The precise model formulation will be given and motivated further in Section 2, but we already note that the models we consider accommodate for behavior similar to a unit root process for large values of the observed series but almost no restrictions are placed on their dynamics for moderate values of the observed series. The conditional variance is allowed to follow a rather general nonlinear ARCH process. In our main result, we show that the considered AR–ARCH processes are, under appropriate conditions, subgeometrically ergodic at a polynomial rate; the convergence rate of β-mixing coefficients and finiteness of certain moments are also obtained (for details, see Section 3.2).

The inclusion of ARCH (instead of IID) errors considerably complicates the proofs of (sub)geometric ergodicity of nonlinear autoregressions. Papers considering subgeometric ergodicity of homoskedastic autoregressions were already listed above. Geometric ergodicity of nonlinear autoregressive models with ARCH (or generalized ARCH) errors has previously been considered by numerous authors; see, e.g., Cline and Pu (2004), Meitz and Saikkonen (2008, 2010), and the many references therein. Compared to these two strands of previous literature, the combination of the subgeometrically ergodic type of nonlinear dynamics in the conditional mean with ARCH errors leads to additional complications in the proofs. To appropriately separate these two sources of dynamics we make use of a (relatively unknown) extension of Bernoulli's inequality due to Fefferman and Shapiro (1972) (combined with Young's inequality), and to control terms arising due to conditional heteroskedasticity we devise a special matrix norm that is of a more complicated type than the norms typically used when analysing the stability of nonlinear time series models.

The rest of the paper is organized as follows. Section 2 introduces the nonlinear AR–ARCH model considered and states the assumptions we employ. Results on subgeometric ergodicity are given in Section 3. In Section 4 we consider an empirical application of our model to a daily time series of an energy sector volatility index. Section 5 concludes. All proofs are collected in an Appendix.

2 Model

2.1 Conditional mean

We consider the univariate process y_t $(t = 1, 2, \ldots)$ generated by

$$y_t = \pi_1 y_{t-1} + \cdots + \pi_{p-1} y_{t-p+1} + g(u_{t-1}) + \sigma_t \varepsilon_t, \tag{3}$$

where $p \geq 1$ is the autoregressive order, $u_t = y_t - \pi_1 y_{t-1} - \cdots - \pi_{p-1} y_{t-p+1}$, g is a real-valued function, ε_t is an IID error term, and $\sigma_t = \sigma(\boldsymbol{y}_{t-1})$ is a positive volatility term that depends on $p + q$ lagged values of y_t, $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p-q})$, where $q \geq 1$ is an ARCH order. For now, one concrete example of the volatility term is a linear ARCH process, where σ_t satisfies

$$\sigma_t^2 = \omega + \alpha_1 e_{t-1}^2 + \cdots + \alpha_q e_{t-q}^2 \tag{4}$$

and $e_t = y_t - \pi_1 y_{t-1} - \cdots - \pi_{p-1} y_{t-p+1} - g(u_{t-1})$, $\omega > 0$, and $\alpha_i \geq 0$ $(i = 1, \ldots, q)$; a more general formulation for the conditional variance will be considered below. Note that a compact

expression for e_t is $e_t = u_t - g(u_{t-1})$ so that equation (3) can be expressed as $u_t = g(u_{t-1}) + \sigma_t \varepsilon_t$. If $\pi_1 = \cdots = \pi_{p-1} = 0$ in equation (3), we have $u_t = y_t$ so that the autoregressive order p reduces to one and equation (3) reduces to $y_t = g(y_{t-1}) + \sigma_t \varepsilon_t$.

Our first assumption contains basic requirements for the error term ε_t and makes clear that the squared volatility, σ_t^2, is the conditional variance of y_t (when appropriate moments exist).

Assumption 1. *$\{\varepsilon_t,\, t = 1, 2, \ldots\}$ is a sequence of IID random variables that is independent of $(y_0, \ldots, y_{1-p-q})$, has zero mean and unit variance, and the distribution of ε_1 has a (Lebesgue) density that is bounded away from zero on compact subsets of $\mathbb{R}$.*

Later on we introduce an assumption on the conditional variance σ_t^2 which further restricts the moments of ε_t.

To further describe the conditional mean of the autoregressions we consider, we next specify the conditions needed for the function g in equation (3). The following assumption is a simplification of Assumption 1 in Meitz and Saikkonen (2022) (the somewhat more general formulation used therein is briefly discussed at the end of this subsection).

Assumption 2.
(i) The roots of the polynomial $\varpi(z) = 1 - \pi_1 z - \cdots - \pi_{p-1} z^{p-1}$ lie outside the unit circle.
(ii) The function $g : \mathbb{R} \to \mathbb{R}$ in (3) is measurable, locally bounded, and satisfies $|g(u)| \to \infty$ as $|u| \to \infty$, and there exist positive constants r, M_0, K_0, and $0 < \rho < 2$ such that for all $u \in \mathbb{R}$

$$|g(u)| \leq \begin{cases} (1 - r\,|u|^{-\rho})\,|u| & \text{for } |u| \geq M_0, \\ K_0 & \text{for } |u| \leq M_0. \end{cases} \tag{5}$$

Assumption 2(i) corresponds to the conventional stationarity condition of a linear autoregression in that it requires the roots of the polynomial $\varpi(z)$ to lie outside the unit circle. In the first-order case $p = 1$, this condition becomes redundant because then $\pi_1 = \cdots = \pi_{p-1} = 0$. Assumption 2(ii) is needed to prove the subgeometric ergodicity of the process y_t, as already done by Fort and Moulines (2003) and Douc et al. (2004) in the first-order case $p = 1$ and by Meitz and Saikkonen (2022) for higher-order autoregressions.

We next provide some intuition and motivation for our model in (3). To clarify the role of inequality (5) restricting the function $g(\cdot)$, suppose Assumptions 1 and 2(i) hold but instead of Assumption 2(ii) suppose the function $g(\cdot)$ were linear with $g(u) = \pi_0 u$ and $\pi_0 \in [-1, 1]$. Using the lag operator L, equation (3) could then be written as

$$u_t - \pi_0 u_{t-1} = (1 - \pi_0 L)(1 - \pi_1 L - \cdots - \pi_{p-1} L^{p-1}) y_t = \sigma_t \varepsilon_t, \tag{6}$$

that is, as the familiar linear AR(p) model (with autoregressive heteroskedasticity). Given Assumptions 1 and 2(i), the case $\pi_0 \in (-1, 1)$ corresponds to geometric ergodicity of y_t and the cases $\pi_0 = \pm 1$ to non-ergodicity. Nonlinear functions $g(\cdot)$ satisfying Assumption 2(ii) provide a middle ground between these extreme cases of geometric ergodicity and non-ergodicity. For instance, if $g(u) = (1 - r\,|u|^{-\rho})u$ for $|u| > r^{1/\rho}$ and $g(u) = 0$ otherwise ($r > 0$, $0 < \rho < 2$), then for any fixed $\pi_0 \in (-1, 1)$ and for all u sufficiently large in absolute value (i.e., for the values of u that are crucial for determining ergodicity),

$$|\pi_0||u| < |g(u)| < |u|.$$

The subgeometrically ergodic autoregressions we consider thus provide one possibility for modeling small departures from unit root autoregressions. Assumption 2(ii) implies that for large values of $|u_{t-1}|$, the conditional mean of model (3) is close to that of an integrated process (of order one). On the other hand, as inequality (5) restricts the function $g(\cdot)$ only for large values of its argument, no restrictions (apart from the boundedness condition in (5)) are imposed when the argument takes values inside some bounded set of values. Thus the autoregressions we consider may exhibit rather arbitrary (stationary, unit root, explosive, nonlinear, etc.) behavior for moderate values of the observed series.

The autoregressions we consider are to some extent related to existing models that have autoregressive roots near unity. To illustrate, when $g(u)$ is as in the previous paragraph and we further set $p = r = \rho = 1$, the model in (3) simplifies to

$$y_t = \Big(1 - \frac{1}{|y_{t-1}|}\Big) y_{t-1} + e_t \quad \text{when } |y_{t-1}| > 1 \quad \text{and} \quad y_t = e_t \quad \text{otherwise}$$

where $e_t = \sigma_t \varepsilon_t$. In comparison, a prototypical local-to-unity autoregression could be expressed as

$$y_t = \Big(1 - \frac{1}{T}\Big) y_{t-1} + e_t, \quad t = 1, \ldots, T, \quad \text{where } T \text{ denotes the sample size.}$$

Both of the above formulations involve an autoregressive coefficient near unity, the former when the observed process takes on large (absolute) values and the latter when the sample size is large. However, the fact that the sample size is an essential part of local-to-unity autoregressions makes them quite different from the autoregressions we consider — in particular, the autoregressions we consider are ergodic. For more details on local-to-unity autoregressions and other related models, we refer the reader to the recent contributions of Lieberman and Phillips (2020) and Phillips (2023) and the references therein.

Homoskedastic subgeometrically ergodic autoregressions satisfying (a somewhat more general version of) Assumption 2 were already considered by Meitz and Saikkonen (2022). As many time series in economics, finance, and other fields exhibit conditional heteroskedasticity, in this paper we consider an extension to ARCH errors. In the homoskedastic case considered in Meitz and Saikkonen (2022), the term $g(u_{t-1})$ in (3) was replaced with the more general formulation $u_{t-1} + \tilde{g}(y_{t-1}, \ldots, y_{t-p})$ (with $\tilde{g}$ a real-valued function) to allow for more general dependence on the past through the variables $y_{t-1}, \ldots, y_{t-p}$ (and not only through the linear combination $u_{t-1} = y_{t-1} - \pi_1 y_{t-2} - \cdots - \pi_{p-1} y_{t-p}$). The present simpler formulation worked well in the empirical application of Section 4 and in some other examples we tried out, and leads to more transparent assumptions and streamlined proofs.

2.2 Companion form

To establish ergodicity, we need the companion form of the $(p+q)$-dimensional process $\boldsymbol{y}_t = (\boldsymbol{y}_{1,t}, \boldsymbol{y}_{2,t})$ with a p-dimensional $\boldsymbol{y}_{1,t} = (y_t, \ldots, y_{t-p+1})$ and a q-dimensional $\boldsymbol{y}_{2,t} = (y_{t-p}, \ldots, y_{t-p-q+1})$.

First we formulate the p-dimensional companion form related to equation (3), which reads as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_{p-1} & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ \vdots \\ y_{t-p} \end{bmatrix} + g(u_{t-1}) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} + \sigma_t \varepsilon_t \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

or, denoting the matrix in this equation with $\boldsymbol{\Phi}$ and setting $\boldsymbol{\iota}_p = (1, 0, \ldots, 0)$ $(p \times 1)$, as

$$\boldsymbol{y}_{1,t} = \boldsymbol{\Phi} \boldsymbol{y}_{1,t-1} + g(u_{t-1}) \boldsymbol{\iota}_p + \sigma_t \varepsilon_t \boldsymbol{\iota}_p \tag{7}$$

(when $p = 1$, $\boldsymbol{\Phi} = 0$ and $u_{t-1} = y_{t-1}$). As $\sigma_t = \sigma(\boldsymbol{y}_{t-1})$ depends on the whole $(p+q)$-dimensional vector $\boldsymbol{y}_{t-1}$, we have to expand (7) to the $(p+q)$-dimensional companion form

$$\begin{bmatrix} \boldsymbol{y}_{1,t} \\ \boldsymbol{y}_{2,t} \end{bmatrix} = \left[\begin{array}{ccc} & \boldsymbol{\Phi} & \mathbf{0}_{p \times q} \\ \hline \mathbf{0}_{q \times (p-1)} & I_q & \mathbf{0}_{q \times 1} \end{array}\right] \begin{bmatrix} \boldsymbol{y}_{1,t-1} \\ \boldsymbol{y}_{2,t-1} \end{bmatrix} + g(u_{t-1}) \boldsymbol{\iota}_{p+q} + \sigma_t \varepsilon_t \boldsymbol{\iota}_{p+q}, \tag{8}$$

where I_q is the $(q \times q)$ identity matrix and $\mathbf{0}_{* \times *}$ denotes a matrix of zeros with the indicated dimensions (and $\boldsymbol{\iota}_{p+q}$ is defined in the obvious way). This shows that $\boldsymbol{y}_t$ is a Markov chain on $\mathbb{R}^{p+q}$.

In order to establish ergodicity we further transform the p-dimensional companion form (7) in a way already used in Meitz and Saikkonen (2022, Sec 4). To this end we define the matrices

$$\mathbf{A} = \begin{bmatrix} 1 & -\pi_1 & -\pi_2 & \cdots & -\pi_{p-1} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Pi} = \mathbf{A} \boldsymbol{\Phi} \mathbf{A}^{-1} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & \pi_1 & \pi_2 & \cdots & \pi_{p-1} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0}_{1 \times (p-1)} \\ \boldsymbol{\iota}_{p-1} & \boldsymbol{\Pi}_1 \end{bmatrix}, \tag{9}$$

where $\mathbf{A}$ is nonsingular and $\boldsymbol{\Pi}_1$ is the $(p-1) \times (p-1)$ dimensional lower right hand corner of $\boldsymbol{\Pi}$ (when $p = 1$, $\mathbf{A} = 1$ and $\boldsymbol{\Pi} = 0$). With these definitions equation (7) can be transformed into

$$\mathbf{A} \boldsymbol{y}_{1,t} = \boldsymbol{\Pi} \mathbf{A} \boldsymbol{y}_{1,t-1} + g(u_{t-1}) \boldsymbol{\iota}_p + \sigma_t \varepsilon_t \boldsymbol{\iota}_p, \tag{10}$$

where $\mathbf{A} \boldsymbol{y}_{1,t} = (u_t, y_{t-1}, \ldots, y_{t-p+1})$. Now, for any p-dimensional vector $\boldsymbol{x}_1$, form the partition $\boldsymbol{x}_1 = (x_{1,1}, \ldots, x_{1,p}) = (x_{1,1}, \boldsymbol{x}_{1,2})$ and define

$$\boldsymbol{z}(\boldsymbol{x}_1) = \begin{bmatrix} z_1(\boldsymbol{x}_1) \\ \boldsymbol{z}_2(\boldsymbol{x}_1) \end{bmatrix} = \mathbf{A} \boldsymbol{x}_1 = \begin{bmatrix} x_{1,1} - \pi_1 x_{1,2} - \cdots - \pi_{p-1} x_{1,p} \\ \boldsymbol{x}_{1,2} \end{bmatrix} \tag{11}$$

(when $p = 1$, $\boldsymbol{x}_{1,2}$ and $\boldsymbol{z}_2(\boldsymbol{x}_1)$ are dropped). Using this notation equation (10) can be expressed

as $\boldsymbol{z}(\boldsymbol{y}_{1,t}) = \boldsymbol{\Pi}\boldsymbol{z}(\boldsymbol{y}_{1,t-1}) + g(z_1(\boldsymbol{y}_{1,t-1}))\boldsymbol{\iota}_p + \sigma(\boldsymbol{y}_{t-1})\varepsilon_t\boldsymbol{\iota}_p$, that is, as

$$\begin{aligned}\begin{bmatrix} z_1(\boldsymbol{y}_{1,t}) \\ \boldsymbol{z}_2(\boldsymbol{y}_{1,t}) \end{bmatrix} &= \begin{bmatrix} 0 & \mathbf{0}_{1\times(p-1)} \\ \boldsymbol{\iota}_{p-1} & \boldsymbol{\Pi}_1 \end{bmatrix} \begin{bmatrix} z_1(\boldsymbol{y}_{1,t-1}) \\ \boldsymbol{z}_2(\boldsymbol{y}_{1,t-1}) \end{bmatrix} + g(z_1(\boldsymbol{y}_{1,t-1}))\boldsymbol{\iota}_p + \sigma(\boldsymbol{y}_{t-1})\varepsilon_t\boldsymbol{\iota}_p \\ &= \begin{bmatrix} g(z_1(\boldsymbol{y}_{1,t-1})) + \sigma(\boldsymbol{y}_{t-1})\varepsilon_t \\ \boldsymbol{\Pi}_1\boldsymbol{z}_2(\boldsymbol{y}_{1,t-1}) + z_1(\boldsymbol{y}_{1,t-1})\boldsymbol{\iota}_{p-1} \end{bmatrix}. \end{aligned} \tag{12}$$

Here the first equation is in a form where the autoregressive order is one and the volatility term is a function of the $(p+q)$-dimensional vector $\boldsymbol{y}_{t-1} = (\boldsymbol{y}_{1,t-1}, \boldsymbol{y}_{2,t-1})$ whereas the second equation involves the p-dimensional vector $\boldsymbol{y}_{1,t-1}$ only.

