On Markov-Switching ARMA Processes-Stationarity, Existence of Moments, and Geometric Ergodicity

Author(s): Robert Stelzer

Source: *Econometric Theory*, Feb., 2009, Vol. 25, No. 1 (Feb., 2009), pp. 43-62

Published by: Cambridge University Press

Stable URL: https://www.jstor.org/stable/20532432

Econometric Theory, **25**, 2009, 43–62. Printed in the United States of America.
doi:10.1017/S0266466608090026

ON MARKOV-SWITCHING ARMA PROCESSES—STATIONARITY, EXISTENCE OF MOMENTS, AND GEOMETRIC ERGODICITY

ROBERT STELZER
Technische Universität München

The probabilistic properties of $\mathbb{R}^d$-valued Markov-switching autoregressive moving average (ARMA) processes with a general state space parameter chain are analyzed. Stationarity and ergodicity conditions are given, and an easy-to-check general sufficient stationarity condition based on a tailor-made norm is introduced. Moreover, it is shown that causality of all individual regimes is neither a necessary nor a sufficient criterion for strict negativity of the associated Lyapunov exponent.

Finiteness of moments is also considered and geometric ergodicity and strong mixing are proven. The easily verifiable sufficient stationarity condition is extended to ensure these properties.

1. INTRODUCTION

To model time series that exhibit structural breaks but behave locally linear, a vast number of modifications of the classical autoregressive moving average (ARMA) model (see, e.g., Brockwell and Davis, 1991) using time-dependent ARMA coefficients have been introduced, including Markov-switching ARMA (MS-ARMA) processes, where the ARMA coefficients are allowed to change over time according to a Markov chain. In this paper we extend the well-known MS-ARMA processes with the ARMA parameters being a Markov chain with finitely many states (cf., e.g., Francq and Zakoïan, 2001; Yao, 2001) by allowing for an arbitrary (i.e., possibly uncountable) state space of the parameter process, study various probabilistic properties, and introduce a new feasible criterion for these properties to hold.

Since the seminal paper by Hamilton (1989) MS-ARMA models have been used actively in econometrics to model various time series (see, e.g., Hamilton, 1990; Krolzig, 1997; Hamilton and Raj, 2002; and the references therein).

The research leading to this paper was originally undertaken for the author's diploma thesis. The author thanks his supervisors Claudia Klüppelberg and Alexander Lindner for all their advice. Furthermore, financial support by the Deutsche Forschungsgemeinschaft through the graduate program "Angewandte Algorithmische Mathematik" at the Munich University of Technology is gratefully acknowledged. Address correspondence to Robert Stelzer, Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, D-85747 Garching, Germany; e-mail: rstelzer@ma.tum.de.

 0266-4666/09 $15.00

Moreover, they have also been used extensively in electrical engineering (see Tugnait, 1982; Doucet, Logothetis, and Krishnamurthy, 2000; and references therein). In all applications so far the Markov parameter chain has had only finitely many states, and only the theoretical statistical paper by Douc, Moulines, and Rydén (2004) has allowed for infinitely many. However, it may often be advantageous to use an MS-ARMA model with an uncountable state space of the ARMA coefficients where the Markovian structure is described by only a few parameters instead of a model with a discrete but large state space. One natural model, for example, is an MS-ARMA process where the ARMA coefficients are chosen from a distribution centered around the old coefficients (see Examples 5.1 and 5.2 later in this paper for concrete univariate MS-AR(1) processes of this type). Thus the comprehensive probabilistic study of MS-ARMA processes with a general state space presented in the following discussion provides the basis for interesting new specifications of MS-ARMA processes in applications. Moreover, it should be noted that our general model includes random coefficient ARMA models and that the effects of a heavy-tailed noise in MS-ARMA models of this form are studied in Stelzer (2008).

The outline of this paper is as follows. We start in Section 2 by defining MS-ARMA processes with a general state space parameter chain and consider throughout vector-valued processes. Here we mainly discuss the literature on the finite state space case and the extension to infinite (noncountable) state spaces. In particular, we show that the sufficient stationarity and ergodicity criteria from the finite state space case extend to our general model. In Section 3 we analyze the relation between causality of the individual regimes (the possible ARMA coefficient sets) and the stationarity of the MS-ARMA process. Furthermore, we establish as our main result a feasible sufficient stationarity condition, which is based on a general result on the norm of matrices of a special structure. The existence of moments is discussed in Section 4, and finally we establish V-uniform ergodicity and thereby geometric ergodicity and strong mixing in Section 5.

2. THE MARKOV-SWITCHING ARMA MODEL

In defining MS-ARMA processes, one starts from a (multivariate) ARMA equation (see, e.g., Brockwell and Davis, 1991) with drift and allows for random coefficients that are modeled as a Markov chain. We denote the real $d \times d$ ($m \times n$) matrices by $M_d(\mathbb{R})$ ($M_{m,n}(\mathbb{R})$). Moreover, "stationarity" always means strict stationarity.

DEFINITION 2.1 (MS-ARMA(p,q) process). *Let* $p,q \in \mathbb{N}_0$, $p+q \geq 1$ *and* $\Delta = (\mu_t, \Sigma_t, \Phi_{1t}, \ldots, \Phi_{pt}, \Theta_{1t}, \ldots, \Theta_{qt})_{t\in\mathbb{Z}}$ *be a stationary and ergodic Markov chain with some (measurable) subset* S *of* $\mathbb{R}^d \times M_d(\mathbb{R})^{1+p+q}$ *as state space. Moreover, let* $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ *be an independent and identically distributed (i.i.d.) sequence of* $\mathbb{R}^d$*-valued random variables independent of* Δ *and set* $Z_t := \Sigma_t \epsilon_t \in \mathbb{R}^d$.

A stationary process $(X_t)_{t\in\mathbb{Z}}$ *in* $\mathbb{R}^d$ *is called an* MS-ARMA(p,q,Δ,ϵ) process *if it satisfies*

$$X_t - \Phi_{1t}X_{t-1} - \cdots - \Phi_{pt}X_{t-p} = \mu_t + Z_t + \Theta_{1t}Z_{t-1} + \cdots + \Theta_{qt}Z_{t-q} \qquad \textbf{(2.1)}$$

for all $t \in \mathbb{Z}$. *Equation (2.1) is referred to as the* MS-ARMA (p,q,Δ,ϵ) equation.

Furthermore, a stationary process $(X_t)_{t\in\mathbb{Z}}$ *is said to be an* MS-ARMA(p,q) *process if it is an MS-ARMA*(p,q,Δ,ϵ) *process for some* Δ *and* ϵ *satisfying the preceding conditions.*

Remark 2.1.

(a) The elements of S are called *regimes*, extending the notion from the finite state space literature. The state space S is assumed to be equipped with a metric inherited from some norm on $\mathbb{R}^d \times M_d(\mathbb{R})^{1+p+q}$ and the Borel σ-algebra $\mathcal{S}$.

(b) "Ergodic" is to be understood in its general measure theoretic meaning, namely, that the back-shift invariant σ-algebra over the sequence space is trivial; see, e.g., Ash and Gardner (1975) and the comprehensive monograph by Krengel (1985).

(c) The preceding definition extends the one from the case with only finitely many regimes (see, e.g., Francq and Zakoïan, 2001). It includes random coefficient autoregressions (i.e., autoregressive [AR] processes with i.i.d. random coefficients) as analyzed, e.g., in Nicholls and Quinn (1982), Feigin and Tweedie (1985), and Klüppelberg and Pergamenchtchikov (2004).

(d) Sometimes it may be of interest to consider a setup with the dimensions of X and ϵ being different. To this end one can simply take ϵ to be an $\mathbb{R}^k$-valued sequence and Σ to be $M_{d,k}(\mathbb{R})$-valued. All results of this paper except Proposition 5.2 extend immediately to this setup. Yet, Proposition 5.2 also remains valid when assuming $k \geq d$ and that Σ_t is always of full rank.

Given some i.i.d. noise (ϵ_t) and parameter chain (Δ_t), the natural question arising is whether there exists a stationary and ergodic solution (X_t) to (2.1). In what follows, the zeros denote zeros in $M_{m,n}(\mathbb{R})$ or $\mathbb{R}^d$ with the appropriate dimensions m, n, and d being obvious from the context.

