Central European Journal of Economic Modelling and Econometrics

Autocovariance and Linear Transformations of Markov Switching VARMA Processes

Maddalena Cavicchioli*

Submitted: 1.08.2014, Accepted: 22.11.2014

Abstract

We study the autocovariance structure of a general Markov switching second-order stationary VARMA model. Then we give stable finite order VARMA(p^*, q^*) representations for those M-state Markov switching VARMA(p, q) processes where the observables are uncorrelated with the regime variables. This allows us to obtain sharper bounds for p^* and q^* with respect to the ones existing in literature. Our results provide new insights into stochastic properties and facilitate statistical inference about the orders of MS-VARMA models and the underlying number of hidden states.

Keywords: time series, multivariate ARMA, state-space models, Markov chains, changes in regime, autocovariance, linear representations

JEL Classification: C01, C05, C32

*University of Modena and Reggio Emilia; e-mail: maddalena.cavicchioli@unimore.it

1 Introduction

Investigations of economic systems and models make use of linear transformations of multivariate stochastic processes in various ways, for instance, when linear aggregates or subprocesses are considered. It is known that linear transformations of VARMA models are again VARMA processes, as proved by Lütkepohl (1984). In the present paper we consider dynamic models whose parameters can change as a result of a regime shift variable, governed by an unobserved Markov chain. Such models have attracted much interest in the literature for their applications in areas as economics, statistics, and finance. As general references see, for example, Hamilton (2005), Krolzig (1997), and Guidolin (2012). The achievements of the Markov switching (MS) models in fitting empirical data have been confirmed in many economic studies. For this reason, it is important to characterize the theoretical properties of these models. Stationarity conditions and the autocovariance structure for MS-VARMA models were derived by Yang (2000) and Francq and Zakoïan (2001). An interesting question arising in this context is to investigate the state dimension of the MS process. The current methods for determining the number of regimes are based either on complexity-penalized likelihood criteria (see, for example, Psaradakis and Spagnolo (2003), Olteanu and Rynkiewicz (2007), Ríos and Rodríguez (2008)) or on finite-order stable VARMA representations of the initial switching models (see, for example, Krolzig (1997), Zhang and Stine (2001), Francq and Zakoïan (2001), and Cavicchioli (2014)). The parameters of the VARMA representations can be determined by evaluating the autocovariance function of the MS 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/or the orders of the switching moving-average autoregression. Some bounds are previously determined by Krolzig (1997), Zhang and Stine (2001), Francq and Zakoïan (2001), and Cavicchioli (2014) for some Markov regime switching models of different type. Now we state the main results of the paper (the specifics of the models and the standard regularity conditions will be explained in detail in the next sections).

Theorem 1.1 *Let* $\mathbf{y} = (\mathbf{y_t})$ *be a* K*-dimensional second-order stationary process which satisfies the* M*-state Markov switching (also in the intercept term) VARMA*(p,q) *model, in short MS(*M*)-VARMA*(p,q)*:*

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

where $\boldsymbol{\phi}_{s_t}(L) = \sum_{i=0}^{p} \boldsymbol{\phi}_{s_t,i} L^i$, $\boldsymbol{\phi}_{s_t,0} = \mathbf{I}_K$*,* $\boldsymbol{\phi}_{s_t,p} \neq \mathbf{0}$*,* $\boldsymbol{\Theta}_{s_t}(L) = \sum_{j=0}^{q} \boldsymbol{\Theta}_{s_t,j} L^j$*,* $\boldsymbol{\Theta}_{s_t,0} = \boldsymbol{\Sigma}_{s_t}$ *(non-singular* $K \times K$ *matrix),* $\boldsymbol{\Theta}_{s_t,q} \neq \mathbf{0}$*, and* $\mathbf{u}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$*.*

Suppose that the regime variable is uncorrelated with the observables. Under standard regularity conditions, $\mathbf{y} = (\mathbf{y}_t)$ *admits a stable VARMA(*p^*, q^**) representation, where* $p^* \leq M + p - 1$ *and* $q^* \leq M + q - 1$*. 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* $q^* \leq M + (K-1)p + q - 1$*. If the AR and MA polynomials of the former stable representation have top degrees and no common roots, the above relations become equalities, that is,* $p^* = M + p - 1$ *and* $q^* = M + q - 1$*.*

Here the hypothesis assuming uncorrelated regime variable with the observables is always satisfied when the intercept term is absent (when, of course, the innovation process is independent of the Markov chain). However, it is also a reasonable assumption if the researcher believes changing in regime is due to an event outside the economic system. The above result allows us to determine a lower bound for the number of states.

