Journal of Multivariate Analysis 157 (2017) 124–135

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotic Fisher information matrix of Markov switching VARMA models

Maddalena Cavicchioli *

Dipartimento di Economia "Marco Biagi", Università degli Studi di Modena e Reggio Emilia, Viale Berengario 51, 41121 Modena, Italy
Dipartimento di Scienze Economiche, Università degli Studi di Verona, Via Cantarane 24, 37129 Verona, Italy

ARTICLE INFO

Article history:
Received 23 February 2016
Available online 23 March 2017

AMS subject classifications:
62B10
62F12
62M02
62M10

Keywords:
Time series with changes in regime
Information matrix computation
Markov switching VARMA
Asymptotic covariance matrix
Gaussian maximum likelihood estimator

ABSTRACT

We study the Fisher information (FI) matrix of Markov switching vector autoregressive moving average (MS VARMA) models and derive an explicit expression in closed form for the asymptotic FI matrix of the underlying model. Our result is more general than the available one in the literature for linear VARMA models, which has been recently studied in Bao and Hua (2014), in two respects. First, we treat the variance of the error term in a more general setting rather than considering it as a nuisance parameter. Then, we consider non-trivial intercept in the MS VARMA model. Under general conditions, the asymptotic FI matrix can be used to derive the asymptotic covariance matrix of the Gaussian maximum likelihood estimator of the model parameters. Some examples and numerical applications illustrate the results.

1. Introduction and motivations

Markov switching vector autoregressive moving average (MS VARMA) models have been extensively used in statistics and economics to model nonlinear multivariate time series. For information concerning the stationarity, estimation, and consistency of such models, along with other statistical inference procedures, we refer to [2,5,8,10,11,23,29]. See also [3], where explicit matrix expressions for the maximum likelihood (ML) estimator of the parameters in MS VAR(CH) models have been derived, and [27] for a recent application of structural vector autoregressive models with Markov switching to the financial crisis. Higher-order moments of MS VARMA models are provided in [4].

As stated in [23, Section 6.6.2], it is often impractical to evaluate analytically the asymptotic covariance matrix of a generic MS VAR model. In this paper we study the Fisher information (FI) matrix of MS VARMA models and derive a closed form expression for the asymptotic FI matrix by using appropriate techniques from matrix calculus. Invoking standard asymptotic theory (see, e.g., [11, Section 21], [23, Section 6.6.2], and [24, Section C.4]), one can use the inverse of such a matrix as the asymptotic covariance matrix of the Gaussian ML estimator of the model parameters.

Our result is more general than what is available in the literature for linear VARMA models, which was recently studied in [1], in two respects. First, we treat the variance of the error term in a more general setting rather than considering it as a nuisance parameter. Then, we consider the case of a non-trivial intercept in the MS VARMA model.

* Correspondence to: Dipartimento di Economia "Marco Biagi", Università degli Studi di Modena e Reggio Emilia, Viale Berengario 51, 41121 Modena, Italy.

E-mail address: maddalena.cavicchioli@unimore.it.

http://dx.doi.org/10.1016/j.jmva.2017.03.004
0047-259X/

Despite the increasing interest in multivariate Markov switching processes, an explicit matrix expression for the asymptotic covariance matrix of the ML estimators of MS VARMA models has heretofore not been provided, at least so far as we know. Asymptotic properties of the ML estimator for MS VAR models have been studied in [2,5], even for continuous-valued switching processes, but the expression of the limiting covariance of the ML estimator has a rather complicated form; see the proof of Lemma 1 in [2] and Theorem 3 in [5]. An algorithm was provided in [26] for computing the asymptotic FI matrix of a VARMA process at the scalar-level. It is based on a frequency domain representation of the FI matrix, known as Whittle's formula. That approach has been generalized and put in matrix-level form in a series of papers due to Klein, Spreij and Mélard who have been working on algorithms for inverting asymptotic and exact FI matrices for many years. See [16,18,20,21] for the asymptotic and exact FI matrices of a VARMA process. These authors then extend the proposed methods to the class of stationary vector autoregressive moving average processes with exogenous or input variables, called VARMAX models. See [17,22] on the inversion of the exact FI matrix for VARMAX processes. The algorithm described in the latter paper is composed of Chandrasekhar recursion equations at a vector–matrix level, and some of these recursions consist of derivatives based on appropriate differential rules applied to a state space representation of the underlying model. For the asymptotic properties of the ML estimator of the coefficients of VARMAX models, see [12]. Further recursive filtering methods for computing asymptotic and exact FI matrices for a Gaussian VARMA model expressed in state space form have been developed in [19,30]. Finally, [28] studies the FI matrix estimation in generalized linear mixed models.

Our main contributions are as follows. First, we use recursion equations at a vector–matrix level instead of writing recursions for each element of the FI matrix. For this purpose, we apply matrix differential rules from [7,25], and use the recursive filter and smoother algorithms for computing the full-sample conditioned regime probabilities as described in [23, Section 5.2]. We also consider the underlying model in wider generality through changes in regime and non-trivial intercept, and the variance of the error term is treated as an unknown parameter. Moreover, our matrix expressions improve computational performance since they are readily programmable and greatly reduce the computational cost.

The matrix formulas we obtain are useful for statistical inference of MS VARMA models. To show their applicability, we consider a testing hypothesis problem (e.g., Wald test) and perform the test using the asymptotic covariance matrix of the ML estimator of the unknown parameters involved in a MS VARMA model. Also, via a numerical study we use the results to construct asymptotic confidence intervals for the unknown parameters of a certain MS VARMA model.

The paper is organized as follows. In Section 2 we introduce the model, give some preliminaries and notations, and formulate the main result concerning the closed form expression for the asymptotic FI matrix of an MS VARMA model. The proof of the main theorem is given in Appendix A. Section 3 provides some examples and Section 4 illustrates the computation of the asymptotic FI matrix via numerical simulation (the obtained expressions are given in Appendix B). We also consider a testing hypothesis problem and the construction of asymptotic confidence intervals for the unknown parameters of a 2-state bivariate ARMA(1, 1) model. Section 5 concludes with remarks. For the basic identities and results on matrix calculus the reader is referred to [7,25].

2. Main result

Consider the M-state Markov switching K-dimensional ARMA(p, q) model

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

where $\mathbf{y}_t$ is a random vector with values in $\mathbb{R}^K$, $\boldsymbol{\nu}_{s_t}$ is a $K \times 1$ real state-dependent vector, and $\mathbf{\Phi}_{s_t,i}$ and $\mathbf{\Theta}_{s_t,j}$ are $K \times K$ real state-dependent matrices ($\mathbf{\Phi}_{s_t,p} \neq \mathbf{0}$ and $\mathbf{\Theta}_{s_t,q} \neq \mathbf{0}$, where $\mathbf{0}$ denotes the null matrix).

Assumption 1. The process (s_t) is an irreducible, aperiodic and ergodic Markov chain with finite space $\Xi = \{1, \ldots, M\}$, stationary transition probabilities $p_{ij} = \Pr(s_t = j|s_{t-1} = i)$ and unconditional (or steady-state) probabilities $\pi_i = \Pr(s_t = i)$ for all $i, j \in \{1, \ldots, M\}$.

Collect the transition probabilities p_{ij} in an $M \times M$ matrix $\mathbf{P}$, known as the transition probability matrix. To allow for the possibility of change in variance, we assume that $\mathbf{u}_t = \mathbf{\Sigma}_{s_t}\boldsymbol{\epsilon}_t$, where $\mathbf{\Sigma}_{s_t}$ is a $K \times K$ real state-dependent matrix and $(\boldsymbol{\epsilon}_t)$ is a stationary and ergodic sequence of K-dimensional centered and uncorrelated variables with $\mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top) = \mathbf{I}_K$ (here $\mathbf{I}_K$ is the $K \times K$ identity matrix).

Assumption 2. The innovation $(\boldsymbol{\epsilon}_t)$ is independent of (s_t), and $(\mathbf{y}_t)$ is second-order stationary, i.e., it satisfies Theorem 2 from [8].

Since $(s_t, \boldsymbol{\epsilon}_t)$ is an ergodic process, the MS VARMA is also ergodic. So, for instance, the ergodic theorem could be used to obtain the asymptotic variance of the ML estimator. Using the lag operator L, such that $L^j\mathbf{y}_t = \mathbf{y}_{t-j}$, model (1) can be written as

$$\mathbf{\Phi}_{s_t}(L)\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \mathbf{\Theta}_{s_t}(L)\mathbf{u}_t, \tag{2}$$

where $\mathbf{\Phi}_{s_t}(L) = \mathbf{I}_K - \sum_{i=1}^{p} \mathbf{\Phi}_{s_t,i}L^i$ and $\mathbf{\Theta}_{s_t}(L) = \mathbf{I}_K - \sum_{j=1}^{q} \mathbf{\Theta}_{s_t,j}L^j$. Let $\mathbf{\Phi}_m(L)$ and $\mathbf{\Theta}_m(L)$ be the matrix lag polynomials obtained by replacing s_t by m in the definition of $\mathbf{\Phi}_{s_t}(L)$ and $\mathbf{\Theta}_{s_t}(L)$, respectively.

Assumption 3. Given $s_t = m$, $\mathbf{\Phi}_m(L)$ and $\mathbf{\Theta}_m(L)$ have no common roots, and model (2) is casual and invertible, for all $m \in \{1, \ldots, M\}$, i.e., $\det \mathbf{\Phi}_m(z) \neq 0$ and $\det \mathbf{\Theta}_m(z) \neq 0$, for all $z \in \mathbb{C}$ such that $|z| \leq 1 + \epsilon$, for some $\epsilon > 0$.

Notice that the above model is far more general than a linear VARMA model, as recently studied in [1]. First, we allow changes in regime and the linear model is just a particular case. Second, we treat the variance of the error term in a more general setting rather than considering it as a nuisance parameter. Finally, we consider the case of non-trivial intercept.

