Likelihood Ratio Test and Information Criteria for Markov Switching Var Models: An Application to the Italian Macroeconomy

Maddalena Cavicchioli¹

Received: 23 December 2014 / Accepted: 13 April 2015 / Published online: 24 April 2015
© Società Italiana degli Economisti (Italian Economic Association) 2015

Abstract In this work we consider multivariate autoregressions subject to Markovian changes in regime. Estimation methods and filtering techniques for such processes are well established in the literature as well as the asymptotic distribution of the maximum likelihood estimators. Assuming the conditions under which the standard asymptotic distribution theory holds, the likelihood ratio (LR) has the null distribution. We give explicit formulae for LR tests of various hypotheses of interest in the context of Markov switching VAR models. The proposed LR statistic has a rather simple form as it reduces to the use of the estimated unrestricted and restricted variance-covariance matrices. Moreover, we derive simple expressions for some information criteria to address the question of linearity versus nonlinearity. An application to Italian macroeconomic data gives new insights on the number of regimes and the dynamics characterizing the economy.

Keywords Markov-switching VAR models · Filtering · Smoothing · MLE · LR tests · Information criteria · Italian economy

JEL Classification C01 · C32 · C51

This article is based on the author's PhD dissertation, that has been awarded by SIE (Italian Economic Association) as best PhD Thesis in 2014.

✉ Maddalena Cavicchioli
maddalena.cavicchioli@unimore.it

¹ Department of Economics “Marco Biagi”, University of Modena and Reggio E., Viale Berengario 51, 41121 Modena, Italy

1 Introduction

In this paper we consider Markov switching vector autoregression (in short, MS VAR) models and deal with the problem of model specification. Despite the interest of applied financial and economic researchers in these models and the abundance of empirical works, there have been no clear guide on the issue of model selection in order to verify adequacy of the model to the data. In other context, such as regression-type models, a variety of tests is routinely performed, so that it seems natural to demand same standards for MS models. The first issue to be considered relates to estimation and filtering techniques. A fundamental contribution in this field is due to Hamilton (1990, 1993). In his work the focus is on expectation maximization (EM) algorithm. The main limitation of this procedure is the implementation in the case of models with autoregressive dynamic, which conduces to some problems. This depends on the fact that the log likelihood is typically multimodal and convergence problems might appear. Thus, any optimization algorithm, including EM, may converge towards a local maximum or even a saddle point. Furthermore, starting values can have a profound impact on which local optimum is selected. Today there are no methods which guarantee to find the maximum likelihood estimators (MLE), but the best advice available is to start the optimization algorithm from several different (possibly random) points in the parameter set. A further drawback of the EM algorithm is its rate of convergence, which is only linear in the vicinity of the MLE. Various modifications of the basic algorithm have been proposed in the literature to improve it (see, for example, Bickel et al. 1998 and references therein). However, little has been published at present on which of these modifications perform well. For practical implementation Krolzig (1997) relied on EM algorithm in conjunction with filter recursions (Baum–Lindgren–Hamilton–Kim, in short BLHK, filter, presented in Chp. 5, p.79) in order to determine smoothed probabilities and MLE of model's parameter vector. A major improvement in estimation is due to Cavicchioli (2014b) in which explicit expressions of the MLE of MS VAR models and their corresponding limiting covariance matrices have been derived. The advantage is twofold. Firstly, the implementation does not require the EM algorithm. We need only the computation of the smoothed probabilities that can be done by iterating backward the BLHK filter, mentioned above. This reduces calculation times by as much as several orders of magnitude and overpasses the above problems arising from EM algorithm, improving computational performances. Secondly, classical test procedures could be implemented for the selection of lag order of the autoregressive polynomial as well as the linearity of some parameters of the model, assuming that standard asymptotic theory holds. However, a critical decision remains, that is, the choice of the number of regimes in the specification of the MS VAR. In fact, testing procedures on regime's number suffer from non-standard asymptotic distributions of the likelihood ratio test statistic due to existence of nuisance parameters under the null hypothesis. In particular, when nuisance parameters are not identified under the null, the likelihood function is not locally quadratic with respect to these parameters under the null at the optimum and the corresponding scores are identically zero. As a result, the information matrix is singular under the null hypothesis and usual tests do not have an asymptotic standard distribution. Thus, we have two problems at hand. From one side the determination of the regime number, that is ultimately, the prob-

lem of linearity versus nonlinearity. From the other side, having set the number of regimes, we want to investigate parameters stability avoiding the use of EM algorithm, given the drawbacks discussed above. Concerning the first problem, the most employed methods for determining the state dimension are mainly based on either recursive complexity-penalized likelihood criteria (see, for example, [Psaradakis and Spagnolo 2003, 2006](#); [Olteanu and Rynkiewicz 2007](#); [Rios and Rodriguez 2008](#)) or on finite-order vector autoregressive moving average (VARMA) representations of the initial Markov switching models (see, for example, [Krolzig 1997](#); [Zhang and Stine 2001](#); [Francq and Zakoian 2001](#); [Cavicchioli 2014a](#)). Furthermore, [Hansen \(1992\)](#), [Garcia \(1998\)](#), [Cho and White \(2007\)](#), and [Carter and Steigerwald \(2012\)](#) examined the distribution of the likelihood ratio statistics which is non-standard. However, their procedures can be quite involved to be implemented and, for this reason, not widely used in practice. Other likelihood ratio tests for parameter stability based on bootstrapping methods were proposed by [Di Sanzo \(2009\)](#). See also the class of optimal tests for the constancy of parameters in random coefficients models (including also Markov switching models) proposed recently by [Carrasco et al. \(2014\)](#). These authors used Bartlett-type identities for the construction of the test statistics, and bootstrap critical values. Another option is to conduct generic specification tests (precisely, for omitted autocorrelation, omitted ARCH, and omitted explanatory variables) developed by [Hamilton \(1996\)](#) on the hypothesis that an m -regime model accurately describes the data to avoid test indeterminacy. Finally, [Huang \(2014\)](#) avoids the problem of having nuisance parameters studying the special case of testing two regimes versus the alternative of one ultimate absorbing state. This paper responds to both questions. With regard to the second problem, we give simple formulae for likelihood ratio (LR) tests of various hypotheses of interest in the general case of MS VAR models to check parameters stability. Our formulae are in close forms and their implementation does not require the EM algorithm, improving computational performances. Moreover, it is shown that the maximum likelihood theory for MS VARs we develop generalizes the classical theory for linear VAR models described in [Hamilton \(1994\)](#), reconciling in a unified framework estimation and testing procedures for linear and non-linear VAR models. Furthermore, in response to the first problem, we give simple formulae to address the well-known problem of testing linearity versus nonlinearity or versus several states alternative with relatively simple methods which could be easily applied in practice. This relies on specifying penalized likelihood criteria using our close form expression for the likelihood which avoids burdensome and recursive likelihood computations. A simulation experiment shows the goodness of this approach. The rest of the paper is organized as follows. In Sect. 2 we introduce the MS VAR model and recall the matrix formulae for the MLE based on the results proved in [Cavicchioli \(2014b\)](#). Then we give an explicit formula in close form for the likelihood function of these models. In Sect. 3 we explicitly specify the likelihood ratio test in this context and propose various LR tests for several hypotheses of interest. Such tests require that the standard asymptotic theory for MLE holds and that the number of states is unaltered under the null hypothesis. Penalized likelihood criteria procedures for those hypotheses altering the number of regimes are given in Sect. 4. Section 5 is devoted to illustrate the presence of nonlinearities in Italian macroeconomic data. Finally, a brief summary and conclusion are given in Sect. 6.

