ELSEVIER

Journal of Econometrics 102 (2001) 339–364

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Stationarity of multivariate Markov–switching ARMA models

C. Francq[a], J.-M. Zakoïan[b, *]

[a] *Laboratoire de Mathématiques Pures et Appliquées Joseph Liouville, Université du Littoral, 62228 Calais, France*

[b] *GREMARS, Université Lille 3 and CREST, 15 Boulevard Gabriel Péri, F-92245 Malakoff Cedex, France*

Received 11 February 1999; revised 2 October 2000; accepted 23 January 2001

Abstract

In this article we consider multivariate ARMA models subject to Markov switching. In these models, the parameters are allowed to depend on the state of an unobserved Markov chain. A natural idea when estimating these models is to impose local stationarity conditions, i.e. stationarity within each regime. In this article we show that the local stationarity of the observed process is neither sufficient nor necessary to obtain the global stationarity. We derive stationarity conditions and we compute the autocovariance function of this nonlinear process. Interestingly, it turns out that the autocovariance structure coincides with that of a standard ARMA. Some examples are proposed to illustrate the stationarity conditions. Using Monte Carlo simulations we investigate the consequences of accounting for the stationarity conditions in statistical inference.

JEL classification: C32

Keywords: Multivariate ARMA models; Regime-switching models; Markov-switching models; Strict and second-order stationary time series

* Corresponding author. Tel.: +33-01-4117-7825; fax: +33-01-4117-7666.

E-mail addresses: francq@lmpa.univ-littoral.fr (C. Francq), zakoian@ensae.fr (J.M. Zakoïan).

PII: S0304-4076(01)00057-4

1. Introduction

In this paper we are interested in multivariate ARMA models with random coefficients. The general class of models is of the following type:

$$X_t = c(\Delta_t) + \sum_{i=1}^{p} a_i(\Delta_t)X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j(\Delta_t)\varepsilon_{t-j}, \tag{1}$$

where X_t is a random vector with values in $\mathbb{R}^K$, (Δ_t) is an irreducible, aperiodic, Markov chain with finite state-space $\mathscr{E}=\{1,2,\ldots,d\}$ and stationary transition probabilities denoted $p(i,j):=\mathrm{pr}(\Delta_t=j|\Delta_{t-1}=i)$, the $a_i(\Delta_t)$ and $b_j(\Delta_t)$ are $K\times K$ real random matrices, $c(\Delta_t)$ is a $K\times 1$ vector. It is assumed that (Δ_t) is stationary, with stationary probabilities denoted $\pi(i):=\mathrm{pr}(\Delta_1=i)$, $1\leqslant i\leqslant d$.

Write M' for the transpose of any matrix M. To allow for the possibility of change in variance, we assume that

$$\varepsilon_t=\varepsilon_t(\Delta_t)=\sigma(\Delta_t)\eta_t$$

where $\sigma(\Delta_t)$ is a $K\times K$ random matrix and (η_t) is a stationary and ergodic sequence of K-dimensional centered and uncorrelated variables, with $\mathrm{E}(\eta_t\eta_t')=\Omega$, the covariance matrix Ω being nonsingular. In addition, assume that (η_t) is independent of (Δ_t). Hence (ε_t) is a white noise. A statistical analysis of this process has been proposed in the univariate case by Francq and Roussignol (1997). Given the state, or regime, of the chain, say $\{\Delta_t=k\}$, $k\in\mathscr{E}$, X_t satisfies a vector ARMA(p,q) equation with coefficients $c(k)$, $a_i(k)$, $1\leqslant i\leqslant p$, $b_j(k)$, $1\leqslant j\leqslant q$.

Model (1) can be viewed as a Markov mixture of dynamic models. In the last decades, general mixture models have been extensively studied in the statistical literature (see e.g. the monograph by Titterington et al. (1985)). In the statistical applications, neither the noise (ε_t) nor the Markov chain (Δ_t) are observed and the latter is said to be hidden. Specifically, model (1) belongs to the class of Markov-switching models in which an increasing interest has been shown in the recent econometric literature since the seminal paper by Hamilton (1989) (see also Hamilton, 1988, 1990, 1994; McCulloch and Tsay, 1994; Chib, 1996). Such models extend the class of hidden Markov models, introduced by Baum and Petrie (1966) and analyzed by many authors (see e.g. Rabiner and Juang, 1986; Leroux, 1992; Rydèn, 1994, Elliott et al., 1995), in which the observations are assumed to be independent conditional on a hidden Markov chain. As far as modeling issues are concerned, the main advantages of model (1) are its flexibility and ability to approximate a broad range of different dynamics. It can be used to provide dynamics with frequent regime changes and, likewise, it can account for sudden unusual changes. However the statistical inference is delicate. Due to the presence of a MA structure, standard recursive algorithms (see e.g. Hamilton, 1989; Holst et al., 1994) fail to compute the likelihood functions. Recently, following the

Bayesian approach developed by Albert and Chib (1993) for the Hamilton (1989) model, a method based on MCMC techniques (see e.g. Kim and Nelson (1999) or Casella and Robert (1999)) for estimating switching ARMA models has been proposed by Billio et al. (1999). Other approaches based on simulated EM algorithms (Shephard, 1994) or the partial Kalman filter (Billio and Monfort, 1995) can also be considered.

A key assumption common to these inference procedures is that of stationarity. Therefore, a natural idea used by some of the above-mentioned authors, is to impose the classical stationarity constraints (i.e. the roots of the AR polynomials are outside the unit circle) within *each* regime. In effect, it seems natural to think that local stationarity should imply global stationarity. Unfortunately, this is not true as we will see below. Furthermore, these local stationarity restrictions will be too strong in other cases: imposing them would constitute a serious limitation to the range of dynamics permitted by the class of models.

The main purpose of this paper is to give conditions for the existence of strictly and second-order stationary Markov-switching ARMA processes, which do not require the local stationarity. This is particularly interesting for it allows to mimic typical behaviors with explosive phases followed by quiet periods. Section 2 provides a Markovian representation, which is used to derive a condition for strict stationarity. In Section 3 a discussion of the second-order stationarity properties of the models leads to necessary and sufficient conditions. The autocovariance structure is analyzed in Section 4. Finally, examples, illustrations and a Monte Carlo study are proposed in the last two sections.

2. Markovian representation and strict stationarity

In order to investigate the stationarity properties of the process (X_t), we use the following vectorial form of the model:

$$\underline{z}_t = \underline{\omega}_t + \Phi_t \underline{z}_{t-1}, \tag{2}$$

where $\underline{\omega}_t = \underline{c}_t + \underline{\varepsilon}_t = \underline{c}_t + \Sigma_t \eta_t$

$$\underline{c}_t = \begin{pmatrix} c(\Delta_t) \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{K(p+q)}, \quad \underline{z}_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q+1} \end{pmatrix} \in \mathbb{R}^{K(p+q)},$$

$$\Sigma_t = \begin{pmatrix} \sigma(\Delta_t) \\ 0 \\ \vdots \\ 0 \\ \sigma(\Delta_t) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is a $K(p+q)\times K$ matrix and

$$\Phi_t = \begin{pmatrix} a_1(\Delta_t) & & \cdots & a_p(\Delta_t) & b_1(\Delta_t) & & \cdots & & b_q(\Delta_t) \\ I_K & 0 & \cdots & 0 & 0 & & \cdots & & 0 \\ 0 & I_K & \cdots & 0 & 0 & & \cdots & & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \ldots\ I_K & 0 & 0 & & \ldots & 0 & 0 \\ 0 & & \cdots & 0 & 0 & & \cdots & & 0 \\ 0 & & \cdots & 0 & I_K & 0 & \cdots & & 0 \\ 0 & & \cdots & 0 & 0 & I_K & \cdots & & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \ldots\ 0 & 0 & 0 & & \ldots & I_K & 0 \end{pmatrix}$$

is a $K(p+q)\times K(p+q)$ matrix.

In this vectorial representation, we have implicitly assumed that $p \geqslant 1$, $q \geqslant 1$, without loss of generality because $a_p(\cdot)$ and $b_q(\cdot)$ can be equal to zero in (1).

Since (Δ_t, η_t) is an ergodic process, $(\Phi_t, \underline{\omega}_t)$ is also a strictly stationary ergodic process. Let $||\cdot||$ denote any operator norm on the sets of $K(p+q)\times K(p+q)$ and $K(p+q)\times 1$ matrices and let, for $x>0$, $\log^+ x = \max\{\log x, 0\}$. It is clear that both $\mathrm{E}\log^+||\Phi_t||$ and $\mathrm{E}\log^+||\underline{\omega}_t||$ are finite. Therefore, from Brandt (1986) (see Bougerol and Picard, 1992), the unique stationary solution of (2) is given by

$$\underline{z}_t = \underline{\omega}_t + \sum_{k=1}^{\infty} \Phi_t\Phi_{t-1}\cdots\Phi_{t-k+1}\underline{\omega}_{t-k}, \tag{3}$$

whenever the top Lyapunov exponent γ defined by

$$\gamma = \inf_{t\in\mathbb{N}^*}\left\{E\frac{1}{t}\log||\Phi_t\Phi_{t-1}\cdots\Phi_1||\right\}$$

is strictly negative.