PROPOSITION 2.1 (State space representation). *Define*

$$\begin{aligned}
\mathbf{X}_t &= (X_t^\top, X_{t-1}^\top, \ldots, X_{t-p+1}^\top, Z_t^\top, \ldots, Z_{t-q+1}^\top)^\top \in \mathbb{R}^{d(p+q)},\\
\Sigma_t &= (\Sigma_t^\top, \underbrace{0^\top, \ldots, 0^\top}_{p-1}, \Sigma_t^\top, \underbrace{0^\top, \ldots, 0^\top}_{q-1})^\top \in M_{d(p+q),d}(\mathbb{R}), \qquad \textbf{(2.2)}\\
\mathbf{m}_t &= (\mu_t^\top, 0^\top, \ldots, 0^\top)^\top \in \mathbb{R}^{d(p+q)}, \qquad \mathbf{C}_t = \mathbf{m}_t + \Sigma_t\epsilon_t,
\end{aligned}$$

$$\mathbf{\Phi}_t = \begin{pmatrix} \Phi_{1t} & \cdots & \Phi_{(p-1)t} & \Phi_{pt} \\ I_d & 0\cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots 0 & I_d & 0 \end{pmatrix} \in M_{dp}(\mathbb{R}), \qquad \mathbf{J} = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ I_d & 0 & \cdots & 0 \\ 0 & \ddots & 0\cdots & \vdots \\ 0 & \cdots 0 & I_d & 0 \end{pmatrix} \in M_{dq}(\mathbb{R}),$$

$$\Theta_t = \begin{pmatrix} \Theta_{1t} & \cdots & \Theta_{(q-1)t} & \Theta_{qt} \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \in M_{dp,dq}(\mathbb{R}), \qquad \mathbf{A}_t = \begin{pmatrix} \mathbf{\Phi}_t & \Theta_t \\ 0 & \mathbf{J} \end{pmatrix} \in M_{d(p+q)}(\mathbb{R}). \tag{2.3}$$

Then (2.1) has a stationary and ergodic solution if and only if

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{C}_t \tag{2.4}$$

has one.

Proof. We obviously have that any stationary solution of (2.1) leads via (2.2) to one of (2.4) and, vice versa, that the first d components of a stationary solution of (2.4) are one for (2.1). That an ergodic solution of (2.4) gives an ergodic solution of (2.1) and vice versa follows from standard ergodicity theory (see, e.g., Brandt, Franken, and Lisek, 1990, Lem. A.1.2.7). ■

Remark 2.2.

(a) To avoid degeneracies in the state space representation, we presume without loss of generality that $p \geq 1$ from now on. In the case of a purely autoregressive MS-ARMA equation, i.e., $q = 0$, it is implicitly understood that $\mathbf{J}_t$ and Θ_t vanish, $\mathbf{X}_t = (X_t^\mathsf{T}, X_{t-1}^\mathsf{T}, \ldots, X_{t-p+1}^\mathsf{T})^\mathsf{T}$, $\Sigma_t = (\Sigma_t^\mathsf{T}, 0^\mathsf{T}, \ldots, 0^\mathsf{T})^\mathsf{T}$, and $\mathbf{A}_t = \mathbf{\Phi}_t$.

(b) This proposition shows also that any d-dimensional MS-ARMA(p,q) process can be represented as a $d(p+q)$-dimensional MS-AR(1) process.

Regarding notation, $\|\cdot\|$ will denote any norm on $\mathbb{R}^{d(p+q)}$ and also the induced operator norm and $\xrightarrow{\mathscr{D}}$ convergence in distribution. If $k = 0$, the product $\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k+1}$ is understood to be identical to the identity $I_{d(p+q)}$ on $\mathbb{R}^{d(p+q)}$, a convention to be used throughout for products of this structure.

THEOREM 2.1.

(i) (Stationary solution) Equation (2.4) and the MS-ARMA(p,q,Δ,ϵ) equation (2.1) have a unique stationary and ergodic solution if

$\mathrm{E}(\log^+ \|\mathbf{A}_0\|)$ *and* $\mathrm{E}(\log^+ \|\mathbf{C}_0\|)$ *are finite and the Lyapunov exponent* $\gamma := \inf_{t\in\mathbb{N}_0} \{\mathrm{E}(\log(\|\mathbf{A}_0\mathbf{A}_{-1}\ldots\mathbf{A}_{-t}\|))/(t+1)\}$ *is strictly negative. The unique stationary solution* $\mathbf{X} = (\mathbf{X}_t)_{t\in\mathbb{Z}}$ *of (2.4) is given by*

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t\mathbf{A}_{t-1}\ldots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}, \tag{2.5}$$

and this series converges absolutely almost surely (a.s.).

(ii) (Convergence to the stationary solution) Let $\mathbf{V}_0$ *be an arbitrary* $\mathbb{R}^{d(p+q)}$*-valued random variable defined on the same probability space as* $(\Delta_t, \epsilon_t)_{t\in\mathbb{Z}}$ *and define* $(\mathbf{V}_t)_{t\in\mathbb{N}}$ *recursively via (2.4) (or let* $V_0, \ldots, V_{-p+1}, Z_0, \ldots, Z_{-q+1}$ *be arbitrary* $\mathbb{R}^d$*-valued random variables and define* $(V_t)_{t\in\mathbb{N}}$ *via (2.1),* $\mathbf{V}_t := (V_t, \ldots, V_{t-p+1}, Z_t, \ldots, Z_{t-q+1})^{\mathsf{T}}$*).*
Then $\|\mathbf{X}_t - \mathbf{V}_t\| \overset{\text{a.s.}}{\to} 0$ *as* $t \to \infty$ *and, in particular,* $\mathbf{V}_t \overset{\mathcal{D}}{\to} \mathbf{X}_0$ *as* $t \to \infty$.

Proof. $(\epsilon_t)_{t\in\mathbb{Z}}$ is i.i.d. and thereby mixing. As, moreover, $(\Delta_t)_{t\in\mathbb{Z}}$ is ergodic, Brandt et al. (1990, Thm. A.1.2.6) implies that the joint random sequence $(\Delta, \epsilon) = (\Delta_t, \epsilon_t)_{t\in\mathbb{Z}}$ is stationary and ergodic, which in turn gives that the transformed sequence $(\mathbf{A}_t, \mathbf{C}_t)_{t\in\mathbb{Z}}$ is stationary and ergodic (Brandt et al., 1990, Lem. A.1.2.7). Hence, we obtain (i) from the multidimensional extension of Theorem 1 of Brandt (1986) by Bougerol and Picard (1992, Thm. 1.1). Part (ii) is now also immediate from Brandt (1986, Thm. 1). ■

For a finite state space of Δ, Theorem 2.1 (i) has been given in Francq and Zakoïan (2001) together with a proof along the same lines. The results in (ii) will later be extended to geometric ergodicity of $(\mathbf{X}_t, \Delta_t)$, but this requires considerably more involved conditions.

Remark 2.3. Let $(A_t)_{t\in\mathbb{Z}}$ be any stationary and ergodic random sequence in $M_d(\mathbb{R})$ and $\gamma = \inf_{t\in\mathbb{N}_0}\{\mathrm{E}(\log \|A_0A_{-1}\ldots A_{-t}\|)/(t+1)\}$ its Lyapunov exponent. Then γ is independent of the algebra norm. Consequently, one can work with some algebra norm that makes it rather straightforward to show $\gamma < 0$. Observe also that $\mathrm{E}(\log \|A_0\|) < 0$ suffices to ensure $\gamma < 0$.

Although in our case matrices of the structure of $\mathbf{A}_t$ are of norm greater than or equal to one in all usual matrix norms, the latter is used in the next section to obtain a feasible condition.

A classical result from Furstenberg and Kesten (1960, Thm. 1) states that the infimum can be replaced by a limit, i.e.,

$$\gamma = \lim_{n\to\infty} \frac{1}{n+1}\mathrm{E}(\log \|A_0A_{-1}\ldots A_{-n}\|). \tag{2.6}$$

Actually it is only the AR part $\mathbf{\Phi}_t$ of the matrix $\mathbf{A}_t$ that determines the Lyapunov exponent. Francq and Zakoïan (2001, p. 343) showed this for a finite state space Markov parameter chain, but their proof is also valid in our general case.