2 Markov Switching VAR Models

In this section we introduce the MS VAR model and recall some estimation results proved in Cavicchioli (2014b). Then we obtain a close form expression for the maximum value of the likelihood function of the MS VAR model. It contains as a particular (linear) case the derivation by Hamilton (1994). Let $\mathbf{y}_t$ be a $(K \times 1)$ random vector which follows a M -regime Markov switching (MS) VAR(p) process, with $p > 0$:

$$\mathbf{y}_t + \sum_{i=1}^p \Phi_{s_t, i} \mathbf{y}_{t-i} = \mathbf{v}_{s_t} + \Sigma_{s_t} \mathbf{u}_t \quad (1)$$

where $(s_t)_{t \geq 0}$ is an M -state, homogeneous, irreducible and ergodic Markov chain, $\mathbf{u}_t \sim \text{NID}(\mathbf{0}, \mathbf{I}_K)$, and $\Phi_{m, i}$ is a $(K \times K)$ matrix for every $i = 1, \dots, p$ and $m = 1, \dots, M$. Let $\mathbf{P} = (p_{ij})_{i, j=1, \dots, M}$ denote the transition matrix of the chain, where $p_{ij} = \text{Pr}(s_t = j | s_{t-1} = i)$. Ergodicity implies the existence of a stationary vector of probabilities $\boldsymbol{\pi} = (\pi_1, \dots, \pi_M)'$ satisfying $\mathbf{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$ and $\mathbf{i}_M'\boldsymbol{\pi} = 1$, where $\mathbf{i}_M$ denotes a $(M \times 1)$ vector of ones. Irreducibility implies that $\pi_i > 0$ for $i = 1, \dots, M$, meaning that all unobservable states are possible. We also assume that the Markov chain $(s_t)_{t \geq 0}$ is stationary (or, initialized from its stationary distribution). This implies that the MS VAR process is strictly stationary and hence ergodic. If we set $\mathbf{Y}_{t-1} = (\mathbf{y}_{t-1}', \dots, \mathbf{y}_{t-p}')'$, then we have

$$\mathbf{y}_t | s_t, \mathbf{Y}_{t-1} \sim \text{NID}\left(\mathbf{v}_{s_t} - \sum_{i=1}^p \Phi_{s_t, i} \mathbf{y}_{t-i}, \boldsymbol{\Omega}_{s_t}\right) \quad (2)$$

where $\boldsymbol{\Omega}_{s_t} = \Sigma_{s_t} \Sigma_{s_t}'$ (positive definite matrix). The unknown parameter $\boldsymbol{\theta}$ consists of the elements of the intercept vectors $(\mathbf{v}_1, \dots, \mathbf{v}_M)$, the variance-covariance matrices $(\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_M)$ (or, equivalently, their inverses) and the matrices $\Phi_{m, i}$ for $m = 1, \dots, M$ and $i = 1, \dots, p$. The vector $\boldsymbol{\rho}$ is formed by the transition probabilities $(p_{11}, p_{12}, \dots, p_{MM})$. Collect the unknown parameters in a single vector $\boldsymbol{\lambda} = (\boldsymbol{\theta}', \boldsymbol{\rho}')'$. Model (1) implies the estimation of a number of parameters which is equal to $M[K + pK^2 + K(K + 1)/2 + (M - 1)]$ for the most general situation. In the case $s_t = m$, the log of the conditional density $\eta_{mt}(\boldsymbol{\theta}) = p(\mathbf{y}_t | s_t = m, \mathbf{Y}_{t-1}; \boldsymbol{\theta})$ becomes

$$\begin{aligned} \log_e \eta_{mt}(\boldsymbol{\theta}) &= \log_e p(\mathbf{y}_t | s_t = m, \mathbf{Y}_{t-1}; \boldsymbol{\theta}) = -\frac{K}{2} \log_e(2\pi) - \frac{1}{2} \log_e |\boldsymbol{\Omega}_m| \\ &\quad - \frac{1}{2} \left(\mathbf{y}_t - \mathbf{v}_m + \sum_{i=1}^p \Phi_{m, i} \mathbf{y}_{t-i} \right)' \boldsymbol{\Omega}_m^{-1} \left(\mathbf{y}_t - \mathbf{v}_m + \sum_{i=1}^p \Phi_{m, i} \mathbf{y}_{t-i} \right). \end{aligned}$$

Set $\boldsymbol{\eta}_t(\boldsymbol{\theta}) = (\eta_{1t}(\boldsymbol{\theta}) \dots \eta_{Mt}(\boldsymbol{\theta}))'$. A useful representation for a Markov chain is obtained by letting $\boldsymbol{\xi}_t$ denote a random $(M \times 1)$ vector whose j th element is one if $s_t = j$ and zero otherwise. Then the chain can be expressed as AR (1) model $\boldsymbol{\xi}_{t+1} = \mathbf{P}'\boldsymbol{\xi}_t + \mathbf{v}_{t+1}$, where $\mathbf{v}_{t+1} = \boldsymbol{\xi}_{t+1} - E(\boldsymbol{\xi}_{t+1} | \boldsymbol{\xi}_t)$ is a zero mean martingale

difference sequence. Following Krolzig (1997), Formula (6.21), p. 106, the expected log likelihood function for (1) can be approximated (up to a constant term) by

$$\mathcal{L}(\lambda) = \log_e L(\lambda | \mathbf{Y}_T) = \sum_{t=1}^T \sum_{m=1}^M \log_e \eta_{mt}(\theta) \widehat{\xi}_{mt|T}(\lambda) \quad (3)$$

where $\widehat{\xi}_{mt|T} = E(\xi_t | \mathbf{Y}_T)$ are the smoothed probabilities. An algorithm for calculating such probabilities can be found in Hamilton (1994), Chp. 22 and Krolzig (1997), Chp. 5. Define the following matrices:

$$\begin{aligned} \Phi_m &= (\Phi_{m,1} \cdots \Phi_{m,p}) & \mathbf{A}_m &= (\mathbf{A}_m(i, j)) & \mathbf{B}_m &= (\mathbf{B}_m(i)) \\ \mathbf{C}_m &= (\mathbf{C}_m(i)) & \mathbf{S}_m &= \left(\sum_{t=1}^T \widehat{\xi}_{mt|T} \right) & \mathbf{T}_m &= \left(- \sum_{t=1}^T \mathbf{y}_t \widehat{\xi}_{mt|T} \right) \end{aligned}$$

where

$$\mathbf{A}_m(i, j) = \sum_{t=1}^T \mathbf{y}_{t-i} \mathbf{y}'_{t-j} \widehat{\xi}_{mt|T} \quad \mathbf{B}_m(i) = \sum_{t=1}^T \mathbf{y}'_{t-i} \widehat{\xi}_{mt|T}$$

and

$$\mathbf{C}_m(i) = - \sum_{t=1}^T \mathbf{y}_t \mathbf{y}'_{t-i} \widehat{\xi}_{mt|T}$$

for $m = 1, \dots, M$ and $i = 1, \dots, p$. Let us consider the matrices

$$\mathbf{X}_m = \mathbf{B}_m' \mathbf{B}_m - \mathbf{S}_m \mathbf{A}_m \quad \mathbf{W}_m = \mathbf{T}_m \mathbf{B}_m - \mathbf{S}_m \mathbf{C}_m$$

for every $m = 1, \dots, M$. Then $\mathbf{X}_m = (\mathbf{X}_m(i, j))$ and $\mathbf{W}_m = (\mathbf{W}_m(i))$, where

$$\begin{aligned} \mathbf{X}_m(i, j) &= \left[\sum_{t=1}^T \mathbf{y}_{t-i} \widehat{\xi}_{mt|T} \right] \left[\sum_{t=1}^T \mathbf{y}'_{t-j} \widehat{\xi}_{mt|T} \right] - \left[\sum_{t=1}^T \widehat{\xi}_{mt|T} \right] \left[\sum_{t=1}^T \mathbf{y}_{t-i} \mathbf{y}'_{t-j} \widehat{\xi}_{mt|T} \right] \\ \mathbf{W}_m(i) &= - \left[\sum_{t=1}^T \mathbf{y}_t \widehat{\xi}_{mt|T} \right] \left[\sum_{t=1}^T \mathbf{y}'_{t-i} \widehat{\xi}_{mt|T} \right] + \left[\sum_{t=1}^T \widehat{\xi}_{mt|T} \right] \left[\sum_{t=1}^T \mathbf{y}_t \mathbf{y}'_{t-i} \widehat{\xi}_{mt|T} \right] \end{aligned}$$

for every $m = 1, \dots, M$ and $i, j = 1, \dots, p$. The following result was proved in Cavicchioli (2014b), Theorem 5.

Theorem 1 Assume that $\mathbf{X}_m$ is non-singular for every $m = 1, \dots, M$. With the above notations, the maximum likelihood estimates of the unknown parameters from MS VAR(p) model (1), $p > 0$, are given by

$$\widehat{\mathbf{v}}_m = S_m^{-1}(\mathbf{W}_m \mathbf{X}_m^{-1} \mathbf{B}_m' - \mathbf{T}_m) \quad (4)$$

$$\widehat{\Phi}_m = \mathbf{W}_m \mathbf{X}_m^{-1} \quad (5)$$

and

$$\widehat{\Omega}_m = S_m^{-1} \left[\sum_{t=1}^T (\widehat{\Sigma}_m \widehat{\mathbf{u}}_t)(\widehat{\Sigma}_m \widehat{\mathbf{u}}_t)' \widehat{\xi}_{mt|T} \right] \quad (6)$$

where

$$\widehat{\Sigma}_m \widehat{\mathbf{u}}_t = \mathbf{y}_t - \widehat{\mathbf{v}}_m + \sum_{i=1}^p \widehat{\Phi}_{m,i} \mathbf{y}_{t-i}.$$

For the consistency and the asymptotic normality of the MLE as well as the asymptotic covariance matrix and the Wald tests see the quoted paper. To perform a likelihood ratio test, we need to calculate the maximum value achieved for (3). Thus consider

$$\mathcal{L}(\widehat{\lambda}) = \sum_{t=1}^T \sum_{m=1}^M \left\{ -\frac{K}{2} \log_e(2\pi) - \frac{1}{2} \log_e |\widehat{\Omega}_m| - \frac{1}{2} \widehat{\mathbf{u}}_t' \widehat{\Sigma}_m' \widehat{\Omega}_m^{-1} \widehat{\Sigma}_m \widehat{\mathbf{u}}_t \right\} \widehat{\xi}_{mt|T} \quad (7)$$

using the estimates from Theorem 1. We prove in the Appendix that the last term in (7) is

$$\frac{1}{2} \sum_{t=1}^T \sum_{m=1}^M \widehat{\mathbf{u}}_t' \widehat{\Sigma}_m' \widehat{\Omega}_m^{-1} \widehat{\Sigma}_m \widehat{\mathbf{u}}_t \widehat{\xi}_{mt|T} = \frac{TK}{2} \quad (8)$$

Substituting (8) into (7) produces

$$\begin{aligned} \mathcal{L}(\widehat{\lambda}) &= -\frac{K}{2} \log_e(2\pi) \sum_{t=1}^T \sum_{m=1}^M \widehat{\xi}_{mt|T} - \frac{1}{2} \sum_{t=1}^T \sum_{m=1}^M \log_e |\widehat{\Omega}_m| \widehat{\xi}_{mt|T} - \frac{TK}{2} \\ &= -\frac{TK}{2} \log_e(2\pi) - \frac{1}{2} \sum_{m=1}^M S_m \log_e |\widehat{\Omega}_m| - \frac{TK}{2} \end{aligned} \quad (9)$$

This formula makes likelihood ratio tests particularly simple to perform. What is interesting is that it generalizes Formula (11.1.32) from [Hamilton \(1994\)](#), Chp. 11, p. 296 obtained for state-invariant VAR(p) model. In fact, for the case $M = 1$, (9) reduces to

$$\mathcal{L}(\widehat{\lambda}) = -\frac{TK}{2} \log_e(2\pi) - \frac{T}{2} \log_e |\widehat{\Omega}| - \frac{TK}{2}.$$