By Assumption 2(i), the roots of the polynomial $\varpi(z)$ lie outside the unit circle, so that the eigenvalues of the matrix $\boldsymbol{\Pi}_1$ in the second equation in (12) are smaller than one in absolute value. As is well known, this implies the existence of a matrix norm of $\boldsymbol{\Pi}_1$ that is also smaller than one. Specifically, for any vector norm $\|\cdot\|$, denote by $|||\cdot|||$ the corresponding induced matrix norm (Horn and Johnson, 2013, Defn 5.6.1); that is, for any conformable square matrix A, set

$$|||A||| = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|.$$

Then we obtain the following result (Horn and Johnson, 2013, Lemma 5.6.10).

Lemma 1. *There exists a vector norm $\|\cdot\|_*$ and a corresponding induced matrix norm $|||\cdot|||_*$ such that $|||\boldsymbol{\Pi}_1|||_* = \varpi < 1$.*

The existence of an induced matrix norm with the property in the above lemma is essential in our proofs. (When $p = 1$, Assumption 2(i) and Lemma 1 are redundant.) The norms $\|\cdot\|_*$ and $|||\cdot|||_*$ are defined on $\mathbb{R}^{p-1}$ and $\mathbb{R}^{(p-1)\times(p-1)}$, respectively, and they have been commonly used in time series models. In the next subsection we introduce norms which are of a different type.

2.3 Conditional variance

The root condition of Assumption 2(i) and inequality (5) of Assumption 2(ii) are of major importance for establishing the stability of our model. However, as these conditions only concern the conditional mean, we need additional assumptions restricting the conditional variance σ_t^2. As an extension of the basic ARCH model (4) we consider a nonlinear formulation of the conditional variance defined as

$$\sigma_t^2 = \zeta_{0,t-1}\omega + \alpha_1\zeta_{1,t-1}e_{t-1}^2 + \cdots + \alpha_q\zeta_{q,t-1}e_{t-q}^2, \tag{13}$$

where $\zeta_{i,t-1} = \zeta_i(\boldsymbol{y}_{t-1})$ is a function of $\boldsymbol{y}_{t-1}$ $(i = 0, \ldots, q)$ and otherwise the notation is as in equation (4) (including the conditions $\omega > 0$ and $\alpha_1, \ldots, \alpha_q \geq 0$). When the functions $\zeta_{i,t-1}$ are the same for all $i = 0, \ldots, q$ we remove the index i and use the notations ζ_{t-1} and $\zeta(\cdot)$. This is the case in our empirical example where $\zeta_{t-1} = \zeta(y_{t-1}) = 1/(1+e^{-\gamma(y_{t-1}-a)})$ is a logistic function depending only on y_{t-1}. For possible alternatives we consider a more general formulation and introduce the following assumption.

Assumption 3. *In equation (13), the following conditions are assumed. (i) The parameters* $\omega, \alpha_1, \ldots, \alpha_q$ *satisfy* $\omega > 0$, $\alpha_1, \ldots, \alpha_q \geq 0$, *and* $\sum_{i=1}^q \alpha_i < 1$. *(ii) For each* $i = 0, \ldots, q$, *the function* ζ_i *takes values in* $(0, 1]$.

The above assumption includes the case $\zeta_i \equiv 1$ for all i, which corresponds to the linear ARCH model (4). It covers also the above-mentioned logistic function.

Consider the q-dimensional process $\boldsymbol{\xi}_t = (e_t^2, e_{t-1}^2, \ldots, e_{t-q+1}^2)$ $(t \geq 1)$ with initial values $\boldsymbol{\xi}_0 = (e_0^2, \ldots, e_{-q+1}^2)$ where $e_0^2, \ldots, e_{-q+1}^2$ are functions of $\boldsymbol{y}_0$. Inspired by Cline and Pu (2004, Example 4.2) we now introduce the following equation which is a straightforward implication of equation (13) and the fact $\sigma_t \varepsilon_t = e_t$:

$$\begin{bmatrix} e_t^2 \\ e_{t-1}^2 \\ \vdots \\ \vdots \\ e_{t-q+1}^2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \zeta_{1,t-1} \varepsilon_t^2 & \alpha_2 \zeta_{2,t-1} \varepsilon_t^2 & \cdots & \alpha_{q-1} \zeta_{q-1,t-1} \varepsilon_t^2 & \alpha_q \zeta_{q,t-1} \varepsilon_t^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} e_{t-1}^2 \\ e_{t-2}^2 \\ \vdots \\ \vdots \\ e_{t-q}^2 \end{bmatrix} + \begin{bmatrix} \zeta_{0,t-1} \varepsilon_t^2 \omega \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

$(t = 1, 2, \ldots)$ or, more briefly,

$$\boldsymbol{\xi}_t = \Lambda_{\zeta,t} \boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_{\zeta,t}, \quad t = 1, 2, \ldots; \tag{14}$$

as $\boldsymbol{\xi}_t$ is a function of $\boldsymbol{y}_t$, we occasionally write $\boldsymbol{\xi}_t = \boldsymbol{\xi}(\boldsymbol{y}_t)$. For later purposes we also note that due to the identities $\sigma_t \varepsilon_t = e_t$ and $e_t^2 = \boldsymbol{\iota}_q' \boldsymbol{\xi}(\boldsymbol{y}_t)$ we have

$$\sigma_t^2 = \sigma^2(\boldsymbol{y}_{t-1}) = E[\boldsymbol{\iota}_q' \boldsymbol{\xi}(\boldsymbol{y}_t) \mid \boldsymbol{y}_{t-1}]. \tag{15}$$

When there is need to make the dependence of $\Lambda_{\zeta,t}$ on $\boldsymbol{y}_{t-1}$ explicit we use the notation $\Lambda_{\zeta,t}(\boldsymbol{y}_{t-1})$ and replace the (random) argument $\boldsymbol{y}_{t-1}$ by a fixed counterpart when needed. Specifically, $\Lambda_{\zeta,t}(\boldsymbol{x})$ means that the functions $\zeta_{i,t-1} = \zeta_{i,t-1}(\boldsymbol{y}_{t-1})$ used in $\Lambda_{\zeta,t}(\boldsymbol{y}_{t-1})$ are replaced by $\zeta_{i,t}(\boldsymbol{x})$ for all $i = 1, \ldots, q$, and the notations $\sigma^2(\boldsymbol{x})$ and $\boldsymbol{\xi}(\boldsymbol{x})$ are used similarly.

We also define the matrices

$$\Lambda_t = \begin{bmatrix} \alpha_1 \varepsilon_t^2 & \alpha_2 \varepsilon_t^2 & \cdots & \alpha_{q-1} \varepsilon_t^2 & \alpha_q \varepsilon_t^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \Lambda = E[\Lambda_t] = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{q-1} & \alpha_q \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}. \tag{16}$$

Note that Λ_t is obtained from the matrix $\Lambda_{\zeta,t}$ by choosing $\zeta_{i,t} = 1$ for all $i = 1, \ldots, q$. Similarly, we denote $\boldsymbol{\omega}_t = (\omega_t, 0, \ldots, 0)'$ with $\omega_t = \omega \varepsilon_t^2$ and $\boldsymbol{\omega} = E[\boldsymbol{\omega}_t] = (\omega, 0, \ldots, 0)'$.

In our proofs we need to appropriately control the size of the random matrix Λ_t, and not just the size of the non-random matrix $\Lambda = E[\Lambda_t]$. This is the reason why we next consider vector and matrix norms more complicated than those in Lemma 1. To this end, we first recall the definition of an L^p-norm (for convenience, in this subsection only, we use the notation p in L^p-norms; elsewhere in the paper p stands for the autoregressive order in model (3)). If $\|\cdot\|$ is

any vector norm on $\mathbb{R}^q$ and $\boldsymbol{v}$ is a q-dimensional random vector, equation

$$\|\boldsymbol{v}\|_{L^p} = (E[\|\boldsymbol{v}\|^p])^{1/p} \qquad (1 \leq p < \infty)$$

defines an L^p-norm on the set of (equivalence classes of almost surely equal) $q \times 1$ random vectors that are p-integrable (see, e.g., Dudley, 2004, Secs 5.1 and 5.2). It may be worth noting that for nonrandom vectors there is no difference between the norms $\|\cdot\|$ and $\|\cdot\|_{L^p}$ but for random vectors the outcome of $\|\cdot\|$ is random and that of $\|\cdot\|_{L^p}$ is nonrandom. This L^p-norm can be used to induce a norm for random matrices; for the conventional non-random matrix case and for the terminology used below, see Horn and Johnson (2013, Def. 5.6.1 and Sec 5.6.). Specifically, to define a generalized (non-submultiplicative) matrix norm $|||\cdot|||_{L^p}$, for any $q \times q$ random matrix A set

$$|||A|||_{L^p} = \max_{\|\boldsymbol{x}\|_{L^p}=1} \|A\boldsymbol{x}\|_{L^p} = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|_{L^p} \qquad (\boldsymbol{x} \in \mathbb{R}^q), \tag{17}$$

where the latter equality holds as $\boldsymbol{x}$ is nonrandom. This defines a generalized matrix norm[1] on the set of (equivalence classes of almost surely equal) $q \times q$ random matrices with p-integrable entries; moreover, the norms $\|\cdot\|$, $\|\cdot\|_{L^p}$, and $|||\cdot|||_{L^p}$ are related by the inequality[2]

$$\|A\boldsymbol{x}\|_{L^p} \leq |||A|||_{L^p}\|\boldsymbol{x}\| \qquad (\boldsymbol{x} \in \mathbb{R}^q). \tag{18}$$

We next state a high-level condition that assumes the existence of a vector norm on $\mathbb{R}^q$ with particular additional properties. (Primitive conditions ensuring this high-level assumption will be given momentarily.) One of these properties is monotonicity in the sense of Definition 5.4.18 of Horn and Johnson (2013): a vector norm $\|\cdot\|$ is monotone if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^q$ satisfying $|x_i| \leq |y_i|$ for $i = 1, \ldots, q$ always implies that $\|\boldsymbol{x}\| \leq \|\boldsymbol{y}\|$. For clarity, we use the notation $\|\cdot\|_\bullet$ for the specific vector norm in the assumption below; similarly, we denote the related L^p-norm by $\|\cdot\|_{\bullet L^p}$ and the generalized matrix norm by $|||\cdot|||_{\bullet L^p}$. We also introduce two constants, $s_0 \geq 1$ and $b \geq 1$, such that

$$b = 1 \text{ when } s_0 = 1 \quad \text{and} \quad b > (2s_0 - \rho)/[s_0(2-\rho)] > 1 \text{ when } s_0 > 1$$

(recall from Assumption 2 that $\rho \in (0, 2)$ so that $2s_0 > \rho$). These constants are used in the next section where we establish our ergodicity result and there the size of s_0 will have an effect on the rate of convergence obtained and the order of moments that are finite. The rather complex conditions required from the constant b are due to the connection between the conditional mean and ARCH errors (this connection disappears when $s_0 = 1$ as it also does in subgeometric homoskedastic autoregressions).

Assumption 4. *Suppose there exists a vector norm $\|\cdot\|_\bullet$ on $\mathbb{R}^q$ that is (i) monotone and (ii) such that $|||\Lambda_t|||_{\bullet L^{bs_0}} = \lambda < 1$, where b and s_0 are as described above.*

[1] Axioms (1), (1a), (2), and (3) of a generalized matrix norm (see Horn and Johnson, 2013, pp. 340–341) can be checked similarly as in the proof of Theorem 5.6.2(c) of the same reference (replacing the norms $\|\cdot\|$ and $|||\cdot|||$ therein with $\|\cdot\|_{L^p}$ and $|||\cdot|||_{L^p}$, replacing appropriate statements therein with their almost sure counterparts, and using Minkowski's inequality as an additional justification for axiom (3)).

[2] Inequality (18) can be verified analogously to Theorem 5.6.2(b) of Horn and Johnson (2013).

This assumption tacitly requires that $E[|\varepsilon_t|^{2bs_0}]$ is finite, thereby strengthening Assumption 1 (when $s_0 > 1$). Assumption 4 is formulated in a way that is convenient in our proofs but is not very transparent. The following lemma gives primitive conditions ensuring that Assumption 4 holds (for a proof, see the appendix).

Lemma 2. *Suppose that Assumptions 1 and 3 hold and also that the parameters $\alpha_1, \ldots, \alpha_q$ in Assumption 3 satisfy $\sum_{i=1}^{q} \alpha_i < 1/\bar{\mu}_{2bs_0}$ where $\bar{\mu}_{2bs_0} = (E[|\varepsilon_1^2|^{bs_0}])^{1/bs_0}$. Then Assumption 4 holds.*

To illustrate, consider the case $q = s_0 = b = 1$ so that the condition in Lemma 2 reduces to the requirement $\alpha_1 < 1$. In geometrically ergodic AR models with linear ARCH(1) errors, $\alpha_1 < 1$ is the usual requirement for covariance stationarity while geometric ergodicity can hold under even weaker conditions, such as $E[\ln(\alpha_1 \varepsilon_t^2)] < 0$ (see, e.g., Meitz and Saikkonen, 2010, Assumption 3 and Thm 1 for further details). In the present setting the situation is different: as will be seen in Section 3.2, condition $\alpha_1 < 1$ does not guarantee a finite variance.

