AStA Advances in Statistical Analysis (2021) 105:431–449
https://doi.org/10.1007/s10182-020-00383-4

ORIGINAL PAPER

OLS Estimation of Markov switching VAR models: asymptotics and application to energy use

Maddalena Cavicchioli[1]

Received: 27 January 2020 / Accepted: 12 October 2020 / Published online: 22 October 2020

Abstract

We show that the ordinary least squares (OLS) estimates of population parameters for Markov switching vector autoregressive (MS VAR) models coincide with the maximum likelihood estimates. Then, we propose an algorithm in matrix form for the estimation of model parameters, and derive an explicit expression in closed-form for the asymptotic covariance matrix of the OLS estimator of such models. The obtained characterization of the asymptotic variance is new to our knowledge. It is easier to program than the usual approach based on second derivatives, and more accurate. Our theorems generalize the classical results known for a linear VAR process, and complete those existing in the literature on the estimation of the asymptotic covariance matrix for multivariate stationary time series. Numerical simulations are provided to illustrate the obtained theoretical results. Finally, an application on energy use and economic growth in the Euro area gives some insights on the nonlinear nature of the corresponding time series, and reproduces the major stylized facts.

Keywords Markov switching VAR model · OLS estimator · Asymptotic covariance matrix · Energy use · Economic growth

JEL Classification C32 · C34 · C13

1 Introduction

Economic time series prediction deals with the task of modelling the underlying data generation process using past observations and using the model to extrapolate the time series into the future. In the literature, two principal classes of models were

✉ Maddalena Cavicchioli
maddalena.cavicchioli@unimore.it

1 Department of Economics "Marco Biagi", University of Modena and Reggio Emilia, Viale Berengario 51, 41121 Modena, Italy

studied for the purpose of forecasting i.e., the statistical time series models and the structural econometric models.

Linear multivariate time series models, like vector autoregressive moving-average (VARMA) models, were among the first to be developed and widely used in statistics and economics to model simultaneously the dynamics of a set of endogenous variables. A leading class is formed by vector autoregressive (VAR) models, as special case of VARMA, pioneered by Sims (1980). More recently, Fernández-Villaverde et al. (2007) showed that linearized dynamic stochastic general equilibrium (DSGE) models imply that the variables of interest are generated by a VARMA process. For textbook discussions on several model specifications as well as estimation and inference procedures for linear time series see Hamilton (1994) and Lütkepohl (2007).

Another important aspect of the modelling process is the evaluation of the fit of a selected model. Several authors were interested in goodness-of-fit tests for VARMA models i.e., tests which can be applied when no *a priori* information is available about what departures from the null should be anticipated. The null hypothesis is that the observed multivariate process is a VARMA process with fixed and minimal autoregressive and moving-average orders. A goodness-of-fit approach for univariate ARMA models has been developed in Ubierna and Velilla (2007), and then generalized to a multivariate setting by Velilla and Thu (2018). The idea is to consider a stochastic process based on a modified residual autocorrelation sequence, that is shown to converge to the Brownian bridge. The associated test statistics, constructed as goodness-of-fit functionals, have then standard limit distributions, that are free of unknown parameters. Visual goodness-of-fit criteria for fitting a linear bivariate VAR model, yielding also indications on the necessary order, were developed by Ioannidis (2007). This author gave a characterization of the range of spectral density matrices, which are feasible for bivariate VAR models.

Despite their simplicity and versatily in modelling various types of economic time series, VARMA models were constrained by their linear scope. However, real-world systems are seldom linear, and it has been shown empirically that some financial and economic series typically exhibit changes in the dynamics of the conditional distribution.

One popular approach to modelling changes in regimes is the class of Markov switching models. Since the seminal papers by Goldfeld and Quandt (1973) and Hamilton (1989, 1990), VAR models subject to Markov switching have been used actively in econometrics to model various nonlinear time series. The Markov process is not observable and is often referred to as the regime. Flexibility is one of the main advantages of Markov switching models. The changes in regime can be smooth or abrupt, and they can occur frequently or occasionally depending on the transition probabilities of the Markov chain.

For information concerning the stationarity, maximum likelihood (ML) estimation, consistency, statistical inference, model selection and specification testing of Markov switching (MS) VAR models see Hamilton (1989, 1990, 1994, 1996), Kim (1994), Krolzig (1997), Yang (2000), Francq and Zakoïan (2001), Stelzer (2009) and Cavicchioli (2014a). Explicit matrix expressions for the ML estimator of the parameters in MS VAR models have been derived in Cavicchioli

(2014b). Higher-order moments and asymptotic Fisher information matrix of MS VARMA models are provided in Cavicchioli (2017a) and (2017b), respectively. Tractable methods to derive the spectral representations of a general class of MS VARMA models have been proposed by Pataracchia (2011), Cavicchioli (2013), and Cheng (2016). An advantage of spectral analysis is that it can reveal detailed features of financial data without using a parametric model. The only assumption needed is the stationarity of the series. A fresh insight can be gained into the time series structure and cyclic behaviour at different time scale, such as seasonal patterns and business cycle. This serves to derive some goodness-of-fit tests on MS models based on spectral density functions. A goodness-of-fit testing scheme for the marginal distribution of MS models has been derived by Janczura and Weron (2013).

A substantial literature exists on Bayesian inference for MS VARs. A Bayesian approach for estimating Markov mixture models has been proposed by Albert and Chib (1993), Chib (1996), Billio et al. (1999), Kim and Nelson (1999), Wang and Zivot (2000), Sims and Zha (2006), Sims et al. (2008), and Hahn et al. (2009). Maximum likelihood estimation of MS regressions models with endogenous switching is presented in Kim et al. (2008). An efficient estimation of MS models with an application to electricity spot prices is described in Janczura and Weron (2012).

In this paper, we present an alternative estimation approach for MS VAR models, including a closed-form expression for the asymptotic covariance matrix of the parameter estimator. More precisely, we derive a neat and compact estimation procedure for MS VAR models using OLS methodology and show that the OLS estimates coincide with the maximum likelihood (ML) ones. The new proposed estimation procedure generalizes the linear case as in Hamilton (1994). Our recursion equations, written as concisely as possible at the vector-matrix level, improve computational performance with respect to the expectation-maximization (EM) algorithm since they are readily programmable and in addition greatly reduce the computational cost. Moreover, as noted in Newton (1978) and Wong and Li (2000), one of the most difficult computational problem is to derive the limiting covariance matrix in closed-form of the ML estimators in the framework of the EM algorithm. The first author proposed using the integral formula of Whittle (1953) to evaluate the Fisher information matrix, whose inverse gives the asymptotic covariance matrix when the process is Gaussian. Other authors, like Douc et al. (2004) and Bao and Hua (2014), have studied asymptotic properties of the ML estimators for autoregressive models with Markov regime, but their expression of the limiting covariance has a rather complicated form. The results given by the present paper also solve this problem, providing an explicit matrix expression in closed-form for the asymptotic covariance matrix of the OLS (and hence ML) estimators for MS VAR models under a normality assumption. This gives a characterization of the asymptotic variance, which is new to our knowledge. Furthermore, it is easier to program than the usual approach based on second derivatives, and more accurate. These results may be helpful to solve the dimensionality problem of the maximum likelihood estimates (MLE) for large MS VAR models. The theoretical results are then complemented by some Monte Carlo simulations and an illustrative application to energy use. Finally, our estimation method can be easily applied when studying structural representations of

MS VAR models by using a reduced form in the sense of Karamé (2015) and Lanne et al. (2010).

The plan of the rest of the paper is as follows. In Sect. 2 we introduce the model, describe the main assumptions, and provide a vectorial representation of it, which is used to derive the main results of the paper. The model is estimated by use of the OLS method. The estimation procedure in matrix form for MS VAR models is described in Sect. 3, where explicit matrix formulas for the OLS (and hence ML) estimators of the parameters are derived. Indeed, we show that an OLS algorithm to estimate a MS VAR model coincides with MLE estimates. The deviation of the OLS estimate from the true value is displayed in Sect. 4. Then, we prove consistency of these estimators and their asymptotic properties. Finally, we provide explicit matrix expression in closed-form for the asymptotic covariance matrix of the OLS (and hence ML) estimates of the parameters for MS VAR models under a normality assumption. Our theorems generalize the classical results known in the literature for linear VAR processes (which become particular subcases) to the setting of time series with changes in regime. Section 5 presents a short numerical illustration by some Monte Carlo simulations. Section 6 describes an empirical example of the model that uses energy consumption and economic growth in the Euro area for the period 1968–2015. Section 7 gives some concluding remarks. A final Appendix collects all technical proofs of the results.

