


## SUBGEOMETRIC ERGODICITY AND $\beta$ -MIXING

MIKA MEITZ \* and  
PENTTI SAIKKONEN,\*\* *University of Helsinki*

### Abstract

It is well known that stationary geometrically ergodic Markov chains are  $\beta$ -mixing (absolutely regular) with geometrically decaying mixing coefficients. Furthermore, for initial distributions other than the stationary one, geometric ergodicity implies  $\beta$ -mixing under suitable moment assumptions. In this note we show that similar results hold also for subgeometrically ergodic Markov chains. In particular, for both stationary and other initial distributions, subgeometric ergodicity implies  $\beta$ -mixing with subgeometrically decaying mixing coefficients. Although this result is simple, it should prove very useful in obtaining rates of mixing in situations where geometric ergodicity cannot be established. To illustrate our results we derive new subgeometric ergodicity and  $\beta$ -mixing results for the self-exciting threshold autoregressive model.

*Keywords:* Markov chain; rate of convergence; mixing coefficient; subgeometric rate; subexponential rate; polynomial rate; SETAR model

2010 Mathematics Subject Classification: Primary 60J05  
Secondary 37A25

### 1. Introduction

Let  $X_t$  ( $t = 0, 1, 2, \dots$ ) be a Markov chain on the state space  $X$  with  $n$ -step transition probability measure  $P^n$  and stationary distribution  $\pi$ . If the  $n$ -step probability measures  $P^n$  converge in total variation norm to the stationary probability measure  $\pi$  at rate  $r^n$  (for some  $r > 1$ ), that is,

$$\lim_{n \rightarrow \infty} r^n \|P^n(x; \cdot) - \pi(\cdot)\| = 0, \quad \pi\text{-a.e.}, \quad (1)$$

the Markov chain is said to be geometrically ergodic. It is well known that for stationary Markov chains, geometric ergodicity implies that so-called  $\beta$ -mixing coefficients (or coefficients of absolute regularity)  $\beta(n)$ , to be defined formally in Section 2, converge to zero at the same rate,  $\lim_{n \rightarrow \infty} r^n \beta(n) = 0$  (see e.g. [8, page 89], [2, Theorem 3.7], or [3, Theorem 21.19]). For initial distributions other than the stationary one, a similar mixing result has been obtained by Liebscher [16, Proposition 4].

---

Received 16 April 2019; revision received 9 November 2020.

The supplementary material for this article can be found at <http://doi.org/10.1017/jpr.2020.108>.

\* Postal address: Department of Economics, University of Helsinki, PO Box 17, FI-00014 University of Helsinki, Finland. Email address: [mika.meitz@helsinki.fi](mailto:mika.meitz@helsinki.fi)

\*\* Postal address: Department of Mathematics and Statistics, University of Helsinki, PO Box 68, FI-00014 University of Helsinki, Finland. Email address: [pentti.saikkonen@helsinki.fi](mailto:pentti.saikkonen@helsinki.fi)

© The Author(s), 2021. Published by Cambridge University Press on behalf of Applied Probability Trust.

We are interested in counterparts of these mixing results when the convergence in (1) takes place at a rate  $r(n)$  slower than geometric, that is,

$$\lim_{n \rightarrow \infty} r(n) \|P^n(x; \cdot) - \pi(\cdot)\| = 0, \quad \pi\text{-a.e.} \quad (2)$$

When (2) holds with suitably defined rates  $r(n)$  slower than geometric, the Markov chain is called subgeometrically ergodic. The main result of this note establishes that for both stationary and other initial distributions, subgeometric ergodicity implies  $\beta$ -mixing with subgeometrically decaying mixing coefficients, that is,  $\lim_{n \rightarrow \infty} \tilde{r}(n)\beta(n) = 0$  for some rate function  $\tilde{r}(n)$ .

To illustrate some common rate functions, consider the expression

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

In the case  $\alpha, \beta, c, d > 0$  the four terms above satisfy  $e^{dn}/e^{cn^\gamma} \rightarrow \infty$ ,  $e^{cn^\gamma}/(1+n)^\beta \rightarrow \infty$ , and  $(1+n)^\beta/(1+\ln(n))^\alpha \rightarrow \infty$  as  $n \rightarrow \infty$ , and this hierarchy can be used to define different growth rates. Ordered from the fastest to the slowest growth rate, a growth rate is called geometric (sometimes also exponential) if the dominant term is  $e^{dn}$  (with  $d > 0$ ; note that  $e^{dn} = r^n$  with  $r > 1$  if and only if  $d > 0$ ), subexponential if the dominant term is  $e^{cn^\gamma}$  ( $c > 0$  and above  $d = 0$ ), polynomial if the dominant term is  $(1+n)^\beta$  ( $\beta > 0, c = d = 0$ ), and logarithmic if the dominant term is  $(1 + \ln(n))^\alpha$  ( $\alpha > 0, \beta = c = d = 0$ ).

We next provide some brief background on  $\beta$ -mixing and subgeometric ergodicity. The notion of  $\beta$ -mixing (or absolute regularity) was introduced by Volkonskii and Rozanov [29, 30], who attributed it to Kolmogorov. The surveys by Bradley [1, 2], the monograph by Doukhan [8], and the three-volume series by Bradley [3] specialized on (the various different concepts of) mixing and contain a wealth of further references. As for subgeometric ergodicity, the first subgeometric ergodicity results for general state space Markov chains were obtained by Nummelin and Tuominen [21] and Tweedie [24]; the subgeometric rate functions  $r(n)$  considered were introduced by Stone and Wainger [22]. Tuominen and Tweedie [23] gave a set of conditions that imply the convergence in (2) and, in particular, formulated a sequence of so-called drift conditions to establish subgeometric ergodicity. Subsequent work by Fort and Moulines [9, 10], Jarner and Roberts [11], and Douc *et al.* [6] led to a formulation of a single drift condition to ensure subgeometric ergodicity, paralleling the use of a Foster–Lyapunov drift condition to establish geometric ergodicity (see e.g. [19, Chapter 15]).

The rest of the paper proceeds as follows. Section 2 contains necessary mathematical preliminaries. Section 3 reviews the relation of geometric ergodicity and  $\beta$ -mixing, while the corresponding results in the subgeometric case are given in Section 4. The general results obtained are exemplified in Section 5, where subgeometric ergodicity and  $\beta$ -mixing results for the self-exciting threshold autoregressive model are presented. Section 6 concludes the paper and all proofs are given in an Appendix.

## 2. Preliminaries

To formalize the discussion in the Introduction, consider  $X_t$  ( $t = 0, 1, 2, \dots$ ), a time-homogeneous discrete-time Markov chain on a general measurable state space  $(X, \mathcal{B}(X))$ . Comprehensive treatments of the relevant Markov chain theory can be found in [19] or [7]. Let  $\mu$  be any initial measure on  $\mathcal{B}(X)$ , and suppose that  $X_0$  has distribution  $\mu$ . Denote the transition probabilities by  $P(x; A)$  ( $x \in X, A \in \mathcal{B}(X)$ ) and let  $(\Omega, \mathcal{F}, \mathbb{P}_\mu)$  denote the probability

space of the Markov process  $\{X_0, X_1, \dots\}$ . As usual,  $\mathbb{P}_x$  denotes the probability measure corresponding to a fixed initial value  $X_0 = x$  and  $P^n(x; A) = \mathbb{P}_x(X_n \in A)$  ( $x \in X, A \in \mathcal{B}(X)$ ) signifies the  $n$ -step transition probability measure.