3 Subgeometric ergodicity at a polynomial rate

3.1 Main result

We now consider the stability of the model introduced in the previous section. We begin with a brief account of some necessary Markov chain concepts (for more comprehensive discussions, see Meyn and Tweedie (2009) and Douc et al. (2018) and also Meitz and Saikkonen (2022, Sec 2)). Let X_t $(t = 0, 1, 2, \ldots)$ be a Markov chain on a general measurable state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ (with $\mathcal{B}(\mathsf{X})$ the Borel σ-algebra) and let $P^n(x\,;\, A) = \Pr(X_n \in A \mid X_0 = x)$ signify its n-step transition probability measure. For an arbitrary fixed measurable function $f : \mathsf{X} \to [1, \infty)$ and for any signed measure μ, define the f-norm $\|\mu\|_f$ as

$$\|\mu\|_f = \sup_{f_0 : |f_0| \leq f} |\mu(f_0)|\,, \tag{19}$$

where $\mu(f_0) = \int_{x \in \mathsf{X}} f_0(x)\mu(dx)$ and the supremum in (19) runs over all measurable functions $f_0 : \mathsf{X} \to \mathbb{R}$ such that $|f_0(x)| \leq f(x)$ for all $x \in \mathsf{X}$ (when $f \equiv 1$, the f-norm $\|\mu\|_f$ reduces to the total variation norm $\|\mu\|_{TV} = \sup_{f_0 : |f_0| \leq 1} |\mu(f_0)|$ used in (1) and (2)). When the n-step probability measures $P^n(x\,;\,\cdot)$ converge in f-norm and at rate $r(n)$ to the stationary probability measure π satisfying $\pi(f) < \infty$, that is,

$$\lim_{n \to \infty} r(n) \|P^n(x\,;\,\cdot) - \pi\|_f = 0 \qquad \text{for } \pi\text{-almost all } x \in \mathsf{X},\ ^{3} \tag{20}$$

we say that the Markov chain X_t is (f, r)-ergodic; this implicitly entails the existence of π as well as certain moments as $\pi(f) < \infty$. In the conventional geometrically ergodic case, $r(n) = r^n$ for some $r > 1$. To establish (f, r)-ergodicity, we use a so-called drift condition defined as follows (here $\mathbf{1}_S(x)$ denotes the indicator function taking value one when x belongs to the set S and zero elsewhere).

[3] That is, the convergence in (20) is required to hold for all $x \in \mathsf{X}$ except for those x in a set that has probability zero with respect to the stationary measure π.

Condition D. There exist a measurable function $V : \mathsf{X} \to [1, \infty)$, a concave increasing continuously differentiable function $\phi : [1, \infty) \to (0, \infty)$, a measurable set C, and a finite constant $\tilde{b}$ such that

$$E\left[V(X_1) \mid X_0 = x\right] \leq V(x) - \phi\left(V(x)\right) + \tilde{b}\mathbf{1}_C(x), \qquad x \in \mathsf{X}. \tag{21}$$

The idea is to verify this condition with suitable functions V and ϕ, which together with some additional conditions ensures the (f, r)-ergodicity of the process X_t; for more details, see Meitz and Saikkonen (2022, Thm 1).

Now consider the stability of the Markov chain $\boldsymbol{y}_t$ on $\mathbb{R}^{p+q}$ given in (8). To define the function V in (21), we use the functions $z_1(\cdot)$, $\boldsymbol{z}_2(\cdot)$, and $\boldsymbol{\xi}(\cdot)$ in (11)–(12) and (14) and the norms $\|\cdot\|_*$ and $\|\cdot\|_\bullet$ in Lemma 1 and Assumption 4. Set $\boldsymbol{x} = (x_1, \ldots, x_{p+q}) \in \mathbb{R}^{p+q}$ and decompose $\boldsymbol{x}$ to its p- and q-dimensional components as $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$. We define the function V as

$$V(\boldsymbol{x}) = 1 + |z_1(\boldsymbol{x}_1)|^{2s_0} + s_1\|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha} + s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}, \tag{22}$$

where s_0 and b are defined above Assumption 4, s_1 and s_2 are positive constants to be specified later (with s_1 small and s_2 large), and $\alpha = 1 - \rho/2s_0$ (recall from Assumption 2 that $\rho \in (0, 2)$ so that $\alpha \in (0, 1)$). It may be clarifying to note that when $p = 1$, model (3) reduces to $y_t = g(y_{t-1}) + \sigma_t\varepsilon_t$; then we can set $s_1 = 0$ and drop $\boldsymbol{z}_2(\boldsymbol{x}_1)$ so that the function V in (22) becomes $V(\boldsymbol{x}) = 1 + |x_1|^{2s_0} + s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}$.

To verify Condition D, we need to consider the conditional expectation

$$\begin{aligned} E\left[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}\right] = 1 &+ E\left[|z_1(\boldsymbol{y}_{1,1})|^{2s_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}\right] + s_1 E\left[\|\boldsymbol{z}_2(\boldsymbol{y}_{1,1})\|_*^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}\right] \\ &+ s_2 E\left[\|\boldsymbol{\xi}(\boldsymbol{y}_1)\|_\bullet^{bs_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}\right], \end{aligned} \tag{23}$$

bound the conditional expectations on the right hand side of (23), and express these bounds in a way which conforms to inequality (21) with the function ϕ satisfying the conditions required in Condition D. These considerations, combined with the checking of some additional technical conditions, lead to the following theorem (the proof can be found in the Appendix).

Theorem 1. *Consider the Markov chain $\boldsymbol{y}_t$ defined in (8). Suppose that Assumptions 1–4 hold and that $V(\boldsymbol{x})$ is as in (22). Then $\boldsymbol{y}_t$ is (f, r)-ergodic with the polynomial convergence rate $r(n) = n^{\delta-1}$ and the function f given by $f(\boldsymbol{x}) = V(\boldsymbol{x})^{1-\delta\rho/2s_0}$; this result holds for any choice of $\delta \in [1, 2s_0/\rho]$ and for some (small enough) $s_1 > 0$ and some (large enough) $s_2 > 0$.*

Theorem 1 provides the first subgeometric ergodicity results for autoregressions with autoregressive conditional heteroskedasticity. In this theorem, the convergence rate $r(n)$ shows the speed at which the n-step transition probability measures of the process $\boldsymbol{y}_t$ converge to the stationary probability measure. Due to the polynomial convergence rate we therefore call the process $\boldsymbol{y}_t$ polynomially ergodic. Note also that the choice of δ in Theorem 1 allows for a trade-off between the rate of convergence and the size of the f-norm.

3.2 Discussion

Geometric ergodicity. In previous literature, geometric ergodicity of nonlinear autoregressions with ARCH errors has been considered using a variety of different assumptions for the allowed nonlinear dynamics and for the required moment conditions for the innovations; see, e.g., Cline and Pu (2004), Meitz and Saikkonen (2010), and the many references therein.

Homoskedastic case. Theorem 1 remains valid also in the homoskedastic case (obtained by setting $\alpha_1 = \cdots = \alpha_q = 0$). Previous polynomial ergodicity results for homoskedastic autogressions were obtained by Fort and Moulines (2003, Sec 2.2) and Meitz and Saikkonen (2022, Thm 3), and the above Theorem 1 provides partial improvements over these earlier results in certain cases. Assumptions and notation are slightly different in all the papers, but (in the notation of the present paper) Theorem 1 improves earlier results when $1 \leq \rho < 2$ and $1 < s_0 < 2$.

Proof strategy. The proof of Theorem 1 is also somewhat different from the previous polynomial ergodicity results in Fort and Moulines (2003, Sec 2.2) and Meitz and Saikkonen (2022, Thm 3). A rather obvious difference is that these earlier results deal with homoskedastic autoregressions whereas our model contains a nonlinear ARCH term, the size of which is controlled with the special matrix norm defined in Assumption 4. Regarding the conditional expectation, the mentioned earlier results rely on Lemma 3 in Fort and Moulines (2003) while our proof of Theorem 1 avoids the use of this lemma, and instead makes use of a (relatively unknown) extension of Bernoulli's inequality due to Fefferman and Shapiro (1972) (combined with Young's inequality).

Mixing and moment results. As already indicated in the Introduction, the polynomial ergodicity result of Theorem 1 also implies that the process $\boldsymbol{y}_t$ is β-mixing (and hence α-mixing). Moreover, the convergence rate of the β-mixing coefficients $\beta(n)$ is given by the fastest convergence rate, that is, $\lim_{n\to\infty} n^{2s_0/\rho-1}\beta(n) = 0$. For further details and justifications of these mixing results, see Meitz and Saikkonen (2021, Thm 2) and Meitz and Saikkonen (2022, Sec 2).

Another consequence of Theorem 1 is that the stationary distribution of $\boldsymbol{y}_t$ has finite moments up to order $2s_0 - \rho$ (for a proof, see the Appendix). Note that depending on the values of $s_0 \geq 1$ and $\rho \in (0, 2)$, the order of these finite moments may be very small; in particular, when $s_0 = 1$ we do not obtain a finite variance.

Subexponential ergodicity. Theorem 1 concerns only polynomial ergodicity of subgeometric AR–ARCH models, and does not consider subexponential ergodicity (where the rate $r(n)$ in (2) equals, say, e^{cn^γ} with $c > 0$ and $0 < \gamma < 1$). The reason for this is that the properties of ARCH-type models do not seem compatible with the moment requirements needed for subexponential ergodicity. To elaborate on this, first note that the previous results of Douc et al. (2004, Sec 3.3) and Meitz and Saikkonen (2022, Sec 4.1) on subexponential ergodicity of homoskedastic nonlinear autoregressions (i) require the IID error term to possess moments of *all* orders and (ii) imply that the observed process y_t also has finite moments of *all* orders. (To

provide some further details, (ii) is given as Corollary to Theorem 2 in Meitz and Saikkonen (2022). As for (i), see Assumptions 3.3 and 2(a) of Douc et al. (2004) and Meitz and Saikkonen (2022), respectively. These assumptions require the IID error terms to be sub-Weibull random variables, which in turn entails they possess moments of all orders; see Vladimirova et al. (2020, Defn 1 and Thm 1) or Wong et al. (2020, Defn 3 and Lemma 5).)

The abovementioned moment requirements are in stark contrast to ARCH-type models. For instance, in the simplest ARCH(1) model ($e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \omega + \alpha_1 e_{t-1}^2$, and ε_t IID N(0,1)), the finiteness of moments of order $2r$ for e_t ($E[|e_t|^{2r}] < \infty$) is known to require the condition $\alpha_1^r E[|\varepsilon_t|^{2r}] < 1$ (see, e.g., Ling and McAleer, 2002, Thm 2.1 and Ling, 1999, Example 6.1). For integer values of r, this condition is equivalent with $\alpha_1 < [(2r-1)!!]^{-1/r} = [1 \cdot 3 \cdot \ldots \cdot (2r-1)]^{-1/r}$ and consequently *all* moments of the ARCH process e_t cannot be finite unless $\alpha_1 = 0$. The situation is similar also in more complicated (G)ARCH and AR–(G)ARCH models (see, e.g., Meitz and Saikkonen, 2008a, Thm 2 and Meitz and Saikkonen, 2008b, Thm 1, respectively). This suggests that ARCH-type heteroskedastic errors may not be compatible with the moment requirements needed for subexponential ergodicity.

Potential extensions. Extending our results to allow for GARCH (and not only ARCH) errors would be interesting. However, previous literature suggests that studying the stability of nonlinear AR–GARCH models can be challenging. Geometric ergodicity of nonlinear AR–GARCH models has previously been studied by Liu, Li, and Li (1997), Ling (1999), Cline (2007), and Meitz and Saikkonen (2008b); of these articles, the former two are confined to threshold AR–GARCH models, whereas the latter two consider more general nonlinear autoregressions. In the present setting, the autoregressive part of the model we consider is rather general (the restrictions imposed on function $g(\cdot)$ in Assumption 2(ii) are quite mild, essentially restricting $g(\cdot)$ only for large values of its argument) and techniques used for threshold models can not be applied. Using an approach similar to Cline's (2007) appears challenging as the assumptions he employs are quite general and appear difficult to verify (in fact, a threshold AR–GARCH model is the only example that is explicitly treated in his article). On the other hand, Meitz and Saikkonen (2008b) require certain structure and smoothness of the conditional mean (see Assumption 2 of their paper) and it is not clear how to apply these results in the current setting. As the extension to GARCH errors appears challenging, we leave it for future research.

Another useful extension would be to consider the subgeometric ergodicity of multivariate autoregressions with autoregressive conditional heteroskedasticity. Fort and Moulines (2003, Sec 2.2) and Douc et al. (2004, Sec 3.3) already studied multivariate first-order autoregressions with IID errors and obtained results for polynomial and subexponential ergodicity, respectively. In principle, generalizing these results to the higher-order case with multivariate ARCH errors should be possible but it is not immediate how to formulate a general model that would be both theoretically manageable as well as useful in practical applications. We hope to return to this issue in subsequent work.

3.3 Examples

The conditional mean of the model we have so far discussed is very general, and we next consider some concrete illustrating examples. The following two special cases were introduced in Meitz and Saikkonen (2022, Sec 5) in the case of a homoskedastic error term. We first consider a model with a time-varying intercept term based on a logistic function and specified as

$$y_t = \nu_1 L(u_{t-1};\gamma,a_1) + \nu_2(1-L(u_{t-1};\gamma,a_2)) + y_{t-1} + \pi_{t-1}\Delta y_{t-1} + \cdots + \pi_{p-1}\Delta y_{t-p+1} + \sigma_t\varepsilon_t, \quad (24)$$

where $L(u;\gamma,a) = 1/(1+e^{-\gamma(u-a)})$ is the logistic function and the parameters γ, a_1, a_2 are assumed to satisfy $\gamma > 0$ and $a_1 \leq a_2$, and ν_1, ν_2 are assumed to satisfy $\nu_1 < 0 < \nu_2$. Moreover, Δ signifies the difference operator (so that $\Delta y_{t-1} = y_{t-1} - y_{t-2}$) and the remaining notation is as in model (3). Arguments similar to those in Meitz and Saikkonen (2022, proof of Proposition 1) can now be used to prove the following result (for details, see the Appendix).

Proposition 1. *Consider the process y_t defined in equation (24) and suppose that Assumptions 1, 2(i), 3, and 4 hold. Then, $\boldsymbol{y}_t$ is polynomially ergodic with convergence rate $r(n) = n^{2s_0-1}$ and finite moments up to order $2s_0 - 1$.*

The convergence rate presented in Proposition 1 also shows the rate of β-mixing coefficients.

As another special case, we consider a model with a time-varying slope term defined as

$$y_t = \pi_{t-1}y_{t-1} + \cdots + \pi_{p-1}y_{t-p+1} + S(u_{t-1})u_{t-1} + \sigma_t\varepsilon_t, \quad (25)$$

where $S(u_{t-1})$ is either $S_1(u_{t-1}) = 1 - r_0/h(u_{t-1})$ or $S_2(u_{t-1}) = \exp\{-r_0/h(u_{t-1})\}$ (with $r_0 > 0$) and the function $h : \mathbb{R} \to (0,\infty)$ as defined in Proposition 2 of Meitz and Saikkonen (2022, Sec 5.2). In addition to a general formulation of the function h that proposition provides six special cases of which two are $h(u) = 1 + |u-a|^\rho$ and $h(u) = (1+(u-a)^2)^{\rho/2}$ (where $a \in \mathbb{R}$ and $\rho \in (0,2)$; see Assumption 2). Regarding the remaining notation, it is as in model (3).

The following result can be established by using arguments similar to those in the proof of Proposition 2 in Meitz and Saikkonen (2022, Sec 5.2) (for details, see the Appendix).