2 Markov switching VAR models

Let us consider a M–state Markov switching K–dimensional AR(p) model (in short, MS(M) VAR(p)) of the following type:

$$\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \sum_{i=1}^{p} \boldsymbol{\Phi}_{i,s_t} \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_{t,s_t} \tag{1}$$

where $\mathbf{y}_t$ is a K–dimensional random vector with values in $\mathbb{R}^K$, (s_t) is an irreducible, aperiodic and ergodic Markov chain with values in $\Xi = \{1, 2, \ldots, M\}$, stationary transition probabilities $p_{ij} = Pr(s_t = j | s_{t-1} = i)$ for $i, j = 1, \ldots, M$, and unconditional (or steady state) probabilities $\pi_i = Pr(s_t = i)$ for $i \in \Xi$. Let $\mathbf{P} = (p_{ij})$ denote the transition probability matrix of the chain. Of course, we have $\sum_{i=1}^{M} \pi_i = 1$ and $\sum_{j=1}^{M} p_{ij} = 1$ for $i \in \Xi$. The Markov chain is *ergodic* if exactly one of the eigenvalues of $\mathbf{P}$ is unity and all other eigenvalues are inside the unit circle. Under this condition, there exists a stationary or unconditional probability distribution of the regimes, that is, $\boldsymbol{\pi} = (\pi_1 \ \cdots \ \pi_M)' \in \mathbb{R}^M$. The Markov chain is *irreducible* if the unconditional probability $\pi_i > 0$ for $i \in \Xi$. In Model (1) $\boldsymbol{\Phi}_{i,s_t} \in \mathbb{R}^{K \times K}$ and $\boldsymbol{\nu}_{s_t} \in \mathbb{R}^K$, where $\mathbb{R}^{m \times n}$ denotes the class of real $m \times n$ matrices and $\mathbb{R}^n$ the class of n-dimensional real vectors, that is, $\mathbb{R}^n = \mathbb{R}^{n \times 1}$.

To allow for the possibility of change in variance, it is assumed that $\boldsymbol{\epsilon}_{t,s_t} = \boldsymbol{\Sigma}_{s_t} \mathbf{u}_t$, where $\boldsymbol{\Sigma}_{s_t}$ is a $K \times K$ real nonsingular random matrix and $(\mathbf{u}_t)$ is a zero-mean white noise process with $E(\mathbf{u}_t \mathbf{u}_t') = \mathbf{I}_K$ and $E(\mathbf{u}_t \mathbf{u}_\tau') = \mathbf{0}$, $t \neq \tau$ (here $\mathbf{I}_K$ is the $K \times K$ identity matrix). Set $\boldsymbol{\Omega}_{s_t} = \boldsymbol{\Sigma}_{s_t} \boldsymbol{\Sigma}_{s_t}'$. In addition, we assume that $(\mathbf{u}_t)$ is independent of (s_t), and that $(\mathbf{y}_t)$ is second–order stationary i.e., it satisfies Theorem 2 from Francq and Zakoïan (2001).

Following Krolzig (1997, §2), model (1) can be written as

$$\mathbf{y}_t = (\mathbf{x}_t' \otimes \mathbf{I}_K)\,\boldsymbol{\beta}_{s_t} + \boldsymbol{\epsilon}_{t,s_t} \tag{2}$$

where

$$\mathbf{x}_t = (1\ \mathbf{y}_{t-1}' \ \dots\ \mathbf{y}_{t-p}')' \in \mathbb{R}^R \qquad \boldsymbol{\beta}_{s_t} = (\boldsymbol{\nu}_{s_t}' \ \ [\mathrm{vec}\boldsymbol{\Phi}_{1,s_t}]' \ \dots\ [\mathrm{vec}\boldsymbol{\Phi}_{p,s_t}]')' \in \mathbb{R}^{RK}$$

with $R = pK + 1$. We now explain the tensor product notation and the vec operator used from (2) onwards, mainly in the proofs (see Appendix). For $\mathbf{A} = (a_{ij})$ an $m \times n$ matrix and $\mathbf{B}$ a $p \times q$ matrix, the *Kronecker product* (also called the *tensor product*) of $\mathbf{A}$ and $\mathbf{B}$ is defined as the following $(mp) \times (nq)$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}.$$

Given $\mathbf{A}$ as above, vec$\mathbf{A}$ is an $(mn) \times 1$ column vector, obtained by stacking the columns of $\mathbf{A}$, one below the other, with the columns ordered from left to right. For the main properties of the tensor product and the vec operator see, for example, Hamilton (1994, p.732), Lütkepohl (2007, p.660), and Yang and He (2008).

Let $\boldsymbol{\beta}_m$ and $\boldsymbol{\Omega}_m$ be obtained from $\boldsymbol{\beta}_{s_t}$ and $\boldsymbol{\Omega}_{s_t}$, respectively, by setting $s_t = m$, for every $m \in \Xi$. Then, the population parameters of model (2) are $\boldsymbol{\theta}_m = (\boldsymbol{\beta}_m' \ \ [\mathrm{vec}\boldsymbol{\Omega}_m]' \ \ \pi_m)'$, for every $m \in \Xi$, together with the vector of transition probabilities $\boldsymbol{\rho} = \mathrm{vec}\mathbf{P}$. Collect them in a vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1' \ \cdots\ \boldsymbol{\theta}_M')'$, and set $\boldsymbol{\lambda} = (\boldsymbol{\theta}' \ \ \boldsymbol{\rho}')'$.

To estimate the model parameters by OLS, the sum of squares based on a sample of size T can be written (for more details to derive (3) see Appendix) as

$$SSE_T = \sum_{m=1}^{M}\sum_{t=1}^{T} \boldsymbol{\epsilon}_{t,m}'\,\boldsymbol{\epsilon}_{t,m} = \sum_{m=1}^{M}\sum_{t=1}^{T} [\mathbf{y}_t - (\mathbf{x}_t' \otimes \mathbf{I}_K)\,\boldsymbol{\beta}_m]'\,[\mathbf{y}_t - (\mathbf{x}_t' \otimes \mathbf{I}_K)\,\boldsymbol{\beta}_m]\,\xi_{mt|T} \tag{3}$$

where $\xi_{mt|T} = E(\xi_{mt}|\mathbf{Y}_T)$ denotes the *smoothed regime probabilities* and $\mathbf{Y}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$ is the information set. Furthermore, $\boldsymbol{\xi}_t = (\xi_{1t} \ \cdots\ \xi_{Mt})'$ is the random $M \times 1$ vector whose mth element is equal to unity if $s_t = m$ and zero otherwise. Note that $\mathbf{i}_M' \boldsymbol{\xi}_t = 1$, $E(\boldsymbol{\xi}_t) = \boldsymbol{\pi}$, $E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t') = \mathbf{D}$, and $E(\boldsymbol{\xi}_t \boldsymbol{\xi}_\tau') = \mathbf{0}$ for $t \neq \tau$, where $\mathbf{D} = \mathrm{diag}(\pi_1 \ \cdots\ \pi_M)$ and $\mathbf{i}_M$ is the $M \times 1$ vector of ones.

Typically, the expectation step of the ML algorithm is the conditional expectation of the "complete likelihood" (as if $\mathbf{y}_t$ and s_t were observed) conditioned on $\mathbf{Y}_T$. This gives the SSE_T from (3) but also an additional term, which is the logarithm of the conditional distribution of s_t conditioned on $\mathbf{Y}_T$. This term is extremely important as it allows one to obtain a consistent estimate of the parameters $\boldsymbol{\pi}$ and p_{ij}, for all $i, j \in \Xi$. It is often evaluated with the Baum-Welch algorithm (see, for example, Baggenstoss 2001). Then, the conditional on $\mathbf{Y}_T$ log-likelihood function can be written

$$\mathcal{L}(\boldsymbol{\theta}) = -SSE_T + \sum_{t=1}^{T} \sum_{m=1}^{M} [\log_e \pi_m]\, \xi_{mt|T}. \tag{4}$$

Thus, we are going to do a constrained maximization of $\mathcal{L}(\boldsymbol{\theta})$ subject to the restriction $\pi_1 + \pi_2 + \cdots + \pi_M = 1$, that is, we form the Lagrangean

$$J(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \lambda(1 - \pi_1 - \cdots - \pi_M)$$

and set the derivatives with respect to $\boldsymbol{\theta}_m$ equal to zero, for all $m \in \Xi$.

A fast recursive algorithm to evaluate $\xi_{mt|T}$ was described in Krolzig (1997, §5). See also Hamilton (1994), Equations [22.4.5], [22.4.6] and [22.4.14]. This algorithm requires specification of a particular distribution for the innovations. Therefore, suppose that $\mathbf{u}_t$ is normally distributed. Note, however, that the results given below can be extended to many other distributions.