Next consider the rate of convergence of the  $n$ -step probability measures  $P^n$  to the stationary probability measure  $\pi$ . To this end, for any two probability measures  $\lambda_1$  and  $\lambda_2$  on  $(X, \mathcal{B}(X))$ , the total variation distance is defined as

$$\|\lambda_1 - \lambda_2\| = 2 \sup_{B \in \mathcal{B}(X)} |\lambda_1(B) - \lambda_2(B)| = \sup_{|h| \leq 1} |\lambda_1(h) - \lambda_2(h)|,$$

where the last supremum runs over all  $\mathcal{B}(X)$ -measurable functions  $h: X \rightarrow \mathbb{R}$  bounded in absolute value by 1 and  $\lambda_i(h) = \int_X \lambda_i(dx)h(x) < \infty$ . The  $n$ -step probability measures  $P^n$  converge in total variation norm to the stationary probability measure  $\pi$  at rate  $r(n), n \geq 0$ , if

$$\lim_{n \rightarrow \infty} r(n) \|P^n(x; \cdot) - \pi(\cdot)\| = 0, \quad \pi\text{-a.e.} \tag{3}$$

If (3) holds, we say that the Markov chain  $X_t$  is ergodic with rate  $r(n)$ ; geometric ergodicity obtains when  $r(n) = r^n$  for some  $r > 1$ .

To define the  $\beta$ -mixing coefficients, let  $\mathcal{F}_k^l, 0 \leq k \leq l < \infty$ , signify the  $\sigma$ -algebra generated by  $\{X_k, \dots, X_l\}$ . For the stochastic process  $\{X_0, X_1, \dots\}$ , the  $\beta$ -mixing coefficients  $\beta(n), n = 1, 2, \dots$ , are defined by ([8, Section 1.1], [3, Chapter 3])

$$\begin{aligned} \beta(n) &= \frac{1}{2} \sup_{m \in \mathbb{N}} \sup \sum_{i=1}^I \sum_{j=1}^J |\mathbb{P}_\mu(A_i \cap B_j) - \mathbb{P}_\mu(A_i)\mathbb{P}_\mu(B_j)| \\ &= \sup_{m \in \mathbb{N}} \mathbb{E}_\mu \left[ \sup_{B \in \mathcal{F}_{n+m}^\infty} |\mathbb{P}_\mu(B | \mathcal{F}_0^m) - \mathbb{P}_\mu(B)| \right], \end{aligned}$$

where  $\mathbb{N} = \{0, 1, 2, \dots\}$ , and in the first expression for  $\beta(n)$  the second supremum is taken over all pairs of (finite) partitions  $\{A_1, A_2, \dots, A_I\}$  and  $\{B_1, B_2, \dots, B_J\}$  of  $\Omega$  such that  $A_i \in \mathcal{F}_0^m$  for each  $i$  and  $B_j \in \mathcal{F}_{n+m}^\infty$  for each  $j$ . For our purposes it is convenient to use the following alternative expression obtained by Davydov ([5, Proposition 1; note that his definition of  $\beta(n)$  includes an additional factor of 2]):

$$\beta(n) = \frac{1}{2} \sup_{m \in \mathbb{N}} \int_X \mu P^m(dx) \|P^n(x; \cdot) - \mu P^{n+m}(\cdot)\|, \quad n = 1, 2, \dots, \tag{4}$$

where  $\mu P^m(\cdot) = \int_X \mu(dx) P^m(x; \cdot)$  denotes the distribution of  $X_m$  ( $m = 1, 2, \dots; \mu P^0 = \mu$ ). For a stationary Markov chain (i.e. one with initial distribution  $\pi$ ), the  $\beta$ -mixing coefficients can be expressed simply as

$$\beta(n) = \frac{1}{2} \int_X \pi(dx) \|P^n(x; \cdot) - \pi(\cdot)\|, \quad n = 1, 2, \dots \tag{5}$$

Process  $X_t$  is said to be  $\beta$ -mixing (or sometimes absolutely regular) if  $\lim_{n \rightarrow \infty} \beta(n) = 0$ . As with the convergence in (3), the rate of this convergence is of interest, and in what follows we seek for results of the form  $\lim_{n \rightarrow \infty} r(n)\beta(n) = 0$  with some rate function  $r(n)$ .

### 3. The geometric case

We start by briefly discussing the relation of geometric ergodicity and  $\beta$ -mixing; although these results are well known, comparing them with the subgeometric case will be illuminating. For a stationary Markov chain (i.e. one with initial distribution  $\pi$ ), this relation is particularly simple. As was first shown by Nummelin and Tuominen [20, Theorem 2.1], a geometrically ergodic Markov chain satisfies, for some  $r > 1$ ,  $\lim_{n \rightarrow \infty} r^n \int \pi(dx) \|P^n(x; \cdot) - \pi(\cdot)\| = 0$ ; given expression (5), the  $\beta$ -mixing property immediately follows and the mixing coefficients satisfy  $\lim_{n \rightarrow \infty} r^n \beta(n) = 0$ . Statements of this result can be found for instance in [8, page 89], [2, Theorem 3.7], and [3, Theorem 21.19]. For initial distributions other than the stationary one, a corresponding result seems to have first appeared in [16, Proposition 4].

To facilitate comparison with the subgeometric case, we present the ergodicity and mixing results as consequences of a particular drift criterion; as is discussed in [19], this is how geometric ergodicity is often established. We use the following traditional Foster–Lyapunov-type geometric drift condition (see [19, Theorem 15.0.1]). Here  $\mathbf{1}_C(\cdot)$  signifies the indicator function of a set  $C$ . As a technical remark, note that in this condition we assume the function  $V$  to be everywhere finite (i.e.  $V: X \rightarrow [1, \infty)$ ) and such that  $\sup_{x \in C} V(x) < \infty$ . In contrast, in [19, Theorem 15.0.1] it is only assumed that  $V$  is extended-real-valued (i.e.  $V: X \rightarrow [1, \infty]$ ) and finite at some point  $x_0 \in X$ . Our stronger requirements hold in most practical applications and lead to more transparent exposition and proofs.

**Condition Drift–G.** Suppose there exist a petite set  $C$ , constants  $b < \infty$ ,  $\beta > 0$ , and a measurable function  $V: X \rightarrow [1, \infty)$  such that  $\sup_{x \in C} V(x) < \infty$ , satisfying

$$\mathbb{E}[V(X_1) | X_0 = x] \leq V(x) - \beta V(x) + b \mathbf{1}_C(x), \quad x \in X.$$

For the definition of a ‘petite set’ appearing in this condition, and for the concepts of irreducibility and aperiodicity in the theorem below, we refer the reader to [19]. Theorem 1 summarizes the relation between geometric ergodicity and  $\beta$ -mixing.

**Theorem 1.** *Suppose  $X_t$  is a  $\psi$ -irreducible and aperiodic Markov chain and that Condition Drift–G holds. Then*

- (a)  $X_t$  is geometrically ergodic, that is, for some  $r_1 > 1$ ,  $\lim_{n \rightarrow \infty} r_1^n \|P^n(x; \cdot) - \pi(\cdot)\| = 0$  for all  $x \in X$ .

Suppose further that the initial state  $X_0$  has distribution  $\mu$  such that  $\int_X \mu(dx) V(x) < \infty$ . Then

- (b) for some  $r_2 > 1$ ,  $\lim_{n \rightarrow \infty} r_2^n \int_X \mu(dx) \|P^n(x; \cdot) - \pi(\cdot)\| = 0$ ,
- (c)  $X_t$  is  $\beta$ -mixing and the mixing coefficients satisfy, for some  $r_3 > 1$ ,  $\lim_{n \rightarrow \infty} r_3^n \beta(n) = 0$ .