Write Φ_t as

$$\Phi_t = \begin{pmatrix} A_t & B_t \\ 0 & J \end{pmatrix},$$

where A_t and B_t are $Kp \times Kp$ and $Kp \times Kq$ matrices, respectively. We find that

$$\Phi_t \Phi_{t-1} \cdots \Phi_{t-k+1} = \begin{pmatrix} A_t A_{t-1} \ldots A_{t-k+1} & \sum_{l=0}^{k-1} A_t A_{t-1} \ldots A_{t-k+l+2} B_{t-k+l+1} J^l \\ 0 & J^k \end{pmatrix} \tag{4}$$

with by convention $J^0 = I_{Kq}$. Note that $J^q = 0$. If we let $\tilde{\Phi}_t$ denote the matrix obtained by replacing B_t and J by 0 in Φ_t we then have, for $k \geqslant q$

$$\Phi_t \Phi_{t-1} \cdots \Phi_{t-k+1} = \tilde{\Phi}_t \tilde{\Phi}_{t-1} \cdots \tilde{\Phi}_{t-(k-q)+1} \Phi_{t-(k-q)} \cdots \Phi_{t-k+1}.$$

Similarly, we have

$$\tilde{\Phi}_t \tilde{\Phi}_{t-1} \cdots \tilde{\Phi}_{t-k+1} = \Phi_t \Phi_{t-1} \cdots \Phi_{t-(k-q)+1} \tilde{\Phi}_{t-(k-q)} \cdots \tilde{\Phi}_{t-k+1}.$$

Therefore, because the top Lyapunov exponent is independent of the norm, by choosing a multiplicative norm it is straightforward to obtain that

$$\gamma = \tilde{\gamma} := \inf_{t \in \mathbb{N}^*} \left\{ E \frac{1}{t} \log ||\tilde{\Phi}_t \tilde{\Phi}_{t-1} \cdots \tilde{\Phi}_1|| \right\} = \inf_{t \in \mathbb{N}^*} \left\{ E \frac{1}{t} \log ||A_t A_{t-1} \cdots A_1|| \right\}.$$

We have shown our first result which gives a sufficient condition for strict stationarity.

Theorem 1. Suppose that $\tilde{\gamma} < 0$. Then for all $t \in \mathbb{Z}$ the series

$$\underline{z}_t = \underline{\omega}_t + \sum_{k=1}^{\infty} \Phi_t \Phi_{t-1} \cdots \Phi_{t-k+1} \underline{\omega}_{t-k}$$

converges a.s. and the process (X_t), *defined as the first component of* $\underline{z}_t$, *is the unique strictly stationary solution of* (1).

It is seen from this theorem that, as in the standard ARMA models, the key element governing strict stationarity is the autoregressive part of the model. In particular, observe that in the case of a univariate ($K = 1$) ARMA$(1, q)$, a sufficient condition for the existence of a strictly stationary solution is that

$$\sum_{i=1}^{d} \pi(i) \log |a_1(i)| < 0. \tag{5}$$

It is worth noting that the existence of explosive regimes (i.e. $|a_1(i)| > 1$) does not preclude strict stationarity. In contrast with standard ARMA processes, however, the fact that the sufficient condition is also necessary is an open issue. Note that the technique introduced by Bougerol and Picard (1992) for establishing a converse to Theorem 1 in the single-regime case, cannot be used because the sequence $(\Phi_t, \underline{\omega}_t)$ is not independent. In the next section we give conditions for the existence of second-order stationary solutions.

3. Necessary and sufficient second-order stationarity conditions

The problem of finding conditions ensuring second-order stationarity has been addressed by Tjøstheim (1986), Karlsen (1990), Francq and Roussignol (1998), Yao and Attali (1998) and Yang (2000) in the case of autoregressive-type Markov-switching models. General results on the existence of second-order moments for Markov-switching ARMA(p,q) models have not been obtained, to our knowledge. To verify that the process

$$\underline{z}_t = \underline{\omega}_t + \Phi_t\underline{\omega}_{t-1} + \Phi_t\Phi_{t-1}\underline{\omega}_{t-2} + \cdots \tag{6}$$

is well-defined in L^2, it is sufficient to show that

$$\underline{z}_{t,k} := \Phi_t\Phi_{t-1}\cdots\Phi_{t-k+1}\underline{\omega}_{t-k}$$

(with $\underline{z}_{t,0} = \underline{\omega}_t$) converges to zero in L^2 (endowed with any matrix norm), at an exponential rate, as k goes to infinity. Indeed, we have $\|\underline{z}_t\|_{L^2} \leqslant \sum_{k=0}^{\infty}\|\underline{z}_{t,k}\|_{L^2}$.

In this paper we are only interested in solutions of (1) such that there exists a measurable function f with $X_t = f(\Delta_t, \eta_t, \Delta_{t-1}, \eta_{t-1}, \ldots)$ *a.s.* for all $t \in \{0, \pm 1, \ldots\}$. Such solutions are called *nonanticipative*. Note that the existence of a nonanticipative stationary solution of (1) is equivalent to the existence of a nonanticipative stationary solution of (2).

Let $\otimes$ denote the matrix tensor product. From the independence between (η_t) and (Δ_t) and the fact that $\mathrm{E}\eta_t = 0$, we have

$$\begin{aligned}
&\mathrm{E}\{\mathrm{vec}(\underline{z}_{t,k}\underline{z}'_{t,k})\}\\
&\quad = \mathrm{E}\{(\Phi_t \otimes \Phi_t)\cdots(\Phi_{t-k+1} \otimes \Phi_{t-k+1})(\underline{\omega}_{t-k} \otimes \underline{\omega}_{t-k})\}\\
&\quad = E\{(\Phi_t \otimes \Phi_t)\cdots(\Phi_{t-k+1} \otimes \Phi_{t-k+1})(\Sigma_{t-k} \otimes \Sigma_{t-k})\}\mathrm{E}\{\eta_{t-k} \otimes \eta_{t-k}\}\\
&\qquad + E\{(\Phi_t \otimes \Phi_t)\cdots(\Phi_{t-k+1} \otimes \Phi_{t-k+1})(\underline{c}_{t-k} \otimes \underline{c}_{t-k})\}.
\end{aligned} \tag{7}$$

Let $\Phi(k)$ (resp. $\underline{c}(k)$) (resp. $\Sigma(k)$) be the matrix obtained by replacing Δ_t by k in Φ_t (resp. $\underline{c}_t$) (resp. Σ_t). Let the $dK^2(p+q)^2 \times dK^2(p+q)^2$ matrix

$$P = \begin{pmatrix}
p(1,1)\{\Phi(1)\otimes\Phi(1)\} & p(2,1)\{\Phi(1)\otimes\Phi(1)\} & \cdots & p(d,1)\{\Phi(1)\otimes\Phi(1)\}\\
p(1,2)\{\Phi(2)\otimes\Phi(2)\} & p(2,2)\{\Phi(2)\otimes\Phi(2)\} & \cdots & p(d,2)\{\Phi(2)\otimes\Phi(2)\}\\
\vdots & \vdots & & \vdots\\
p(1,d)\{\Phi(d)\otimes\Phi(d)\} & p(2,d)\{\Phi(d)\otimes\Phi(d)\} & \cdots & p(d,d)\{\Phi(d)\otimes\Phi(d)\}
\end{pmatrix}$$

and the $dK^2(p+q)^2 \times K^2(p+q)^2$ and $dK^2(p+q)^2 \times 1$ matrices

$$S = \begin{pmatrix} \pi(1)\{\Sigma(1)\otimes\Sigma(1)\} \\ \pi(2)\{\Sigma(2)\otimes\Sigma(2)\} \\ \vdots \\ \pi(d)\{\Sigma(d)\otimes\Sigma(d)\} \end{pmatrix}, \quad C = \begin{pmatrix} \pi(1)\{\underline{c}(1)\otimes\underline{c}(1)\} \\ \pi(2)\{\underline{c}(2)\otimes\underline{c}(2)\} \\ \vdots \\ \pi(d)\{\underline{c}(d)\otimes\underline{c}(d)\} \end{pmatrix}.$$

We have

$$\mathrm{E}\{\mathrm{vec}(\underline{z}_{t,k}\underline{z}'_{t,k})\} = \mathbb{1}' P^k (S \ \mathrm{vec}\, \Omega + C), \tag{8}$$

where

$$\mathbb{1}' = (I_{K^2(p+q)^2}, \dots, I_{K^2(p+q)^2}) \quad \text{and} \quad I_{K^2(p+q)^2}$$

are, respectively, a $K^2(p+q)^2 \times dK^2(p+q)^2$ matrix and the $K^2(p+q)^2 \times K^2(p+q)^2$ identity matrix.

Now let $\rho(M)$ denote the spectral radius of any matrix M ($\rho(M) = \max_{\lambda \in E} |\lambda|$, where E is the set of eigenvalues of the matrix M) and let $||\cdot||$ denote the matrix norm $||M|| = \sum |m_{ij}|$. If $\rho(M)$ is strictly less than 1 then it can be seen, by using the Jordan decomposition, that $||M^k||$ converges to zero at an exponential rate as k goes to infinity. Therefore, because the matrix norm is multiplicative, we have

$$||\underline{z}_{t,k}||_{L^2} \leqslant ||\mathbb{1}' P^k (S\,\mathrm{vec}\,\Omega + C)||^{1/2} \leqslant ||\mathbb{1}'||^{1/2} ||P^k||^{1/2} ||S\,\mathrm{vec}\,\Omega + C||^{1/2}.$$

Hence, a sufficient condition for second-order stationarity is that

$$\rho(P) < 1.$$

Now let $A(k)$ be the matrix obtained by replacing Δ_t by k in A_t and let the $dK^2p^2 \times dK^2p^2$ matrix

$$\tilde{P} = \begin{pmatrix} p(1,1)\{A(1)\otimes A(1)\} & p(2,1)\{A(1)\otimes A(1)\} & \cdots & p(d,1)\{A(1)\otimes A(1)\} \\ p(1,2)\{A(2)\otimes A(2)\} & p(2,2)\{A(2)\otimes A(2)\} & \cdots & p(d,2)\{A(2)\otimes A(2)\} \\ \vdots & \vdots & & \vdots \\ p(1,d)\{A(d)\otimes A(d)\} & p(2,d)\{A(d)\otimes A(d)\} & \cdots & p(d,d)\{A(d)\otimes A(d)\} \end{pmatrix}.$$

In the appendix we show that the nonzero eigenvalues of matrices P and $\tilde{P}$ are the same. Hence, the MA part in (1) does not matter for the sufficient stationarity condition.

