Ca' Foscari University of Venice

Department of Economics

Working Paper

Maddalena Cavicchioli

Determining the number of regimes in Markov-switching VAR and VMA models

ISSN: 1827-3580
No. 03/WP/2013

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 03/WP/2013
ISSN 1827-3580

DETERMINING THE NUMBER OF REGIMES IN MARKOV-SWITCHING VAR AND VMA MODELS

Maddalena Cavicchioli

Advanced School of Economics – Cà Foscari University of Venice

06/03/2013

Abstract. We give stable finite order VARMA(p*; q*) representations for M-state Markov switching second-order stationary time series whose autocovariances satisfy a certain matrix relation. The upper bounds for p* and q* are elementary functions of the dimension K of the process, the number M of regimes, the autoregressive and moving average orders of the initial model. If there is no cancellation, the bounds become equalities, and this solves the identification problem. Our class of time series include every M-state Markov switching multivariate moving average models and autoregressive models in which the regime variable is uncorrelated with the observable. Our results include, as particular cases, those obtained by Krolzig (1997), and improve the bounds given by Zhang and Stine (2001) and Francq and Zakoian (2001) for our classes of dynamic models. Data simulations and an application on foreign exchange rates complete the paper.

Keywords. Second-order stationary time series, VMA models, VAR models, State-Space models, Markov chains, changes in regime, regime number.

JEL Codes. C01, C32, C50, C52.

Address for correspondence:
Maddalena Cavicchioli
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: maddalena.cavicchioli@unive.it

The Working Paper Series
is available only on line
(http://www.unive.it/nqcontent.cfm?a_id=86302)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

1. Introduction

In this paper we consider dynamic models whose parameters can change as a result of a regime-shift variable, described as the outcome of an unobserved Markov chain. Such models have attracted much interest in the literature for their applications in areas as economics, statistics, and finance. A key problem arising in applications is to determine the number of Markov regimes for which a switching model gives an adequate representation of the observed data. In practice, the state dimension of the Markov chain is sometimes dictated by the actual application or it is determined in an informal manner by visual inspection of plots. However, there exists in the literature likelihood ratio test developed under non-standard conditions which help testing Markov switching models (see Hansen (1992)). The current methods for determining the state dimension are mainly based either on complexity-penalized likelihood criteria (see, for example, Psaradakis and Spagnolo (2003), Olteanu and Rynkiewicz (2007), and Ríos and Rodríguez (2008)) or on finite order VARMA representations of the initial switching models (see, for example, Krolzig (1997), Zhang and Stine (2001) and Francq and Zakoïan (2001)). The parameters of the VARMA representations can be determined by evaluating the autocovariance function of the Markov-switching models. It turns out that the above parameters are elementary functions of the dimension of the dynamic process, the number of regimes and the orders of the switching autoregressive moving-average model. As the sample autocovariances are more easily calculated than maximum (penalized) likelihood estimates of the model parameters, the bounds arising from the above-mentioned elementary functions are very useful for selecting the number of regimes and the orders of the switching moving-average autoregression. Some bounds are previously determined by Krolzig (1997), Zhang and Stine (2001) and Francq and Zakoïan (2001) for some Markov regime switching models of different type. Surprisingly, we show that the bounds given by Krolzig maintain their validity for Markov switching time series whose autocovariances satisfy a matrix relation specified in the statement of Theorem 2.2. This allows us to improve the bounds obtained by Zhang and Stine (2001) in Theorem 4 and Francq and Zakoïan (2001) in Section 4.3 for a large class of dynamic models. This class includes every multivariate regime switching Moving Average (MA) process and multivariate regime switching Autoregressive (AR) processes, in which the regime variable is uncorrelated with the observable. The main results of the paper are Theorems 2.2, 3.5, 4.2 and we are going to illustrate them (the specifics of the models and the regularity assumptions will be given in the next sections). The first result states that a second-order stationary dynamic process, whose autocovariances satisfy a certain matrix relation, has a stable VARMA representation whose AR and MA orders are well-specified elementary functions. The second result relates to M-state switching multivariate Moving Average model MA(q) of the type $\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Theta}_{s_t}(L)\mathbf{u}_t$.

We will show that it admits a stable VARMA(p^*, q^*) representation, whose autoregressive lag polynomial is scalar, and where the orders of the stable VARMA satisfy $p^* \leq M-1$ and $q^* \leq M+q-1$. Finally, the last result regards to a M-state switching K-dimensional Autoregressive model AR(p) of the type $\boldsymbol{\phi}_{s_t}(L)\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Sigma}_{s_t}(L)\mathbf{u}_t$. Assuming that the regime variable is uncorrelated with the observable, we prove that it admits a stable VARMA(p^*, q^*) representation, whose autoregressive lag polynomial is scalar, and where the orders of the stable VARMA satisfy $p^* \leq M+Kp-1$ and $q^* \leq M+(K-1)p-1$. The assumption of uncorrelated regimes with the observable is, of course, satisfied when the autoregressive part is regime-invariant (as done in Krolzig (1997)). However, it is also a reasonable assumption when coinceving the change in regimes as an outside event from the economic system as, for instance, abrupt natural events or unexpected wars. This leads us to the fact that if the lag polynomials of the autoregressive and moving-average parts of the stable VARMA(p^*, q^*) are coprime, then equalities hold in the previous relations. In other words, if there is no cancellation, then the identification problem is completely solved. In any case, the above result allows us to determine a lower bound for the number of states. This means that a VARMA representation, and hence a VARMA structure in the autocovariance function, may reveal the characteristics of a data generating MS(M)-VMA and MS(M)-VAR processes. More precisely, the determination of the number of regimes, as well as the number of autoregressive (or moving-average) parameters, can be based on currently available procedures to estimate the orders of VARMA models. In principle, any of the existing model selection criteria may be applied for identifying M and p or q. In the case of MS(M)-VMA(q) model, evaluating estimates $(\widehat{p^*}, \widehat{q^*})$ of the orders of the stable VARMA(p^*, q^*) representation, we get the estimates $\widehat{M} = \widehat{p^*}+1$ and $\widehat{q} = \widehat{q^*} - \widehat{p^*}$, for M and q, respectively, when there is no cancellation (see Theorem 3.5). Furthermore, for MS(M)-VAR(p) model, in which the regime variable is uncorrelated with the observable, having estimates $(\widehat{p^*}, \widehat{q^*})$ of the orders of the stable VARMA(p^*, q^*) representation, gives us $\widehat{M} = K(\widehat{q^*}+1)-(K-1)(\widehat{p^*}+1)$ and $\widehat{p} = \widehat{p^*} - \widehat{q^*}$, for M and p, respectively, when there is no cancellation and the autoregressive lag polynomial of the stable representation is scalar (see Theorem 4.2). The rest of the paper is organized as follows. In Section 2 we recall a characterization of VARMA process in terms of autocovariances, as given by Zhang and Stine (2001). Then we extend Lemma 1 of the quoted paper, and give a different proofs of Krolzig's results on VARMA representations of certain MS processes (see Krolzig (1997), Propositions 2, 3, and 4, Chp. 3) and of some results of Zhang and Stine (2001) and Francq and Zakoïan (2001). In Section 3 (resp. 4) we show that an MS(M)-MA(q), $q \geq 0$, (resp. MS(M)-VAR(p), $p \geq 0$, for which the regime variable is uncorrelated with the observable) has a VARMA(p^*, q^*) representation, where the upper bounds for p^* and q^* are induced by specific elementary functions. In Section 5 we discuss the implications of our results for model selection, and illustrate some simulation experiments and numerical applications. In Section 6 we include our results on the exchange rata data from Engle and Hamilton (1990). Section 7 concludes. The proofs of theorems in Section 3 and 4 are reported in the Appendix which completes the paper.

2. VARMA Representations

In this section, we introduce the model and the basic notation concerning with it. In particular, we prove new algebraic results (Theorem 2.2 and Corollary 2.3) which will be used in the next Sections 3 and 4 for the determination of the number of regimes in Markov-switching VMA and VAR models. As a consequence, we give different and simple proofs of some results, previously obtained by several authors, for classes of Markov switching time series which are included in our models.

Let $\mathbf{y} = (\mathbf{y}_t)$ be a second-order stationary K-dimensional process. Then $\mathbf{y}$ is said to have a stable and invertible VARMA(p, q) *representation* if it satisfies a finite difference equation $\boldsymbol{\phi}(L)\mathbf{y}_t = \boldsymbol{\Theta}(L)\mathbf{u}_t$, where $\boldsymbol{\phi}(L) = \sum_{i=0}^{p} \boldsymbol{\phi}_i L^i$ and $\boldsymbol{\Theta}(L) = \sum_{j=0}^{q} \boldsymbol{\Theta}_j L^j$ are $K \times K$ matrix polynomials in the lag operator L, $\boldsymbol{\phi}_0 = \boldsymbol{\Theta}_0 = \mathbf{I}_K$, $\boldsymbol{\phi}_p \neq \mathbf{0}, \boldsymbol{\Theta}_q \neq \mathbf{0}$. Here the variables $\mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p}$, $\mathbf{u}_{-1}, \ldots, \mathbf{u}_{-q}$ are assumed to be uncorrelated with $\mathbf{u}_t$ for every $t \geq 0$. The process $\mathbf{u} = (\mathbf{u}_t)$ is a zero mean white noise, i.e., $E(\mathbf{u}_t \mathbf{u}_\tau^{'}) = \delta_\tau^t G$ with $|G| \neq 0$ (through the paper, the symbol $|A|$ denotes the determinant of a square matrix A, and δ_τ^t denotes the Kronecker symbol, i.e., $\delta_\tau^t = 1$ if $t = \tau$ and zero otherwise). To avoid redundancy, $\boldsymbol{\phi}(L)$ and $\boldsymbol{\Theta}(L)$ are coprime, and to guarantee invertibility, we assume that the polynomials $|\boldsymbol{\phi}(z)|$ and $|\boldsymbol{\Theta}(z)|$, $z \in \mathbb{C}$, have all their roots strictly outside the unit circle. This definition implies that the orders p and q are minimal in the usual sense. Finally, the process $\mathbf{y} = (\mathbf{y}_t)$ is second-order stationary if the mean $E(\mathbf{y}_t)$ and the autocovariances $\boldsymbol{\Gamma}_{\mathbf{y}}(t, h)$ are independent of t. So we can write $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}_t)$ and $\boldsymbol{\Gamma}_{\mathbf{y}}(h) = \boldsymbol{\Gamma}_{\mathbf{y}}(t, h)$. We start with the following well-knowm result which characterizes the minimal VARMA(p, q) model in terms of its autocovariance function (through the paper we always assume that the process is not deterministic). For the proof see Zhang and Stine (2001), Theorem 1. Let L be the backward shift operator, $L\Gamma_{\mathbf{y}}(h) = \Gamma_{\mathbf{y}}(h-1)$, where $\Gamma_{\mathbf{y}}$ is the autocovariance function of the observed process $\mathbf{y} = (\mathbf{y}_t)$.

Theorem 2.1. *Suppose that the K-dimensional process $\mathbf{y} = (\mathbf{y}_t)$ is second-order stationary (or equivalently, weakly stationary) and the covariances $\Gamma_{\mathbf{y}}(h)$, $h \in \mathbb{Z}$, satisfy a finite difference equation of order p and rank $q + 1$, that is, there exist $K \times K$ matrices $\mathbf{A}_i$, $i = 0, \ldots, p$, with $\mathbf{A}_0 = \mathbf{I}_K$ and $\mathbf{A}_p \neq \mathbf{0}$, such that $B(L)\Gamma_{\mathbf{y}}(h) \neq 0$ for $h = q$ and vanishes for every $h \geq q + 1$, where $B(L) = \sum_{i=0}^{p} \mathbf{A}_i L^i$. Then $\mathbf{y} = (\mathbf{y}_t)$ has a VARMA(p^*, q^*) representation, where $p^* \leq p$ and $q^* \leq q$. If the pair (p, q) is minimal, then we have equalities $p^* = p$ and $q^* = q$.*

Note that the dimension K is absent from p^* and q^*. This depends on the fact that the autoregressive part of VARMA(p^*, q^*) consists of matrices, in general. In what follows, we also seek special VARMA(p^*, q^*) representation whose autoregressive part consists of scalars. This corresponds to the *final equation form* usually considered in the statements of Krolzig's book (1997); see, for example, Formula 3.10, p.58.