Proposition 2. *Consider the process y_t defined in equation (25) and suppose that Assumptions 1, 2(i), 3, and 4 hold. Then, $\boldsymbol{y}_t$ is polynomially ergodic with convergence rate $r(n) = n^{2s_0/\rho-1}$ and finite moments up to order $2s_0 - \rho$.*

The rate of β-mixing coefficients coincides with the rate given in the proposition. As the function h depends on the parameter $\rho \in (0,2)$, the convergence rate in Proposition 2 differs from that obtained in Proposition 1 except in the case $\rho = 1$.

4 Empirical application

Although theoretical work on subgeometric ergodicity has been ongoing for four decades, practical illustrations of (homoskedastic) subgeometrically ergodic autoregressions have been scarce; we are not aware of *any* previous empirical applications of subgeometrically ergodic autoregressions using real data. A small illustration of simulated data from one subgeometrically ergodic autoregression is given in Fort and Moulines (2003, Sec 3). Meitz and Saikkonen (2022, Sec 5)

provide examples of some concrete subgeometrically ergodic autoregressive time series models and illustrations of a few simulated data series from them. These simulation exercises suggest that subgeometrically ergodic autoregressions could be useful when the observed time series bears some resemblance to unit root type behavior and the autocorrelation function indicates very strong persistence, but when the time series nevertheless exhibits eventual mean-reverting behavior. The discussion in Section 2.1 around equation (6) had a similar message, suggesting these models could be seen as a middle ground between the extreme cases of geometric ergodicity and non-ergodicity.

We next illustrate the use of subgeometrically ergodic AR–ARCH models in a small empirical example. Our aim is simply to provide a proof of concept for the applicability of subgeometrically ergodic AR–ARCH models, illustrating that the model used fits the data well. Further work is certainly needed to judge the usefulness of these models in practical applications but we leave such more comprehensive empirical applications for future research.

The data we employ consists of daily observations on the Chicago Board Options Exchange energy sector volatility index (`fred.stlouisfed.org/series/VXXLECLS`) over the period 16 March 2011 through 31 December 2021 (a total of 2719 observations). This data series reflects energy sector risk and is displayed in the top left graph of Figure 1 (the solid graph; the dashed horizontal line shows the estimate $\hat{a} = 25.366$, see (26) and (27) below). The time series plot shows signs of strong persistence, which is also reflected in the autocorrelation function of the data shown in the top right graph of Figure 1.

We model this data series using the parametric specification in (24). As for the error distribution, after some experimentation a skew version of the t-distribution due to Jones and Faddy (2003) was found to provide a good fit (in contrast, estimation with normal errors lead to a distinct discrepancy between the residual distribution and the Gaussian one). The density function of this distribution is

$$f(x;c,d) = C_{c,d}^{-1}\left\{1+\frac{x}{(c+d+x^2)^{1/2}}\right\}^{c+1/2}\left\{1-\frac{x}{(c+d+x^2)^{1/2}}\right\}^{d+1/2},$$

where c and d are positive parameters and $C_{c,d} = 2^{c+d-1}B(c,d)(c+d)^{1/2}$ (with $B(\cdot,\cdot)$ denoting the beta function); the case $c = d$ results in a symmetric t-distribution with $2c$ degrees of freedom and the cases $c < d$ and $c > d$ imply skewness to the left and right, respectively. In our application, we use this distribution centralized to have mean zero and standardized to have unit variance (i.e., in the density function x is replaced by $sx + m$ and $C_{c,d}^{-1}$ by $sC_{c,d}^{-1}$ where m and s^2 denote the mean and variance, see Jones and Faddy, 2003, Sec 2.1; this requires that $c > 1$ and $d > 1$, for a moment of order k is finite when $c > k/2$ and $d > k/2$).

We estimate the model parameters using the method of maximum likelihood and employ optimization routines in R. (We simply assume that standard properties of maximum likelihood estimators hold and calculate standard errors based on the standard formulas.) Trying out different model orders lead to model (24) with order $p = 1$ and with a nonlinear ARCH term of order $q = 3$ (with these choices the residual diagnostics shown in Figure 2 in Appendix B indicated a very good fit). Specifically, the considered model is

$$\begin{aligned} y_t &= y_{t-1} - \nu L(y_{t-1};\gamma,a) + \nu(1 - L(y_{t-1};\gamma,a)) + \sigma_t\varepsilon_t \\ \sigma_t^2 &= (\omega + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \alpha_3 e_{t-3}^2)L(y_{t-1};\gamma,a), \end{aligned} \tag{26}$$

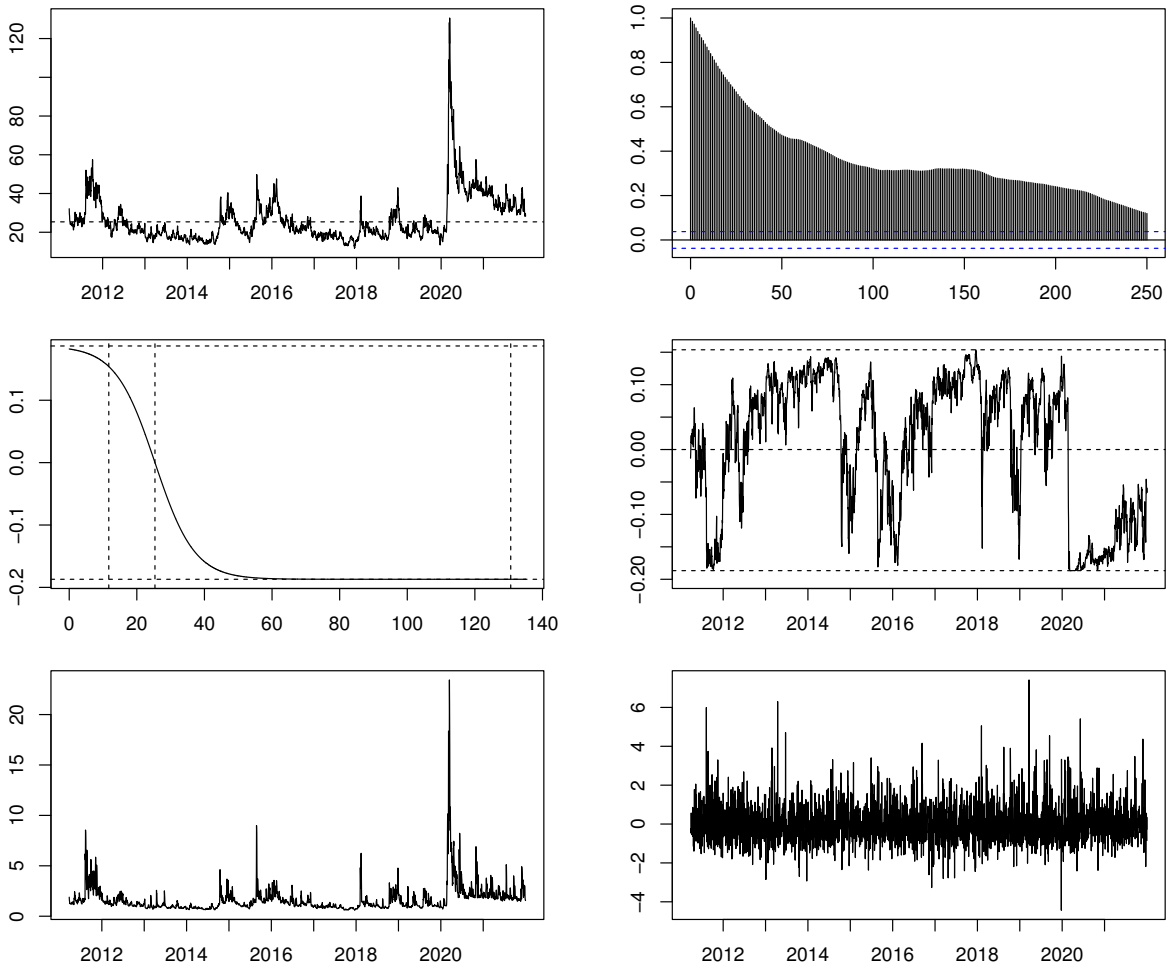

Figure 1: Top row: Daily observations on the Chicago Board Options Exchange energy sector volatility index, 16 March 2011 – 31 December 2021 (left); the corresponding autocorrelation function (right). Middle row: The function $I(x) = -\nu L(x;\gamma,a) + \nu(1 - L(x;\gamma,a))$ (left) and the corresponding time-varying intercept term $I(y_{t-1})$ (right), based on parameter estimates in (27). Bottom row: The estimated volatility series $\hat{\sigma}_t$ (left) and residual series $\hat{\varepsilon}_t$ (right), based on parameter estimates in (27).

where the errors ε_t are IID$(0,1)$ and follow the above described (centralized and standardized) skew t-distribution, $L(y;\gamma,a) = 1/(1+e^{-\gamma(y-a)})$ is the logistic function, the parameters ν and γ are positive, and $a \in \mathbb{R}$. (We also tried a model where the logistic functions in the conditional expectation and in the ARCH term were different but this extension had only a minor effect on the results.) ML estimation (with the constraints $\nu,\gamma,\omega > 0$, $\alpha_1,\alpha_2,\alpha_3 \geq 0$, and $\alpha_1+\alpha_2+\alpha_3 < 1$ that ensure polynomial ergodicity by Proposition 1) leads to the following results:

$$\begin{aligned} y_t &= y_{t-1} - \underset{(0.040)}{0.187} L(y_{t-1}; \underset{(0.018)}{0.171}, \underset{(1.434)}{25.366}) + \underset{(0.040)}{0.187}(1 - L(y_{t-1}; \underset{(0.018)}{0.171}, \underset{(1.434)}{25.366})) + \hat{\sigma}_t\hat{\varepsilon}_t \\ \hat{\sigma}_t^2 &= (\underset{(0.493)}{3.259} + \underset{(0.081)}{0.406} e_{t-1}^2 + \underset{(0.066)}{0.310} e_{t-2}^2 + \underset{(0.052)}{0.149} e_{t-3}^2) L(y_{t-1}; \underset{(0.018)}{0.171}, \underset{(1.434)}{25.366}), \end{aligned} \tag{27}$$

where the numbers in parenthesis are standard errors; estimates for the parameters in the error distribution are $\hat{c} = 3.551$ (0.422) and $\hat{d} = 2.138$ (0.197).

To illustrate the conditional mean of the estimated model, consider the function $I(x) = -\nu L(x;\gamma,a)+\nu(1-L(x;\gamma,a))$ and the corresponding time-varying intercept term $I(y_{t-1})$ based on the above parameter estimates. These are shown in the middle row of Figure 1. In the left panel, the two horizontal dashed lines show the minimum and maximum $I(x)$ attains, while the three vertical dashed lines indicate the minimum of the observed data series y_t (11.71), the estimate $\hat{a} = 25.366$, and the maximum of y_t (130.61). On the right, the three horizontal dashed lines show the minimum and maximum $I(y_{t-1})$ attains (-0.187 and 0.154) and the origin. Intuitively, when y_{t-1} is close to $\hat{a}$ the time-varying intercept term $I(y_{t-1})$ is close to zero and the conditional mean of (27) corresponds to unit root type behavior (without drift); when y_{t-1} takes values clearly below/above $\hat{a}$ the intercept $I(y_{t-1})$ is positive/negative and behavior akin to a unit root process with increasing/decreasing drift occurs.

The left panel in the bottom row of Figure 1 displays the estimated volatility series $\hat{\sigma}_t$. The variation of the volatility over time is strong, and the large spikes in the volatility series coincide with the large values in the observed series. The logistic formulation of the conditional variance in (27) makes it possible for large observations to amplify volatility more than a standard linear ARCH model would allow for.

The right panel in the bottom row of Figure 1 shows the residual series $\hat{\varepsilon}_t$. Four additional graphs analyzing the residuals are available in Figure 2 in Appendix B: autocorrelation functions of the residuals and of the squared residuals, together with a histogram and a Q-Q plot. The autocorrelation functions reveal that the very strong persistence present in the original series has been quite well captured by the estimated subgeometrically ergodic AR–ARCH model (only three of the shown 100 autocorrelation coefficients are barely outside the displayed critical values). The histogram and the Q-Q plot indicate that the employed skew version of the t-distribution fits well as only a few outlying observations deviate from the estimated density function and the 45 degree line.

Note also that the estimated AR–ARCH model satisfies the requirements of a stationary and polynomially ergodic process with finite absolute moments.[4] It may be interesting to note that estimation attempts using standard linear ARMA(1,1)–GARCH(1,1) models (with skew t errors) lead to estimated autoregressive coefficients in excess of 0.999, reflecting the very persistent nature of the data series apparent from the time series and autocorrelation plots in the top row of Figure 1.

The primary purpose of this small empirical example was to demonstrate what kind of time series could be modeled with subgeometrically ergodic AR–ARCH models. It is worth pointing out that such models may work well even in cases where the graphs of the employed time series and related autocorrelation functions look very different from those displayed in Figure 1.

5 Conclusions

In this paper, we examined the subgeometric ergodicity of nonlinear autoregressive models with autoregressive conditional heteroskedasticity. We provided conditions that ensured polynomial ergodicity of the considered AR–ARCH models. Our results generalized existing results that

[4]That is, the parameter estimates in (27) correspond to a process satisfying the requirements of Proposition 1 with $s_0 = 1$. (Note that these requirements are not satisfied with $s_0 = 1.5$ which would correspond to finite second moments of y_t.)

assumed the error terms to be IID. The use of subgeometrically ergodic AR–ARCH models was illustrated in an empirical example using energy sector volatility index data.

Several future research topics could be entertained. In this paper we have only considered ARCH-type conditional heteroskedasticity, and extending the results to the generalized ARCH (GARCH) case would be of interest. Subgeometric ergodicity of multivariate autoregressions with autoregressive conditional heteroskedasticity is another interesting topic left for future work. On the empirical side, further applied work is certainly needed to judge the usefulness of subgeometrically ergodic autoregressions (with or without ARCH) in practical applications. For instance, providing more concrete advice on when to use subgeometrically (rather than geometrically) ergodic autoregressions would be useful for practitioners. Another question future applications should address is whether subgeometrically ergodic autoregressions can outperform relevant competing models in out-of-sample forecasting exercises.

Appendix A

Appendix A contains the proofs of Lemma 2, Theorem 1, and Propositions 1 and 2 as well as details for the finiteness of moments in Section 3.2.