Now, let us assume that $(\underline{z}_t)$ is a second-order stationary nonanticipative solution of (2). We have

$$\underline{z}_t = \underline{\varepsilon}_t^{(k)} + \underline{c}_t^{(k)} + \Phi_t \cdots \Phi_{t-k}\underline{z}_{t-k-1} \tag{9}$$

with

$$\underline{\varepsilon}_t^{(k)} = \underline{\varepsilon}_t + \Phi_t \underline{\varepsilon}_{t-1} + \cdots + \Phi_t \cdots \Phi_{t-k+1} \underline{\varepsilon}_{t-k}$$

and

$$\underline{c}_t^{(k)} = \underline{c}_t + \Phi_t \underline{c}_{t-1} + \cdots + \Phi_t \cdots \Phi_{t-k+1} \underline{c}_{t-k}.$$

Note that $\underline{\varepsilon}_t^{(k)}$ and $\underline{c}_t^{(k)} + \Phi_t \cdots \Phi_{t-k} \underline{z}_{t-k-1}$ are uncorrelated and that $\mathrm{E}(\underline{\varepsilon}_t^{(k)}) = 0$. Hence, for all $k \geqslant 0$,

$$\begin{aligned} \mathrm{E}(\underline{z}_t' \underline{z}_t) \ \geqslant \ \mathrm{E}(\underline{\varepsilon}_t^{(k)\prime} \underline{\varepsilon}_t^{(k)}) &= \mathrm{E}(\underline{\varepsilon}_t' \underline{\varepsilon}_t) + \mathrm{E}(\underline{\varepsilon}_{t-1}' \Phi_t' \Phi_t \underline{\varepsilon}_{t-1}) \\ &+ \cdots + \mathrm{E}(\underline{\varepsilon}_{t-k}' \Phi_{t-k+1}' \cdots \Phi_t' \Phi_t \cdots \Phi_{t-k+1} \underline{\varepsilon}_{t-k}). \end{aligned}$$

Hence, we must have

$$\sum_{k=0}^{\infty} ||\Phi_t \cdots \Phi_{t-k+1} \underline{\varepsilon}_{t-k}||_{L^2}^2 < \infty,$$

which implies that

$$\sum_{k=0}^{\infty} ||E(\Phi_t \cdots \Phi_{t-k+1} \underline{\varepsilon}_{t-k} \underline{\varepsilon}_{t-k}' \Phi_{t-k+1}' \cdots \Phi_t')|| = \sum_{k=0}^{\infty} ||\mathbb{1}' P^k S \operatorname{vec} \Omega|| < \infty. \tag{10}$$

We will show that the converse is true when $c(\cdot) = 0$ in (1), i.e. that (10) is a necessary and sufficient condition for the existence of a second-order stationary nonanticipative solution in that particular case. In view of (6), we have that $\underline{z}_t \in L^2$ if and only if

$$\left|\left| \sum_{k=n}^{m} \underline{z}_{t,k} \right|\right|_{L^2} \to 0 \quad \text{as} \quad n \to \infty, \ n \leqslant m.$$

Moreover, we have

$$\left|\left| \sum_{k=n}^{m} \underline{z}_{t,k} \right|\right|_{L^2}^2 = \sum_{k=n}^{m} ||\underline{z}_{t,k}||_{L^2}^2$$

because the $\underline{z}_{t,k}$'s are centered and uncorrelated when $c(\cdot) = 0$. The conclusion follows from the fact that

$$\sum_{k=n}^{m} ||\underline{z}_{t,k}||_{L^2}^2 \leqslant \sum_{k=n}^{m} ||\mathbb{1}' P^k S \ \operatorname{vec} \Omega||$$

which goes to zero as m and n go to infinity, when (10) holds.

The results on the second-order stationarity are collected in the following theorem.

Theorem 2. Suppose that

$$\rho(\tilde{P}) < 1. \tag{11}$$

Then, for all $t \in \mathbb{Z}$, *the series*

$$\underline{z}_t = \underline{\omega}_t + \sum_{k=1}^{\infty} \Phi_t \Phi_{t-1} \cdots \Phi_{t-k+1} \underline{\omega}_{t-k} \tag{12}$$

converges in L^2 *and the process* (X_t), *defined as the first component of* $\underline{z}_t$, *is the unique nonanticipative second-order stationary solution of* (1).

Suppose that (2) *admits a nonanticipative second-order stationary solution. Then we have*

$$\sum_{k=0}^{\infty} \|\mathbb{I}' P^k S \; vec\, \Omega\| < \infty. \tag{13}$$

Finally, if $c(\cdot) = 0$ *in* (1), *a necessary and sufficient condition for the existence of a nonanticipative second-order stationary solution to* (1) *is given by* (13).

Remark 1. The authors mentioned at the beginning of this section also obtained sufficient stationarity conditions for general classes of processes, including the Markov-switching AR of this paper. However, when applied to the AR(p) model with $p > 1$, the comparison of these conditions with our sufficient condition (11) reveals that the former are more restrictive. Moreover, these results do not apply to ARMA models.

Remark 2. Examples presented in the last section will show that the sufficient condition (11) is also necessary in important particular cases, though not in the general case.

Remark 3. As pointed out in Section 2 for the strict stationarity, the sufficient condition (11) is only constraining the autoregressive part (i.e. the coefficients $a_i(k)$) of the model. However, this condition is not necessary in general, and a remarkable point is that the presence of MA terms can modify the stationary region of the AR coefficients. The same remark holds concerning the intercepts of the model. See Examples 6 and 7 below.

Remark 4. In contrast with the strict stationarity condition, (11) provides a rather easy route for verifying whether a solution of (1) is second-order stationary. The eigenvalues computation can often be avoided since (11) holds when the sum of the absolute terms of each line of $\tilde{P}$ is strictly less than one.

4. Computation of the second-order moments

Once second-order stationarity is ensured, it can be useful to compute the expectation and the autocovariance function of (X_t). Throughout the section, (11) is assumed to hold.

4.1. Expectation and variance

Starting from (2) we have

$$\begin{aligned}\pi(k)\mathrm{E}(\underline{z}_t|\Delta_t=k) &= \pi(k)\underline{c}(k)+\Phi(k)\pi(k)\mathrm{E}(\underline{z}_{t-1}|\Delta_t=k)\\ &= \pi(k)\underline{c}(k)+\Phi(k)\sum_{j=1}^{d}\mathrm{E}(\underline{z}_{t-1}|\Delta_{t-1}=j)p(j,k)\pi(j).\end{aligned}$$

Let $U=(\mathrm{E}(\underline{z}_t'|\Delta_t=1)\pi(1),\dots,\mathrm{E}(\underline{z}_t'|\Delta_t=d)\pi(d))'$, $\underline{c}=(\pi(1)\underline{c}'(1),\dots,\pi(d)\underline{c}'(d))'$. We have

$$U=P^*U+\underline{c},$$

where P^* is obtained by replacing the matrices $\Phi(k)\otimes\Phi(k)$ by $\Phi(k)$ in the definition of P. Therefore, provided $I_{dK(p+q)}-P^*$ is invertible, we have

$$U=(I_{dK(p+q)}-P^*)^{-1}\underline{c}.$$

The expectation of X_t is then obtained by

$$\mathrm{E}(X_t)=(e'\otimes f')U, \tag{14}$$

where $e=(1,\dots,1)'\in\mathbb{R}^d$ and $f'=(I_K,0,\dots,0)$ is a $K\times K(p+q)$ matrix. Note that, in the particular case where $c(\cdot)=0$ in (1), the process (X_t) is centered.

Another way to obtain the expectation of X_t, which does not require the invertibility of $I_{dK(p+q)}-P^*$, is by using the expansion in (6). Indeed, because (11) implies the L^2 convergence of the partial sums of the $\underline{z}_{t,k}$'s, we have $\mathrm{E}(\underline{z}_t)=\sum_{k=0}^{\infty}\mathrm{E}(\underline{z}_{t,k})$. Moreover, a formula similar to (8) holds for $\mathrm{E}(\underline{z}_{t,k})$, namely $\mathrm{E}(\underline{z}_{t,k})=(I_{K(p+q)},\dots,I_{K(p+q)})P^{*k}\underline{c}$. When $P^{*k}\to 0$ we can write

$$\begin{aligned}\mathrm{E}(\underline{z}_t) &= (I_{K(p+q)},\dots,I_{K(p+q)})\left(\sum_{k=0}^{\infty}P^{*k}\right)\underline{c}\\ &= (I_{K(p+q)},\dots,I_{K(p+q)})(I_{dK(p+q)}-P^*)^{-1}\underline{c}\end{aligned}$$

from which (14) can be deduced. On the other hand, when P^{*k} does not tend to 0 (in particular if $I_{dK(p+q)}-P^*$ is a singular matrix), we just have

$$\mathrm{E}(\underline{z}_t)=\sum_{k=0}^{\infty}(I_{K(p+q)},\dots,I_{K(p+q)})P^{*k}\underline{c},$$

which gives the expectation of X_t, though not in closed form.

Now, turning to the second-order moments, similar computations show that

$$\begin{aligned}&\pi(k)\mathrm{E}\{\mathrm{vec}(\underline{z}_t\underline{z}_t')|\Delta_t=k\}\\ &\quad=\pi(k)[\mathrm{vec}\{\underline{c}(k)\underline{c}(k)'\}+\{\Sigma(k)\otimes\Sigma(k)\}\mathrm{vec}\,\Omega]\end{aligned}$$

$$+\{\Phi(k)\otimes \underline{c}(k)+\underline{c}(k)\otimes \Phi(k)\}\sum_{j=1}^{d} p(j,k)\pi(j)\mathrm{E}(\underline{z}_{t-1}|\Delta_{t-1}=j)$$

$$+\{\Phi(k)\otimes \Phi(k)\}\sum_{j=1}^{d} p(j,k)\pi(j)\mathrm{E}\{\mathrm{vec}(\underline{z}_{t-1}\underline{z}'_{t-1})|\Delta_{t-1}=j\}.$$

Let V defined as U but with $\underline{z}_t$ replaced by $\mathrm{vec}(\underline{z}_t\underline{z}'_t)$. Let D, defined by replacing the terms $\Phi(k)\otimes\Phi(k)$ by $\underline{c}(k)\otimes\Phi(k)+\Phi(k)\otimes\underline{c}(k)$ in P. Note that $I_{dK^2(p+q)^2}-P$ is an invertible matrix under (11). We then have

$$V=(I_{dK^2(p+q)^2}-P)^{-1}[C+S\ \mathrm{vec}\,\Omega+DU]$$

from which we can compute

$$\mathrm{vec}(\mathrm{E}X_tX'_t)=(e'\otimes f'\otimes f')V,$$

The variance of X_t is easily deduced.

