Econometric Theory, **38**, 2022, 959–985.
doi:10.1017/S0266466620000419

SUBGEOMETRICALLY ERGODIC AUTOREGRESSIONS

MIKA MEITZ
University of Helsinki

PENTTI SAIKKONEN
University of Helsinki

In this paper, we discuss how the notion of subgeometric ergodicity in Markov chain theory can be exploited to study stationarity and ergodicity of nonlinear time series models. Subgeometric ergodicity means that the transition probability measures converge to the stationary measure at a rate slower than geometric. Specifically, we consider suitably defined higher-order nonlinear autoregressions that behave similarly to a unit root process for large values of the observed series but we place almost no restrictions on their dynamics for moderate values of the observed series. Results on the subgeometric ergodicity of nonlinear autoregressions have previously appeared only in the first-order case. We provide an extension to the higher-order case and show that the autoregressions we consider are, under appropriate conditions, subgeometrically ergodic. As useful implications, we also obtain stationarity and β-mixing with subgeometrically decaying mixing coefficients.

1. INTRODUCTION

Markov chain theory and the notion of geometric ergodicity have become standard tools in econometrics and statistics when analyzing the stationarity and ergodicity of nonlinear autoregressions or other nonlinear time series models. A detailed discussion of the relevant Markov chain theory will be given in Section 2. For now, consider a Markov chain X_t $(t = 0, 1, 2, \ldots)$ on the state space X and initialized from X_0 following some initial distribution (that is not necessarily the stationary distribution). Geometric ergodicity of X_t entails that the n-step probability measures $P^n(x; \cdot) = \Pr(X_n \in \cdot \,| X_0 = x)$ converge in total variation norm $\|\cdot\|_{TV}$ to the stationary probability measure π at rate r^n (for some $r > 1$), that is,

$$\lim_{n\to\infty} r^n \|P^n(x; \cdot) - \pi\|_{TV} = 0 \qquad \textbf{(1)}$$

(the definition of $\|\cdot\|_{TV}$ and a formulation of (1) using a more general norm are given in Section 2). A common and convenient way to establish geometric

The authors thank the Academy of Finland for financial support, and Peter Phillips, Donald Andrews, and two anonymous referees for useful comments and suggestions. Address correspondence to Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FI-00014 Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

ergodicity involves the verification of a so-called drift condition. Useful implications obtained with this approach include the existence of a stationary probability distribution π of X_t as well as the geometric β-mixing of X_t. (For a definition of β-mixing, see Doukhan, 1994, Sect. 1.1 or Bradley, 2007, Chap. 3.) The authoritative and classic reference to Markov chain theory is the monograph of Meyn and Tweedie (1993, 2009). Recent papers establishing geometric ergodicity of different nonlinear time series models include Francq and Zakoïan (2006), Ling (2007), Meitz and Saikkonen (2008), and Fokianos, Rahbek, and Tjøstheim (2009), among others.

In this paper, we consider autoregressions that may exhibit rather arbitrary (stationary, unit root, explosive, nonlinear, etc.) behavior for moderate values of the observed series and that behave similarly to a unit root process for large values of the observed series. What this exactly means will be clarified shortly, but first we would like to emphasize that the autoregressions we consider will not necessarily be geometrically ergodic. Under appropriate conditions, they will, nevertheless, satisfy a weaker form of so-called subgeometric ergodicity. A Markov chain is said to be subgeometrically ergodic when the convergence in (1) takes place at a rate $r(n)$ slower than geometric, that is,

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

In the geometric case $r(n)=r^n$ with $r>1$ or, equivalently, $r(n)=e^{cn}$ with $c>0$. Examples of rates slower than geometric include subexponential rates (say, $r(n)=e^{cn^\gamma}$ with $c>0$ and $\gamma\in(0,1)$) and polynomial rates (say, $r(n)=(1+n)^\beta$ with $\beta>0$). For an up-to-date treatment of subgeometric ergodicity, we refer to Chaps. 16 and 17 of Douc et al. (2018) (further references will be given below).

As will be discussed in Section 2, subgeometric ergodicity can conveniently be established by verifying a suitably formulated drift condition and useful implications analogous to those in the case of geometric ergodicity again follow. In particular, the existence of a stationary probability distribution π of X_t as well as the finiteness of certain moments are obtained. Moreover, in a companion paper Meitz and Saikkonen (2019) we show that subgeometric ergodicity implies β-mixing with subgeometrically decaying mixing coefficients. Subgeometric ergodicity therefore allows one to use limit theorems developed for β-mixing processes.

The main aims of this paper are to establish subgeometric ergodicity of certain higher-order nonlinear autoregressions and to illustrate the potential of the concept of subgeometric ergodicity for nonlinear time series models. To facilitate discussion, first consider a simple special case at an informal level. Specifically, consider the univariate first-order nonlinear autoregressive model

$$y_t=g(y_{t-1})+\varepsilon_t,\quad t=1,2,\ldots, \tag{3}$$

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

where the error term ε_t is a sequence of independent and identically distributed (IID) zero-mean random variables and g is a real-valued function. For now, assume that g is such that

$$|g(x)| \leq \left(1 - r|x|^{-\rho}\right)|x| \quad \text{for } |x| \geq M_0 \quad [r > 0, M_0 > r^{1/\rho}, 0 < \rho \leq 2], \tag{4}$$

and that $g(x)$ is bounded for $|x| \leq M_0$. A concrete example where (4) can be easily verified is

$$y_t = \left(1 - \frac{r_0}{1 + |y_{t-1}|^{\rho}}\right) y_{t-1} + \varepsilon_t \quad [r_0 > 0, 0 < \rho \leq 2]. \tag{5}$$

The model defined in (5) can be thought of as first-order autoregression with a time-varying autoregressive coefficient. For large values of $|y_{t-1}|$, the autoregressive coefficient takes values that are close to one and the generation mechanism of y_t is close to a random walk. On the other hand, for small values of $|y_{t-1}|$ the autoregressive coefficient is close to $1 - r_0$ and, depending on the value of r_0, very different behaviors can be accommodated (e.g., stationary behavior for $0 < r_0 < 2$ or explosive behavior for $2 < r_0$). Overall, the generation mechanism of y_t fluctuates between these two borderline cases. Simulated examples in Section 5 demonstrate that processes of the type described in (5) can exhibit behavior close to a random walk for rather long times before returning to a less persistent regime.

When $\rho \geq 1$ the model defined by equation (5) can be viewed as a special case of the model

$$y_t = y_{t-1} + \tilde{g}(y_{t-1}) + \varepsilon_t, \tag{6}$$

where the function $\tilde{g}$ is bounded (but not constant).[1] In Section 3, a higher-order version of equation (6) (without assuming boundedness) is used as a starting point of the formulation of our general model. Our main results in Section 4 show that, depending on the assumptions made, either geometric, subexponential, or polynomial ergodicity is obtained.

The preceding discussion illustrates what kind of behavior the autoregressions we consider may exhibit especially for large values of the observed series. This point could be useful when modeling data which seem to have related characteristics. For instance, interest rate and inflation series might be potential examples. However, it should be emphasized that inequality (4) restricts the regression function g only for large values of its argument. As long as the assumed boundedness condition imposed on the function g is satisfied, no restrictions are required when the process evolves in the vicinity of the origin. Allowing unit root type behavior for large (absolute) values of the observed series is the main feature which distinguishes the models we consider from most previous nonlinear

[1] This model belongs to a class of models referred to as "random-walk-type Markov chains" by Jarner and Tweedie (2003); see particularly equation (3) of their paper.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

autoregressions where stationary behavior is related to large (absolute) values of the process.[2]

We also note that the autoregressions we consider are to some extent related to existing models designed to capture small departures from unit root autoregressions. In first-order autoregressions, these departures appear as time-dependent autoregressive coefficients. For instance, in local-to-unity and mildly (or moderately) integrated models the autoregressive coefficients are deterministic and depend on the sample size of the considered series, whereas in stochastic unit root models the autoregressive coefficients are stochastic and (in some versions) also functions of the sample size. (For more details and for a lucid discussion of these models, see the introduction of Lieberman and Phillips, 2020.) However, the fact that the sample size is an essential part of these models makes them quite different from the autoregressions we consider—in particular, the autoregressions we consider are ergodic.

Previously results on subgeometric ergodicity of nonlinear autoregressions have been obtained in the probability literature 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 (further discussion on these and some related papers will be provided in Section 4). To the best of our knowledge, all of the previous results concern only first-order models. We contribute to this literature by obtaining results for more general higher-order autoregressions. This is achieved using techniques similar to those in the aforementioned papers, especially in Fort and Moulines (2003) and Douc et al. (2004). Depending on the assumptions imposed on the moments of the error term, the resulting rate of ergodicity is either geometric or subexponential or polynomial.

The rest of the paper is organized as follows. Section 2 contains basic concepts of Markov chains and summarizes existing results on subgeometric ergodicity. Section 3 introduces the nonlinear autoregressive model considered and states the assumptions used to obtain the results of the paper. The main results on subexponential and polynomial ergodicity are given in Section 4. In Section 5, we provide examples of our general model. Section 6 concludes. All proofs are collected in an Appendix and an Online Supplementary Appendix.

Finally, a few notational conventions are given. The minimum (maximum) of the real numbers x and y is denoted by $x \wedge y$ ($x \vee y$), and L and Δ signify the lag operator and the difference operator, respectively (so that $\Delta x_t = (1-L)x_t = x_t - x_{t-1}$). The notation $\mathbf{1}_S(x)$ is used for the indicator function which takes the value one when x belongs to the set S and zero elsewhere, and $|\cdot|$ is used for both an absolute value and Euclidean norm. Furthermore, $\mathbf{0}_k$ denotes a $k \times 1$ vector of zeros and $\boldsymbol{\iota}_k = (1, 0, \ldots, 0)$ $(k \times 1)$.

[2]For instance, Lu (1998), Gouriéroux and Robert (2006), and Bec, Rahbek, and Shephard (2008), among others, establish geometric ergodicity (and thus the existence of a stationary distribution) for autoregressions whose behavior approaches stationarity when the process moves away from the origin while in the vicinity of the origin its behavior can be rather arbitrary.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

2. MARKOV CHAINS AND SUBGEOMETRIC ERGODICITY

In this section, we discuss basic concepts of Markov chains needed to obtain our results. More comprehensive discussions can be found in Meyn and Tweedie (2009) and Douc et al. (2018). 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. As in Fort and Moulines (2003) and Douc et al. (2004) our goal is to establish the convergence—in a suitably defined norm and at rate $r(n)$—of the n-step probability measures $P^n(x; \cdot)$ to the stationary distribution π. To this end, let $f : \mathsf{X} \to [1, \infty)$ be an arbitrary fixed measurable function and, for any signed measure μ, define the f-norm $\|\mu\|_f$ as

$$\|\mu\|_f = \sup_{f_0 : |f_0| \le f} |\mu(f_0)|, \tag{7}$$

where $\mu(f_0) = \int_{x \in \mathsf{X}} f_0(x)\mu(dx)$ (and the supremum in (7) runs over all measurable functions $f_0 : \mathsf{X} \to \mathbb{R}$ such that $|f_0(x)| \le f(x)$ for all $x \in \mathsf{X}$). When $f \equiv 1$, the f-norm $\|\mu\|_f$ reduces to the total variation norm $\|\mu\|_{TV} = \sup_{f_0 : |f_0| \le 1} |\mu(f_0)|$ used in (1) and (2).