Corollary 1.2 *The number of states* M *of the Markov chain* $s = (s_t)$ *in the model above satisfies* $M \geq \max\{p^* - p + 1, q^* - q + 1\}$*, where* p^* *and* q^* *are the orders of the stable VARMA representation of* $\mathbf{y} = (\mathbf{y}_t)$*.*

The rest of the paper is organized as follows. In Section 2 we define notations and give some preliminary results on the autocovariance structure of the MS(M)-VMA(q) process which will be useful to prove the main theorem in the following section. In fact, in Section 3 we state general results for a second–order stationary MS(M)-VARMA(p, q), $p, q \geq 0$, in terms of its autocovariances and linear representation. An empirical application is shown in Section 4. Finally, Section 5 concludes. Proofs are given in the Appendix.

2 Notations and preliminary results

This section is devoted to introduce notations and some preliminary results which we shall use to prove the main theorem. First, we consider Markov switching vector moving-average process, in short MS(M)-VMA(q):

$$\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Theta}_{s_t}(L)\mathbf{u}_t \tag{1}$$

Here we allow 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, $\boldsymbol{\Theta}_{s_t}(L) = \sum_{j=0}^{q} \boldsymbol{\Theta}_{s_t,j} L^j$ is $K \times K$ matrix polynomial in the lag operator L, with $\boldsymbol{\Theta}_{s_t,0} = \boldsymbol{\Sigma}_{s_t}$ (non-singular $K \times K$ matrix) and $\boldsymbol{\Theta}_{s_t,q} \neq \mathbf{0}$. The process $\mathbf{u} = (\mathbf{u}_t)$ is a zero-mean white noise with $E(\mathbf{u}_t \mathbf{u}_\tau^{'}) = \delta_\tau^t \mathbf{I}_K$ (through the paper, δ_τ^t denotes the Kronecker symbol). 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 $m = 1, \dots, M$, 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 $\boldsymbol{\Theta}_{s_t,j}\mathbf{u}_{t-j}$, $j = 0, \dots, q$ in (1) belong to the L^2 space of square-summable vector functions. The Markov chain follows an AR(1) process

$$\boldsymbol{\xi}_t = \mathbf{P}^{'}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \tag{2}$$

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 process $\mathbf{v} = (\mathbf{v}_t)$ is a martingale difference sequence with zero mean. By direct computations, we have

$$E(\boldsymbol{\xi}_t) = \boldsymbol{\pi} \qquad E(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}^{'}) = \mathbf{D}\mathbf{P}^h \qquad E(\mathbf{v}_t\mathbf{v}_\tau^{'}) = \delta_\tau^t(\mathbf{D} - \mathbf{P}^{'}\mathbf{D}\mathbf{P})$$

where $\mathbf{D} = diag(\pi_1, \dots, \pi_M)$ 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 $\boldsymbol{\Lambda} = (\boldsymbol{\nu}_1 \dots \boldsymbol{\nu}_M)$ and $\boldsymbol{\Theta}_j = (\boldsymbol{\Theta}_{1,j} \dots \boldsymbol{\Theta}_{M,j})$ for $j = 0, \dots, q$, where $\boldsymbol{\Theta}_0 = \boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \dots \boldsymbol{\Sigma}_M)$, the process $\mathbf{y} = (\mathbf{y}_t)$ in (1) admits the following state-space representation:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\Lambda}\boldsymbol{\xi}_t + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)L^j\mathbf{u}_t \\ \boldsymbol{\xi}_t &= \mathbf{P}^{'}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \end{aligned} \tag{3}$$