Now we reparametrize model (2) by using absolutely summable sequences of matrices. Then we have

$$\mathbf{u}_t = \mathbf{A}_m(L)\mathbf{y}_t - \mathbf{\Theta}_m^{-1}(1)\boldsymbol{\nu}_m, \qquad \mathbf{y}_t = \mathbf{B}_m(L)\mathbf{u}_t + \mathbf{\Phi}_m^{-1}(1)\boldsymbol{\nu}_m, \tag{3}$$

where $\mathbf{A}_m(L) = \mathbf{\Theta}_m^{-1}(L)\mathbf{\Phi}_m(L)$ and $\mathbf{B}_m(L) = \mathbf{\Phi}_m^{-1}(L)\mathbf{\Theta}_m(L)$, for all $m \in \{1, \ldots, M\}$. The matrix functions $\mathbf{A}_m(z)$ and $\mathbf{B}_m(z)$ can be determined by expanding $\mathbf{\Theta}_m^{-1}(z)\mathbf{\Phi}_m(z)$ and $\mathbf{\Phi}_m^{-1}(z)\mathbf{\Theta}_m(z)$ into power series over some open region containing the complex unit disk and equating the matrix coefficients. We can set $\mathbf{A}_m(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{m\ell}L^\ell$ and $\mathbf{B}_m(L) = \sum_{\ell=0}^{\infty} \mathbf{B}_{m\ell}L^\ell$, where the matrix sequences $(\mathbf{A}_{m\ell})_\ell$ and $(\mathbf{B}_{m\ell})_\ell$ are absolutely (and also square) summable in a component-by-component sense; see, e.g., [14]. We also write

$$\mathbf{\Theta}_m^{-1}(L) = \sum_{\ell=0}^{\infty} \mathbf{\Psi}_{m\ell}L^\ell \tag{4}$$

for an absolutely summable sequence $(\mathbf{\Psi}_{m\ell})_\ell$. Following [11, Section 10], the matrices $\mathbf{\Psi}_{m\ell}$ can be evaluated by requiring

$$(\mathbf{I}_K - \mathbf{\Theta}_{m1}L^1 - \mathbf{\Theta}_{m2}L^2 - \cdots - \mathbf{\Theta}_{mq}L^q)(\mathbf{I}_K + \mathbf{\Psi}_{m1}L^1 + \mathbf{\Psi}_{m2}L^2 + \cdots) = \mathbf{I}_K.$$

Setting the coefficient of L^ℓ equal to zero produces

$$\mathbf{\Psi}_{m\ell} = \mathbf{\Theta}_{m1}\mathbf{\Psi}_{m,\ell-1} + \mathbf{\Theta}_{m2}\mathbf{\Psi}_{m,\ell-2} + \cdots + \mathbf{\Theta}_{mq}\mathbf{\Psi}_{m,\ell-q} \tag{5}$$

for all $\ell \in \{1, 2, \ldots\}$ with $\mathbf{\Psi}_{m\ell} = \mathbf{0}$ for $\ell < 0$ and $\mathbf{\Psi}_{m0} = \mathbf{I}_K$.

Denote, as usual, by $\mathbf{\Gamma}_{\mathbf{y}}(h) = \mathrm{cov}(\mathbf{y}_t, \mathbf{y}_{t-h})$ the autocovariance function of the process $\mathbf{y} = (\mathbf{y}_t)$ driven by (1). An explicit computation of $\mathbf{\Gamma}_{\mathbf{y}}(h)$ was given in [8, Section 4]. Define

$$\begin{aligned} \boldsymbol{\gamma}_m &= \left((\mathrm{vec}\,\mathbf{\Phi}_{m1})^\top \cdots (\mathrm{vec}\,\mathbf{\Phi}_{mp})^\top\right)^\top, & \boldsymbol{\sigma}_m &= \mathrm{vec}\,\mathbf{\Omega}_m^{-1}, \\ \boldsymbol{\delta}_m &= \left((\mathrm{vec}\,\mathbf{\Theta}_{m1})^\top \cdots (\mathrm{vec}\,\mathbf{\Theta}_{mq})^\top\right)^\top, & \boldsymbol{\alpha}_m &= \left(\boldsymbol{\nu}_m^\top\ \boldsymbol{\gamma}_m^\top\ \boldsymbol{\delta}_m^\top\right)^\top, \end{aligned}$$

where $\mathbf{\Omega}_m = \mathbf{\Sigma}_m\mathbf{\Sigma}_m^\top$, for all $m \in \{1, \ldots, M\}$. Then $\boldsymbol{\gamma}_m$ is $(pK^2) \times 1$, $\boldsymbol{\delta}_m$ is $(qK^2) \times 1$, $\boldsymbol{\sigma}_m$ is $K^2 \times 1$, and $\boldsymbol{\alpha}_m$ is $b \times 1$, where $b = K + (p+q)K^2$. Set $\boldsymbol{\theta}_m = (\boldsymbol{\alpha}_m^\top\ \boldsymbol{\sigma}_m^\top)^\top$. Collect $\boldsymbol{\theta}_m$ into a vector $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top \cdots \boldsymbol{\theta}_M^\top)^\top$. Let $\boldsymbol{\rho} = (p_{11} \cdots p_{1M} \cdots p_{M1} \cdots p_{MM})^\top$ be the $M^2 \times 1$ vector of transition probabilities. Then $\boldsymbol{\lambda} = (\boldsymbol{\theta}^\top\ \boldsymbol{\rho}^\top)^\top$ is the parameter vector of model (1). Under quite general regularity conditions (such as identifiability, stability and the fact that the true parameter vector does not fall on the boundaries, which we assume here), a ML estimator $\hat{\boldsymbol{\lambda}}$ for $\boldsymbol{\lambda}$ is consistent and asymptotically normal; see [2,5], [23, Section 6.6.2], [24, Section C.4]. Then $\sqrt{T}\,(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathcal{F}_a^{-1}(\boldsymbol{\lambda}))$, where $\mathcal{F}_a(\boldsymbol{\lambda})$ is the asymptotic FI matrix.

The Gaussian conditional density of $(\mathbf{y}_t)$ in (1) given $s_t = m$ and $\mathbf{Y}_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots\}$ is

$$\eta_{m,t}(\boldsymbol{\theta}) = p(\mathbf{y}_t | s_t = m, \mathbf{Y}_{t-1}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{K/2}|\mathbf{\Omega}_m|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{u}_t^\top \mathbf{\Omega}_m^{-1}\mathbf{u}_t\right).$$

As proved in [23, Eq. (6.8)], the derivative of the log-likelihood function $\mathcal{L}(\boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$ for a K-dimensional time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ of length T (sample size), generated by model (1), is

$$\frac{\partial \mathcal{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T}\sum_{m=1}^{M} \frac{\partial \ln \eta_{mt}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\, \xi_{mt|T},$$

where

$$\ln \eta_{mt}(\boldsymbol{\theta}) = -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}_m| - \frac{1}{2}\mathbf{u}_t^\top \mathbf{\Omega}_m^{-1}\mathbf{u}_t.$$

Here $\xi_{mt|T} = \mathrm{E}(\xi_{mt}|\mathbf{Y}_T)$ are the smoothed regime probabilities and $\mathbf{Y}_T = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$ is the information set. Furthermore, $\boldsymbol{\xi}_t = (\xi_{1t} \cdots \xi_{Mt})^\top$ is the random $M \times 1$ vector whose mth element is equal to unity if $s_t = m$ and zero otherwise. Note that $\mathrm{E}(\boldsymbol{\xi}_t) = \mathrm{E}(\boldsymbol{\xi}_{t|T}) = \boldsymbol{\pi} = (\pi_1 \cdots \pi_M)^\top$, where $\boldsymbol{\xi}_{t|T} = (\xi_{1t|T} \cdots \xi_{Mt|T})^\top$. A fast recursive algorithm to evaluate $\xi_{mt|T}$ was described in [23, Section 5]. See also [11, Eqs. (22.4.5), (22.4.6) and (22.4.14)]. 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 $\boldsymbol{\xi}_{t|t} = \mathrm{E}(\boldsymbol{\xi}_t|\mathbf{Y}_t)$ can be computed by iterating on the following pair of recursive formulas

$$\boldsymbol{\xi}_{t|t} = \frac{\boldsymbol{\eta}_t \odot \boldsymbol{\xi}_{t|t-1}}{\boldsymbol{\eta}_t^\top \boldsymbol{\xi}_{t|t-1}}, \qquad \boldsymbol{\xi}_{t+1|t} = \mathbf{P}^\top \boldsymbol{\xi}_{t|t},$$

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

$$\boldsymbol{\xi}_{t|T} = \{\mathbf{P}(\boldsymbol{\xi}_{t+1|T}(\div)\boldsymbol{\xi}_{t+1|t})\} \odot \boldsymbol{\xi}_{t|t},$$

where the symbol $(\div)$ denotes the element-by-element division. For model (1), set

$$\mathbf{x}_t = (\mathbf{y}_{t-1}^\top \cdots \mathbf{y}_{t-p}^\top)^\top \in \mathbb{R}^{pK}, \qquad \mathbf{w}_{mt} = (\mathbf{z}_{m,t-1}^\top \cdots \mathbf{z}_{m,t-q}^\top)^\top \in \mathbb{R}^{qK}$$

where $\mathbf{z}_{m,t-j} = \boldsymbol{\Phi}_m(L)\mathbf{y}_{t-j} - \boldsymbol{\nu}_m$ for all $j \in \{1, \ldots, q\}$ and $m =\in \{1, \ldots, M\}$. Then we have the following result, whose proof is given in Appendix A.