4.2. Autocovariance function

Let us compute the autocovariance function of X_t. For all $h\geqslant 0$, let $W(h)$ be the matrix of size $dK(p+q)\times K(p+q)$ whose kth block $(k=1,\dots,d)$ is the $K(p+q)\times K(p+q)$ matrix $\pi(k)\mathrm{E}(\underline{z}_t\underline{z}'_{t-h}|\Delta_t=k)$. We have, for $h>0$,

$$\pi(k)\mathrm{E}(\underline{z}_t\underline{z}'_{t-h}|\Delta_t=k) = \pi(k)\underline{c}(k)\mathrm{E}(\underline{z}'_{t-h}|\Delta_t=k)$$

$$+\sum_{j=1}^{d}\mathrm{E}\{\Phi(k)\underline{z}_{t-1}\underline{z}'_{t-h}|\Delta_{t-1}=j\}p(j,k)\pi(j).$$

Moreover,

$$\pi(k)\mathrm{E}\{\underline{c}(k)\underline{z}'_{t-h}|\Delta_t=k\}=\sum_{j=1}^{d}\mathrm{E}\{\underline{c}(k)\underline{z}'_{t-h}|\Delta_{t-h}=j\}p^{(h)}(j,k)\pi(j),$$

where $p^{(h)}(j,k)=P(\Delta_t=k|\Delta_{t-h}=j)$. Let $\mathbb{P}$ denote the transition matrix of the chain (Δ_t). Writing $Y(h)$ (resp. $\tilde{U}$) the matrix obtained by replacing $\mathrm{E}(\underline{z}_t\underline{z}'_{t-h}|\Delta_t=k)$ by $\mathrm{E}\{\underline{c}(k)\underline{z}'_{t-h}|\Delta_t=k\}$ (resp. $\mathrm{E}(\underline{z}'_t|\Delta_t=k)$) in $W(h)$, we have

$$Y(h)=\dot{P}(\mathbb{P}')^{h-1}\tilde{U},$$

where $\dot{P}$ is obtained by replacing $\Phi(k)\otimes\Phi(k)$ by $\underline{c}(k)$ in P. Finally we have, for all $h>0$

$$W(h)=Y(h)+P^*W(h-1)$$

and

$$\mathrm{E}(X_tX'_{t-h})=(e'\otimes f')W(h)f \tag{15}$$

The autocovariance function of (X_t) follows.

4.3. ARMA representation

Now we aim to show that the (X_t) process defined by (1) is also a standard ARMA. This is not very surprising: a variety of examples of nonlinear processes admitting ARMA representations in the weak sense (i.e. under weak assumptions on the noise structure) can be found in Francq and Zakoïan (1998). It is clearly sufficient to verify that the autocovariance structure of $(\underline{z}_t)$ is that of an ARMA. For computational simplicity, let us assume that $c(\cdot)=0$. Then we have $Y(h)=0$ and $W(h)=P^{*h}W(0)$. Using the Jordan decomposition (see e.g. Lancaster and Tismenetsky, 1985), it can be seen that $W(h)$ has the following form:

$$W(h)=\sum_{i=1}^{k}\sum_{j=0}^{r_i-1} h^j\lambda_i^h A_{ij}, \quad h>\max_i r_i,$$

where the A_{ij}'s are $dK(p+q)\times K(p+q)$ matrices, the λ_i's denote the (possibly equal) eigenvalues of P^*, and $r_1+\cdots+r_k=dK(p+q)$. In view of (15) we then have

$$\gamma(h):=\mathrm{Cov}(X_t,X_{t-h})=\sum_{i=1}^{k}\sum_{j=0}^{r_i-1} h^j\lambda_i^h(e'\otimes f')A_{ij}f, \quad h>\max_i r_i.$$

Hence, there exists a recursive relation between the autocovariances, of the form

$$\gamma(h)-\sum_{l=1}^{dK(p+q)}\mu_l\gamma(h-i)=0,$$

where $-\mu_l$ is the coefficient of x^l in the polynomial $Q(x)=\prod_{i=1}^{k}(1-\lambda_i x)^{r_i}$. Since this form characterizes the autocovariance function of ARMA processes (see Brockwell and Davis, 1991), the conclusion follows. The ARMA representation can be used to obtain the linear predictions of the observed process. An explicit example of such a weak ARMA representation will be given in the last section.

5. Examples

In this section, we consider some important special cases aimed to illustrate the second-order stationarity results.

Example 1. The standard univariate ARMA model. When there is no regime change, with $K=1$, we have $P=\Phi\otimes\Phi$ and $\tilde{P}=A\otimes A$, where Φ and A are obtained by replacing the $a_i(\Delta_t)$ (resp., $b_i(\Delta_t)$) by a_i (resp., b_i) in Φ_t and A_t. From Theorem 2 a sufficient condition for second-order stationarity is that $\rho(A)<1$, which is equivalent to the well-known condition on the roots of the autoregressive polynomial. We further assume that the AR and MA polynomials have no common root and that $a_p \neq 0$ or $b_q \neq 0$. Then we can prove that the sufficient condition is also necessary for the existence of a nonanticipative second-order stationary solution (X_t). If we let Γ the covariance matrix of $\underline{z}_t$, we have

$$\Gamma = \Sigma + \Phi\Gamma\Phi', \tag{16}$$

where Σ is the variance of the vector $(\varepsilon_t, 0, \ldots, \varepsilon_t, 0, \ldots, 0)'$. Now let λ denote any eigenvalue of matrix Φ' and let $v = (v_1, \ldots, v_{p+q})'$ any corresponding eigenvector. In view of (16) we have

$$\bar{v}'\Gamma v = \bar{v}'\Sigma v + \lambda\bar{\lambda}\bar{v}'\Gamma v$$

and then

$$0 \leqslant \bar{v}'\Sigma v = (1 - |\lambda|^2)\bar{v}'\Gamma v.$$

With the assumptions made on the AR and MA polynomials and on the variance of the noise ε_t, the matrix Γ is positive definite (otherwise, one could find a fixed linear combination equal to zero between the components of $\underline{z}_t$). Then we can deduce that $|\lambda| \leqslant 1$. Moreover, if $|\lambda| = 1$ then $\bar{v}'\Sigma v = 0$, which implies that $v_1 = v_{p+1} = 0$; from the equality $\Phi' v = \lambda v$ it can be easily deduced that $v = 0$. Therefore $|\lambda| < 1$. To conclude, it can be observed that $\rho(A) = \rho(\Phi)$.

Example 2. Markov-switching MA. In the case where there is no AR part, the vector representation (2) is useless. It is sufficient to remark that the terms $c(\Delta_t)$, ε_t and $b_j(\Delta_t)\varepsilon_{t-j}$ in (1) belong to L^2. Therefore, a Markov-switching MA(q) is always second-order stationary.

Example 3. $\rho(\tilde{P}) < 1$ is a necessary and sufficient second-order stationarity condition for the univariate ARMA(1,1). When $K = p = q = 1$, we show that the sufficient second-order stationarity condition (11) is also necessary in the case where, for at least one regime, the AR and MA polynomials have no common root, i.e. $a_1(\cdot) \neq = -b_1(\cdot)$. First consider the AR(1) case. The result is a straightforward consequence of the fact that, from the assumptions made on $(\Delta_t), \sigma(\cdot)$ and (η_t), all the coefficients of the matrices

$$\mathbb{1}' = (1, \ldots, 1) \quad \text{and} \quad S \operatorname{vec} \Omega = (\pi(1)\sigma^2(1)\Omega, \ldots, \pi(d)\sigma^2(d)\Omega)'$$

in (13) are strictly positive. Now, consider the general univariate ARMA (1, 1). Because

$$\Phi_t = \begin{pmatrix} a_1(\Delta_t) & b_1(\Delta_t) \\ 0 & 0 \end{pmatrix},$$

it is easily seen that, in P, only the rows $1, 5, \ldots, 4i + 1, \ldots$ are different from zero. The same property holds for the P^k. Thus, it can be verified that

$PS\,\mathrm{vec}\,\Omega$ is a vector with component $4i+1$ equal to $\sum_{j=1}^{d} p(j,i+1)(a(i+1)+b(i+1))^2\pi(j+1)\sigma^2(j+1)$, for $i=0,\dots,d-1$, and with other components equal to zero. Therefore, we can assert that

$$P^k S\,\mathrm{vec}\,\Omega = \ddot{P}^{k-1} PS\,\mathrm{vec}\,\Omega \tag{17}$$

where $\ddot{P}$ is obtained by replacing $b_1(\cdot)$ by zero in P. Note that, in $\ddot{P}^k$, all coefficients are equal to zero apart from those located at the intersections of the rows and columns $1,5,\dots,4i+1,\dots$. Now, because in the first row of $\mathbb{I}'$ the elements of the columns $1,5,\dots,4i+1,\dots$ are positive (equal to 1), by choosing the matrix norm $||M|| = \sum |m_{ij}|$, we can deduce from (17) that

$$\sum_{k=1}^{\infty} ||\mathbb{I}' P^k S\,\mathrm{vec}\,\Omega|| \geqslant \inf_{i=1,5,\dots} \{PS\,\mathrm{vec}\,\Omega\}_i \sum_{k=1}^{\infty} ||\ddot{P}^{k-1}||.$$

Therefore, (13) requires that $\rho(\ddot{P}) = \rho(\tilde{P}) < 1$. Hence, we have shown that (11) is a necessary and sufficient stationary condition.