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 2.1 *The autocovariance function of the process* $\mathbf{y} = (\mathbf{y}_t)$ *in* (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 $h = 0, \dots, q$;

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

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

Now we use an argument discussed in Krolzig (1997), Section 2.3. The transition equation in (3) 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), and we consider a new, $(M-1)$-dimensional state vector defined by $\boldsymbol{\delta}_t = (\xi_{1,t} - \pi_1 \ \ldots \ \xi_{M-1,t} - \pi_{M-1})^{'}$. The transition matrix, say $\mathbf{F}$, 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 $(M-1) \times (M-1)$ matrix with all eigenvalues inside the unit circle. Then we have

$$\boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \tag{4}$$

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

$$E(\boldsymbol{\delta}_t) = \mathbf{0} \qquad E(\boldsymbol{\delta}_t\boldsymbol{\delta}_{t+h}^{'}) = \widetilde{\mathbf{D}}(\mathbf{F}^{'})^h \qquad E(\mathbf{w}_t\mathbf{w}_\tau^{'}) = \delta_\tau^t(\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'})$$

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) can be reformulated as

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

Then the process $\mathbf{y} = (\mathbf{y}_t)$ in (1) admits another state-space representation:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\Lambda}\boldsymbol{\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_t + \sum_{j=0}^{q} \widetilde{\boldsymbol{\Theta}_j}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_{t-j} \\ \boldsymbol{\delta}_t &= \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{aligned} \tag{5}$$

where

$$\widetilde{\boldsymbol{\Lambda}} = (\boldsymbol{\nu}_1 - \boldsymbol{\nu}_M \ \cdots \ \boldsymbol{\nu}_{M-1} - \boldsymbol{\nu}_M) \qquad \widetilde{\boldsymbol{\Theta}_j} = (\boldsymbol{\Theta}_{1,j} - \boldsymbol{\Theta}_{M,j} \ \cdots \ \boldsymbol{\Theta}_{M-1,j} - \boldsymbol{\Theta}_{M,j})$$

for every $j = 0, \ldots, q$. Equations given by (5) are also called the unrestricted state-space representation of $\mathbf{y}$, where $\mathbf{w} = (\mathbf{w}_t)$ is a martingale difference sequence with a non-singular covariance matrix and the innovation sequence in the measurement equation is unaltered. Note that the measurement equation in (5) can be written in short as

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

where $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}_t) = \boldsymbol{\Lambda}\boldsymbol{\pi}$ and $\widetilde{\boldsymbol{\pi}} = (\pi_1 - \pi_M \ \cdots \ \pi_{M-1} - \pi_M)^{'}$. Using representation (5) and doing computations similar to those in the proof of Theorem 2.1, we get

Theorem 2.2 *The autocovariance function of the process* $\mathbf{y} = (\mathbf{y}_t)$ *in* (1) *is given by*

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

$$\text{for} \quad h = 0, \ldots, q;$$

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

3 MS(M)-VARMA(p, q) models

Let us consider $\mathbf{y} = (\mathbf{y}_t)$ be a K-dimensional second-order stationary dynamic process satisfying the following Markov switching autoregressive moving average model

$$\boldsymbol{\phi}_{s_t}(L)\,\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Theta}_{s_t}(L)\,\mathbf{u}_t \tag{6}$$

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 in Section 2, $\boldsymbol{\Theta}_{s_t}(L) = \sum_{j=0}^{q} \boldsymbol{\Theta}_{s_t,j} L^j$, with $\boldsymbol{\Theta}_{s_t,0} = \boldsymbol{\Sigma}_{s_t}$ (a non-singular $K \times K$ matrix) and $\boldsymbol{\Theta}_{s_t,q} \neq \mathbf{0}$. 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 Yang (2000) and Francq and Zakoïan (2001), respectively. 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 (6) admits the following state-space representation:

$$\begin{aligned} &\boldsymbol{\phi}(L)(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{y}_t = \boldsymbol{\Lambda}\boldsymbol{\xi}_t + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t \\ &\boldsymbol{\xi}_t = \mathbf{P}^{'}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \end{aligned} \tag{7}$$