Main theorem. *Under Assumptions 1–3, the partitioned form of the mth asymptotic Fisher information matrix $\mathcal{F}_a(\boldsymbol{\theta}_m)$ for the* MS VARMA *model (1) has the following blocks pertaining to $\boldsymbol{\alpha}_m$ and $\boldsymbol{\sigma}_m$:*

$$\mathcal{F}_a(\boldsymbol{\theta}_m) = \begin{pmatrix} \mathcal{F}_a(\boldsymbol{\alpha}_m) & \mathbf{0}^\top \\ \mathbf{0} & \mathcal{F}_a(\boldsymbol{\sigma}_m) \end{pmatrix},$$

where

$$\mathcal{F}_a(\boldsymbol{\sigma}_m) = \frac{\pi_m}{2}(\boldsymbol{\Omega}_m \otimes \boldsymbol{\Omega}_m). \tag{6}$$

The block matrix $\mathcal{F}_a(\boldsymbol{\alpha}_m)$ is

$$\mathcal{F}_a(\boldsymbol{\alpha}_m) = \begin{pmatrix} \mathcal{F}_a(\boldsymbol{\nu}_m) & \mathcal{F}_a(\boldsymbol{\gamma}_m, \boldsymbol{\nu}_m)^\top & \mathbf{0}^\top \\ \mathcal{F}_a(\boldsymbol{\gamma}_m, \boldsymbol{\nu}_m) & \mathcal{F}_a(\boldsymbol{\gamma}_m) & \mathcal{F}_a(\boldsymbol{\delta}_m, \boldsymbol{\gamma}_m)^\top \\ \mathbf{0} & \mathcal{F}_a(\boldsymbol{\delta}_m, \boldsymbol{\gamma}_m) & \mathcal{F}_a(\boldsymbol{\delta}_m) \end{pmatrix}.$$

The diagonal blocks are

$$\mathcal{F}_a(\boldsymbol{\nu}_m) = \pi_m \boldsymbol{\Theta}_m^{-1}(1)^\top \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Theta}_m^{-1}(1), \tag{7}$$

$$\mathcal{F}_a(\boldsymbol{\gamma}_m) = \sum_{\ell,n\geq 0}^{\infty} (\boldsymbol{\Psi}_{m\ell}^\top \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Psi}_{mn}) \otimes \mathbf{Q}_m(\ell, n), \tag{8}$$

$$\mathcal{F}_a(\boldsymbol{\delta}_m) = \sum_{\ell,n,h,k\geq 0}^{\infty} (\boldsymbol{\Psi}_{m\ell} \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Psi}_{mh}^\top) \otimes \{(\mathbf{I}_q \otimes \boldsymbol{\Psi}_{mn}) \mathbf{R}_m(\ell+n, h+k)(\mathbf{I}_q \otimes \boldsymbol{\Psi}_{mk}^\top)\}, \tag{9}$$

where $\mathbf{Q}_m(\ell, n) = \mathrm{E}(\mathbf{x}_{t-\ell}\mathbf{x}_{t-n}^\top \xi_{mt|T})$ and $\mathbf{R}_m(\ell, n) = \mathrm{E}(\mathbf{w}_{m,t-\ell}\mathbf{w}_{m,t-n}^\top \xi_{mt|T})$. The cross components are

$$\mathcal{F}_a(\boldsymbol{\gamma}_m, \boldsymbol{\nu}_m) = \sum_{\ell=0}^{\infty} \{\boldsymbol{\Psi}_{m\ell}^\top \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Theta}_m^{-1}(1)\} \otimes \mathrm{E}(\mathbf{x}_{t-\ell}\xi_{mt|T}), \tag{10}$$

$$\mathcal{F}_a(\boldsymbol{\delta}_m, \boldsymbol{\gamma}_m) = -\sum_{\ell,n,r\geq 0}^{\infty} (\boldsymbol{\Psi}_{mr} \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Psi}_{m\ell}) \otimes \{(\mathbf{I}_q \otimes \boldsymbol{\Psi}_{mn}) \mathrm{E}(\mathbf{w}_{m,t-r-n}\mathbf{x}_{t-\ell}^\top \xi_{mt|T})\}. \tag{11}$$

The matrix coefficients $\boldsymbol{\Psi}_{m\ell}$ can be computed by the recursive expression in (5) and the smoothed regime probabilities $\xi_{mt|T}$ are derived by the smoothing algorithm described in this section.

For practical inference purposes, expectations and infinite series in the above statement can be approximated by sample means and summations up to order N, respectively, for T (sample size) and N sufficiently large. The matrices $\boldsymbol{\Psi}_{m\ell}$ are replaced by their ML estimates $\widehat{\boldsymbol{\Psi}}_{m\ell}$ which can be obtained from the recurrence formula in (5) by using the ML estimates $\widehat{\boldsymbol{\Theta}}_{mj}$ of $\boldsymbol{\Theta}_{mj}$ for all $j \in \{1, \ldots, q\}$. Of course, we replace $\boldsymbol{\Omega}_m^{-1}$ by its ML estimate. This gives a convenient plug-in approach to approximate the block matrices above. Recall that [10] introduced an EM algorithm for obtaining ML estimates of parameters for discrete-valued Markov switching processes. The simplicity of the EM algorithm permits potential application of the approach to large vector systems. Further developments related to this algorithm have been recently proposed in [3]. Now the asymptotic covariance matrix of the Gaussian ML estimator $\widehat{\boldsymbol{\theta}}_m$ of $\boldsymbol{\theta}_m$ is given by

$$\mathrm{var}_a(\widehat{\boldsymbol{\theta}}_m) = \frac{1}{T}\begin{pmatrix} \mathcal{F}_a(\boldsymbol{\alpha}_m)^{-1} & \mathbf{0} \\ \mathbf{0}^\top & \mathcal{F}_a(\boldsymbol{\sigma}_m)^{-1} \end{pmatrix}.$$

Let us consider a constraint of the form $\mathbf{R}\boldsymbol{\theta}_m = \mathbf{r}$, with rank $(\mathbf{R}) = r$, involving the coefficients of the MS VARMA at the regime $s_t = m$. The Wald test has an asymptotic χ^2 distribution with r degrees of freedom, i.e., $\chi_r^2 = (\mathbf{R}\widehat{\boldsymbol{\theta}}_m -$

$\mathbf{r})^\top\{\mathbf{R}\,\mathrm{var}_a(\widehat{\boldsymbol\theta}_m)\mathbf{R}^\top\}^{-1}(\mathbf{R}\widehat{\boldsymbol\theta}_m-\mathbf{r})$. Let $\widehat{\theta}_{mi}$ and $\widehat{\upsilon}_{mi}$ denote the ith component of $\widehat{\boldsymbol\theta}_m$ and the ith element on the principal diagonal of $\mathrm{var}_a(\widehat{\boldsymbol\theta}_m)$, respectively. The sample standard error of $\widehat{\theta}_{mi}$ is given by $\widehat{s.e.}(\widehat{\theta}_{mi})=\sqrt{\widehat{\upsilon}_{mi}}$, and the 95% confidence interval for the true parameter value θ_{mi} is $\widehat{\theta}_{mi}\pm 1.96\sqrt{\widehat{\upsilon}_{mi}}$. A classical procedure to test a null hypothesis $\theta_{mi}=0$ can be performed by using the t-statistic $t=\widehat{\theta}_{mi}/\sqrt{\widehat{\upsilon}_{mi}}$. Section 4 is devoted to illustrate such procedures via numerical evaluation. Note that $\partial^2\mathcal{L}(\boldsymbol\lambda)/\partial\boldsymbol\theta_m\partial\boldsymbol\theta_n^\top=\mathbf{0}$ for every $m,n\in\{1,\ldots,M\}$ with $m\neq n$. So, as done above, we can always consider the asymptotic Fisher information matrix of parameters $\boldsymbol\theta_m$ in each component, separately, that is, $\mathcal{F}_a(\boldsymbol\theta)=\mathrm{diag}\{\mathcal{F}_a(\boldsymbol\theta_1),\ldots,\mathcal{F}_a(\boldsymbol\theta_M)\}$.

Finally, note that a path dependence problem occurs for the case of an MS VARMA model because an unobservable variable at date t (the lagged error term in VARMA models or the lagged conditional variance in GARCH models) depends on the entire path of states that have been followed until that date. This is an estimation problem, but it will not affect our results because the FI matrix is derived analytically, not involving practical estimation problems. However, when one estimates the FI matrix, an approximation will be made as, for example, the usual Kim approximation proposed in [15].

3. Some examples

Example 3.1 (*MS VAR(p) models*). Suppose $p\geq 1$. In this case, we have $\boldsymbol\Theta_m(L)=\mathbf{I}_K$ and $\boldsymbol\delta_m=\mathbf{0}$ for every $m\in\{1,\ldots,M\}$. Then $\boldsymbol\theta_m=(\boldsymbol\nu_m^\top\boldsymbol\gamma_m^\top\boldsymbol\sigma_m^\top)^\top$. The asymptotic FI matrix $\mathcal{F}_a(\boldsymbol\theta_m)$ reduces to the symmetric block matrix

$$\mathcal{F}_a(\boldsymbol\theta_m)=\begin{pmatrix}\mathcal{F}_a(\boldsymbol\nu_m) & \mathcal{F}_a(\boldsymbol\gamma_m,\boldsymbol\nu_m)^\top & \mathbf{0}^\top\\ \mathcal{F}_a(\boldsymbol\gamma_m,\boldsymbol\nu_m) & \mathcal{F}_a(\boldsymbol\gamma_m) & \mathbf{0}^\top\\ \mathbf{0} & \mathbf{0} & \mathcal{F}_a(\boldsymbol\sigma_m)\end{pmatrix},$$

where $\mathcal{F}_a(\boldsymbol\sigma_m)$ is given in (6). From (7), (8) and (10) it follows that

$$\mathcal{F}_a(\boldsymbol\nu_m)=\pi_m\boldsymbol\Omega_m^{-1},\qquad \mathcal{F}_a(\boldsymbol\gamma_m)=\boldsymbol\Omega_m^{-1}\otimes \mathrm{E}(\mathbf{x}_t\mathbf{x}_t^\top\xi_{mt|T}),\qquad \mathcal{F}_a(\boldsymbol\gamma_m,\boldsymbol\nu_m)=\boldsymbol\Omega_m^{-1}\otimes\mathrm{E}(\mathbf{x}_t\xi_{mt|T}).$$

These formulas extend those given in Proposition 11.1 from [11], case $M=1$, and are simpler than the existing expressions which involve integration over the frequency domain as in [2,5]. The inverse matrix of $\mathcal{F}_a(\boldsymbol\theta_m)$ can be derived using Woodbury's formula. In particular, if $\boldsymbol\nu_m=\mathbf{0}$, then $\mathrm{var}_a(\widehat{\boldsymbol\gamma}_m)=\boldsymbol\Omega_m\otimes\mathbf{Q}_m^{-1}$, where $\mathbf{Q}_m=\mathrm{E}(\mathbf{x}_t\mathbf{x}_t^\top\xi_{mt|T})$, for $m=1,\ldots,M$.