More precisely, let $\boldsymbol{\eta}_t = \boldsymbol{\eta}_t(\boldsymbol{\theta})$ denote the $M \times 1$ vector of the densities of $\mathbf{y}_t$ conditional on s_t and $\mathbf{Y}_{t-1}$, i.e, $\boldsymbol{\eta}_t = (\eta_{1t}(\boldsymbol{\theta}) \cdots \eta_{Mt}(\boldsymbol{\theta}))$. The filter inference $\boldsymbol{\xi}_{t|t} = E(\boldsymbol{\xi}_t|\mathbf{Y}_t)$ can be computed by iterating on the following pair of recursive formulas

$$\xi_{t|t} = \frac{\boldsymbol{\eta}_t \odot \xi_{t|t-1}}{\boldsymbol{\eta}_t' \xi_{t|t-1}} \qquad\qquad \xi_{t+1|t} = \mathbf{P}' \xi_{t|t} \tag{5}$$

where the symbol $\odot$ denotes the element-by-element multiplication. The iteration is started by assuming that the initial state vector is drawn from the stationary unconditional probability distribution of the Markov chain i.e., $\boldsymbol{\xi}_{1|0} = \boldsymbol{\pi}$. Finally, the full-sample smoothed regime probabilities $\boldsymbol{\xi}_{t|T}$ can be found by iterating backward from $t = T-1, \ldots, 1$ by starting from the last output $\boldsymbol{\xi}_{T|T}$ of the filter using the formula

$$\xi_{t|T} = \left[\mathbf{P}(\xi_{t+1|T} (\div) \xi_{t+1|t})\right] \odot \xi_{t|t} \tag{6}$$

where the symbol $(\div)$ denotes the element-by-element division. These filter and smoother recursions also provide the analytical tool to construct and evaluate the likelihood function of the model. See Hamilton (1989, (1990), Krolzig (1997), and Cavicchioli (2014b).

A consistent estimate of π_m, for all $m \in \Xi$, is given by

$$\hat{\pi}_m = \frac{1}{T} \sum_{t=1}^{T} \xi_{mt|T} \tag{7}$$

as proved in Appendix.

If the initial probability of $\boldsymbol{\xi}_{1|0}$ is taken to be a fixed value of $\boldsymbol{\pi}$, unrelated to the other parameters, then it is shown in Hamilton (1990) that a consistent estimate for the transition probabilities satisfies

$$\widehat{p}_{ij} = \frac{\sum_{t=1}^{T} Pr(s_t = j,\, s_{t-1} = i|\mathbf{Y}_T;\widehat{\lambda})}{\sum_{t=1}^{T} Pr(s_{t-1} = i|\mathbf{Y}_T;\widehat{\lambda})} \tag{8}$$

for $i, j = 1, \dots, M$.

3 OLS estimation

The first-order conditions (FOC) from (3) give the following result:

Theorem 1 *Suppose that the matrix* $\mathbf{X}_{mT} = \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \xi_{mt|T}$ *is invertible. Then, the* OLS *estimates of the population parameters* $\boldsymbol{\beta}_m$ *and* $\boldsymbol{\Omega}_m$, $m = 1, \dots, M$, *for the* MS(M) VAR(p) *model in* (2) *are obtained from a solution of the following M-step equations:*

$$\hat{\boldsymbol{\beta}}_m = \left[\mathbf{X}_{mT}^{-1} \otimes \mathbf{I}_K\right] \left[\sum_{t=1}^{T} (\mathbf{x}_t \otimes \mathbf{I}_K)\, \mathbf{y}_t\, \xi_{mt|T}\right] \tag{9}$$

$$\hat{\boldsymbol{\Omega}}_m = \left[\sum_{t=1}^{T} \xi_{mt|T}\right]^{-1} \left[\sum_{t=1}^{T} \hat{\boldsymbol{\epsilon}}_{t,m}\, \hat{\boldsymbol{\epsilon}}_{t,m}'\, \xi_{mt|T}\right] \tag{10}$$

where

$$\hat{\boldsymbol{\epsilon}}_{t,m} = \mathbf{y}_t - (\mathbf{x}_t' \otimes \mathbf{I}_K)\, \hat{\boldsymbol{\beta}}_m.$$

Corollary 1 *The OLS estimates of the population parameters of the* MS(M)VAR(p) *model in* (1) *coincide with the ML estimates (with Gaussian disturbances).*

A source for the MLE formulas is Cavicchioli (2014b). Compare with Hamilton (2016). Furthermore, Formulas (9) and (10) generalize Formulas [11.1.11] and [11.1.27], respectively, given in Hamilton (1994, §11) for a linear VAR(p) model. The assumption in Theorem 1 (X_{mT} invertible, for every $m = 1, \dots, M$) is standard, and extends the analogous one (case $M = 1$) for linear VAR(p) models. See Formula [11.1.11] from Hamilton (1994). We also require in Assumption (12) below that $T^{-1}X_{mT}$ converges in probability to a finite nonsingular matrix for every $m \in \Xi$. This guarantees the existence of the asymptotic covariance matrix of the OLS estimators under a normality assumption (see Sect. 4).

The matrix expressions in Theorem 1 depend on the observable variables and the smoothed regime probabilities. They solve the M-step of the EM algorithm while the computation of the smoothed probabilities is part of the usual E-step. Hamilton (2016, §2.5) describes how the EM algorithm can be implemented for finding ML (hence OLS) estimates of MS VARs. See Filardo (1994) for an extension of the Hamilton estimation procedure to incorporate time-varying probabilities. Here, we adapt the method by using Theorem 1.

The Algorithm. By using Equations (7) and (8) and Theorem 1 we suggest a simple iterative algorithm for getting OLS (or ML) estimates of model parameters. The implementation of this algorithm does not require the numerical maximization of the log-likelihood function; hence, it improves the EM algorithm. In fact, the maximization step is solved analytically by using our simple matrix formulas. Start the procedure with an arbitrary guess for the model parameters, say $\boldsymbol{\beta}_m := \boldsymbol{\beta}_m^{(0)}$, $\boldsymbol{\Omega}_m := \boldsymbol{\Omega}_m^{(0)}$, $\pi_m := \pi_m^{(0)}$, $m = 1, \ldots, M$, and $\boldsymbol{\rho} := \boldsymbol{\rho}^{(0)}$. For this guess, we calculate the smoothed probabilities $\xi_{mt|T}^{(0)}$ from (5) and (6). Then, one can calculate the magnitudes on the right sides of formulas (7), (8), (9) and (10) with $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\Omega}^{(0)}$, $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\rho}^0$ and $\xi_{mt|T}^{(0)}$ in place of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\Omega}}$, $\boldsymbol{\pi}$, $\boldsymbol{\rho}$ and $\xi_{mt|T}$. The left sides of such formulas then produce new estimates $\boldsymbol{\beta}_m^{(1)}$, $\boldsymbol{\Omega}_m^{(1)}$ and $\pi_m^{(1)}$, for $m = 1, \ldots, M$, and $\boldsymbol{\rho}^{(1)}$. These estimates can be used to compute the smoothed probabilities $\xi_{mt|T}^{(1)}$ and recalculate the expressions on the right sides of formulas (7), (8), (9) and (10). One continues iterating in this fashion until the change between the current estimates at step $\ell + 1$ and those at step ℓ is smaller than some specified convergence criterion. This fixed-point algorithm has typically linear convergence. Since the innovations are Gaussian, the EM algorithm converges to a maximum of the true likelihood function (or attains a local maximum), which is asymptotically consistent.

4 Asymptotic properties

In this section, the probability statements implicitly assume that the number of regimes in the data generating process remains fixed.

The deviation of the OLS estimate $\hat{\boldsymbol{\beta}}_m$ from its true value is

$$\hat{\boldsymbol{\beta}}_m = \boldsymbol{\beta}_m + \left[\mathbf{X}_{mT}^{-1} \otimes \mathbf{I}_K\right] \left[\sum_{t=1}^{T} (\mathbf{x}_t \otimes \mathbf{I}_K)\, \boldsymbol{\Sigma}_m\, \mathbf{u}_t\, \xi_{mt|T}\right]. \tag{11}$$

Assume that

$$\mathbf{Q}_m = \operatorname{plim}_{T\to\infty} T^{-1}\, \mathbf{X}_{mT} = E(\mathbf{x}_t\, \mathbf{x}_t^{'}\, \xi_{mt|T}) \tag{12}$$

is finite and nonsingular for every $m \in \Xi$, where plim means convergence in probability. From (10) and (11) we get

Theorem 2 *Under Assumption* (12) *and* $\mathbf{u}_t \sim \text{i.i.d}\mathcal{N}(\mathbf{0}, \mathbf{I}_K)$, *the estimators* $\hat{\boldsymbol{\beta}}_m$ *and* $\hat{\boldsymbol{\Omega}}_m$ *of* $\boldsymbol{\beta}_m$ *and* $\boldsymbol{\Omega}_m$ *are consistent for every* $m \in \Xi$.