Moreover:

- (d) In the stationary case ( $\mu = \pi$ ) condition  $\int_X \pi(dx) V(x) < \infty$  is not needed, (b) and (c) hold with  $r_2 = r_3$ , and (b) and (c) are equivalent.

Parts (a) and (b) are very well known (see e.g. [19, Theorem 15.0.1] for part (a) and [20, Theorem 2.3] for part (b)) and so is also the mixing result in the stationary case (see the references given earlier). Part (c) for general initial distributions was obtained by Liebscher [16, Proposition 4], although our formulation is somewhat different from his (our formulation and proof avoid the use of so-called ‘ $Q$ -geometric ergodicity’ employed by Liebscher; for

completeness, our proof of Theorem 1, which may be of independent interest, is provided in the supplementary material). Part (d) elaborates parts (b) and (c) as well as their relation in the stationary case.

### 4. The subgeometric case

We seek a counterpart of Theorem 1 in which the geometric rate  $r^n$  is replaced by some slower rate function; such rate functions were already exemplified in the Introduction. More formally, the subgeometric rate functions we consider are defined as follows (see e.g. [21] and [6]). Let  $\Lambda_0$  be the set of positive non-decreasing functions  $r_0: \mathbb{N} \rightarrow [1, \infty)$  such that  $\ln[r_0(n)]/n$  decreases to zero as  $n \rightarrow \infty$ . The class of subgeometric rate functions, denoted by  $\Lambda$ , consists of positive functions  $r: \mathbb{N} \rightarrow (0, \infty)$  for which there exists some  $r_0 \in \Lambda_0$  such that

$$0 < \liminf_{n \rightarrow \infty} \frac{r(n)}{r_0(n)} \leq \limsup_{n \rightarrow \infty} \frac{r(n)}{r_0(n)} < \infty. \tag{6}$$

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  $\alpha > 0$ .

In analogy with the geometric case, subgeometric ergodicity and mixing results are most conveniently obtained by verifying an appropriate drift condition. The following drift condition for subgeometric ergodicity is adapted from [7, Definition 16.1.7]. A somewhat more general drift condition, for instance allowing for  $V$  to be extended-real-valued, is given in [6].

**Condition Drift–SubG.** Suppose there exist a petite set  $C$ , a constant  $b < \infty$ , a concave increasing continuously differentiable function  $\phi: [1, \infty) \rightarrow (0, \infty)$  satisfying  $\lim_{v \rightarrow \infty} \phi'(v) = 0$ , and a measurable function  $V: X \rightarrow [1, \infty)$  such that  $\sup_{x \in C} V(x) < \infty$  and

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

Note that if  $\phi(v) = \eta v$  ( $\eta > 0$ ), then one obtains Condition Drift–G (but assumption  $\lim_{v \rightarrow \infty} \phi'(v) = 0$  rules this out; as we are interested in subgeometric rates of ergodicity, assuming this means no loss of generality; see [7, Remark 16.1.8]).

Following Douc *et al.* [6] we next introduce a rate function, denoted by  $r_\phi$ . First define the function

$$H_\phi(v) = \int_1^v \frac{dx}{\phi(x)},$$

where  $\phi$  is as in Condition Drift–SubG. The definition implies that  $H_\phi$  is a non-decreasing, concave, and differentiable function on  $[1, \infty)$ , and it has an inverse  $H_\phi^{-1}: [0, \infty) \rightarrow [1, \infty)$  which is increasing and differentiable (see [6, Section 2.1]). Thus we can define the rate function

$$r_\phi(z) = (H_\phi^{-1})'(z) = \phi \circ H_\phi^{-1}(z).$$

Douc *et al.* [6, Lemma 2.3 and Proposition 2.5] showed that this rate function is subgeometric and that Condition Drift–SubG implies the convergence (3) at rate  $r_\phi(n)$ .

Theorem 2 summarizes the relation between subgeometric ergodicity and  $\beta$ -mixing. Here  $[k]$  denotes the integer part of the real number  $k$ .

**Theorem 2.** *Suppose  $X_t$  is a  $\psi$ -irreducible and aperiodic Markov chain and that Condition Drift–SubG holds. Then*

- (a)  $X_t$  is subgeometrically ergodic with rate  $r_\phi(n)$ , that is,  $\lim_{n \rightarrow \infty} r_\phi(n) \|P^n(x; \cdot) - \pi(\cdot)\| = 0$  for all  $x \in \mathcal{X}$ .

Suppose further that the initial state  $X_0$  has distribution  $\mu$  such that  $\int_{\mathcal{X}} \mu(dx)V(x) < \infty$ . Then

- (b)  $\lim_{n \rightarrow \infty} r_\phi(n) \int \mu(dx) \|P^n(x; \cdot) - \pi(\cdot)\| = 0$ ,
- (c)  $X_t$  is  $\beta$ -mixing and the mixing coefficients satisfy  $\lim_{n \rightarrow \infty} \tilde{r}_\phi(n)\beta(n) = 0$  for any rate function  $\tilde{r}_\phi(n)$  such that  $\limsup_{n \rightarrow \infty} \tilde{r}_\phi(n)/r_\phi(n_1) < \infty$  where  $n_1 = [n/2]$ .

Moreover:

- (d) In the stationary case ( $\mu = \pi$ ) condition  $\int_{\mathcal{X}} \pi(dx)V(x) < \infty$  is not needed, (b) and (c) hold with  $r_\phi(n) = \tilde{r}_\phi(n)$ , and (b) and (c) (with  $r_\phi(n) = \tilde{r}_\phi(n)$ ) are equivalent.
- (e) If  $r_\phi(n)$  satisfies (6) with  $r_{\phi,0}(n) = (1 + \ln(n))^\alpha \cdot (1 + n)^\beta \cdot e^{cn^\gamma}$  and  $\tilde{r}_\phi(n)$  satisfies (6) with  $\tilde{r}_{\phi,0}(n) = (1 + \ln(n))^\alpha \cdot (1 + n)^\beta \cdot e^{\tilde{c}n^\gamma}$  for some  $0 < \tilde{c} < c2^{-\gamma}$ , then  $\limsup_{n \rightarrow \infty} \tilde{r}_\phi(n)/r_\phi(n_1) < \infty$ .

Of the results in Theorem 2, part (a) is given in Proposition 2.5 of [6]. Part (b) can be obtained by combining Theorem 4.1 of [23] and Proposition 2.5 of [6], but in the proof we make use of [21]. Part (c) is new and illuminates the relation between subgeometrically ergodic Markov chains and their  $\beta$ -mixing properties, thereby providing a counterpart of a result obtained by Liebscher [16, Proposition 4] in the case of geometric ergodicity. Part (d) is analogous to its counterpart in Theorem 1 and provides further insight into parts (b) and (c), whereas part (e) makes part (c) more concrete in the case of the most common rate functions. For completeness, we give a detailed proof in the Appendix.