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 $\mathbf{x}_t = \boldsymbol{\Lambda}\boldsymbol{\xi}_t + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t$. For every $h \geq 0$ and assuming that the regime variable $\boldsymbol{\xi}_{t+h}$ is uncorrelated with $\mathbf{y}_t$, we have

$$\begin{aligned} cov(\mathbf{x}_{t+h}, \mathbf{y}_t) &= cov(\boldsymbol{\phi}(L)(\boldsymbol{\xi}_{t+h} \otimes \mathbf{I}_K)\mathbf{y}_{t+h}, \mathbf{y}_t) = \\ &= \boldsymbol{\phi}(L)[E(\boldsymbol{\xi}_{t+h}) \otimes cov(\mathbf{y}_{t+h}, \mathbf{y}_t)] = \\ &= \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)[1 \otimes cov(\mathbf{y}_{t+h}, \mathbf{y}_t)] = \\ &= B(L)\Gamma_{\mathbf{y}}(h) \end{aligned} \tag{8}$$

where $B(L) = \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)$ is a $K \times K$ matrix lag polynomial of degree p. Here L denotes also the backward shift operator, i.e., $L\,\boldsymbol{\Gamma}_{\mathbf{y}}(h) = \boldsymbol{\Gamma}_{\mathbf{y}}(h-1)$. By explicit computations, we can see that $B(L) = \sum_{i=0}^{p} B_i L^i$, 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 2, we can substitute $\boldsymbol{\xi}_t$ with the $(M-1) \times 1$ state vector $\boldsymbol{\delta}_t$ in order to obtain the unrestricted state-space representation

$$\begin{aligned} &\widetilde{\boldsymbol{\phi}(L)}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{y}_t + \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{y}_t = \boldsymbol{\Lambda}\boldsymbol{\pi} + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\delta}_t + \\ &\qquad + \sum_{j=0}^{q} \widetilde{\boldsymbol{\Theta}}_j(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t + \sum_{j=0}^{q} \boldsymbol{\Theta}_j(\boldsymbol{\pi} \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t \\ &\boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{aligned} \tag{9}$$

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 (9) 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

$$\begin{aligned}\mathbf{x}_{t+h} =& \boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\boldsymbol{\delta}_t + \sum_{r=0}^{h-1}\widetilde{\boldsymbol{\Lambda}}\mathbf{F}^r\mathbf{w}_{t+h-r} + \sum_{j=0}^{q}\widetilde{\boldsymbol{\Theta}}_j[(\mathbf{F}^h\boldsymbol{\delta}_t)\otimes\mathbf{I}_K]\,L^j\,\mathbf{u}_{t+h} \\ &+ \sum_{j=0}^{q}\sum_{r=0}^{h-1}\widetilde{\boldsymbol{\Theta}}_j[(\mathbf{F}^r\mathbf{w}_{t+h-r})\otimes\mathbf{I}_K]\,L^j\,\mathbf{u}_{t+h} + \sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\pi}\otimes\mathbf{I}_K)\,L^j\,\mathbf{u}_{t+h}\end{aligned}$$

By this formula, we obtain

$$cov(\mathbf{x}_{t+h},\mathbf{y}_t) = cov(\boldsymbol{\Lambda\pi} + \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\boldsymbol{\delta}_t,\mathbf{y}_t) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h cov(\boldsymbol{\delta}_t,\mathbf{y}_t) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h E(\boldsymbol{\delta}_t\mathbf{y}_t^{'}) \qquad (10)$$

for every $h \geq q+1$.
Now we explicitly compute $E(\boldsymbol{\delta}_t\mathbf{y}_t^{'})$. Postmultiplying the measurement equation in (9) by $\boldsymbol{\delta}_t^{'}$ and taking expectation give the relation $\boldsymbol{\phi}(1)(\boldsymbol{\pi}\otimes\mathbf{I}_K)E(\mathbf{y}_t\boldsymbol{\delta}_t^{'}) = \widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{D}}$, hence

$$E(\boldsymbol{\delta}_t\mathbf{y}_t^{'}) = \widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'}[R^{'}]^{-1} \qquad (11)$$