PROPOSITION 2.2. *Let* $\|\cdot\|$ *denote arbitrary algebra norms on* $M_{d(p+q)}(\mathbb{R})$ *and* $M_{dp}(\mathbb{R})$ *and* $\mathrm{E}\left(\log^+\|\mathbf{A}_0\|\right) < \infty$*; then* $\tilde{\gamma} := \inf_{t\in\mathbb{N}_0}\{\mathrm{E}(\log(\|\boldsymbol{\Phi}_0\boldsymbol{\Phi}_{-1}\ldots\boldsymbol{\Phi}_{-t}\|))/(t+1)\}$.

"Causality" is an important concept in the analysis of ARMA processes. The following definition gives an appropriate extension to MS-ARMA processes.

DEFINITION 2.2 (Causality). *An MS-ARMA*(p,q,Δ,ϵ) *process* $(X_t)_{t\in\mathbb{Z}}$ *is said to be* causal *if there is some measurable function* f *such that* $X_t = f(\Delta_t,\Delta_{t-1},\ldots,\epsilon_t,\epsilon_{t-1},\ldots)\,\forall t\in\mathbb{Z}$.

Remark 2.4.

(a) The unique stationary solution to an MS-ARMA equation constructed in Theorem 2.1 is causal.
(b) If Δ is an i.i.d. sequence, the results of Bougerol and Picard (1992) show under technical conditions that the strict negativity of the Lyapunov coefficient is also necessary for the existence of a causal solution to an MS-ARMA equation. See also Goldie and Maller (2000) for a general discussion of the one-dimensional case.

3. GLOBAL AND LOCAL STATIONARITY

The preceding discussion has shown that it is important to find criteria ensuring strict negativity of the Lyapunov exponent that can be easily used in practice. In this section we discuss the relation to causality in the sense of Brockwell and Davis (1991, Def. 3.1.3, p. 468) of the individual regimes.

DEFINITION 3.1. *An MS-ARMA process is called* locally stationary *if a.s. all the eigenvalues of* $\boldsymbol{\Phi}_0$ *are strictly less than one in modulus, and it is said to be* globally stationary *if the Lyapunov exponent* γ *is strictly negative.*

We use the term *local stationarity*, extending a notion introduced in Francq and Zakoïan (2001) regarding L^2-stationarity. Note, however, that this term is also used in a very different sense in the literature.

Intuitively local stationarity means that, whenever we fix the ARMA coefficients to one set of possible values (the same one for all times!), we obtain a causal ARMA process.

By Theorem 2.1 and Remark 2.4 (a) global stationarity implies that the MS-ARMA process is causal, provided the logarithmic moment conditions are satisfied. Before giving a theorem on simultaneous local and global stationarity, we show that the relation between local and global stationarity is highly non-trivial, as in general neither of the two is sufficient or necessary for the other.

PROPOSITION 3.1 (MS-ARMA$(1,q)$). *Let a one-dimensional MS-ARMA $(1,q)$ process be given and assume that* $\mathrm{E}(\log^+ \|\mathbf{A}_0\|)$ *is finite. Then local stationarity is a sufficient condition for global stationarity.*

Proof. For $s \in S$ let $\mathbf{\Phi}_0(s) = \mathbf{\Phi}_0 | \Delta_0 = s$. Local stationarity gives $|\mathbf{\Phi}_0(s)| < 1$ a.s., and thus $\gamma = \tilde{\gamma} = \mathrm{E}(\log|\mathbf{\Phi}_0|) < 0$. ■

In view of Remark 2.2 (b) and the upcoming Example 3.2 it is clear that extending the result to $d > 1$ is not possible.

Example 3.1 (Nonnecessity of local stationarity)

Consider an MS-ARMA $(1,q)$ process in one dimension and let Δ_t have two states $\Delta^{(1)}, \Delta^{(2)}$ and stationary distribution $(\pi^{(1)}, \pi^{(2)})$. Then $\mathrm{E}(\log|\mathbf{\Phi}_0|) < 0$ translates into $\pi^{(1)} \log|\mathbf{\Phi}^{(1)}| + \pi^{(2)} \log|\mathbf{\Phi}^{(2)}| < 0$, where $\mathbf{\Phi}^{(1)}$ and $\mathbf{\Phi}^{(2)}$ are the two possible values for $\mathbf{\Phi}_t$. This is equivalent to $|\mathbf{\Phi}^{(1)}|^{\pi^{(1)}} < |\mathbf{\Phi}^{(2)}|^{-\pi^{(2)}}$. From the last equation it is immediate to see that $|\mathbf{\Phi}^{(1)}|$ can be arbitrarily large provided that $|\mathbf{\Phi}^{(2)}|$ is close enough to zero. So, local stationarity is not necessary for global stationarity.

For a similar example but with an uncountable state space see Example 5.2, which follows.

Example 3.2 (Nonsufficiency of local stationarity)

Take a stationary and ergodic Markov chain Δ with two states and transition matrix

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}.$$

Further, let the regimes $\Delta^{(1)}$ and $\Delta^{(2)}$ be given by the two equations

$$X_t = \Phi_1^{(1)} X_{t-1} + \Phi_2^{(1)} X_{t-2} + \epsilon_t \qquad \text{and} \qquad X_t = \Phi_1^{(2)} X_{t-1} + \epsilon_t,$$

where $\Phi_1^{(1)} = \frac{9}{5}$, $\Phi_2^{(1)} = -\frac{9}{10}$, and $\Phi_1^{(2)} = -\frac{1}{5}$. So, the possible states of $\mathbf{A}_t$ are

$$\mathbf{A}^{(1)} = \begin{pmatrix} \frac{9}{5} & -\frac{9}{10} \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{A}^{(2)} = \begin{pmatrix} -\frac{1}{5} & 0 \\ 1 & 0 \end{pmatrix}.$$

As one obtains $\rho(\mathbf{A}^{(1)}) = |(\frac{9}{10}) \pm (\frac{3}{10})i| = \frac{3}{\sqrt{10}} < 1$ and $\rho(\mathbf{A}^{(2)}) = \frac{1}{5}$ for the spectral radii, both regimes correspond to causal AR processes.

The crucial observation is that $R := \mathbf{A}^{(1)}\mathbf{A}^{(2)}$ and $T := \mathbf{A}^{(2)}\mathbf{A}^{(1)}$ both have spectral radius $\frac{63}{50} > 1$. Fixing p_{12} and p_{21} to the value one, we obtain an ergodic and periodic Markov chain Δ, which has stationary distribution $(\pi^{(1)}, \pi^{(2)}) = (0.5, 0.5)$. Observe that aperiodicity is not required for ergodicity in our sense, as any stationary, irreducible, and positive recurrent countable state space Markov chain is ergodic in our sense (see Ash and Gardner, 1975, Sect. 3.5). Let us further assume temporarily that the noise ϵ is not random at all, but $\epsilon_t = 1$ for all times. So $\mathbf{C}_t = (1,0)^{\mathsf{T}}$. One readily calculates for $n \in \mathbb{N}$

$$R^n \mathbf{C}_0 = \begin{pmatrix} \left(-\frac{63}{50}\right)^n \\ -\frac{1}{5}\left(-\frac{63}{50}\right)^{n-1} \end{pmatrix} \qquad \text{and} \qquad \mathbf{A}^{(2)} R^n \mathbf{C}_0 = \begin{pmatrix} -\frac{1}{5}\left(-\frac{63}{50}\right)^n \\ \left(-\frac{63}{50}\right)^n \end{pmatrix}.$$

Thus, both $R^n \mathbf{C}_0$ and $\mathbf{A}^{(2)} R^n \mathbf{C}_0$ diverge to infinity in norm for $n \to \infty$, and hence it is straightforward to see that the series $\mathbf{X}_0 = \sum_{k=0}^{\infty} \mathbf{A}_t \mathbf{A}_{-1} \ldots \mathbf{A}_{-k+1} \mathbf{C}_{-k}$ is almost surely divergent. Therefore, Theorem 2.1 implies that the Lyapunov coefficient associated with the chosen parameter chain Δ cannot be strictly negative. This shows that causality of all regimes does not ensure global stationarity.

Regarding L^2-stationarity similar results have been given in Francq and Zakoïan (2001). Actually, Example 3.2 is a deeper analysis of their Example 5.

The following general result on sets of matrices of the special structure of $\mathbf{A}_t$ or $\mathbf{\Phi}_t$ provides the necessary insight to obtain a condition ensuring local and global stationarity.