Example 3.2 (*MS VMA(q) models*). Suppose $q\geq 1$. In this case, we have $\boldsymbol\Phi_m(L)=\mathbf{I}_K$ and $\boldsymbol\gamma_m=\mathbf{0}$ for every $m\in\{1,\ldots,M\}$. Then $\boldsymbol\theta_m=(\boldsymbol\nu_m^\top\boldsymbol\delta_m^\top\boldsymbol\sigma_m^\top)^\top$. The asymptotic FI matrix $\mathcal{F}_a(\boldsymbol\theta_m)$ reduces to the diagonal block matrix $\mathcal{F}_a(\boldsymbol\theta_m)=\mathrm{diag}\{\mathcal{F}_a(\boldsymbol\nu_m),\mathcal{F}_a(\boldsymbol\delta_m),\mathcal{F}_a(\boldsymbol\sigma_m)\}$, whose entries are given by (6), (7) and (9) in that order. For these models, the components of $\mathbf{w}_{m,t}$ are $\mathbf{z}_{m,t-j}=\mathbf{y}_{t-j}-\boldsymbol\nu_m$ for all $j\in\{1,\ldots,q\}$.

4. A numerical illustration

In this section some numerical results are displayed for an MS VARMA process such that $M=K=2$ and $p=q=1$, i.e., a 2-state bivariate ARMA(1, 1) model, viz

$$\mathbf{y}_t-\boldsymbol\Phi_{s_t}\mathbf{y}_{t-1}=\boldsymbol\nu_{s_t}+\mathbf{u}_t-\boldsymbol\Theta_{s_t}\mathbf{u}_{t-1},\quad \mathbf{u}_t\sim\mathrm{NID}(\mathbf{0},\boldsymbol\Omega_{s_t}),$$

where

$$\boldsymbol\nu_{s_t}=\begin{pmatrix}-0.6(s_t-2)\\-0.1s_t+0.4\end{pmatrix},\qquad \boldsymbol\Phi_{s_t}=\begin{pmatrix}0.6(s_t-1) & 0.4\\ -0.3(s_t-2) & -0.9s_t+2.1\end{pmatrix},$$

$$\boldsymbol\Theta_{s_t}=\begin{pmatrix}-0.9(s_t-2) & 0.1s_t\\ 0.1(s_t+1) & 0.8(s_t-1)\end{pmatrix},$$

$$\boldsymbol\Omega_{s_t}=\begin{pmatrix}0.2s_t & -0.3s_t+0.7\\ -0.3s_t+0.7 & -3.3s_t+7.8\end{pmatrix},$$

and $s_t\in\{1,2\}$. The transition probability matrix is

$$\mathbf{P}=\begin{pmatrix}p_{11} & 1-p_{11}\\ 1-p_{22} & p_{22}\end{pmatrix},$$

where $p_{11}=0.2$ and $p_{22}=0.7$. The unconditional probabilities are $\pi_1=0.2727$ and $\pi_2=0.7273$. For $m\in\{1,2\}$, we set $\boldsymbol\gamma_m=\mathrm{vec}\,\boldsymbol\Phi_m$, $\boldsymbol\delta_m=\mathrm{vec}\,\boldsymbol\Theta_m$ and $\boldsymbol\sigma_m=\mathrm{vec}\,\boldsymbol\Omega_m^{-1}$. The parameter vector $\boldsymbol\theta_m=(\boldsymbol\alpha_m^\top\,\boldsymbol\sigma_m^\top)^\top$ is 14×1, where $\boldsymbol\alpha_m=(\boldsymbol\nu_m^\top\,\boldsymbol\gamma_m^\top\,\boldsymbol\delta_m^\top)^\top$ is 10×1. The basic assumption that the eigenvalues of $\boldsymbol\Theta_{s_t}$ are less than 1 in modulus is fulfilled because its spectral radius is $0.9744-0.0527s_t$.

We generate $T=1000$ observations of $\mathbf{u}_t$ by using Gaussian deviates with zero mean and variance $\boldsymbol\Omega_{s_t}$. Then the process $\mathbf{y}_t$ is obtained from the above model by using true parameters. A partitioned form of the asymptotic FI matrix is considered in Appendix B. The latter is evaluated at the ML estimates of true parameters by using the matrix expressions in the statement of the main theorem. Note that the summations in Eqs. (8)–(11) are taken up to order $N = 100$. The matrix coefficients in (4) for this example are given by $\boldsymbol{\Psi}_{m\ell} = \boldsymbol{\Theta}_m^{\ell}$, for each $m \in \{1, 2\}$ and $\ell \geq 0$ (here we set $\boldsymbol{\Theta}_m^{\ell} = \mathbf{I}_2$ for $\ell = 0$).

Now we consider a testing hypothesis problem and show how the results can be used to perform the test using the asymptotic covariance matrix of the ML estimator. Let the null hypothesis be $\mathbf{v}_m = \mathbf{0}$, where $\mathbf{v}_m = (\boldsymbol{\nu}_m^\top\, 0 \cdots 0)^\top$ is 10×1, for a fixed $m \in \{1, 2\}$. The Wald test has an asymptotic χ^2 distribution with 2 degrees of freedom whose 95% critical value is 0.103. Since $\widehat{\mathbf{v}}_m^\top \operatorname{var}_a(\widehat{\boldsymbol{\alpha}}_m)\widehat{\mathbf{v}}_m$ is 0.116 (resp. 0.004) for $m = 1$ (resp. $m = 2$), we reject (resp. accept) the null hypothesis that the mth intercept is zero. Let us consider the null hypothesis $\mathbf{r}_m = \mathbf{0}$, where $\mathbf{r}_m = (0\ 0\ \boldsymbol{\gamma}_m^\top\ 0\ 0\ 0\ 0)^\top$ is 10×1. The 95% critical value for the χ^2 distribution with 4 degrees of freedom is 0.711. Since $\widehat{\mathbf{r}}_m^\top \operatorname{var}_a(\widehat{\boldsymbol{\alpha}}_m)\widehat{\mathbf{r}}_m$ is 46.73 (resp. 10.91) for $m = 1$ (resp. $m = 2$), we reject the null hypothesis.

Finally, we use the results to construct asymptotic confidence intervals for the unknown parameters of the MS VARMA model. The sample standard error of the $(1, 1)$ element ν_{11} of $\boldsymbol{\nu}_1$ is $\widehat{s.e.}(\widehat{\nu}_{11}) = \sqrt{0.0203} = 0.1425$, where $\widehat{\nu}_{11} = 0.7210$ is the ML estimator of ν_{11}. Then the 95% confidence interval for the true parameter value $\nu_{11} = 0.6$ is $\widehat{\nu}_{11} \pm 1.96\widehat{s.e.}(\widehat{\nu}_{11}) = 0.7210 \pm 0.2793$, i.e., $[0.4417, 1.0003]$. The sample standard error of the element $\phi_{2,12}$ (resp. $\theta_{2,22}$) of $\boldsymbol{\Phi}_2$ (resp. $\boldsymbol{\Theta}_2$) is $\widehat{s.e.}(\widehat{\phi}_{2,12}) = \sqrt{0.0101} = 0.1005$ (resp. $\widehat{s.e.}(\widehat{\theta}_{2,22}) = \sqrt{0.0672} = 0.2592$), where $\widehat{\phi}_{2,12} = 0.35$ (resp. $\widehat{\theta}_{2,22} = 0.78$). Then the 95% confidence interval for the true parameter $\phi_{2,12} = 0.4$ (resp. $\theta_{2,22} = 0.8$) is $\widehat{\phi}_{2,12} \pm 1.96\widehat{s.e.}(\widehat{\phi}_{2,12}) = 0.35 \pm 0.1969$ (resp. $\widehat{\theta}_{2,22} \pm 1.96\widehat{s.e.}(\widehat{\theta}_{2,22}) = 0.78 \pm 0.5080$), i.e., $[0.153, 0.547]$ (resp. $[0.272;\ 1.288]$).

5. Conclusions

In this paper we derived a closed form expression for the asymptotic FI matrix of the Gaussian ML estimator for MS VARMA models by using matrix calculus. Then we used this matrix to derive the asymptotic covariance matrix of the ML estimator of the model parameters. Its applicability relates with testing hypothesis problems and easier computation of asymptotic confidence intervals for the unknown parameters of MS VARMA models.

Our approach may be potentially useful for application in other situations such as MS VARMA models with autoregressive conditional heteroskedastic (ARCH) innovations. For example, one can consider such models in the framework of the Baba–Engle–Kraft–Kroner (BEKK) formulation; see also [6]. Also, these methods might be applied for general classes of finite mixture models, as considered, e.g., in [13]. The obtained results could be applied to further models such as multivariate MS GARCH. These models suffer from path-dependence and their estimation needs approximated filters as, e.g., in [15]. Simulation results on univariate MS GARCH models can be found in [9] and further investigations for the multivariate case would be of interest.

Acknowledgments

Work performed under the auspices of the Cournot Foundation, Mont-sous-Vaudrey, France, and financially supported by a research grant from the University of Modena and Reggio Emilia, and the University of Verona, Italy. We are grateful to the Editor-in-Chief, Prof. Christian Genest, the Associate Editor and three anonymous referees for their very useful suggestions and remarks which improved the final version of the paper.