As discussed in [6, Section 2.3] and [17, Theorem 1], there is a connection between the function  $\phi$  and the rate function  $r_\phi$ , which can be used to find out the latter in particular cases. For instance, polynomial rate functions are associated with cases where the function  $\phi$  is of the form  $\phi(v) = cv^\alpha$  with  $\alpha \in (0, 1)$  and  $c \in (0, 1]$ , and then the rate obtained is  $r_\phi(n) = n^{\alpha/(1-\alpha)}$  (an alternative form is  $r_\phi(n) = n^{\kappa-1}$  with  $\kappa = 1 + \alpha/(1 - \alpha)$  already given by Järner and Roberts [11]). In the subexponential case the function  $\phi$  is such that  $v/\phi(v)$  goes to infinity slower than polynomially so that a possibility, given in [17, Theorem 1], is  $\phi(v) = c(v + v_0)/[\ln(v + v_0)]^\alpha$  for some  $\alpha, c, v_0 > 0$ . This results in the rate  $r_\phi(n) = (e^d)^{n^{1/(1+\alpha)}}$  for some  $d > 0$ , which is faster than polynomial. A logarithmic rate is an example of a rate slower than polynomial. Then the function  $\phi$  is of the form  $\phi(v) = c[1 + \ln(v)]^\alpha$  for some  $\alpha > 0$  and  $c \in (0, 1]$ , and the resulting rate is  $r_\phi(n) = [\ln(n)]^\alpha$  (see [6, Section 2.3]).

Theorem 2 (or 1) also provides information about the moments of the stationary distribution of  $X_t$ . Specifically, once part (a) of Theorem 2 (or 1) has been established, one can deduce from Condition Drift–SubG (or Drift–G) and Theorem 14.3.7 of [19] that  $\int_{\mathcal{X}} \pi(dx)\phi(V(x)) < \infty$  (or  $\int_{\mathcal{X}} \pi(dx)V(x) < \infty$ ). This can be very useful when one aims to apply limit theorems developed for  $\beta$ -mixing processes, where moment conditions are typically assumed.

We close this section by noting that Condition Drift–SubG can also be used to obtain more general ergodicity results than provided in Theorem 2. Without going into details, we only mention that Theorem 2.8 of [6] and Theorem 1 of [17] show how a stronger form of ergodicity, called  $(f,r)$ -ergodicity, can be established.

### 5. Example

To illustrate our results we consider the self-exciting threshold autoregressive (SETAR) model studied by Chan *et al.* [4]. These authors analyzed the model

$$X_t = \varphi(j) + \theta(j)X_{t-1} + W_t(j), \quad X_{t-1} \in (r_{j-1}, r_j], \quad (7)$$

where  $-\infty = r_0 < \dots < r_M = \infty$ , and for each  $j = 1, \dots, M$ ,  $\{W_t(j)\}$  is an independent and identically distributed mean zero sequence independent of  $\{W_t(i)\}$ ,  $i \neq j$ , and with  $W_t(j)$  having a density that is positive on the whole real line. They considered the following conditions:

$$\theta(1) < 1, \quad \theta(M) < 1, \quad \theta(1)\theta(M) < 1, \quad (8a)$$

$$\theta(1) = 1, \quad \theta(M) < 1, \quad 0 < \varphi(1), \quad (8b)$$

$$\theta(1) < 1, \quad \theta(M) = 1, \quad \varphi(M) < 0, \quad (8c)$$

$$\theta(1) = 1, \quad \theta(M) = 1, \quad \varphi(M) < 0 < \varphi(1), \quad (8d)$$

$$\theta(1) < 0, \quad \theta(1)\theta(M) = 1, \quad \varphi(M) + \varphi(1)\theta(M) > 0, \quad (8e)$$

and showed that the SETAR model is ergodic if and only if one of the conditions (8a)–(8e) holds [4, Theorem 2.1]. Moreover, if  $\mathbb{E}[|W_t(j)|] < \infty$  for each  $j$ , they showed that condition (8a) ensures geometric ergodicity [4, Theorem 2.3]. To our knowledge, in the cases (8b)–(8e) no results regarding the rate of ergodicity have as yet appeared in the literature and our Theorem 4(b) below indicates that geometric ergodicity may not always hold without stronger assumptions. Related to this, we note that Meyn and Tweedie [19, Section 11.4.3 and Section B.2] discussed the (geometric) ergodicity of the SETAR model (7), reproducing the ergodicity result of [4, Theorem 2.1] as their Proposition 11.4.5. On their page 541, Meyn and Tweedie [19] also stated that (our additions in brackets) ‘in the interior of the parameter space [the union of (8a)–(8e)] we are able to identify geometric ergodicity in Proposition 11.4.5 . . . the stronger form [geometric ergodicity] is actually proved in that result’ but no formal proof is given for this statement.

We consider rates of ergodicity and  $\beta$ -mixing in case (8d) when the autoregressive coefficients  $\theta(1)$  and  $\theta(M)$  equal unity. For intuition, note that due to non-zero intercept terms  $\varphi(1)$  and  $\varphi(M)$ , both the first and the last regimes exhibit non-stationary random-walk-type behavior with a drift. As the intercept terms satisfy  $\varphi(M) < 0 < \varphi(1)$ , the drift is increasing in the first regime and decreasing in the last regime. This feature prevents the process  $y_t$  from exploding to (plus or minus) infinity, thereby providing intuition why ergodicity can hold true. It is noteworthy that ergodicity is in no way dependent on the behavior of the process in the middle regimes  $(2, \dots, M-1)$ , which can exhibit stationary, random-walk-type (with or without drift), or even explosive behavior.

In their results, Chan *et al.* [4] allowed for regime-dependent distributions for the error term  $W_t(j)$ . To obtain our results for the case (8d), we strengthen the assumptions on the error term

and, in particular, assume that the error distribution is the same in each regime (this stronger assumption is needed to apply the results mentioned in the proof of Theorem 3 below, and relaxing it appears less than straightforward). To compensate, we obtain results for a model more general than the SETAR model (7) with (8d). Specifically, we formulate our results in terms of the general non-linear autoregressive model

$$X_t = g(X_{t-1}) + \varepsilon_t, \quad t = 1, 2, \dots, \tag{9}$$

where the function  $g: \mathbb{R} \rightarrow \mathbb{R}$  and the error term  $\varepsilon_t$  satisfy the following conditions:

- (A1)  $g$  is a measurable function with the property  $|g(x)| \rightarrow \infty$  as  $|x| \rightarrow \infty$  and such that there exist positive constants  $r$  and  $M_0$  such that

$$|g(x)| \leq (1 - r/|x|)|x| \quad \text{for } |x| \geq M_0 \text{ and } \sup_{|x| \leq M_0} |g(x)| < \infty;$$

- (A2)  $\{\varepsilon_t, t = 1, 2, \dots\}$  is a sequence of independent and identically distributed mean zero random variables that is independent of  $X_0$  and the distribution of  $\varepsilon_1$  has a (Lebesgue) density that is bounded away from zero on compact subsets of  $\mathbb{R}$ .

Model (9) with conditions (A1) and (A2) is a special case of models considered by Fort and Moulines [10, Section 2.2], Douc *et al.* [6, Section 3.3], and Meitz and Saikkonen [17, Sections 3–4]. These authors considered much more general models, but for clarity of presentation we have simplified the model as much as possible while still being able to obtain results for the SETAR model (7) with (8d) (papers [10] and [6] considered a multivariate version of (9) whereas [17] considered a higher-order generalization of (9); the inequality constraint for the function  $g$  in condition (A1) is also more general in these papers, where it is only required that  $|g(x)| \leq (1 - r|x|^{-\rho})|x|$  for some  $0 < \rho \leq 2$ ).

The following theorem establishes ergodicity and  $\beta$ -mixing results for model (9) with varying rates of convergence. The proof (in the Appendix) makes use of results in [10], [6], and [17] to obtain rates of ergodicity, as well as Theorems 1 and 2 above to obtain rates of  $\beta$ -mixing (only the subgeometric mixing results in parts (b) and (c) are new).