THEOREM 3.1. *Let $d, p \in \mathbb{N}$, $q \in \mathbb{N}_0$, and $\mathcal{A} \subset M_{d(p+q)}(\mathbb{R})$ be a set of matrices such that for each $A \in \mathcal{A}$ there are matrices $A_1(A), \ldots, A_p(A)$, $B_1(A), \ldots, B_q(A) \in M_d(\mathbb{R})$ such that*

$$A = \begin{pmatrix} A_1(A) & \cdots & A_{p-1}(A) & A_p(A) & B_1(A) & \cdots & B_{q-1}(A) & B_q(A) \\ I_d & 0\cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & \cdots 0 & I_d & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & 0 & \ddots & 0\cdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots 0 & I_d & 0 \end{pmatrix}.$$

Assume, moreover, that there is a norm $\|\cdot\|_d$ on $\mathbb{R}^d$ and $c < 1$ such that $\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \|A_i(A)\|_d < c$ and $\sup_{A\in\mathcal{A}} \sum_{i=1}^{q} \|B_i(A)\|_d < \infty$ hold for the induced operator norm.

Then there is a norm $\|\cdot\|$ on $\mathbb{R}^{d(p+q)}$ and $c' < 1$ such that $\sup_{A\in\mathcal{A}} \|A\| < c'$ in the induced operator norm. Especially, $\|x_0 x_1 \ldots x_k\| < (c')^{k+1}$ for any $k \in \mathbb{N}$ and sequence $(x_n)_{n\in\mathbb{N}_0}$ with elements in $\mathcal{A}$.

Proof. Choose $c_1, \ldots, c_p \in \mathbb{R}$ such that $1 = c_1 > c_2 > \cdots > c_p > c$. Then

$$\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_i} \le \sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} < \frac{c}{c_p} < 1.$$

Next choose $M \in (c/c_p, 1)$ and $\tilde{c} \in \mathbb{R}^+$ such that

$$\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} + \sup_{A\in\mathcal{A}} \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{\tilde{c}} < M < 1$$

and $c_{p+1}, \ldots, c_{p+q} \in \mathbb{R}$ with $c_{p+1} > \cdots > c_{p+q} > \tilde{c}$. Define a norm $\|\cdot\|$ on $\mathbb{R}^{d(p+q)}$ by

$$\begin{aligned}
&\|(x_1^\top, \ldots, x_p^\top, y_1^\top, \ldots, y_q^\top)^\top\| \\
&\quad = \max\{c_1\|x_1\|_d, \ldots, c_p\|x_p\|_d, c_{p+1}\|y_1\|_d, \ldots, c_{p+q}\|y_q\|_d\}.
\end{aligned}$$

For any $(x_1^\top, \ldots, x_p^\top, y_1^\top, \ldots, y_q^\top)^\top \in \mathbb{R}^{d(p+q)}$ and $A \in \mathcal{A}$ we have

$$\begin{aligned}
&\left\|A(x_1^\top, \ldots, x_p^\top, y_1^\top, \ldots, y_q^\top)^\top\right\| \\
&\quad = \left\|\left(\sum_{i=1}^{p}(A_i(A)x_i)^\top + \sum_{i=1}^{q}(B_i(A)y_i)^\top, x_1^\top, \ldots, x_{p-1}^\top, 0^\top, y_1^\top, \ldots, y_{q-1}^\top\right)^\top\right\| \\
&\quad = \max\left\{\left\|\sum_{i=1}^{p} A_i(A)x_i + \sum_{i=1}^{q} B_i(A)y_i\right\|_d, \frac{c_2}{c_1}c_1\|x_1\|_d, \ldots, \frac{c_p}{c_{p-1}}c_{p-1}\|x_{p-1}\|_d,\right. \\
&\qquad\qquad \left. 0, \frac{c_{p+2}}{c_{p+1}}c_{p+1}\|y_1\|_d, \ldots, \frac{c_{p+q}}{c_{p+q-1}}c_{p+q-1}\|y_{q-1}\|_d\right\} \\
&\quad \le \max\left\{\left\|\sum_{i=1}^{p} A_i(A)x_i + \sum_{i=1}^{q} B_i(A)y_i\right\|_d,\right. \\
&\qquad\qquad \left. \max_{2\le k\le p+q, k\ne p+1}\left\{\frac{c_k}{c_{k-1}}\right\} \|(x_1^\top, \ldots, x_p^\top, y_1^\top, \ldots, y_q^\top)^\top\|\right\}
\end{aligned}$$

and, moreover,

$$\left\|\sum_{i=1}^{p} A_i(A)x_i + \sum_{i=1}^{q} B_i(A)y_i\right\|_d \leq \sum_{i=1}^{p} \|A_i(A)\|_d \|x_i\|_d + \sum_{i=1}^{p} \|B_i(A)\|_d \|y_i\|_d$$

$$\leq \left(\sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_i} + \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{c_{p+i}}\right)$$

$$\times \left\|(x_1^{\mathsf{T}}, \ldots, x_p^{\mathsf{T}}, y_1^{\mathsf{T}}, \ldots, y_q^{\mathsf{T}})^{\mathsf{T}}\right\|.$$

From this one deduces

$$\sup_{A\in\mathcal{A}} \|A\| \leq \max\left\{ \sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} \right.$$

$$\left. + \sup_{A\in\mathcal{A}} \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{\tilde{c}}, \max_{2\leq k\leq p+q, k\neq p+1} \left\{\frac{c_k}{c_{k-1}}\right\}\right\}$$

$$\leq \max\left\{M, \max_{2\leq k\leq p+q, k\neq p+1} \left\{\frac{c_k}{c_{k-1}}\right\}\right\} =: c' < 1,$$

which concludes the proof. ■

Note that q can also taken to be zero in Theorem 3.1. Then the second condition $\sup_{A\in\mathcal{A}} \sum_{i=1}^{q} \|B_i(A)\|_d < \infty$ vanishes, and matrices with the structure of $\mathbf{\Phi}_t$ are analyzed.

This immediately leads to a feasible condition for the strict negativity of the Lyapunov exponent.

COROLLARY 3.1. *Consider an MS-ARMA(p,q,Δ,ϵ) equation with* $\mathrm{E}(\log^+ \|\mathbf{A}_0\|) < \infty$ *and assume that there is a norm* $\|\cdot\|_d$ *on* $\mathbb{R}^d$ *and* $\bar{c} < 1$ *such that* $\sum_{i=1}^{p} \|\Phi_{i0}\|_d \leq \bar{c}$ *a.s. Then the MS-ARMA process is globally and locally stationary.*

Proof. Apply Theorem 3.1 on the subset $\mathcal{A} = \{\mathbf{\Phi}_0 : \sum_{i=1}^{p} \|\Phi_{i0}\|_d \leq \bar{c}\}$ of the state space of $\mathbf{\Phi}_0$ to obtain an operator norm $\|\cdot\|$ that ensures $\|\mathbf{\Phi}_0\| < c'$ a.s. for some $c' < 1$. This ensures $\mathrm{E}(\log \|\mathbf{\Phi}_0\|) < 0$ and so implies the preceding claim immediately. ■

Remark 3.1. For $d = 1$ the condition on $\sum_{i=1}^{p} \|\Phi_{i0}\|_d$ corresponds to the general stationarity condition for threshold AR (TAR) models (i.e., a piecewise AR model where the parameter set is chosen dependent on the current value of the process) as given in An and Huang (1996). Actually, using the basic setup of An and Huang (1996) one can immediately give a direct proof of the TAR stationarity condition using only our Theorem 3.1 and Tweedie's drift criterion (cf. An and Huang, 1996, Lem. 2.2). This illustrates that Theorem 3.1 can be applied to various piecewise ARMA processes, as no particular features of MS-ARMA are needed.

4. EXISTENCE OF MOMENTS

In this section we give sufficient conditions for the finiteness of moments of MS-ARMA processes using the following notion of r-times integrability for multivariate random variables.