The following result generalizes Lemma 1 of Zhang and Stine (2001), proved for the case $p = q = 0$.

Theorem 2.2. *Suppose that the K-dimensional process $\mathbf{y} = (\mathbf{y}_t)$ is second-order stationary and the autocovariances of $\mathbf{y}$ satisfy*

$$B(L)\Gamma_{\mathbf{y}}(h) = A^{'}Q^h B$$

for every $h \geq q \geq 0$, where all the matrices on the right hand-side are nonzero matrices, Q is $M \times M$, A and B are $M \times K$ matrices, $B(L) = \sum_{i=0}^{p} \mathbf{B}_i L^i$ is a $K \times K$ matrix polynomial of degree $p \geq 0$, with $\mathbf{B}_0 = I_K$ and $\mathbf{B}_p \neq \mathbf{0}$. Then $\mathbf{y} = (\mathbf{y}_t)$ has a stable VARMA(p^*, q^*) representation, where $p^* \leq M+p$ and $q^* \leq M+q-1$. If we require that the autoregressive part of such a representation consists of scalars (not matrices) and assume the usual regularity conditions on the roots of the polynomial $|B(z)|$, $z \in \mathbb{C}$, to guarantee the invertibility of $B(L)$, the bounds become $p^* \leq M + Kp$ and $q^* \leq M + (K-1)p + q - 1$.

Proof. The Cayley-Hamilton theorem implies that there exist real numbers $f_1 \dots f_M \in \mathbb{R}$ such that

$$Q^M - f_1 Q^{M-1} - \cdots - f_M I_M = \varphi_Q(Q) = \mathbf{0} \tag{2.1}$$

where $\varphi_Q(\lambda) = \lambda^M - f_1\lambda^{M-1} - \cdots - f_M$ is the characteristic polynomial of Q, that is, $\varphi_Q(\lambda) = |\lambda I_M - Q|$. The hypothesis of the statement implies the following relations

$$\begin{aligned}
\Gamma_{\mathbf{y}}(q+h+M) + \mathbf{B}_1\Gamma_{\mathbf{y}}(q+h+M-1) + \cdots + \mathbf{B}_p\Gamma_{\mathbf{y}}(q+h+M-p) &= A^{'}Q^{q+h+M}B \\
\Gamma_{\mathbf{y}}(q+h+M-1) + \mathbf{B}_1\Gamma_{\mathbf{y}}(q+h+M-2) + \cdots + \mathbf{B}_p\Gamma_{\mathbf{y}}(q+h+M-p-1) &= A^{'}Q^{q+h+M-1}B \\
&\ \ \vdots \\
\Gamma_{\mathbf{y}}(q+h) + \mathbf{B}_1\Gamma_{\mathbf{y}}(q+h-1) + \cdots + \mathbf{B}_p\Gamma_{\mathbf{y}}(q+h-p) &= A^{'}Q^{q+h}B
\end{aligned}$$

for every $h \geq 0$. Multiplying the last M lines with $-f_1, \dots, -f_M$ and adding all equations, we get

$$\Gamma_{\mathbf{y}}(q+h+M) + \sum_{j=1}^{M+p} \mathbf{C}_j \Gamma_{\mathbf{y}}(q+h+M-j) = A^{'}Q^{q+h}\varphi_Q(Q)B = \mathbf{0} \tag{2.2}$$

for some matrices $\{\mathbf{C}_j\}$ and for every $h \geq 0$. The right hand-side of (2.2) is zero by (2.1). Formula (2.2) can be written in the equivalent form

$$\Gamma_{\mathbf{y}}(h) + \sum_{j=1}^{M+p} \mathbf{C}_j \Gamma_{\mathbf{y}}(h-j) = \mathbf{0} \tag{2.3}$$

for every $h \geq M + q$. From (2.3) and Theorem 2.1 we can conclude that

$$p^* \leq M + p \qquad \text{and} \qquad q^* \leq M + q - 1.$$

To get a stable VARMA(p^*, q^*) representation whose autoregressive part consists of scalars (not matrices), we multiply (2.1) by Q^{q-p}

$$Q^{M+q-p} - f_1 Q^{M+q-p-1} - \cdots - f_M Q^{q-p} = \mathbf{0}. \tag{2.4}$$

Premultiplying (2.4) by $A^{'}$ and postmultiplying (2.4) by B yield

$$A^{'} Q^{M+q-p} B - f_1 A^{'} Q^{M+q-p-1} B - \cdots - f_M A^{'} Q^{q-p} B = \mathbf{0}$$

hence

$$B(L)\Gamma_{\mathbf{y}}(M+q-p) - f_1 B(L)\Gamma_{\mathbf{y}}(M+q-p-1) - \cdots - f_M B(L)\Gamma_{\mathbf{y}}(q-p) = \mathbf{0}$$

by using the matrix relation of the statement. Premultiplying the last equation by the adjoint $B(L)^*$ of $B(L)$ defined as $B(L)^* = |B(L)|B(L)^{-1}$ recalling that $B(L)$ is invertible by hypothesis (or equivalently, $B(L)^* B(L) = |B(L)|\mathbf{I}_K$), we get

$$|B(L)|(\Gamma_{\mathbf{y}}(M+q-p) - f_1 \Gamma_{\mathbf{y}}(M+q-p-1) - \cdots - f_M \Gamma_{\mathbf{y}}(q-p)) = \mathbf{0}$$

where the polynomial $|B(L)|$ has degree Kp. Doing the matrix products term-by-term, taking in mind the definition of the operator L and collecting similar terms, we get a finite scalar difference equation of the form

$$\Gamma_{\mathbf{y}}(M+q+Kp-p) + \eta_1 \Gamma_{\mathbf{y}}(M+q+Kp-p-1) + \cdots + \eta_{M+Kp} \Gamma_{\mathbf{y}}(q-p) = \mathbf{0} \tag{2.5}$$

where the coefficients $\{\eta_j\}$ are scalars. Now the last result in the statement follows from (2.5) and Theorem 2.1, that is, we get $p^* \leq M + Kp$, and $q^* \leq M + q + (K-1)p - 1$. $\square$

Corollary 2.3. *Under the hypothesis of Theorem 2.2, if the autocovariances of $\mathbf{y} = (\mathbf{y}_t)$ satisfy*

$$B(L)\Gamma_{\mathbf{y}}(h) = \sum_{i=1}^{r} A_i^{'} Q_i^h B_i$$

for every $h \geq q \geq 0$, where A_i and B_i are $M_i \times K$ nonzero matrices, Q_i is an $M_i \times M_i$ nonzero matrix, for $i = 1, \ldots, r$, and $B(L)$ is a $K \times K$ matrix polynomial in L of degree $p \geq 0$, with $\mathbf{B}_0 = \mathbf{I}_K$. Then $\mathbf{y} = (\mathbf{y}_t)$ has a stable VARMA(p^*, q^*) representation, where $p^* \leq \sum_{i=1}^r M_i + p$ and $q^* \leq \sum_{i=1}^r M_i + q - 1$. If $B(L)$ is invertible and we require that the autoregressive part of such a representation consists of scalars (not matrices), the bounds become $p^* \leq \sum_{i=1}^r M_i + Kp$ and $q^* \leq \sum_{i=1}^r M_i + (K-1)p + q - 1$.

Proof. Setting $A^{'} = [A_1^{'} \dots A_r^{'}]$, $B = [B_1^{'} \dots B_r^{'}]^{'}$ and $Q = diag(Q_1 \dots Q_r)$, we get $B(L)\Gamma_{\mathbf{y}}(h) = A^{'}Q^h B$. The result now follows from Theorem 2.2. □

To complete the section we give different proofs of Krolzig's results on VARMA representations of certain MS processes (see Krolzig (1997), Chp. 3).

Proposition 2.4. (Krolzig(1997), Proposition 2, p.56) *Let $\mathbf{y} = (\mathbf{y}_t)$ be a K-dimensional Hidden Markov-chain process (called MSI(M)-VAR(0) process in Krolzig (1997),p.50)*

$$\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \mathbf{u}_t \qquad \mathbf{u}_t \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_u).$$

Then $\mathbf{y}$ admits a stable VARMA(p^*, q^*) representation with $p^* = q^* \leq M-1$.

Proof. The autocovariances of $\mathbf{y}$ satisfy $\Gamma_{\mathbf{y}}(h) = A^{'}F^h B$ for every $h \geq 1$, where A and B are nonzero $(M-1)\times K$ matrices, and F is $(M-1)\times(M-1)$ (see Krolzig (1997), Section 3.3.2, Formula (3.21)). Now apply Theorem 2.2 for $p=0$, $q=1$ and $M-1$ instead of M. □

Proposition 2.5. (Krolzig(1997), Proposition 3, p.57) *Let $\mathbf{y} = (\mathbf{y}_t)$ be a K-dimensional M-state Markov switching process (called MSI(M)-VAR(p) process, $p > 0$, in Krolzig (1997),p.50)*

$$A(L)\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \mathbf{u}_t \qquad \mathbf{u}_t \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_u).$$

where $A(L) = I_K - A_1 L - \cdots - A_p L^p$, $A_p \neq 0$, is invertible and regime invariant. Then $\mathbf{y} = (\mathbf{y}_t)$ admits a stable VARMA(p^*, q^*) representation

$$C(L)(\mathbf{y}_t - \mu_{\mathbf{y}}) = B(L)\boldsymbol{\epsilon}_t$$

where $p^* \leq M+p-1$ and $q^* \leq M-1$, $C(L)$ is a $K\times K$ dimensional lag polynomial of order $M+p-1$, $B(L)$ is a $K\times K$ dimensional lag polynomial of order $M-1$, and $\boldsymbol{\epsilon}_t$ is a zero mean vector white noise process. If we require that the autoregressive part of such a stable representation is scalar, the bounds become $p^* \leq M+Kp-1$ and $q^* \leq M+(K-1)p-1$, as pointed out in Krolzig (1997), p.58, Formula (3.10).

Proof. The autocovariances of $\mathbf{y}$ satisfy $\Gamma_{\mathbf{y}}(h) - \sum_{j=1}^{p} A_j \Gamma_{\mathbf{y}}(h-j) = A^{'}F^h B$ for every $h \geq 1$, where A and B are nonzero $(M-1)\times K$ matrices, and F is $(M-1)\times(M-1)$ (see Krolzig (1997), Section 3.3.4, Formula (3.27)). The result now follows from Theorem 2.2 for $q=1$ and $M-1$ instead of M. □

Proposition 2.6. (Krolzig(1997), Proposition 4, p.58) *Let $\mathbf{y} = (\mathbf{y}_t)$ be a K-dimensional M-state Markov switching process (called MSM(M)-VAR(p) process, $p > 0$, in Krolzig (1997),p.50)*

$$A(L)(\mathbf{y}_t - \boldsymbol{\mu}_{s_t}) = \mathbf{u}_t \qquad \mathbf{u}_t \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_u).$$

where $A(L)$ is as above. Then there exists a final equation form VARMA(p^*, q^*) representation

$$\gamma(L)(\mathbf{y}_t - \mu_{\mathbf{y}}) = B(L)\boldsymbol{\epsilon}_t$$

where $p^* \leq M + Kp - 1$ and $q^* \leq M + Kp - 2$, $\gamma(L)$ is a scalar lag polyonomial of order $M + Kp - 1$, $B(L)$ is a $(K \times K)$ dimensional lag polynomial of order $M + Kp - 2$, and $\boldsymbol{\epsilon}_t$ is a zero mean vector white noise process.