Notice that the process $\{(\mathbf{x}_t \otimes \mathbf{I}_K)\, \boldsymbol{\Sigma}_m\, \mathbf{u}_t\, \xi_{mt|T}\}$ in (11) is a martingale difference sequence with zero mean and asymptotic covariance matrix $\mathbf{Q}_m \otimes \boldsymbol{\Omega}_m$. For this, we use the fact that $\mathbf{u}_t$ is uncorrelated with $\xi_{mt|T}$ as $\mathbf{u}_t$ is independent of s_τ for any t and τ, by standard assumption.

From Hamilton (1994, p.474) it follows that

$$\frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right]\xrightarrow{L}\mathcal{N}(\mathbf{0},\mathbf{Q}_m\otimes\boldsymbol{\Omega}_m) \tag{13}$$

for every $m\in\Xi$. Then, we have

Theorem 3 *Under assumption* (12) *and* $\mathbf{u}_t\sim$ i.i.d$\mathcal{N}(\mathbf{0},\mathbf{I}_K)$, *the asymptotic distribution of the* OLS *estimate* $\hat{\boldsymbol{\beta}}_m$ *of the parameter vector* $\boldsymbol{\beta}_m$ *in model* (2) *is given by*

$$\sqrt{T}\,(\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)\xrightarrow{L}\mathcal{N}(\mathbf{0},\mathbf{Q}_m^{-1}\otimes\boldsymbol{\Omega}_m)$$

for every $m\in\Xi$.

Thus, each $\hat{\boldsymbol{\beta}}_m$ individually is asymptotically Gaussian and $O_p(\frac{1}{\sqrt{T}})$. Furthermore, the usual t and F tests about $\boldsymbol{\beta}_m$ calculated in the usual way are all asymptotically valid. For example, the Wald form of the OLS χ^2 test about the null hypothesis $H_0\,:\,\mathbf{R}\,\boldsymbol{\beta}_m=\mathbf{r}$ is

$$\chi^2_{mT}=(\mathbf{R}\,\hat{\boldsymbol{\beta}}_m-\mathbf{r})'\left[\mathbf{R}(\mathbf{X}_{mT}^{-1}\otimes\hat{\boldsymbol{\Omega}}_m)\,\mathbf{R}'\right]^{-1}(\mathbf{R}\,\hat{\boldsymbol{\beta}}_m-\mathbf{r}).$$

Under the null hypothesis this expression can be rewritten as

$$\begin{aligned}\chi^2_{mT}&=\left[\mathbf{R}(\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)\right]'\left[\mathbf{R}(\mathbf{X}_{mT}^{-1}\otimes\hat{\boldsymbol{\Omega}}_m)\,\mathbf{R}'\right]^{-1}\left[\mathbf{R}(\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)\right]\\&=\left[\mathbf{R}\,\sqrt{T}\,(\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)\right]'\left[\mathbf{R}(T\,\mathbf{X}_{mT}^{-1}\otimes\hat{\boldsymbol{\Omega}}_m)\,\mathbf{R}'\right]^{-1}\left[\mathbf{R}\,\sqrt{T}\,(\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)\right].\end{aligned}$$

But

$$\mathbf{R}\,\sqrt{T}\,(\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)\xrightarrow{L}\mathcal{N}(\mathbf{0},\mathbf{R}\,(\mathbf{Q}_m^{-1}\otimes\boldsymbol{\Omega}_m)\,\mathbf{R}')$$

and

$$\mathbf{R}(T\,\mathbf{X}_{mT}^{-1}\otimes\hat{\boldsymbol{\Omega}}_m)\,\mathbf{R}'\xrightarrow{p}\mathbf{R}\,(\mathbf{Q}_m^{-1}\otimes\boldsymbol{\Omega}_m)\,\mathbf{R}'.$$

Thus χ^2_{mT} is a quadratic form in asymptotically Gaussian variable. Therefore, it is asymptotically $\chi^2(k)$.

If $(\mathbf{u}_t)$ are Gaussian, Cavicchioli (2014b) proved that the estimates of $\boldsymbol{\beta}_m$ and $\boldsymbol{\Omega}_m$ are uncorrelated and $\mathrm{var}[\mathrm{vec}(\hat{\boldsymbol{\Omega}}_m)]=2\,\pi_m^{-1}\,\boldsymbol{\Omega}_m\otimes\boldsymbol{\Omega}_m$ for every $m\in\Xi$. Then

$$\widetilde{\boldsymbol{\Sigma}}_m=\mathrm{var}[\mathrm{vech}(\hat{\boldsymbol{\Omega}}_m)]=2\,\pi_m^{-1}\,\mathbf{D}_K^{+}\,(\boldsymbol{\Omega}_m\otimes\boldsymbol{\Omega}_m)\,(\mathbf{D}_K^{+})'$$

where $\mathbf{D}_K^{+}$ is the Moore-Penrose inverse of the $K^2\times n$ duplication matrix $\mathbf{D}_K$ with $n=K(K+1)/2$. This extends Formula [11.1.48] of Hamilton (1994) for linear VAR models. Recall that $\mathbf{D}_K\,\mathrm{vech}\boldsymbol{\Omega}_m=\mathrm{vec}\boldsymbol{\Omega}_m$ and $\mathbf{D}_K^{+}\,\mathrm{vec}\boldsymbol{\Omega}_m=\mathrm{vech}\boldsymbol{\Omega}_m$, where $\mathbf{D}_K^{+}=(\mathbf{D}_K'\,\mathbf{D}_K)^{-1}\,\mathbf{D}_K'$.

The next result provides explicit matrix expression in closed-form for the asymptotic covariance matrix of a MS VAR model. This characterization of the asymptotic covariance is new to our knowledge. It also generalizes Proposition 11.2 from Hamilton (1994) given for linear case.

Theorem 4 *Under assumption* (12) *and* $\mathbf{u}_t \sim$ *i.i.d.* $\mathcal{N}(\mathbf{0}, \mathbf{I}_K)$, *the asymptotic distribution of the* OLS (*or* ML) *estimates of the parameters in stationary* MS VAR *models as in* (2) *is given by*

$$\begin{pmatrix} \sqrt{T}\,(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) \\ \sqrt{T}\left[\mathrm{vech}(\hat{\boldsymbol{\Omega}}_m) - \mathrm{vech}(\boldsymbol{\Omega}_m)\right] \end{pmatrix} \xrightarrow{L} \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Q}_m^{-1} \otimes \boldsymbol{\Omega}_m & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Sigma}}_m \end{pmatrix}\right]$$

for every $m \in \Xi$.

Calculating the asymptotic covariance matrix from Theorem 4 is better than the usual approach based on second derivatives. In fact, the proposed expression is easily tractable and directly programmable in addition of greatly reducing the computational cost. Furthermore, it is more accurate, and allows for more precise inference.

A matrix expression in closed-form to evaluate the asymptotic Fisher information matrix of MS VARMA models is derived in Cavicchioli (2017b); its inverse gives the asymptotic covariance matrix when the process is Gaussian.

5 Numerical illustrations

Let us consider Model (1) with $M = K = 2$ and $p = 1$, that is, a bivariate 2-regime MS VAR(1). The true parameters are

$$\boldsymbol{\nu}_1 = [0.15 \quad 0.3]' \qquad \boldsymbol{\nu}_2 = [0.7 \quad 0.9]'$$

$$\boldsymbol{\Phi}_{1,1} = \begin{pmatrix} 0.2 & 0.4 \\ 0.3 & 0.2 \end{pmatrix} \qquad \boldsymbol{\Phi}_{1,2} = \begin{pmatrix} 0.25 & 0.15 \\ 0.3 & 0.1 \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}.$$

The transition probabilities are fixed to be $p_{11} = 0.6$ and $p_{22} = 0.2$, hence the unconditional probabilities are $\pi_1 = 0.67$ and $\pi_2 = 0.33$. The parameters are chosen to proxy a first regime of low means but more stable. On the contrary, the second regime is more unstable both in terms of probability and volatility, but with higher mean returns. The estimation procedure is performed for processes of size $T = 500$ by using the OLS estimation algorithm proposed in Sect. 3. We consider simulations for numbers of replications equal to 50, 500 and 1500 (first 50 observations are discarded). The innovations are normally distributed. A typical realization of the bivariate process is plotted in Fig. 1. Monte Carlo simulation results in terms of replication means are reported in Table 1 which contains the estimated parameters and the corresponding standard errors in parenthesis. The standard errors are computed by using the asymptotic matrix expressions given in Theorem 4. Mean estimates are rather close to the true parameter values and they tend to be closer as the number of replications increases. However, the standard errors remain very similar so that precision in mean increases but variability does not change as much. Moreover, we point out that our algorithm is very fast since it takes only 0.09 seconds for each iteration. Overall, the estimation procedure works very well giving satisfactory conclusions.