**Theorem 3.** *Consider model (9) with conditions (A1) and (A2).*

- (a) *If  $\mathbb{E}[e^{z_0|\varepsilon_1|}] < \infty$  for some  $z_0 > 0$ , then  $X_t$  is geometrically ergodic with convergence rate  $r(n) = r_1^n$  for some  $r_1 > 1$ . Moreover, if the initial state  $X_0$  has a distribution such that  $\mathbb{E}[e^{z|X_0|}] < \infty$  for some  $z > 0$ , then  $X_t$  is also  $\beta$ -mixing and the mixing coefficients satisfy, for some  $r_3 > 1$ ,  $\lim_{n \rightarrow \infty} r_3^n \beta(n) = 0$ .*
- (b) *If  $\mathbb{E}[e^{z_0|\varepsilon_1|^{\kappa_0}}] < \infty$  for some  $z_0 > 0$  and  $\kappa_0 \in (0, 1)$ , then  $X_t$  is subexponentially ergodic with convergence rate  $r(n) = (e^c)^{n^{\kappa_0}}$  (for some  $c > 0$ ). Moreover, if the initial state  $X_0$  has a distribution such that  $\mathbb{E}[e^{z|X_0|^{\kappa_0}}] < \infty$  for some  $z > 0$ , then  $X_t$  is also  $\beta$ -mixing and the mixing coefficients satisfy, for some  $\tilde{c} > 0$ ,  $\lim_{n \rightarrow \infty} (e^{\tilde{c}})^{n^{\kappa_0}} \beta(n) = 0$ .*
- (c) *If  $\mathbb{E}[|\varepsilon_1|^{s_0}] < \infty$  for either  $s_0 = 2$  or  $s_0 \geq 4$ , then  $X_t$  is polynomially ergodic with convergence rate  $r(n) = n^{s_0-1}$ . Moreover, if the initial state  $X_0$  has a distribution such that  $\mathbb{E}[|X_0|^{s_0}] < \infty$ , then  $X_t$  is also  $\beta$ -mixing and the mixing coefficients satisfy  $\lim_{n \rightarrow \infty} n^{s_0-1} \beta(n) = 0$ .*

Theorem 3 shows that there is a trade-off between rates of ergodicity and  $\beta$ -mixing and finiteness of moments of the error term. The fastest geometric rate is obtained when  $\mathbb{E}[e^{z_0|\varepsilon_1|}] < \infty$  ( $z_0 > 0$ ), so that  $\varepsilon_1$  has finite moments of all orders and the slowest polynomial



rate is obtained when only  $\mathbb{E}[\varepsilon_1^2] < \infty$ . As discussed after Theorem 2, we also have  $\int_{\mathcal{X}} \pi(dx)\phi(V(x)) < \infty$  so that there is a similar trade-off between these convergence rates and finiteness of moments of the stationary distribution (expressions of  $V$  and  $\phi$  are available in the proof of Theorem 3).

Above it was mentioned that [10], [6], and [17] considered (subgeometric) ergodicity of models more general than (9) with conditions (A1) and (A2). Making use of our Theorems 1 and 2, subgeometric rates of  $\beta$ -mixing can easily be obtained for these more general models too. We omit the details for brevity.

In a series of papers, Veretennikov and co-authors also considered the model (9) with function  $g$  satisfying  $|g(x)| \leq (1 - r|x|^{-\rho})|x|$  for some  $1 \leq \rho \leq 2$ . Using methods very different from ours, they obtained results on subgeometric ergodicity and subgeometric rates for  $\beta$ -mixing coefficients. The cases  $1 < \rho < 2$  and  $\rho = 2$  are considered in [27], [14, 15], and [13] and are shown to lead to subgeometric rates. For the case  $\rho = 1$  relevant for the SETAR example, these papers refer to [25, 26] and [28]. A result corresponding to our Theorem 3(a) can be found in [28, Theorem 1], but subgeometric rates, such as those in our Theorem 3 parts (b) and (c), do not seem to be established in the case  $\rho = 1$ .

We now specialize the results above to the SETAR model (7) with (8d). It is easy to see that this model, with the function  $g$  in (9) defined as

$$g(x) = \sum_{j=1}^M [\varphi(j) + \theta(j)x] \mathbf{1}\{x \in (r_{j-1}, r_j]\}$$

(with  $\mathbf{1}\{\cdot\}$  denoting the indicator function), satisfies the condition in (A1). Namely, for  $x$  large enough and positive we have  $|g(x)| = g(x) = x + \varphi(M) = |x| - (-\varphi(M))$ , whereas for  $x$  small enough and negative we have  $|g(x)| = -g(x) = -x - \varphi(1) = |x| - \varphi(1)$ , so that the inequality in (A1) holds for  $M_0 > \max\{|r_1|, |r_{M-1}|\}$  and  $r = \min\{\varphi(1), -\varphi(M)\}$  (and the supremum condition is obviously satisfied).

Part (a) of the next theorem simply restates the result of Theorem 3 for the SETAR model (7) with (8d), whereas part (b) establishes that geometric ergodicity cannot hold under the weaker moment assumptions of Theorem 3 parts (b) and (c).

**Theorem 4.** *Consider the SETAR model (7) with the parameters satisfying (8d) and the error terms satisfying  $W_t(j) = \varepsilon_t$  ( $j = 1, \dots, M$ ) with  $\varepsilon_t$  as in (A2).*

- (a) *Sufficient conditions for geometric, subexponential, and polynomial ergodicity and  $\beta$ -mixing of  $X_t$  are as in parts (a), (b), and (c) of Theorem 3, respectively.*
- (b) *If  $\mathbb{E}[e^{z_0|\varepsilon_1|}] = \infty$  for all  $z_0 > 0$ , then  $X_t$  is not geometrically ergodic.*

Theorem 4(b) shows that for the SETAR model (7) with (8d), the subgeometric rates of Theorem 3 parts (b) and (c) cannot be improved to a geometric rate unless stronger moment assumptions are made regarding the error term. This result is obtained by making use of a necessary condition for geometric ergodicity of certain specific type of Markov chains in [12] (using their necessary condition to obtain this result appears possible only in case (8d) out of (8a)–(8e)).

## 6. Conclusion

In this note we have shown that subgeometrically ergodic Markov chains are  $\beta$ -mixing with subgeometrically decaying mixing coefficients. Although this result is simple, it should

prove very useful in obtaining rates of mixing in situations where geometric ergodicity cannot be established. An illustration using the popular self-exciting threshold autoregressive model showed how our results can yield new subgeometric rates of mixing.

### Appendix A. Proofs

This Appendix contains the proofs of Theorems 2–4; proof of Theorem 1 is provided in the supplementary material. Proofs of Theorems 1 and 2 make use of the following handy inequality for the  $\beta$ -mixing coefficients due to Liescher [16, Proposition 3]. (Note that our Lemma A.1 below includes an additional factor of  $\frac{1}{2}$  compared to Liescher’s Proposition 3; see our expression for  $\beta(n)$  in (4) and his equation (27).) Again,  $[k]$  denotes the integer part of the real number  $k$ .

**Lemma A.1.** *Suppose  $X_t$  is a Markov chain with stationary distribution  $\pi$  and that the initial state  $X_0$  has distribution  $\mu$ . Then*

$$\beta(n) \leq \frac{1}{2} \int \pi(dx) \|P^{n_1}(x; \cdot) - \pi\| + \frac{3}{2} \int \mu(dx) \|P^{n_1}(x; \cdot) - \pi\|, \quad n = 1, 2, \dots,$$

where  $n_1 = \lfloor n/2 \rfloor$ .