Proof. The autocovariances of $\mathbf{y}$ satisfy $\Gamma_{\mathbf{y}}(h) = V^{'} F^h W + X^{'} A^h Z$, for every $h \geq 0$, where V and W are nonzero $(M-1) \times K$ matrices, X and Z are nonzero $(Kp) \times K$ nonzero matrices, F is $(M-1) \times (M-1)$ and A is $(Kp) \times (Kp)$ (see Krolzig (1997), Section 3.3.3, Formula (3.24)). Now apply Corollary 2.3 by setting $r = 2$ and $p = q = 0$ in the last statement of it. □

Note that there is a typographic error in the statement of Proposition 4, Krolzig (1997), p.58, that is, one reads $q^* \leq M - Kp - 2$. However, the first minus sign in this inequality is a typo as remarked at line -4, p.58, in the same reference.

For completeness, we also recall the bounds for the regime number obtained by Francq and Zakoïan (2001) in Section 4.3. Here we give a different proof of their result by using Theorem 2.2 above.

Proposition 2.7. (Francq and Zakoïan (2001), Sec. 4.3) *Let $\mathbf{y} = (\mathbf{y}_t)$ be a K-dimensional second-order stationary centered dynamic process which satisfies the following MS(M) VARMA(p, q) model (in the notation of the quoted paper)*

$$\mathbf{y}_t = \sum_{i=1}^{p} a_i(s_t)\mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t + \sum_{j=1}^{q} b_j(s_t)\boldsymbol{\epsilon}_{t-j}$$

where $a_i(s_t)$ and $b_j(s_t)$ are $K \times K$ random matrices, $\boldsymbol{\epsilon}_t = \sigma(s_t)\boldsymbol{\eta}_t$, where $\sigma(s_t)$ is a $K \times K$ random matrix and $\boldsymbol{\eta}_t$ is a centered white noise with $E(\boldsymbol{\eta}_t\boldsymbol{\eta}_\tau^{'}) = \delta_\tau^t \Omega$ (Ω non singular). Then $\mathbf{y} = (\mathbf{y}_t)$ admits a stable VARMA(p^*, q^*) representation, where $p^*, q^* \leq MK(p+q)$.

Proof. It was shown in Francq and Zakoïan, Section 4.3, that the autocovariance of $\mathbf{y} = (\mathbf{y}_t)$ is given by $\Gamma_{\mathbf{y}}(h) = E(\mathbf{y}_t\mathbf{y}_{t-h}^{'}) = (\mathbf{e}^{'} \otimes \mathbf{f}^{'})(P^*)^h W(0)\mathbf{f}$, for every $h > 0$, where $(\mathbf{e}^{'} \otimes \mathbf{f}^{'})$ and $W(0)\mathbf{f}$ are nontrivial $K \times [MK(p+q)]$ and $[MK(p+q)] \times K$ matrices, respectively, and P^* is $[MK(p+q)] \times [MK(p+q)]$. Now we apply Theorem 2.2 where on the right side $Q = P^*$ is a square matrix of order $MK(p+q)$ and on the left side $B(L) = \mathbf{I}_K$, i.e., we must set $p = 0$ and $q = 1$ according to the notation in the statement of Theorem 2.2. □

Finally, we obtain the main result in Zhang and Stine (2001) by using Theorem 2.2 above.

Proposition 2.8. (Zhang and Stine (2001), Theorem 4) *Let* $\mathbf{y} = (\mathbf{y}_t)$ *be a* K*-dimensional second-order stationary Markov regime switching VAR(p)*

$$\mathbf{y}_t = A_{s_t}^{(1)}\mathbf{y}_{t-1} + \cdots + A_{s_t}^{(p)}\mathbf{y}_{t-p} + \mathbf{\Sigma}_{s_t}\mathbf{v}_t$$

where $\mathbf{v}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$*,* $\mathbf{\Sigma}_{s_t}$ *is a* $K \times K$ *positive definitive matrix,* $A_{s_t}^{(i)}$ *is* $K \times K$ *and* $\{s_t\}$ *is independent of* $\{\mathbf{v}_t\}$*. Then* $\mathbf{y} = (\mathbf{y}_t)$ *admits a stable VARMA(*p^*, q^**) representation, where* $p^* \leq M(Kp)^2$ *and* $q^* \leq M(Kp)^2 - 1$*.*

Proof. It was shown in Zhang and Stine (2001), Formula (29), that the autocovariance of $\mathbf{y} = (\mathbf{y}_t)$ is given by $\operatorname{vec}\Gamma_{\mathbf{y}}(h) = \mathbf{Q}F_1^h\mathbf{R}$, for every $h \geq 0$, where $\mathbf{Q}$ and $\mathbf{R}$ are nontrivial $K^2 \times [M(Kp)^2]$ and $[M(Kp)^2] \times 1$ matrices, respectively, and F_1 is $[M(Kp)^2] \times [M(Kp)^2]$. Now we apply Theorem 2.2 where on the right side $Q = F_1$ is a square matrix of order $M(Kp)^2$ and on the left side $B(L) = \mathbf{I}_K$, i.e., we must set $p = 0$ and $q = 0$ according to the notation in the statement of Theorem 2.2. $\square$

In the next sections, we improve the bounds given by the previous authors, and show that the bounds given by Krolzig (1997) maintain their validity in the general case of MS(M)-VMA(q) models and for the class of MS(M)-VAR(p) models in which the regime variable is uncorrelated with the observable.

3. Markov Switching Moving Average Models

In this section we consider Markov-switching models with the following moving-average form (in short, MS(M)-MA(q)):

$$\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \mathbf{\Theta}_{s_t}(L)\mathbf{u}_t \tag{3.1}$$

Here we allows Markovian shifts in the intercept term; the case with regime changes in the mean can be treated in a similar manner. As usual, $\mathbf{y} = (\mathbf{y}_t)$ is a K-dimensional random process, $\mathbf{\Theta}_{s_t}(L) = \sum_{j=0}^{q} \mathbf{\Theta}_{s_t,j}L^j$, with $\mathbf{\Theta}_{s_t,0} = \mathbf{\Sigma}_{s_t}$ (nonsingular symmetric $K \times K$ matrix) and $\mathbf{\Theta}_{s_t,q} \neq \mathbf{0}$. The process $\mathbf{u} = (\mathbf{u}_t)$ is a zero mean white noise with $E(\mathbf{u}_t\mathbf{u}_\tau^{'}) = \delta_\tau^t\mathbf{I}_K$. The M-state Markov chain $s = (s_t)$ is irreducible, stationary and ergodic with transition matrix $\mathbf{P} = (p_{ij})$, where $p_{ij} = P(s_{t+1} = j | s_t = i)$, and stationary distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)^{'}$. Irreducibility implies that $\pi_m > 0$, for $m = 1, \ldots, M$, meaning that all unobservable states are possible. As remarked in Francq and Zakoïan (2001), Example 2, p.351, a Markov-switching moving-average process is always second-order stationary. It is sufficient to observe that the terms $\boldsymbol{\nu}_{s_t}$ and $\mathbf{\Theta}_{s_t,j}\mathbf{u}_{t-j}$, $j = 0, \ldots, q$ in (3.1) belong to the space of square summable vector function L^2. The Markov chain follows an AR(1) model

$$\boldsymbol{\xi}_t = \mathbf{P}^{'}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \tag{3.2}$$

where $\boldsymbol{\xi}_t$ is the random $M \times 1$ vector whose mth element is equal to 1 if $s_t = m$ and zero otherwise. The innovation $\mathbf{v} = (\mathbf{v}_t)$ is a zero mean martingale difference sequence with respect to an increasing σ-field (for more details, see Krolzig (1997), p.34). By direct computations, we have

$$\begin{aligned} &E(\boldsymbol{\xi}_t) = \boldsymbol{\pi} \\ &E(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t+h}^{'}) = \mathbf{D}\mathbf{P}^h \\ &E(\mathbf{v}_t \mathbf{v}_{\tau}^{'}) = \delta_{\tau}^{t}(\mathbf{D} - \mathbf{P}^{'}\mathbf{D}\mathbf{P}) \end{aligned} \tag{3.3}$$

where $\mathbf{D} = diag(\pi_1, \ldots, \pi_M)$ and $h \geq 0$ (here, and in the sequel, we use the convention that $A^h = \mathbf{I}$, identity matrix, if $h = 0$ for every square matrix A). We also assume that $(s_t, \mathbf{u}_t)$ is a strictly stationary process defined on some probability space, and that (s_t) is independent of $(\mathbf{u}_t)$. Our formulation includes the Hidden Markov chain processes of Krolzig (1997), Chp.3, and the Markov mean-variance switching models of Zhang and Stine (2001), Section 3.1, which is the case $q = 0$. Setting $\boldsymbol{\Lambda} = (\boldsymbol{\nu}_1 \ldots \boldsymbol{\nu}_M)$ and $\boldsymbol{\Theta}_j = (\boldsymbol{\Theta}_{1,j} \ldots \boldsymbol{\Theta}_{M,j})$ for $j = 0, \ldots, q$, where $\boldsymbol{\Theta}_0 = \boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_M)$, the process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) admits the following state-space representation

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\Lambda}\boldsymbol{\xi}_t + \sum_{j=0}^{q} \boldsymbol{\Theta}_j (\boldsymbol{\xi}_t \otimes \mathbf{I}_K) L^j \mathbf{u}_t \\ \boldsymbol{\xi}_t &= \mathbf{P}^{'}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t. \end{aligned} \tag{3.4}$$

Taking expectation gives $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}_t) = \boldsymbol{\Lambda} E(\boldsymbol{\xi}_t) = \boldsymbol{\Lambda}\boldsymbol{\pi}$ as $E(\boldsymbol{\xi}_t) = \boldsymbol{\pi}$. In the next theorem we compute the autocovariance function of the process $\mathbf{y} = (\mathbf{y}_t)$. This extends Theorem 3 from Zhang and Stine (2001) proved for the case $q = 0$.

Theorem 3.1. *The autocovariance function of the process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) is given by*

i) $$\Gamma_{\mathbf{y}}(h) = \boldsymbol{\Lambda}(\mathbf{Q}^{'})^h \mathbf{D}(\mathbf{I}_M - \delta_h^0 \mathbf{P}_{\infty})\boldsymbol{\Lambda}^{'} + \sum_{j=h}^{q} \boldsymbol{\Theta}_j ((\mathbf{P}^{'})^h \mathbf{D} \otimes \mathbf{I}_K)\boldsymbol{\Theta}_{j-h}^{'}$$

for $h = 0, \ldots, q$; *and*

ii) $\Gamma_{\mathbf{y}}(h) = \boldsymbol{\Lambda}(\mathbf{Q}^{'})^h \mathbf{D}\boldsymbol{\Lambda}^{'}$ *for every* $h \geq q + 1$

where $\mathbf{Q} = \mathbf{P} - \mathbf{P}_{\infty}$, $\mathbf{P}_{\infty} = \lim_n \mathbf{P}^n = \mathbf{i}_M \boldsymbol{\pi}^{'}$ *and* $\mathbf{i}_M = (1, 1, \ldots, 1)^{'}$.