6 Energy use and economic growth

Some authors have documented the nonlinear nature of energy use e.g., Seifritz and Hodgkin (1991) and Lee and Chang (2007). As a consequence, the relationship between energy use and economic growth could also be subject to regime shift. To explore this relationship, we use annual time series for the Euro area in the period 1968–2015. The series are energy use (kt of oil equivalent) and GDP (constant 2010 US$) as obtained from *World Development Indicators*. The series are transformed to natural logarithms and plotted in Fig. 2. To deal with stationary series, we take first difference of the log-series and select the order of MS VAR using BIC, detecting $p = 1$ as autoregressive coefficient. The model to be estimated is then a bivariate 2-regime MS VAR(1). The estimates of the parameters obtained by using OLS procedure are reported in Table 2. The estimated probabilities are $p_{11} = 0.8127$

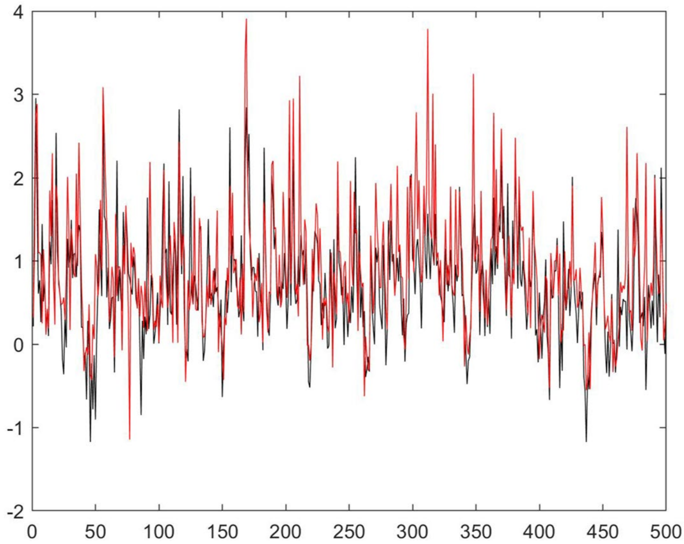

Fig. 1 Plot of a typical realization of the bivariate 2-regime MS VAR(1) as described in Sect. 5 (black and red lines identify the bivariate process) (colour figure online)

Table 1 OLS estimates (MC replication means) of simulation experiments described in Sect. 5. Corresponding standard errors are in parenthesis

	true	repl=50	repl=500	repl=1500
$\boldsymbol{\nu}_1'$	[0.15 0.3]	[0.1391 0.2613] (0.0541) (0.0642)	[0.1445 0.2688] (0.0524) (0.0422)	[0.1490 0.2894] (0.0568) (0.0722)
$\boldsymbol{\nu}_2'$	[0.7 0.9]	[0.6526 0.9360] (0.1110) (0.1210)	[0.6543 0.9333] (0.1090) (0.1573)	[0.6858 0.9037] (0.1103) (0.1471)
$\boldsymbol{\Phi}_{1,1}$	$\begin{bmatrix} 0.2 & 0.4 \\ 0.3 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0.1690 & 0.2607 \\ 0.2719 & 0.1169 \end{bmatrix}$ (0.1014) (0.2052) (0.1016) (0.1025)	$\begin{bmatrix} 0.1774 & 0.2667 \\ 0.2766 & 0.1531 \end{bmatrix}$ (0.1037) (0.1450) (0.1008) (0.1020)	$\begin{bmatrix} 0.1970 & 0.3678 \\ 0.2936 & 0.1808 \end{bmatrix}$ (0.1074) (0.1566) (0.1069) (0.1010)
$\boldsymbol{\Phi}_{1,2}$	$\begin{bmatrix} 0.25 & 0.15 \\ 0.3 & 0.1 \end{bmatrix}$	$\begin{bmatrix} 0.2141 & 0.1370 \\ 0.2883 & 0.0707 \end{bmatrix}$ (0.1004) (0.1028) (0.1738) (0.0887)	$\begin{bmatrix} 0.2189 & 0.1465 \\ 0.2975 & 0.0786 \end{bmatrix}$ (0.0973) (0.0764) (0.0861) (0.0622)	$\begin{bmatrix} 0.2431 & 0.1506 \\ 0.2998 & 0.0860 \end{bmatrix}$ (0.1075) (0.0932) (0.1824) (0.0985)
$\boldsymbol{\Sigma}_1$	$\begin{bmatrix} 0.2 & 0.1 \\ - & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0.1455 & 0.0707 \\ - & 0.1476 \end{bmatrix}$ (0.0232) (0.0173) (0.0233)	$\begin{bmatrix} 0.1666 & 0.0801 \\ - & 0.1482 \end{bmatrix}$ (0.0335) (0.0218) (0.0301)	$\begin{bmatrix} 0.1867 & 0.0904 \\ - & 0.1767 \end{bmatrix}$ (0.0882) (0.0366) (0.0893)
$\boldsymbol{\Sigma}_2$	$\begin{bmatrix} 0.5 & 0.3 \\ - & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.4450 & 0.2674 \\ - & 0.4935 \end{bmatrix}$ (0.0876) (0.0639) (0.0855)	$\begin{bmatrix} 0.4544 & 0.2694 \\ - & 0.4941 \end{bmatrix}$ (0.0913) (0.0535) (0.0944)	$\begin{bmatrix} 0.4962 & 0.2789 \\ - & 0.4999 \end{bmatrix}$ (0.1045) (0.0871) (0.1093)

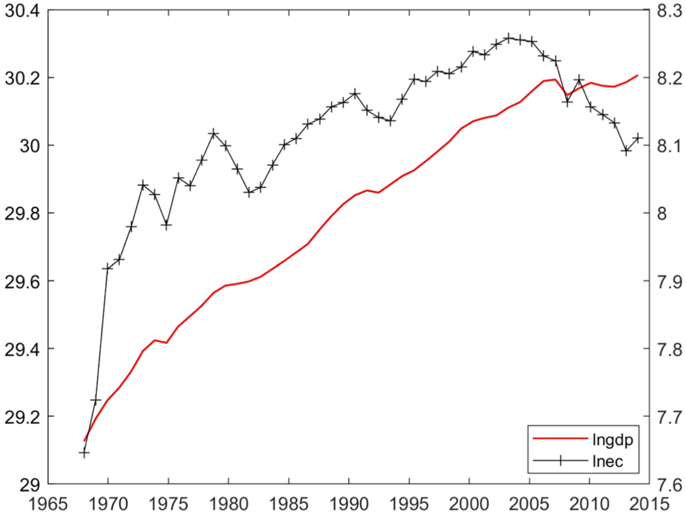

Fig. 2 Natural logarithm of energy use and GDP in the Euro area for the period 1968–2015, obtained from *World Development Indicators*

Table 2 OLS estimates of energy use and GDP (first difference of the natural logarithms) in the Euro area modelled as described in Sect. 6. Corresponding standard errors are in parenthesis

	Δlnec	Δlngdp
Regime 1		
Intercept	0.5106 (0.6806)	0.8153 (0.5068)
Δlnec(−1)	− 0.0965 (0.2325)	− 0.5479 (0.1732)
Δlngdp(−1)	− 0.2321 (0.3671)	0.2519 (0.2734)
$\mathbf{\Sigma}_1 = \begin{pmatrix} 7.0017(1.6936) & 4.6905(1.1478) \\ - & 3.2898(0.7958) \end{pmatrix}$		
Regime 2		
Intercept	0.8838 (0.8935)	1.7365 (0.6125)
Δlnec(−1)	0.1497 (0.1436)	0.9387 (0.0984)
Δlngdp(−1)	0.0205 (0.3212)	0.4183 (0.2201)
$\mathbf{\Sigma}_2 = \begin{pmatrix} 2.4787(0.9791) & 0.9402(0.5786) \\ - & 1.3746(0.5430) \end{pmatrix}$		

and $p_{22} = 0.6873$. Thus, both regimes are persistent. Figure 3 reports the smoothed probabilities of being the first regime. The first regime identifies GDP slow-down with lower growth and higher volatililty due to instability. This regime includes the energy crisis in the mid-70s, the recessions in the early 80s and early 90s and the great recession starting from 2001 toward the end of the sample. The model is then capable of detecting changes in the relationship of these variables and to identify separate phases in which the behaviour is different.