Substituting Formulae (8) and (11) into (10), we get the following result:

Theorem 3.1 *The autocovariance function of the second-order stationary process* $\mathbf{y} = (\mathbf{y}_t)$ *in* (6) *satisfies the matrix relation*

$$\mathbf{B}(L)\Gamma_{\mathbf{y}}(h) = \widetilde{\boldsymbol{\Lambda}}\mathbf{F}^h\widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Lambda}}^{'}[R^{'}]^{-1}$$

for every $h \geq q+1$, *where* $R = \boldsymbol{\phi}(1)(\boldsymbol{\pi}\otimes\mathbf{I}_K)$ *has been assumed to be non-singular.*

Now we apply Theorem 4 of Cavicchioli (2014) replacing q by $q+1$ and M by $M-1$ (bearing in mind that $\mathbf{F}$ is $(M-1)\times(M-1)$). This gives the following result:

Theorem 3.2 *Suppose that the regime variable is uncorrelated with the observables and* $\widetilde{\boldsymbol{\Lambda}} \neq \mathbf{0}$. *Then the process* $\mathbf{y} = (\mathbf{y}_t)$ *in* (6) *admits a stable VARMA*(p^*, q^*) *representation with* $p^* \leq M+p-1$ *and* $q^* \leq M+q-1$. *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* $q^* \leq M+(K-1)p+q-1$. *If the AR and MA polynomials of the former stable representation have top degrees and no common roots, the above relations become equalities, that is,* $p^* = M+p-1$ *and* $q^* = M+q-1$.

Now we compute explicitly a stable VARMA representation for the process $\mathbf{y} = (\mathbf{y}_t)$ in (6). This gives a new proof of Theorem 3.2 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 3.3 *Let us consider the process* $\mathbf{y} = (\mathbf{y}_t)$ *in (6), with* $\boldsymbol{\phi}_{s_t}(L) = A(L) = \sum_{i=0}^{p} A_i L^i$ *being state independent and* $A_0 = \mathbf{I}_K$*. Assume that the reduced transition matrix* $\mathbf{F}$ *and the pth autoregressive matrix* A_p *are non-singular. Then* $\mathbf{y}$ *has a VARMA(*p^*, q^**) representation with* $p^* \leq M + p - 1$ *and* $q^* \leq M + q - 1$*. 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* $q^* \leq M + (K-1)p + q - 1$*. More precisely, there exists a stable finite order VARMA(*p^*, q^**) representation*

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

where $\gamma(L) = |F(L)||A(L)|$ *is the scalar AR operator of degree* $M + Kp - 1$*,* $\mathbf{C}(L)$ *is a matrix lag polynomial of degree* $M + (K-1)p + q - 1$*, and the process* $\boldsymbol{\epsilon}_t = (\boldsymbol{w}_t^{'} \quad \boldsymbol{u}_t^{'}(\boldsymbol{w}_t^{'} \otimes \mathbf{I}_K) \quad \ldots \quad \boldsymbol{u}_t^{'}(\boldsymbol{w}_{t+q}^{'} \otimes \mathbf{I}_K) \qquad \boldsymbol{u}_t^{'})^{'}$ *is a zero-mean vector white noise with covariance matrix given by*

$$\mathrm{var}(\boldsymbol{\epsilon}_t) = \mathrm{diag}(\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}, (\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}) \otimes \mathbf{I}_K, \ldots, (\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}) \otimes \mathbf{I}_K, \mathbf{I}_K).$$

In the general case in which the autoregression part of the initial process is state dependent (assuming again that the regime variable is uncorrelated with the observables), we can proceed as follows. By Theorem 3.1 the autocovariances of the process satisfy a finite difference equation of order $p^* = M + p - 1$ and rank $q^* + 1 = M + q$. Then the process can be represented by a stable VARMA(p^*, q^*) model. 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 3.2 with equalities.

4 An empirical application

Here we show how Theorem 3.2 can be used for model selection in the case where q (resp. p) and M are all unknown, given the value of p (resp. q). We propose the following algorithm to estimate simultaneously the number of regimes and the order of the autoregressive (AR) (resp. moving-average (MA)) part of the considered MS-VARMA process, given the MA (resp. AR) order.