DEFINITION 4.1. *Denote by* $L^r_{\mathbb{R}}$ *with* $r \in (0,\infty]$ *the usual space of* r*-times integrable real-valued random variables and let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^d$ *(or* $M_d(\mathbb{R})$*). Then* $L^r_{\mathbb{R}^d}$ *(or* $L^r_{M_d(\mathbb{R})}$*) is defined as the space of all* $\mathbb{R}^d$*- (or* $M_d(\mathbb{R})$*-) valued random variables* X *with* $\|X\| \in L^r_{\mathbb{R}}$*. For short we often omit the space subscript and write* L^r*.*

Moreover, $\|\cdot\|_{L^r} : L^r \to \mathbb{R}^+_0,\ X \mapsto \mathrm{E}(\|X\|^r)^{1/r}$ *defines (up to almost sure identity) a norm on* L^r *for* $r \geq 1$ *and* $d_{L^r}(\cdot,\cdot) : L^r \times L^r \to \mathbb{R}^+_0,\ (X,Y) \mapsto \mathrm{E}(\|X-Y\|^r)$ *a metric on* L^r *for* $0 < r < 1$*.*

The L^r spaces are independent of the norm $\|\cdot\|$ used on $\mathbb{R}^d$ (or $M_d(\mathbb{R})$). However, different norms $\|\cdot\|$ on $\mathbb{R}^d$ (or $M_d(\mathbb{R})$) lead to different norms $\|\cdot\|_{L^r}$ and metrics $d_{L^r}(\cdot,\cdot)$. Yet, because of the equivalence of all norms on $\mathbb{R}^d$ (or $M_d(\mathbb{R})$) it is immediate to see that for different norms $\|\cdot\|$ the induced norms and metrics on L^r are equivalent. This implies that the results of this section do not depend on the norm used.

All results from the well-known theory of the $L^r_{\mathbb{R}}$ spaces extend immediately to the multidimensional L^r spaces.

THEOREM 4.1. *Assume that* $\mathrm{E}(\log^+\|\mathbf{A}_0\|), \mathrm{E}(\log^+\|\mathbf{C}_0\|) < \infty$ *and* $\gamma < 0$*. If, moreover, for some* $r \in [1,\infty]$

$$\sum_{k=0}^{\infty} \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-k+1}\mathbf{C}_{-k}\|_{L^r} \tag{4.1}$$

or for some $r \in (0,1)$

$$\sum_{k=0}^{\infty} \mathrm{E}\left(\|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^r\right) \tag{4.2}$$

converges, then the unique stationary solution X_t *of the MS-ARMA equation (2.1) given in Theorem 2.1 (i) and its state space representation* $\mathbf{X}_t$ *are in* L^r*. Moreover, the series (2.5) defining* $\mathbf{X}_t$ *converges in* L^r*.*

Proof. We assume that $t = 0$ without loss of generality. For $r \in [1;\infty]$ L^r is a Banach space, and thus the absolute convergence in (4.1) implies the convergence of the series (2.5) in L^r and that $\mathbf{X}_t \in L^r$. Using the norm $\|(x_1, x_2, \ldots, x_i)^{\mathsf{T}}\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_i|\}$ on $\mathbb{R}^{d(p+q)}$ and $\mathbb{R}^d$, this immediately gives $X_t \in L^r$ for the MS-ARMA process.

For $r \in (0,1)$ we observe that L^r is a complete metric space and for $m, n \in \mathbb{N}$, $m > n$,

$$d_{L^r}\left(\sum_{k=0}^{m} \mathbf{A}_0 \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}, \sum_{k=0}^{n} \mathbf{A}_0 \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\right)$$

$$= d_{L^r}\left(\sum_{k=n+1}^{m} \mathbf{A}_0 \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}, 0\right) \le \sum_{k=n+1}^{m} \mathrm{E}\left(\|\mathbf{A}_0 \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^r\right).$$

Therefore, (4.2) implies that $\left(\sum_{k=0}^{m} \mathbf{A}_0 \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\right)_{m\in\mathbb{N}}$ is a Cauchy sequence in L^r and thus convergent. Now proceed as in the case $r \in [1,\infty]$. ■

Remark 4.1.

(a) Using the root criterion, we have that (4.1) or (4.2) holds if

$$\limsup_{k\to\infty} \|\mathbf{A}_0\mathbf{A}_{-1} \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|_{L^r}^{1/k} < 1 \quad \text{or}$$

$$\limsup_{k\to\infty} \mathrm{E}\left(\|\mathbf{A}_0\mathbf{A}_{-1} \ldots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^r\right)^{1/k} < 1.$$

(b) It is immediate that Theorem 4.1 remains valid when replacing the MS-ARMA equation with a multivariate stochastic difference equation $\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{C}_t$ with arbitrary stationary and ergodic input $(\mathbf{A}_t, \mathbf{C}_t)$ and referring to the results of Brandt (1986) and Bougerol and Picard (1992) instead of Theorem 2.1. Then it is the multidimensional extension of the results of Karlsen (1990).

The following proposition gives a decomposition of the preceding conditions into an asymptotic condition for the sequence $\mathbf{A}_t$ and an integrability condition on $\mathbf{C}_t$.

PROPOSITION 4.1. *Let $r \in (0,\infty)$ and assume that there exist $u, v \in [1,\infty]$ with $1/u + 1/v = 1$ such that $\mathbf{A}_0 \ldots \mathbf{A}_{-k+1} \in L^{ru}\ \forall k \in \mathbb{N}$ and $\mathbf{C}_0 \in L^{rv}$. If either*

$$\limsup_{k\to\infty} \mathrm{E}(\|\mathbf{A}_0\mathbf{A}_{-1} \ldots \mathbf{A}_{-k+1}\|^{ru})^{1/k} < 1 \tag{4.3}$$

for $0 < u < \infty$ or

$$\lim_{k\to\infty} \|\mathbf{A}_0\mathbf{A}_{-1} \ldots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1 \tag{4.4}$$

for $u = \infty$, then $\gamma < 0$ and (4.1) for $r \ge 1$ or (4.2) for $0 < r < 1$ holds.

Proof. From the standard result on the limit of subadditive sequences (see, e.g., Hille and Phillips, 1957, Lem. 4.7.1) it can be straightforwardly deduced that $\lim_{k\to\infty} \|\mathbf{A}_0 \dots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k}$ exists and equals $\inf_{k\in\mathbb{N}} \|\mathbf{A}_0 \dots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k}$, if $\mathbf{A}_0 \in L^\infty$.

The inequality $\gamma < 0$ is obvious for $u = \infty$ using (4.4) and otherwise follows from Jensen's inequality and (4.3).

Finally, (4.1) for $r \geq 1$ and (4.2) for $0 < r < 1$ are established by using Remark 4.1 (a), applying Hölder's inequality, and observing that $\lim_{k\to\infty} \mathrm{E}(\|\mathbf{C}_{-k}^{rv}\|)^{1/k} = 1$ (unless $\mathbf{C}_t = 0$ a.s.). ∎

For $r \in [1,\infty)$ it is immediate that $\limsup_{k\to\infty} \mathrm{E}(\|\mathbf{A}_0 \dots \mathbf{A}_{-k+1}\|^r)^{1/k} < 1$ is equivalent to $\limsup_{k\to\infty} \|\mathbf{A}_0 \dots \mathbf{A}_{-k+1}\|_{L^r}^{1/k} < 1$.

COROLLARY 4.1. *If* $\mathbf{A}_0 \in L^\infty$, $\lim_{k\to\infty} \|\mathbf{A}_0 \dots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$, *and* $\mathbf{C}_0 \in L^r$, $r \in (0,\infty]$, *then the MS-ARMA process* X_t *and its state space representation* $\mathbf{X}_t$ *are in* L^r.

Proof. For $r = \infty$ this is obvious from Theorem 4.1; otherwise it is a direct consequence of Proposition 4.1. ∎

The following result extends the feasible stationarity criterion of Section 3 to a condition enabling one to deduce finiteness of the moments of the MS-ARMA process from the moments of $\mathbf{C}_0$.