7 Concluding remarks

We consider estimation in Markov switching vector autoregressive (MS VAR) models, where the underlying Markov chain starts with its stationary distribution. The model parameters are estimated by ordinary least squares (OLS). To this end, a vectorial

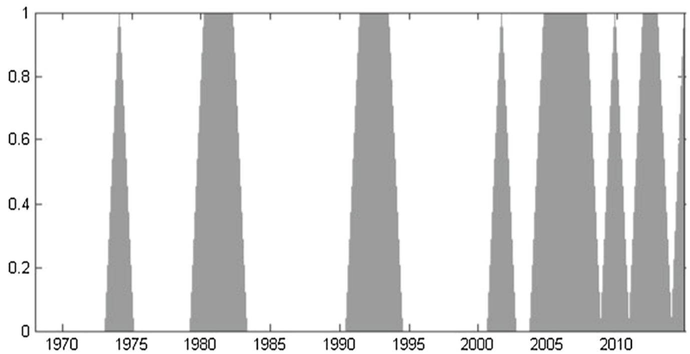

Fig. 3 Plot of the smoothed probabilities from the bivariate MS VAR(1) of energy use and GDP growth in the first regime

representation of the model is introduced using tensor product notation. This allows to derive explicit OLS estimators in matrix form, showing that such estimators coincide with the maximum likelihood estimators (with Gaussian disturbances). An iterative algorithm based on these is proposed, and the differences to the classical expectation maximization (EM) algorithm are discussed in detail, the main benefit being that the maximization step is solved analytically while the maximization of the conditional log-likelihood in this model requires a numerical maximization in the EM algorithm. Furthermore, we derive an explicit expression in closed-form for the asymptotic covariance matrix of the OLS estimators of MS VAR models under a normality assumption. The neat matrix formula solves one of the most difficult computational problem in obtaining the limiting covariance matrix in closed-form, and allows for more precise inference. The theoretical results are complemented by Monte Carlo simulation experiments whose conclusions are precise and easy to obtain since they use generalization of the classical results known for linear VAR processes. Our approach is useful for empirical applications whose time series are subject to nonlinear relationship. The analysis of the nonlinear interaction of energy use and economic growth shows the capability of the model to capture recessionary subperiods whose dynamics are different.

Acknowledgements Work financially supported by FAR research Grant (2019) of the University of Modena and Reggio Emilia, Italy. The author would like to thank the Editor-in-Chief of the journal, Professor Yarema Okhrin, and two anonymous reviewers for their valuable and constructive comments which improved the final version of the paper.

Appendix

Derivation of (3).

$$\begin{aligned}
\mathrm{SSE}_T &= \sum_{m=1}^{M}\sum_{t=1}^{T} \boldsymbol{\epsilon}'_{t,m}\,\boldsymbol{\epsilon}_{t,m} = \sum_{m=1}^{M} T\,E_T(\boldsymbol{\epsilon}'_{t,s_t}\,\boldsymbol{\epsilon}_{t,s_t}|s_t = m)\\
&= \sum_{m=1}^{M} T\,E_T\Big[E(\boldsymbol{\epsilon}'_{t,s_t}\,\boldsymbol{\epsilon}_{t,s_t}|s_t = m, \mathbf{Y}_T)\,\Pr(s_t = m|\mathbf{Y}_T)\Big]\\
&= T\sum_{m=1}^{M} E_T\Big[E(\boldsymbol{\epsilon}'_{t,m}\,\boldsymbol{\epsilon}_{t,m}|\mathbf{Y}_T)\,\xi_{mt|T}\Big]\\
&= \sum_{m=1}^{M}\sum_{t=1}^{T} [\mathbf{y}_t - (\mathbf{x}'_t \otimes \mathbf{I}_K)\,\boldsymbol{\beta}_m]'\,[\mathbf{y}_t - (\mathbf{x}'_t \otimes \mathbf{I}_K)\,\boldsymbol{\beta}_m]\,\xi_{mt|T}.
\end{aligned}$$

Derivation of (7).

The first derivative of $J(\boldsymbol{\theta})$ with respect to π_m gives

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \pi_m} = \sum_{t=1}^{T} \pi_m^{-1}\,\xi_{mt|T} - \lambda = 0$$

hence

$$\lambda = \sum_{t=1}^{T} \pi_m^{-1} \xi_{mt|T}.$$

Summing up over $m = 1, \ldots, M$, produces

$$\lambda = \lambda \sum_{m=1}^{M} \pi_m = \sum_{m=1}^{M} \sum_{t=1}^{T} \xi_{mt|T} = \sum_{t=1}^{T} \left(\sum_{m=1}^{M} \xi_{mt|T} \right) = \sum_{t=1}^{T} 1 = T.$$

Now Equation (7) follows. Finally, we have $\operatorname{plim}_{T\to\infty} \hat{\pi}_m = E(\xi_{mt|T}) = \pi_m$, for all $m \in \Xi$. This proves the consistency of $\hat{\pi}_m$.

Derivation of (9).

Taking the first derivative of SSE_T with respect to $\boldsymbol{\beta}_m$ gives

$$\frac{\partial\, \mathrm{SSE}_T}{\partial\, \boldsymbol{\beta}_m} = \sum_{t=1}^{T} (-\mathbf{x}_t \otimes \mathbf{I}_K) \left[\mathbf{y}_t - (\mathbf{x}_t' \otimes \mathbf{I}_K)\, \boldsymbol{\beta}_m \right] \xi_{mt|T} = \mathbf{0}$$

hence

$$\left[\sum_{t=1}^{T} (\mathbf{x}_t\, \mathbf{x}_t') \otimes \mathbf{I}_K\, \xi_{mt|T} \right] \hat{\boldsymbol{\beta}}_m = \sum_{t=1}^{T} (\mathbf{x}_t \otimes \mathbf{I}_K)\, \mathbf{y}_t\, \xi_{mt|T}$$

or equivalently

$$(\mathbf{X}_{mT} \otimes \mathbf{I}_K)\, \hat{\boldsymbol{\beta}}_m = \sum_{t=1}^{T} (\mathbf{x}_t \otimes \mathbf{I}_K)\, \mathbf{y}_t\, \xi_{mt|T}.$$

Since $\mathbf{X}_{mT}$ is invertible, we get

$$\hat{\boldsymbol{\beta}}_m = (\mathbf{X}_{mT} \otimes \mathbf{I}_K)^{-1} \left[\sum_{t=1}^{T} (\mathbf{x}_t \otimes \mathbf{I}_K)\, \mathbf{y}_t\, \xi_{mt|T} \right]$$

which gives (9). Since

$$\frac{\partial^2\, \mathrm{SSE}_T}{\partial\, \boldsymbol{\beta}_m\, \partial\, \boldsymbol{\beta}_m'} = 2 \sum_{t=1}^{T} (\mathbf{x}_t\, \mathbf{x}_t') \otimes \mathbf{I}_K\, \xi_{mt|T} = 2\, \mathbf{X}_{mT} \otimes \mathbf{I}_K$$

is positive definite, the OLS estimates $\hat{\boldsymbol{\beta}}_m$ is a minimizer of the function SSE_T for every $m = 1, \ldots, M$.

Derivation of (10).

$$
\begin{aligned}
E_T(\hat{\epsilon}_{t,s_t}\,\hat{\epsilon}^{'}_{t,s_t}|s_t = m) &= E_T\Big[E(\hat{\epsilon}_{t,m}\,\hat{\epsilon}^{'}_{t,m}|\mathbf{Y}_T)\,\xi_{mt|T}\Big] = E_T(\hat{\boldsymbol{\Omega}}_m\,\xi_{mt|T}) \\
&= \hat{\boldsymbol{\Omega}}_m\left(T^{-1}\sum_{t=1}^{T}\xi_{mt|T}\right) = E_T\Big(\hat{\epsilon}_{t,m}\,\hat{\epsilon}^{'}_{t,m}\,\xi_{mt|T}\Big) \\
&= T^{-1}\sum_{t=1}^{T}[\mathbf{y}_t - (\mathbf{x}^{'}_t\otimes\mathbf{I}_K)\,\hat{\boldsymbol{\beta}}_m]\,[\mathbf{y}_t - (\mathbf{x}^{'}_t\otimes\mathbf{I}_K)\,\hat{\boldsymbol{\beta}}_m]^{'}\,\xi_{mt|T}.
\end{aligned}
$$

Derivation of (11).