Step 1. Use one of the existing model selection criteria (for example, the Box-Jenkins strategy or the Choi method (1992)) to estimate the orders of the stable VARMA from Theorem 3.2. Let $\widehat{p}^*$ and $\widehat{q}^*$ denote such estimates.

Step 2. If the AR and MA polynomials of the stable VARMA representation have no roots in common, Theorem 3.2 gives the estimates $\widehat{M} = \widehat{p}^* - p + 1$ and $\widehat{q} = \widehat{q}^* - \widehat{p}^* + p$ for M and q unknown, given the value of p. Analogously, Theorem 3.2 gives the estimates $\widehat{M} = \widehat{q}^* - q + 1$ and $\widehat{p} = \widehat{p}^* - \widehat{q}^* + q$ for M and p unknown, given the value of q.

Moreover, note that if the researcher has a clear prior belief on the number of regimes (states of nature) to be considered in a particular application, then the algorithm is convenient to correctly specify the lag order of the AR and MA parts before proceeding to the estimation and forecast exercises. This is the starting point in the following application of the above algorithm, in which we focus on model selection issue introduced in the paper by Bergman and Hansson (2005). These authors propose to model the real exchange rate between the major currencies in the post-Bretton Woods period as a stationary 2-state Markov switching AR(1) model. In particular, they argue that unit roots can be rejected when allowing structural instability in the time series. A natural implication is to associate the parameters of the model to different regimes or states of nature. In the context of nominal exchange rates, it is straightforward to distinguish between two regimes corresponding to exchange rate appreciation and depreciation. Having this in mind, Bergman and Hansson (2005) estimate exchange rate series by means of univariate MS(2)-AR(1) models where only the intercept varies across regimes. Then they show that this model outperforms both the single regime random walk and the 2-regimes random walk models (where the autoregressive coefficients are equal to unity). Now we want to test whether the data are correctly modeled via autoregressive processes and how many lags should be considered. The data consist of quarterly observations on the period average spot exchange rates (in units of foreign currency per US dollar) and the consumer price index for six major industrialized countries (UK, France, Germany, Switzerland, Canada, Japan and USA) taken from the IMF International Financial Statistics CD-ROM. The sample runs from the second quarter of 1973 to the fourth quarter of 1997. The effective sample of observations is 99. The real exchange rate is normalized to unity in the second quarter of 1973 and we use 100 times the natural logarithm of the real exchange rate, just as in Bergman and Hansson (2005). Table 1 reports the estimated p^* and q^* orders using Box-Jenkins procedure together with the implied switching model using the bounds described above. Since it is reasonable to assume 2-regimes, we evaluate the AR and MA orders to asses the correct model and to discriminate whether only the intercept switches or also the other parameters.

Table 1: Estimated switching model orders for the univariate series of the six real exchange rates, 1973:2-1990:4.

Rate	$ARMA(\hat{p}^*, \hat{q}^*)$	Bounds	MS-Model
GBP	ARMA(1,0)	$1 \le 2+p-1, 0 \le 2+q-1$	MS(2)-ARMA(0,0)
FRF	ARMA(2,0)	$2 \le 2+p-1, 0 \le 2+q-1$	MS(2)-ARMA(1,0)
DEM	ARMA(2,0)	$2 \le 2+p-1, 0 \le 2+q-1$	MS(2)-ARMA(1,0)
CHF	ARMA(2,0)	$2 \le 2+p-1, 0 \le 2+q-1$	MS(2)-ARMA(1,0)
CAD	ARMA(1,0)	$1 \le 2+p-1, 0 \le 2+q-1$	MS(2)-ARMA(0,0)
JPY	ARMA(2,0)	$2 \le 2+p-1, 0 \le 2+q-1$	MS(2)-ARMA(1,0)

This analysis suggests that four (FRF, DEM, CHF, JPY) out of six series should be correctly modeled using a process in which both the intercept and the first lag autoregressive coefficient switch, while GBP and CAD can be considered by means of the specification used in Bergman and Hansson (2005). However, this analysis is in line with the conclusions of Bergman and Hansson (2005), which are in favour of a switching model. It also supports the theoretical predictions by Dumas (1992), which imply a Markov switching model whose intercept, autoregressive parameter and variance should all depend on the state.