Theorem 3.2. *Assume $\boldsymbol{\Lambda} \neq \mathbf{0}$. Then the process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) has a VARMA(p^*, q^*) representation, where $p^* \leq M - 1$ and $q^* \leq M + q - 1$.*

Note that the remaining case $\boldsymbol{\Lambda} = \mathbf{0}$ (and hence $\widetilde{\boldsymbol{\Lambda}} = \mathbf{0}$) will be included in the next Theorem 3.5. Now we use an argument discussed in Krolzig (1997), Section 2.3. The transition equation in (3.4) differs from a stable linear AR(1) process by the fact that one eigenvalue of $\mathbf{P}^{'}$ is equal to one and the covariance matrix of $\mathbf{v}_t$ is singular, due to the adding-up restriction $\mathbf{i}_M^{'}\boldsymbol{\xi}_t = 1$. For analytical purposes, a slightly different formulation of the transition equation is more useful, where the above restriction is eliminated. This procedure alters representation (3.4), and we consider a new state $(M-1)$-dimensional vector defined by $\boldsymbol{\delta}_t = (\xi_{1,t} - \pi_1 \ldots \xi_{M-1,t} - \pi_{M-1})^{'}$. The transition matrix, $\mathbf{F}$ say, associated with the state vector $\boldsymbol{\delta}_t$ is given by

$$\mathbf{F} = \begin{pmatrix} p_{11} - p_{M1} & \cdots & p_{M-1,1} - p_{M1} \\ \vdots & & \vdots \\ p_{1,M-1} - p_{M,M-1} & \cdots & p_{M-1,M-1} - p_{M,M-1} \end{pmatrix}$$

which is an $(M-1) \times (M-1)$ nonsingular matrix with all eigenvalues inside the unit circle. Then we have

$$\boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \tag{3.5}$$

where $\mathbf{w}_t = [\mathbf{I}_{M-1} \quad -\mathbf{i}_{M-1}]\mathbf{v}_t$. By direct computations, we have

$$\begin{aligned} &E(\boldsymbol{\delta}_t) = \mathbf{0} \\ &E(\boldsymbol{\delta}_t\boldsymbol{\delta}_{t+h}^{'}) = \widetilde{\mathbf{D}}(\mathbf{F}^{'})^h \\ &E(\mathbf{w}_t\mathbf{w}_\tau^{'}) = \delta_\tau^t(\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}) \end{aligned} \tag{3.6}$$

where $\widetilde{\mathbf{D}} = \mathbf{A}\mathbf{D}(\mathbf{I} - \mathbf{P}_\infty)\mathbf{A}^{'}$ and $\mathbf{A} = [\mathbf{I}_{M-1} \quad \mathbf{o}_{M-1}]$ is $(M-1) \times M$ (here $\mathbf{o}_{M-1}$ is the $(M-1) \times 1$ vector of zeros). More explicitly, we get

$$\widetilde{\mathbf{D}} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_{M-1} \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_{M-1} \\ \vdots & \vdots & & \vdots \\ -\pi_{M-1}\pi_1 & -\pi_{M-1}\pi_2 & \cdots & \pi_{M-1}(1-\pi_{M-1}) \end{pmatrix}.$$

We can see that $|\widetilde{\mathbf{D}}| = |\mathbf{D}| = \pi_1\pi_2\cdots\pi_M \neq 0$ as the Markov chain is irreducible. Now the measurement equation in (3.4) can be reformulated as

$$\mathbf{y}_t = \boldsymbol{\Lambda\pi} + \boldsymbol{\Lambda}(\boldsymbol{\xi}_t - \boldsymbol{\pi}) + \sum_{j=0}^{q} \boldsymbol{\Theta}_j[(\boldsymbol{\xi}_t - \boldsymbol{\pi}) \otimes \mathbf{I}_K]\mathbf{u}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_{t-j}.$$

Then the process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) admits a second state-space representation

$$\begin{aligned} &\mathbf{y}_t = \boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_t + \sum_{j=0}^{q} \widetilde{\boldsymbol{\Theta}_j}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_{t-j} \\ &\boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{aligned} \tag{3.7}$$

where $\widetilde{\mathbf{\Lambda}} = (\boldsymbol{\nu}_1 - \boldsymbol{\nu}_M \ \cdots \ \boldsymbol{\nu}_{M-1} - \boldsymbol{\nu}_M)$ and $\widetilde{\mathbf{\Theta}_j} = (\mathbf{\Theta}_{1,j} - \mathbf{\Theta}_{M,j} \ \cdots \ \mathbf{\Theta}_{M-1,j} - \mathbf{\Theta}_{M,j})$ for every $j = 0, \dots, q$. Equation (3.7) is also called the *unrestricted* state-space representation of $\mathbf{y}$, where $\mathbf{w} = (\mathbf{w}_t)$ is a martingale difference sequence with a nonsingular covariance matrix and the innovation sequence in the measurement equation is unaltered. Note that (3.7) can be written in short as

$$\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}} = \widetilde{\mathbf{\Lambda}}\boldsymbol{\delta}_t + \sum_{j=0}^{q} \widetilde{\mathbf{\Theta}_j}[(\boldsymbol{\delta}_t + \widetilde{\boldsymbol{\pi}}) \otimes \mathbf{I}_K] L^j \mathbf{u}_t$$

where $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}_t) = \mathbf{\Lambda}\boldsymbol{\pi}$ and $\widetilde{\boldsymbol{\pi}} = (\pi_1 - \pi_M \ \cdots \ \pi_{M-1} - \pi_M)^{'}$. Using representation (3.7) and doing computations similar to those in the proof of Theorem 3.1, we get

Theorem 3.3. *The autocovariance function of the process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) is given by*

$$i) \quad \Gamma_{\mathbf{y}}(h) = \widetilde{\mathbf{\Lambda}}\mathbf{F}^h\widetilde{\mathbf{D}}\widetilde{\mathbf{\Lambda}}^{'} + \sum_{j=h}^{q} \widetilde{\mathbf{\Theta}}_j[(\mathbf{F}^h\widetilde{\mathbf{D}}) \otimes \mathbf{I}_K]\widetilde{\mathbf{\Theta}}^{'}_{j-h} + \sum_{j=h}^{q} \mathbf{\Theta}_j[(\mathbf{D}\mathbf{P}_\infty) \otimes \mathbf{I}_K]\mathbf{\Theta}^{'}_{j-h}$$

$$\textit{for} \quad h = 0, \dots, q; \ \textit{and}$$

$$ii) \quad \Gamma_{\mathbf{y}}(h) = \widetilde{\mathbf{\Lambda}}\mathbf{F}^h\widetilde{\mathbf{D}}\widetilde{\mathbf{\Lambda}}^{'} \qquad \textit{for every} \qquad h \geq q+1.$$

Now $\Gamma_{\mathbf{y}}(h) = \widetilde{\mathbf{\Lambda}}\mathbf{F}^h\widetilde{\mathbf{D}}\widetilde{\mathbf{\Lambda}}^{'}$ for $h \geq q+1$ is in the form of Theorem 2.2, with $p = 0$ and $q+1$ instead of q. Since $\mathbf{F}$ is $(M-1) \times (M-1)$, we can apply directly Theorem 2.2 to get an alternative proof of Theorem 3.2, assuming that $\widetilde{\mathbf{\Lambda}} \neq \mathbf{0}$. However, the next two results are valid in the general case.

Theorem 3.4. *The process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) admits the MA(∞) representation*

$$\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}} = \widetilde{\mathbf{\Lambda}}\mathbf{F}(L)^{-1}\mathbf{w}_t + \sum_{j=0}^{q} \widetilde{\mathbf{\Theta}_j}[(\mathbf{F}(L)^{-1}\mathbf{w}_t) \otimes \mathbf{I}_K] L^j \mathbf{u}_t + \sum_{j=0}^{q} \mathbf{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K) L^j \mathbf{u}_t$$

where $\mathbf{F}(L) = \mathbf{I}_{M-1} - \mathbf{F}L$.

Now we compute explicitly a VARMA representation for the process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1). The final equation form of the stable representation could be very useful also when dealing with inference problems. Moreover, this gives a new proof of Theorem 3.2 and generalizes Proposition 2 from Krolzig (1997), Section 3.2.3.

Theorem 3.5. *The process $\mathbf{y} = (\mathbf{y}_t)$ in (3.1) admits a final equation form VARMA(p^*, q^*) representation with $p^* \leq M-1$ and $q^* \leq M+q-1$*

$$\gamma(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \mathbf{C}(L)\boldsymbol{\epsilon}_t$$

where $\gamma(L) = |F(L)|$ is the scalar AR operator of degree $M-1$, $\mathbf{C}(L)$ is the matrix lag polynomial of degree $M+q-1$ given by

$$\mathbf{C}(L) = [\widetilde{\boldsymbol{\Lambda}}\mathbf{F}^*(L) \quad \widetilde{\boldsymbol{\Theta}}_0(F^*(L)\otimes \mathbf{I}_K) \quad \cdots \quad \widetilde{\boldsymbol{\Theta}}_q(F^*(L)\otimes \mathbf{I}_K)L^q \quad |\mathbf{F}(L)|\sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\pi}\otimes \mathbf{I}_K)L^j]$$

and $\boldsymbol{\epsilon}_t = (\mathbf{w}_t^{'} \quad \mathbf{u}_t^{'}(\mathbf{w}_t^{'}\otimes \mathbf{I}_K) \cdots \mathbf{u}_t^{'}(\mathbf{w}_{t+q}^{'}\otimes \mathbf{I}_K) \quad \mathbf{u}_t^{'})^{'}$ is a zero mean vector white noise with $var(\boldsymbol{\epsilon}_t) = diag(\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'} \quad (\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'})\otimes \mathbf{I}_K \quad \cdots \quad (\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'})\otimes \mathbf{I}_K \quad \mathbf{I}_K)$. Here $\mathbf{F}^*(L)$ is the adjoint of $\mathbf{F}(L) = \mathbf{I}_{M-1} - \mathbf{F}L$. Note that $p^* = M-1$ and $q^* = M+q-1$ are satisfied if $\gamma(L)$ and $\mathbf{C}(L)$ are coprime, so the identification problem is completely solved, that is, $M = p^*+1$ and $q = q^*-p^*$ (hence $q^* \geq p^*$ in this case).

To end the section we treat the forecasting for the above Markov switching moving average model. Predictions of Markov switching VARMA models can be based on the state–space representations obtained above. By ignoring the parameter estimation problem, i.e., the fact that the parameters of the multivariate Markov switching process are unknown and must therefore be estimated, the mean squared prediction error optimal forecast can be generated by the conditional expectation $\widehat{\mathbf{y}}_{t+h|t} = E(\mathbf{y}_{t+h}|\mathbf{Y}_t)$, for $h \geq 1$, where $\mathbf{Y}_t = (\mathbf{y}_t^{'}, \mathbf{y}_{t-1}^{'}, \dots)^{'}$. From (3.2) we get the forecast of the hidden Markov chain, that is, $\widehat{\boldsymbol{\xi}}_{t+h|t} = (\mathbf{P}^{'})^h\widehat{\boldsymbol{\xi}}_{t|t}$, for $h \geq 1$, where $\widehat{\boldsymbol{\xi}}_{t|t} = E(\boldsymbol{\xi}_t|\mathbf{Y}_t)$. Equivalently, from (3.5) we obtain $\widehat{\boldsymbol{\delta}}_{t+h|t} = \mathbf{F}^h\widehat{\boldsymbol{\delta}}_{t|t}$, for $h \geq 1$, where $\widehat{\boldsymbol{\delta}}_{t|t}$ is the $(M-1)$ vector formed by the columns, but the last one, of $\widehat{\boldsymbol{\xi}}_{t|t} - \boldsymbol{\pi}$. Inserting this formula into Equation (3.7) gives the h–step predictor in the case of MS(M) MA(q) models. More precisely, we have

$$\begin{aligned}\widehat{\mathbf{y}}_{t+h|t} &= E(\boldsymbol{\Lambda}\boldsymbol{\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_{t+h} + \sum_{j=0}^{q}\widetilde{\boldsymbol{\Theta}}_j(\boldsymbol{\delta}_{t+h}\otimes \mathbf{I}_K)\mathbf{u}_{t+h-j} + \sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\pi}\otimes \mathbf{I}_K)\mathbf{u}_{t+h-j}|\mathbf{Y}_t)\\ &= \boldsymbol{\Lambda}\boldsymbol{\pi} + \widetilde{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\delta}}_{t+h|t} = \boldsymbol{\Lambda}\boldsymbol{\pi} + \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\widehat{\boldsymbol{\delta}}_{t|t}.\end{aligned}$$

Since the eigenvalues of $\mathbf{F}$ are all inside the unit circle, the forecasts of $\mathbf{y}_{t+h}$ converge to the unconditional mean of the process as h goes to infinity, that is, we have

$$\lim_{h\to\infty}\widehat{\mathbf{y}}_{t+h|t} = \boldsymbol{\Lambda}\boldsymbol{\pi} = \boldsymbol{\mu}_{\mathbf{y}}.$$

Our prediction formula $\widehat{\mathbf{y}}_{t+h|t}$ is based on the quantity $\widehat{\boldsymbol{\delta}}_{t|t}$, or equivalently, on $\widehat{\boldsymbol{\xi}}_{t|t}$. It is known that $\widehat{\boldsymbol{\xi}}_{t|t}$ can be computed by iterating on the following pair of recursive formulae

$$\widehat{\boldsymbol{\xi}}_{t|t} = \frac{\boldsymbol{\eta}_t\odot\widehat{\boldsymbol{\xi}}_{t|t-1}}{\boldsymbol{\eta}_t^{'}\widehat{\boldsymbol{\xi}}_{t|t-1}} \qquad\qquad \widehat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P}^{'}\widehat{\boldsymbol{\xi}}_{t|t}$$

where the symbol $\odot$ denotes the element-by-element multiplication and $\boldsymbol{\eta}_t$ is the $(M\times 1)$ vector whose jth component is the conditional density of $\mathbf{y}_t$ given $s_t = j$ and $\mathbf{Y}_{t-1}$. See, for example, Krolzig (1997), Chp.5, Formulae (5.4) and (5.5). The iteration is started by assuming that the initial state vector is drawn from the stationary unconditional probability distribution of the Markov chain, that is, $\widehat{\boldsymbol{\xi}}_{1|0} = \boldsymbol{\pi}$.