Appendix A. Proof of the main theorem

A.1. Derivation of $\mathcal{F}_a(\alpha_m)$

The block matrix $\mathcal{F}_a(\boldsymbol{\alpha}_m)$ is

$$\mathcal{F}_a(\boldsymbol{\alpha}_m) = \mathrm{E}\left\{\left(\frac{\partial \mathbf{u}_t^\top}{\partial \boldsymbol{\alpha}_m}\right) \boldsymbol{\Omega}_m^{-1} \xi_{mt|T} \left(\frac{\partial \mathbf{u}_t}{\partial \boldsymbol{\alpha}_m^\top}\right)\right\}. \tag{A.1}$$

From (3) and (4) we get

$$\begin{aligned}
\frac{\partial \mathbf{u}_t}{\partial \boldsymbol{\Phi}_{mi}^\top} &= \frac{\partial \mathbf{u}_t}{\partial (\operatorname{vec} \boldsymbol{\Phi}_{mi}^\top)^\top} = \boldsymbol{\Theta}_m^{-1}(L) \frac{\partial \{\boldsymbol{\Phi}_m(L)\mathbf{y}_t\}}{\partial (\operatorname{vec} \boldsymbol{\Phi}_{mi}^\top)^\top} \\
&= \boldsymbol{\Theta}_m^{-1}(L) \frac{\partial \operatorname{vec}(-\boldsymbol{\Phi}_{mi}\mathbf{y}_{t-i})}{\partial (\operatorname{vec} \boldsymbol{\Phi}_{mi}^\top)^\top} = -\boldsymbol{\Theta}_m^{-1}(L) \frac{\partial \operatorname{vec}(\mathbf{y}_{t-i}^\top \boldsymbol{\Phi}_{mi}^\top)}{\partial (\operatorname{vec} \boldsymbol{\Phi}_{mi}^\top)^\top} \\
&= -\boldsymbol{\Theta}_m^{-1}(L)(\mathbf{I}_K \otimes \mathbf{y}_{t-i}^\top) \frac{\partial \operatorname{vec} \boldsymbol{\Phi}_{mi}^\top}{\partial (\operatorname{vec} \boldsymbol{\Phi}_{mi}^\top)^\top} = -\boldsymbol{\Theta}_m^{-1}(L)(\mathbf{I}_K \otimes \mathbf{y}_{t-i}^\top) \\
&= -\left(\sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_{m\ell} L^\ell\right)(\mathbf{I}_K \otimes \mathbf{y}_{t-i}^\top) = -\sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_{m\ell}(\mathbf{I}_K \otimes \mathbf{y}_{t-i-\ell}^\top) \\
&= -\sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_{m\ell} \otimes \mathbf{y}_{t-i-\ell}^\top
\end{aligned} \tag{A.2}$$

for all $i \in \{1, \ldots, p\}$ and $m \in \{1, \ldots, M\}$. Substituting (A.2) into (A.1) yields

$$\begin{aligned}
\mathcal{F}_a(\boldsymbol{\Phi}_{mi}, \boldsymbol{\Phi}_{mj}) &= \mathrm{E}\left\{\left(\frac{\partial \mathbf{u}_t^\top}{\partial \boldsymbol{\Phi}_{mi}}\right) \boldsymbol{\Omega}_m^{-1} \xi_{mt|T} \left(\frac{\partial \mathbf{u}_t}{\partial \boldsymbol{\Phi}_{mj}^\top}\right)\right\} \\
&= \mathrm{E}\left\{\left(\sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_{m\ell}^\top \otimes \mathbf{y}_{t-i-\ell}\right) \boldsymbol{\Omega}_m^{-1} \xi_{mt|T} \left(\sum_{n=0}^{\infty} \boldsymbol{\Psi}_{mn} \otimes \mathbf{y}_{t-j-n}^\top\right)\right\} \\
&= \sum_{\ell=0}^{\infty} \sum_{n=0}^{\infty} (\boldsymbol{\Psi}_{m\ell}^\top \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Psi}_{mn}) \otimes \mathrm{E}(\mathbf{y}_{t-i-\ell} \mathbf{y}_{t-j-n}^\top \xi_{mt|T})
\end{aligned}$$

for all $i, j \in \{1, \ldots, p\}$. Define

$$\mathbf{Q}_m^T(\ell, n) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_{t-\ell} \mathbf{x}_{t-n}^\top \xi_{mt|T} \in \mathbb{R}^{(pK)\times(pK)}.$$

Then

$$\mathrm{plim}_{T\to\infty}\, \mathbf{Q}_m^T(\ell, n) = \mathrm{E}(\mathbf{x}_{t-\ell} \mathbf{x}_{t-n}^\top \xi_{mt|T}) = \mathbf{Q}_m(\ell, n),$$

so $\mathbf{Q}_m(\ell, n)$ can be approximated by $\mathbf{Q}_m^T(\ell, n)$ for T sufficiently large. Then the $(pK^2) \times (pK^2)$ block matrix $\mathcal{F}_a(\boldsymbol{\gamma}_m)$ is given by (8). This result generalizes Proposition 11.1 from [11] for the case of a pth Gaussian vector autoregression without changes in regime (i.e., $M = 1$ and $\boldsymbol{\Theta}_m(L) = \mathbf{I}_K$), i.e., $\mathcal{F}_a(\widehat{\boldsymbol{\gamma}}) = \boldsymbol{\Omega}^{-1} \otimes \mathbf{Q}$, where $\mathbf{Q} = \mathrm{E}(\mathbf{x}_t \mathbf{x}_t^\top)$.

From (3) we get

$$\begin{aligned}
\frac{\partial \mathbf{u}_t^\top}{\partial \boldsymbol{\Theta}_{mj}} &= \left(\frac{\partial \mathbf{u}_t}{\partial \boldsymbol{\Theta}_{mj}^\top}\right)^\top = \left[\frac{\partial\, \mathrm{vec}(\mathbf{u}_t^\top)}{\partial \{\mathrm{vec}(\boldsymbol{\Theta}_{mj}^\top)\}^\top}\right]^\top \\
&= \left[\frac{\partial\, \mathrm{vec}\{\boldsymbol{\Theta}_m^{-1}(L)^\top\}}{\partial \{\mathrm{vec}(\boldsymbol{\Theta}_{mj}^\top)\}^\top}\right]^\top [\mathbf{I}_K \otimes \{\boldsymbol{\Phi}_m(L)\mathbf{y}_t - \boldsymbol{\nu}_m\}] \\
&= \left[\{\boldsymbol{\Theta}_m^{-1}(L)^\top \otimes \boldsymbol{\Theta}_m^{-1}(L)^\top\} L^j \frac{\partial\, \mathrm{vec}(\boldsymbol{\Theta}_{mj}^\top)}{\partial \{\mathrm{vec}(\boldsymbol{\Theta}_{mj}^\top)\}^\top}\right]^\top [\mathbf{I}_K \otimes \{\boldsymbol{\Phi}_m(L)\mathbf{y}_t - \boldsymbol{\nu}_m\}] \\
&= \{\boldsymbol{\Theta}_m^{-1}(L) \otimes \boldsymbol{\Theta}_m^{-1}(L)\}(\mathbf{I}_k \otimes \mathbf{z}_{m,t-j}) \\
&= \sum_{\ell=0}^{\infty} \sum_{n=0}^{\infty} (\boldsymbol{\Psi}_{m\ell} \otimes \boldsymbol{\Psi}_{mn})(\mathbf{I}_K \otimes \mathbf{z}_{m,t-j-\ell-n}) \\
&= \sum_{\ell=0}^{\infty} \sum_{n=0}^{\infty} \boldsymbol{\Psi}_{m\ell} \otimes (\boldsymbol{\Psi}_{mn} \mathbf{z}_{m,t-j-\ell-n})
\end{aligned} \tag{A.3}$$

for all $j \in \{1, \ldots, q\}$ and $m \in \{1, \ldots, M\}$. Substituting (A.3) into (A.1) yields

$$\begin{aligned}
\mathcal{F}_a(\boldsymbol{\Theta}_{mi}, \boldsymbol{\Theta}_{mj}) &= \mathrm{E}\left\{\left(\frac{\partial \mathbf{u}_t^\top}{\partial \boldsymbol{\Theta}_{mi}}\right) \boldsymbol{\Omega}_m^{-1} \xi_{mt|T} \left(\frac{\partial \mathbf{u}_t}{\partial \boldsymbol{\Theta}_{mj}^\top}\right)\right\} \\
&= \mathrm{E}\Big[\Big\{\sum_{\ell,n\ge 0} \boldsymbol{\Psi}_{m\ell} \otimes (\boldsymbol{\Psi}_{mn} \mathbf{z}_{m,t-i-\ell-n})\Big\} \boldsymbol{\Omega}_m^{-1} \xi_{mt|T} \times \Big\{\sum_{h,k\ge 0} \boldsymbol{\Psi}_{mh}^\top \otimes (\mathbf{z}_{m,t-j-h-k}^\top \boldsymbol{\Psi}_{mk}^\top)\Big\}\Big] \\
&= \sum_{\ell,m,h,k\ge 0} (\boldsymbol{\Psi}_{m\ell} \boldsymbol{\Omega}_m^{-1} \boldsymbol{\Psi}_{mh}^\top) \otimes \{\boldsymbol{\Psi}_{mn} \mathrm{E}(\mathbf{z}_{m,t-i-\ell-n} \mathbf{z}_{m,t-j-h-k}^\top \xi_{mt|T}) \boldsymbol{\Psi}_{mk}^\top\}
\end{aligned}$$

for all $i, j \in \{1, \ldots, q\}$. Set

$$\mathbf{R}_m^T(\ell, n) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{w}_{m,t-\ell} \mathbf{w}_{m,t-n}^\top \xi_{mt|T} \in \mathbb{R}^{(qK)\times(qK)}.$$

Then

$$\mathrm{plim}_{T\to\infty}\, \mathbf{R}_m^T(\ell, n) = \mathrm{E}(\mathbf{w}_{m,t-\ell} \mathbf{w}_{m,t-n}^\top \xi_{mt|T}) = \mathbf{R}_m(\ell, n),$$

so $\mathbf{R}_m(\ell, n)$ can be approximated by $\mathbf{R}_m^T(\ell, n)$ for T sufficiently large. This proves that the $(qK^2) \times (qK^2)$ block matrix $\mathcal{F}_a(\boldsymbol{\delta}_m)$ is given by (9).