Using (2) for $s_t = m$ with $\boldsymbol{\epsilon}_{tm} = \boldsymbol{\Sigma}_m\,\mathbf{u}_t$, we get

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_m &= [\mathbf{X}^{-1}_{mT}\otimes\mathbf{I}_K]\left[\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\mathbf{y}_t\,\xi_{mt|T}\right] \\
&= [\mathbf{X}_{mT}\otimes\mathbf{I}_K]^{-1}\left\{\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\left[(\mathbf{x}^{'}_t\otimes\mathbf{I}_K)\,\boldsymbol{\beta}_m + \boldsymbol{\Sigma}_m\,\mathbf{u}_t\right]\xi_{mt|T}\right\} \\
&= [\mathbf{X}_{mT}\otimes\mathbf{I}_K]^{-1}\left[\sum_{t=1}^{T}(\mathbf{x}_t\,\mathbf{x}^{'}_t)\otimes\mathbf{I}_K)\,\boldsymbol{\beta}_m\,\xi_{mt|T} + \sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right] \\
&= [\mathbf{X}_{mT}\otimes\mathbf{I}_K]^{-1}\,[\mathbf{X}_{mT}\otimes\mathbf{I}_K]\,\boldsymbol{\beta}_m + [\mathbf{X}_{mT}\otimes\mathbf{I}_K]^{-1}\left[\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right] \\
&= \boldsymbol{\beta}_m + [\mathbf{X}^{-1}_{mT}\otimes\mathbf{I}_K]\left[\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right].
\end{aligned}
$$

Proof of Theorem 2 *Consistency of* $\hat{\boldsymbol{\beta}}_m$.

$$
\begin{aligned}
\mathrm{plim}_{T\to\infty}\hat{\boldsymbol{\beta}}_m &= \boldsymbol{\beta}_m + \mathrm{plim}_{T\to\infty}\left[\mathbf{X}_{mT}\otimes\mathbf{I}_K\right]^{-1}\mathrm{plim}_{T\to\infty}\left[\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right] \\
&= \boldsymbol{\beta}_m + \mathrm{plim}_{T\to\infty}\left[\frac{1}{T}\,\mathbf{X}_{mT}\otimes\mathbf{I}_K\right]^{-1}\mathrm{plim}_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right] \\
&= \boldsymbol{\beta}_m + \left[\mathbf{Q}^{-1}_m\otimes\mathbf{I}_K\right]E\left[(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right]
\end{aligned}
$$

where $\mathbf{Q}^{-1}_m\otimes\mathbf{I}_K$ is finite by (12). Now, the second summand vanishes. In fact, we have

$$
\begin{aligned}
E\left[(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\right] &= E\left\{E\left[(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}|\mathbf{Y}_T\right]\right\} \\
&= E\left[(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,E(\mathbf{u}_t\,\xi_{mt|T}|\mathbf{Y}_T)\right] \\
&= E\left[(\mathbf{x}_t\otimes\mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,E(\mathbf{u}_t)\,E(\xi_{mt|T}|\mathbf{Y}_T)\right] = \mathbf{0}
\end{aligned}
$$

as $E(\mathbf{u}_t) = \mathbf{0}$. Here, we use the fact that $\mathbf{u}_t$ is independent of s_t, and hence $\boldsymbol{\xi}_{t|T}$. The claim follows.

Consistency of $\hat{\boldsymbol{\Omega}}_m$.

$$
\begin{aligned}
\hat{\boldsymbol{\Omega}}_m &= \frac{\sum_{t=1}^{T} [\mathbf{y}_t - (\mathbf{x}_t' \otimes \mathbf{I}_K)\hat{\boldsymbol{\beta}}_m][\mathbf{y}_t - (\mathbf{x}_t' \otimes \mathbf{I}_K)\hat{\boldsymbol{\beta}}_m]' \xi_{mt|T}}{\sum_{t=1}^{T} \xi_{mt|T}} \\
&= \frac{\sum_{t=1}^{T} [(\mathbf{x}_t' \otimes \mathbf{I}_K)\boldsymbol{\beta}_m - (\mathbf{x}_t' \otimes \mathbf{I}_K)\hat{\boldsymbol{\beta}}_m + \boldsymbol{\epsilon}_{t,m}][(\mathbf{x}_t' \otimes \mathbf{I}_K)\boldsymbol{\beta}_m - (\mathbf{x}_t' \otimes \mathbf{I}_K)\hat{\boldsymbol{\beta}}_m + \boldsymbol{\epsilon}_{t,m}]' \xi_{mt|T}}{\sum_{t=1}^{T} \xi_{mt|T}} \\
&= \frac{\sum_{t=1}^{T} [(\mathbf{x}_t' \otimes \mathbf{I}_K)(\boldsymbol{\beta}_m - \hat{\boldsymbol{\beta}}_m) + \boldsymbol{\epsilon}_{t,m}][(\mathbf{x}_t' \otimes \mathbf{I}_K)(\boldsymbol{\beta}_m - \hat{\boldsymbol{\beta}}_m) + \boldsymbol{\epsilon}_{t,m}]' \xi_{mt|T}}{\sum_{t=1}^{T} \xi_{mt|T}} \\
&= \frac{\sum_{t=1}^{T} (\mathbf{x}_t' \otimes \mathbf{I}_K)(\boldsymbol{\beta}_m - \hat{\boldsymbol{\beta}}_m)\boldsymbol{\epsilon}_{t,m}' \xi_{mt|T} + \sum_{t=1}^{T} \boldsymbol{\epsilon}_{t,m}(\boldsymbol{\beta}_m - \hat{\boldsymbol{\beta}}_m)'(\mathbf{x}_t \otimes \mathbf{I}_K)\xi_{mt|T}}{\sum_{t=1}^{T} \xi_{mt|T}} \\
&\quad + \frac{\sum_{t=1}^{T} (\mathbf{x}_t' \otimes \mathbf{I}_K)(\boldsymbol{\beta}_m - \hat{\boldsymbol{\beta}}_m)(\boldsymbol{\beta}_m - \hat{\boldsymbol{\beta}}_m)'(\mathbf{x}_t \otimes \mathbf{I}_K)\xi_{mt|T} + \sum_{t=1}^{T} \boldsymbol{\epsilon}_{t,m}\boldsymbol{\epsilon}_{t,m}' \xi_{mt|T}}{\sum_{t=1}^{T} \xi_{mt|T}}.
\end{aligned}
$$

Since $\text{plim}_{T\to\infty} \hat{\boldsymbol{\beta}}_m = \boldsymbol{\beta}_m$, it follows that

$$
\begin{aligned}
\text{plim}_{T\to\infty} \hat{\boldsymbol{\Omega}}_m &= \text{plim}_{T\to\infty} \frac{T^{-1} \sum_{t=1}^{T} \boldsymbol{\epsilon}_{t,m}\boldsymbol{\epsilon}_{t,m}' \xi_{mt|T}}{T^{-1} \sum_{t=1}^{T} \xi_{mt|T}} \\
&= \frac{E(\boldsymbol{\epsilon}_{t,m}\boldsymbol{\epsilon}_{t,m}' \xi_{mt|T})}{E(\xi_{mt|T})} = \frac{E(\boldsymbol{\Sigma}_m \mathbf{u}_t \mathbf{u}_t' \boldsymbol{\Sigma}_m' \xi_{mt|T})}{E[E(\xi_{mt}|\mathbf{Y}_T)]} \\
&= \frac{E[E(\boldsymbol{\Sigma}_m \mathbf{u}_t \mathbf{u}_t' \boldsymbol{\Sigma}_m' \xi_{mt|T}|\mathbf{Y}_T)]}{E(\xi_{mt})} \\
&= \frac{E[\boldsymbol{\Sigma}_m E(\mathbf{u}_t \mathbf{u}_t')\boldsymbol{\Sigma}_m' E(\xi_{mt|T}|\mathbf{Y}_T)]}{\pi_m} \\
&= \frac{E[\boldsymbol{\Sigma}_m\boldsymbol{\Sigma}_m' E(\xi_{mt|T}|\mathbf{Y}_T)]}{\pi_m} = \frac{\boldsymbol{\Sigma}_m\boldsymbol{\Sigma}_m' E[E(\xi_{mt|T}|\mathbf{Y}_T)]}{\pi_m} \\
&= \frac{\boldsymbol{\Omega}_m \pi_m}{\pi_m} = \boldsymbol{\Omega}_m.
\end{aligned}
$$

as $\mathbf{u}_t$ is independent of s_t (and hence $\boldsymbol{\xi}_{t|T}$), $E(\mathbf{u}_t \mathbf{u}_t') = \mathbf{I}_K$, and $E(\xi_{mt}) = \pi_m$, for every $m = 1, \dots, M$.

Derivation of (13).