THEOREM 4.2. *Assume that there are a* $\bar{c} < 1$, $M \in \mathbb{R}^+$, *and a norm* $\|\cdot\|_d$ *on* $\mathbb{R}^d$ *such that* $\sum_{i=1}^p \|\Phi_{i0}\|_d \leq \bar{c}$ *and* $\sum_{i=1}^q \|\Theta_{i0}\|_d \leq M$ *a.s. Let, moreover,* $\mathrm{E}(\log^+ \|\mathbf{C}_0\|)$ *be finite.*

(i) Then $\mathrm{E}(\log^+ \|\mathbf{A}_0\|) < \infty$, $\gamma < 0$, *and thus there is a unique stationary and ergodic solution* $(X_t)_{t\in\mathbb{Z}}$ *to the MS-ARMA(*p, q, Δ, ϵ*) equation (2.1) given by Theorem 2.1 (i).*

(ii) If $\mathbf{C}_0 \in L^r$ *for some* $r \in (0,\infty]$, *then the solution* X_t *of the MS-ARMA equation (2.1) and its state space representation* $\mathbf{X}_t$ *are in* L^r. *Moreover, the series defining* $\mathbf{X}_t$ *(as given in Theorem 2.1 (i)) converges in* L^r.

Proof. The conditions give $\mathbf{A}_0 \in L^\infty$ and thereby $\mathrm{E}(\log^+ \|\mathbf{A}_0\|) < \infty$. The inequality $\gamma < 0$ follows from Corollary 3.1. Regarding (ii), it only remains to show in view of Corollary 4.1 that $\lim_{k\to\infty} \|\mathbf{A}_0 \dots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ holds, but this is immediate using Theorem 3.1 as in the proof of Corollary 3.1. ∎

5. GEOMETRIC ERGODICITY AND STRONG MIXING

It is immediate to see that the joint sequence $(\mathbf{X}_t, \Delta_t)$ is a Markov chain. In this section we analyze (V-uniform) geometric ergodicity and strong mixing of

$(\mathbf{X}_t, \Delta_t)$ and thereby of MS-ARMA processes. We start with recalling some notions on Markov chains (for a comprehensive discussion, see Meyn and Tweedie, 1993).

Consider a Markov chain $X = (X_t)_{t \in \mathbb{N}}$ with topological state space S equipped with the Borel σ-algebra $\mathcal{S}$ and denote by $P^n(\cdot, \cdot)$ with $n \in \mathbb{N}$ its n-step transition kernel. The chain X is said to be a weak Feller chain if $\mathrm{E}(g(X_1)|X_0 = y)$ is continuous in $y \in S$ for all bounded and continuous $g : S \to \mathbb{R}$. If μ is some nondegenerate measure on $(S, \mathcal{S})$ and $\mu(A) > 0$ implies $\sum_{n=1}^{\infty} P^n(x, A) > 0$ for all $x \in S$ and $A \in \mathcal{S}$, then X is called *μ-irreducible*. Assume that $V : S \to \mathbb{R}$ is measurable and $V(x) \geq 1\ \forall x \in S$. If there is a probability measure π on $(S, \mathcal{S})$ such that

$$\|P^n - \pi\|_V := \sup_{x \in S} \sup_{g \in F_V} \frac{\left|\int_S g(y)(P^n(x, dy) - \pi(dy))\right|}{V(x)} \to 0 \qquad \text{as } n \to \infty, \quad \mathbf{(5.1)}$$

where $F_V := \{f : S \to \mathbb{R}, \text{ measurable}, |f(x)| \leq V(x)\, \forall x \in S\}$, then the Markov chain X is said to be V-uniformly ergodic. Moreover, V-uniform ergodicity implies geometric ergodicity.

A discrete time stationary stochastic process $X = (X_n)_{n \in \mathbb{Z}}$ is called strongly mixing if

$$\alpha_l := \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_l^{\infty}\} \to 0$$

as $l \to \infty$, where $\mathcal{F}_{-\infty}^0 := \sigma(\ldots, X_{-2}, X_{-1}, X_0)$ and $\mathcal{F}_l^{\infty} = \sigma(X_l, X_{l+1}, X_{l+2}, \ldots)$. The values α_l are called mixing coefficients. If there are constants $C \in \mathbb{R}^+$ and $a \in (0,1)$ such that $\alpha_l \leq Ca^l$, X is said to be strongly mixing with geometric rate. Finally, it should be noted that many results regarding statistical properties hold under strong mixing.

As it is most convenient when analyzing stationary MS-ARMA processes, we have, apart from Theorem 2.1 (i), always considered processes starting in the infinite past so far. The geometric ergodicity results of this section are useful both when $(\mathbf{X}_t, \Delta_t)$ is started in the infinite past and also at time zero with arbitrary initial values $(\mathbf{X}_0, \Delta_0)$.

The next theorem studies the V-uniform ergodicity of MS-ARMA processes. Regarding the topological properties recall Remark 2.1 (a) and observe that it means, in particular, that one cannot use the discrete metric/topology for a countable and nonfinite state space S of Δ, as this contradicts the required compactness. On $\mathbb{R}^{d(p+q)} \times S$ the metric/topology is, of course, understood to be the product metric/topology.

THEOREM 5.1.

(i) *(Geometric ergodicity) Assume that $(\mathbf{X}_t, \Delta_t)$ is a μ-irreducible and aperiodic weak Feller chain, the support of μ has nonempty interior, and the state space S of Δ is compact. If, moreover, there are $\eta \in (0,1]$ and $c < 1$ such that*

$$\mathrm{E}(\|\mathbf{A}_1\|^\eta | \Delta_0 = \delta) \leq c \ \ \forall\, \delta \in S \tag{5.2}$$

for some norm $\|\cdot\|$ *on* $\mathbb{R}^{d(p+q)}$ *and* $\epsilon_1 \in L^\eta$, *then* $(\mathbf{X}_t, \Delta_t)$ *is* V*-uniformly ergodic with* $V : \mathbb{R}^{d(p+q)} \times S \to \mathbb{R}$ *given by* $(x, \delta) \mapsto 1 + \|x\|^\eta$.

(ii) (Infinite past, strong mixing) If $(\Delta_t)_{t\in\mathbb{Z}}$ *is additionally stationary and ergodic, then* $\mathrm{E}(\log^+ \|\mathbf{A}_0\|), \mathrm{E}(\log^+ \|\mathbf{C}_0\|) < \infty$, $\gamma < 0$, *and thus there is a unique stationary and ergodic solution* $X = (X_t)_{t\in\mathbb{Z}}$ *to the MS-ARMA* (p, q, Δ, ϵ) *equation (2.1) given by Theorem 2.1 (i). Moreover,* $(\mathbf{X}_t, \Delta_t)_{t\in\mathbb{Z}}$, *the state space representation* $\mathbf{X}$, *and the MS-ARMA process* X *itself are strongly mixing with geometric rate.*

Proof.

(i) Let ψ denote a maximal irreducibility measure for $(\mathbf{X}_t, \Delta_t)$ in the sense of Meyn and Tweedie (1993, Prop. 4.2.2). Thus, μ is especially absolutely continuous with respect to ψ, i.e., $\psi(A) = 0$ implies $\mu(A) = 0$, and therefore $\operatorname{supp}\psi \supseteq \operatorname{supp}\mu$, which shows that the support of ψ has nonempty interior.

As $0 < \eta \leq 1$, we have $\|a+b\|^\eta \leq \|a\|^\eta + \|b\|^\eta$ for all $a, b \in \mathbb{R}^{d(p+q)}$. Thus, for any $x \in \mathbb{R}^{d(p+q)}$ and $\delta \in S$

$$\mathrm{E}(V(\mathbf{X}_1, \Delta_1) | \mathbf{X}_0 = x, \Delta_0 = \delta) = \mathrm{E}(\|\mathbf{A}_1 x + \mathbf{C}_1\|^\eta + 1 | \mathbf{X}_0 = x, \Delta_0 = \delta)$$

$$\leq \mathrm{E}(\|\mathbf{A}_1\|^\eta | \Delta_0 = \delta)\|x\|^\eta + \mathrm{E}(\|\mathbf{C}_1\|^\eta | \Delta_0 = \delta) + 1,$$

because Δ_1 only depends on Δ_0. As S is compact and $\epsilon_1 \in L^\eta$ and is independent of Δ, there is an $M > 0$ such that $\mathrm{E}(\|\mathbf{C}_1\|^\eta | \Delta_0 = \delta) < M - 1$ for all $\delta \in S$. Hence, $\mathrm{E}(\|\mathbf{X}_1\|^\eta + 1 | \mathbf{X}_0 = x, \Delta_0 = \delta) \leq c\|x\|^\eta + M$. Choose $\tau > 0$ with $1 - \tau > c$ and then set $R = ((M)/(1-\tau-c))^{1/\eta}$ and $C = B_R(0)$ (the ball with radius R in $\mathbb{R}^{d(p+q)}$). For all $x \in C^c = \mathbb{R}^{d(p+q)} \backslash C$ we have $(1-\tau-c)\|x\|^\eta \geq M$, and therefore