Since $\mathbf{z}_{m,t-j} = \boldsymbol{\Theta}_m(L)\mathbf{u}_{t-j} = \boldsymbol{\Theta}_m(L)\boldsymbol{\Sigma}_{s_{t-j}}\boldsymbol{\epsilon}_{t-j}$, we could derive further matrix expressions for $\mathcal{F}_a(\boldsymbol{\Theta}_{mj})$ and $\mathcal{F}_a(\boldsymbol{\delta}_m)$, as indicated below. In fact, we have

$$\begin{aligned}\operatorname{vec}\mathrm{E}(\mathbf{z}_{m,t-j}\mathbf{z}_{m,t-j}^\top\xi_{mt|T}) &= \mathrm{E}\left[\{\boldsymbol{\Theta}_m(L)\otimes\boldsymbol{\Theta}_m(L)\}(\boldsymbol{\Sigma}_{s_{t-j}}\otimes\boldsymbol{\Sigma}_{s_{t-j}})\,\xi_{mt|T}\operatorname{vec}(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top)\right]\\ &= \mathrm{E}\left[\{\boldsymbol{\Theta}_m(L)\otimes\boldsymbol{\Theta}_m(L)\}(\boldsymbol{\Sigma}_{s_{t-j}}\otimes\boldsymbol{\Sigma}_{s_{t-j}})\xi_{mt|T}\right]\operatorname{vec}\mathrm{E}(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top)\\ &= \mathrm{E}\left[\{\boldsymbol{\Theta}_m(L)\otimes\boldsymbol{\Theta}_m(L)\}(\boldsymbol{\Sigma}_{s_{t-j}}\otimes\boldsymbol{\Sigma}_{s_{t-j}})\xi_{mt|T}\right]\operatorname{vec}\mathbf{I}_K\end{aligned}$$

because $(\boldsymbol{\epsilon}_t)$ is independent of (s_t) (and hence $\xi_{mt|T}$). For $r \geq 1$, we have

$$\begin{aligned}\mathrm{E}(\boldsymbol{\Omega}_{s_{t-r}}\xi_{mt|T}) &= \mathrm{E}[\mathrm{E}(\boldsymbol{\Omega}_{s_{t-r}}\xi_{mt|T}|s_{t-r})]\\ &= \mathrm{E}\left\{\sum_{n=1}^{M}\boldsymbol{\Omega}_n E(\xi_{mt|T}|s_{t-r}=n)p_{nm}^{(r)}\pi_n\right\}\\ &= \sum_{n=1}^{M}\boldsymbol{\Omega}_n\mathrm{E}\{\mathrm{E}(\xi_{mt|T}|s_{t-r}=n)\}p_{nm}^{(r)}\pi_n\\ &= \sum_{n=1}^{M}\boldsymbol{\Omega}_n\mathrm{E}(\xi_{mt|T})p_{nm}^{(r)}\pi_n = \pi_m\sum_{n=1}^{M}\boldsymbol{\Omega}_n p_{nm}^{(r)}\pi_n,\end{aligned}$$

where $p_{nm}^{(r)} = \Pr(s_t = m|s_{t-r} = n)$ is the (n, m)th element of the matrix $\mathbf{P}^r$. Here $\mathbf{P}^r = \mathbf{P}$ if $r = 1$, and $\mathbf{P}^r = \mathbf{P}\cdots\mathbf{P}$ (r times) if $r > 1$. Putting such formulas together yields the claim.

From (3) we get

$$\begin{aligned}\frac{\partial\mathbf{u}_t}{\partial\boldsymbol{\nu}_m^\top} &= \frac{\partial\mathbf{u}_t}{\partial\{\operatorname{vec}(\boldsymbol{\nu}_m^\top)\}^\top} = -\frac{\partial\boldsymbol{\Theta}_m^{-1}(1)\boldsymbol{\nu}_m}{\partial\{\operatorname{vec}(\boldsymbol{\nu}_m^\top)\}^\top}\\ &= -\frac{\partial\operatorname{vec}\{\boldsymbol{\nu}_m^\top\boldsymbol{\Theta}_m^{-1}(1)^\top\}}{\partial\{\operatorname{vec}(\boldsymbol{\nu}_m^\top)\}^\top} = -\boldsymbol{\Theta}_m^{-1}(1).\end{aligned}\tag{A.4}$$

Substituting (A.4) into (A.1) yields the matrix expression (7) for $\mathcal{F}_a(\boldsymbol{\nu}_m)$.

For each $i \in \{1, \ldots, p\}$, we have

$$\begin{aligned}\mathcal{F}_a(\boldsymbol{\Phi}_{mi},\boldsymbol{\nu}_m) &= \mathrm{E}\left\{\left(\frac{\partial\mathbf{u}_t^\top}{\partial\boldsymbol{\Phi}_{mi}}\right)\boldsymbol{\Omega}_m^{-1}\xi_{mt|T}\left(\frac{\partial\mathbf{u}_t}{\partial\boldsymbol{\nu}_m^\top}\right)\right\}\\ &= \mathrm{E}\left\{\left(\sum_{\ell=0}^{\infty}\boldsymbol{\Psi}_{m\ell}^\top\otimes\mathbf{y}_{t-i-\ell}\right)\boldsymbol{\Omega}_m^{-1}\xi_{mt|T}\boldsymbol{\Theta}_m^{-1}(1)\right\}\\ &= \sum_{\ell=0}^{\infty}\{\boldsymbol{\Psi}_{m\ell}^\top\boldsymbol{\Omega}_m^{-1}\boldsymbol{\Theta}_m^{-1}(1)\}\otimes\mathrm{E}(\mathbf{y}_{t-i-\ell}\xi_{mt|T})\end{aligned}$$

hence the $(pK^2)\times K$ cross component $\mathcal{F}_a(\boldsymbol{\gamma}_m,\boldsymbol{\nu}_m)$ is given by (10).

We now show that the $(qK^2)\times K$ cross component $\mathcal{F}_a(\boldsymbol{\delta}_m,\boldsymbol{\nu}_m)$ is zero.

In fact,

$$\begin{aligned}\mathcal{F}_a(\boldsymbol{\Theta}_{mj},\boldsymbol{\nu}_m) &= \mathrm{E}\left\{\left(\frac{\partial\mathbf{u}_t^\top}{\partial\boldsymbol{\Theta}_{mj}}\right)\boldsymbol{\Omega}_m^{-1}\xi_{mt|T}\left(\frac{\partial\mathbf{u}_t}{\partial\boldsymbol{\nu}_m^\top}\right)\right\}\\ &= -\mathrm{E}\left[\left\{\sum_{\ell,n\geq 0}(\boldsymbol{\Psi}_{m\ell}\otimes\boldsymbol{\Psi}_{mn})(\mathbf{I}_K\otimes\mathbf{z}_{m,t-j-\ell-n})\right\}\boldsymbol{\Omega}_m^{-1}\xi_{mt|T}\boldsymbol{\Theta}_m^{-1}(1)\right]\\ &= -\sum_{\ell,n\geq 0}(\boldsymbol{\Psi}_{m\ell}\otimes\boldsymbol{\Psi}_{mn})\left\{\boldsymbol{\Omega}_m^{-1}\otimes\mathrm{E}(\mathbf{z}_{m,t-j-\ell-n}\xi_{mt|T})\right\}\boldsymbol{\Theta}_m^{-1}(1)\end{aligned}$$

for all $j \in \{1, \ldots, q\}$. Now the middle term vanishes because

$$\begin{aligned}\mathrm{E}(\mathbf{z}_{m,t-r}\xi_{mt|T}) &= \mathrm{E}\{\boldsymbol{\Theta}_m(L)\boldsymbol{\Sigma}_{s_{t-r}}\boldsymbol{\epsilon}_{t-r}\xi_{mt|T}\}\\ &= \mathrm{E}\{\boldsymbol{\Theta}_m(L)\boldsymbol{\Sigma}_{s_{t-r}}\xi_{mt|T}\}\mathrm{E}(\boldsymbol{\epsilon}_{t-r}) = \mathbf{0}.\end{aligned}$$

Here we use the fact that $\mathrm{E}(\boldsymbol{\epsilon}_t) = \mathbf{0}$ and $(\boldsymbol{\epsilon}_t)$ is independent of (s_t), and hence $\xi_{mt|T}$. Thus $\mathcal{F}_a(\boldsymbol{\delta}_m,\boldsymbol{\nu}_m) = \mathbf{0}$.

For every $i \in \{1, \dots, p\}$ and $j \in \{1, \dots, q\}$, we have

$$\begin{aligned}
\mathcal{F}_a(\mathbf{\Theta}_{mj}, \mathbf{\Phi}_{mi}) &= \mathrm{E}\left\{\left(\frac{\partial \mathbf{u}_t^\top}{\partial \mathbf{\Theta}_{mj}}\right) \mathbf{\Omega}_m^{-1} \xi_{mt|T} \left(\frac{\partial \mathbf{u}_t}{\partial \mathbf{\Phi}_{mi}^\top}\right)\right\} \\
&= -\mathrm{E}\left[\left\{\sum_{r,n\geq 0} (\mathbf{\Psi}_{mr} \otimes \mathbf{\Psi}_{mn})(\mathbf{I}_K \otimes \mathbf{z}_{m,t-j-r-n})\right\} \mathbf{\Omega}_m^{-1} \xi_{mt|T} \left(\sum_{\ell=0}^{\infty} \mathbf{\Psi}_{m\ell} \otimes \mathbf{y}_{t-i-\ell}^\top\right)\right] \\
&= -\sum_{\ell,n,r\geq 0} (\mathbf{\Psi}_{mr} \otimes \mathbf{\Psi}_{mn})(\mathbf{\Omega}_m^{-1} \otimes \mathbf{I}_K) \left\{\mathbf{\Psi}_{m\ell} \otimes \mathrm{E}(\mathbf{z}_{m,t-j-r-n} \mathbf{y}_{t-i-\ell}^\top \xi_{mt|T})\right\} \\
&= -\sum_{\ell,n,r\geq 0} (\mathbf{\Psi}_{mr} \mathbf{\Omega}_m^{-1} \mathbf{\Psi}_{m\ell}) \otimes \left\{\mathbf{\Psi}_{mn} \mathrm{E}(\mathbf{z}_{m,t-j-r-n} \mathbf{y}_{t-i-\ell}^\top \xi_{mt|T})\right\}
\end{aligned}$$

hence the $(qK^2) \times (pK^2)$ cross component $\mathcal{F}_a(\boldsymbol{\delta}_m, \boldsymbol{\gamma}_m)$ is given by (11).