4. Markov Switching Autoregressive Models

Let $\mathbf{y} = (\mathbf{y}_t)$ be a K-dimensional second-order stationary dynamic process satisfying the following Markov switching autoregressive model (in short, MS(M)-VAR(p)):

$$\boldsymbol{\phi}_{s_t}(L)\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Sigma}_{s_t}\mathbf{u}_t \tag{4.1}$$

where $\mathbf{u}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$ and $\boldsymbol{\phi}_{s_t}(L) = \sum_{i=0}^{p} \boldsymbol{\phi}_{s_t,i}L^i$ with $\boldsymbol{\phi}_{s_t,0} = \mathbf{I}_k$ and $\boldsymbol{\phi}_{s_t,p} \neq \mathbf{0}$. As usual, we assume that the polynomials $|\boldsymbol{\phi}_{s_t}(z)|$ have all their roots strictly outside the unit circle. Sufficient conditions ensuring second-order stationarity for Markov-switching VAR models and Markov-switching VARMA models can be found, for example, in Karlsen (1990a and 1990b) and Francq and Zakoïan (2001), respectively. Moreover, in Francq and Zakoïan (2002) it was shown that, under appropriate moment conditions, the powers of the stationary solutions admit weak ARMA representations, which are potentially useful for statistical applications. Define

$$\boldsymbol{\Lambda} = (\boldsymbol{\nu}_1 \cdots \boldsymbol{\nu}_M) \qquad \boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \cdots \boldsymbol{\Sigma}_M)$$

and

$$\boldsymbol{\phi}(L) = (\sum_{i=0}^{p} \boldsymbol{\phi}_{1,i}L^i \cdots \sum_{i=0}^{p} \boldsymbol{\phi}_{M,i}L^i).$$

Then the process $\mathbf{y} = (\mathbf{y}_t)$ in (4.1) admits the following state-space representation

$$\begin{aligned} &\boldsymbol{\phi}(L)(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{y}_t = \boldsymbol{\Lambda}\boldsymbol{\xi}_t + \boldsymbol{\Sigma}(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_t \\ &\boldsymbol{\xi}_t = \mathbf{P}^{'}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t. \end{aligned} \tag{4.2}$$

Taking expectation gives $\boldsymbol{\phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)\boldsymbol{\mu}_{\mathbf{y}} = \boldsymbol{\Lambda}\boldsymbol{\pi}$. Assuming the invertibility of the $K \times K$ matrix $R = \boldsymbol{\phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)$, we can write $\boldsymbol{\mu}_{\mathbf{y}} = R^{-1}\boldsymbol{\Lambda}\boldsymbol{\pi}$. Set $\mathbf{x}_t = \boldsymbol{\Lambda}\boldsymbol{\xi}_t + \boldsymbol{\Sigma}(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_t$. For every $h \geq 0$ and assuming that the regime variable $\boldsymbol{\xi}_{t+h}$ is uncorrelated with $\mathbf{y}_t$, we have

$$\begin{aligned} cov(\mathbf{x}_{t+h}, \mathbf{y}_t) &= cov(\boldsymbol{\phi}(L)(\boldsymbol{\xi}_{t+h} \otimes \mathbf{I}_K)\mathbf{y}_{t+h}, \mathbf{y}_t) \\ &= \boldsymbol{\phi}(L)[E(\boldsymbol{\xi}_{t+h}) \otimes cov(\mathbf{y}_{t+h}, \mathbf{y}_t)] \\ &= \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)[1 \otimes cov(\mathbf{y}_{t+h}, \mathbf{y}_t)] \\ &= B(L)\Gamma_{\mathbf{y}}(h) \end{aligned}$$

where $B(L) = \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)$ is a $K \times K$ matrix lag polynomial of degree p. By explicit computations, we can see that $B(L) = \sum_{i=0}^{p} B_iL^i$, with $B_0 = \mathbf{I}_K$, where $B_i = \boldsymbol{\phi}_i(\boldsymbol{\pi} \otimes \mathbf{I}_K)$

is $K \times K$ and $\boldsymbol{\phi}_i = (\boldsymbol{\phi}_{1,i} \cdots \boldsymbol{\phi}_{M,i})$ is $K \times (KM)$ for every $i = 1, \ldots, p$. As done in Section 3, we can substitute $\boldsymbol{\xi}_t$ with the state $(M-1) \times 1$ vector $\boldsymbol{\delta}_t$ in order to obtain the unrestricted state-space representation

$$\begin{aligned} \widetilde{\boldsymbol{\phi}(L)}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{y}_t + \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{y}_t = \boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_t + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t \\ \boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{aligned} \tag{4.4}$$

where $\widetilde{\boldsymbol{\Lambda}} = (\boldsymbol{\nu}_1 - \boldsymbol{\nu}_M \cdots \boldsymbol{\nu}_{M-1} - \boldsymbol{\nu}_M)$, $\widetilde{\boldsymbol{\Sigma}} = (\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_M \cdots \boldsymbol{\Sigma}_{M-1} - \boldsymbol{\Sigma}_M)$ and $\widetilde{\boldsymbol{\phi}(L)} = (\sum_{i=1}^{p}(\boldsymbol{\phi}_{1,i} - \boldsymbol{\phi}_{M,i})L^i \cdots \sum_{i=1}^{p}(\boldsymbol{\phi}_{M-1,i} - \boldsymbol{\phi}_{M,i})L^i)$. From the transition equation in (4.4) we obtain $\boldsymbol{\delta}_{t+h} = \mathbf{F}^h\boldsymbol{\delta}_t + \sum_{j=0}^{h-1}\mathbf{F}^j\mathbf{w}_{t+h-j}$. Using this relation, $\mathbf{x}_{t+h}$ can be expressed as

$$\begin{aligned} \mathbf{x}_{t+h} =& \boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\boldsymbol{\delta}_t + \sum_{j=0}^{h-1}\widetilde{\boldsymbol{\Lambda}}\mathbf{F}^j\mathbf{w}_{t+h-j} + \widetilde{\boldsymbol{\Sigma}}[(\mathbf{F}^h\boldsymbol{\delta}_t) \otimes \mathbf{I}_K]\mathbf{u}_{t+h} \\ &+ \sum_{j=0}^{h-1}\widetilde{\boldsymbol{\Sigma}}[(\mathbf{F}^j\mathbf{w}_{t+h-j}) \otimes \mathbf{I}_K]\mathbf{u}_{t+h} + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_{t+h}. \end{aligned} \tag{4.5}$$

By (4.5), we obtain

$$cov(\mathbf{x}_{t+h}, \mathbf{y}_t) = cov(\boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\boldsymbol{\delta}_t, \mathbf{y}_t) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h cov(\boldsymbol{\delta}_t, \mathbf{y}_t) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h E(\boldsymbol{\delta}_t\mathbf{y}_t^{'}) \tag{4.6}$$

for every $h > 0$. For $h = 0$, we have

$$\begin{aligned} cov(\mathbf{x}_t, \mathbf{y}_t) &= cov(\boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_t + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t, \mathbf{y}_t) \\ &= \widetilde{\boldsymbol{\Lambda}}E(\boldsymbol{\delta}_t\mathbf{y}_t^{'}) + \widetilde{\boldsymbol{\Sigma}}E[(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t\mathbf{y}_t^{'}] + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)E(\mathbf{u}_t\mathbf{y}_t^{'}). \end{aligned} \tag{4.7}$$

Now we are going to compute $E(\boldsymbol{\delta}_t\mathbf{y}_t^{'})$, $E(\mathbf{u}_t\mathbf{y}_t^{'})$ and $E[(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t\mathbf{y}_t^{'}]$. Postmultiplying the measurement equation in (4.4) by $\boldsymbol{\delta}_t^{'}$ and taking expectation give

$$\boldsymbol{\phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)E(\mathbf{y}_t\boldsymbol{\delta}_t^{'}) = \widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{D}}$$

hence

$$E(\boldsymbol{\delta}_t\mathbf{y}_t^{'}) = \widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'}[R^{'}]^{-1}. \tag{4.8}$$

Postmultiplying the measurement equation in (4.2) and (4.4) by $\mathbf{u}_t^{'}$ and equating them, we get

$$\boldsymbol{\phi}(L)(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{y}_t\mathbf{u}_t^{'} = \boldsymbol{\Lambda\pi}\mathbf{u}_t^{'} + \widetilde{\boldsymbol{\Lambda}}\delta_t\mathbf{u}_t^{'} + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t\mathbf{u}_t^{'} + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t\mathbf{u}_t^{'}.$$

Taking expectation gives

$$\boldsymbol{\phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)E(\mathbf{y}_t\mathbf{u}_t^{'}) = \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)$$

hence

$$E(\mathbf{u}_t\mathbf{y}_t^{'}) = (\boldsymbol{\pi}^{'} \otimes \mathbf{I}_K)\boldsymbol{\Sigma}^{'}[R^{'}]^{-1}. \tag{4.9}$$

Reasoning as above by using $\mathbf{u}_t^{'}(\boldsymbol{\delta}_t^{'} \otimes \mathbf{I}_K)$ instead of $\mathbf{u}_t^{'}$, we get

$$\boldsymbol{\phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)E[\mathbf{y}_t\mathbf{u}_t^{'}(\boldsymbol{\delta}_t^{'} \otimes \mathbf{I}_K)] = \widetilde{\boldsymbol{\Sigma}}(\widetilde{\mathbf{D}} \otimes \mathbf{I}_K)$$

hence

$$E[(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t\mathbf{y}_t^{'}] = (\widetilde{\mathbf{D}} \otimes \mathbf{I}_K)\widetilde{\boldsymbol{\Sigma}}^{'}[R^{'}]^{-1}. \tag{4.10}$$

Substituting Formulae (4.8), (4.9) and (4.10) into (4.6) and (4.7), we get

$$cov(\mathbf{x}_{t+h}, \mathbf{y}_t) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'}[R^{'}]^{-1} \tag{4.11}$$

for every $h > 0$, and

$$cov(\mathbf{x}_t, \mathbf{y}_t) = [\widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'} + \boldsymbol{\Sigma}(\mathbf{D}\mathbf{P}_\infty \otimes \mathbf{I}_K)\boldsymbol{\Sigma}^{'} + \widetilde{\boldsymbol{\Sigma}}(\widetilde{\mathbf{D}} \otimes \mathbf{I}_K)\widetilde{\boldsymbol{\Sigma}}^{'}][R^{'}]^{-1}. \tag{4.12}$$

Collecting Formulae (4.3), (4.11) and (4.12) gives the following result:

Theorem 4.1. *Under the hypothesis that the regime variable is uncorrelated with the observable, the autocovariance function of the second-order stationary process* $\mathbf{y} = (\mathbf{y}_t)$ *in (4.1) is given by*

i) $\mathbf{B}(L)\Gamma_{\mathbf{y}}(0) = [\widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'} + \boldsymbol{\Sigma}(\mathbf{D}\mathbf{P}_\infty \otimes \mathbf{I}_K)\boldsymbol{\Sigma}^{'} + \widetilde{\boldsymbol{\Sigma}}(\widetilde{\mathbf{D}} \otimes \mathbf{I}_K)\widetilde{\boldsymbol{\Sigma}}^{'}][R^{'}]^{-1};$ *and*

ii) $\mathbf{B}(L)\Gamma_{\mathbf{y}}(h) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'}[R^{'}]^{-1}$ *for every* $h > 0$,

where $R = \boldsymbol{\phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)$ *has been assumed to be nonsingular.*

Applying Theorem 2.2 for $q = 1$ and taking in mind that $\mathbf{F}$ is $(M-1) \times (M-1)$, we get

Theorem 4.2. *Suppose that the regime variable* $\boldsymbol{\xi}_{t+h}$ *is uncorrelated with* $\mathbf{y}_t$ *for every* $h \geq 0$ *and* $\widetilde{\boldsymbol{\Lambda}} \neq \mathbf{0}$*. Then the MS(M)-VAR(p) process* $\mathbf{y} = (\mathbf{y}_t)$ *in (4.1) admits a stable VARMA*(p^*, q^*) *representation with* $p^* \leq M + p - 1$ *and* $q^* \leq M - 1$*. If we require that the autoregressive lag polynomial of such a stable representation is scalar, then the bounds become* $p^* \leq M + Kp - 1$ *and* $q^* \leq M + (K-1)p - 1$*. If there is no cancellation, then the identification problem is completely solved and the above relations become equalities. In particular, the last formulae imply* $M = K(q^*+1) - (K-1)(p^*+1)$ *and* $p = p^* - q^*$*.*

Theorem 4.2 can be reformulated by using the hypothesis $\boldsymbol{\Lambda} \neq \mathbf{0}$ as done in Theorem 3.2. This arises from the autocovariances expressed by using the matrix $\mathbf{Q}$ instead of $\mathbf{F}$, similarly as in Theorem 4.1. Moreover, to justify the hypothesis of having $\widetilde{\boldsymbol{\Lambda}} \neq \mathbf{0}$

(respectively, $\mathbf{\Lambda} \neq \mathbf{0}$), we refer to Krolzig (1997), p.53, line 15, where it was assumed the identifiability of the regimes, $\boldsymbol{\nu}_i \neq \boldsymbol{\nu}_j$ for $i \neq j$, in order to render the results unique. To end the section we compute explicitly a VARMA representation for the process $\mathbf{y} = (\mathbf{y}_t)$ in (4.1). This gives a new proof of Theorem 4.2 and extends Proposition 3 from Krolzig (1997), Section 3.2.4. We start with the more simple case in which the autoregressive lag polynomial of the initial process is state independent.

Theorem 4.3. *Under quite general regularity conditions, the process* $\mathbf{y} = (\mathbf{y}_t)$ *in (4.1), with* $\mathbf{\Phi}_{s_t}(L) = A(L)$ *is state independent, has a VARMA*(p^*, q^*) *representation with* $p^* \leq M + Kp - 1$ *and* $q^* \leq M + (K-1)p - 1$

$$\gamma(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \mathbf{C}(L)\boldsymbol{\epsilon}_t$$

where $\gamma(L) = |F(L)||A(L)|$ *is the scalar AR operator of degree* $M + Kp - 1$, $\mathbf{C}(L) = [A(L)^*\widetilde{\Lambda}F(L)^* \quad A(L)^*\widetilde{\Sigma}(F(L)^* \otimes \mathbf{I}_K) \quad |F(L)|A(L)^*\Sigma(\pi \otimes \mathbf{I}_K)]$ *is a matrix lag polynomial of degree* $M + (K-1)p - 1$, *and* $\boldsymbol{\epsilon}_t = (\boldsymbol{w}_t^{'} \quad \boldsymbol{u}_t^{'}(\boldsymbol{w}_t^{'} \otimes \mathbf{I}_K) \quad \boldsymbol{u}_t^{'})^{'}$ *is a zero mean vector white noise process with* $\mathrm{var}(\boldsymbol{\epsilon}_t) = \mathrm{diag}(\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}, (\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}) \otimes \mathbf{I}_K, \mathbf{I}_K)$.

In the general case in which the autoregression part of the initial process is state dependent but the regime variable is uncorrelated with the observable, we can proceed as follows. By Theorem 4.1 the autocovariances of the process satisfy a finite difference equation of order $p^* = M + Kp - 1$ and rank $q^* + 1 = M + (K-1)p$. Then the process can be represented by a stable VARMA(p^*, q^*), whose autoregression lag polynomial is assumed to be scalar. Given the process $(\mathbf{y}_t)$, we can estimate the coefficients of the stable VARMA(p^*, q^*) via OLS. If there is no cancellation between the AR and MA part of the estimated VARMA(p^*, q^*), then we get the representation of Theorem 4.2 with equalities.

To complete the section we also discuss the forecasting for our Markov switching autoregressive model. So let us consider the MS(M)-VAR(p) model in (4.4). Then we can write

$$\mathbf{y}_t + \sum_{i=1}^{p} \widetilde{\mathbf{\Phi}}_i(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{y}_{t-i} + \sum_{i=1}^{p} \mathbf{\Phi}_i(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{y}_{t-i} = \mathbf{\Lambda}\boldsymbol{\pi} + \widetilde{\mathbf{\Lambda}}\boldsymbol{\delta}_t + \widetilde{\mathbf{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \mathbf{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t$$

where $\mathbf{\Phi}_i = (\mathbf{\Phi}_{1,i} \cdots \mathbf{\Phi}_{M,i})$ and $\widetilde{\mathbf{\Phi}}_i = (\mathbf{\Phi}_{1,i} - \mathbf{\Phi}_{M,i} \cdots \mathbf{\Phi}_{M-1,i} - \mathbf{\Phi}_{M,i})$, for every $i = 1, \ldots, p$ (see Section 4). The one-step predictor $\widehat{\mathbf{y}}_{t+1|t}$ can be calculated as above, so we get

$$\widehat{\mathbf{y}}_{t+1|t} = \mathbf{\Lambda}\boldsymbol{\pi} + \widetilde{\mathbf{\Lambda}}\mathbf{F}\widehat{\boldsymbol{\delta}}_{t|t} - \sum_{i=1}^{p} \widetilde{\mathbf{\Phi}}_i(\mathbf{F} \otimes \mathbf{I}_K)(\widehat{\boldsymbol{\delta}}_{t|t} \otimes \mathbf{I}_K)\mathbf{y}_{t+1-i} - \sum_{i=1}^{p} \mathbf{\Phi}_i(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{y}_{t+1-i}.$$

For h–step predictions, $h > 1$, the task is much more complicated, and the last formula generalizes as follows

$$\widehat{\mathbf{y}}_{t+h|t} = \mathbf{\Lambda}\boldsymbol{\pi} + \widetilde{\mathbf{\Lambda}}\mathbf{F}^h\widehat{\boldsymbol{\delta}}_{t|t} - \sum_{i=1}^{p} \widetilde{\mathbf{\Phi}}_i(\mathbf{F} \otimes \mathbf{I}_K)^h(\widehat{\boldsymbol{\delta}}_{t|t} \otimes \mathbf{I}_K)\widehat{\mathbf{y}}_{t+h-i|t} - \sum_{i=1}^{p} \mathbf{\Phi}_i(\boldsymbol{\pi} \otimes \mathbf{I}_K)\widehat{\mathbf{y}}_{t+h-i|t}$$

which in practice gives a recursive formula. Also in this case, taking the limits for $h \to \infty$ on both sides yields

$$\lim_{h\to\infty} \widehat{\mathbf{y}}_{t+h|t} = \mathbf{\Lambda}\boldsymbol{\pi} - \sum_{i=1}^{p} \mathbf{\Phi}_i(\boldsymbol{\pi} \otimes \mathbf{I}_K) \lim_{h\to\infty} \widehat{\mathbf{y}}_{t+h-i|t}$$

hence

$$\lim_{h\to\infty} \widehat{\mathbf{y}}_{t+h|t} = R^{-1}\mathbf{\Lambda}\boldsymbol{\pi} = \boldsymbol{\mu}_{\mathbf{y}}.$$

where $R = \mathbf{\Phi}(1)(\boldsymbol{\pi} \otimes \mathbf{I}_K)$. However, in applied work it is customary to follow a suggestion of Doan, Litterman and Sims (1984) for which the sequence of predicted values $\{\widehat{\mathbf{y}}_{t+1|t}, \ldots, \widehat{\mathbf{y}}_{t+h|t}, \ldots\}$ is substituted by the sequence $\{\widehat{\mathbf{y}}_{t+1|t}, \ldots, \widehat{\mathbf{y}}_{t+h|t+h-1}, \ldots\}$. See, for example, Krolzig (1997), Section 4.4, for more details on this construction. Of course, the calculation of the filtered regime probabilities $\widehat{\boldsymbol{\xi}}_{t|t}$ (and hence $\widehat{\boldsymbol{\delta}}_{t|t}$) can be performed by the recursive formulae listed at the end of Section 3.

5. Data Simulation

In this section, we perform some MonteCarlo experiment for the estimation of the number of states given by the estimated lower bound obtained in the previous sections and penalized likelihood criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC). For the computation of the orders of the stable VARMA we use the 3-pattern method (TPM) proposed by Choi (1992). We perform three experiments. The first is taken from Zhang and Stine (2001) in which we generate three different Poisson Markov regime switching models with two states. The second is a Markov switching process with two states and autoregressive dynamic with one lag (in short, MS(2)-AR(1) process) and the third is a two-states Markov switching with two lags in the autoregressive part (in short, MS(2)-AR(2)). The data-generating processes have gaussian i.i.d. errors and the parameters are reported below in Table 1, 2 and 3.

With regards to the first experiment, we consider three different Markov switching Poisson models (Table 1). The first model, denoted as case A, is a one-state model; that is, y_t is an i.i.d. sequence of Poisson random variables with λ= 4. Cases B and C are two-state models and share the same transition matrix but different means. With λ_1= 4 and $\lambda_2 = 12$, the two states of case B are more distinct than those of case C with $\lambda_1 = 4$ and $\lambda_2 = 6$.

Parameters	1.A		1.B		1.C	
P	1	0	0.8	0.2	0.8	0.2
	1	0	0.2	0.8	0.2	0.8
λ	4	4	4	12	4	6

Table 1 – First experiment: Poisson Markov regime switching models with two states. In table we report the transition matrices Ps and Poisson Means λ for cases A–C.

Concerning with the second experiment, the parameters are set in accordance to Table 2. It is a Markov switching process with two states and autoregressive dynamic with one lag and it can be written as $y_t = \mu_{s_t} + \phi_{s_t} y_{t-1} + \sigma_{s_t} u_t$ with $s_t \in \{1, 2\}$. The three models in A, B and C share the same transition probability matrix and the values for the intercept and the variance but there are diffences in the autoregressive parameters (this follow some experiments as in Psaradakis and Spagnolo, 2003). The last case D, instead, has a different transition matrix which gives a more persistent chain.

Parameters	**2.A**		**2.B**		**2.C**		**2.D**	
P	0.6	0.4	0.6	0.4	0.6	0.4	0.8	0.2
	0.4	0.6	0.4	0.6	0.4	0.6	0.2	0.8
μ	0	3	0	3	0	3	0	3
σ	1	$\sqrt{1.5}$	1	$\sqrt{1.5}$	1	$\sqrt{1.5}$	1	$\sqrt{1.5}$
ϕ	0.3	0.3	0.3	0.6	0.3	0.9	0.3	0.6

Table 2 – Second experiment: Markov switching process with two states and autoregressive dynamic with one lag. In table we report the transition matrices Ps and the parameters of the process (means μs, standard deviation σs and autoregressive coefficients ϕs) for cases A–D.