3 Likelihood Ratio Tests

The asymptotic normal distribution of the ML estimators ensures that the classical tests, known from time-invariant VAR models, can be applied. Unfortunately for one important exception, standard asymptotic distribution theory cannot be invoked, namely, hypothesis tests of the number of regimes of the Markov chain. Further specification procedures for those hypotheses altering the number of regimes under the null will be discussed in the next section. Here we are interested in likelihood ratio (LR) tests for MS VAR models as (1) under the validity of standard asymptotic theory (see Krolzig 1997, Chp. 7; Lutkepohl 1991, Chp. 4). As well-known, the LR test is based on the statistic

$$LR = 2[\mathcal{L}(\hat{\lambda}) - \mathcal{L}(\hat{\lambda}_r)]$$

where $\hat{\lambda}$ denotes the unconstrained ML estimator and $\hat{\lambda}_r$ the restricted ML estimator under the null hypothesis $H_0 : \phi(\lambda) = 0$. Here $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^r$ is a continuously differentiable function with rank $r \leq n$ (where n is the dimension of the parameter space). Under the null hypothesis H_0 , LR statistic has an asymptotic χ^2 -distribution with r degrees of freedom [see Krolzig 1997, Chp. 7, Formula (7.3)]. A necessary condition for the validity of this standard result is that the number of regimes M is unaltered under the null. Suppose we want to test the null hypothesis H_0 that a set of variables was generated from a Gaussian MS VAR model with p_0 lags against the alternative specification H_1 of $p_1 > p_0$ lags (with M fixed). To estimate the system under the null hypothesis, we perform a set of OLS regressions of each variable in the system on a constant term and on p_0 lags of all the variable in the system for each regime $s_t = m \in \{1, \dots, M\}$. Let

$$\hat{\Omega}_m^{(0)} = \left(S_m^{(0)}\right)^{-1} \left[\sum_{t=1}^T \hat{\Sigma}_m^{(0)} \hat{u}_t^{(0)} \left(\hat{\Sigma}_m^{(0)} \hat{u}_t^{(0)}\right)' \hat{\xi}_{mt|T}^{(0)} \right]$$

be the variance-covariance matrix of the residuals from these regressions. Then the maximum value for the log likelihood under H_0 is

$$\mathcal{L}_0^* = \mathcal{L}(\hat{\lambda}^{(0)}) = -\frac{TK}{2} \log_e(2\pi) - \frac{1}{2} \sum_{m=1}^M S_m^{(0)} \log_e |\hat{\Omega}_m^{(0)}| - \frac{TK}{2}.$$

Similarly, the system is estimated under the alternative hypothesis by OLS regressions that include p_1 lags of all the variables. The maximized log likelihood under the alternative is

$$\mathcal{L}_1^* = \mathcal{L}(\hat{\lambda}^{(1)}) = -\frac{TK}{2} \log_e(2\pi) - \frac{1}{2} \sum_{m=1}^M S_m^{(1)} \log_e |\hat{\Omega}_m^{(1)}| - \frac{TK}{2}$$

where $\widehat{\mathbf{\Omega}}_m^{(1)}$ is the variance-covariance matrix of the residuals from this second set of regressions. Twice the log likelihood ratio is then

$$\begin{aligned} LR &= 2(\mathcal{L}_1^* - \mathcal{L}_0^*) = 2 \left\{ -\frac{1}{2} \sum_{m=1}^M S_m^{(1)} \log_e |\widehat{\mathbf{\Omega}}_m^{(1)}| + \frac{1}{2} \sum_{m=1}^M S_m^{(0)} \log_e |\widehat{\mathbf{\Omega}}_m^{(0)}| \right\} \\ &= \sum_{m=1}^M \left[S_m^{(0)} \log_e |\widehat{\mathbf{\Omega}}_m^{(0)}| - S_m^{(1)} \log_e |\widehat{\mathbf{\Omega}}_m^{(1)}| \right] = \sum_{m=1}^M \log_e \frac{|\widehat{\mathbf{\Omega}}_m^{(0)}|^{S_m^{(0)}}}{|\widehat{\mathbf{\Omega}}_m^{(1)}|^{S_m^{(1)}}} \\ &= \log_e \left(\prod_{m=1}^M \frac{|\widehat{\mathbf{\Omega}}_m^{(0)}|^{S_m^{(0)}}}{|\widehat{\mathbf{\Omega}}_m^{(1)}|^{S_m^{(1)}}} \right) \end{aligned} \quad (10)$$

which generalizes Formula (11.1.33) from Hamilton (1994), Chp. 11, p. 297 for VAR models (case $M = 1$):

$$LR = 2(\mathcal{L}_1^* - \mathcal{L}_0^*) = T \left\{ \log_e |\widehat{\mathbf{\Omega}}^{(0)}| - \log_e |\widehat{\mathbf{\Omega}}^{(1)}| \right\}.$$

Under the null hypothesis, this statistic asymptotically has a χ^2 distribution with degrees of freedom equal to the number of restrictions imposed under H_0 . Each equation in the specification restricted by H_0 has $(p_1 - p_0)$ fewer lags on each on K variables compared with H_1 . Thus the null imposes $K(p_1 - p_0)$ restrictions on each equation. Since there are MK such equations, H_0 imposes $MK^2(p_1 - p_0)$ restrictions. Thus the magnitude calculated in (10) is asymptotically χ^2 with $r = MK^2(p_1 - p_0)$ degrees of freedom. See again Hamilton (1994), Chp. 11, p. 297, for the linear case which has $r = K^2(p_1 - p_0)$ degrees of freedom. As long as the number of regimes remains unchanged and the identifiability assumption of standard asymptotic theory hold, LR tests concerning linear restrictions of the MS VAR coefficient vector θ can be performed as in linear models. It is of interest in applications knowing which parameters of the model have to be considered as switching and which instead are constant in different states. Hereafter we list some tests to be performed when the following assumption are satisfied. For Test (i) we assume $\mathbf{v}_n \neq \mathbf{v}_m$ for all $n \neq m$. For Tests (ii), (iii) and (iv) we assume that $\text{vech } \mathbf{\Omega}_n \neq \text{vech } \mathbf{\Omega}_m$ for all $n \neq m$.

- (i) *Testing for regime-dependent heteroskedasticity* ($r = (M - 1)[K(K + 1)/2]$).
 $H_0 : \text{vech } \mathbf{\Omega}_n = \text{vech } \mathbf{\Omega}_m$ for all $n, m = 1, \dots, M$ vs. $H_1 : \text{vech } \mathbf{\Omega}_n \neq \text{vech } \mathbf{\Omega}_m$ for at least one $n \neq m$.
- (ii) *Testing for regime-dependent intercepts* ($r = (M - 1)K$).
 $H_0 : \mathbf{v}_n = \mathbf{v}_m$ for all $n, m = 1, \dots, M$ vs. $H_1 : \mathbf{v}_n \neq \mathbf{v}_m$ for at least one $n \neq m$.
- (iii) *Testing for regime-dependent autoregressive parameters* ($r = (M - 1)pK^2$).
 $H_0 : \Phi_{ni} = \Phi_{mi}$ for all $n, m = 1, \dots, M$ ($i = 1, \dots, p$) vs. $H_1 : \Phi_{ni} \neq \Phi_{mi}$ for at least one $n \neq m$.
- (iv) *Testing the order of lags in the MS VAR model* ($r = MK^2$).
 $H_0 : \Phi_{m,p+1} = \mathbf{0}$ for all $m = 1, \dots, M$ vs. $H_1 : \Phi_{m,p+1} \neq \mathbf{0}$ for at least one m .