Example 4. $\rho(\tilde{P}) < 1$ *is only a sufficient second-order stationarity condition in the general case.* This example is aimed to show that, in certain cases, condition (11) is not necessary. Consider the following Markov-switching AR(2) process:

$$X_t = \begin{cases} \eta_t & \text{if } \Delta_t = 1, \\ \eta_t + aX_{t-2} & \text{if } \Delta_t = 2, \end{cases}$$

where a is a constant and $d=2$. Then, we have

$$P = \begin{pmatrix} 0_2 & 0_2 & 0_2 & 0_2 \\ P_1 & 0_2 & P_2 & 0_2 \\ 0_2 & aP_3 & 0_2 & aP_4 \\ P_3 & 0_2 & P_4 & 0_2 \end{pmatrix},$$

where 0_2 is the 2×2 null matrix and

$$P_1 = \begin{pmatrix} 0 & 0 \\ p(1,1) & 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0 & 0 \\ p(2,1) & 0 \end{pmatrix},$$

$$P_3 = \begin{pmatrix} 0 & ap(1,2) \\ p(1,2) & 0 \end{pmatrix}, \quad P_4 = \begin{pmatrix} 0 & ap(2,2) \\ p(2,2) & 0 \end{pmatrix}.$$

Then, it is straightforward to compute the spectral radius of $\tilde{P}$:

$$\rho(\tilde{P}) = \max\{|a|\,p(2,2)^{1/2}, |a|\{p(2,2)^2 + p(1,2)p(2,1)\}^{1/2}\}.$$

However, direct computation based on the expansion

$$X_t = \eta_t + \sum_{k=1}^{\infty} a^k \eta_{t-2k} 1_{\Delta_t = 2,\dots,\Delta_{t-2k+2}=2},$$

reveals that the necessary and sufficient condition for second-order stationarity is simply

$$|a|\{p(2,2)^2 + p(1,2)p(2,1)\}^{1/2} < 1,$$

which proves that condition (11) can be unnecessary.

Example 5. Local stationarity does not imply global stationarity. It is clear that the second-order stationarity conditions within *each* regime are not necessary to obtain the global stationarity (see e.g. Example 4). More surprisingly, it can be seen that the local stationarity conditions are neither sufficient. An example where the chain switches deterministically can be found in Karlsen (1998). We propose here an example based on a general two-state Markov chain. Consider the Markov-switching AR(2)

$$X_t = \begin{cases} a_1(1)X_{t-1} + a_2(1)X_{t-2} + \eta_t & \text{if } \Delta_t = 1, \\ a_1(2)X_{t-1} + \eta_t & \text{if } \Delta_t = 2, \end{cases} \tag{18}$$

where (η_t) i.i.d. $\mathcal{N}(0,1)$. If (X_t) is second-order stationary then $\mathrm{E}(X_t^2|\Delta_t = 1, \Delta_{t-1} = 2)$ exists and is independent of t. Therefore,

$$\begin{aligned} &\mathrm{E}(X_t^2|\Delta_t = 1, \Delta_{t-1} = 2) \\ &\quad = \mathrm{E}([\{a_1(1)a_1(2) + a_2(1)\}X_{t-2} + \eta_t + a_1(1)\eta_{t-1}]^2|\Delta_t = 1, \Delta_{t-1} = 2) \\ &\quad \geqslant \{a_1(1)a_1(2) + a_2(1)\}^2\mathrm{E}(X_{t-2}^2|\Delta_t = 1, \Delta_{t-1} = 2) \\ &\quad \geqslant \{a_1(1)a_1(2) + a_2(1)\}^2\mathrm{E}(X_{t-2}^2|\Delta_t = 1, \Delta_{t-1} = 2, \Delta_{t-2} = 1, \Delta_{t-3} = 2) \\ &\qquad \times P(\Delta_{t-2} = 1, \Delta_{t-3} = 2|\Delta_t = 1, \Delta_{t-1} = 2) \\ &\quad = \{a_1(1)a_1(2) + a_2(1)\}^2 p(2,1)p(1,2)\mathrm{E}(X_{t-2}^2|\Delta_{t-2} = 1, \Delta_{t-3} = 2). \end{aligned}$$

This is not possible when

$$\{a_1(1)a_1(2) + a_2(1)\}^2 p(2,1)p(1,2) > 1. \tag{19}$$

Therefore, it is easy to see that (19) is compatible with stationarity of each regime. As an illustration we take $a_1(1) = 1.8$, $a_2(1) = -0.9$, $a_1(2) = -0.2$, $p(1,1) = 0.2$ and $p(2,2) = 0.1$. Both regimes are stationary but, from (19), the (X_t) process is explosive. A trajectory of size 100 is displayed in Fig. 1. To give an idea of the trend, Fig. 2 displays a trajectory of size 300 of the series $\mathrm{e}^{-t/10}X_t$. Since the magnitude of the transformed series seems to be stable, we can conclude that the explosion of (X_t) is exponential. This is rather counter-intuitive: the series is strongly nonstationary whereas the two regimes are stationary. In fact, a careful examination of the graphs reveals that the increase in magnitude arises when the chain switches. Because the probability of staying in the same regime is small, the local stationarity does not compensate the explosive nature of the regime changes.

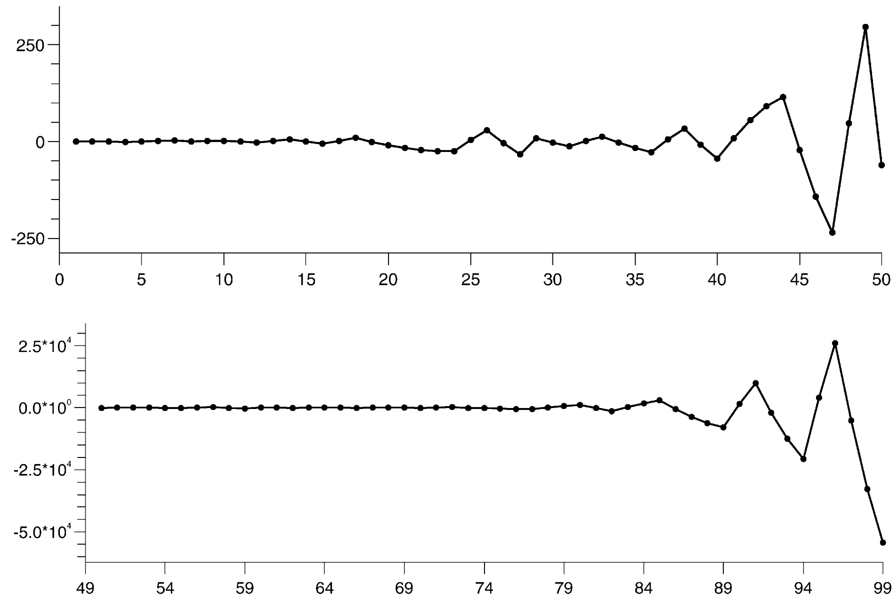

Fig. 1. A simulation of length 100 of (18) with $a_1(1) = 1.8$, $a_2(1) = -0.9$, $a_1(2) = -0.2$, $p(1,1) = 0.2$ and $p(2,2) = 0.1$.

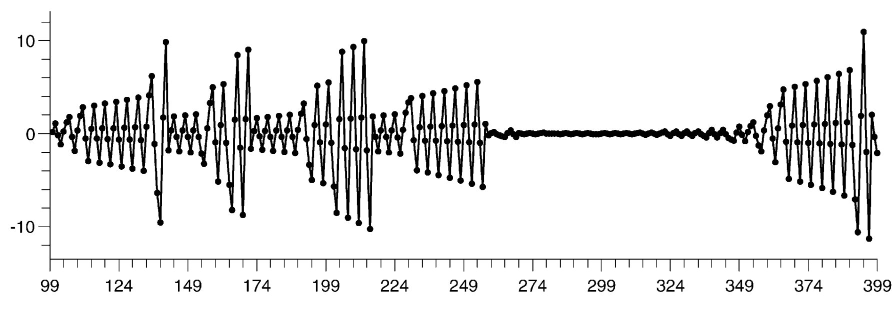

Fig. 2. A simulation of length 300 of $\mathrm{e}^{-t/10}X_t$, where (X_t) is generated as in Fig. 1.