A.2. Derivation of $\mathcal{F}_a(\boldsymbol{\sigma}_m)$ and $\mathcal{F}_a(\boldsymbol{\sigma}_m, \boldsymbol{\alpha}_m)$

Now we derive the elements of the FI matrix $\mathcal{F}_a(\boldsymbol{\theta}_m)$ which involve $\boldsymbol{\sigma}_m$. The derivative of $\mathcal{L}(\boldsymbol{\lambda})$ with respect to $\boldsymbol{\sigma}_m$ is

$$\frac{\partial \mathcal{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\sigma}_m} = \sum_{t=1}^{T} \frac{\partial \ln \eta_{mt}(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_m} \xi_{mt|T},$$

where

$$\frac{\partial \ln \eta_{mt}(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_m} = \frac{1}{2} \operatorname{vec} \mathbf{\Omega}_m - \frac{1}{2} \mathbf{u}_t \otimes \mathbf{u}_t$$

and $\mathbf{u}_t$ is given in (3). Then we get

$$\frac{\partial^2 \ln \eta_{mt}(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_m \partial \boldsymbol{\sigma}_m^\top} = \frac{1}{2} \frac{\partial \operatorname{vec} \mathbf{\Omega}_m}{\partial (\operatorname{vec} \mathbf{\Omega}_m^{-1})^\top} = -\frac{1}{2} (\mathbf{\Omega}_m \otimes \mathbf{\Omega}_m).$$

Substituting this relation into

$$\mathcal{F}_a(\boldsymbol{\sigma}_m) = -\mathrm{E}\left\{\frac{\partial^2 \ln \eta_{mt}(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_m \partial \boldsymbol{\sigma}_m^\top} \xi_{mt|T}\right\}$$

gives (6). The first derivatives of the function $\ln \eta_{mt} = \ln \eta_{mt}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\alpha}_m = (\boldsymbol{\nu}_m^\top \boldsymbol{\gamma}_m^\top \boldsymbol{\delta}_m^\top)^\top$ are given by

$$\begin{aligned}
\frac{\partial \ln \eta_{mt}}{\partial \boldsymbol{\nu}_m} &= \mathbf{\Theta}_m^{-1}(1)^\top \mathbf{\Omega}_m^{-1} \mathbf{u}_t, \\
\frac{\partial \ln \eta_{mt}}{\partial \mathbf{\Phi}_{mi}} &= \sum_{\ell=0}^{\infty} \mathbf{\Psi}_{m\ell}^\top \mathbf{\Omega}_m^{-1} \mathbf{u}_t \mathbf{y}_{t-i-\ell}^\top, \\
\frac{\partial \ln \eta_{mt}}{\partial \mathbf{\Theta}_{mj}} &= \sum_{\ell,n\geq 0} \mathbf{\Psi}_{mn} \mathbf{z}_{m,t-j-\ell-n} \mathbf{u}_t^\top \mathbf{\Omega}_m^{-1} \mathbf{\Psi}_{m\ell}
\end{aligned}$$

for all $i \in \{1, \dots, p\}$ and $j \in \{1, \dots, q\}$. The second derivatives of $\ln \eta_{mt}$ with respect to $\boldsymbol{\sigma}_m$ and the components of $\boldsymbol{\alpha}_m$ are

$$\begin{aligned}
\frac{\partial^2 \ln \eta_{mt}}{\partial \boldsymbol{\nu}_m \partial \boldsymbol{\sigma}_m^\top} &= \mathbf{u}_t^\top \otimes \mathbf{\Theta}_m^{-1}(1)^\top, \\
\frac{\partial^2 \ln \eta_{mt}}{\partial \mathbf{\Phi}_{mi}\, \partial \boldsymbol{\sigma}_m^\top} &= \sum_{\ell=0}^{\infty} (\mathbf{y}_{t-i-\ell} \mathbf{u}_t^\top) \otimes \mathbf{\Psi}_{m\ell}^\top, \\
\frac{\partial^2 \ln \eta_{mt}}{\partial \mathbf{\Theta}_{mj} \partial \boldsymbol{\sigma}_m^\top} &= \sum_{\ell,n\geq 0} \{\mathbf{\Psi}_{m\ell}^\top \otimes \mathbf{\Psi}_{mn}) \left(\mathbf{I}_K \otimes (\mathbf{z}_{m,t-j-\ell-n} \mathbf{u}_t^\top)\right\}.
\end{aligned}$$

Then we have

$$\begin{aligned}
\mathcal{F}_a(\boldsymbol{\nu}_m, \boldsymbol{\sigma}_m) &= -\mathrm{E}\left(\frac{\partial^2 \ln \eta_{mt}}{\partial \boldsymbol{\nu}_m \partial \boldsymbol{\sigma}_m^\top} \xi_{mt|T}\right) \\
&= -\mathrm{E}\{\mathbf{u}_t^\top \otimes \mathbf{\Theta}_m^{-1}(1)^\top \xi_{mt|T}\} \\
&= -\mathrm{E}(\mathbf{u}_t^\top \xi_{mt|T}) \otimes \mathbf{\Theta}_m^{-1}(1)^\top = \mathbf{0}.
\end{aligned}$$

This follows from

$$\mathrm{E}(\mathbf{u}_t \xi_{mt|T}) = \mathrm{E}(\boldsymbol{\Sigma}_{s_t} \boldsymbol{\epsilon}_t \xi_{mt|T}) = \mathrm{E}(\boldsymbol{\Sigma}_{s_t} \xi_{mt|T})\mathrm{E}(\boldsymbol{\epsilon}_t) = \mathbf{0}.$$

For all $i \in \{1, \ldots, p\}$, we get

$$\begin{aligned}\mathcal{F}_a(\boldsymbol{\Phi}_{mi}, \boldsymbol{\sigma}_m) &= -\mathrm{E}\left(\frac{\partial^2 \ln \eta_{mt}}{\partial \boldsymbol{\Phi}_{mi} \partial \boldsymbol{\sigma}_m^\top} \xi_{mt|T}\right) \\ &= -\sum_{\ell=0}^{\infty} \mathrm{E}(\mathbf{y}_{t-i-\ell} \mathbf{u}_t^\top \xi_{mt|T}) \otimes \boldsymbol{\Psi}_{m\ell}^\top = \mathbf{0}.\end{aligned}$$

In fact, by using (3) we see that

$$\mathrm{E}(\mathbf{y}_{t-i-\ell} \mathbf{u}_t^\top \xi_{mt|T}) = \sum_{n=0}^{\infty} \mathbf{B}_{mn} \mathrm{E}(\mathbf{u}_{t-i-\ell-n} \mathbf{u}_t^\top \xi_{mt|T}) = \mathbf{0}$$

for every $i \in \{1, \ldots, p\}$ and $\ell \geq 0$. Recall that

$$\begin{aligned}\mathrm{vec}\{\mathrm{E}(\mathbf{u}_{t-r} \mathbf{u}_t^\top \xi_{mt|T})\} &= \mathrm{E}\{(\boldsymbol{\Sigma}_{s_t} \otimes \boldsymbol{\Sigma}_{s_t})\, \mathrm{vec}(\boldsymbol{\epsilon}_{t-r} \boldsymbol{\epsilon}_t^\top) \xi_{mt|T}\} \\ &= \mathrm{E}\{(\boldsymbol{\Sigma}_{s_t} \otimes \boldsymbol{\Sigma}_{s_t}) \xi_{mt|T}\}\, \mathrm{vec}\, \mathrm{E}(\boldsymbol{\epsilon}_{t-r} \boldsymbol{\epsilon}_t^\top) = \mathbf{0}\end{aligned}$$

for every $r \geq 1$. Thus the $K^2 \times (pK^2)$ cross component $\mathcal{F}_a(\boldsymbol{\sigma}_m, \boldsymbol{\gamma}_m)$ is zero. For all $j \in \{1, \ldots, q\}$, we have

$$\begin{aligned}\mathcal{F}_a(\boldsymbol{\Theta}_{mj}, \boldsymbol{\sigma}_m) &= -\mathrm{E}\left\{\frac{\partial^2 \ln \eta_{mt}}{\partial \boldsymbol{\Theta}_{mj} \partial \boldsymbol{\sigma}_m^\top} \xi_{mt|T}\right\} \\ &= -\mathrm{E}\left\{\sum_{\ell,n \geq 0} \boldsymbol{\Psi}_{m\ell}^\top \otimes (\boldsymbol{\Psi}_{mn} \mathbf{z}_{m,t-j-\ell-n} \mathbf{u}_t^\top) \xi_{mt|T}\right\} \\ &= -\sum_{\ell,n \geq 0} (\boldsymbol{\Psi}_{m\ell}^\top \otimes \boldsymbol{\Psi}_{mn})\{\mathbf{I}_K \otimes \mathrm{E}(\mathbf{z}_{m,t-j-\ell-n} \mathbf{u}_t^\top \xi_{mt|T})\} = \mathbf{0}.\end{aligned}$$

This follows from the fact that

$$\begin{aligned}\mathrm{E}(\mathbf{z}_{m,t-j-\ell-n} \mathbf{u}_t^\top \xi_{mt|T}) &= \mathrm{E}\{\boldsymbol{\Theta}_m(L) \mathbf{u}_{t-j-\ell-n} \mathbf{u}_t^\top \xi_{mt|T}\} \\ &= \mathrm{E}\left(\sum_{h=0}^{\infty} \boldsymbol{\Psi}_{mh} \mathbf{u}_{t-j-\ell-n-h} \mathbf{u}_t^\top \xi_{mt|T}\right) \\ &= \sum_{h=0}^{\infty} \boldsymbol{\Psi}_{mh} \mathrm{E}(\mathbf{u}_{t-j-\ell-n-h} \mathbf{u}_t^\top \xi_{mt|T}) = \mathbf{0}\end{aligned}$$

for all $j \in \{1, \ldots, q\}$ and $\ell, n \geq 0$. Thus the $K^2 \times (qK^2)$ cross component $\mathcal{F}_a(\boldsymbol{\sigma}_m, \boldsymbol{\delta}_m)$ is zero. From above, the asymptotic covariance matrix of the Gaussian ML estimator $\hat{\boldsymbol{\sigma}}_m$ of $\boldsymbol{\sigma}_m$ is given by

$$\mathrm{var}_a(\hat{\boldsymbol{\sigma}}_m) = \frac{2}{\pi_m}(\boldsymbol{\Omega}_m^{-1} \otimes \boldsymbol{\Omega}_m^{-1}).$$

It follows that $\mathrm{var}_a(\hat{\boldsymbol{\Omega}}_m) = 2\pi_m{}^{-1}(\boldsymbol{\Omega}_m \otimes \boldsymbol{\Omega}_m)$. For example, consider the simple univariate ARMA(p, q), which is a special case with $M = 1$ and $K = 1$. Then the asymptotic variance of the estimated error variance $\hat{\boldsymbol{\Omega}} = \hat{\sigma}^2$ is $2\sigma^4$.