Finally, in the third experiment, we consider a Markov switching process with two states and autoregressive dynamic with two lags, written as $y_t = \mu_{s_t} + \phi_{1s_t} y_{t-1} + \phi_{2s_t} y_{t-2} + \sigma_{s_t} u_t$ with $s_t \in \{1, 2\}$. Here we want to compare the performance of the bounds proposed in the present paper (we will denote it by CAV) with those proposed by Zhang and Stine (2001) (in short, ZS) and Francq and Zakoïan (2001) (in short, FZ) for those autoregressive markov switching models. Case B considers a more persistent process compared to the baseline case A and case C instead considers same autoregressive coefficients in different states.

Parameters	**3.A**		**3.B**		**3.C**	
P	0.6	0.4	0.9	0.1	0.6	0.4
	0.4	0.6	0.1	0.9	0.4	0.6
μ	0	3	0	3	0	3
σ	1	$\sqrt{1.5}$	1	$\sqrt{1.5}$	1	$\sqrt{1.5}$
ϕ_1	0.3	0.6	0.3	0.6	0.3	0.3
ϕ_2	0.4	0.8	0.4	0.8	0.1	0.1

Table 3 – Third experiment: Markov switching process with two states and autoregressive dynamic with two lags. In table we report the transition matrices Ps and the parameters of the process (means μs, standard deviation σs and autoregressive coefficients ϕ_1s and ϕ_2s) for cases A–C.

The experiments simulate artificial time series of length $T+50$ with $T \in \{100, 500, 1000\}$; the first 50 initial data points are discarded to minimize the effect of initial conditions. 100 MonteCarlo replications are carried out for each trial. When complexity-penalized likelihood criteria are computed, we use the recursive procedure discussed by Hamilton and the penalization constants are the usual proposed in that literature (1 for AIC, $\frac{1}{2}lnN$ for BIC and $lnlnN$ for HQC).

The simulation results from the first experiment are reported in Table 4. With respect to the one-state case (A) only the TPM applied on our bounds seems to correctly predict the exact number of states most of the times, while the likelihood criteria seem to overestimate that. The same happens for case B and in case C for larger samples. Overall, the TPM does better than any other likelihood methods.

		1.A				1.B				1.C			
N	$\hat{M}$	AIC	BIC	HQC	TPM	AIC	BIC	HQC	TPM	AIC	BIC	HQC	TPM
100	1	0	0	0	45	0	0	0	5	0	0	0	39
	2	53	74	66	15	22	41	26	64	59	77	67	26
	3	47	26	34	40	78	59	74	31	41	23	33	35
500	1	0	0	0	44	0	0	0	0	0	0	0	26
	2	28	58	44	25	4	6	5	62	11	51	27	36
	3	72	42	56	31	96	94	95	38	89	49	73	38
1000	1	0	0	0	40	0	0	0	0	0	0	0	8
	2	19	58	31	23	0	2	2	60	8	27	11	62
	3	81	42	69	37	100	98	98	40	92	73	89	28

Table 4 – Simulation results of the first experiment: Poisson Markov regime switching models with two states for cases A–C. We report the number of regimes chosen by the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC). TPM denotes the number of regimes chosen in accordance to the bounds presented in the present paper.

With regards to the second experiment, the results are shown in Table 5. Here close conclusions can be drawn. In fact, the likelihood methods overestimate the number of states, with the exception of the Bayesian Criterion (BIC) in small sample, while the TPM detects it most of the times. These conclusions are robust to the change in the transition probability matrix, when choosing a more persistent chain (case D).

		2.A				2.B			
N	$\hat{M}$	AIC	BIC	HQC	TPM	AIC	BIC	HQC	TPM
100	1	0	0	0	0	0	0	0	0
	2	38	87	65	75	24	63	40	76
	3	62	13	35	25	76	37	60	24
500	1	0	0	0	0	0	0	0	0
	2	0	21	1	74	1	3	1	70
	3	100	79	99	26	99	97	99	30
1000	1	0	0	0	0	0	0	0	0
	2	0	0	0	68	0	0	0	72
	3	100	100	100	32	100	100	100	28

		2.C				2.D			
N	$\hat{M}$	AIC	BIC	HQC	TPM	AIC	BIC	HQC	TPM
100	1	0	0	0	0	0	0	0	0
	2	13	48	21	74	10	58	24	75
	3	87	52	79	26	90	42	76	25
500	1	0	0	0	0	0	0	0	0
	2	9	16	13	75	0	2	1	82
	3	91	84	87	25	100	98	99	18
1000	1	0	0	0	0	0	0	0	0
	2	10	17	14	76	0	0	0	74
	3	90	83	86	24	100	100	100	26

Table 5 – Simulation results of the second experiment: Markov switching process with two states and autoregressive dynamic with one lag (in short, MS(2)-AR(1) process) for cases A–D. We report the number of regimes chosen by the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC). TPM denotes the number of regimes chosen in accordance to the bounds presented in the present paper.

The results of the third experiment are in Table 6. As expected from theoretical aspects, using the bounds of Zhang and Stine (2001) (ZS in short) or Francq and Zakoïan (2001) (FZ in short) we tend to underestimate the number of states since these bounds are larger and then less informative. Whereas when using the bounds obtained in this work (denote by CAV), we are able to detect the exact number of regimes most of the time; and this choice is robust to the change of the transition probability matrix (case B) or of the values of the autoregressive coefficients (case C).

		3.A			3.B			3.C		
N	$\hat{M}$	CAV	ZS	FZ	CAV	ZS	FZ	CAV	ZS	FZ
100	1	0	100	100	0	100	100	1	100	100
	2	73	0	0	96	0	0	75	0	0
	3	27	0	0	4	0	0	24	0	0
500	1	0	100	100	0	100	100	0	100	100
	2	75	0	0	20	0	0	70	0	0
	3	25	0	0	80	0	0	30	0	0
1000	1	0	100	100	0	100	100	0	100	100
	2	75	0	0	50	0	0	60	0	0
	3	25	0	0	50	0	0	40	0	0

Table 6 – Simulation results of the third experiment: Markov switching process with two states and autoregressive dynamic with two lags (in short, MS(2)-AR(2) process) for cases A–C. We report the number of regimes chosen by using the three-pattern method and applying different bounds either those obtained in Zhang and Stine (2001) (ZS), Francq and Zakoïan (2001) (FZ) or in the present paper (CAV).

6. Application on foreign exchange rates

As an application on real data, we want to consider and complete the example of Zhang and Stine (2001) on foreign exchange rates. The data are the same used in Engle and

Hamilton (1990), who consider quarterly data for French franc, British pound and German mark for the period from 1973:Q3 to 1988:Q4. Engle and Hamilton (1990) proposed to model the logarithm of exchange rates as a two states Markov-switching autoregressive of order one. In line with Zhang and Stine (2001), when we fit Gaussian regime switching models, penalized likelihood criteria as Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC) choose $M = 1$ for all three currencies while Akaike Information Criterion (AIC) chooses 2 regimes for franc and pound and only one regime for the mark. What is interesting is that, when we compute the orders of the stable VARMA and evaluate the lower bounds as proposed in the papers of Zhang and Stine (2001) (ZS), Francq and Zakoïan (2001) (FZ) and in the present one (CAV), we find that our bounds are more precise and then more informative. In particular, our bounds propose the existence of two regimes for franc and mark and one regime for the pound (as reported in Table 7), while using the bounds of ZS and FZ we are not able to infer any information on regime switching from the data. Finally, note that the last methodology compared to penalized methods is less demanding and computationally faster since it does not request likelihood calculations.

Criteria	**Franc**	**Pound**	**Mark**
AIC	2	2	1
BIC	1	1	1
HQC	1	1	1
CAV	2	1	2
ZS	0	0	0
FZ	1	0	1

Table 7 – Estimates of the number of regimes for quarterly data on French franc, British pound and German mark for the period from 1973:Q3 to 1988:Q4 based on penalized likelihood criteria (Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC)) and based on three-pattern method computing the lower bounds as in Zhang and Stine (2001) (ZS), Francq and Zakoïan (2001) (FZ) or in the present paper (CAV).

7. Conclusion

In this paper, for M-state Markov switching multivariate moving average models and autoregressive models in which the regime variable is uncorrelated with the observable, we give finite order VARMA(p^*, q^*) representations. The parameters of the VARMA can be determined by evaluating the autocovariance function of the Markov-switching models. It turns out that upper bounds for p^* and q^* are elementary functions of the dimension K of the process, the number M of regimes, and the orders p and q. In particular, the order of the stable VARMA admits a simple form: $p^* \leq M-1$, $q^* \leq M+q-1$ for M-state switching

VMA(q) models and $p^* \leq M + Kp - 1$, $q^* \leq M + p(K-1) + q - 1$ for M-state switching VAR(p) models. This result yields an easily computed method for setting a lower bound on the number of regimes from an estimated autocovariance function. Our results include, as particular cases, those obtained by Krolzig (1997), and improve the bounds found in the literature in the works of Zhang and Stine (2001) and Francq and Zakoïan (2001) for our classes of dynamic models. Our simulation results indicate the procedure is more precise than penalized likelihood criteria such as AIC, BIC and HQC which require more elaborate procedures and assumption of a specific probability model and the associated likelihood calculations. Moreover, having bounds for the number of states which are small than those of Zhang and Stine (2001) or Francq and Zakoïan (2001) give estimates for the number of states which are more precise and then more informative. This is shown both with simulated experiments and with real data application on exchange rates.

Acknowledgements. I would like to thank my PhD supervisor Professor Monica Billio and the Department of Economics at University of Venice, Italy, where this work was carried out. Furthermore, many thanks are due to the participants at the workshop held at CEMFI (Madrid) and at the CFE 2012 conference (Oviedo) for useful comments and remarks. I also thank Christian Francq and Jean-Michel Zakoïan for their kind suggestions and comments which were most valuable in writing the final version of the paper.

References

[1] Choi, B. S. (1992) ARMA Model Identification, New York, Springer-Verlag.

[2] Doan, T., Litterman, R. and Sims, C. (1984) *Forecasting and conditional projection using realistic prior distributions*, Econometric Reviews **3**, 1–100.

[3] Engle, C. and Hamilton, J.D. (1990) *Long swings in the dollar: are they in the data and do markets know it?*, American Economic Review **80**, 689–713.

[4] Francq, C. and Zakoïan, J.M. (2001) *Stationarity of Multivariate Markov-Switching ARMA Models*, Journal of Econometrics **102**, 339–364.

[5] Francq, C. and Zakoïan, J.M. (2002) *Autocovariance Structure of Powers of Switching-Regime ARMA Processes*, ESAIM: Probability and Statistics **6**, 259–270.

[6] Hansen B.E., (1992) *The likelihood Ratio Test under Nonstandard Conditions: testing the Markov Switching Model of GNP*, Journal of Applied Econometrics **7**, S61–S82.

[7] Karlsen, H.A (1990a) *Doubly stochastic vector AR(1) processes*, Technical Report, Department of Mathematics, University of Bergen, Norway.

[8] Karlsen, H.A (1990b) *A class of non-linear time series models*, Ph.D. Thesis, University of Bergen, Norway.

[9] Krolzig, H.M. (1997) *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis*, Springer Verlag, Berlin-Heidelberg-New York.

[10] Olteanu, M. and Rynkiewicz, J. (2007) *Estimating the number of regimes in a switching autoregressive model*, HAL (Hyper Articles en Ligne), CCSD (Centre pour la Communication Scientifique Directe).

[11] Psaradakis, Z. and Spagnolo, N. (2003) *On the determination of the number of regimes in Markov-switching autoregressive models*, Journal of Time Series Analysis **24** (2), 237–252.

[12] Psaradakis, Z. and Spagnolo, N. (2006) *Joint determination of the state dimension and autoregressive order for models with in Markov regime switching*, Journal of Time Series Analysis **27** (5), 753–766.

[13] Ríos, R. and Rodríguez, L.A. (2008) *Penalized estimate of the number of states in Gaussian linear AR with Markov regime*, Electronic Journal of Statistics, **2**, 1111–1128.

[14] Zhang, J. and Stine, R.A. (2001) *Autocovariance structure of Markov regime switching models and model selection.* Journal of Time Series Analysis **22**(1), 107–124.