Example 6. The presence of MA terms is material to the stationarity question. It is well known that in a standard ARMA model, the existence of a stationary solution only depends of the autoregressive coefficients. In addition, we have seen in Example 2 that a Markov-switching MA model is always second-order stationary. Therefore, it could seem that the MA coefficients are not material to the second-order stationarity of Markov-switching ARMA models. The following example shows that it is not true. Let L denote the lag operator and consider the three-regime AR(p) model

$$X_t = \begin{cases} \phi_1(L)X_{t-1} + \eta_t & \text{if } \Delta_t = 1, \\ \phi_2(L)X_{t-1} + \eta_t & \text{if } \Delta_t = 2, \\ \eta_t & \text{if } \Delta_t = 3, \end{cases} \tag{20}$$

where $\phi_1(\cdot)$, $\phi_2(\cdot)$ are polynomials, and the transition probabilities are such that $p(1,1) = p(2,2) = p(3,2) = 0$ and $p(1,3) \neq 0$. Moreover, assume that the model admits no second-order stationary solution (in view of (5) there exists a stricly stationary solution because $\pi(3) \neq 0$). Now, keeping the same transition probabilities, let us add MA coefficients to the model and consider

$$X_t = \begin{cases} \phi_1(L)X_{t-1} + \eta_t + \psi_1(L)\eta_{t-1} & \text{if } \Delta_t = 1, \\ \phi_2(L)X_{t-1} + \eta_t + \psi_2(L)\eta_{t-1} & \text{if } \Delta_t = 2, \\ \eta_t + \psi_3(L)\eta_{t-1} & \text{if } \Delta_t = 3, \end{cases} \tag{21}$$

where $\psi_1(\cdot)$, $\psi_2(\cdot)$, $\psi_3(\cdot)$ are polynomials such that $I - \phi_i(L)$ and $I - \psi_i(L)$ $(i = 1, 2)$ have no common root and the following equality holds:

$$\psi_3(L)\{1 - \phi_1(L)\phi_2(L)L^2\} = \psi_2(L) + \phi_2(L) + \phi_2(L)\{\psi_1(L) + \phi_1(L)\}L. \tag{22}$$

To see why the strictly stationary solution to model (21) is also second-order stationary, suppose that the chain is in the third state at time t. Then, with probability 1, at time $t+1$ the chain will either stay in regime 3, or it will move to regime 1. In the latter case, the process X will take the value $X_{t+1} = \eta_{t+1} + \psi_4(L)\eta_t$, with $\psi_4(L) = \phi_1(L) + \psi_1(L) + \phi_3(L)\psi_1(L)$, and at time $t+2$ it will either return to regime 3 or go to regime 2 and take the value

$$\begin{aligned} X_{t+2} &= \phi_2(L)[\phi_1(L)\{\eta_t + \psi_3(L)\eta_{t-1}\} \\ &\quad + \eta_{t+1} + \psi_1(L)\eta_t] + \eta_{t+2} + \psi_2(L)\eta_{t+1} \\ &= \eta_{t+2} + \psi_3(L)\eta_{t+1}. \end{aligned}$$

Therefore, it is easily seen that, for all $s \geqslant t$,

$$P[X_s \in \{\eta_s + \psi_3(L)\eta_{s-1}, \eta_s + \psi_4(L)\eta_{s-1}\} | \Delta_t = 3] = 1.$$

Hence, the strictly stationary solution is also second-order stationary. To be more specific, consider the Markov-switching ARMA(1,2) model given by

$$X_t = \begin{cases} 1.5X_{t-1} + \eta_t + 0.5\eta_{t-1} - 1.35\eta_{t-2} & \text{if } \Delta_t = 1, \\ 0.9X_{t-1} + \eta_t - 1.8\eta_{t-2} & \text{if } \Delta_t = 2, \\ \eta_t + 0.9\eta_{t-1} & \text{if } \Delta_t = 3 \end{cases} \tag{23}$$

with the same assumptions on the transition probabilities. It is straightforward to verify that (22) holds, hence there exists a second-order stationary solution to model (23). However, if we further assume that $1.35^2\, p(1,2)\, p(2,1) > 1$, the model obtained by cancelling the MA part in (23) has no second-order stationary solution because otherwise, a computation similar to that of Example 5 yields

$$\begin{aligned} &\mathrm{E}(X_t^2 \mid \Delta_t = 1, \Delta_{t-1} = 2) \\ &\quad \geqslant (0.9 \times 1.5)^2 p(2,1) p(1,2) \mathrm{E}(X_{t-2}^2 \mid \Delta_{t-2} = 1, \Delta_{t-3} = 2), \end{aligned}$$

which is absurd. Another way to verify the nonstationarity of the AR process is to compute $\rho(\tilde{P}) = 1.35\sqrt{p(1,2)p(2,1)}$ (recall that, from Example 3, $\rho(\tilde{P}) < 1$ is a necessary and sufficient condition for second-order stationarity in this case). Finally, it is interesting to remark that the existence of a second-order stationary solution to model (23) can be established by Theorem 2. Because there is no intercept in this model, (13) is a necessary and sufficient condition for the second-order stationarity. The conclusion follows from the fact that $P^k S = 0$ for $k \geqslant 2$.

Example 7. The presence of intercepts is material to the stationarity question. Another important difference with standard ARMA models is that the addition of a constant in a single regime can be sufficient to transform a second-order stationary process into an explosive process. Take for example model (23) and suppose that the constant 1 is added to the last equation, i.e. $X_t = 1 + \eta_t + 0.9\eta_{t-1}$ if $\Delta_t = 3$. Then it is easily seen that, for all $k \geqslant 0$

$$\begin{aligned} \mathrm{E}(X_{t+2k}^2) &\geqslant \mathrm{E}(X_{t+2k}^2 \mid \Delta_{t+2k} = 2, \Delta_{t+2k-1} = 1, \ldots, \Delta_{t+2} = 2, \Delta_{t+1} = 1, \Delta_t = 3) \\ &\quad P(\Delta_{t+2k} = 2, \Delta_{t+2k-1} = 1, \ldots, \Delta_{t+2} = 2, \Delta_{t+1} = 1, \Delta_t = 3) \\ &= \mathrm{E}((1.35^k + \eta_{t+2k} + 0.9\eta_{t+2k-1})^2 \mid \Delta_{t+2k} = 2, \ldots, \Delta_t = 3) \\ &\quad \{p(1,2)p(2,1)\}^k p(3,1) \\ &\geqslant \{1.35^2 p(1,2)p(2,1)\}^k p(3,1). \end{aligned}$$

Therefore, if we assume that $1.35^2 p(1,2)p(2,1) > 1$, there is no second-order stationary solution to the model.

6. Numerical illustrations and conclusion

In this section, the general results of the previous sections are particularized for a specific model, and we briefly examine their implications for the statistical inference. Consider the following two-regimes AR(1) model:

$$X_t = \begin{cases} \eta_t & \text{if } \Delta_t = 1, \\ 4\eta_t + 1 + 1.11X_{t-1} & \text{if } \Delta_t = 2, \end{cases} \tag{24}$$

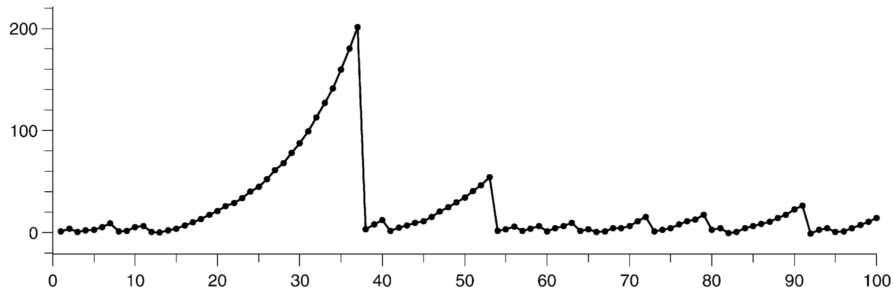

Fig. 3. A simulation of length 100 of model (24).

where $p(1,1)=0.1, p(2,2)=0.8$ and (η_t) i.i.d. $\mathcal{N}(0,1)$. It is interesting to note that, since $\pi(1)\log|a_1(1)|=-\infty$ in (5), there always exists a nonanticipative strictly stationary solution. From Section 5, this solution admits second-order moments because $\rho(\tilde{P})=p(2,2)a^2(2)<1$. Fig. 3 displays a trajectory of length 100 of the process. The graph presents periods with exponential growth, followed by sudden falls. Similar trajectories can be observed in various application fields, in particular for financial time series (see e.g. some of the emerging markets interest rates series presented in Frankel and Okongwu (1998)). Such behaviors are precluded when stationarity is assumed within each regime.

6.1. Maximum-likelihood estimation

It seems from the previous example that, in order to mimic particular features of time series, imposing the stationary condition of this paper could be more appropriate than the local stationarity conditions. We therefore compare via Monte Carlo experiments the effect of imposing the stationarity conditions of both types on the estimation of the two-regimes AR(1) model:

$$X_t=\begin{cases} c(1)+a(1)X_{t-1}+\sigma(1)\eta_t & \text{if } \Delta_t=1,\\ c(2)+a(2)X_{t-1}+\sigma(2)\eta_t & \text{if } \Delta_t=2. \end{cases} \tag{25}$$

The maximum likelihood (ML) approach requires specifying a particular distribution. Therefore suppose that η_t is normally (0,1) distributed. Let $\theta=(p(1,1),p(2,1),a(1),a(2),c(1),c(2),\sigma(1),\sigma(2))$ be the parameter. Let $(X_1,\dots,X_n)$ be a realization of length n of the nonanticipative stationary solution of (25), with parameter value $\theta_0=(0.1,0.2,0,1.11,0,1,1,4)$. We have purposely kept the coefficients of model (24). Given one initial value X_0, the conditional likelihood $L_\theta(X_1,\dots,X_n)$ is given by summing over all the possible paths of the Markov chain, the probability density at the point $(X_1,\dots,X_n)$

given a particular path:

$$L_\theta(X_1,\dots,X_n) = \sum_{(\delta_1,\dots,\delta_n)\in\{\mathscr{E}\}^n} \pi(\delta_1)\left[\prod_{t=2}^{n} p(\delta_{t-1},\delta_t)\right]\left[\prod_{t=1}^{n} f_{\delta_t}(X_{t-1},X_t)\right], \tag{26}$$

where

$$f_{\delta_t}(X_{t-1},X_t) = \frac{1}{\sqrt{2\pi}\sigma(\delta_t)}\exp\left[-\frac{\{X_t - c(\delta_t) - a(\delta_t)X_{t-1}\}^2}{2\sigma(\delta_t)^2}\right].$$

The likelihood can be rewritten as a product of matrices.