We aim to establish that 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, that[3]

$$\lim_{n \to \infty} r(n) \|P^n(x; \cdot) - \pi\|_f = 0 \quad \text{for } \pi\text{-almost all } x \in \mathsf{X}. \tag{8}$$

If (8) holds 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 $\pi(f) < \infty$. (For instance, if $\mathsf{X} = \mathbb{R}$ and $f(x) = 1 + x^2$, then (f, r)-ergodicity implies that the stationary distribution of X_t has finite second moments.)

In the probability literature, the preceding definition of (f, r)-ergodicity is standard. However, an equivalent and more transparent formulation is obtained by replacing equation (8) with

$$\lim_{n \to \infty} r(n) \sup_{f_0 : |f_0| \le f} |E[f_0(X_n) \mid X_0 = x] - \pi(f_0)| = 0 \quad \text{for } \pi\text{-almost all } x \in \mathsf{X}$$

(see Tuominen and Tweedie, 1994, p. 776). For instance, if $f(x) = 1 + |x|$, the above equation shows that, for almost any initial value x, the conditional expectation $E[X_n \mid X_0 = x]$ converges to $\int_{x \in \mathsf{X}} x\pi(dx)$, the expectation of the stationary distribution of X_t, and the rate of the convergence is given by $r(n)$.

Most of the recent ergodicity results obtained for nonlinear autoregressions have established geometric ergodicity so that the rate of convergence in (8) is given by $r(n) = r^n$, $r > 1$. The subgeometric rate functions we consider are defined as follows (cf., e.g., Nummelin and Tuominen, 1983 and Douc et al., 2004). Let Λ_0 be

[3]That is, the convergence in (8) 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.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

the set of positive nondecreasing functions $r_0 : \mathbb{N} \to [1, \infty)$ such that $\ln[r_0(n)]/n$ decreases to zero as $n \to \infty$. The class of subgeometric rate functions, denoted by Λ, consists of positive functions $r : \mathbb{N} \to (0, \infty)$ for which there exists some $r_0 \in \Lambda_0$ such that

$$0 < \liminf_{n\to\infty} \frac{r(n)}{r_0(n)} \leq \limsup_{n\to\infty} \frac{r(n)}{r_0(n)} < \infty.$$

Typical examples are obtained of rate functions r for which these inequalities hold with (for notational convenience, we set $\ln(0) = 0$)

$$r_0(n) = (1 + \ln(n))^{\alpha} \cdot (1 + n)^{\beta} \cdot e^{cn^{\gamma}}, \quad \alpha, \beta, c \geq 0, \gamma \in (0, 1).$$

The rate function $r_0(n)$ is called subexponential when $c > 0$, polynomial when $c = 0$ and $\beta > 0$, and logarithmic when $\beta = c = 0$ and $\alpha > 0$. Douc et al. (2004, Sect. 3.3) consider subexponential convergence rates whereas Fort and Moulines (2003, Sect. 2.2) consider polynomial convergence rates in model (3) (see also the related references mentioned in these papers).

The proofs of our results make use of the following condition adapted from Douc et al. (2018, Def. 16.1.7).[4]

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 b such that

$$E[V(X_1) \mid X_0 = x] \leq V(x) - \phi(V(x)) + b\mathbf{1}_C(x), \quad x \in \mathsf{X}. \tag{9}$$

Conditions of this kind are known as drift conditions; when $\phi(v) = \lambda v$ for some $\lambda > 0$ the so-called Foster–Lyapunov drift condition used to establish geometric ergodicity is obtained. For ease of discussion and reference, the following theorem summarizes geometric, subexponential, and polynomial ergodicity results that can be obtained using Condition D. (For the definitions of irreducibility, aperiodicity, and petite sets appearing in the theorem we refer the reader to Meyn and Tweedie, 2009.)

THEOREM 1 (Meyn and Tweedie, 2009; Douc et al., 2004). *Suppose X_t is a ψ-irreducible and aperiodic Markov chain on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and that Condition D holds with a petite set C such that $\sup_{x\in C} V(x) < \infty$ and the function ϕ being either*

(i; geometric case) $\phi(v) = \lambda v$ for some $\lambda > 0$,

(ii; subexponential case) $\phi(v) = c(v + v_0)/[\ln(v + v_0)]^{\alpha}$ for some $c, \alpha, v_0 > 0$, or

(iii; polynomial case) $\phi(v) = cv^{\alpha}$ for some $\alpha \in [0, 1)$ and $c \in (0, 1]$.

[4] A somewhat more general version which allows V to be extended-real-valued (i.e., $V : \mathsf{X} \to [1, \infty]$) is given in Douc et al. (2004).

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

Then X_t is (f,r)-ergodic with either

(i) $f = V$ *and* $r(n) = r^n$ *for some* $r > 1$ *(or, equivalently,* $r(n) = (e^c)^n$ *for some* $c > 0$*),*
(ii) $f = V^\delta$ *and* $r(n) = (e^d)^{n^{1/(1+\alpha)}}$ *for any* $\delta \in (0,1)$ *and any* $d \in (0, (1-\delta)\{c(1+\alpha)\}^{1/(1+\alpha)})$*, or*
(iii) $f = V^{1-\delta(1-\alpha)}$ *and* $r(n) = n^{\delta-1}$ *for any* $\delta \in [1, 1/(1-\alpha)]$.

In the geometric case, the result of Theorem 1 is given in Meyn and Tweedie (2009), and in the subexponential and polynomial cases, the result can be obtained from Douc et al. (2004); some further details are provided in the proof of Theorem 1 in the Supplementary Appendix. Note that in the subexponential case choosing v_0 sufficiently large ensures the concavity of ϕ required in Condition D and also that (in the subexponential case) results with a faster rate of convergence and/or larger f-norm could be obtained at the expense of more complex notation; see Douc et al. (2004, Sect. 2.3).

An essential feature of the subgeometric ergodicity results in Theorem 1 is that there is a trade-off between the rate of convergence and the size of the f-norm; in Theorem 1, the choice of δ reflects this. If a fast rate of convergence is desired one has to accept a small f-norm (recall from (8) that the size of the f-norm is directly proportional to the order of finite moments the stationary distribution is guaranteed to have). For instance, in the polynomial case choosing $\delta = 1/(1-\alpha)$ gives the fastest rate of convergence and with this choice the f-norm reduces to the total variation norm (so that $f \equiv 1$); the extreme case $\alpha = 0$ results in $r(n) \equiv 1$ and standard ergodicity. In the subexponential case, values of δ that are close to zero (one) correspond to small (large) f-norms.

It is also worth noting that Condition D is only sufficient, not necessary, for (f,r)-ergodicity. It is therefore possible that with another drift condition (not necessarily a special case of Condition D), a better rate function could be obtained, but presumably at the cost of a smaller norm. Being able to obtain necessary conditions for particular subgeometric ergodicity rates would be of interest but we will not pursue this issue. Necessary conditions for geometric and polynomial ergodicity in the context of random-walk-type Markov chains (see (6)) are given in Jarner and Tweedie (2003) (for an application of this result to a threshold autoregressive model, see Meitz and Saikkonen, 2019).

As already indicated in the Introduction, the ergodicity results of Theorem 1 imply results on β-mixing or, more specifically, on convergence rates of β-mixing coefficients $\beta(n)$ $(n = 1, 2, \ldots)$ (for a definition of $\beta(n)$ and properties of β-mixing, see Doukhan, 1994, Sect. 1.1, Bradley, 2007, Chap. 3, or Meitz and Saikkonen, 2019). To illustrate this point, let μ signify the distribution of X_0, the initial value of the Markov chain X_t, and assume that $\int_{x\in\mathsf{X}} V(x)\mu(dx) < \infty$ (with V as in Theorem 1). Then, using Thm. 1 and 2 of Meitz and Saikkonen (2019) the

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

three cases in Theorem 1 imply the following convergence rates for β-mixing coefficients (here c and α are as in Theorem 1):[5]

(i) geometric case: $\lim_{n\to\infty} \tilde{r}^n \beta(n) = 0$ for some $\tilde{r} > 1$;

(ii) subexponential case: $\lim_{n\to\infty} (e^{\tilde{d}})^{n^{1/(1+\alpha)}} \beta(n) = 0$ for any $\tilde{d} \in (0, \{c(1+\alpha)/2\}^{1/(1+\alpha)})$;

(iii) polynomial case: $\lim_{n\to\infty} n^{\alpha/(1-\alpha)} \beta(n) = 0$.

Thus, the convergence rates of the β-mixing coefficients are qualitatively similar to the fastest convergence rates of ergodicity obtained in Theorem 1 (as indicated above, a slight improvement can be achieved in the subexponential case). These results, combined with the fact that the (f, r)-ergodicity given in Theorem 1 implies finiteness of moments, make possible to use limit theorems developed for β-mixing processes (and also for α-mixing processes because β-mixing is known to imply α-mixing).

3. MODEL AND ASSUMPTIONS

We now introduce a higher-order generalization of the model discussed in the Introduction. Suppose the process y_t $(t = 1, 2, \ldots)$ is generated by

$$y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \tilde{g}(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t, \tag{10}$$

where $\tilde{g}$ is a real-valued function, the error term ε_t is a sequence of IID random variables, and exactly one of the roots of the polynomial $\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$ is equal to unity and (when $p \geq 2$) all others lie outside the unit circle. Thus, the regression function of the model has a linear part and a nonlinear part, and without the nonlinear part we have a standard linear pth order autoregression with a single unit root (cf. model (6)).