$$\mathrm{E}(V(\mathbf{X}_1, \Delta_1) | \mathbf{X}_0 = x, \Delta_0 = \delta)$$

$$\leq c\|x\|^\eta + (1-\tau-c)\|x\|^\eta \leq (1-\tau)V(x, \delta) \tag{5.3}$$

for all $(x, \delta) \in C^c \times S$. Setting $K := C \times S$ we obtain a compact set. Hence, Meyn and Tweedie (1993, Prop. 6.2.8(ii)) ensures that K is a petite set (for a definition, see Meyn and Tweedie, 1993, Sect. 5.5.2). Combining (5.3) with the observation $\mathrm{E}(V(\mathbf{X}_1) | \mathbf{X}_0 = x, \Delta = \delta) \leq c\|x\|^\eta + M \leq (cM)/(1-\tau-c) + M =: b$ for all $x \in C$, we obtain $\mathrm{E}(V(\mathbf{X}_1) | \mathbf{X}_0 = x, \Delta = \delta) \leq (1-\tau)V(x, \delta) + 1_K(x, \delta)b$. An application of Theorem 16.0.1 of Meyn and Tweedie (1993) concludes the proof.

(ii) The compactness of S and $\epsilon_1 \in L^\eta$ ensure the finiteness of $\mathrm{E}(\|\mathbf{C}_0\|^\eta)$ and thus $\mathrm{E}(\log^+ \|\mathbf{C}_0\|)$. Likewise, (5.2) gives $\mathrm{E}(\|\mathbf{A}_1\|^\eta) \leq c$, which implies $\mathrm{E}(\log^+ \|\mathbf{A}_0\|) < \infty$ and $\gamma < 0$. So, there is a unique stationary and ergodic solution $(X_t)_{t\in\mathbb{Z}}$ to the MS-ARMA(p,q,Δ,ϵ) equation (2.1) given by Theorem 2.1 (i). The strong mixing properties are implied by the V-uniform ergodicity (see Meyn and Tweedie, 1993, Ch. 16) and the fact that strong mixing of a joint random sequence $(A_t,C_t)_{t\in\mathbb{Z}}$ implies this property for the individual sequences $(A_t)_{t\in\mathbb{Z}}$ and $(C_t)_{t\in\mathbb{Z}}$, which is obvious from the definition. ■

Remark 5.1.

(a) A straightforward sufficient condition for (5.2) is the existence of a norm $\|\cdot\|$ and $c<1$ such that $\|\mathbf{A}_1\|\le c$ for all possible states of Δ_1. Moreover, Jensen's inequality shows that $\mathrm{E}(\|\mathbf{A}_1\|^\gamma\,|\,\Delta_0=\delta)\le c\ \forall\ \delta\in S$ for some $\gamma\ge 1$ implies the validity of (5.2) for all $\eta\in(0,1]$.

(b) Yao and Attali (2000) gave criteria for geometric ergodicity of nonlinear Markov-switching autoregressions with finitely many regimes, which were extended in Lee (2005).

Next we examine conditions for an MS-ARMA process to be weakly Fellerian.

PROPOSITION 5.1.

(i) Assume that there is some measurable function F such that $\Delta_t = F(\Delta_{t-1},u_t)$, where (u_t) is an i.i.d. sequence assuming values in a measurable space $(G,\mathcal{G})$ and $F(\cdot,u)$ is continuous for any fixed $u\in G$. Then $(\mathbf{X}_t,\Delta_t)$ is a weak Feller chain.

(ii) If $(\mathbf{X}_t,\Delta_t)$ is weakly Fellerian, then (Δ_t) is a weak Feller chain.

Proof.

(i) Because projections are continuous, there are functions $F_{\mathbf{A}}, F_{\mathbf{m}}, F_{\Sigma}$ such that $\mathbf{A}_t=F_{\mathbf{A}}(\Delta_{t-1},u_t)$, $\mathbf{m}_t=F_{\mathbf{m}}(\Delta_{t-1},u_t)$, $\Sigma_t=F_{\Sigma}(\Delta_{t-1},u_t)$, and $F_{\mathbf{A}}$, $F_{\mathbf{m}}$, F_{Σ} are continuous in Δ_{t-1}. Thus, we obtain that

$$(\mathbf{X}_t,\Delta_t)=(F_{\mathbf{A}}(\Delta_{t-1},u_t)\mathbf{X}_{t-1}+F_{\mathbf{m}}(\Delta_{t-1},u_t)$$
$$+F_{\Sigma}(\Delta_{t-1},u_t)\epsilon_t, F(\Delta_{t-1},u_t))$$

is a continuous function of $(\mathbf{X}_{t-1},\Delta_{t-1})$.

Let $g:\mathbb{R}^{d(p+q)}\times S\to\mathbb{R}$ be bounded and continuous and denote $P(\epsilon,u)$ the joint distribution of (ϵ_1,u_1); then

$$\begin{aligned}&\mathrm{E}\,(g(\mathbf{X}_1,\Delta_1)|\mathbf{X}_0=x,\Delta_0=\delta)\\&\quad=\int\limits_{\mathbb{R}^d\times G} g\,(F_{\mathbf{A}}(\delta,u)x+F_{\mathbf{m}}(\delta,u)+F_{\Sigma}(\delta,u)\epsilon, F(\delta,u))\,dP(\epsilon,u)\end{aligned}$$

is a continuous function of (x,δ), as the continuity lemma from standard integration theory (see, e.g., Bauer, 1992, Lem. 16.1) shows.

(ii) Let $g : S \to \mathbb{R}$ be bounded and continuous. Define $\tilde{g} : \mathbb{R}^{d(p+q)} \times S \to \mathbb{R}$ by $\tilde{g}(x, \delta) = g(\delta)$. Then $\tilde{g}$ is bounded and continuous and $\mathrm{E}(g(\Delta_1) | \Delta_0 = \delta) = \mathrm{E}(\tilde{g}(\mathbf{X}_1, \Delta_1) | \mathbf{X}_0 = x, \Delta_0 = \delta)$ is continuous, because Δ_1 only depends on Δ_0 and $(\mathbf{X}_t, \Delta_t)$ is weakly Fellerian. Thus, Δ is a weak Feller chain. ∎

Demanding the existence of such a function F is still a rather weak condition, as many Markov chains are of this type (cf., e.g., Meyn and Tweedie, 1993, Sect. 2.2 and Ch. 7). Compared to the nonlinear state space models studied in Meyn and Tweedie (1993) our assumptions are even weaker, because we do not impose any differentiability restrictions on F.

Now we turn to studying μ-irreducibility and aperiodicity. Denoting the Lebesgue measure on $\mathbb{R}^r$ by λ^r the following proposition covers most cases of practical relevance.

PROPOSITION 5.2. *Let P_Δ^n denote the n-step transition kernel of the Markov chain Δ and μ_Δ be a nondegenerate measure on $(S, \mathcal{S})$ such that for any $A \in \mathcal{S}$ with $\mu_\Delta(A) > 0$ and all $x \in S$*

$$\sum_{n=p+q}^{\infty} P_\Delta^n(x, A) > 0 \tag{5.4}$$

holds. Assume that ϵ_0 has a strictly positive density with respect to λ^d and, moreover, that Σ_t is invertible for all possible states of Δ_t.

(i) Then (Δ_t) is μ_Δ- and $(\mathbf{X}_t, \Delta_t)$ is $\lambda^{d(p+q)} \otimes \mu_\Delta$-irreducible.

(ii) If the support of μ_Δ has nonempty interior, then the same holds for $\lambda^{d(p+q)} \otimes \mu_\Delta$.

(iii) Assume that Δ is also aperiodic; then so is $(\mathbf{X}_t, \Delta_t)$.

Proof. Condition (5.4) immediately implies that Δ is μ_Δ-irreducible. Inspecting the iteration $\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{C}_t$, it is obvious that under the preceding assumptions $\mathbf{X}_{p+q+k}$ can reach any set of positive Lebesgue measure for all $k \in \mathbb{N}_0$ with strictly positive probability regardless of the value $(\mathbf{X}_0, \Delta_0)$ and the evolution of the chain (Δ_t), because ϵ_t has a strictly positive density and Σ_t is invertible. Combining this with the fact that for every set A with positive measure μ_Δ there is an $n \geq p+q$ such that $P_\Delta^n(x, A) > 0$, yields (i).