For illustration purposes, the LR Test (i) which corresponds to Formula (10) becomes (see the Appendix)

$$LR = 2(\mathcal{L}_1^* - \mathcal{L}_0^*) = \log_e \left(\frac{|\widehat{\Omega}^{(0)}|^T}{\prod_{m=1}^M |\widehat{\Omega}_m^{(1)}|^{S_m^{(1)}}} \right) \quad (11)$$

Now we illustrate our testing procedures with two examples. Firstly, suppose that a bivariate 2-state VAR model is estimated with three and four lags. So we have $K = M = 2$, $p_0 = 3$ and $p_1 = 4$. Say that the original sample contains 50 observations on each variable, denoted $\mathbf{y}_{-3}, \mathbf{y}_{-2}, \dots, \mathbf{y}_{46}$ and that observations 1 through 46 were used to estimate both the three and four-lag specifications so that $T = 46$. Set $\widehat{\epsilon}_{m,t}^{(0)} = \widehat{\Sigma}_m^{(0)} \widehat{\mathbf{u}}_t^{(0)}$ for $m = 1, 2$. Suppose that

$$\begin{aligned} \widehat{\Omega}_1^{(0)} &= [S_1^{(0)}]^{-1} \sum_{t=1}^T \widehat{\epsilon}_{1,t}^{(0)} \widehat{\epsilon}_{1,t}^{(0)'} \widehat{\xi}_{1t|T}^{(0)} = \begin{pmatrix} 2.0 & 1.0 \\ 1.0 & 2.5 \end{pmatrix} \\ \widehat{\Omega}_2^{(0)} &= [S_2^{(0)}]^{-1} \sum_{t=1}^T \widehat{\epsilon}_{2,t}^{(0)} \widehat{\epsilon}_{2,t}^{(0)'} \widehat{\xi}_{2t|T}^{(0)} = \begin{pmatrix} 3.0 & 2.0 \\ 2.0 & 4.0 \end{pmatrix} \end{aligned}$$

where $S_1^{(0)} = 10$ and $S_2^{(0)} = 20$. Suppose that when a fourth lag is added to each regression, the residual covariances matrices are reduced to

$$\widehat{\Omega}_1^{(1)} = \begin{pmatrix} 1.8 & 0.9 \\ 0.9 & 2.2 \end{pmatrix} \quad \widehat{\Omega}_2^{(1)} = \begin{pmatrix} 2.8 & 1.9 \\ 1.9 & 3.5 \end{pmatrix}$$

with $S_1^{(1)} = 8$ and $S_2^{(1)} = 18$. Then we have

$$\begin{aligned} S_1^{(0)} \log_e |\widehat{\Omega}_1^{(0)}| + S_2^{(0)} \log_e |\widehat{\Omega}_2^{(0)}| &= 10 \log_e 4 + 20 \log_e 8 = 55.45 \\ S_1^{(1)} \log_e |\widehat{\Omega}_1^{(1)}| + S_2^{(1)} \log_e |\widehat{\Omega}_2^{(1)}| &= 8 \log_e 3.15 + 18 \log_e 6.19 = 41.99 \end{aligned}$$

Thus the LR ratio gives

$$LR = 2(\mathcal{L}_1^* - \mathcal{L}_0^*) = 55.45 - 41.99 = 13.46.$$

The degrees of freedom for this test are $r = MK^2(p_1 - p_0) = 8$. Since the 0.05 critical value for a $\chi^2(8)$ variable is 15.5, we get $13.46 < 15.5$, hence the null hypothesis is accepted. The dynamics are completely captured by a three-lag MS VAR and a four-lag specification does not seem preferable. Secondly, we illustrate Test (i) for $K = M = 2$, $p = 4$ and $T = 46$. Suppose that

$$\widehat{\Omega}^{(0)} = \begin{pmatrix} 2.0 & 1.0 \\ 1.0 & 2.5 \end{pmatrix}$$

under H_0 and $\widehat{\Omega}_1^{(1)}$ and $\widehat{\Omega}_2^{(1)}$ under H_1 are as in the previous example with $S_1^{(1)} = 3$ and $S_2^{(1)} = 5$. Thus the LR ratio gives

$$LR = 2(\mathcal{L}_1^* - \mathcal{L}_0^*) = 46 \log_e 4 - [3 \log_e 3.15 + 5 \log_e 6.19] = 51.21.$$

The degrees of freedom for this test are given by $r = (M - 1)[K + pK^2 + K(K + 1)/2] = 21$. Since the 0.05 critical value for a $\chi^2(21)$ variable is 32.7, we get $51.21 > 32.7$, hence the null hypothesis is rejected, so that the dynamics seems to be not homoskedastic.

4 Penalized Likelihood Criteria

In this section we deal with the problem of determining the number of states, which includes also the case of testing the null hypothesis of a single state (linearity) versus the alternative of several states (nonlinearity). As mentioned before, standard asymptotic theory cannot be invoked in general for testing with different number of states in the Markov chain. Denote by λ the vector of the unknown parameters of a multivariate MS VAR model as in (1). We follow the criteria introduced by [Psaradakis and Spagnolo \(2003\)](#) to estimate the number of states M . However, we further specify the information criteria (IC) to avoid recursive procedures. In particular, we select M such that it minimizes a criterion function of the form:

$$C = -\log_e L(\widehat{\lambda}) + P(\dim(\Lambda))$$

where $\widehat{\lambda}$ is the MLE of λ and P is a non-negative real-valued random variable which depends on the dimension of the parameter space Λ so that it measures the complexity of the model. Using (9) we are now able to give a simpler form of criterion function:

$$C = \left[\frac{TK}{2} \log_e (2\pi) + \frac{1}{2} \sum_{m=1}^M S_m \log_e |\widehat{\Omega}_m| + \frac{TK}{2} \right] + P(\dim(\Lambda)).$$

Surely, it has the advantage of greatly reducing the computational burden and to speed up the final decision. The expression for C and, consequently, the expressions for the information criteria given below can be simplified further by noting that the first and the third terms are constant for every $m = 1, \dots, M$, and hence they do not play a role when minimizing the information criterion. Three popular choices of penalty term are the following:

$$P(\dim(\Lambda)) = \dim(\Lambda)$$

$$P(\dim(\Lambda)) = \frac{1}{2} \log_e (T) \dim(\Lambda)$$

$$P(\dim(\Lambda)) = \log_e (\log_e (T)) \dim(\Lambda)$$

where $\dim(\Lambda) = M[K + pK^2 + K(K + 1)/2 + (M - 1)]$ for the most general case. The penalty terms correspond to the Akaike's ([1973, 1974](#)) information criterion (AIC),