In the proof below, notation  $\mathbb{E}_\mu[\cdot]$  is used for the conditional expectation of a  $\mathcal{F}_0^\infty$ -measurable random variable conditioned on the initial state  $X_0$  with distribution  $\mu$ . When conditioning is on  $X_0 = x$  the notation  $\mathbb{E}_x[\cdot]$  is used; these are connected via  $\mathbb{E}_\mu[\cdot] = \int \mu(dx) \mathbb{E}_x[\cdot]$ . We also define the concept of return time to a measurable set  $A$  as  $\tau_A = \inf\{n \geq 1 : X_n \in A\}$ .

*Proof of Theorem 2.* First note that, due to the assumed irreducibility and aperiodicity, the petite set  $C$  in Condition Drift–SubG is small [19, Theorem 5.5.7]. We first show that

$$\sup_{x \in C} \mathbb{E}_x \left[ \sum_{k=0}^{\tau_C-1} r_\phi(k) \right] < \infty; \tag{10}$$

by Theorem 2.1 of Tuominen and Tweedie [23] this implies the subgeometric ergodicity in (a) (for related results implying (a); see also [21, Theorem 2.7(i)], [24, Theorem 1(iii)], [6, Proposition 2.5]). Douc *et al.* [6, Proposition 2.1 and Lemma 2.3] showed that Condition Drift–SubG implies the existence of a sequence of drift functions  $V_k(x)$ ,  $k = 0, 1, 2, \dots$ , such that, for  $k \geq 0$ ,

$$\mathbb{E}[V_{k+1}(X_1) | X_0 = x] \leq V_k(x) - r_\phi(k) + \tilde{b}r_\phi(k)\mathbf{1}_C(x),$$

where  $\tilde{b} = br_\phi(1)(r_\phi(0))^{-2}$  (see their Proposition 2.1 and the top of page 1358) and  $r_\phi \in \Lambda$  (see their Lemma 2.3). Applying Proposition 11.3.2 of [19] with  $Z_k = V_k(X_k)$ ,  $f_k(x) = r_\phi(k)$ ,  $s_k(x) = \tilde{b}r_\phi(k)\mathbf{1}_C(x)$ , and stopping time  $\tau_C$ , we obtain ([6, Proposition 2.2] also states this conclusion; note also that by their equation (2.2) we have  $V_0(x) \leq V(x)$ )

$$\begin{aligned} \mathbb{E}_x \left[ \sum_{k=0}^{\tau_C-1} r_\phi(k) \right] &\leq V(x) + \mathbb{E}_x \left[ \sum_{k=0}^{\tau_C-1} \tilde{b}r_\phi(k)\mathbf{1}_C(x) \right] \\ &= V(x) + \tilde{b}r_\phi(0)\mathbf{1}_C(x) \\ &= V(x) + b \frac{r_\phi(1)}{r_\phi(0)} \mathbf{1}_C(x). \end{aligned} \tag{11}$$

By condition  $\sup_{x \in C} V(x) < \infty$  (in Condition Drift–SubG), we obtain (10). Now Theorem 2.1 of [23] ensures that  $\lim_{n \rightarrow \infty} r_\phi(n) \|P^n(x; \cdot) - \pi(\cdot)\| = 0$ , so the subgeometric ergodicity in (a) is established (note that as  $V_0(x) \leq V(x)$  holds with  $V(x)$  assumed everywhere finite, the set  $S(f, r)$  in Theorem 2.1 of [23] coincides with  $X$ , so the aforementioned convergence holds for all  $x \in X$ ).

To prove (b), suppose the initial state  $X_0$  has distribution  $\mu$  such that  $\int_X \mu(dx) V(x) < \infty$ . We will use Theorems 2.7(i,ii) and 2.2 of [21], but first we obtain a property of the rate function  $r_\phi(z)$  (which is well known for members of  $\Lambda_0$  but not for members of  $\Lambda$ ). Recall that  $r_\phi(z) = (H_\phi^{-1})'(z) = \phi \circ H_\phi^{-1}(z)$  so that  $r_\phi'(z)/r_\phi(z) = \phi' \circ H_\phi^{-1}(z)$ . As  $\phi'$  is non-increasing (see [6, first paragraph of Section 2.1]) and  $H_\phi^{-1}$  is increasing, it follows that  $r_\phi'(z)/r_\phi(z) = \phi' \circ H_\phi^{-1}(z)$  is non-increasing. Therefore the function

$$\ln(r_\phi(x))/x = \frac{1}{x} \int_0^x (r_\phi'(s)/r_\phi(s)) ds \quad (x > 0)$$

is also non-increasing. Following the proof of Lemma 1 of Stone and Wainger [22] (which relies only on their property (iii) on their page 326) yields the desired property  $r_\phi(m+n) \leq r_\phi(m)r_\phi(n)$  for all  $m, n > 0$ .

Using this property we now obtain

$$r_\phi(\tau_C) \leq r_\phi(1)r_\phi(\tau_C - 1) \leq r_\phi(1) \sum_{k=0}^{\tau_C-1} r_\phi(k),$$

and further that

$$\mathbb{E}_x \left[ \sum_{k=0}^{\tau_C} r_\phi(k) \right] \leq (r_\phi(1) + 1) \mathbb{E}_x \left[ \sum_{k=0}^{\tau_C-1} r_\phi(k) \right] \quad \text{and} \quad \mathbb{E}_x[r_\phi(\tau_C)] \leq r_\phi(1) \mathbb{E}_x \left[ \sum_{k=0}^{\tau_C-1} r_\phi(k) \right]$$

(see [23, equations (5) and (14)]). The former result together with (10) implies that condition (2.12) of Theorem 2.7(i) of [21] is satisfied. The latter result together with (11) yields

$$\mathbb{E}_x[r_\phi(\tau_C)] \leq r_\phi(1) [V(x) + b \frac{r_\phi(1)}{r_\phi(0)} \mathbf{1}_C(x)]$$

and, as  $\mathbb{E}_\mu[r_\phi(\tau_C)] = \int_X \mu(dx) \mathbb{E}_x[r_\phi(\tau_C)]$ , the assumed bound  $\int_X \mu(dx) V(x) < \infty$  implies

$$\mathbb{E}_\mu[r_\phi(\tau_C)] < \infty,$$

so the condition in Theorem 2.7(ii) of [21] is satisfied. Therefore, by Theorems 2.7(i,ii) and 2.2 of [21],

$$\lim_{n \rightarrow \infty} r_\phi(n) \int \mu(dx) \|P^n(x; \cdot) - \pi(\cdot)\| = 0.$$

Next consider part (d). In the stationary case ( $\mu = \pi$ ), the desired result

$$\lim_{n \rightarrow \infty} r_\phi(n) \int \pi(dx) \|P^n(x; \cdot) - \pi(\cdot)\| = 0$$

follows from the last remark in Theorem 2.2 of [21] (and condition  $\int_X \pi(dx) V(x) < \infty$  is not needed). Thus (b) holds in the stationary case. Regarding part (c) in the stationary case, note from (5) that now  $\beta(n) = \int \pi(dx) \|P^n(x; \cdot) - \pi\|$ ,  $n = 1, 2, \dots$ , so (b) and (c) are clearly equivalent (and hold with the same rate  $r_\phi(n)$ ).

To prove (c), use Lemma A.1 to obtain the inequality

$$\begin{aligned} & \tilde{r}_\phi(n)\beta(n) \\ & \leq \frac{\tilde{r}_\phi(n)}{r_\phi(n_1)} \left[ \frac{1}{2} r_\phi(n_1) \int \pi(dx) \|P^{n_1}(x; \cdot) - \pi\| + \frac{3}{2} r_\phi(n_1) \int \mu(dx) \|P^{n_1}(x; \cdot) - \pi\| \right]. \end{aligned}$$

The term in square brackets converges to zero as  $n \rightarrow \infty$  by parts (b) and (d) and, by assumption,  $\limsup_{n \rightarrow \infty} \tilde{r}_\phi(n)/r_\phi(n_1) < \infty$ . This establishes (c).