To express (10) in a different way, set $\pi_j = -\sum_{i=j+1}^{p} \varphi_i$ $(j = 1, \ldots, p-1$; when $p = 1$, set $\pi_1 = \cdots = \pi_{p-1} = 0)$ so that we can express the polynomial $\varphi(z)$ as

$$\varphi(z) = (1 - z)(1 - \pi_1 z - \cdots - \pi_{p-1} z^{p-1}),$$

where the roots of the polynomial $\varpi(z) = 1 - \pi_1 z - \cdots - \pi_{p-1} z^{p-1}$ lie outside the unit circle. This shows that we can write equation (10) alternatively as

$$y_t = y_{t-1} + \pi_1 \Delta y_{t-1} + \cdots + \pi_{p-1} \Delta y_{t-p+1} + \tilde{g}(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t. \tag{11}$$

Denoting $u_t = y_t - \pi_1 y_{t-1} - \cdots - \pi_{p-1} y_{t-p+1}$, equation (11) can be written as

$$y_t = \pi_1 y_{t-1} + \cdots + \pi_{p-1} y_{t-p+1} + u_{t-1} + \tilde{g}(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t \tag{12}$$

[5] See also the discussion following Thm. 2 of Meitz and Saikkonen (2019), and note that their Theorem 2(e) is also used to obtain the subexponential rate shown here.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

or as $u_t = u_{t-1} + \tilde{g}(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t$; when $p = 1$ we obtain $y_t = y_{t-1} + \tilde{g}(y_{t-1}) + \varepsilon_t$ as in (6). The formulation in (12) is convenient in our theoretical developments and will therefore be used instead of (11). One reason for this convenience is that in cases where the function $\tilde{g}$ depends on $y_{t-1}, \ldots, y_{t-p}$ only through the linear combination $u_{t-1} = y_{t-1} - \pi_1 y_{t-2} - \cdots - \pi_{p-1} y_{t-p}$ we can write equation (12) (with a slight abuse of notation) in a more compact way as $u_t = u_{t-1} + \tilde{g}(u_{t-1}) + \varepsilon_t$. Then the process u_t can be treated as the first-order model (6) and, as will be discussed shortly, with a suitable assumption, we can make use of results in Fort and Moulines (2003, Sect. 2.2) and Douc et al. (2004, Sect. 3.3) (this turns out to be the case even when $\tilde{g}$ is not a function of the process u_{t-1} only).

Next we introduce the assumptions needed to prove our results. Our first assumption restricts the dynamics in equation (12).

Assumption 1. Suppose the polynomial $\varpi(z) = 1 - \pi_1 z - \cdots - \pi_{p-1} z^{p-1}$ and the function $\tilde{g} : \mathbb{R}^p \to \mathbb{R}$ in (12) satisfy the following conditions:

(i) The roots of $\varpi(z)$ lie outside the unit circle.

(ii) The function $\tilde{g}$ is measurable, bounded on compact subsets of $\mathbb{R}^p$, and there exists a measurable function $g : \mathbb{R} \to \mathbb{R}$ with the property $|g(x)| \to \infty$ as $|x| \to \infty$ such that the following two conditions hold.

(ii.a) With $\boldsymbol{x} = (x_1, \ldots, x_p)$ and $u = x_1 - \pi_1 x_2 - \cdots - \pi_{p-1} x_p$, the function $\tilde{g}$ satisfies

$$|u + \tilde{g}(\boldsymbol{x}) - g(u)| \leq |\epsilon(\boldsymbol{x})\boldsymbol{x}|, \tag{13}$$

where $\epsilon(\boldsymbol{x})$ is a real-valued function such that $|\epsilon(\boldsymbol{x})| = o(|\boldsymbol{x}|^{-d})$ as $|\boldsymbol{x}| \to \infty$ for some $d > 0$.

(ii.b) There exist positive constants r, M_0, K_0, and $0 < \rho \leq 2$ such that for all $u \in \mathbb{R}$

$$|g(u)| \leq \begin{cases} (1 - r|u|^{-\rho})\,|u| & \text{for } |u| \geq M_0, \\ K_0 & \text{for } |u| \leq M_0. \end{cases} \tag{14}$$

Assumption 1(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 1(ii) requires the function $\tilde{g}$ to be bounded on compact subsets and links it to another function g. Condition (ii.a) controls the difference between the functions $u + \tilde{g}(\boldsymbol{x})$ and $g(u)$ or, in model (12), the difference between the processes $u_{t-1} + \tilde{g}(y_{t-1}, \ldots, y_{t-p})$ and $g(u_{t-1})$. In the special case where the function $\tilde{g}$ depends on u only, condition (ii.a) becomes obvious because then one can choose $u + \tilde{g}(\boldsymbol{x}) = g(u)$ and $\epsilon(\boldsymbol{x}) = 0$, and it suffices to check condition (ii.b) only. In this case we can use results in Fort and Moulines (2003, Sect. 2.2) and Douc et al. (2004, Sect. 3.3) directly in our proofs. However, we can do the same,

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

albeit in a more complicated way, also when the function $\tilde{g}$ depends on the whole p-dimensional vector $\boldsymbol{x}$, but then the difference between the functions $u+\tilde{g}(\boldsymbol{x})$ and $g(u)$ may not increase "too fast" when $|\boldsymbol{x}|$ gets large. What is "too fast" is controlled by the function $\epsilon(\boldsymbol{x})$ (for instance, when $d \geq 1$ the difference between $u+\tilde{g}(\boldsymbol{x})$ and $g(u)$ becomes negligible when $|\boldsymbol{x}|$ increases).

To illustrate condition (ii.a), note that it implies $|u+\tilde{g}(\boldsymbol{x})| \leq |g(u)| + |\epsilon(\boldsymbol{x})\boldsymbol{x}|$ which, combined with condition (ii.b), yields

$$|u+\tilde{g}(\boldsymbol{x})| \leq \left(1 - r\,|u|^{-\rho}\right)|u| + o(|\boldsymbol{x}|^{-d})\,|\boldsymbol{x}| \quad \text{for } |u| \geq M_0. \tag{15}$$

Therefore $|u+\tilde{g}(\boldsymbol{x})|$ has an upper bound similar to that $|g(u)|$ has in (14) but with some "slackness" allowed through the term $o(|\boldsymbol{x}|^{-d})\,|\boldsymbol{x}|$. This fact is used in our proofs.

A convenient way to verify condition (ii.a) is to note that it is implied by the equality $u+\tilde{g}(\boldsymbol{x}) = g(u) + \tilde{\epsilon}(\boldsymbol{x})\theta'\boldsymbol{x}$ (where θ is a p-dimensional parameter vector) together with the assumption $\tilde{\epsilon}(\boldsymbol{x}) = o(|\boldsymbol{x}|^{-d})$. These conditions ensure that (ii.a) holds with $\epsilon(\boldsymbol{x}) = |\theta|\,\tilde{\epsilon}(\boldsymbol{x})$. We illustrate this approach for checking condition (ii.a) in Section 5.3.

Condition (ii.b) is similar to its first-order counterpart (4) to which it reduces when $p = 1$. Note that apart from the boundedness condition, no restrictions are placed on $g(u)$ for moderate values of u. In the higher-order case this assumption concerns the filtered process $u_t = \varpi(L)y_t$. In the first-order case, we also have $\boldsymbol{x} = u$ and the easiest way to verify Assumption 1(ii) may then be to define the function g as $g(x) = x + \tilde{g}(x)$ (and $\epsilon(\boldsymbol{x}) = 0$), and verify condition (ii.b) directly. When $p \geq 2$, the fact that the domain of the function $\tilde{g}$ is larger than that of g complicates the situation in that then no simple connection between inequalities (13) and (14) can generally be found. An example of this case is provided in Section 5.

Our second assumption gives conditions required of the error term in equation (12).

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

(a) $E\left[e^{\beta_0|\varepsilon_1|^{\kappa_0}}\right] < \infty$ for some $\beta_0 > 0$ and $\kappa_0 \in (0, 1]$, and $E[\varepsilon_1] = 0$; or
(b) $E\left[|\varepsilon_1|^{s_0}\right] < \infty$ for some $s_0 > \rho$ (with ρ as in Assumption 1), and $E[\varepsilon_1] = 0$ holds if $\rho \geq 1$.

Assumption 2(a) corresponds to Assump. 3.3 of Douc et al. (2004, Sect. 3.3), whereas Assumption 2(b) is a combination of the conditions imposed in (NSS 1), (NSS 4), and Lem. 3 of Fort and Moulines (2003, Sect. 2.2). The boundedness condition imposed in Assumption 2 on the density of the error term is stronger than would be needed but is used for simplicity (see Assump. 3.3 of Douc et al., 2004, Sect. 3.3 for a more general alternative).

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

Note that finiteness of the first expectation in Assumption 2(a) holds with $\kappa_0 = 1$ if the distribution of ε_1 has a moment generating function in some interval of the origin. Although many widely used distributions satisfy this condition some heavy tailed distributions are ruled out (this applies to distributions whose densities cannot be bounded by a term of the form $c_1 e^{-c_2|x|}$ with c_1 and c_2 positive constants). An example is Student's t-distribution irrespective of the value of the degrees of freedom parameter. The condition in Assumption 2(b) is used to address this issue. In this condition, the case $0 < \rho < s_0 < 1$ is rather extreme in that not even the expectation $E[\varepsilon_1]$ is assumed to exist.

4. RESULTS

We now present our ergodicity results which we base on model (12). In Section 4.1, the rate of ergodicity established is subexponential whereas a slower polynomial rate of ergodicity is obtained in Section 4.2. The difference between these two cases stems from the assumed moment conditions: in Section 4.1, the condition in Assumption 2(a) is assumed whereas in Section 4.2, the weaker condition in Assumption 2(b) is employed. First we have to present the companion form of model (12) which applies to both of these cases and will be needed in the proofs of our theorems.

To simplify notation, denote $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})$ and define the function $\overline{g} : \mathbb{R}^p \to \mathbb{R}$ as

$$\overline{g}(\boldsymbol{x}) = x_1 - \pi_1 x_2 - \cdots - \pi_{p-1} x_p + \tilde{g}(\boldsymbol{x}) = u + \tilde{g}(\boldsymbol{x})$$

so that $\overline{g}(\boldsymbol{y}_{t-1}) = u_{t-1} + \tilde{g}(\boldsymbol{y}_{t-1})$. It is readily seen that the companion form related to equation (12) reads as

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

or, with obvious matrix notation,

$$\boldsymbol{y}_t = \boldsymbol{\Phi} \boldsymbol{y}_{t-1} + \overline{g}(\boldsymbol{y}_{t-1}) \boldsymbol{\iota}_p + \varepsilon_t \boldsymbol{\iota}_p \tag{16}$$

(when $p = 1$, $\boldsymbol{\Phi} = 0$). Thus, Assumption 2 ensures that $\boldsymbol{y}_t$ is a Markov chain on $\mathbb{R}^p$. For later purposes it is convenient to transform the companion form (16). To this