Proof of Lemma 2. Define the vector $(\bar{\alpha}_1, \ldots, \bar{\alpha}_q) = (\alpha_1\bar{\mu}_{2s_0b}, \ldots, \alpha_q\bar{\mu}_{2s_0b})$ and let $\bar{\Lambda}$ denote the $q \times q$ matrix obtained by replacing the first row of the matrix Λ_t by $(\bar{\alpha}_1, \ldots, \bar{\alpha}_q)$. By assumption, $\bar{\alpha}_1, \ldots, \bar{\alpha}_q \geq 0$ and $\sum_{i=1}^q \bar{\alpha}_i < 1$. These conditions ensure that the polynomial $p(t) = t^q - \bar{\alpha}_1 t^{q-1} - \cdots - \bar{\alpha}_q$ has all its roots inside the unit circle (if a root t with $|t| \geq 1$ existed, the contradiction $1 = \bar{\alpha}_1/t + \cdots + \bar{\alpha}_q/t^q \leq \bar{\alpha}_1 + \cdots + \bar{\alpha}_q$ would follow); this in turn implies that the matrix $\bar{\Lambda}$ has spectral radius $\rho(\bar{\Lambda}) < 1$ (see Horn and Johnson, 2013, pp. 194–195). Therefore the matrix $I_q - \bar{\Lambda}$ is invertible with $(I_q - \bar{\Lambda})^{-1} = \sum_{i=0}^\infty \bar{\Lambda}^i$. Set $\mathbf{1}_q = (1, \ldots, 1)$ $(q \times 1)$ and let $(\boldsymbol{x})^{abs} = (|x_1|, \ldots, |x_q|)$ $(q \times 1)$ denote the elementwise absolute value of a vector $\boldsymbol{x} \in \mathbb{R}^q$. We define the vector norm $\|\cdot\|_\bullet$ on $\mathbb{R}^q$ as $\|\boldsymbol{x}\|_\bullet = \mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}(\boldsymbol{x})^{abs}$. Note that as $(I_q - \bar{\Lambda})^{-1} = \sum_{i=0}^\infty \bar{\Lambda}^i$ with $\bar{\Lambda}$ having nonnegative (and also some strictly positive) entries, $\|\boldsymbol{x}\|_\bullet = \mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}(\boldsymbol{x})^{abs} > \mathbf{1}_q'(\boldsymbol{x})^{abs} = \|\boldsymbol{x}\|_1$ whenever $\boldsymbol{x} \neq 0$ (here $\|\cdot\|_1$ denotes the usual l_1 vector norm).

Now let $\boldsymbol{x} \neq 0$ be arbitrary and consider $\|\Lambda_t\boldsymbol{x}\|_\bullet$. To this end, note that $|\alpha_1\varepsilon_t^2 x_1 + \cdots + \alpha_q\varepsilon_t^2 x_q| \leq \alpha_1\varepsilon_t^2|x_1| + \cdots + \alpha_q\varepsilon_t^2|x_q|$, which implies that the elementwise inequality $(\Lambda_t\boldsymbol{x})^{abs} \leq \Lambda_t(\boldsymbol{x})^{abs}$ holds (with probability one; note that only the first elements differ). Thus also

$$\|\Lambda_t\boldsymbol{x}\|_\bullet = \mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}(\Lambda_t\boldsymbol{x})^{abs} \leq \mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}\Lambda_t(\boldsymbol{x})^{abs}.$$

As $bs_0 \geq 1$, Minkowski's inequality and the definition of the vector $(\bar{\alpha}_1, \ldots, \bar{\alpha}_q)$ yield

$$E[\{\mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}\Lambda_t(\boldsymbol{x})^{abs}\}^{bs_0}]^{1/bs_0} \leq \mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}\bar{\Lambda}(\boldsymbol{x})^{abs},$$

where, as $(I_q - \bar{\Lambda})^{-1}\bar{\Lambda} = (I_q - \bar{\Lambda})^{-1} - I_q$ and $\|\boldsymbol{x}\|_\bullet > \|\boldsymbol{x}\|_1$,

$$\mathbf{1}_q'(I_q - \bar{\Lambda})^{-1}\bar{\Lambda}(\boldsymbol{x})^{abs} = \|\boldsymbol{x}\|_\bullet - \|\boldsymbol{x}\|_1 = \|\boldsymbol{x}\|_\bullet(1 - \|\boldsymbol{x}\|_1/\|\boldsymbol{x}\|_\bullet) < \|\boldsymbol{x}\|_\bullet.$$

These derivations establish that

$$\|\Lambda_t \boldsymbol{x}\|_{\bullet L^{bs_0}} = (E[\|\Lambda_t \boldsymbol{x}\|_\bullet^{bs_0}])^{1/bs_0} < \|\boldsymbol{x}\|_\bullet$$

and that

$$\|\!|\!|\Lambda_t|\!|\!|_{\bullet L^{bs_0}} = \max_{\|\boldsymbol{x}\|_{\bullet L^{bs_0}}=1} \|\Lambda_t \boldsymbol{x}\|_{\bullet L^{bs_0}} = \max_{\|\boldsymbol{x}\|_\bullet=1} \|\Lambda_t \boldsymbol{x}\|_{\bullet L^{bs_0}} < 1.$$

Finally, by its definition, it is clear that the vector norm $\|\cdot\|_\bullet$ is monotone. ■

Proof of Theorem 1: For clarity, we break down the long proof into several intermediate steps.

Step 1: Preliminaries. We first consider the function V defined in (22) and the conditional expectation $E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}]$. As before, we decompose an $\boldsymbol{x} \in \mathbb{R}^{p+q}$ to its p- and q-dimensional components as $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$; similarly, we decompose $\boldsymbol{y}_1$ as $\boldsymbol{y}_1 = (\boldsymbol{y}_{1,1}, \boldsymbol{y}_{2,1})$. For any $\boldsymbol{x} \in \mathbb{R}^{p+q}$, it is convenient to define

$$V_1(\boldsymbol{x}_1) = |z_1(\boldsymbol{x}_1)|^{2s_0}, \qquad V_2(\boldsymbol{x}_1) = s_1 \|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha}, \qquad \text{and} \qquad V_3(\boldsymbol{x}) = s_2 \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}$$

so that $V(\boldsymbol{x}) = 1 + V_1(\boldsymbol{x}_1) + V_2(\boldsymbol{x}_1) + V_3(\boldsymbol{x})$ (when $p = 1$, we can set $s_1 = 0$ and drop $\boldsymbol{z}_2(\boldsymbol{x}_1)$ and V_2). We next consider the three conditional expectations

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] = E[|z_1(\boldsymbol{y}_{1,1})|^{2s_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \tag{28}$$
$$E[V_2(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] = E[s_1 \|\boldsymbol{z}_2(\boldsymbol{y}_{1,1})\|_*^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \tag{29}$$
$$E[V_3(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] = E[s_2 \|\boldsymbol{\xi}(\boldsymbol{y}_1)\|_\bullet^{bs_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \tag{30}$$

related to functions V_1, V_2, and V_3. In Steps 2–4 below we establish that these conditional expectations can be bounded from above using the following upper bounds

$$E[|z_1(\boldsymbol{y}_{1,1})|^{2s_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \le |z_1(\boldsymbol{x}_1)|^{2s_0} - \tilde{r}|z_1(\boldsymbol{x}_1)|^{2s_0\alpha} + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} + C \tag{31}$$
$$E[s_1 \|\boldsymbol{z}_2(\boldsymbol{y}_{1,1})\|_*^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \le s_1 \|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha} - \tilde{\varpi} s_1^\alpha \|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha^2} + \tilde{s}_1 |z_1(\boldsymbol{x}_1)|^{2s_0\alpha} + C \tag{32}$$
$$E[s_2 \|\boldsymbol{\xi}(\boldsymbol{y}_1)\|_\bullet^{bs_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \le s_2 \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} - \tilde{\lambda} s_2 \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} + C, \tag{33}$$

where $\tilde{r}, \tilde{\varpi}, \tilde{s}_1, \tilde{\lambda} > 0$ with $\tilde{\lambda} < 1$ and where $\tilde{s}_1$ can be made as close to zero as desired by choosing a small enough s_1 (and $\tilde{s}_1 = 0$ when $p = 1$). Moreover, here and in what follows, for simplicity we use C to denote a finite positive constant whose value may change from occurence to occurence (alternatively, we could use $C_1, C_2, \ldots$). For brevity, we also often (but not always) drop the argument $\boldsymbol{x}_1$ from $z_1(\boldsymbol{x}_1)$ and $\boldsymbol{z}_2(\boldsymbol{x}_1)$ and simply write z_1 and $\boldsymbol{z}_2$.

For ease of reference, we also note here that Assumption 4 allows us to bound the conditional variance as follows. By the definition of σ_t^2 in (13), Assumption 3, and definition of $\boldsymbol{\xi}(\boldsymbol{y}_{t-1})$ in (14), $\sigma_t^2 \le \omega + e_{t-1}^2 + \cdots + e_{t-q}^2 = \omega + \|\boldsymbol{\xi}(\boldsymbol{y}_{t-1})\|_1$ (with $\|\cdot\|_1$ denoting the usual l_1 vector norm). The equivalence of vector norms on $\mathbb{R}^q$ and the fact that ω is a (finite) constant implies that (for some finite constant C)

$$\sigma_t^2 = \sigma^2(\boldsymbol{y}_{t-1}) \le C(1 + \|\boldsymbol{\xi}(\boldsymbol{y}_{t-1})\|_\bullet) \text{ a.s.} \quad \text{and} \quad \sigma^2(\boldsymbol{x}) \le C(1 + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet) \text{ for all fixed } \boldsymbol{x}. \tag{34}$$

Step 2: Upper bound for V_1. Using (12) the conditional expectation in (28) can be expressed as

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] = E[|g(z_1(\boldsymbol{x}_1)) + \sigma(\boldsymbol{x})\varepsilon_1|^{2s_0}].$$

For any positive real number Z, define the set $S_1(Z) = \{\boldsymbol{x} \in \mathbb{R}^{p+q} : |z_1(\boldsymbol{x}_1)| \leq Z\}$ and let $S_1^c(Z)$ denote the complement of this set.

First consider values of $\boldsymbol{x}$ such that $\boldsymbol{x} \in S_1^c(Z)$ so that $|z_1| = |z_1(\boldsymbol{x}_1)| > Z$. Choose Z large enough to ensure that $g(z_1) \neq 0$ (Assumption 2) so that $|g(z_1) + \sigma(\boldsymbol{x})\varepsilon_1|^{2s_0}$ can be written as

$$|g(z_1)|^{2s_0} |1 + \sigma(\boldsymbol{x})\varepsilon_1/g(z_1)|^{2s_0}. \tag{35}$$

We first bound the latter term in this expression using the following extension of Bernoulli's inequality due to Fefferman and Shapiro (1972): for any $a \geq 2$, there exists positive numbers A and B such that

$$|1 + u|^a \leq 1 + au + Au^2 + B|u|^a \tag{36}$$

for all $u \in \mathbb{R}$. Using (36), the latter term in (35) is dominated by

$$1 + 2s_0 \frac{\sigma(\boldsymbol{x})}{g(z_1)}\varepsilon_1 + A\frac{\sigma(\boldsymbol{x})^2}{g(z_1)^2}\varepsilon_1^2 + B\frac{\sigma(\boldsymbol{x})^{2s_0}}{|g(z_1)|^{2s_0}}|\varepsilon_1|^{2s_0}.$$

This upper bound, (35), the facts $E[\varepsilon_1] = 0$ and $E[\varepsilon_1^2] = 1$ (Assumption 1), and the notation $\mu_{2s_0} = E[|\varepsilon_1|^{2s_0}]$, now yield

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq |g(z_1)|^{2s_0} + A|g(z_1)|^{2s_0-2}\sigma(\boldsymbol{x})^2 + B\sigma(\boldsymbol{x})^{2s_0}\mu_{2s_0}.$$

Choose Z large enough to ensure that $0 < 1 - r|z_1|^{-\rho} < 1$ and $|g(z_1)| \leq (1 - r|z_1|^{-\rho})|z_1|$ (Assumption 2). Using the elementary inequalities $(1-u)^{a_1} \leq 1 - u$ and $(1-u)^{a_2} \leq 1$ for all $0 < u < 1$, $a_1 \geq 1$, and $a_2 \geq 0$, and recalling that $s_0 \geq 1$ and $\sigma^2(\boldsymbol{x}) \leq C\,(1 + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet)$ (see (34)), we obtain

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq |z_1|^{2s_0} - r|z_1|^{2s_0-\rho} + C|z_1|^{2s_0-2} + C|z_1|^{2s_0-2}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0}\mu_{2s_0}$$

for some positive C (by choosing Z large enough, the constant term on the dominant side has been absorbed into $|z_1|^{2s_0}$). To merge the terms $-r|z_1|^{2s_0-\rho}$ and $C|z_1|^{2s_0-2}$, by choosing Z large enough to ensure that $C/r|z_1|^{2-\rho} < 1$ we have

$$-r|z_1|^{2s_0-\rho} + C|z_1|^{2s_0-2} = -r|z_1|^{2s_0-\rho}(1 - C/r|z_1|^{2-\rho}) \leq -\hat{r}|z_1|^{2s_0-\rho}$$

for some positive constant $\hat{r}$. Hence,

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq |z_1|^{2s_0} - \hat{r}|z_1|^{2s_0-\rho} + C|z_1|^{2s_0-2}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0}\mu_{2s_0} \text{ for all } \boldsymbol{x} \in S_1^c(Z). \tag{37}$$

Now consider values of $\boldsymbol{x}$ such that $\boldsymbol{x} \in S_1(Z)$. As inequality (5) implies that $g(z_1)$ is

bounded on $S_1(Z)$, triangle inequality and the elementary inequality

$$\left|\sum_{i=i}^{m} a_i\right|^r \leq c_r \sum_{i=i}^{m} |a_i|^r \quad \text{where } c_r = 1 \text{ for } 0 < r \leq 1 \text{ and } c_r = m^{r-1} \text{ for } r > 1 \tag{38}$$

for any real numbers $a_1, \ldots, a_m$ (see, e.g., Davidson, 1994, p. 140) imply that $|g(z_1)+\sigma(\boldsymbol{x})\varepsilon_1|^{2s_0}$ is dominated by $C(1 + \sigma^{2s_0}(\boldsymbol{x})|\varepsilon_1|^{2s_0})$ (for some $C > 0$; we omit this statement from now on) for all $\boldsymbol{x} \in S_1(Z)$. As $\mu_{2s_0} = E[|\varepsilon_1|^{2s_0}]$ and $\sigma^2(\boldsymbol{x}) \leq C\,(1 + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet)$, it is seen that

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq C(1 + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0}\mu_{2s_0}) \quad \text{for all} \quad \boldsymbol{x} \in S_1(Z). \tag{39}$$

Combining (37) and (39), noting that $2s_0 - \rho = 2s_0\alpha$ (see the discussion following (22)), and merging constants, we can conclude that for all $\boldsymbol{x} \in \mathbb{R}^{p+q}$,

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq |z_1|^{2s_0} - \hat{r}|z_1|^{2s_0\alpha} + C|z_1|^{2s_0-2}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0} + C. \tag{40}$$

For future developments, it is convenient to further manipulate this upper bound. First, consider the product $|z_1|^{2s_0-2}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet$ appearing in (40) and momentarily focus on the case $s_0 > 1$ (when also $b > 1$). Using Young's inequality (with exponents $bs_0/(bs_0 - 1)$ and bs_0) yields

$$|z_1|^{2s_0-2}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet \leq \frac{bs_0 - 1}{bs_0}|z_1|^{2s_0 b(s_0-1)/(bs_0-1)} + \frac{1}{bs_0}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}.$$

Simple calculations show that the assumption $b > (2s_0 - \rho)/[s_0(2 - \rho)]$ implies that $2s_0 b(s_0 - 1)/(bs_0 - 1) < 2s_0\alpha$. Therefore for some small positive ς

$$|z_1|^{2s_0-2}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet \leq C(1 + |z_1|^{2s_0\alpha-\varsigma} + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}); \tag{41}$$

clearly this upper bound also holds in the case $s_0 = 1$.