Part (ii) is now a trivial consequence of Part (i), because we are using the product topology and $\mathrm{supp}\lambda^{d(p+q)} = \mathbb{R}^{d(p+q)}$. Furthermore, the preceding considerations on the sets that $\mathbf{X}_t$ can reach give immediately that $(\mathbf{X}_t, \Delta_t)$ cannot exhibit any cyclic behavior when Δ is aperiodic. This gives Part (iii). ∎

Finally we extend the feasible sufficient stationarity criterion of Corollary 3.1 to one ensuring (5.2). Again, this is an immediate consequence of Theorem 3.1.

PROPOSITION 5.3. *Assume that S is compact and that there is a norm $\|\cdot\|_d$ on $\mathbb{R}^d$ and $\bar{c} < 1$ such that $\sum_{i=1}^{p} \|\Phi_{i1}\|_d \leq \bar{c}$ for all possible states of Δ_1; then there is a norm $\|\cdot\|$ on $\mathbb{R}^{d(p+q)}$ and $c < 1$ with $\|\mathbf{A}_1\| \leq c$ for all possible states of Δ_1. In particular, (5.2) is satisfied for all $\eta \in (0, 1]$.*

Remark 5.2 (Finite state space)**.** For Markov chains with finite state space the usual construction (given, e.g., in Resnick, 1992, Sect. 2.1) automatically implies weak Fellerianity via Proposition 5.1. However, as we may not use the discrete metric, this does not extend to a nonfinite countable state space; then one has to check the continuity at accumulation points of S in detail.

Likewise, we take the counting measure on S as μ_Δ in Proposition 5.2 in the case of a finite state space of Δ, because this conforms with the standard notion of irreducibility. The counting measure always has a nonempty interior of the support. Moreover, elementary arguments show that irreducibility already implies (5.4).

To conclude this paper let us give a concrete example of a Markov-switching process with an uncountable state space for the parameter chain.

Example 5.1

Assume that a Markov-switching AR(1) process (X_t) is given by

$$X_t = \Phi_{1t} X_{t-1} + \epsilon_t, \tag{5.5}$$

where the noise ϵ is an i.i.d. sequence ϵ_t with a standard normal distribution and the parameter chain Φ_{1t} is given as follows.

Let a, b, c be such that $-1 < a < b < 1$ and $c > 0$ and (u_t) be an i.i.d. sequence uniformly distributed on the interval $[-1, 1]$. Then the evolution of the AR coefficient is given by $\Phi_{1t} = \max\left(\min\left(\Phi_{1,t-1} + cu_t, b\right), a\right)$; i.e., we choose the new parameter uniformly from the neighborhood with radius c of the old one but do not allow it to leave the interval $[a; b]$.

Using Corollary 3.1 it is clear that Theorem 2.1 implies the existence of a unique stationary and ergodic solution to (5.5). Likewise, Theorem 4.2 gives that X_t has a finite moment of any order. Moreover, looking at the definition it is immediate that (Φ_{1t}) is aperiodic and irreducible with respect to the Lebesgue measure restricted to $[a, b]$ and that (5.4) is satisfied. Now that we have observed this, Propositions 5.1–5.3 imply that Theorem 5.1 is also applicable and thus the MS-ARMA process is geometrically ergodic/strong mixing.

The easiest way to see that the Markov parameter chain satisfies the conditions needed is, of course, to use Corollary 3.1 or Proposition 5.3. But when these are applicable there are no explosive regimes. However, in applications the presence of explosive regimes is often desirable. To show that models with explosive regimes have some desirable probabilistic properties one can often simply use the general conditions we have given directly. Let us illustrate this with a concrete variant of Example 5.1 that has explosive regimes.

Example 5.2

Let an MS-ARMA process be given by the setup of Example 5.1 with $a = -1.2$, $b = 1.2$, and $c = 1.5$. Then $\mathrm{E}(|\Phi_{1,1}|\,|\Phi_{1,0} = \delta) \leq \mathrm{E}(|\Phi_{1,1}|\,|\,\Phi_{1,0} = 1.2)$ for all $\delta \in [-1.2, 1.2]$ is obvious, and one calculates $\mathrm{E}(|\Phi_{1,1}|\,|\,\Phi_{1,0} = 1.2) = 0.5 \cdot 1.2 + 0.5 \int_{-0.3}^{1.2} (x/1.5) dx = 0.825$.

Hence, condition **(5.2)** is satisfied with $\eta = 1$, and this implies that we have V-uniform ergodicity and strong mixing, because the other conditions of Theorem 5.1(i) are fulfilled using the same arguments as for Example 5.1. This gives immediately that the Markov parameter chain can be chosen to be stationary and ergodic. If this is done, Theorem 5.1(ii) applies and, hence, shows that the MS-ARMA process given by (5.5) is stationary and ergodic.

REFERENCES

An, H.Z. & F.C. Huang (1996) The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6, 943–956.

Ash, R.B. & M.F. Gardner (1975) *Topics in Stochastic Processes*. Academic Press.

Bauer, H. (1992) *Maß- und Integrationstheorie*, 2nd ed. de Gruyter.

Bougerol, P. & Picard, N. (1992) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1730.

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

Brandt, A., Franken, P. & Lisek, B. (1990) *Stationary Stochastic Models*. Wiley.

Brockwell, P.J. & Davis, R.A. (1991) *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag.

Douc, R., E. Moulines, & T. Rydén (2004) Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Annals of Statistics* 32, 2254–2304.

Doucet, A., Logothetis, A. & Krishnamurthy, V. (2000) Stochastic sampling algorithms for state estimation of jump Markov linear systems. *IEEE Transactions on Automatic Control* 45, 188–202.

Feigin, P.D. & Tweedie, R.L. (1985) Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* 6, 1–14.

Francq, C. & Zakoïan, J.-M. (2001) Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics* 102, 339–364.

Furstenberg, H. & H. Kesten (1960) Products of random matrices. *Annals of Mathematical Statistics* 31, 457–469.

Goldie, C.M. & R.A. Maller (2000) Stability of perpetuities. *Annals of Probability* 28, 1195–1218.

Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

Hamilton, J.D. (1990) Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.

Hamilton, J.D. & Raj, B. (eds.) (2002) *Advances in Markov-Switching Models—Applications in Business Cycle Research and Finance*. Physica-Verlag.

Hille, E. & Phillips, R.S. (1957) *Functional Analysis and Semi-Groups*, rev. ed. American Mathematical Society.

Karlsen, H.A. (1990) Existence of moments in a stationary stochastic difference equation. *Advances in Applied Probability* 22, 129–146.

Klüppelberg, C. & Pergamenchtchikov, S. (2004) The tail of the stationary distribution of a random coefficient AR(q) model. *Annals of Applied Probability* 14, 971–1005.

Krengel, U. (1985) *Ergodic Theorems*. de Gruyter.

Krolzig, H.-M. (1997) *Markov-Switching Vector Autoregressions*. Lecture Notes in Economics and Mathematical Systems 454. Springer-Verlag.

Lee, O. (2005) Probabilistic properties of a nonlinear ARMA process with Markov switching. *Communications in Statistics—Theory and Methods* 34, 193–204.

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

Nicholls, D.F. & Quinn, B.G. (1982) *Random Coefficient Autoregressive Models: An Introduction*. Lecture Notes in Statistics 11. Springer-Verlag.

Resnick, S.I. (1992) *Adventures in Stochastic Processes*. Birkhäuser.

Stelzer, R. (2008) Multivariate Markov-switching ARMA processes with regularly varying noise. *Journal of Multivariate Analysis* 99, 1177–1190.

Tugnait, J.K. (1982) Adaptive estimation and identification for discrete systems with Markov jump parameters. *IEEE Transactions on Automatic Control* 27, 1054–1065. Correction Published in 1984, vol. 29, 1984, p. 286.

Yao, J. (2001) On square-integrability of an AR process with Markov switching. *Statistics and Probability Letters* 52, 265–270.

Yao, J.F. & J.G. Attali (2000) On stability of nonlinear AR processes with Markov switching. *Advances in Applied Probability* 32, 394–407.