Table 1 Parameters specification of the data generating processes (DGP)

DGP	ν	ϕ	σ	$\mathbf{P}$
AR(1)	2	0.9	1	–
MS(2)-AR(1)	[–1 2]	[0.1 0.8]	[4 2]	$\begin{bmatrix} 0.2 & 0.3 \\ 0.8 & 0.7 \end{bmatrix}$
MS(3)-AR(1)	[–1 2 5]	[0.1 0.5 0.9]	[4 2 1]	$\begin{bmatrix} 0.47 & 0.08 & 0.04 \\ 0.14 & 0.29 & 0.02 \\ 0.39 & 0.63 & 0.94 \end{bmatrix}$

In the table we report the transition matrices $\mathbf{P}$ s and the model parameters (means ν s, standard deviation σ s and autoregressive coefficients ϕ s)

Schwarz's (1978) Bayesian information criterion (BIC), and Hannan and Quinn's (1979) criterion (HQC), respectively. Finally, we obtain a simplified form for the previous information criteria:

$$\begin{aligned}
 AIC &= \frac{1}{2} \sum_{m=1}^M S_m \log_e |\hat{\boldsymbol{\Omega}}_m| + \dim(\boldsymbol{\Lambda}) \\
 BIC &= \frac{1}{2} \sum_{m=1}^M S_m \log_e |\hat{\boldsymbol{\Omega}}_m| + \frac{1}{2} \log_e(T) \dim(\boldsymbol{\Lambda}) \\
 HQC &= \frac{1}{2} \sum_{m=1}^M S_m \log_e |\hat{\boldsymbol{\Omega}}_m| + \log_e(\log_e(T)) \dim(\boldsymbol{\Lambda}).
 \end{aligned}$$

To conclude the section, we conduct a simulation experiment to evaluate the performance of penalized information criteria (AIC, BIC and HQC) in detecting nonlinearity of some processes. We simulate three different data generating processes (DGP): a linear one, a 2-state process (an MS(2)-AR(1)) and a 3-state process (an MS(3)-AR(1)), whose parameters are reported in Table 1, and Gaussian i.i.d. errors. The experiments simulate artificial time series of length $T + 50$ with $T = \{50, 100, 250, 500\}$; the first 50 initial data points are discarded to minimize the effect of initial conditions. For each experiment, we evaluate the performance of the methods over 3,000 replications. Results are shown in Table 2 for AIC, BIC and HQC. All the criteria correctly detect the data generating processes and their precision increases with the length of the time series. Moreover, our specification of the criteria are convenient and less time-consuming since they have a rather simple form and avoid recursive likelihood computations.

5 The Italian Macroeconomy

In this section we employ the proposed methods to investigate the presence of nonlinearities in the Italian economy. In order to do this, we consider five Italian time series taken from Datastream: GDP, private consumption, gross investments, exports

Table 2 Percentage of correct detection of the regime number over 3000 replications

DGP	IC	$T = 50$	$T = 100$	$T = 250$	$T = 500$
AR(1)	<i>AIC</i>	100	100	100	100
	<i>BIC</i>	100	100	100	100
	<i>HQC</i>	100	100	100	100
MS(2)-AR(1)	<i>AIC</i>	86	92	93	100
	<i>BIC</i>	96	98	99	100
	<i>HQC</i>	91	96	99	100
MS(3)-AR(1)	<i>AIC</i>	61	73	84	91
	<i>BIC</i>	69	74	88	96
	<i>HQC</i>	65	71	87	95

Penalized information criteria: Akaike information criterion (AIC), Bayesian information criterion (BIC) and Hannan-Quinn criterion (HQC) as defined in Sect. 4. The DGPs are described in Table 1

Table 3 Model specification

Series	AIC	BIC	HQC	ν_{s_t}	ϕ_{s_t}	σ_{s_t}	Lags
GDP	3	2	2	✓	✓	✓	1
Consumption	2	2	2	✗	✗	✓	2
Investment	2	2	2	✓	✓	✓	3
Imports	3	3	3	✓	✓	✓	1
Exports	3	3	3	✓	✓	✓	1

Testing the number of regimes by penalized likelihood criteria (AIC, BIC and HQC) (columns 2, 3 and 4), summary of LR tests for stability of the model parameters (columns 5, 6, 7) and choice of the number of lags (column 8) for Italian macroeconomic time series from 1980:Q1 to 2014:Q2 (Source: Datastream)

and imports of goods and services. The series are taken at 2010 constant prices and at quarterly frequency for the period from 1980:Q1 to 2014:Q2. The log-series are made stationary by taken the first difference. Penalized information criteria indicate substantial evidence of nonlinearities for Italian macroeconomic data as shown in Table 3 (columns 2, 3, 4). In particular, two regimes are appropriate for GDP, consumption and investments while three regimes are more suitable for imports and exports. In fact, the latter variables are very likely influenced by cyclical movements outside the domestic system. Then we use LR tests in the form presented in Sect. 3 to select the appropriate parametric model in terms of parameter stability and number of lags before conducting estimation. A graphical overview of the chosen switching parameters is given in Table 3 (columns 5, 6, 7) together with the number of lags chosen in the autoregressive parameter—being switching or not (column 8). It can be seen that all the series require switching parameters, except for the consumption which is more stable and only retains some variability in the uncertainty parameter. After this preliminary analysis, we can now proceed with the estimation. However, for the purpose of this work we only focus on some of the series. Firstly, we analyze the dynamic of Italian GDP whose estimates are reported in the upper part of Table 4. The chosen parametric model for GDP captures in an efficient way the main facts of the business cycle. In fact, expansion and recession are respectively characterized by high

Table 4 Model estimation

	ν	ϕ	σ	p	Exp.Dur.
GDP					
Regime 1	0.08 (0.29)	0.34 (0.20)	1.62 (0.21)	0.75 (0.08)	4.05
Regime 2	0.15 (0.07)	0.52 (0.12)	0.4 (0.09)	0.95 (0.03)	18.43
Export					
Regime 1	-0.11 (0.07)	-0.14 (0.11)	0.62 (0.09)	0.10 (0.22)	1.01
Regime 2	0.08 (0.04)	0.20 (0.13)	0.21 (0.11)	0.75 (0.09)	4.06
Regime 3	0.12 (0.08)	0.23 (0.15)	0.12 (0.05)	0.92 (0.03)	13.20

Parameters estimates, regime probabilities (standard errors in parentheses) and expected durations for Italian GDP and exports (goods and services) series modelled as in (1) and Table 3 for the period from 1980:Q1 to 2014:Q2 (Source: Datastream)

Fig. 1 Plot of the smoothed probabilities corresponding to the recessionary regime for Italian GDP series in the period from 1980:Q1 to 2014:Q2 (Source: Datastream)

and low, but positive, growth with higher variability in the recessionary regime. Probabilities show asymmetric features with expansions being more persistent and high expected duration with respect to recessions. Smoothed probabilities corresponding to the recessionary regime are reported in Fig. 1 where it can be recognized the 82–83 crisis, the 92–93 and 97–99 currency crisis, and the last 2008–2011 financial crisis.