Second, consider the terms involving $|z_1|^{2s_0\alpha}$ in (40) and $|z_1|^{2s_0\alpha-\varsigma}$ in (41). By considering values of $|z_1|$ larger and smaller than some large bound, it is straightforward to see that

$$C|z_1|^{2s_0\alpha-\varsigma} - \hat{r}|z_1|^{2s_0\alpha} = -\hat{r}(1 - C/[\hat{r}|z_1|^{\varsigma}])|z_1|^{2s_0\alpha} \leq C - \tilde{r}|z_1|^{2s_0\alpha}$$

for some positive constant $\tilde{r}$. Third, the term $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0}$ appearing in (40) is clearly dominated by a term of the form $C + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}$.

Inequality (40) together with these additional manipulations leads to the final upper bound

$$E[V_1(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq |z_1|^{2s_0} - \tilde{r}|z_1|^{2s_0\alpha} + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} + C \tag{42}$$

which holds for all $\boldsymbol{x} \in \mathbb{R}^{p+q}$.

Step 3: Upper bound for V_2. Using (12), we can express the conditional expectation in (29) as

$$E[V_2(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] = s_1\|\boldsymbol{\Pi}_1\boldsymbol{z}_2(\boldsymbol{x}_1) + z_1(\boldsymbol{x}_1)\boldsymbol{\iota}_{p-1}\|_*^{2s_0\alpha}.$$

Recall that $\alpha = 1 - \rho/2s_0 \in (0,1)$ (because $s_0 \geq 1$ and $\rho \in (0,2)$ by assumption) and that $|||\boldsymbol{\Pi}_1|||_* \leq \varpi$ for some $0 < \varpi < 1$ (by Lemma 1). These facts together with elementary inequalities (and dropping the argument $\boldsymbol{x}_1$ from $z_1(\boldsymbol{x}_1)$ and $\boldsymbol{z}_2(\boldsymbol{x}_1)$) imply that

$$\|\boldsymbol{\Pi}_1 \boldsymbol{z}_2 + z_1 \boldsymbol{\iota}_{p-1}\|_*^\alpha \leq \|\boldsymbol{\Pi}_1 \boldsymbol{z}_2\|_*^\alpha + \|z_1 \boldsymbol{\iota}_{p-1}\|_*^\alpha \leq \varpi^\alpha \|\boldsymbol{z}_2\|_*^\alpha + \|\boldsymbol{\iota}_{p-1}\|_*^\alpha |z_1|^\alpha .$$

This, together with the convexity of the function $|x| \mapsto |x|^{2s_0}$ (recall that $s_0 \geq 1$ by assumption), imply that for any $\tau_1 \in (0,1)$ and $\tau_2 = 1 - \tau_1$,

$$\begin{aligned} s_1 \|\boldsymbol{\Pi}_1 \boldsymbol{z}_2 + z_1 \boldsymbol{\iota}_{p-1}\|_*^{2s_0\alpha} &\leq \left(\tau_2 \frac{s_1^{1/2s_0} \varpi^\alpha}{\tau_2} \|\boldsymbol{z}_2\|_*^\alpha + \tau_1 \frac{s_1^{1/2s_0} \|\boldsymbol{\iota}_{p-1}\|_*^\alpha}{\tau_1} |z_1|^\alpha \right)^{2s_0} \\ &\leq \tau_2 \frac{s_1 \varpi^{2s_0\alpha}}{\tau_2^{2s_0}} \|\boldsymbol{z}_2\|_*^{2s_0\alpha} + \tau_1 \frac{s_1 \|\boldsymbol{\iota}_{p-1}\|_*^{2s_0\alpha}}{\tau_1^{2s_0}} |z_1|^{2s_0\alpha} . \end{aligned} \tag{43}$$

Consider the former term on the dominant side of (43). Fix a τ_2 such that $\tau_2 \in (\varpi^\alpha, 1)$ and set $\tilde{\varpi} = 1 - (\varpi^\alpha/\tau_2)^{2s_0} \in (0,1)$. Then the former term on the dominant side of (43) satisfies

$$\tau_2 \frac{s_1 \varpi^{2s_0\alpha}}{\tau_2^{2s_0}} \|\boldsymbol{z}_2\|_*^{2s_0\alpha} = \tau_2 s_1 (1 - \tilde{\varpi}) \|\boldsymbol{z}_2\|_*^{2s_0\alpha} < s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} - \tilde{\varpi} s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} . \tag{44}$$

Suppose now that s_1 is any fixed (but potentially arbitrarily small) positive number. If $\|\boldsymbol{z}_2\|_*$ is large enough to ensure that $s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} \geq 1$, then $s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} \geq s_1^\alpha \|\boldsymbol{z}_2\|_*^{2s_0\alpha^2}$ as $\alpha \in (0,1)$ and the right side of (44) is dominated by $s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} - \tilde{\varpi} s_1^\alpha \|\boldsymbol{z}_2\|_*^{2s_0\alpha^2}$. On the other hand, if $s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} < 1$ the right side of (44) is bounded by a constant. Therefore

$$\tau_2 \frac{s_1 \varpi^{2s_0\alpha}}{\tau_2^{2s_0}} \|\boldsymbol{z}_2\|_*^{2s_0\alpha} < s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} - \tilde{\varpi} s_1^\alpha \|\boldsymbol{z}_2\|_*^{2s_0\alpha^2} + C.$$

Now consider the latter term on the dominant side of (43). Choosing a small enough fixed s_1, this term can be made smaller than $\tilde{s}_1 |z_1|^{2s_0\alpha}$ where $\tilde{s}_1$ can be chosen as close to zero as desired. To summarize, it holds that

$$E[V_2(\boldsymbol{y}_{1,1}) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq s_1 \|\boldsymbol{z}_2\|_*^{2s_0\alpha} - \tilde{\varpi} s_1^\alpha \|\boldsymbol{z}_2\|_*^{2s_0\alpha^2} + \tilde{s}_1 |z_1|^{2s_0\alpha} + C \tag{45}$$

where $\tilde{\varpi} \in (0,1)$ and the value of $\tilde{s}_1 > 0$ can be chosen as close to zero as desired.

Step 4: Upper bound for V_3. By the definition of the function $\boldsymbol{\xi}$ in (14), $\boldsymbol{\xi}(\boldsymbol{y}_1) = \Lambda_{\zeta,1}(\boldsymbol{y}_0)\boldsymbol{\xi}(\boldsymbol{y}_0) + \boldsymbol{\omega}_{\zeta,1}$. We start by bounding both terms on the right hand side of this equality and, for simplicity, remove the argument $\boldsymbol{y}_0$ and instead use the notations $\Lambda_{\zeta,1}$ and $\boldsymbol{\xi}_0$.

First denote $v_{\zeta,1} = \varepsilon_1^2(\alpha_1 \zeta_{1,0} e_0^2 + \cdots + \alpha_q \zeta_{q,0} e_{1-q}^2)$ and $v_1 = \varepsilon_1^2(\alpha_1 e_0^2 + \cdots + \alpha_q e_{1-q}^2)$. Using the definitions of the matrices $\Lambda_{\zeta,1}$ and Λ_1 and the vector $\boldsymbol{\xi}_0$ (see (14) and (16)) we then have

$$\Lambda_{\zeta,1}\boldsymbol{\xi}_0 = (v_{\zeta,1}, e_0^2, \ldots, e_{1-q}^2) \quad \text{and} \quad \Lambda_1 \boldsymbol{\xi}_0 = (v_1, e_0^2, \ldots, e_{1-q}^2),$$

where all components of both vectors are nonnegative. As $\zeta_{i,0} \in (0,1]$ for all $i = 1, \ldots, q$ by assumption, we have $v_{\zeta,1} \leq v_1$ (a.s.). The monotonicity of the norm $\|\cdot\|_\bullet$ required in Assumption

4 now implies that $\|\Lambda_{\zeta,1}\boldsymbol{\xi}_0\|_\bullet \le \|\Lambda_1\boldsymbol{\xi}_0\|_\bullet$ (a.s.) (see the discussion preceding Assumption 4). Regarding the vector $\boldsymbol{\omega}_{\zeta,1}$, its first component is $\zeta_{0,0}(\boldsymbol{y}_0)\varepsilon_1^2\omega \le \varepsilon_1^2\omega$ (a.s.) and the other components are zero, so that the monotonicity of the norm $\|\cdot\|_\bullet$ shows that $\|\boldsymbol{\omega}_{\zeta,1}\|_\bullet \le \|\boldsymbol{\omega}_1\|_\bullet = \varepsilon_1^2\|\boldsymbol{\omega}\|_\bullet$ (a.s.) where $\boldsymbol{\omega} = (\omega, 0, \ldots, 0)$.

The preceding discussion together with the triangle inequality now yields, with probability one,

$$\|\boldsymbol{\xi}(\boldsymbol{y}_1)\|_\bullet = \|\Lambda_{\zeta,1}(\boldsymbol{y}_0)\boldsymbol{\xi}(\boldsymbol{y}_0) + \boldsymbol{\omega}_{\zeta,1}\|_\bullet \le \|\Lambda_{\zeta,1}(\boldsymbol{y}_0)\boldsymbol{\xi}(\boldsymbol{y}_0)\|_\bullet + \|\boldsymbol{\omega}_{\zeta,1}\|_\bullet \le \|\Lambda_1\boldsymbol{\xi}(\boldsymbol{y}_0)\|_\bullet + \varepsilon_1^2\|\boldsymbol{\omega}\|_\bullet.$$

Using the notation $\bar{\mu}_{2bs_0} = (E[|\varepsilon_0|^{2bs_0}])^{1/bs_0}$ and Minkowski's inequality we find that

$$\left(E[\|\boldsymbol{\xi}(\boldsymbol{y}_1)\|_\bullet^{bs_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}]\right)^{1/bs_0} \le \left(E[\|\Lambda_1\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}]\right)^{1/bs_0} + \bar{\mu}_{2bs_0}\|\boldsymbol{\omega}\|_\bullet.$$

By inequality (18) and Assumption 4, the first term on the dominant side satisfies

$$\left(E[\|\Lambda_1\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}]\right)^{1/bs_0} = \|\Lambda_1\boldsymbol{\xi}(\boldsymbol{x})\|_{\bullet L^{bs_0}} \le |||\Lambda_1|||_{\bullet L^{bs_0}}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet = \lambda\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet$$

with $\lambda < 1$. The preceding steps imply that

$$\left(E[\|\boldsymbol{\xi}(\boldsymbol{y}_1)\|_\bullet^{bs_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}]\right)^{1/bs_0} \le \lambda\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet + \bar{\mu}_{2bs_0}\|\boldsymbol{\omega}\|_\bullet = \lambda\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet\left(1 + \frac{\bar{\mu}_{2bs_0}\|\boldsymbol{\omega}\|_\bullet}{\lambda\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet}\right)$$

and, using (30) and the notation $\hat{\lambda} = \lambda^{bs_0} < 1$, we obtain

$$\begin{aligned} E[V_3(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] &\le s_2\hat{\lambda}\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}\left(1 + \frac{\bar{\mu}_{2bs_0}\|\boldsymbol{\omega}\|_\bullet}{\lambda\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet}\right)^{bs_0} \\ &= s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} - \left\{1 - \hat{\lambda}\left(1 + \frac{\bar{\mu}_{2bs_0}\|\boldsymbol{\omega}\|_\bullet}{\lambda\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet}\right)^{bs_0}\right\} s_2\left\|\boldsymbol{\xi}(\boldsymbol{x})\right\|_\bullet^{bs_0}. \end{aligned}$$

As $\hat{\lambda} \in (0,1)$, we can choose a $\xi > 0$ such that the term in curly brackets is larger than some $\tilde{\lambda} \in (0,1)$ whenever $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet > \xi$. Therefore, whenever $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet > \xi$, we have

$$E[V_3(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \le s_2\left\|\boldsymbol{\xi}(\boldsymbol{x})\right\|_\bullet^{bs_0} - \tilde{\lambda}s_2\left\|\boldsymbol{\xi}(\boldsymbol{x})\right\|_\bullet^{bs_0}.$$

On the other hand, whenever $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet \le \xi$ the previous derivations also make it clear that $E[V_3(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}]$ is bounded by some constant. Therefore for all $\boldsymbol{x} \in \mathbb{R}^{p+q}$,

$$E[V_3(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \le s_2\left\|\boldsymbol{\xi}(\boldsymbol{x})\right\|_\bullet^{bs_0} - \tilde{\lambda}s_2\left\|\boldsymbol{\xi}(\boldsymbol{x})\right\|_\bullet^{bs_0} + C. \tag{46}$$

Step 5: Upper bound for V. We next combine the upper bounds (42), (45), and (46) derived in Steps 2–4 to obtain

$$\begin{aligned} E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \le 1 &+ |z_1|^{2s_0} - \tilde{r}|z_1|^{2s_0\alpha} + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} \\ &+ s_1\left\|\boldsymbol{z}_2\right\|_*^{2s_0\alpha} - \tilde{\varpi}s_1^\alpha\left\|\boldsymbol{z}_2\right\|_*^{2s_0\alpha^2} + \tilde{s}_1\left|z_1\right|^{2s_0\alpha} \\ &+ s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} - \tilde{\lambda}s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} + C. \end{aligned} \tag{47}$$

To combine the terms involving $|z_1|^{2s_0\alpha}$, set $\bar{r} = \tilde{r} - \tilde{s}_1$ so that

$$-\tilde{r}|z_1|^{2s_0\alpha} + \tilde{s}_1|z_1|^{2s_0\alpha} = -\bar{r}|z_1|^{2s_0\alpha};$$

recalling that $\tilde{s}_1$ can be chosen as close to zero as desired, we have $\bar{r} > 0$ by a suitable choice of $\tilde{s}_1$. On the other hand, to manipulate the terms involving $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}$, choose s_2 large enough to ensure that $\tilde{\lambda} - C/s_2 \in (0,1)$ and set $\bar{\lambda} = \tilde{\lambda} - C/s_2$. As now $-\bar{\lambda}s_2 = C - \tilde{\lambda}s_2$, the terms involving $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}$ in (47) can be written as

$$s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} + C\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} - \tilde{\lambda}s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} = s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} - \bar{\lambda}s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}. \tag{48}$$

Whenever $s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} < 1$, (48) is bounded by the constant 1; for $s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} \geq 1$, we have $s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} \geq s_2^\alpha\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}$ as $\alpha \in (0,1)$. Thus the expression in (48) is always bounded by $1 + s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0} - \bar{\lambda}s_2^\alpha\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}$. As $V(\boldsymbol{x}) = 1 + |z_1|^{2s_0} + s_1\|\boldsymbol{z}_2\|_*^{2s_0\alpha} + s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0}$, we obtain from (47) that

$$E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq V(\boldsymbol{x}) - (1 + \bar{r}|z_1|^{2s_0\alpha} + \tilde{\varpi}s_1^\alpha\|\boldsymbol{z}_2\|_*^{2s_0\alpha^2} + \bar{\lambda}s_2^\alpha\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}) + C. \tag{49}$$

The inequality in (38) implies that

$$[V(\boldsymbol{x})]^\alpha = (1 + |z_1|^{2s_0} + s_1\|\boldsymbol{z}_2\|_*^{2s_0\alpha} + s_2\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0})^\alpha \leq 1 + |z_1|^{2s_0\alpha} + s_1^\alpha\|\boldsymbol{z}_2\|_*^{2s_0\alpha^2} + s_2^\alpha\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}$$

so that, setting $e = \min\{\bar{r}, \tilde{\varpi}, \bar{\lambda}\} \in (0,1)$, we have

$$e[V(\boldsymbol{x})]^\alpha \leq 1 + \bar{r}|z_1|^{2s_0\alpha} + \tilde{\varpi}s_1^\alpha\|\boldsymbol{z}_2\|_*^{2s_0\alpha^2} + \bar{\lambda}s_2^\alpha\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}.$$

Therefore, setting $\tilde{e} = e/2$,

$$E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq V(\boldsymbol{x}) - \tilde{e}[V(\boldsymbol{x})]^\alpha + \{C - \tilde{e}[V(\boldsymbol{x})]^\alpha\}.$$

Now, define the set

$$A_N = \big\{\boldsymbol{x} \in \mathbb{R}^{p+q} : |z_1(\boldsymbol{x}_1)|^{2s_0} \leq N, \quad \|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha} \leq N, \quad \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha} \leq N\big\}, \tag{50}$$

where N is so large that A_N is nonempty (see (4)). The complement of A_N is denoted by A_N^c so that $\boldsymbol{x} \in A_N^c$ if either $|z_1(\boldsymbol{x}_1)|^{2s_0} > N$, $\|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha} > N$ or $\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha} > N$. By choosing a large enough N, for all $\boldsymbol{x} \in A_N^c$ it holds that $C - \tilde{e}[V(\boldsymbol{x})]^\alpha < 0$ so that $E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq V(\boldsymbol{x}) - \tilde{e}[V(\boldsymbol{x})]^\alpha$ for all $\boldsymbol{x} \in A_N^c$. On the other hand, the function $V(\boldsymbol{x}) - e[V(\boldsymbol{x})]^\alpha + C$ is clearly bounded by some positive constant $\tilde{b}$ on A_N. Therefore we can conclude that there exists an N and a positive constant $\tilde{b}$ such that

$$E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq V(\boldsymbol{x}) - \phi_1(V(\boldsymbol{x})) + \tilde{b}\mathbf{1}_{A_N}(\boldsymbol{x}), \tag{51}$$

where $\phi_1(v) = \tilde{e}v^\alpha$. This implies that Condition D holds with $\phi = \phi_1$ and $C = A_N$.