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

end, we define the matrices

$$\mathbf{A} = \begin{bmatrix} 1 & -\pi_1 & -\pi_2 & \cdots & -\pi_{p-1} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad \text{and} \quad \Pi = \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^{-1} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & \pi_1 & \pi_2 & \cdots & \pi_{p-1} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \mathbf{0}'_{p-1} \\ \boldsymbol{\iota}_{p-1} & \boldsymbol{\Pi}_1 \end{bmatrix}, \tag{17}$$

where $\mathbf{A}$ is nonsingular and $\boldsymbol{\Pi}_1$ is the $(p-1)\times(p-1)$ dimensional lower right hand corner of Π (when $p=1$, $\mathbf{A}=1$ and $\Pi=0$). With these definitions (16) can be transformed into

$$\mathbf{A}\boldsymbol{y}_t = \Pi\mathbf{A}\boldsymbol{y}_{t-1} + \overline{g}(\boldsymbol{y}_{t-1})\boldsymbol{\iota}_p + \varepsilon_t\boldsymbol{\iota}_p, \tag{18}$$

where $\mathbf{A}\boldsymbol{y}_t = (u_t, y_{t-1}, \ldots, y_{t-p+1})$. Now, for any p-dimensional vector $\boldsymbol{x}$, form the partition $\boldsymbol{x} = (x_1, \ldots, x_p) = (x_1, \boldsymbol{x}_2)$ and define $\boldsymbol{z}(\boldsymbol{x}) = (z_1(\boldsymbol{x}), \boldsymbol{z}_2(\boldsymbol{x})) = \boldsymbol{A}\boldsymbol{x}$, where (due to (17)) $z_1(\boldsymbol{x}) = x_1 - \pi_1 x_2 - \cdots - \pi_{p-1}x_p$ and $\boldsymbol{z}_2(\boldsymbol{x}) = \boldsymbol{x}_2$ (when $p=1$, $\boldsymbol{x}_2$ and $\boldsymbol{z}_2(\boldsymbol{x})$ are dropped). With this notation equation (18) can be expressed as

$$\begin{bmatrix} z_1(\boldsymbol{y}_t) \\ \boldsymbol{z}_2(\boldsymbol{y}_t) \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0}'_{p-1} \\ \boldsymbol{\iota}_{p-1} & \boldsymbol{\Pi}_1 \end{bmatrix} \begin{bmatrix} z_1(\boldsymbol{y}_{t-1}) \\ \boldsymbol{z}_2(\boldsymbol{y}_{t-1}) \end{bmatrix} + \overline{g}(\boldsymbol{y}_{t-1})\boldsymbol{\iota}_p + \varepsilon_t\boldsymbol{\iota}_p$$

$$= \begin{bmatrix} \overline{g}(\boldsymbol{y}_{t-1}) + \varepsilon_t \\ \boldsymbol{\Pi}_1\boldsymbol{z}_2(\boldsymbol{y}_{t-1}) + z_1(\boldsymbol{y}_{t-1})\boldsymbol{\iota}_{p-1} \end{bmatrix}. \tag{19}$$

The first equation in (19) is now in a form that can be analyzed by using the results in Fort and Moulines (2003) and Douc et al. (2004). As for the second equation, by Assumption 1(i), the roots of the polynomial $\varpi(z)$ lie outside the unit circle, so that the eigenvalues of the matrix $\boldsymbol{\Pi}_1$ are smaller than one in absolute value. As is well known, this implies the existence of a matrix norm $\|\cdot\|_*$ induced by a vector norm, also denoted by $\|\cdot\|_*$, such that $\|\boldsymbol{\Pi}_1\|_* \le \eta$ for some $\eta < 1$ (see, e.g., Def. 5.6.1 and Lem. 5.6.10 in Horn and Johnson, 2013). These facts will be useful in our proofs.

4.1. Subexponential Case

Our results make use of Condition D which requires choosing the function V. To this end, let b_1, b_2, and b_3 be positive constants whose values (to be specified later) depend on the constants β_0, κ_0, and ρ introduced in Assumptions 1 and 2; for b_3, we already mention that it will always satisfy $b_3 \in (0,1]$. When $p \ge 2$, we define the function V as

$$V(\boldsymbol{x}) = \tfrac{1}{2}\exp\{b_1 |z_1(\boldsymbol{x})|^{b_3}\} + \tfrac{1}{2}\exp\{b_2 \|\boldsymbol{z}_2(\boldsymbol{x})\|_*^{b_3}\} \tag{20}$$

and when $p=1$, we define $V(x) = \exp\{b_1 |x|^{b_3}\}$.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

Now we can state the following theorem which makes use of the stronger moment requirement in Assumption 2(a). (The proof is given in the Supplementary Appendix.)

THEOREM 2. *Suppose $p \geq 2$ and consider the Markov chain $\mathbf{y}_t$ defined in equation (16). Let Assumptions 1 and 2(a) hold, suppose that in Assumption 1 the constants ρ and d satisfy $0 < \rho < 2$ and $d = \rho/b_3$, and let $V(\boldsymbol{x})$ be as in (20).*

(i) *If $\rho > \kappa_0$, then $\mathbf{y}_t$ is (f, r)-ergodic with the subexponential convergence rate $r(n) = e^{kn^{b_3/\rho}}$ and the function f given by $f(\boldsymbol{x}) = V(\boldsymbol{x})^{\delta}$; this result holds for any choice of $\delta \in (0, 1)$, for any k such that $0 < k < (1-\delta)\{c\rho/b_3\}^{b_3/\rho}$, and for some (small) $b_1, b_2 \in (0, \beta_0)$, $b_3 = \kappa_0 \wedge (2-\rho) \in (0, 1)$, and some (small) $c > 0$.*

(ii) *If $\rho = \kappa_0$, then $\mathbf{y}_t$ is geometrically ergodic with the convergence rate $r(n) = e^{cn}$ and the function f given by $f(\boldsymbol{x}) = V(\boldsymbol{x})$; this result holds for some (small) $b_1, b_2 \in (0, \beta_0)$, $b_3 = \kappa_0 \in (0, 1]$, and some $c > 0$.*

When $p = 1$, consider the Markov chain y_t defined by $y_t = y_{t-1} + \tilde{g}(y_{t-1}) + \varepsilon_t$. The above results hold for y_t with the function V defined as $V(x) = \exp\{b_1 |x|^{b_3}\}$ (and the constant d becomes redundant).

In Theorem 2, the case $\rho = \kappa_0$ represents a qualitative change in the ergodic behavior of the considered Markov chain: For $\rho > \kappa_0$, a slower subexponential convergence rate is obtained and $\rho = \kappa_0$ is the borderline case where a change to the faster geometric rate occurs. For $\rho < \kappa_0$, geometric ergodicity could also be established but we omit this case for brevity (in the first-order case this result is an immediate consequence of Thm. 3.3 of Douc et al., 2004, Sect 3.3). Note also that, by the definition of the constant b_3, the rate of ergodicity in the subexponential case decreases as the value of ρ increases. Furthermore, the formulation for the subexponential case involves many parameters to allow for a trade-off between the rate of convergence and the size of the f-norm: values of δ that are close to zero (one) correspond to small (large) f-norms and allow for a larger (smaller) choice of k leading to a faster (slower) rate.

Previously, Douc et al. (2004, Sect 3.3) obtained the results of Theorem 2 in the first-order case; our primary purpose here is to provide higher-order analogs of their results. Klokov and Veretennikov (2004, 2005) and Klokov (2007) have also studied the first-order model (3) satisfying inequality (4) with $1 < \rho < 2$ but otherwise their assumptions are rather different from ours. They obtain subexponential bounds for ergodicity in total variation norm (i.e., $(1, r)$-ergodicity) and for β-mixing coefficients but they do not discuss general (f, r)-ergodicity.

As discussed in Section 2, we can also establish β-mixing and, in contrast to Klokov and Veretennikov (2004, 2005) and Klokov (2007), we can permit all initial values with distribution μ such that $\int_{x \in \mathbb{R}^p} V(x)\mu(dx) < \infty$ (and V as in Theorem 2). Specifically, the discussion at the end of Section 2 and Theorem 2 imply β-mixing with the following rates: In case (i), the rate is subexponential,

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

that is, $\lim_{n\to\infty} e^{\tilde{k}n^{b_3/\rho}}\beta(n) = 0$ with any $\tilde{k} \in (0, \{c\rho/2b_3\}^{b_3/\rho})$, and in case (ii) the rate is geometric, that is, $\lim_{n\to\infty} \tilde{r}^n\beta(n) = 0$ for some $\tilde{r} > 1$ (or, equivalently, $\lim_{n\to\infty} e^{\tilde{c}n}\beta(n) = 0$ for some $\tilde{c} > 0$).

The following corollary is an immediate consequence of the discussion after equality (8) (for a formal result, see Thm. 14.0.1 in Meyn and Tweedie, 2009).

COROLLARY TO THEOREM 2. *Let π signify the stationary distribution of $\mathbf{y}_t$ in Theorem 2 and let the function f be as in cases (i) and (ii) of Theorem 2. Then $\pi(f) = \int_{\boldsymbol{x}\in\mathbb{R}^p} f(\boldsymbol{x})\pi(d\boldsymbol{x}) < \infty$; in particular, $\pi(|\boldsymbol{x}|^s) < \infty$ for all $s > 0$ so that the stationary distribution has finite moments of all orders.*

As we remarked after Theorem 1, in the subexponential case of Theorem 2 results with a faster rate of convergence and/or larger f-norm could be obtained at the expense of more complex notation. This means that in the above corollary finiteness of slightly larger moments could be obtained and the subexponential β-mixing rate discussed above could similarly be slightly improved.

4.2. Polynomial Case

Next we consider ergodicity results relying only on the weaker moment requirement in Assumption 2(b). This will below lead to a slower polynomial rate of ergodicity. The key result used to relax the moment requirement is Lem. 3 of Fort and Moulines (2003, Sect. 2.2) (which the authors use in conjunction with their analog of Condition D; we depart from their approach and use Condition D which corresponds to an analogous condition described in Sect. 1.2 in Fort and Moulines, 2003).

The function V employed is now different from the subexponential case. When $p \geq 2$, we define the function V as

$$V(\boldsymbol{x}) = 1 + |z_1(\boldsymbol{x})|^{s_0} + s_1\|\boldsymbol{z}_2(\boldsymbol{x})\|_*^{\alpha s_0}, \tag{21}$$

where s_0 is as in Assumption 2(b), $\alpha = 1 - \rho/s_0$ with ρ as in Assumption 1(ii.b), s_1 is a positive constant (to be specified later), and the norm $\|\cdot\|_*$ is as in the discussion following equation (19); when $p = 1$, we define $V(x) = 1 + |x|^{s_0}$.

The following theorem presents the ergodicity result obtained when using the weaker moment condition in Assumption 2(b). (The proof is given in the Supplementary Appendix.)

THEOREM 3. *Suppose $p \geq 2$ and consider the Markov chain $\mathbf{y}_t$ defined in equation (16). Let Assumptions 1 and 2(b) hold, suppose that in Assumption 1 the constants ρ and d satisfy $0 < \rho \leq 2$ and $d = \rho/s_0$ when $s_0 < 1$ and $d = \rho$ when $s_0 \geq 1$, and let $V(\boldsymbol{x})$ be as in (21). Assume further that either*

(i) $0 < \rho < 1$ *and* $s_0 > \rho$,
(ii) $1 \leq \rho < 2$ *and either* $s_0 = 2$ *or* $s_0 \geq 4$, *or*
(iii) $\rho = 2$ *and* $s_0 \geq 4$ *with* $s_0 r - \frac{1}{2}s_0(s_0 - 1)E[\varepsilon_1^2] > 0$.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

Then $\mathbf{y}_t$ is (f,r)-ergodic with the polynomial convergence rate $r(n) = n^{\delta-1}$ and the function f given by $f(\boldsymbol{x}) = V(\boldsymbol{x})^{1-\delta\rho/s_0}$; this result holds for any choice of $\delta \in [1, s_0/\rho]$ and for some (small) $s_1 > 0$.