Appendix B. Example in Section 4

Let us consider the partitioned form of the asymptotic FI matrix $\mathcal{F}_a(\boldsymbol{\theta}_m)$ as given in the statement of the main theorem. The matrix expressions in that statement yield the following submatrices of $\mathcal{F}_a(\boldsymbol{\theta}_m)$, for $m \in \{1, 2\}$, rounded to 4 decimal places. From Eq. (6), we get

$$\mathcal{F}_a(\boldsymbol{\sigma}_1) = \begin{pmatrix} 0.0055 & 0.0109 & 0.0109 & 0.0218 \\ 0.0109 & 0.1227 & 0.0218 & 0.2455 \\ 0.0109 & 0.0218 & 0.1227 & 0.2455 \\ 0.0218 & 0.2455 & 0.2455 & 2.7614 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\sigma}_2) = \begin{pmatrix} 0.0582 & 0.0145 & 0.0145 & 0.0036 \\ 0.0145 & 0.1745 & 0.0036 & 0.0436 \\ 0.0145 & 0.0036 & 0.1745 & 0.0436 \\ 0.0036 & 0.0436 & 0.0436 & 0.5236 \end{pmatrix}.$$

The entries of the block matrices in the asymptotic FI submatrix $\mathcal{F}_a(\boldsymbol{\alpha}_m)$, for $m \in \{1, 2\}$, are computed according to Eqs. (7)–(11).

$$\mathcal{F}_a(\boldsymbol{\nu}_1) = \begin{pmatrix} 2.5038 & 0.2338 \\ 0.2338 & 0.0225 \end{pmatrix}, \qquad \mathcal{F}_a(\boldsymbol{\nu}_2) = \begin{pmatrix} 5.6843 & 11.2107 \\ 11.2107 & 32.2110 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\gamma}_1) = \begin{pmatrix} 1.3620 & 2.9200 & 0.1483 & 0.3179 \\ 2.9200 & 7.8598 & 0.3180 & 0.8561 \\ 0.1483 & 0.3180 & 0.0161 & 0.0346 \\ 0.3179 & 0.8561 & 0.0346 & 0.0933 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\gamma}_2) = \begin{pmatrix} 0.0454 & 0.0997 & 0.1312 & 0.2875 \\ 0.0997 & 0.2850 & 0.2874 & 0.8170 \\ 0.1312 & 0.2874 & 0.3815 & 0.8366 \\ 0.2875 & 0.8170 & 0.8366 & 2.3793 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\delta}_1) = \begin{pmatrix} 10.6115 & 2.2637 & 2.3129 & 0.4934 \\ 2.2637 & 0.4833 & 0.4934 & 0.1053 \\ 2.3129 & 0.4934 & 0.5041 & 0.1075 \\ 0.4934 & 0.1053 & 0.1075 & 0.0230 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\delta}_2) = \begin{pmatrix} 0.0216 & 0.0938 & 0.0956 & 0.4150 \\ 0.0938 & 0.4074 & 0.4151 & 1.8029 \\ 0.0956 & 0.4151 & 0.4339 & 1.8837 \\ 0.4150 & 1.8029 & 1.8837 & 8.1808 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\gamma}_1, \boldsymbol{\nu}_1) = \begin{pmatrix} 1.6598 & 0.1551 \\ 3.1018 & 0.2898 \\ 0.1807 & 0.0169 \\ 0.3376 & 0.0315 \end{pmatrix}, \qquad \mathcal{F}_a(\boldsymbol{\gamma}_2, \boldsymbol{\nu}_2) = \begin{pmatrix} 0.0484 & 0.1393 \\ 0.0948 & 0.2771 \\ 0.1457 & 0.4039 \\ 0.2903 & 0.8035 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\delta}_1, \boldsymbol{\gamma}_1) = \begin{pmatrix} 1.1827 & 2.8741 & 0.1292 & 0.3134 \\ 0.2439 & 0.5976 & 0.0267 & 0.0652 \\ 0.2576 & 0.6270 & 0.0282 & 0.0684 \\ 0.0531 & 0.1304 & 0.0058 & 0.0142 \end{pmatrix},$$

$$\mathcal{F}_a(\boldsymbol{\delta}_2, \boldsymbol{\gamma}_2) = \begin{pmatrix} -0.0192 & -0.0596 & -0.0558 & -0.1722 \\ -0.0843 & -0.2595 & -0.2456 & -0.7500 \\ -0.0843 & -0.2631 & -0.2475 & -0.7656 \\ -0.3710 & -1.1457 & -1.0893 & -3.3342 \end{pmatrix}.$$

References

[1] Y. Bao, Y. Hua, On the Fisher information matrix of a vector ARMA process, Econom. Lett. 123 (2014) 14–16.
[2] P.J. Bickel, Y. Ritov, T. Ryden, Asymptotic normality of the maximum likelihood estimator for general hidden Markov models, Ann. Statist. 26 (1998) 1614–1635.
[3] M. Cavicchioli, Analysis of the likelihood function for Markov switching VAR(CH) models, J. Time Series Anal. 35 (2014) 624–639.
[4] M. Cavicchioli, Higher order moments of Markov switching VARMA models, Econometric Theory (2017) http://dx.doi.org/10.1017/S0266466616000438, forthcoming.
[5] R. Douc, E. Moulines, T. Ryden, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, Ann. Statist. 32 (2004) 2254–2304.
[6] R.F. Engle, K.F. Kroner, Multivariate simultaneous generalized ARCH, Econometric Theory 11 (1995) 122–150.
[7] P.L. Fackler, Notes on Matrix Calculus, North Carolina State University, USA, 2005.
[8] C. Francq, J.-M. Zakoïan, Stationarity of multivariate Markov switching ARMA models, J. Econometrics 102 (2001) 339–364.
[9] G.M. Gallo, E. Otranto, Forecasting realized volatility with changing average levels, Int. J. Forecast. 31 (2015) 620–634.
[10] J.D. Hamilton, Analysis of time series subject to changes in regime, J. Econometrics 45 (1990) 39–70.
[11] J.D. Hamilton, Time Series Analysis, Princeton University Press, Princeton, NJ, 1994.
[12] E.J. Hannan, W.T.M. Dunsmuir, M. Deistler, Estimation of vector ARMAX models, J. Multivariate Anal. 10 (1980) 275–295.
[13] A. Hatefi, M. Jafari Jozani, Fisher information in different types of perfect and imperfect ranked set samples from finite mixture models, J. Multivariate Anal. 119 (2013) 16–31.
[14] S.H. Holan, R. Lund, G. Davis, The ARMA alphabet soup: A tour of ARMA model variants, Stat. Surv. 4 (2010) 232–274.
[15] C.J. Kim, Dynamic linear models with Markov-switching, J. Econometrics 60 (1994) 1–22.
[16] A. Klein, Matrix algebraic properties of the Fisher information matrix of stationary processes, Entropy 16 (2014) 2023–2055.
[17] A. Klein, G. Mélard, An algorithm for the exact Fisher information matrix of a vector ARMAX time series, Linear Algebra Appl. 446 (2014) 1–24.
[18] A. Klein, G. Mélard, A. Saidi, The asymptotic and exact Fisher information matrices of a vector ARMA process, Statist. Probab. Lett. 78 (2008) 1430–1433.
[19] A. Klein, G. Mélard, T. Zahaf, Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules, Linear Algebra Appl. 321 (2000) 209–232.
[20] A. Klein, P. Spreij, On Fisher's information matrix of an ARMA process, in: I. Csiszar, Gy. Michaletsky (Eds.), Stochastic Differential and Difference Equations, in: Progress in Systems and Control Theory, 23, Birkhauser, Boston, 1997, pp. 273–284.
[21] A. Klein, P. Spreij, An explicit expression for the Fisher information matrix of a multiple time series process, Linear Algebra Appl. 417 (2006) 140–149.
[22] A. Klein, P. Spreij, Matrix differential calculus applied to multiple stationary time series and an extended Whittle formula for information matrices, Linear Algebra Appl. 430 (2009) 674–691.

[23] H.M. Krolzig, Markov-Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis, Springer, Berlin, 1997.
[24] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer, Berlin, 2007.
[25] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, second ed., Wiley, Chichester, New York, 1999.
[26] H.J. Newton, The information matrices of the parameters of multiple mixed time series, J. Multivariate Anal. 8 (1978) 317–323.
[27] A. Velinov, W. Chen, Do stock prices reflect their fundamentals? New evidence in the aftermath of the financial crisis, J. Econ. Bus. 80 (2015) 1–20.
[28] M.P. Wand, Fisher information for generalized linear mixed models, J. Multivariate Anal. 98 (2007) 1412–1416.
[29] M. Yang, Some properties of vector autoregressive processes with Markov-switching coefficients, Econometric Theory 16 (2000) 23–43.
[30] P.A. Zadrozny, S. Mittnik, Kalman filtering methods for computing information matrices for time-invariant periodic and generally time-varying VARMA models and samples, Comput. Math. Appl. 28 (1994) 107–119.