Step 6: Showing that A_N is petite. We first note that the definition of a petite set and other Markov chain concepts we refer to below can be found in Meyn and Tweedie (2009, Chs

4–6). The idea is to establish that the (potentially non-compact) set A_N in (50) is petite for any $N \geq 1$ so large that A_N is nonempty. To this end, we show below that there exists an $M_N < \infty$ such that

$$\sup_{\boldsymbol{x} \in A_N} E[\|\boldsymbol{y}_{p+q}\|_1^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}] < M_N^{2s_0\alpha}, \tag{52}$$

where $\|\cdot\|_1$ denotes the usual l_1 vector norm. Next note that Theorem 2.2(ii) of Cline and Pu (1998) along with our Assumption 1 shows that the Markov chain $\boldsymbol{y}_t$ is a ψ-irreducible and aperiodic T-chain (see also Example 2.1 of the aforementioned paper). Therefore the compact set $B_N = \{\boldsymbol{x} \in \mathbb{R}^{p+q} : \|\boldsymbol{x}\|_1 \leq M_N\}$ is small (see Meyn and Tweedie, 2009, Thms 6.2.5(ii) and 5.5.7). Moreover, due to Markov's inequality,

$$\begin{aligned}\inf_{\boldsymbol{x} \in A_N} \Pr[\boldsymbol{y}_{p+q} \in B_N \mid \boldsymbol{y}_0 = \boldsymbol{x}] &= 1 - \sup_{\boldsymbol{x} \in A_N} \Pr[\|\boldsymbol{y}_{p+q}\|_1 \geq M_N \mid \boldsymbol{y}_0 = \boldsymbol{x}] \\ &\geq 1 - \sup_{\boldsymbol{x} \in A_N} E[\|\boldsymbol{y}_{p+q}\|_1^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}]/M_N^{2s_0\alpha},\end{aligned}$$

where the last expression is positive due to (52). Proposition 5.2.4(i) of Meyn and Tweedie (2009) now implies that the set A_N is small. Proposition 5.5.3 of the same reference therefore implies that the set A_N is also petite.

To complete the proof of petiteness of A_N, it remains to establish (52). First we introduce some notation. We let $|||\cdot|||_1$ denote the maximum column sum norm defined for real square matrices (this norm is induced by the l_1 vector norm, see Horn and Johnson, 2013, Sec 5.6). For brevity, we denote $\boldsymbol{z}_t = \boldsymbol{z}(\boldsymbol{y}_{1,t}) = \mathbf{A}\boldsymbol{y}_{1,t}$ and also partition the p-dimensional $\boldsymbol{z}_t$ as $\boldsymbol{z}_t = (z_{1,t}, \boldsymbol{z}_{2,t})$ (see (11)). This allows us to write the companion form (12) as

$$\boldsymbol{z}_t = \begin{bmatrix} z_{1,t} \\ \boldsymbol{z}_{2,t} \end{bmatrix} = \begin{bmatrix} g(z_{1,t-1}) + \sigma_t \varepsilon_t \\ \boldsymbol{\Pi}_1 \boldsymbol{z}_{2,t-1} + z_{1,t-1} \boldsymbol{\iota}_{p-1} \end{bmatrix}. \tag{53}$$

Finally, we set $\vec{\boldsymbol{z}}_{1,p+q} = (z_{1,p+q}, \ldots, z_{1,1})$ and $\vec{\boldsymbol{z}}_{2,p+q} = (\boldsymbol{z}_{2,p+q}, \ldots, \boldsymbol{z}_{2,1})$.

Now consider the norm $\|\boldsymbol{y}_{p+q}\|_1$ in (52). Using properties of the norms $\|\cdot\|_1$ and $|||\cdot|||_1$ we can write

$$\|\boldsymbol{y}_{p+q}\|_1 \leq \sum_{t=1}^{p+q} \|\boldsymbol{y}_{1,t}\|_1 = \sum_{t=1}^{p+q} \|\boldsymbol{A}^{-1}\boldsymbol{z}_t\|_1 \leq |||\boldsymbol{A}^{-1}|||_1 \sum_{t=1}^{p+q} \|\boldsymbol{z}_t\|_1 = |||\boldsymbol{A}^{-1}|||_1 (\|\vec{\boldsymbol{z}}_{1,p+q}\|_1 + \|\vec{\boldsymbol{z}}_{2,p+q}\|_1)$$

implying that $\|\boldsymbol{y}_{p+q}\|_1 \leq C(\|\vec{\boldsymbol{z}}_{1,p+q}\|_1 + \|\vec{\boldsymbol{z}}_{2,p+q}\|_1)$. Adding terms and making use of inequality (38) and the fact that $\alpha \in (0,1)$ we obtain

$$\begin{aligned}\|\boldsymbol{y}_{p+q}\|_1^{2s_0\alpha} &\leq C((1 + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1)^{2s_0\alpha} + \|\vec{\boldsymbol{z}}_{2,p+q}\|_1^{2s_0\alpha}) \\ &\leq C((1 + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1)^{2s_0} + \|\vec{\boldsymbol{z}}_{2,p+q}\|_1^{2s_0\alpha}) \\ &\leq C(1 + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1^{2s_0} + \|\vec{\boldsymbol{z}}_{2,p+q}\|_1^{2s_0\alpha}).\end{aligned} \tag{54}$$

To obtain an upper bound for the term $\|\vec{\boldsymbol{z}}_{1,p+q}\|_1^{2s_0}$, consider the equality $z_{1,t} = g(z_{1,t-1}) + \sigma_t\varepsilon_t$ from (53) and note that, by Assumption 2, $|g(u)| \leq K_0$ for $|u| \leq M_0$ and $|g(u)| \leq (1 - r\,|u|^{-\rho})|u| \leq |u|$ for $|u| \geq M_0$, so that $|g(u)| \leq K_0 + |u|$ for all $u \in \mathbb{R}$ (note that Assumption 2 requires M_0 to be so large that $r\,|u|^{-\rho} \in (0,1)$ for $|u| \geq M_0$). Using these inequalities,

$|z_{1,t}| \leq K_0 + |z_{1,t-1}| + \sigma_t|\varepsilon_t|$ $(t = 1, \ldots, p+q)$, so that

$$\begin{aligned}
|z_{1,1}| &\leq K_0 + |z_{1,0}| + \sigma_1|\varepsilon_1|,\\
|z_{1,2}| &\leq 2K_0 + |z_{1,0}| + \sigma_1|\varepsilon_1| + \sigma_2|\varepsilon_2|,\\
&\vdots\\
|z_{1,p+q}| &\leq (p+q)K_0 + |z_{1,0}| + \sigma_1|\varepsilon_1| + \cdots + \sigma_{p+q}|\varepsilon_{p+q}|.
\end{aligned}$$

Thus $\|\vec{\boldsymbol{z}}_{1,p+q}\|_1 \leq C\left(1 + |z_{1,0}| + \sum_{i=1}^{p+q}\sigma_i|\varepsilon_i|\right)$ and, making use of inequality (38),

$$\|\vec{\boldsymbol{z}}_{1,p+q}\|_1^{2s_0} \leq C\left(1 + |z_{1,0}|^{2s_0} + \sum_{i=1}^{p+q}\sigma_i^{2s_0}|\varepsilon_i|^{2s_0}\right). \tag{55}$$

Next, to bound the term $\|\vec{\boldsymbol{z}}_{2,p+q}\|_1^{2s_0\alpha}$, consider $\boldsymbol{z}_{2,t} = \boldsymbol{\Pi}_1\boldsymbol{z}_{2,t-1} + z_{1,t-1}\boldsymbol{\iota}_{p-1}$ (see (53)). Setting $\kappa = |||\boldsymbol{\Pi}_1|||_1$ and using the fact $\|\boldsymbol{\iota}_{p-1}\|_1 = 1$ we obtain

$$\|\boldsymbol{z}_{2,t}\|_1 \leq |||\boldsymbol{\Pi}_1|||_1\|\boldsymbol{z}_{2,t-1}\|_1 + |z_{1,t-1}|\|\boldsymbol{\iota}_{p-1}\|_1 = \kappa\|\boldsymbol{z}_{2,t-1}\|_1 + |z_{1,t-1}|$$

and furthermore

$$\begin{aligned}
\|\boldsymbol{z}_{2,1}\|_1 &\leq \kappa\|\boldsymbol{z}_{2,0}\|_1 + |z_{1,0}|,\\
\|\boldsymbol{z}_{2,2}\|_1 &\leq \kappa^2\|\boldsymbol{z}_{2,0}\|_1 + \kappa|z_{1,0}| + |z_{1,1}|,\\
&\vdots\\
\|\boldsymbol{z}_{2,p+q}\|_1 &\leq \kappa^{p+q}\|\boldsymbol{z}_{2,0}\|_1 + \kappa^{p+q-1}|z_{1,0}| + \cdots + |z_{1,p+q-1}|.
\end{aligned}$$

This implies that

$$\|\vec{\boldsymbol{z}}_{2,p+q}\|_1 \leq C(\|\boldsymbol{z}_{2,0}\|_1 + 1 + |z_{1,0}| + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1).$$

As the norms $\|\cdot\|_1$ and $\|\cdot\|_*$ are equivalent, it holds that $\|\boldsymbol{z}_{2,0}\|_1 \leq C\|\boldsymbol{z}_{2,0}\|_*$. Making use of inequality (38) and the fact that $\alpha \in (0,1)$ we obtain

$$\begin{aligned}
\|\vec{\boldsymbol{z}}_{2,p+q}\|_1^{2s_0\alpha} &\leq C(\|\boldsymbol{z}_{2,0}\|_*^{2s_0\alpha} + (1 + |z_{1,0}| + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1)^{2s_0\alpha})\\
&\leq C(\|\boldsymbol{z}_{2,0}\|_*^{2s_0\alpha} + (1 + |z_{1,0}| + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1)^{2s_0})\\
&\leq C(1 + \|\boldsymbol{z}_{2,0}\|_*^{2s_0\alpha} + |z_{1,0}|^{2s_0} + \|\vec{\boldsymbol{z}}_{1,p+q}\|_1^{2s_0}).
\end{aligned} \tag{56}$$

Now combine (54) with the upper bounds obtained for $\|\vec{\boldsymbol{z}}_{1,p+q}\|_1^{2s_0}$ and $\|\vec{\boldsymbol{z}}_{2,p+q}\|_1^{2s_0\alpha}$ in (55) and (56), and recall that $z_{1,0} = z_1(\boldsymbol{y}_{1,0})$ and $\boldsymbol{z}_{2,0} = \boldsymbol{z}_2(\boldsymbol{y}_{1,0})$, to obtain

$$\|\boldsymbol{y}_{p+q}\|_1^{2s_0\alpha} \leq C\left(1 + \|\boldsymbol{z}_2(\boldsymbol{y}_{1,0})\|_*^{2s_0\alpha} + |z_1(\boldsymbol{y}_{1,0})|^{2s_0} + \sum_{i=1}^{p+q}\sigma_i^{2s_0}|\varepsilon_i|^{2s_0}\right).$$

As $\mu_{2s_0} = E[|\varepsilon_1|^{2s_0}]$ is finite, this implies that

$$E[\|\boldsymbol{y}_{p+q}\|_1^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq C\left(1 + \|\boldsymbol{z}_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha} + |z_1(\boldsymbol{x}_1)|^{2s_0} + \mu_{2s_0}\sum_{i=1}^{p+q}E[\sigma_i^{2s_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}]\right). \tag{57}$$

Next consider the terms in (57) involving conditional expectations of the $\sigma_i^{2s_0}$'s. We first derive an inequality which is similar to inequality (11) in Meitz and Saikkonen (2010). Using repeated substitution and the equality $\boldsymbol{\xi}(\boldsymbol{y}_t) = \Lambda_{\zeta,t}\boldsymbol{\xi}(\boldsymbol{y}_{t-1}) + \boldsymbol{\omega}_{\zeta,t}$ we obtain, for any fixed $t \geq 1$, that

$$\boldsymbol{\xi}(\boldsymbol{y}_t) = \prod_{k=0}^{t-1} \Lambda_{\zeta,t-k}\boldsymbol{\xi}(\boldsymbol{y}_0) + \boldsymbol{\omega}_{\zeta,t} + \sum_{k=0}^{t-2}\prod_{l=0}^{k} \Lambda_{\zeta,t-l}\boldsymbol{\omega}_{\zeta,t-k-1}.$$

Now consider the vector norm $\|\cdot\|_\bullet$ in Assumption 4. Denote by $|||\cdot|||_\bullet$ the matrix norm induced by the vector norm $\|\cdot\|_\bullet$; that is, for any $q \times q$ matrix A, set

$$|||A|||_\bullet = \max_{\|\boldsymbol{x}\|_\bullet=1} \|A\boldsymbol{x}\|_\bullet \qquad (\boldsymbol{x} \in \mathbb{R}^q).$$