We have to prove that the process $\{(\mathbf{x}_t \otimes \mathbf{I}_K)\boldsymbol{\Sigma}_m \mathbf{u}_t \xi_{mt|T}\}$ has zero mean and asymptotic covariance matrix $\mathbf{Q}_m \otimes \boldsymbol{\Omega}_m$. For the mean, see the proof of consistency of $\hat{\boldsymbol{\beta}}_m$ given above. For the variance, we have

$$\begin{aligned}
\mathrm{var}\Big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi_{mt|T}\Big] &= E\Big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi^2_{mt|T}\,\mathbf{u}_t'\,\boldsymbol{\Sigma}_m'\,(\mathbf{x}_t' \otimes \mathbf{I}_K)\Big]\\
&= E\Big\{E\Big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\mathbf{u}_t\,\xi^2_{mt|T}\,\mathbf{u}_t'\,\boldsymbol{\Sigma}_m'\,(\mathbf{x}_t' \otimes \mathbf{I}_K)|\mathbf{Y}_T\Big]\Big\}\\
&= E\Big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,E(\mathbf{u}_t\,\mathbf{u}_t'\,\xi^2_{mt|T}|\mathbf{Y}_T)\,\boldsymbol{\Sigma}_m'\,(\mathbf{x}_t' \otimes \mathbf{I}_K)\Big]\\
&= E\Big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,E(\mathbf{u}_t\,\mathbf{u}_t')\,E(\xi^2_{mt|T}|\mathbf{Y}_T)\,\boldsymbol{\Sigma}_m'\,(\mathbf{x}_t' \otimes \mathbf{I}_K)\Big]\\
&= E\Big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Sigma}_m\,\boldsymbol{\Sigma}_m'\,E(\xi_{mt|T}|\mathbf{Y}_T)\,(\mathbf{x}_t' \otimes \mathbf{I}_K)\Big]\\
&= E\big[(\mathbf{x}_t \otimes \mathbf{I}_K)\,\boldsymbol{\Omega}_m\,E(\xi_{mt|T}|\mathbf{Y}_T)\,(\mathbf{x}_t' \otimes \mathbf{I}_K)\big]\\
&= E\big[E(\mathbf{x}_t\,\mathbf{x}_t'\,\xi_{mt|T}|\mathbf{Y}_T)\big] \otimes \boldsymbol{\Omega}_m\\
&= E(\mathbf{x}_t\,\mathbf{x}_t'\,\xi_{mt|T}) \otimes \boldsymbol{\Omega}_m \;=\; \mathbf{Q}_m \otimes \boldsymbol{\Omega}_m.
\end{aligned}$$

Here, we use the fact that $\mathbf{u}_t$ is independent of s_t, and hence $\xi^2_{mt|T}$. Furthermore, $E(\mathbf{u}_t\,\mathbf{u}_t') = \mathbf{I}_K$ and $E(\xi^2_{mt|T}|\mathbf{Y}_T) = E(\xi_{mt|T}|\mathbf{Y}_T)$ as $E(\xi^2_{mt}) = E(\xi_{mt}) = \pi_m$, for every $m = 1, \ldots, M$. □

References

Albert, J.H., Chib, S.: Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. J. Bus. Econ. Stat. **11**(1), 1–15 (1993)

Baggenstoss, P.M.: A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. IEEE Trans. Speech Audio Process. **9**(4), 411–416 (2001)

Bao, Y., Hua, Y.: On the Fisher information matrix of a vector ARMA process. Econ. Lett. **123**, 14–16 (2014)

Billio, M., Monfort, A., Robert, C.P.: Bayesian estimation of switching ARMA models. J. Econom. **93**, 229–255 (1999)

Cavicchioli, M.: Spectral density of Markov-switching VARMA models. Econ. Lett. **121**, 218–220 (2013)

Cavicchioli, M.: Determining the number of regimes in Markov-switching VAR and VMA models. J. Time Ser. Anal. **35**(2), 173–186 (2014a)

Cavicchioli, M.: Analysis of the likelihood function for Markov switching VAR(CH) models. J. Time Ser. Anal. **35**(6), 624–639 (2014b)

Cavicchioli, M.: Higher order moments of Markov switching VARMA models. Econom. Theory **33**(6), 1502–1515 (2017a)

Cavicchioli, M.: Asymptotic Fisher information matrix of Markov switching VARMA models. J. Multivar. Anal. **157**, 124–135 (2017b)

Cheng, J.: Spectral density of Markov switching models: derivation, simulation studies and application. Model Assist. Stat. Appl. **11**(4), 277–291 (2016)

Chib, S.: Calculating posterior distributions and model estimates in Markov mixture models. J. Econom. **75**, 79–97 (1996)

Douc, R., Moulines, É., Ryden, T.: Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. Ann. Stat. **32**(5), 2254–2304 (2004)

Fernández-Villaverde, J., Rubio-Ramirez, J., Sargent, T., Watson, M.: ABCs (and Ds) of understanding VARs. Am. Econ. Rev. **97**, 1021–1026 (2007)

Filardo, A.J.: Business-cycle phases and their transitional dynamics. J. Bus. Econ. Stat. **12**(3), 299–308 (1994)

Francq, C., Zakoïan, J.M.: Stationarity of multivariate Markov switching ARMA models. J. Econom. **102**, 339–364 (2001)

Goldfeld, S.M., Quandt, R.E.: A Markov model for switching regressions. J. Econom. **1**, 3–16 (1973)

Hahn, M., Frühwirth-Schnatter, S., Sass, J.: Estimating models based on Markov jump processes given fragmented observation series. AStA Adv. Stat. Anal. **93**(4), 403–425 (2009)

Hamilton, J.D.: A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica **57**, 357–384 (1989)
Hamilton, J.D.: Analysis of time series subject to changes in regime. J. Econom. **45**, 39–70 (1990)
Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton, N.J. (1994)
Hamilton, J.D.: Specification testing in Markov switching time series models. J. Econom. **70**, 127–157 (1996)
Hamilton, J.D.: Macroeconomic regimes and regime shifts. In: Taylor, J.B., Uhlig, H. (eds) Handbook of Macroeconomics, Elsevier vol. 2, pp. 163–201 (2016)
Ioannidis, E.E.: Spectra of bivariate VAR(p) models. J. Stat. Plan. Inference **137**, 554–566 (2007)
Karamé, F.: Asymmetries and Markov-switching structural VAR. J. Econ. Dyn. Control **53**, 85–102 (2015)
Kim, C.J.: Dynamic linear models with Markov switching. J. Econom. **60**, 1–22 (1994)
Kim, C.J., Nelson, C.R.: Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the Business Cycle. Rev. Econ. Stat. **81**(4), 608–619 (1999)
Kim, C.J., Piger, J., Startz, R.: Estimation of Markov regime-switching regression models with endogenous switching. J. Econom. **143**, 263–273 (2008)
Krolzig, H.M.: Markov-switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis. Springer–Verlag, Berlin–Heidelberg–New York. (1997)
Janczura, J., Weron, R.: Efficient estimation of Markov regime-switching models: An application to electricity spot prices. AStA Adv. Stat. Anal. **96**(3), 385–407 (2012)
Janczura, J., Weron, R.: Goodness-of-fit testing for the marginal distribution of regime-switching models with an application to electricity spot prices. AStA Adv. Stat. Anal. **97**(3), 239–270 (2013)
Lanne, M., Lütkepohl, H., Maciejowska, K.: Structural vector autoregressions with Markov switching. J. Econ. Dyn. Control **34**(2), 121–131 (2010)
Lee, C.C., Chang, C.P.: The impact of energy consumption on economic growth: evidence from linear and nonlinear models in Taiwan. Energy **32**, 2282–2294 (2007)
Lütkepohl, H.: New Introduction to Multiple Time Series Analysis. Springer Verlag, Berlin-Heidelberg-New York (2007)
Newton, H.J.: The information matrices of the parameters of multiple mixed time series. J. Multivar. Anal. **8**, 317–323 (1978)
Pataracchia, B.: The Spectral Representation of Markov Switching ARMA Models. Econ. Lett. **112**, 11–15 (2011)
Seifritz, W., Hodgkin, J.: Nonlinear dynamics of the per capita energy consumption. Energy **16**, 615–620 (1991)
Sims, C.A.: Macroeconomics and reality. Econometrica **48**, 1–48 (1980)
Sims, C.A., Waggoner, D.F., Zha, T.: Methods for inference in large multiple-equation Markov-switching models. J. Econom. **146**, 255–274 (2008)
Sims, C.A., Zha, T.: Were there regime switches in U.S. Monetary Policy? Am. Econ. Rev. **96**, 54–81 (2006)
Stelzer, R.: On Markov-switching ARMA processes-stationarity, existence of moments and geometric ergodicity. Econom. Theory **25**(1), 43–62 (2009)
Ubierna, A., Velilla, S.: A goodness-of-fit process for ARMA(p, q) models based on a modified residual autocorrelation sequence. J. Stat. Plan. Inference **137**, 2903–2919 (2007)
Velilla, S., Thu, H.N.: A goodness-of-fit test for VARMA(p, q) models. J. Stat. Plan. Inference **197**, 126–140 (2018)
Wang, J., Zivot, E.: A Bayesian time series model of multiple structural changes in level, trend, and variance. J. Bus. Econ. Stat. **18**(3), 374–386 (2000)
Whittle, P.: The analysis of multiple stationary time series. J. R. Stat. Soc. Ser. B **15**, 125–139 (1953)
Wong, C.S., Li, W.K.: On a mixture autoregressive model. J. R. Stat. Soc. **B62**, 95–115 (2000)
Yang, M.: Some properties of vector autoregressive processes with Markov-switching coefficients. Econom. Theory **16**, 23–43 (2000)
Yang, H., He, G.: Some properties of matrix product and its applications in nonnegative tensor decomposition. J. Inform. Comput. Sci. **3**(4), 269–280 (2008)