To prove (e), it suffices to note that

$$\frac{\tilde{r}_\phi(n)}{r_\phi(n_1)} = \frac{\tilde{r}_\phi(n)}{\tilde{r}_{\phi,0}(n)} \frac{\tilde{r}_{\phi,0}(n)}{r_{\phi,0}(n_1)} \frac{r_{\phi,0}(n_1)}{r_\phi(n_1)},$$

where the first and the last ratio on the right-hand side are bounded from above uniformly in  $n$  due to (6), and that

$$r_{\phi,0}(n_1) = \left( \frac{1 + \ln(n_1)}{1 + \ln(n)} \right)^\alpha (1 + \ln(n))^\alpha \cdot \left( \frac{1 + n_1}{1 + n} \right)^\beta (1 + n)^\beta \cdot \frac{e^{cn_1^\gamma}}{e^{c(n/2)^\gamma}} e^{(c2^{-\gamma})n^\gamma},$$

where the three ratios on the right-hand side are clearly bounded from below uniformly in  $n$  by some constant larger than zero. □

*Proof of Theorem 3.* The ergodicity results of parts (a) and (b) could be obtained using results in [6, Section 3.3] and those in part (c) using results in [10, Section 2.2]; for clarity of presentation, we will in all parts rely on the results in [17]. Model (9) with conditions (A1) and (A2) is a special case of the model considered in [17] (with  $p = \rho = 1$  in that paper). Of the assumptions made in [17], Assumption 1 holds due to (A1) and either Assumption 2(a) or 2(b) holds due to (A2) and the moment conditions assumed in parts (a)–(c) of Theorem 3. Therefore we can make use of Theorems 2 and 3 in [17] to obtain suitable ergodicity results.

(a) In this case Assumption 2(a) of [17] holds with  $\kappa_0 = 1$  and we apply their Theorem 2(ii). From the proof of that theorem (case  $p = 1$ ) it can be seen that Condition Drift–G holds with  $V(x) = e^{b_1|x|}$  for some  $b_1 \in (0, z_0)$ , which can be chosen as small as desired. From Theorem 2(ii) of [17] we obtain that  $X_t$  is geometrically ergodic with convergence rate  $r(n) = (e^c)^n$  (for some  $c > 0$ ), i.e.  $r(n) = r_1^n$  for some  $r_1 > 1$ . To obtain results on  $\beta$ -mixing, we next apply Theorem 1 of the present paper. If the initial state  $X_0$  has distribution such that  $\mathbb{E}[e^{z|X_0|}] < \infty$  for some  $z > 0$  (and noting that above  $b_1$  can be chosen small enough so that  $b_1 \leq z$  holds), then by Theorem 1  $X_t$  is  $\beta$ -mixing and the mixing coefficients satisfy, for some  $r_3 > 1$ ,  $\lim_{n \rightarrow \infty} r_3^n \beta(n) = 0$ .

(b) In this case Assumption 2(a) of [17] holds with  $\kappa_0 \in (0, 1)$  and we apply their Theorem 2(i). From the proof of that theorem (case  $p = 1$ ) it can be seen that Condition Drift–SubG holds with  $V(x) = e^{b_1|x|^{\kappa_0}}$  (for some  $b_1 \in (0, \beta_0)$ , which can be chosen as small as desired) and  $\phi(v) = c_0(v + v_0)(\ln(v + v_0))^{-\alpha}$  (for some  $c_0, v_0 > 0$  and  $\alpha = 1/\kappa_0 - 1$ ). From Theorem 2(i) of [17] we obtain that  $X_t$  is subexponentially ergodic with convergence rate  $r(n) = (e^c)^{n^{\kappa_0}}$  (for some  $c > 0$ ). To obtain results on  $\beta$ -mixing, we next apply Theorem 2 of the present paper. If the initial state  $X_0$  has distribution such that  $\mathbb{E}[e^{z|X_0|^{\kappa_0}}] < \infty$  for some  $z > 0$  (and noting that above  $b_1$  can be chosen small enough so that  $b_1 \leq z$  holds), then by Theorem 2  $X_t$  is  $\beta$ -mixing and the mixing coefficients satisfy, for any  $\tilde{c} \in (0, z2^{-\kappa_0})$ ,  $\lim_{n \rightarrow \infty} \tilde{r}(n)\beta(n) = 0$  with  $\tilde{r}(n) = (e^{\tilde{c}})^{n^{\kappa_0}}$ .

(c) In this case Assumption 2(b) of [17] holds with either  $s_0 = 2$  or  $s_0 \geq 4$  and we apply their Theorem 3(ii) (in which exactly the cases  $s_0 = 2$  and  $s_0 \geq 4$  are available). From the proof of that theorem (the end of step 4 and case  $p = 1$ ) it can be seen that Condition Drift–SubG holds with  $V(x) = 1 + |x|^{s_0}$  and  $\phi(v) = cv^\alpha$  (for some  $c > 0$  and  $\alpha = 1 - 1/s_0$ ). From Theorem 3(ii) of [17] we obtain that  $X_t$  is polynomially ergodic with convergence rate  $r(n) = n$  ( $s_0 = 2$ ) or  $r(n) = n^{s_0-1}$  ( $s_0 \geq 4$ ). To obtain results on  $\beta$ -mixing, we next apply Theorem 2 of the present paper. If the initial state  $X_0$  has distribution such that  $\mathbb{E}[|X_0|^{s_0}] < \infty$ , then  $X_t$  is  $\beta$ -mixing and the mixing coefficients satisfy  $\lim_{n \rightarrow \infty} n\beta(n) = 0$  ( $s_0 = 2$ ) or  $\lim_{n \rightarrow \infty} n^{s_0-1}\beta(n) = 0$  ( $s_0 \geq 4$ ). □

*Proof of Theorem 4.* Part (a) follows immediately from Theorem 3 and the discussion preceding it, noting that the SETAR model (7) with (8d) satisfies the condition in (A1). To prove (b), assume that  $\mathbb{E}[e^{z_0|\varepsilon_1|}] = \infty$  for all  $z_0 > 0$  but that  $X_t$  would be geometrically ergodic. We will use results of [12] to show that this leads to a contradiction. To this end, note that for the SETAR model (7) with the parameters satisfying (8d), the function  $g$  in our equation (9) equals

$$g(x) = \sum_{j=1}^M [\varphi(j) + \theta(j)x] \mathbf{1}\{x \in (r_{j-1}, r_j]\},$$

which can be written as

$$\begin{aligned} g(x) &= [\varphi(1) + x] \mathbf{1}\{x \leq r_1\} + [\varphi(M) + x] \mathbf{1}\{r_{M-1} < x\} \\ &\quad + \sum_{j=2}^{M-1} [\varphi(j) + \theta(j)x] \mathbf{1}\{x \in (r_{j-1}, r_j]\} \\ &= x + \varphi(1) \mathbf{1}\{x \leq r_1\} + \varphi(M) \mathbf{1}\{r_{M-1} < x\} \\ &\quad + \sum_{j=2}^{M-1} [\varphi(j) + \theta(j)x - x] \mathbf{1}\{x \in (r_{j-1}, r_j]\} \end{aligned}$$

or as  $g(x) = x + \tilde{g}(x)$ , where  $\tilde{g}(x)$  is bounded. Also recall that it is assumed that the error terms satisfy  $W_t(j) = \varepsilon_t$  ( $j = 1, \dots, M$ ) with  $\varepsilon_t$  as in (A2). These facts show that the SETAR model (7) with (8d) can be expressed in the form of equation (3) in Jarner and Tweedie [12] so that  $X_t$  is what [12] call a ‘random-walk-type Markov chain’. (Note also that this holds only in case (8d) out of (8a)–(8e).) Theorem 2.2 of [12] shows that a necessary condition for the geometric ergodicity of a random-walk-type Markov chain  $X_t$  with stationary probability measure  $\pi$  is that there exists a  $z > 0$  such that  $\int_{\mathbb{R}} e^{z|x|} \pi(dx) < \infty$ . This can be shown to be in contradiction with our assumption that  $\mathbb{E}[e^{z_0|\varepsilon_1|}] = \infty$  for all  $z_0 > 0$ , as follows.