When $p = 1$, consider the Markov chain y_t defined by $y_t = y_{t-1} + \tilde{g}(y_{t-1}) + \varepsilon_t$. The above results hold for y_t with the functions V and f defined as $V(x) = 1 + |x|^{s_0}$ and $f(x) = 1 + |x|^{s_0-\delta\rho}$ (and the constant d becomes redundant).

Options (i)–(iii) in Theorem 3 represent the combinations of the values of ρ and s_0 for which the result can be obtained by relying on the corresponding cases (i)–(iii) in Lem. 3 of Fort and Moulines (2003, Sect. 2.2). Unlike in Theorem 2, the case $\rho = 2$ is allowed, but then an additional and rather intricate moment condition is required. A further departure from Theorem 2 is that the same polynomial rate of ergodicity is obtained in all cases. However, similarly to the subexponential case in Theorem 2, the rate of ergodicity decreases as the value of ρ increases. Also, from the discussion at the end of Section 2, we can conclude that the rate of β-mixing implied by Theorem 3 is polynomial and, specifically, $\lim_{n\to\infty} n^{s_0/\rho-1}\beta(n) = 0$.

The first-order case of Theorem 3 was obtained by Fort and Moulines (2003, Sect. 2.2) (with slightly different assumptions). Polynomial ergodicity results for first-order autoregressions similar to that in (3) have previously appeared also in Tuominen and Tweedie (1994, Sect. 5.2) (in the case $0 < \rho < 1$), Tanikawa (2001) (in the case $\rho = 1$), and Veretennikov (2000) and Klokov (2007) (in the case $\rho = 2$; these authors also obtain polynomial bounds for β-mixing coefficients but do not consider general (f,r)-ergodicity).

The following corollary on the moments of the stationary distribution is proved in the Supplementary Appendix. (In contrast to the subexponential case, using the (f,r)-ergodicity result of Theorem 3 would here yield a weaker moment result; hence, some extra steps are needed.)

COROLLARY TO THEOREM 3. *Let π signify the stationary distribution of $\mathbf{y}_t$ in Theorem 3. Then $\pi(f) = \int_{\boldsymbol{x}\in\mathbb{R}^p} f(\boldsymbol{x})\pi(d\boldsymbol{x}) < \infty$ with $f(\boldsymbol{x}) = |\boldsymbol{x}|^{s_0-\rho}$ so that the stationary distribution has finite moments up to order $s_0 - \rho$.*

5. ILLUSTRATIVE EXAMPLES

In this section, we discuss special cases of the general model introduced in Section 3. Using the three equivalent formulations (10)–(12) in Section 3, the model considered can be written as

$$\begin{aligned} y_t - \varphi_1 y_{t-1} - \cdots - \varphi_p y_{t-p} &= \Delta y_t - \pi_1 \Delta y_{t-1} - \cdots - \pi_{p-1}\Delta y_{t-p+1} \\ &= u_t - u_{t-1} = \tilde{g}(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t, \end{aligned}$$

where the polynomial $\varphi(z)$ has precisely one unit root, can be decomposed as $\varphi(z) = (1-z)\varpi(z)$, and $u_t = \varpi(L)y_t$. Further formal assumptions will be stated in Propositions 1 and 2 below.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

First-order subgeometrically ergodic autoregressions were already exemplified, albeit at a rather general level, in Fort and Moulines (2003, Sect. 2.2) and Douc et al. (2004, Sects. 3.3 and 3.4). In Meitz and Saikkonen (2019, Sect. 5), we study rates of subgeometric ergodicity and β-mixing in a first-order multiregime self-exciting threshold autoregressive (SETAR) model; the proof of Theorem 3 in that paper illustrates how Theorems 2 and 3 of the present paper can be applied in a first-order case. In what follows we focus on examples of higher-order subgeometrically ergodic autoregressive models.

We consider three main examples. We first give a heuristic overview of them and then present the formalities. The first example we consider can be expressed as

$$u_t = u_{t-1} + I(u_{t-1}) + \varepsilon_t, \tag{22}$$

where $I(u_{t-1})$ is not constant and will be interpreted as a time-varying drift or intercept term (the case where $I(u_{t-1})$ is a constant is not of interest here because then model (22) reduces to a nonstationary unit root model (with or without a drift)). We consider the case where $I(u_{t-1})$ takes values in a bounded interval and fluctuates suitably between increasing and decreasing drifts. This ensures that model (22) will be (subgeometrically) ergodic and stationary under appropriate conditions (see Proposition 1 below) even though its dynamics involve a unit root component and a (time-varying) drift.

The second example we consider is

$$u_t - \nu = S(u_{t-1})(u_{t-1} - \nu) + \varepsilon_t, \tag{23}$$

where $\nu \in \mathbb{R}$ is an intercept term and $S(u_{t-1})$ will be interpreted as a time-varying slope coefficient. In the cases $S(u_{t-1}) \equiv 0$ and $S(u_{t-1}) \equiv 1$ model (23) reduces to the (linear) stationary model $\varpi(L)y_t = u_t = \varepsilon_t$ and to the nonstationary unit root model $\varphi(L)y_t = \varepsilon_t$, respectively. We consider the case of $S(u_{t-1})$ being time-varying, taking values in some interval $[s, 1)$ ($s < 1$), and attaining values arbitrarily close to 1 for values of u_{t-1} large in absolute value. Under appropriate conditions (see Proposition 2 below), model (23) will be (subgeometrically) ergodic and stationary although exhibiting features similar to a unit root process.

Our third example generalizes the second one and illustrates that one can allow nonlinear dependence not only on u_{t-1} but on the entire $\boldsymbol{y}_{t-1}$. Specifically, we consider model

$$u_t = S(u_{t-1})u_{t-1} + F(\boldsymbol{y}_{t-1}) + \varepsilon_t, \tag{24}$$

where we have omitted the intercept for simplicity, $S(u_{t-1})$ again represents a time-varying slope coefficient, and the term $F(\boldsymbol{y}_{t-1})$ captures the nonlinear dependence on the entire $\boldsymbol{y}_{t-1}$.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

5.1. Example with Time-Varying Intercept Term of LSTAR Type

We consider example (22) with the time-varying intercept term $I(u_{t-1})$ specified as in logistic smooth transition autoregressive (LSTAR) models (see, e.g., van Dijk, Teräsvirta, and Franses, 2002). Specifically, we choose

$$I(u_{t-1}) = \nu_1 L(u_{t-1}; b, a_1) + \nu_2 (1 - L(u_{t-1}; b, a_2)) \tag{25}$$

with $L(u; b, a) = 1/(1 + e^{-b(u-a)})$ denoting the logistic function.[6] The parameters b, a_1, a_2 are assumed to satisfy $b > 0$ and $a_1 \leq a_2$ as usual, and ν_1, ν_2 are assumed to satisfy $\nu_1 < 0 < \nu_2$ to obtain ergodicity below. The time-varying intercept term $I(u_{t-1})$ now takes values in the interval (ν_1, ν_2). Note that for large values of b the logistic function $L(u; b, a)$ is close to the indicator function and then this model provides a close approximation to the above-mentioned threshold autoregressive model where $L(u_{t-1}; b, a_i)$ is replaced with an indicator function.

The following proposition shows the ergodicity of this model.

PROPOSITION 1. *Consider the process y_t defined by $u_t = u_{t-1} + I(u_{t-1}) + \varepsilon_t$ as in (22) (with $u_t = \varpi(L) y_t$ and the roots of $\varpi(z)$ outside the unit circle), and with $I(u_{t-1})$ as in (25) (with $\nu_1 < 0 < \nu_2$). Assume further that either*

(1) *Assumption 2(a) is satisfied with $\kappa_0 \in (0, 1)$,*
(2) *Assumption 2(a) is satisfied with $\kappa_0 = 1$, or*
(3) *Assumption 2(b) is satisfied with either $s_0 = 2$ or $s_0 \geq 4$.*

Then, under condition (1)/(2)/(3), the process $\mathbf{y}_t = (y_t, \ldots, y_{t-p+1})$ is either

(1) *subexponentially ergodic with convergence rate $r(n) = (e^k)^{n^{\kappa_0}}$ (for some $k > 0$),*
(2) *geometrically ergodic with convergence rate $r(n) = (e^c)^n$ (for some $c > 0$), or*
(3) *polynomially ergodic with convergence rate $r(n) = n^{s_0 - 1}$.*

The proof of Proposition 1 is a straightforward application of Theorems 2 and 3 in the case $\rho = 1$ (see the Appendix). Depending on moment assumptions the rate of ergodicity is geometric, subexponential, or polynomial, and similar rates also apply to β-mixing coefficients (see the discussions following Theorems 2 and 3 as well as the corollaries to these theorems for existence of finite moments of the stationary distribution).

We now provide intuitive and graphical illustrations of the behavior processes covered by Proposition 1 can exhibit. An informal description captures the main features concisely: For "values of u_{t-1} in the extreme left tail," the intercept term $I(u_{t-1})$ is close to $\nu_2 > 0$ resulting in increasing drift toward "central values of

[6] The logistic function used in (25) is as in logistic STAR models, and below in (26) the functional form employed is similar to that used in exponential STAR models. However, the STAR models in (at least most of the) previous literature behave like stationary mean-reverting processes when they take large values, contrarily to the processes considered in this paper.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