(For clarity, note that $|||\cdot|||_\bullet$ above and $|||\cdot|||_{\bullet L^p}$ defined in (17) coincide for nonrandom matrices but differ for random ones.) As $\|\cdot\|_\bullet$ in Assumption 4 is assumed to be monotone, it follows from Problems 5.6.P41(c) and 5.6.P42 in Horn and Johnson (2013, p. 368) that the induced matrix norm $|||\cdot|||_\bullet$ is monotone on the positive orthant, meaning that any $q \times q$ matrices A and B that satisfy the (entrywise) inequalities $0 \leq A \leq B$ also satisfy the inequality $|||A|||_\bullet \leq |||B|||_\bullet$. Usual properties of vector norms and matrix norms in conjunction with inequality (38) therefore yield

$$\|\boldsymbol{\xi}(\boldsymbol{y}_t)\|_\bullet^{s_0} \leq C\prod_{k=0}^{t-1} |||\Lambda_{\zeta,t-k}|||_\bullet^{s_0}\|\boldsymbol{\xi}(\boldsymbol{y}_0)\|_\bullet^{s_0} + C\|\boldsymbol{\omega}_{\zeta,t}\|_\bullet^{s_0} + C\sum_{k=0}^{t-2}\prod_{l=0}^{k} |||\Lambda_{\zeta,t-l}|||_\bullet^{s_0}\|\boldsymbol{\omega}_{\zeta,t-k-1}\|_\bullet^{s_0}.$$

By the monotonicity properties of the norms $\|\cdot\|_\bullet$ and $|||\cdot|||_\bullet$ and the definitions of the matrices $\Lambda_{\zeta,t}$ and Λ_t in (14) and (16) we also obtain $|||\Lambda_{\zeta,t}|||_\bullet \leq |||\Lambda_t|||_\bullet$ and $\|\boldsymbol{\omega}_{\zeta,t}\|_\bullet \leq \|\boldsymbol{\omega}_t\|_\bullet = \varepsilon_t^2\|\boldsymbol{\omega}\|_\bullet$ (a.s.) for all $t = 1, 2, \ldots$, implying that

$$\|\boldsymbol{\xi}(\boldsymbol{y}_t)\|_\bullet^{s_0} \leq C\prod_{k=0}^{t-1} |||\Lambda_{t-k}|||_\bullet^{s_0}\|\boldsymbol{\xi}(\boldsymbol{y}_0)\|_\bullet^{s_0} + C\left(|\varepsilon_t|^{2s_0} + \sum_{k=0}^{t-2}\prod_{l=0}^{k} |||\Lambda_{t-l}|||_\bullet^{s_0}|\varepsilon_{t-k-1}|^{2s_0}\right)\|\boldsymbol{\omega}\|_\bullet^{s_0}.$$

Now, denote the expectation $E\left[|||\Lambda_t|||_\bullet^{s_0}\right]$ by χ (this expectation is finite due to Assumption 4). Using the independence of the Λ_t's and independence of $|||\Lambda_{t-l}|||_\bullet$'s and ε_{t-k-1}'s, yields

$$E\left[\|\boldsymbol{\xi}(\boldsymbol{y}_t)\|_\bullet^{s_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}\right] \leq C\chi^t\|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0} + C\left(1 + \sum_{k=0}^{t-2}\chi^{k+1}\right)\|\boldsymbol{\omega}\|_\bullet^{s_0}. \tag{58}$$

Inequality (34) in conjunction with (38) show that $\sigma_i^{2s_0} \leq C(1 + \|\boldsymbol{\xi}(\boldsymbol{y}_{i-1})\|_\bullet^{s_0})$ (a.s.) for all $i = 1, \ldots, p+q$. From (58) it then follows that $E[\sigma_i^{2s_0} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq C(1 + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0})$ which together with (57) implies

$$E[\|\boldsymbol{y}_{p+q}\|_1^{2s_0\alpha} \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq C\left(1 + |z_1(\boldsymbol{x}_1)|^{2s_0} + \|z_2(\boldsymbol{x}_1)\|_*^{2s_0\alpha} + \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{s_0}\right).$$

For any $\boldsymbol{x} \in A_N$, the dominant side is bounded by $C(1 + 2N + N^{1/b\alpha})$ and thus we can find a finite M_N such that (52) holds.

Step 7: Completing the proof. We are now ready to complete the proof by applying Theorem 1(iii) in Meitz and Saikkonen (2022). To this end, in the beginning of Step 6 we already noted that the Markov chain $\boldsymbol{y}_t$ is ψ-irreducible and aperiodic. That Condition D holds was shown in (51) in Step 5. Petiteness of the set A_N was shown in Step 6. We also need to verify that $\sup_{\boldsymbol{x}\in A_N} V(\boldsymbol{x}) < \infty$; this inequality is a straightforward consequence of the definitions of the set A_N and the function V. Thus, applying Theorem 1(iii) in Meitz and Saikkonen (2022) we can complete the proof. ■

Details for the finiteness of moments in Section 3.2. The arguments are similar to those used in the proof of Corollary to Theorem 3 in Meitz and Saikkonen (2022). First note that inequality (45) continues to hold if the term $\tilde{\varpi} s_1^{\alpha} \left\| \boldsymbol{z}_2 \right\|_*^{2s_0\alpha^2}$ on its dominant side is replaced with the term $\tilde{\varpi} s_1 \left\| \boldsymbol{z}_2 \right\|_*^{2s_0\alpha}$ (this can be seen from (44) and the arguments that follow it). Consequently, the same replacement can be done on the dominant sides of inequalities (47) and (49), the latter inequality thus becoming

$$E[V(\boldsymbol{y}_1) \mid \boldsymbol{y}_0 = \boldsymbol{x}] \leq V(\boldsymbol{x}) - (1 + \bar{r}|z_1(\boldsymbol{x}_1)|^{2s_0\alpha} + \tilde{\varpi} s_1 \left\| \boldsymbol{z}_2(\boldsymbol{x}_1) \right\|_*^{2s_0\alpha} + \bar{\lambda} s_2^{\alpha} \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}) + C.$$

Finiteness of certain moments with respect to the stationary distribution π of $\boldsymbol{y}_t$ can now be obtained from Theorem 14.3.7 of Meyn and Tweedie (2009), namely $\int_{\mathbb{R}^{p+q}} (1 + \bar{r}|z_1(\boldsymbol{x}_1)|^{2s_0\alpha} + \tilde{\varpi} s_1 \left\| \boldsymbol{z}_2(\boldsymbol{x}_1) \right\|_*^{2s_0\alpha} + \bar{\lambda} s_2^{\alpha} \|\boldsymbol{\xi}(\boldsymbol{x})\|_\bullet^{bs_0\alpha}) \pi(d\boldsymbol{x}) < \infty$. Noting that $2s_0\alpha = 2s_0 - \rho$ and following the arguments in the proof of Corollary to Theorem 3 in Meitz and Saikkonen (2022) it follows that the stationary version of $\boldsymbol{y}_t$ satisfies $E[|y_t|^{2s_0-\rho}] < \infty$. ■

Proofs of Propositions 1 and 2. For Proposition 1, note that model (24) can be written as $u_t = u_{t-1} + \nu_1 L(u_{t-1}; \gamma, a_1) + \nu_2(1 - L(u_{t-1}; \gamma, a_2)) + \sigma_t \varepsilon_t$ so that the function $g(\cdot)$ in Assumption 2(ii) takes the form $g(u) = u + \nu_1 L(u; \gamma, a_1) + \nu_2(1 - L(u; \gamma, a_2))$. Arguments used in the proof of Proposition 1 in Meitz and Saikkonen (2022) now show that Assumption 2(ii) holds with $\rho = 1$. Applying Theorem 1 with $\delta = 2s_0$ yields the polynomial ergodicity result, and the moment result follows from the remarks made after Theorem 1. As for Proposition 2, model (25) can be written as $u_t = S(u_{t-1})u_{t-1} + \sigma_t \varepsilon_t$ so that now $g(u) = S(u)u$. Assumption 2(ii) can be verified as in the proof of Proposition 2 in Meitz and Saikkonen (2022), and the result follows from Theorem 1 (with $\delta = 2s_0/\rho$). ■

Appendix B

Appendix B contains Figure 2 which displays further analysis of the residuals of model (27).

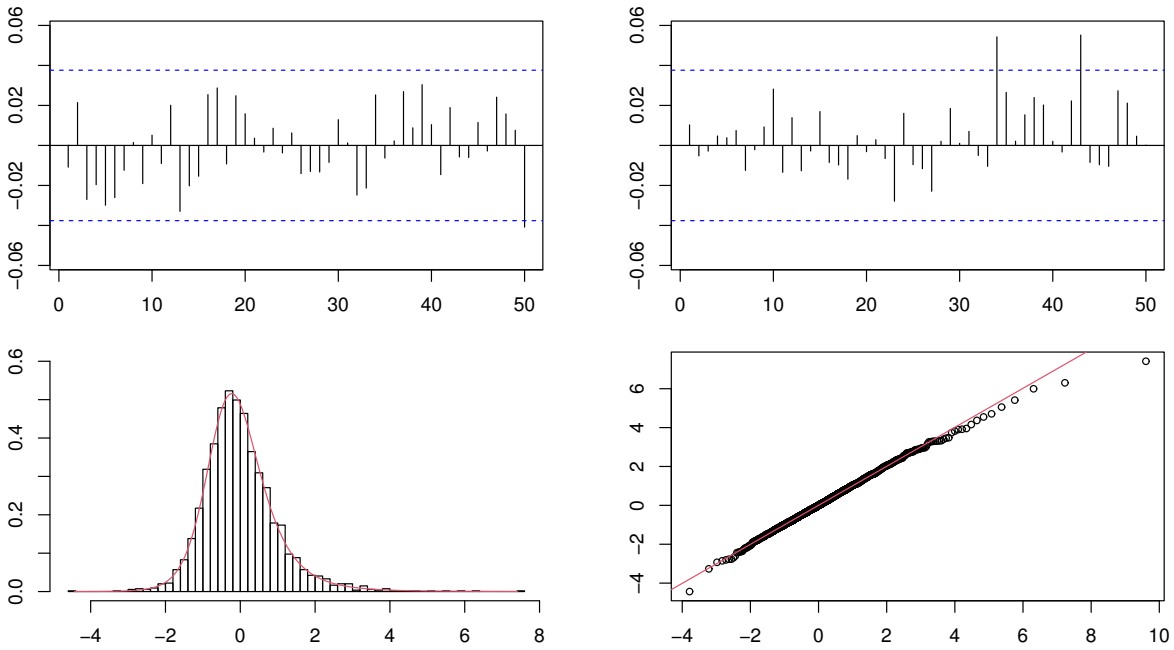

Figure 2: Further analysis of the residuals shown in (the bottom right graph of) Figure 1: autocorrelation function of $\hat{\varepsilon}_t$ (top left), autocorrelation function of $\hat{\varepsilon}_t^2$ (top right), histogram along with the estimated error density (bottom left), and a Q-Q plot (bottom right). The dashed lines in the autocorrelation function graphs show the conventional bounds $\pm 1.96/\sqrt{T} \approx \pm 0.038$ ($T = 2715$; the first four observations are used as initial values).

References

Atchadé, Y. and G. Fort (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli 16*, 116–154.

Cline, D. B. H. (2007). Stability of nonlinear stochastic recursions with application to nonlinear AR–GARCH models. *Advances in Applied Probability 39*, 462–491.

Cline, D. B. H. and H. H. Pu (1998). Verifying irreducibility and continuity of a nonlinear time series. *Statistics & Probability Letters 40*, 139–148.

Cline, D. B. H. and H. H. Pu (2004). Stability and the Lyapounov exponent of threshold AR–ARCH models. *Annals of Applied Probability 14*, 1920–1949.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Douc, R., G. Fort, E. Moulines, and P. Soulier (2004). Practical drift conditions for subgeometric rates of convergence. *Annals of Applied Probability 14*, 1353–1377.

Douc, R., A. Guillin, and E. Moulines (2008). Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques 44*, 239–257.

Douc, R., E. Moulines, P. Priouret, and P. Soulier (2018). *Markov Chains*. Cham: Springer.

Dudley, R. M. (2004). *Real Analysis and Probability*. Cambridge: Cambridge University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Fefferman, C. and H. S. Shapiro (1972). A planar face on the unit sphere of the multiplier space M_p, $1 < p < \infty$. *Proceedings of the American Mathematical Society 36*, 435–439.

Fort, G. and E. Moulines (2000). V-subgeometric ergodicity for a Hastings–Metropolis algorithm. *Statistics & Probability Letters 49*, 401–410.

Fort, G. and E. Moulines (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and their Applications 103*, 57–99.

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis* (2nd ed.). Cambridge University Press.

Jarner, S. F. and G. O. Roberts (2002). Polynomial convergence rates of Markov chains. *Annals of Applied Probability 12*, 224–247.

Jones, M. C. and M. J. Faddy (2003). A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society: Series B 65*, 159–174.

Klokov, S. A. (2007). Lower bounds of mixing rate for a class of Markov processes. *Theory of Probability and Its Applications 51*, 528–535.

Klokov, S. A. and A. Yu. Veretennikov (2004). Sub-exponential mixing rate for a class of Markov chains. *Mathematical Communications 9*, 9–26.

Klokov, S. A. and A. Yu. Veretennikov (2005). On subexponential mixing rate for Markov processes. *Theory of Probability and Its Applications 49*, 110–122.

Lieberman, O. and P. C. B. Phillips (2020). Hybrid stochastic local unit roots. *Journal of Econometrics 215*, 257–285.

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

Ling, S. and M. McAleer (2002). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Econometric Theory 18*, 722–729.

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

Meitz, M. and P. Saikkonen (2008a). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory 24*, 1291–1320.

Meitz, M. and P. Saikkonen (2008b). Stability of nonlinear AR–GARCH models. *Journal of Time Series Analysis 29*, 453–475.

Meitz, M. and P. Saikkonen (2010). A note on the geometric ergodicity of a nonlinear AR–ARCH model. *Statistics & Probability Letters 80*, 631–638.

Meitz, M. and P. Saikkonen (2021). Subgeometric ergodicity and β-mixing. *Journal of Applied Probability 58*, 594–608.

Meitz, M. and P. Saikkonen (2022). Subgeometrically ergodic autoregressions. *Econometric Theory 38*, 959–985.

Merlevède, F., M. Peligrad, and E. Rio (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*, 435–474.

Meyn, S. P. and R. L. Tweedie (2009). *Markov Chains and Stochastic Stability* (2nd ed.). Cambridge: Cambridge University Press.

Nummelin, E. and P. Tuominen (1983). The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Processes and their Applications 15*, 295–311.

Phillips, P. C. B. (2023). Estimation and inference with near unit roots. *Econometric Theory 39*, 221–263.

Tuominen, P. and R. L. Tweedie (1994). Subgeometric rates of convergence of f-ergodic Markov chains. *Advances in Applied Probability 26*, 775–798.

Tweedie, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In J. F. C. Kingman and G. E. H. Reuter (Eds.), *Probability, Statistics and Analysis*, pp. 260–276. Cambridge: Cambridge University Press.

Veretennikov, A. Yu. (2000). On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory of Probability and Its Applications 44*, 361–374.

Vladimirova, M., S. Girard, H. Nguyen, and J. Arbel (2020). Sub-Weibull distributions: Generalizing sub-Gaussian and sub-Exponential properties to heavier tailed distributions. *Stat 9*, e318.

Wong, K. C., Z. Li, and A. Tewari (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Annals of Statistics 48*, 1124–1142.