Let $\mathbf{1} = (1,\dots,1)' \in \mathbf{R}^d$, $\boldsymbol{p}(X_1) = (\pi(1)f_1(X_0,X_1),\dots,\pi(d)f_d(X_0,X_1))' \in \mathbf{R}^d$ and

$$M_\theta(X_{t-1},X_t) = \begin{pmatrix} p(1,1)f_1(X_{t-1},X_t) & p(2,1)f_1(X_{t-1},X_t) & \cdots & p(d,1)f_1(X_{t-1},X_t) \\ p(1,2)f_2(X_{t-1},X_t) & p(2,2)f_2(X_{t-1},X_t) & \cdots & p(d,2)f_2(X_{t-1},X_t) \\ \vdots & \vdots & & \vdots \\ p(1,d)f_d(X_{t-1},X_t) & p(2,d)f_d(X_{t-1},X_t) & \cdots & p(d,d)f_d(X_{t-1},X_t) \end{pmatrix}.$$

Then it is easy to verify that

$$L_\theta(X_1,\dots,X_n) = \mathbf{1}'\left[\prod_{t=2}^{n} M_\theta(X_{n+1-t},X_{n+2-t})\right]\boldsymbol{p}(X_1). \tag{27}$$

From (27) we can compute recursively $L_\theta(X_1,\dots,X_n)$: for $k = 2,\dots,n$

$$\left[\prod_{t=2}^{k} M_\theta(X_{k+1-t},X_{k+2-t})\right] = M_\theta(X_{k-1},X_k)\left[\prod_{t=2}^{k-1} M_\theta(X_{k-t},X_{k+1-t})\right]. \tag{28}$$

Note that the procedure proposed by Hamilton (1989) can also be used in our framework, in order to compute recursively the log-likelihood function.

The experimental design is as follows. We generate 1000 series of size 260, starting with $X_0 = 0$ and then we discard the first 100 observations. The last 110 observations are not used in the estimation but kept for prediction comparisons. Hence the sample size for the estimation is $n = 50$. For each series we perform three ML estimation of model (25). First, we estimate the model without any stationarity constraint. We only constrain the transition probabilities to belong to $(0,1)$ and the $\sigma(i)$'s to be nonnegative. For identifiability reasons, the constraint $a(1) \leqslant a(2)$ is also imposed. Secondly, we impose the stationarity constraint. Recall that, from Example 3,

$\rho(\tilde{P}) < 1$ is the necessary and sufficient condition for second-order stationarity for the univariate ARMA(1,1). In the case of model (25), it takes the form

$$p(1,1)a^2(1) + p(2,2)a^2(2) + [1 - p(1,1) - p(2,2)]a^2(1)a^2(2) < 1,$$
$$p(1,1)a^2(1) + p(2,2)a^2(2) < 2.$$

Finally, the local stationarity conditions are imposed, i.e. $|a(i)| < 1$, $i = 1,2$, although they are not satisfied for the true value θ_0.

The quartiles, means and RMSEs of the distributions of the estimators are reported in Table 1. All the coefficients of the model are poorly estimated when the too restrictive local stationarity conditions are imposed. In particular, the AR coefficient in the second regime is very close to one for most of the replications. The differences between the results of the two other experimental designs are not substantial. The estimations obtained without imposing any stationarity constraint are slightly better, except for $p(2,1)$ and $c(1)$, than those obtained under the global stationarity constraint.

More interesting conclusions can be drawn from the prediction experiments. For each estimated model, we have computed 100 h-step ahead predictions, for $h = 1,\dots,10$. For each value of h, the means and RMSEs are then computed over 100,000 values. To have a gauge, the same statistics are computed using the true model. The results are displayed in Table 2. Again, the results of designs 1 and 2 are very similar and they are also rather close to the results obtained with the true parameter values. For short horizons the estimations without constraints provide higher bias (except for horizon 1) but lower RMSE, whereas for longer horizons (5 to 10 but we do not report all the results) the estimations under the global constraints dominate, both in terms of bias and accuracy. Of particular interest is the fact that, the discrepancies with the third design (local constraints) appear clearly whatever the horizon. Moreover, the predictions deduced from the locally stationary models worsen as the horizon increases.

The same experiments have been conducted for model (25) with $\theta_0 = (0.1, 0.8, 0.25, -1.9, 0, 0, 0.5, 0.5)$. In this model, by contrast with the previous one, the two regimes are AR(1)'s (instead of a noise in regime 1) and the two stationary probabilities are close to 0.5. Moreover the two AR(1) regimes differ by the coefficients $a(\cdot)$ only. As for model (24), the global stationarity condition is met but not the local one. The simulations results are displayed in Tables 3 and 4. The results confirm the conclusions of the first experiments, but it is seen that imposing the global stationarity constraints leads to slightly better results in terms of prediction accuracy than ignoring these constraints.

These experiments lead to seriously question the idea that imposing the local stationarity conditions would not be material to the inference of Markov-switching ARMA models. The results are more balanced concerning the

Table 1
Distribution of the ML estimates of model (25) obtained from 1000 simulations of size $n = 50$[a]

Parameter	Value	Mean	RMSE	Lower	25.0%	Median	75.0%	Upper
Design 1: Estimation without stationarity constraints Number of stationary estimated models: 611 Number of locally stationary estimated models: 227								
$p(1,1)$	0.1	0.247	0.335	0.000	0.001	0.100	0.404	0.983
$p(2,1)$	0.2	0.193	0.121	0.000	0.113	0.176	0.247	1.000
$a(1)$	0	0.044	0.191	−1.314	−0.017	0.004	0.029	1.118
$a(2)$	1.11	1.024	0.219	−0.591	1.013	1.076	1.118	1.661
$c(1)$	0	0.299	7.115	−90.838	−0.461	−0.009	0.423	130.527
$c(2)$	1	1.534	2.539	−4.311	0.578	1.218	1.890	37.842
$\sigma(1)$	1	1.181	2.300	0.001	0.395	0.728	1.117	31.570
$\sigma(2)$	4	4.245	3.581	1.693	3.556	3.920	4.300	101.777
Design 2: Estimation under the global stationarity constraints Number of locally stationary estimated models: 10								
$p(1,1)$	0.1	0.256	0.342	0.000	0.001	0.104	0.428	0.983
$p(2,1)$	0.2	0.202	0.118	0.000	0.132	0.189	0.252	1.000
$a(1)$	0	0.050	0.199	−0.982	−0.017	0.004	0.031	1.118
$a(2)$	1.11	1.009	0.221	−0.591	1.004	1.059	1.100	1.661
$c(1)$	0	0.285	6.949	−90.839	−0.468	−0.007	0.432	130.533
$c(2)$	1	1.664	2.712	−4.313	0.646	1.342	2.052	37.842
$\sigma(1)$	1	1.434	4.351	0.010	0.409	0.747	1.267	71.044
$\sigma(2)$	4	4.355	3.842	1.693	3.576	3.947	4.341	101.777
Design 3: Estimation under the local stationarity constraints								
$p(1,1)$	0.1	0.295	0.376	0.000	0.001	0.182	0.522	0.984
$p(2,1)$	0.2	0.186	0.121	0.000	0.108	0.168	0.236	1.000
$a(1)$	0	0.057	0.216	−1.000	−0.016	0.006	0.034	1.000
$a(2)$	1.11	0.943	0.236	−0.591	0.998	1.000	1.000	1.000
$c(1)$	0	1.282	22.734	−29.556	−0.497	−0.010	0.445	640.272
$c(2)$	1	2.189	3.380	−7.198	0.834	1.736	2.821	37.842
$\sigma(1)$	1	2.492	11.452	0.010	0.424	0.779	1.392	169.585
$\sigma(2)$	4	4.738	4.421	2.024	3.690	4.138	4.629	101.777

[a] *Note*: The estimates are obtained using starting values $p(1,1) = p(2,1) = 0.5$ and $a(1) = a(2) = 0$; the starting values for $\sigma(1)$ and $\sigma(2)$ were chosen equal to the empirical standard deviations of the simulated series.

omission of the global stationarity condition. Let us remark, however, that, even if such a condition is not imposed in the estimation procedure, it is of great importance to know whether the estimated model is stationary or not. Indeed, the second-order stationarity is required to give sense to the concept of autocovariance, or that of ARMA representation.

Table 2
Prediction errors with the simulated models of Table 1. Design 0: True parameter values. Design 1: Estimation without stationarity constraints. Design 2: Estimation under the global stationarity constraints. Design 3: Estimation under the local stationarity constraints

	Design 0		Design 1		Design 2		Design 3	
Horizon	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	−0.088	12.558	−0.008	13.630	0.029	13.848	−0.105	14.237
2	−0.172	16.819	−0.136	18.018	−0.027	18.257	−0.273	18.524
3	−0.247	19.455	−0.274	20.646	−0.086	20.815	−0.466	21.285
4	−0.306	21.251	−0.465	22.418	−0.175	22.499	−0.661	22.955
5	−0.351	22.542	−0.639	23.744	−0.351	23.717	−0.856	24.312
10	−0.447	25.647	−1.721	27.859	−1.582	26.871	−1.625	28.026

6.2. ARMA representation of model (24)

Using the computations of Section 4, we obtain the following relation between the autocovariances $\gamma(h){:=}\mathrm{cov}(X_t,X_{t-h})$:

$$\begin{aligned}\gamma(h) &= \frac{c^2\pi(1)\pi(2)\{p(2,2)-p(1,2)\}^h}{ap(2,2)-1}+ap(2,2)\gamma(h-1)\\ &= -1.328(-0.1)^h+0.888\gamma(h-1), \quad h>0\end{aligned}$$

and

$$\begin{aligned}\gamma(0) = \pi(1) &+ \frac{1}{1-a^2p(2,2)}\\ &\times\left(a^2p(1,2)\pi(1)+\pi(2)(c^2+\sigma^2)+\frac{2ac^2p(2,2)}{1-ap(2,2)}\right)\\ &-\left(\frac{c\pi(2)}{1-ap(2,2)}\right)^2 = 2092.73.\end{aligned}$$