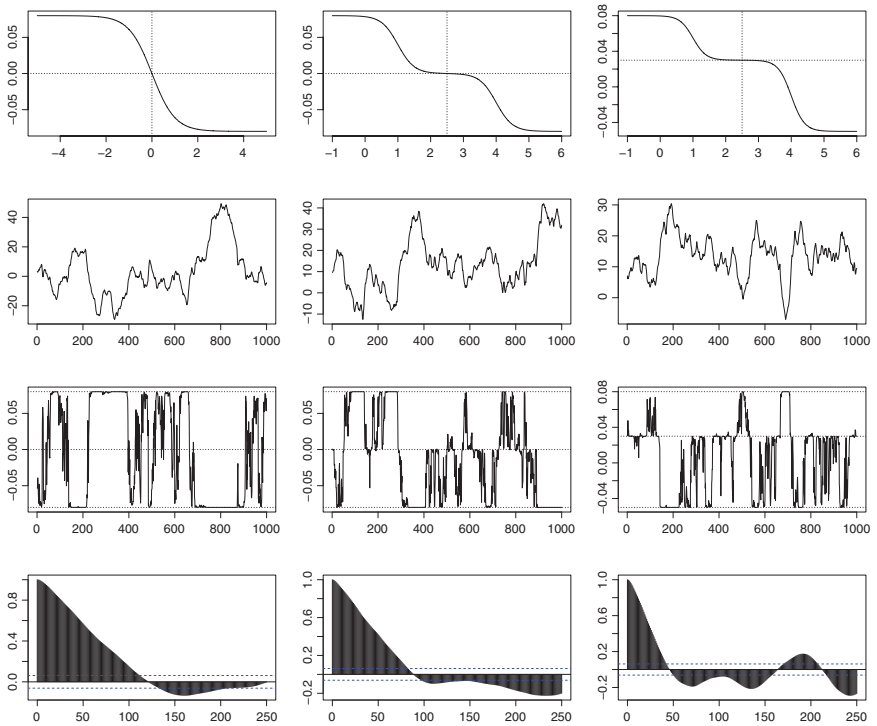


FIGURE 1. *Top row.* Graphs of the function $I(u) = \nu_1 L(u; b, a_1) + \nu_2(1 - L(u; b, a_2))$ in (25). Left: $b = 2$, $\nu_1 = -0.08$, $\nu_2 = 0.08$, and $a_1 = a_2 = 0$. Middle: $b = 4$, $\nu_1 = -0.08$, $\nu_2 = 0.08$, $a_1 = 1$, $a_2 = 4$. Right: $b = 5$, $\nu_1 = -0.05$, $\nu_2 = 0.08$, $a_1 = 1$, and $a_2 = 4$ (the horizontal dotted line at 0.03). *Second row.* Simulated time series of y_t corresponding to the time-varying intercept functions $I(u_t)$ in the top row. The 1,000 observations are generated from model (22) with $p = 2$, $\pi_1 = 0.75$, and ε_t distributed as rescaled Student's t distribution with five degrees of freedom (and $E[\varepsilon_1^2]$ equal to 0.300 on the left, 0.250 in the middle, and 0.125 on the right). *Third row.* The corresponding time series graphs of $I(u_t)$. The dotted lines show the minimum and maximum, and the value zero in the first and second columns and 0.03 in the third column. *Bottom row.* The corresponding autocorrelation functions of the simulated series. The first three autocorrelations are 0.999, 0.996, 0.992 (left); 0.997, 0.991, 0.984 (middle); and 0.996, 0.987, 0.973 (right).

u_{t-1}"; conversely, for "values of u_{t-1} in the extreme right tail," $I(u_{t-1})$ is close to $\nu_1 < 0$ and decreasing drift toward "central values of u_{t-1}" takes place; finally, for such "central values of u_{t-1}," unit root behavior without drift occurs. This informal description is illustrated in the top row of Figure 1 using three different cases of the function $I(\cdot)$ in (25); the precise parameter values used can be found in the caption of Figure 1.

The graph on the left illustrates the simplest possible case. When u takes negative values the function $I(u)$ takes positive values; for $u < -2$ the values of $I(u)$ are

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

very close to its supremum, in this example 0.08. As $u_t = I(u_{t-1}) + u_{t-1} + \varepsilon_t$, the behavior of the process u_t is for $u_{t-1} < 0$ (and particularly for $u_{t-1} < -2$) close to that of a unit root process with an increasing drift. A similar behavior occurs when u_{t-1} takes positive values (and particularly when $u_{t-1} > 2$) but then the drift is decreasing. Thus, large absolute values of u_{t-1} induce a drift toward the origin which provides intuition why the process u_t can be ergodic and stationary even though its behavior shows resemblance to a unit root process with a (negligible) drift near the origin where $I(u) \approx 0$. The other two graphs in the top row of Figure 1 illustrate similar but somewhat more involved possibilities for the function $I(u)$. The middle graph illustrates that the "unit root regime" can take place over a wider range of values of u and these values need not be centered at the origin; in this graph $I(u) \approx 0$ for u roughly between 1.5 and 3.5. The graph on the right illustrates a possibility in which also the "middle regime" corresponds to a unit root with positive drift.

The second row of Figure 1 presents three examples of simulated time series of y_t. These are generated from second-order versions of model (22) using the three time-varying intercept terms $I(u_{t-1})$ depicted in the first row. In all cases the autoregressive coefficient π_1 is equal to 0.75 and the error terms ε_t are generated from rescaled Student's t-distribution with five degrees of freedom (and different variances, see the caption for details). The third and fourth rows of Figure 1 show the time series graphs of the time-varying intercept terms $I(u_t)$ and the autocorrelation functions of y_t, respectively, in these three examples. The time series graphs of y_t in the second row bear a resemblance to those of unit root processes but nevertheless exhibit mean-reverting behavior. However, the mean reversion takes place slower than it would for a geometrically ergodic process. The autocorrelation functions also show very strong persistence—despite being mean-reverting, the processes can have a rather long memory. These features are related to the fact that the considered models are only polynomially ergodic with rate $r(n) = n^{4-\delta}$ for some $\delta > 0$ (this follows from Proposition 1 as the error terms used only have moments of order smaller than five).

We notice from Figure 1 that the periods when u_t takes large absolute values can be rather long and they can contain both increasing and decreasing periods for y_t. An example is the time series in the first column around $t \approx 800$ corresponding to the largest peak of the time series of y_t; similar features occur also in the time series in the second and third columns, often around peaks or troughs of the series. We also note from the second and third columns of Figure 1 that the time series of $I(u_t)$ stays close to 0 or 0.03, respectively, for some time (corresponding to behavior of u_t close to that of a unit root process without a drift or with a drift). An example is the rather long decreasing period in the time series of y_t in the third column, roughly between $t \approx 200$ and $t \approx 500$, that is to large extent due to unit root type behavior.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

5.2. Example with Time-Varying Slope Term of ESTAR Type

Next we consider example (23) with the time-varying slope term $S(u_{t-1})$ being either

$$S_1(u_{t-1}) = 1 - \frac{r_0}{h(u_{t-1})} \quad \text{or} \quad S_2(u_{t-1}) = \exp\Big\{-\frac{r_0}{h(u_{t-1})}\Big\}, \tag{26}$$

where $r_0 > 0$ and the positive-valued function h is such that $h(u)$ is large whenever u is large in absolute value (formal requirements for h are given in Proposition 2 below). Then the time-varying slope term $S(u_{t-1})$ takes values in some interval $[s, 1)$ $(s < 1)$ and attains values arbitrarily close to 1 for values of u_{t-1} large in absolute value. The shapes of $S_1(u)$ and $S_2(u)$ as functions of u resemble the "inverted bell curve form" commonly employed in exponential STAR (ESTAR) models (see, e.g., van Dijk et al., 2002). The main difference between the two functions in (26) is that $S_2(u)$ takes values in the unit interval $(0, 1)$ whereas $S_1(u)$ can also take negative values.

Before providing concrete examples, we state the following proposition which imposes conditions on the function h above to ensure that the results of Theorems 2 and 3 hold for model (23) with $S(\cdot)$ as in (26). The proof is straightforward and available in the Appendix.

PROPOSITION 2. *Consider the process y_t defined by $u_t - \nu = S(u_{t-1})(u_{t-1} - \nu) + \varepsilon_t$ as in (23) (with $u_t = \varpi(L)y_t$ and the roots of $\varpi(z)$ outside the unit circle) with $S(u_{t-1})$ being either $S_1(u_{t-1})$ or $S_2(u_{t-1})$ in (26) (with $r_0 > 0$) and with the function h therein satisfying*

(h) *$h : \mathbb{R} \to (0, \infty)$ is measurable, bounded on compact sets, satisfies $h(u) \to \infty$ as $|u| \to \infty$, and is such that $c_1 h(u) \le |u|^{\rho}$ and $|u|^{\rho+c_2} \le c_3 h^2(u)$ for $|u| \ge M_0$, some $c_1, c_2, c_3, M_0 > 0$, and $0 < \rho \le 2$.*

Assume further that either Assumption

(1) *2(a) is satisfied with $\kappa_0 < \rho$,*
(2) *2(a) is satisfied with $\kappa_0 = \rho$, or*
(3) *2(b) is satisfied with an s_0 such that one of the conditions (i)–(iii) of Theorem 3 holds.*

Then, under condition (1)/(2)/(3), the process $\mathbf{y}_t = (y_t, \ldots, y_{t-p+1})$ is either

(1) *subexponentially ergodic,*
(2) *geometrically ergodic, or*
(3) *polynomially ergodic.*

Moreover, condition (h) above is satisfied for i) $h(u) = 1 + |u - a|^{\rho}$, ii) $h(u) = (1 + |u - a|)^{\rho}$, iii) $h(u) = (1 + (u - a)^2)^{\rho/2}$, iv) $h(u) = 1 + |u - a_1|^{\rho_1} + |u - a_2|^{\rho_2}$, v) $h(u) = 1 + (1 + |u - a_1|)^{\rho_1} + (1 + |u - a_2|)^{\rho_2}$, or vi) $h(u) = 1 + (1 + (u - a_1)^2)^{\rho_1/2} + (1 + (u - a_2)^2)^{\rho_2/2}$ ($\rho, \rho_1, \rho_2 \in (0, 2]$; $a, a_1, a_2 \in \mathbb{R}$).

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

The obtained rate of ergodicity is again either geometric, subexponential, or polynomial depending on the moment assumptions made (the precise convergence rates can be obtained from Theorems 2 and 3; for existence of moments, see the corollaries to Theorems 2 and 3). The last part of the proposition lists several potential concrete choices for the function h, of which cases (iii) and (vi) may be convenient if differentiability of the function h is desired.

To illustrate the type of behavior processes covered by Proposition 2 may exhibit, consider as a simple example model (23) with $\nu = 0$, the slope term $S_1(u_{t-1})$ in (26), and $h(u) = 1 + |u|^\rho$ (cf. (5)). The considered model then becomes

$$u_t = \left(1 - \frac{r_0}{1 + |u_{t-1}|^\rho}\right) u_{t-1} + \varepsilon_t. \tag{27}$$

The shape of the slope coefficient $S_1(u_{t-1})$ as a function of u_{t-1} is now similar to that of an inverted bell curve increasing monotonically to unity as $|u_{t-1}|$ increases, and $S_1(u_{t-1})$ takes values within the interval $[1 - r_0, 1)$. Note that depending on the value of r_0, the slope coefficient takes values potentially only near the unity (say, in the interval $[0.95, 1)$) or even in rather extreme ranges (say, in the interval $[-100, 1)$) so that very different behaviors can be accommodated.

To explain why such processes can still be ergodic and stationary, note that for large values of $|u_{t-1}|$ the slope $S_1(u_{t-1})$ always takes values within $(-1, 1)$ (or a much smaller subset of it near unity). This prevents the process u_t from exploding and ensures mean-reverting behavior. Note that for larger values of ρ (within the permitted range $0 < \rho \leq 2$), the slope $S_1(u_{t-1})$ approaches unity faster as $|u_{t-1}|$ increases, intuitively corresponding to more "wandering" behavior of the observed series. This is reflected in Proposition 2 as slower rates of ergodicity being related to larger values of the parameter ρ (see also Theorems 2 and 3).

Figure 2 illustrates the preceding discussion. The top row depicts examples of the function $S(u)$ with some particular choices of the function h (see the caption of Figure 2 for the details). In the figures on the left and in the middle $S(u) = S_1(u)$ whereas in the figure on the right $S(u) = S_2(u)$. In each example, the shape of the function $S(u)$ is similar to that of an inverted bell curve increasing monotonically to unity as $|u|$ increases.