Suppose  $z > 0$  is such that  $\int_{\mathbb{R}} e^{z|x|} \pi(dx) < \infty$  and assume that  $X_0$ , and hence also  $X_1$ , has the stationary distribution  $\pi$ . Thus  $\mathbb{E}[e^{z|X_0|}] < \infty$  and  $\mathbb{E}[e^{z|X_1|}] < \infty$ . As  $0 < e^{zx} \leq e^{z|x|}$  and  $0 < e^{-zx} \leq e^{z|x|}$ , it follows that  $\mathbb{E}[e^{zX_0}]$ ,  $\mathbb{E}[e^{-zX_0}]$ ,  $\mathbb{E}[e^{zX_1}]$ , and  $\mathbb{E}[e^{-zX_1}]$  are all positive and finite. As  $X_1 = X_0 + \tilde{g}(X_0) + \varepsilon_1$  with  $X_0$  and  $\varepsilon_1$  independent,  $\mathbb{E}[e^{zX_1}] = \mathbb{E}[e^{zX_0} e^{z\tilde{g}(X_0)}] \mathbb{E}[e^{z\varepsilon_1}]$  (due to the non-negativity of the exponential function, this holds whether the expectations involved are finite or equal  $+\infty$ ). As  $0 < \mathbb{E}[e^{zX_0}]$ ,  $\mathbb{E}[e^{zX_1}] < \infty$  and  $\tilde{g}(X_0)$  is bounded, this implies that  $0 < \mathbb{E}[e^{z\varepsilon_1}] < \infty$ . An analogous argument yields that  $0 < \mathbb{E}[e^{-z\varepsilon_1}] < \infty$ . Finally, non-negativity of the random variables involved implies that

$$\mathbb{E}[e^{z|\varepsilon_1|}] = \mathbb{E}[e^{z\varepsilon_1} \mathbf{1}\{\varepsilon_1 \geq 0\} + e^{-z\varepsilon_1} \mathbf{1}\{\varepsilon_1 < 0\}] \leq \mathbb{E}[e^{z\varepsilon_1}] + \mathbb{E}[e^{-z\varepsilon_1}] < \infty,$$

yielding a contradiction. □

## Acknowledgements

The authors thank the Academy of Finland for financial support, and the editors and an anonymous referee for useful comments and suggestions.

## References

- [1] BRADLEY, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics*, eds E. Eberlein and M. S. Taqqu, pp. 165–192. Birkhäuser, Boston.
- [2] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions: a survey and some open questions. *Prob. Surveys* **2**, 107–144.
- [3] BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions*, Vols 1–3. Kendrick Press, Heber City.
- [4] CHAN, K. S., PETRUCCELLI, J. D., TONG, H. AND WOOLFORD, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Prob.* **22**, 267–279.
- [5] DAVYDOV, YU. A. (1973). Mixing conditions for Markov chains. *Theory Prob. Appl.* **18**, 312–328.
- [6] DOUC, R., FORT, G., MOULINES, E. AND SOULIER, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Prob.* **14**, 1353–1377.
- [7] DOUC, R., MOULINES, E., PRIOURET, P. AND SOULIER, P. (2018). *Markov Chains*. Springer, Cham.
- [8] DOUKHAN, P. (1994). *Mixing: Properties and Examples*. Springer, New York.
- [9] FORT, G. AND MOULINES, E. (2000). V-subgeometric ergodicity for a Hastings–Metropolis algorithm. *Statist. Prob. Lett.* **49**, 401–410.
- [10] FORT, G. AND MOULINES, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stoch. Proc. Appl.* **103**, 57–99.
- [11] JARNER, S. F. AND ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Prob.* **12**, 224–247.
- [12] JARNER, S. F. AND TWEEDIE, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9**, 559–578.
- [13] KLOKOV, S. A. (2007). Lower bounds of mixing rate for a class of Markov processes. *Theory Prob. Appl.* **51**, 528–535.
- [14] KLOKOV, S. A. AND VERETENNIKOV, A. YU. (2004). Sub-exponential mixing rate for a class of Markov chains. *Math. Commun.* **9**, 9–26.
- [15] KLOKOV, S. A. AND VERETENNIKOV, A. YU. (2005). On subexponential mixing rate for Markov processes. *Theory Prob. Appl.* **49**, 110–122.
- [16] LIEBSCHER, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Series Anal.* **26**, 669–689.
- [17] MEITZ, M. AND SAIKKONEN, P. (2020). Subgeometrically ergodic autoregressions. To appear in *Econometric Theory*. Available at <http://doi.org/10.1017/S0266466620000419>.
- [18] MEITZ, M. AND SAIKKONEN, P. (2021). Subgeometric ergodicity and  $\beta$ -mixing: supplementary material. Available at <http://doi.org/10.1017/jpr.2020.108>.
- [19] MEYN, S. P. AND TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.
- [20] NUMMELIN, E. AND TUOMINEN, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Appl.* **12**, 187–202.
- [21] NUMMELIN, E. AND TUOMINEN, P. (1983). The rate of convergence in Orey’s theorem for Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Appl.* **15**, 295–311.
- [22] STONE, C. AND WAINGER, S. (1967). One-sided error estimates in renewal theory. *J. Anal. Math.* **20**, 325–352.
- [23] TUOMINEN, P. AND TWEEDIE, R. L. (1994). Subgeometric rates of convergence of  $f$ -ergodic Markov chains. *Adv. Appl. Prob.* **26**, 775–798.
- [24] TWEEDIE, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability, Statistics and Analysis*, eds J. F. C. Kingman and G. E. H. Reuter, pp. 260–276. Cambridge University Press.
- [25] VERETENNIKOV, A. YU. (1988). Bounds for the mixing rate in the theory of stochastic equations. *Theory Prob. Appl.* **32**, 273–281.
- [26] VERETENNIKOV, A. YU. (1991). Estimating the mixing rate for Markov processes. *Lith. Math. J.* **31**, 27–34.
- [27] VERETENNIKOV, A. YU. (2000). On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory Prob. Appl.* **44**, 361–374.

- [28] VERETENNIKOV, A. YU. AND GULINSKII, O. V. (1990). Rate of mixing and the averaging principle for stochastic recursive procedures. *Autom. Remote Control* **51**, 779–788.
- [29] VOLKONSKII, V. A. AND ROZANOV, YU. A. (1959). Some limit theorems for random functions, I. *Theory Prob. Appl.* **4**, 178–197.
- [30] VOLKONSKII, V. A. AND ROZANOV, YU. A. (1961). Some limit theorems for random functions, II. *Theory Prob. Appl.* **6**, 186–198.