It is now interesting to study the dynamic of exports. The estimation reveals the existence of a first regime of severe contraction, a second regime of recession and a third regime of growth. The severe contractionary regime shows negative growth, higher uncertainty and it is unlikely to happen. If we look at the plot of the smoothed probabilities in the first regime (see Fig. 2, upper panel), it identifies two main severe crisis in the considered sample: over the 1990s (economic downturn affecting much of the world in the late 1980s and early 1990s) and the 2008–2010 global financial crisis. Moreover, the plot of the smoothed probabilities in the second regime (see Fig.

Fig. 2 Plot of the smoothed probabilities corresponding to the strong contractionary regime (*upper panel*) and the recessionary regime (*lower panel*) for Italian exports (goods and services) series in the period from 1980:Q1 to 2014:Q2 (Source: Datastream)

2, lower panel) exhibits more peaks. This clearly reveals that exports are also much influenced by external economic shocks besides the domestic ones. However, the estimated parameters for recession (regime 2) and expansion (regime 3) show the same features as described before. This confirms that the preliminary analysis based on IC and LR tests is useful to correctly identify the parametric model in order to conduct proper estimation and inference. Finally, it might be interesting to estimate a bivariate MS(2)-VAR(1) model for the bivariate model of output growth and exports growth to analyse the evolution of the domestic performance. Define y_t to be the first difference in logs of the real GDP, and e_t the first difference in logs of the exports of good and services. The estimated bivariate model looks as follows (the estimation procedure used here is that proposed in [Cavicchioli 2014b](#)):

$$\begin{aligned}
\text{State 1: } \begin{bmatrix} y_t \\ e_t \end{bmatrix} &= \begin{bmatrix} 0.5796 \\ (0.1126) \\ 0.2650 \\ (0.6151) \end{bmatrix} + \begin{bmatrix} 0.8319 & 0.0142 \\ (0.0033) & (0.0012) \\ 1.3105 & 0.4184 \\ (0.0147) & (0.0052) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ e_{t-1} \end{bmatrix} + \mathbf{v}_{1t}, \\
\Omega_1 &= \begin{bmatrix} 0.2125 & 0.4503 \\ (0.0113) & (0.051) \\ 0.4503 & 4.3172 \\ (0.051) & (0.23) \end{bmatrix} \\
\text{State 2: } \begin{bmatrix} y_t \\ e_t \end{bmatrix} &= \begin{bmatrix} -0.0122 \\ (0.0447) \\ 0.1733 \\ (0.2013) \end{bmatrix} + \begin{bmatrix} -0.2415 & 0.0729 \\ (0.0052) & (0.0016) \\ 1.7584 & -0.3565 \\ (0.0287) & (0.0086) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ e_{t-1} \end{bmatrix} + \mathbf{v}_{2t}, \\
\Omega_2 &= \begin{bmatrix} 0.3857 & 0.1364 \\ (0.011) & (0.06) \\ 0.1364 & 11.5199 \\ (0.06) & (0.328) \end{bmatrix}
\end{aligned}$$

where $\mathbf{v}_{s_t,t} \sim NID(\mathbf{0}, \Omega_{s_t})$ and $s_t \in \{1, 2\}$ (standard errors are given in parentheses). Notice that the variance-covariance matrices Ω_1 and Ω_2 are positive definite. Let $\Phi_1 = \begin{bmatrix} 0.8319 & 0.0142 \\ 1.3105 & 0.4184 \end{bmatrix}$ and $\Phi_2 = \begin{bmatrix} -0.2415 & 0.0729 \\ 1.7584 & -0.3565 \end{bmatrix}$ be the matrices of the AR part in State 1 and State 2, respectively. In order to investigate the stationarity properties of the bivariate MS process $(y_t \ e_t)'$, we use the following 8×8 matrix

$$\mathbb{P} = \begin{bmatrix} p_{11}\{\Phi_1 \otimes \Phi_1\} & p_{21}\{\Phi_1 \otimes \Phi_1\} \\ p_{12}\{\Phi_2 \otimes \Phi_2\} & p_{22}\{\Phi_2 \otimes \Phi_2\} \end{bmatrix}$$

where $p_{12} = 1 - p_{11}$, $p_{21} = 1 - p_{22}$, $p_{11} = 0.84$ and $p_{22} = 0.7$. Thus we could identify State 1 as an expansionary regime and State 2 as a recessionary regime. Since the spectral radius of $\mathbb{P}$ is equal to 0.61, we can apply Theorem 2 from [Francq and Zakoian \(2001\)](#). Then the bivariate MS process is second-order stationary. Moreover, except for the intercept in State 2, all the coefficients are significant. Now, with regards to the autoregressive matrices, it can be noted that during normal times the output growth shows a persistent dynamic while exports are less persistent. During difficult times, instead, both variables have a tendency of reversion to the mean but the contribution of one variables to the other is higher than in normal times. This suggests that export can be a major driver of the output growth especially during recessionary times. Finally, the variability of the system is higher in difficult times (see the variance-covariance matrix in State 2) than in normal times. Of course, we are aware that there is a wide body of literature analyzing the theoretical links between exports and economic growth and several different channels have been suggested through which exports affect economic growth (see, for instance, [Pereira and Xu 2000](#)). However, the purpose of this work is not to interpret structurally the model, but to differentiate the co-movement of these two variables in different phases of the business cycle, showing the goodness of the method and suggesting possible ideas for further research.

6 Conclusion

In this paper we provide explicit and simple formulae for LR tests and penalized information criteria in the context of Markov switching vector autoregressive (MS VAR) models. From a practical point of view, this improves the computational process as well as model selection. Indeed, our matrix formulae can be easily implemented for computational estimation purpose avoiding the use of the recursive algorithm, such as EM algorithm. Since LR tests can be used only when the number of regimes does not change, we also specify a new form of penalized information criteria (IC) which avoid the computational burden of recursive likelihood calculations. Their performance is evaluated with a simulated experiment. These procedures are applied for the analysis of Italian macroeconomic data. We find asymmetries in the way the economy behaves at different stages of the business cycle. For reliable results and dating of the cycle, it turns out to be critical the preliminary step where the proposed tests are implemented. Thus these procedures might be useful in real applications that need to take into account nonlinearities in the data.