Rows 2–4 of Figure 2 present simulated time series of y_t, time series graphs of the time-varying slope coefficients $S(u_{t-1})$, and the autocorrelation functions of y_t, respectively, in these three examples. In the first and second columns, the time series graphs of y_t and the related autocorrelation functions indicate unit root type behavior and strong persistence; in these cases, also the slope coefficient $S(u_{t-1})$ takes values mostly larger than 0.90 and often very close to one. Compared to these two cases, the time series graph of y_t in the third column appears less "wandering" and this feature is also reflected in the related autocorrelation function which decays faster and in the slope coefficient $S(u_{t-1})$ which takes values further away from unity. Despite the three examples exhibiting somewhat different behaviors, all of them exhibit mean-reverting behavior and are subexponentially

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

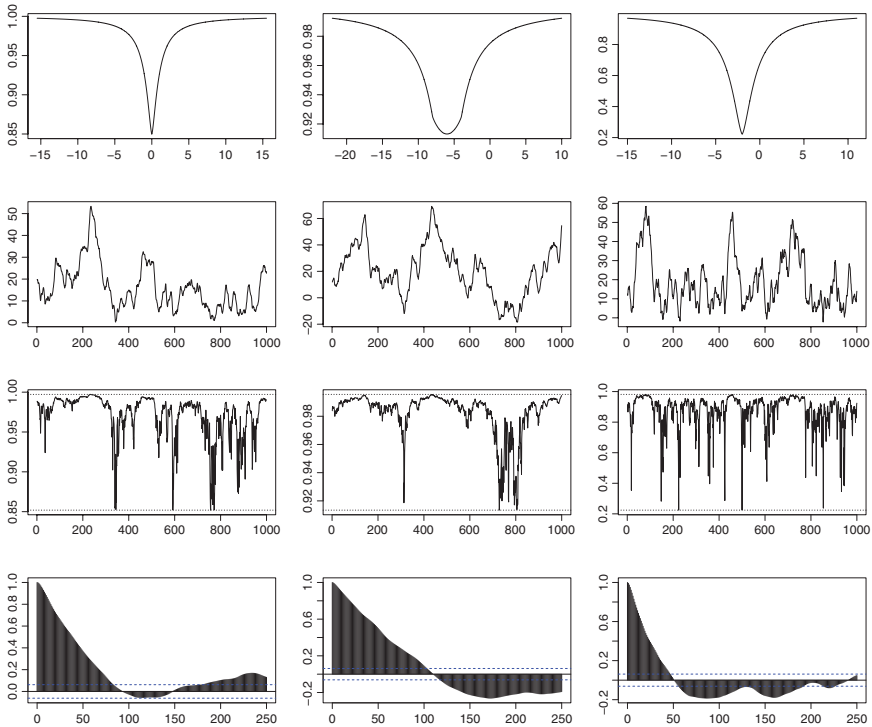


FIGURE 2. *Top row.* Graphs of $S(u)$ in (26). Left: $S_1(u)$ with $r_0 = 0.15$ and $h(u) = 1 + |u|^{1.5}$. Middle: $S_1(u)$ with $r_0 = 0.5$ and $h(u) = 1 + |u+4|^{1.25} + |u+8|^{1.25}$. Right: $S_2(u)$with $r_0 = 1.5$ and $h(u) = 1 + |u+2|^{1.5}$. *Second row.* Simulated time series of y_t corresponding to the time-varying slope functions $S(u_t)$ in the top row. The 1,000 observations are generated from model (23) with $p = 2$, $\pi_1 = 0.75$, $\nu = 2$, and Gaussian ε_t ($\varepsilon_t \sim N(0, 0.25)$ on the left; $\varepsilon_t \sim N(0, 0.75)$ in the middle; $\varepsilon_t \sim N(0, 1.5)$ on the right). *Third row.* The corresponding time series graphs of $S_1(u_{t-1}) = 1 - r_0/h(u_{t-1})$ on the left and in the middle, and $S_2(u_{t-1}) = \exp\{-r_0/h(u_{t-1})\}$ on the right. The dotted lines show the minimum and maximum. *Bottom row.* The corresponding autocorrelation functions of the simulated series. The first three autocorrelations are 0.997, 0.990, 0.981 (left); 0.996, 0.989, 0.980 (middle); and 0.992, 0.974, 0.950 (right).

ergodic (this is due to Proposition 2 because $\rho > 1$ and the error terms used are normally distributed and hence satisfy Assumption 2(a) with $\kappa_0 = 1$). The mean reversion again takes place slower than would be the case for geometrically ergodic processes.

5.3. Example of a More General Formulation

Finally, we briefly consider example (24) and for simplicity set $p = 2$. The time-varying slope term $S(u_{t-1})$ can be either one of the two options in (26). As for

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

$F(\mathbf{y}_{t-1})$, for concreteness we set $F(\mathbf{y}_{t-1}) = \exp\{-\gamma|\boldsymbol{y}_{t-1}|^2\}(\theta_1 y_{t-1} + \theta_2 y_{t-2})$, where $\boldsymbol{y}_{t-1} = (y_{t-1}, y_{t-2})$, $\gamma > 0$, and $\theta = (\theta_1, \theta_2)$ can take any values in $\mathbb{R}^2$. That is, the considered model reads as

$$y_t = \pi_1 y_{t-1} + S(u_{t-1})(y_{t-1} - \pi_1 y_{t-2}) + \exp\{-\gamma|\boldsymbol{y}_{t-1}|^2\}(\theta_1 y_{t-1} + \theta_2 y_{t-2}) + \varepsilon_t. \tag{28}$$

On the right hand side of (28) the term $\exp\{-\gamma|\boldsymbol{y}_{t-1}|^2\}$ has a bell-shaped form (as a function of $|\boldsymbol{y}_{t-1}|$) with maximum at the origin while for choices such as $h(u_{t-1}) = 1 + |u|^\rho$ the shape of $S(u_{t-1})$ is that of an inverted bell curve. Thus, given the shape of the terms $S(u_{t-1})$ and $\exp\{-\gamma|\boldsymbol{y}_{t-1}|^2\}$, model (28) can be viewed as a certain type of three-regime ESTAR model (see, e.g., van Dijk et al., 2002).

In this example, the function $\tilde{g}$ in equation (10) depends on the entire $\boldsymbol{y}_{t-1}$ and not only on u_{t-1} so that condition (ii.a) of Assumption 1 is no longer trivially satisfied. In the Appendix we show that model (28) nevertheless satisfies Assumption 1(ii) for both of the two options for the slope term $S(u_{t-1})$ (and with h as in Proposition 2). For checking condition (ii.a), the approach discussed after inequality (15) is used. The function $\tilde{\epsilon}(\boldsymbol{x})$ in this discussion can now be chosen to be $\tilde{\epsilon}(\boldsymbol{x}) = \exp\{-\gamma|\boldsymbol{x}|^2\}$ (for details, see the Appendix) so that the required condition $\tilde{\epsilon}(\boldsymbol{x}) = o(|\boldsymbol{x}|^{-d})$ holds with any positive d.

This preceding discussion also illustrates how the model in (28) could be modified while still satisfying condition (ii.a): instead of the term $\exp\{-\gamma|\boldsymbol{y}_{t-1}|^2\}$, any term being $o(|\boldsymbol{y}_{t-1}|^{-d})$ (with some $d > 0$) would do. (Note, however, that condition (ii.a) does not necessarily hold if we replace the norm $|\boldsymbol{y}_{t-1}|$ in (28) with a linear function of $\boldsymbol{y}_{t-1}$ such as u_{t-1}.) We also point out that Assumption 1(ii) does not rule out the possibility of setting $\pi_1 = 0$ (and $u_{t-1} = y_{t-1}$) in (28), and similarly for its higher-order counterparts where some or even all of the coefficients $\pi_1, \ldots, \pi_{p-1}$ may be equal to zero.

As a final remark, we note that allowing the parameters θ_1 and θ_2 in model (28) to be totally unrestricted highlights the fact that the autoregressions we consider may exhibit rather arbitrary behavior for moderate values of the observed series. As indicated in the Introduction, geometrically ergodic nonlinear autoregressions with features of this kind have previously been considered by Lu (1998), Gouriéroux and Robert (2006), Bec et al. (2008), and others. However, in most of these previous models stationarity is approached the further away the process moves from the origin whereas in our model unit root type behavior prevails for large absolute values of the process.

6. CONCLUSIONS

In this paper, we examined the subgeometric ergodicity of certain higher-order nonlinear autoregressive models. Generalizing existing first-order results, we provided conditions that ensure subexponential and polynomial ergodicity of the considered autoregressions. These results were established by utilizing suitably formulated drift conditions. Relying on results in a companion paper Meitz

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

and Saikkonen (2019), useful conclusions on the convergence rates of β-mixing coefficients were also obtained.

After obtaining theoretical results for rather general models we considered concrete examples and illustrated them with simulation. However, further work is needed to judge the usefulness of these models in practical applications. Several extensions could also be envisioned. For instance, subgeometric ergodicity of multivariate higher-order autoregressions or of models with conditional heteroskedasticity are interesting topics left for future work.

APPENDIX

This Appendix provides the technical details for the results in Section 5. Proofs of Theorems 1–3 and of Corollary to Theorem 3 are available in the Supplementary Appendix.

Proof of Proposition 1. We first verify that Assumption 1 holds with $\rho = 1$. Condition (i) holds by assumption. As $u_t = u_{t-1} + I(u_{t-1}) + \varepsilon_t$, the function $\tilde{g}$ in condition (ii) depends on u only and equals $I(u)$ (cf. equation (12)) so that condition (ii.a) holds with $g(u) = u + I(u)$ and $\epsilon(\boldsymbol{x}) = 0$ (the condition $|g(u)| \to \infty$ as $|u| \to \infty$ clearly holds). Thus it suffices to check condition (ii.b) with $g(u) = u + I(u)$ and $I(u)$ as in (25). Clearly $g(u)$ is bounded on bounded subsets so the latter part of (14) holds. We now show that the former part of (14) holds with $\rho = 1$, that is, that $|g(u)| \leq |u| - r$ for some $r, M_0 > 0$ and $|u| \geq M_0$. Note that the logistic function $L(u; b, a)$ satisfies $L(u; b, a) \to 0$ as $u \to -\infty$ and $L(u; b, a) \to 1$ as $u \to \infty$. As $I(u) = \nu_1 L(u; b, a_1) + \nu_2(1 - L(u; b, a_2))$ with $\nu_1 < 0 < \nu_2$, we can find an $M_0 > 0$ such that $I(u) > \nu_2/2 > 0$ for $u < -M_0$ and $I(u) < \nu_1/2 < 0$ for $u > M_0$. Setting $r = \min\{-\nu_1/2, \nu_2/2\}$ (and choosing a larger M_0 if necessary) we then have $|g(u)| = u + I(u) < u - (-\nu_1/2) \leq |u| - r$ for $u > M_0$ and $|g(u)| = -u - I(u) < -u - \nu_2/2 \leq |u| - r$ for $u < -M_0$, establishing the former part of (14). Thus Assumption 1 holds with $\rho = 1$.