5 Conclusions

We have investigated the linear representations of Markov switching (MS) VARMA(p, q) models, in which the coefficients are functions of a finite state-space Markov chain. It has already been established in the literature that such models admit finite order VARMA(p^*, q^*) representations, but we obtain sharper bounds for the stable VARMA orders. These linear representations are potentially useful for statistical applications since inference within MS models is notoriously difficult. Determining lower bounds for the number of regimes or the order of AR and MA switching polynomials is the first step in the statistical inference in MS models and in real-world data applications where structural instability should be taken into account.

Acknowledgements

I would like to thank the Editor-in-Chief and the anonymous referee for their useful suggestions and comments which were most valuable in writing the final version of the paper.

References

[1] Bergman, U. M., and Hansson, J. (2005) Real exchange rates and switching regimes, *Journal of International Money and Finance* **24** (1), 121-138.

[2] Cavicchioli, M. (2014) Determining the number of regimes in Markov–switching VAR and VMA models, *Journal of Time Series Analysis* **35**, 173-186.

[3] Dumas, B. (1992) Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**, 153-180.

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

[5] Guidolin, M. (2012) *Markov Switching Models in Empirical Finance*, Advances in Econometrics (D. Drukker et al., editors), Emerald Publishers Ltd.

[6] Hamilton, J.D. (2005) *Regime Switching Models*, New Palgrave Dictionary of Economics. BROOKS, chapter 9, Palgrave Macmillan.

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

[8] Lütkepohl, H. (1984) Linear transformations of vector ARMA processes, *Journal of Econometrics* **26**, 283-293.

[9] 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).

[10] 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.

[11] 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.

[12] Yang, M. (2000) Some properties of vector autoregressive processes with Markov switching coefficients, *Econometric Theory* **16**, 23-43.

[13] 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.

Proof of Theorem 2.1

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

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

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). □

Proof of Theorem 3.3

From (9) we get $\boldsymbol{\delta}_t = F(L)^{-1}\mathbf{w}_t$ as usual. Equating (6) and (9) and substituting the last formula, we get

$$A(L)\mathbf{y}_t = \mathbf{\Lambda}\boldsymbol{\pi} + \widetilde{\mathbf{\Lambda}}\boldsymbol{\delta}_t + \sum_{j=0}^{q} \widetilde{\mathbf{\Theta}}_j (\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t + \sum_{j=0}^{q} \mathbf{\Theta}_j (\boldsymbol{\pi} \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t = A(1)\boldsymbol{\mu}_{\mathbf{y}} +$$

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

hence

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

$$+ \sum_{j=0}^{q} \mathbf{\Theta}_j (\boldsymbol{\pi} \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t.$$

Premultiplying this formula by $|F(L)|$ yields

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

$$+ |F(L)| \sum_{j=0}^{q} \mathbf{\Theta}_j (\boldsymbol{\pi} \otimes \mathbf{I}_K)\, L^j\, \mathbf{u}_t.$$

This equation is a stable VARMA(p^*, q^*) representation with $p^* \leq M + p - 1$ and $q^* \leq M + q - 1$. Since the reduced transition matrix $\mathbf{F}$ is non-singular, the degree of $|F(L)|$ is $M-1$. Now, we can write $A(L)^* A(L) = |A(L)|\mathbf{I}_K$, where the degree of $|A(L)|$ is Kp because the pth autoregressive $K \times K$ matrix A_p is non-singular. Premultiplying the last equation by $A(L)^*$, we get the stable VARMA(p^*, q^*) representation, with $p^* \leq M + Kp - 1$ and $q^* \leq M + (K-1)p + q - 1$, whose autoregression lag polynomial is scalar:

$$|F(L)||A(L)|(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = A(L)^* \widetilde{\boldsymbol{\Lambda}} F(L)^* \mathbf{w}_t +$$

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

which is a model as required in the statement. □