Appendix

In this section, we give proofs of Theorems in Sections 3 and 4.

Proof of Theorem 3.1. The following are well-known facts (see, for example, Zhang and Stine (2001), Section 3.1): $\mathbf{DP}_\infty = \boldsymbol{\pi}\boldsymbol{\pi}'$, $\mathbf{P}_\infty^n = \mathbf{P}^n\mathbf{P}_\infty = \mathbf{P}_\infty\mathbf{P}^n = \mathbf{P}_\infty$ and $\mathbf{Q}^n = \mathbf{P}^n - \mathbf{P}_\infty$ for every $n \geq 1$. First we treat the case $h = 0$. Then we have

$$\begin{aligned}\Gamma_{\mathbf{y}}(0) &= E(\mathbf{y}_t\mathbf{y}_t') - E(\mathbf{y}_t)E(\mathbf{y}_t') \\ &= E(\mathbf{y}_t\mathbf{y}_t') - \boldsymbol{\Lambda}\boldsymbol{\pi}\boldsymbol{\pi}'\boldsymbol{\Lambda}' \\ &= E(\mathbf{y}_t\mathbf{y}_t') - \boldsymbol{\Lambda}\mathbf{DP}_\infty\boldsymbol{\Lambda}'\end{aligned}$$

and

$$\begin{aligned}E(\mathbf{y}_t\mathbf{y}_t') &= E[(\boldsymbol{\Lambda}\boldsymbol{\xi}_t + \sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_{t-j})(\sum_{j=0}^{q}\mathbf{u}_{t-j}'(\boldsymbol{\xi}_t' \otimes \mathbf{I}_K)\boldsymbol{\Theta}_j' + \boldsymbol{\xi}_t'\boldsymbol{\Lambda}')] \\ &= \boldsymbol{\Lambda}E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')\boldsymbol{\Lambda}' + \sum_{j=0}^{q}\boldsymbol{\Theta}_j[E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') \otimes \mathbf{I}_K]\boldsymbol{\Theta}_j' \\ &= \boldsymbol{\Lambda}\mathbf{D}\boldsymbol{\Lambda}' + \sum_{j=0}^{q}\boldsymbol{\Theta}_j(\mathbf{D} \otimes \mathbf{I}_K)\boldsymbol{\Theta}_j'\end{aligned}$$

hence

$$\Gamma_{\mathbf{y}}(0) = \boldsymbol{\Lambda}\mathbf{D}(\mathbf{I}_M - \mathbf{P}_\infty)\boldsymbol{\Lambda}' + \sum_{j=0}^{q}\boldsymbol{\Theta}_j(\mathbf{D} \otimes \mathbf{I}_K)\boldsymbol{\Theta}_j'$$

which proves i) for $h = 0$. For $h = 1, \ldots, q$, we have

$$\Gamma_{\mathbf{y}}(-h) = cov(\mathbf{y}_t, \mathbf{y}_{t+h}) = E(\mathbf{y}_t\mathbf{y}_{t+h}') - E(\mathbf{y}_t)E(\mathbf{y}_{t+h}') = E(\mathbf{y}_t\mathbf{y}_{t+h}') - \boldsymbol{\Lambda}\mathbf{DP}_\infty\boldsymbol{\Lambda}'$$

and

$$\begin{aligned}E(\mathbf{y}_t\mathbf{y}_{t+h}') &= E[(\boldsymbol{\Lambda}\boldsymbol{\xi}_t + \sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_{t-j})(\sum_{i=0}^{q}\mathbf{u}_{t+h-i}'(\boldsymbol{\xi}_{t+h}' \otimes \mathbf{I}_K)\boldsymbol{\Theta}_i' + \boldsymbol{\xi}_{t+h}'\boldsymbol{\Lambda}')] \\ &= \boldsymbol{\Lambda}E(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}')\boldsymbol{\Lambda}' + \sum_{j=0}^{q}\sum_{i=0}^{q}\boldsymbol{\Theta}_j[E(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}') \otimes \delta_{t+h-i}^{t-j}\mathbf{I}_K]\boldsymbol{\Theta}_i' \\ &= \boldsymbol{\Lambda}\mathbf{DP}^h\boldsymbol{\Lambda}' + \sum_{i=0}^{q-h}\boldsymbol{\Theta}_i[(\mathbf{DP}^h) \otimes \mathbf{I}_K]\boldsymbol{\Theta}_{i+h}'\end{aligned}$$

hence

$$\begin{aligned}\Gamma_{\mathbf{y}}(-h) &= \boldsymbol{\Lambda}\mathbf{D}(\mathbf{P}^h - \mathbf{P}_\infty)\boldsymbol{\Lambda}' + \sum_{i=0}^{q-h}\boldsymbol{\Theta}_i[(\mathbf{DP}^h) \otimes \mathbf{I}_K]\boldsymbol{\Theta}_{i+h}' \\ &= \boldsymbol{\Lambda}\mathbf{DQ}^h\boldsymbol{\Lambda}' + \sum_{i=0}^{q-h}\boldsymbol{\Theta}_i[(\mathbf{DP}^h) \otimes \mathbf{I}_K]\boldsymbol{\Theta}_{i+h}'.\end{aligned}$$

Now taking transposition and setting $j = i + h$, we get

$$\Gamma_{\mathbf{y}}(h) = \mathbf{\Lambda}(\mathbf{Q}^{'})^h \mathbf{D}\mathbf{\Lambda}^{'} + \sum_{j=h}^{q} \mathbf{\Theta}_j[((\mathbf{P}^{'})^h \mathbf{D}) \otimes \mathbf{I}_K]\mathbf{\Theta}^{'}_{j-h}$$

which proves i) for $h = 1, \ldots, q$. For every $h \geq q + 1$, we have

$$E(\mathbf{y}_t \mathbf{y}^{'}_{t+h}) = \mathbf{\Lambda} E(\boldsymbol{\xi}_t \boldsymbol{\xi}^{'}_{t+h})\mathbf{\Lambda}^{'} = \mathbf{\Lambda}\mathbf{D}\mathbf{P}^h \mathbf{\Lambda}^{'}$$

and

$$\Gamma_{\mathbf{y}}(-h) = \mathbf{\Lambda}\mathbf{D}(\mathbf{P}^h - \mathbf{P}_\infty)\mathbf{\Lambda}^{'} = \mathbf{\Lambda}\mathbf{D}\mathbf{Q}^h \mathbf{\Lambda}^{'}$$

hence

$$\Gamma_{\mathbf{y}}(h) = \mathbf{\Lambda}(\mathbf{Q}^{'})^h \mathbf{D}\mathbf{\Lambda}^{'}$$

which proves ii). $\Box$

Proof of Theorem 3.2. For $h \geq q + 1$, the autocovariance function $\Gamma_{\mathbf{y}}(h) = \mathbf{\Lambda}(\mathbf{Q}^{'})^h \mathbf{D}\mathbf{\Lambda}^{'}$ is in the form specified in Theorem 2.2 with $p = 0$ and $q + 1$ instead of q. As remarked in Zhang and Stine (2001) (see the proof of Theorem 3), the minimal polynomial of Q has a zero constant term as Q is singular. So the proof is now slightly different from that of Theorem 2.2. Let λ_i, $i = 1, \ldots, M$, be the eigenvalues of $\mathbf{P}^{'}$, where we set $\lambda_1 = 1$ as $\mathbf{P}^{'}\boldsymbol{\pi} = \boldsymbol{\pi}$. Since the Markov chain is ergodic, all other eigenvalues of $\mathbf{P}^{'}$ are inside the unit circle. It follows that the eigenvalues of $\mathbf{Q}^{'}$ are $\mu_1 = 0$ and $\mu_i = \lambda_i - \lim_n \lambda_i^n = \lambda_i$ for $i = 2, \ldots, M$. Since $\mathbf{Q}^{'}$ is an $M \times M$ singular matrix, its minimal polynomial can be written as $\varphi(x) = x^M - f_1 x^{M-1} - \cdots - f_{M-1}x$, where the coefficient f_{M-1} may be zero. An argument similar to that used in the proof of Theorem 2.2 gives

$$\Gamma_{\mathbf{y}}(M + q) - f_1 \Gamma_{\mathbf{y}}(M + q - 1) - \cdots - f_{M-1}\Gamma_{\mathbf{y}}(q + 1) = 0.$$

The result now follows from Theorem 2.1 with $h = M + q$, $p^* \leq M - 1$ and $q^* \leq h - 1$. $\Box$

Proof of Theorem 3.4 From (3.5) we get $\boldsymbol{\delta}_t = (\mathbf{I}_{M-1} - \mathbf{F}L)^{-1}\mathbf{w}_t = \mathbf{F}(L)^{-1}\mathbf{w}_t$ (here we have used the fact that all the eigenvalues of $\mathbf{F}$ are less than 1 in modulus). Inserting the above relation in (3.7) gives the result of the statement. $\Box$

Proof of Theorem 3.5 Premultiplying by $|\mathbf{F}(L)|$ the MA(∞) representation of Theorem 3.4 yields

$$|\mathbf{F}(L)|(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \widetilde{\mathbf{\Lambda}}\mathbf{F}^*(L)\mathbf{w}_t + \sum_{j=0}^{q} \widetilde{\mathbf{\Theta}}_j[(\mathbf{F}^*(L)\mathbf{w}_t) \otimes \mathbf{I}_K]L^j \mathbf{u}_t + |\mathbf{F}(L)| \sum_{j=0}^{q} \mathbf{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K)L^j \mathbf{u}_t$$

$$= \widetilde{\mathbf{\Lambda}}\mathbf{F}^*(L)\mathbf{w}_t + \sum_{j=0}^{q} \widetilde{\mathbf{\Theta}}_j[\mathbf{F}^*(L) \otimes \mathbf{I}_K]L^j(\mathbf{w}_{t+j} \otimes \mathbf{I}_K)\mathbf{u}_t + |\mathbf{F}(L)| \sum_{j=0}^{q} \mathbf{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K)L^j \mathbf{u}_t$$

which is a VARMA(p^*, q^*) representation as claimed in the statement. □

Proof of Theorem 4.3 From (4.4) we get $\boldsymbol{\delta}_t = F(L)^{-1}\mathbf{w}_t$ as usual. Equating (4.1) and (4.4) and substituting the last formula, we get

$$\begin{aligned} A(L)\mathbf{y}_t &= \boldsymbol{\Lambda}\boldsymbol{\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_t + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t \\ &= A(1)\boldsymbol{\mu}_{\mathbf{y}} + \widetilde{\boldsymbol{\Lambda}}F(L)^{-1}\mathbf{w}_t + \widetilde{\boldsymbol{\Sigma}}(F(L)^{-1} \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t \end{aligned}$$

hence

$$A(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \widetilde{\boldsymbol{\Lambda}}F(L)^{-1}\mathbf{w}_t + \widetilde{\boldsymbol{\Sigma}}(F(L)^{-1} \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t. \tag{4.13}$$

Premultiplying (4.13) by $|F(L)|$ yields

$$|F(L)|A(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \widetilde{\boldsymbol{\Lambda}}F(L)^*\mathbf{w}_t + \widetilde{\boldsymbol{\Sigma}}(F(L)^* \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K)\mathbf{u}_t + |F(L)|\boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t. \tag{4.14}$$

Now the regularity conditions of the statement mean that $A(L)$ is invertible, that is, $A(L)^*A(L) = |A(L)|\mathbf{I}_K$. Premultiplying (4.14) by $A(L)^*$, we get the VARMA(p^*, q^*) representation, with $p^* \leq M + Kp - 1$ and $q^* \leq M + (K-1)p - 1$ (use the fact that the degree of $|A(L)|$ is Kp):

$$\begin{aligned} |F(L)||A(L)|(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) =& A(L)^*\widetilde{\boldsymbol{\Lambda}}F(L)^*\mathbf{w}_t + A(L)^*\widetilde{\boldsymbol{\Sigma}}(F(L)^* \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K)\mathbf{u}_t \\ &+ |F(L)|A(L)^*\boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t \end{aligned}$$

which is a model as required in the statement. □