To complete the proof, note that when condition (i) of the Proposition holds, Theorem 2(i) applies (with $b_3 = \kappa_0$) and process $\boldsymbol{y}_t$ is subexponentially ergodic with convergence rate $r(n) = (e^k)^{n^{\kappa_0}}$ (for some $k > 0$). Similarly, when condition (ii) of the Proposition holds, Theorem 2(ii) applies and process $\boldsymbol{y}_t$ is geometrically ergodic with convergence rate $r(n) = (e^c)^n$ (for some $c > 0$). Finally, when condition (iii) of the Proposition holds, Theorem 3 applies and process $\boldsymbol{y}_t$ is polynomially ergodic with convergence rate $r(n) = n^{s_0 - 1}$. □

Proof of Proposition 2. Assumption 1(i) again holds by assumption. For Assumption 1(ii), note that (23) can be written as $u_t = u_{t-1} + [1 - S(u_{t-1})]\nu + [S(u_{t-1}) - 1]u_{t-1} + \varepsilon_t$ (cf. equation (12)) and choosing $\tilde{g}(\boldsymbol{x}) = [1 - S(u)]\nu + [S(u) - 1]u$ and $g(u) = [1 - S(u)]\nu + S(u)u$ it is seen that Assumption 1(ii.a) holds with $\epsilon(\boldsymbol{x}) = 0$. For Assumption 1(ii.b), it again suffices to consider the former part of (14). The assumptions $c_1 h(u) \leq |u|^{\rho}$ and $|u|^{\rho + c_2} \leq c_3 h^2(u)$ made of function h imply that, for large enough $|u|$,

$$0 \leq 1 - \frac{r_0}{h(u)} \leq 1 - \frac{c_1 r_0}{|u|^{\rho}}, \quad \frac{1}{h(u)\,|u|} \leq \frac{c_3^{1/2}}{|u|^{1 - \rho/2 + c_2/2}\,|u|^{\rho}}, \quad \text{and}$$
$$0 < \frac{r_0^2}{2h^2(u)} \leq \frac{c_3 r_0^2}{2} \frac{1}{|u|^{\rho + c_2}}, \tag{A1}$$

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

where (in the middle inequality) $1-\rho/2+c_2/2>0$ as $\rho\le 2$ and $c_2>0$. Thus for the $S_1(u)$ in (26) and large enough $|u|$,

$$|g(u)|\le \frac{r_0\,|\nu|}{h(u)\,|u|}\,|u|+\left|1-\frac{r_0}{h(u)}\right||u|\le\left[1+\left(\frac{r_0\,|\nu|\,c_3^{1/2}}{|u|^{\,1-\rho/2+c_2/2}}-c_1r_0\right)\frac{1}{|u|^{\,\rho}}\right]|u|,$$

where the upper bound is dominated by $\left(1-\frac{c_1r_0/2}{|u|^{\rho}}\right)|u|$ for sufficiently large $|u|$ (so that Assumption 1(ii.b) holds with $r=c_1r_0/2$). For $S_2(u)=\exp\{-r_0/h(u)\}$ in (26), note that the inequality $1+x\le e^x\le 1+x+x^2/2$ (for $x\le 0$) yields $1-S_2(u)\le r_0/h(u)$ and $S_2(u)\le 1-r_0/h(u)+r_0^2/2h^2(u)$ (for all u). These inequalities, together with the fact that $0<S_2(u)<1$, imply that

$$|g(u)|\le|1-S_2(u)|\,|\nu|+|S_2(u)|\,|u|\le\frac{r_0\,|\nu|}{h(u)\,|u|}\,|u|+\left(1-\frac{r_0}{h(u)}+\frac{r_0^2}{2h^2(u)}\right)|u|.$$

Using the inequalities in (A1) implies that, for large enough $|u|$, the above upper bound is dominated by

$$\left[1+\left(\frac{r_0\,|\nu|\,c_3^{1/2}}{|u|^{\,1-\rho/2+c_2/2}}-c_1r_0+\frac{c_3r_0^2}{2\,|u|^{\,c_2}}\right)\frac{1}{|u|^{\,\rho}}\right]|u|$$

which in turn is again dominated by $\left(1-\frac{c_1r_0/2}{|u|^{\rho}}\right)|u|$ for sufficiently large $|u|$ (so that Assumption 1(ii.b) holds with $r=c_1r_0/2$).

The statements (1)–(3) now follow from Theorems 2(i), 2(ii), and 3, respectively.

Finally we show that condition (h) in the Proposition is satisfied for the six choices of $h(u)$ given. For case (i) note that, for $|u|>a$,

$$h(u)=\left(1+\frac{1}{|u-a|^{\rho}}\right)\cdot\left|1-\frac{a}{|u|}\right|^{\rho}\cdot|u|^{\rho},$$

where the product of the first two terms on the right hand side converges to one as $|u|\to\infty$, implying that, for some $c_1\in(0,1)$ and large enough $|u|$, $h(u)\le|u|^{\,\rho}/c_1$ or, equivalently, $c_1h(u)\le|u|^{\,\rho}$. Furthermore, let $c_2\in(0,\rho)$ and write $|u|^{2\rho}=|u|^{\rho-c_2}|u|^{\rho+c_2}$. Then the preceding discussion implies that, for some $c_3\in(0,1)$ and large enough $|u|$, $h^2(u)\ge|u|^{\,\rho+c_2}/c_3$ or, equivalently, $|u|^{\,\rho+c_2}\le c_3h^2(u)$. Thus, condition (h) holds in case (i).

Regarding cases (ii) and (iii), as $\rho\in(0,2]$, Loève's c_r-inequality shows that $(1+|u-a|)^{\rho}\le 2(1+|u-a|^{\rho})$, and the arguments used in case (i) above yield, for $|u|$ large enough, $(c_1/2)h(u)\le|u|^{\,\rho}$ with c_1 as in case (i). Similarly in case (iii), $(1+(u_{t-1}-a)^2)^{\rho/2}\le 1+|u_{t-1}-a|^{\rho}$ so that $c_1h(u)\le|u|^{\,\rho}$ holds with c_1 as in case (i). As for the inequality concerning $h^2(u)$, consider case (ii) and note that, for large enough $|u|$ and some $c_2\in(0,\rho)$,

$$h^2(u)=\left(\frac{1+|u-a|}{|u|}\right)^{2\rho}\cdot|u|^{\rho-c_2}\cdot|u|^{\rho+c_2},$$

where the product of the first two terms on the right hand side tends to infinity as $|u|\to\infty$. Thus, as in case (i) we obtain $|u|^{\,\rho+c_2}\le c_3h^2(u)$, and similar arguments can be used by replacing the definition of h in the above equality with the one in case (iii).

Now consider case (iv) and let $\rho=\rho_1\vee\rho_2$. We then have, for large enough $|u|$, $h(u)\le(1+|u-a_1|^{\rho})+(1+|u-a_2|^{\rho})$, and arguments used in case (i) above show that we can find $c_1\in(0,1)$ such that $c_1h(u)\le|u|^{\,\rho}$ holds. Verifying the desired inequality for $h^2(u)$ can

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

be established by using arguments used in case (ii) above. Cases (v) and (vi) can be handled with arguments similar to those already used; we omit the details. □

Checking Assumption 1(ii) for model (28). Suppose that $S(u_{t-1})$ is either one of the two options in (26) and that the function h satisfies condition (h) in Proposition 2. Clearly, we can write model (28) as $u_t = u_{t-1} + [S(u_{t-1}) - 1]u_{t-1} + \exp\{-\gamma|\boldsymbol{y}_{t-1}|^2\}(\theta_1 y_{t-1} + \theta_2 y_{t-2}) + \varepsilon_t$. Choosing $\tilde{g}(\boldsymbol{x}) = [S(u) - 1]u + \exp\{-\gamma|\boldsymbol{x}|^2\}\theta'\boldsymbol{x}$ ($\boldsymbol{x} = (x_1, x_2)$) and $g(u) = S(u)u$ yields $u + \tilde{g}(\boldsymbol{x}) - g(u) = \exp\{-\gamma|\boldsymbol{x}|^2\}\theta'\boldsymbol{x}$, implying that Assumption 1(ii.a) holds with any positive d (see inequality (15) and the following discussion). As $g(u) = S(u)u$, validity of Assumption 1(ii.b) can be checked as in the proof of Proposition 2 by setting the intercept term ν therein to zero. □

SUPPLEMENTARY MATERIAL

To view the online supplementary material for this article, please visit: http://doi.org/10.1017/S0266466620000419.

REFERENCES

Bec, F., A. Rahbek, & N. Shephard (2008) The ACR model: A multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics & Statistics* 70(5), 583–618.

Bradley, R. C. (2007) *Introduction to Strong Mixing Conditions*, vol. 1–3. Kendrick Press.

Douc, R., G. Fort, E. Moulines, & P. Soulier (2004) Practical drift conditions for subgeometric rates of convergence. *Annals of Applied Probability* 14(3), 1353–1377.

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

Doukhan, P. (1994) *Mixing: Properties and Examples*. Springer.

Fokianos, K., A. Rahbek, & D. Tjøstheim (2009) Poisson autoregression. *Journal of the American Statistical Association* 104, 1430–1439.

Fort, G. & E. Moulines (2003) Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and Their Applications* 103(1), 57–99.

Francq, C. & J.-M. Zakoïan (2006) Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory* 22(5), 815–834.

Gouriéroux, C. & C. Y. Robert (2006) Stochastic unit root models. *Econometric Theory* 22(6), 1052–1090.

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

Jarner, S. F. & R. L. Tweedie (2003) Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* 9(4), 559–578.

Klokov, S. A. (2007) Lower bounds of mixing rate for a class of Markov processes. *Theory of Probability and Its Applications* 51(3), 528–535.

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

Klokov, S. A. & A. Yu. Veretennikov (2005) On subexponential mixing rate for Markov processes. *Theory of Probability and Its Applications* 49(1), 110–122.

Lieberman, O. & P. C. B. Phillips (2020) Hybrid stochastic local unit roots. *Journal of Econometrics* 215(1), 257–285.

Ling, S. (2007) A double AR(p) model: Structure and estimation. *Statistica Sinica* 17(1), 161–175.

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

Meitz, M. & P. Saikkonen (2008) Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory* 24(5), 1291–1320.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press

Meitz, M. & P. Saikkonen (2019) Subgeometric ergodicity and β-mixing, arXiv:1904.07103.

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

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

Nummelin, E. & 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(3), 295–311.

Tanikawa, A. (2001) Markov chains satisfying simple drift conditions for subgeometric ergodicity. *Stochastic Models* 17(2), 109–120.

Tuominen, P. & R. L. Tweedie (1994) Subgeometric rates of convergence of f-ergodic Markov chains. *Advances in Applied Probability* 26(3), 775–798.

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

Veretennikov, A. Yu. (2000) On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory of Probability and Its Applications* 44(2), 361–374.

https://doi.org/10.1017/S0266466620000419 Published online by Cambridge University Press