Acknowledgments I would like to thank the Editor-in-Chief Professor Roberto Cellini, the Managing Editor Professor Michele Polo and the anonymous referee for their useful suggestions and comments which were most valuable in writing the final version of the paper. Moreover, many thanks are due to the Italian Economic Association for awarding me with the 2014 best PhD thesis. This was a great honor for me. Lastly, but not least important, I wish to express my gratitude to Professor Monica Billio for having supervised my PhD work during the doctoral years at the University Cà Foscari in Venice.

Appendix

Derivation of Formula (8).

$$\begin{aligned}
 & \frac{1}{2} \sum_{t=1}^T \sum_{m=1}^M \hat{\mathbf{u}}_t' \hat{\Sigma}_m' \hat{\Omega}_m^{-1} \hat{\Sigma}_m \hat{\mathbf{u}}_t \hat{\xi}_{mt|T} \\
 &= \frac{1}{2} \text{trace} \left[\sum_{t=1}^T \sum_{m=1}^M \hat{\mathbf{u}}_t' \hat{\Sigma}_m' \hat{\Omega}_m^{-1} \hat{\Sigma}_m \hat{\mathbf{u}}_t \hat{\xi}_{mt|T} \right] \\
 &= \frac{1}{2} \text{trace} \left[\sum_{t=1}^T \sum_{m=1}^M \hat{\Omega}_m^{-1} \hat{\Sigma}_m \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \hat{\Sigma}_m' \hat{\xi}_{mt|T} \right] \\
 &= \frac{1}{2} \text{trace} \left[\sum_{m=1}^M S_m \hat{\Omega}_m^{-1} S_m^{-1} \sum_{t=1}^T \hat{\Sigma}_m \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \hat{\Sigma}_m' \hat{\xi}_{mt|T} \right] \\
 &= \frac{1}{2} \text{trace} \left[\sum_{m=1}^M S_m \hat{\Omega}_m^{-1} \hat{\Omega}_m \right] = \frac{1}{2} \text{trace} \left[\sum_{m=1}^M S_m \mathbf{I}_K \right] \\
 &= \frac{1}{2} \text{trace} \left[\sum_{m=1}^M \sum_{t=1}^T \hat{\xi}_{mt|T} \mathbf{I}_K \right] = \frac{1}{2} \text{trace} \left[\sum_{t=1}^T \sum_{m=1}^M \hat{\xi}_{mt|T} \mathbf{I}_K \right] = \frac{TK}{2}.
 \end{aligned}$$

Derivation of Formula (11).

$$\begin{aligned}
LR &= 2(\mathcal{L}_1^* - \mathcal{L}_0^*) = \left[\sum_{m=1}^M S_m^{(0)} \right] \log_e |\hat{\boldsymbol{\Omega}}^{(0)}| - \sum_{m=1}^M S_m^{(1)} \log_e |\hat{\boldsymbol{\Omega}}_m^{(1)}| \\
&= \left[\sum_{t=1}^T \sum_{m=1}^M \hat{\xi}_{mt|T}^{(0)} \right] \log_e |\hat{\boldsymbol{\Omega}}^{(0)}| - \sum_{m=1}^M S_m^{(1)} \log_e |\hat{\boldsymbol{\Omega}}_m^{(1)}| \\
&= T \log_e |\hat{\boldsymbol{\Omega}}^{(0)}| - \sum_{m=1}^M S_m^{(1)} \log_e |\hat{\boldsymbol{\Omega}}_m^{(1)}| \\
&= \log_e \left(\frac{|\hat{\boldsymbol{\Omega}}^{(0)}|^T}{\prod_{m=1}^M |\hat{\boldsymbol{\Omega}}_m^{(1)}|^{S_m^{(1)}}} \right).
\end{aligned}$$

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Proceedings of the 2nd International Symposium on Information Theory. Academiai Kiado, Budapest, p 267–281 (reprinted In: Kotz S, Johnson NL (eds) Breakthroughs in Statistics 1. Springer, New York, p 610–624, 1992)
- Akaike H (1974) A new look at the statistical model identification. IEEE Trans Auto Control 19:716–723
- Bickel P, Ritov Y, Ryden T (1998) Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. Ann Stat 26:1614–1635
- Carrasco M, Hu L, Ploberger W (2014) Optimal test for markov switching parameters. Econometrica 82(2):765–784
- Carter AV, Steigerwald DG (2012) Testing for regime switching: a comment. Econometrica 80(4):1809–1812
- Cavicchioli M (2014a) Determining the number of regimes in Markov-switching VAR and VMA models. J Time Ser Anal 35(2):173–186
- Cavicchioli M (2014b) Analysis of the likelihood function for Markov-switching VAR(CH) models. J Time Ser Anal 35(6):624–639
- Cho JS, White H (2007) Testing for regime switching. Econometrica 75:1671–1720
- Di Sanzo S (2009) Testing for linearity in Markov switching models: a bootstrap approach. Stat Methods Appl 18(2):153–168
- Franco C, Zakoian JM (2001) Stationarity of multivariate Markov-switching ARMA models. J Econom 102:339–364
- Garcia R (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models. Int Econ Rev 39(3):763–788
- Hamilton JD (1990) Analysis of time series subject to changes in regime. J Econom 45:39–70
- Hamilton JD (1993) Estimation, inference and forecasting of time series subject to changes in regime. In: Maddala GS, Rao CR, Vinod HD (eds.) Handbook of Statistics, vol 11. Elsevier, Amsterdam
- Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton
- Hamilton JD (1996) Specification testing in markov-switching time-series models. J Econom 70:127–157
- Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. J Royal Stat Soc B 41:190–195
- Hansen BE (1992) The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. J Appl Econom 7(S1):S61–S82
- Huang YL (2014) Testing Markov switching models. Appl Econ 46(17):2047–2051
- Krolzig HM (1997) Markov-switching vector autoregressions: modelling. Statistical inference and application to business cycle analysis. Springer, Berlin-Heidelberg-New York
- Lütkepohl H (1991) Introduction to multiple time series analysis. Springer, Berlin-Heidelberg-New York

- Olteanu M, 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)
- Pereira AM, Xu Z (2000) Export growth and domestic performance. *Rev Int Econ* 8(1):60–73
- Psaradakis Z, Spagnolo N (2003) On the determination of the number of regimes in Markov-switching autoregressive models. *J Time Ser Anal* 24(2):237–252
- Psaradakis Z, Spagnolo N (2006) Joint determination of the state dimension and autoregressive order for models with Markov Regime Switching. *J Time Ser Anal* 27:753–766
- Ríos R, Rodríguez LA (2008) Penalized estimate of the number of states in Gaussian linear AR with Markov regime. *Electron J Stat* 2:1111–1128
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Zhang J, Stine RA (2001) Autocovariance structure of Markov regime switching models and model selection. *J Time Ser Anal* 22:107–124