The ARMA(2,1) representation follows:

$$X_t-0.788X_{t-1}-0.088X_{t-2}=0.900+e_t+0.100e_{t-1},\ \mathrm{var}(e_t)=431.04$$

where (e_t) is a sequence of uncorrelated and centered random variables. This representation provides a simple device to obtain the best linear predictions. These predictions are not optimal in the mean square sense because e_t is obviously not the strong innovation. From a theoretical point of view, the multi-regime representation is naturally more informative than the single-regime one. However, from a practical point of view, the standard ARMA representations are more easily fitted. Moreover these representations contain valuable information about the number of regimes and the orders of the ARMA regimes. For instance, in the example, if the two regimes were white noises ($a=0$) then the coefficient of X_{t-2} in the ARMA(2,1) equation

Table 3
Distribution of the ML estimates of model (25) obtained from 1000 simulations of size $n = 50$[a]

Parameter	Value	Mean	RMSE	Lower	25.0%	Median	75.0%	Upper
Design 1: Estimation without stationarity constraints								
Number of stationary estimated models: 704								
Number of locally stationary estimated models: 106								
$p(1,1)$	0.1	0.190	0.284	0.000	0.000	0.100	0.210	0.982
$p(2,1)$	0.8	0.738	0.245	0.025	0.713	0.799	0.873	1.000
$a(1)$	0.25	0.208	0.131	−0.391	0.182	0.236	0.281	0.450
$a(2)$	−1.9	−1.718	0.527	−2.437	−1.966	−1.870	−1.761	−0.125
$c(1)$	0	0.002	0.130	−0.385	−0.087	−0.005	0.091	0.706
$c(2)$	0	0.028	0.296	−1.878	−0.074	0.014	0.100	2.878
$\sigma(1)$	0.5	0.506	0.216	0.092	0.410	0.475	0.538	2.418
$\sigma(2)$	0.5	0.809	1.523	0.139	0.422	0.479	0.554	17.674
Design 2: Estimation under the global stationarity constraints								
Number of locally stationary estimated models: 126								
$p(1,1)$	0.1	0.191	0.284	0.000	0.000	0.101	0.207	0.982
$p(2,1)$	0.8	0.754	0.240	0.251	0.758	0.808	0.873	1.000
$a(1)$	0.25	0.206	0.131	−0.391	0.183	0.235	0.279	0.450
$a(2)$	−1.9	−1.695	0.542	−2.437	−1.957	−1.866	−1.745	−0.124
$c(1)$	0	0.003	0.130	−0.385	−0.088	−0.005	0.095	0.706
$c(2)$	0	0.028	0.295	−1.878	−0.072	0.013	0.100	2.878
$\sigma(1)$	0.5	0.508	0.216	0.092	0.414	0.476	0.540	2.418
$\sigma(2)$	0.5	0.820	1.525	0.139	0.424	0.483	0.560	17.674
Design 3: Estimation under the local stationarity constraints								
$p(1,1)$	0.1	0.140	0.285	0.000	0.000	0.000	0.102	0.982
$p(2,1)$	0.8	0.691	0.260	0.251	0.633	0.737	0.845	1.000
$a(1)$	0.25	0.211	0.131	−0.391	0.183	0.239	0.286	0.461
$a(2)$	−1.9	−0.933	0.988	−1.000	−1.000	−1.000	−1.000	−0.125
$c(1)$	0	0.003	0.149	−0.645	−0.095	−0.002	0.100	0.706
$c(2)$	0	0.027	0.303	−1.878	−0.088	0.017	0.100	2.878
$\sigma(1)$	0.5	0.491	0.224	0.092	0.388	0.458	0.533	2.418
$\sigma(2)$	0.5	1.250	1.611	0.338	0.763	0.931	0.151	17.674

[a]*Note*: The estimates are obtained using starting values $p(1,1) = p(2,1) = 0.5$ and $a(1) = a(2) = 0$; the starting values for $\sigma(1)$ and $\sigma(2)$ were chosen equal to the empirical standard deviations of the simulated series.

would be equal to zero. Similarly, if the constant c in the second regime was zero then the linear representation would be an AR(1) with no constant term. We conclude with these informal remarks on the links between the initial non linear model and its linear representation, and we leave this important issue for future researches.

Table 4
Prediction errors with the simulated models of Table 3. Design 0: True parameter values. Design 1: Estimation without stationarity constraints. Design 2: Estimation under the global stationarity constraints. Design 3: Estimation under the local stationarity constraints

	Design 0		Design 1		Design 2		Design 3	
Horizon	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	0.005	1.460	−0.002	1.545	−0.002	1.537	−0.005	1.607
2	0.001	1.613	−0.006	1.679	−0.007	1.663	−0.005	1.676
3	0.002	1.657	−0.004	1.685	−0.005	1.680	−0.004	1.689
4	0.003	1.685	−0.003	1.699	−0.003	1.698	−0.004	1.695
5	0.002	1.689	−0.004	1.697	−0.005	1.696	−0.005	1.695
10	0.000	1.706	−0.006	1.709	−0.006	1.708	−0.007	1.708

Appendix A. Proof that $\rho(P) = \rho(\tilde{P})$

In this appendix, we show that the nonzero eigenvalues of matrices P and $\tilde{P}$ are the same. To see this, observe that the blocks of the partitioned matrix P^k are linear combinations of terms of the form $\Phi(i_1)\cdots\Phi(i_k)\otimes\Phi(i_1)\cdots\Phi(i_k)$, with $i_1,\cdots i_k\in\mathscr{E}$; thus, in view of (4), when $k>q$,

$$\Phi(i_1)\cdots\Phi(i_k)\otimes\Phi(i_1)\cdots\Phi(i_k)$$

$$= \begin{pmatrix} A(i_1)\cdots A(i_k)\otimes\Phi(i_1)\cdots\Phi(i_k) & M \\ 0 & 0 \end{pmatrix}, \tag{29}$$

where M is a $K^2p(p+q)\times K^2q(p+q)$ matrix. Moreover, it can be seen that in the upper blocks of (29), the elements of rows number $Ki(p+q)+Kp+j$ ($i=0,\ldots,Kp-1$; $j=1,\ldots,Kq$) are equal to zero. As a consequence, we obtain straightforwardly that for all $\lambda\in\mathbb{R}$,

$$|P^k - \lambda I_{dK^2(p+q)^2}| = (-\lambda^{dK^2q(2p+q)})|\tilde{P}^k - \lambda I_{dK^2p^2}|$$

which proves that the spectral radius of P and that of $\tilde{P}$ coincide.

References

Albert, J.H., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business and Economic Statistics 11, 1–15.

Baum, L.E., Petrie, T., 1966. Statistical inference for probabilistic functions of finite state Markov chains. Annals of Mathematical Statistics 30, 1554–1563.

Billio, M., Monfort, A., 1995. Switching state space models. Doc. travail CREST no. 9557. Insee, Paris.

Billio, M., Monfort, A., Robert, C.P., 1999. Bayesian estimation of switching ARMA models. Journal of Econometrics 93, 229–255.

Bougerol, P., Picard, N., 1992. Strict stationarity of generalized autoregressive processes. Annals of Probability 20, 1714–1729.

Brandt, A., 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. Advances in Applied Probability 18, 221–254.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer, New York.

Casella, G., Robert, C.P., 1999. Monte-Carlo statistical methods. Springer, New-York.

Chib, S., 1996. Calculating posterior distributions and model estimates in Markov mixture models. Journal of Econometrics 75, 79–97.

Elliott, R.J., Lakhdar, A., Moore, J.B., 1995. Hidden Markov Models: Estimation and Control (Application of Mathematics). Springer, Berlin.

Francq, C., Roussignol, M., 1997. On white noises driven by hidden Markov chains. Journal of Time Series Analysis 18, 553–578.

Francq, C., Roussignol, M., 1998. Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. Statistics 32, 151–173.

Francq, C., Zakoïan, J.M., 1998. Estimating linear representations of nonlinear processes. Journal of Statistical Planning and Inference 68, 145–165.

Frankel, J.A., Okongwu, C., 1998. Liberalized portfolio capital inflows in emerging markets: sterilization, expectations and the incompleteness of interest rate convergence. International Journal of Finance Economics 1, 1–23.

Hamilton, J.D., 1988. Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. Journal of Economic Dynamics and Control 12, 385–423.

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., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Holst, U., Lindgren, G., Holst, J., Thuvesholmen, M., 1994. Recursive estimation in switching autoregressions with a Markov regime. Journal of Time Series Analysis 15, 489–506.

Karlsen, H., 1990. A class of non-linear time series models. Ph.D. Thesis, University of Bergen, Norway.

Karlsen, H., 1998, Second order stationarity for some doubly stochastic processes where the parameter process is a finite Markov chain—some examples. Preprint, University of Bergen, Norway.

Kim, C.J., Nelson, C.R., 1999. State Space Models with Regime Switching. MIT Press, Cambridge, MA.

Lancaster, P., Tismenetsky, M., 1985. The Theory of Matrices. Academic Press, New York.

Leroux, B.G., 1992. Maximum-likelihood estimation for hidden Markov models. Stochastic Process. Appl. 40, 127–143.

McCulloch, R.E., Tsay, R.S., 1994. Statistical analysis of economic time series via Markov switching models. Journal of Time Series Analysis 15, 523–539.

Rabiner, L.R., Juang, B.H., 1986. An introduction to hidden Markov models. IEEE ASSP Mag. 1 (1986) 4–16.

Rydèn, T., 1994. Consistent and asymptotically normal parameter estimates for hidden Markov models. Annals of Statistics 22, 1884–1895.

Shephard, N., 1994. Partial non-Gaussian state space. Biometrika 81, 528–550.

Titterington, D.M., Smith, A.F.M., Makov, U.E., 1985. Statistical analysis of finite mixture distributions. Wiley, New York.

Tjøstheim, D., 1986. Some doubly stochastic time series models. Journal of Time Series Analysis 7, 51–72.

Yao, J., Attali, J.G., 1998. Stabilité des modèles AR non-linéaires à régime markovien. C. R. Acad. Sci. 327, 593–596.

Yang, M., 2000. Some properties of vector autoregressive processes with Markov-switching coefficients. Econometric Theory 16, 23–43.