



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

LINEAR ALGEBRA  
AND ITS  
APPLICATIONS

Linear Algebra and its Applications 403 (2005) 291–313

[www.elsevier.com/locate/laa](http://www.elsevier.com/locate/laa)

## On the resultant property of the Fisher information matrix of a vector ARMA process

André Klein <sup>a,\*</sup>, Guy Mélard <sup>b</sup>, Peter Spreij <sup>c</sup>

<sup>a</sup>*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11,  
1018 WB Amsterdam, The Netherlands*

<sup>b</sup>*ECARES CP 114, Université Libre de Bruxelles, Avenue Franklin Roosevelt 50,  
B-1050 Bruxelles, Belgium*

<sup>c</sup>*Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidersgracht 24,  
1018 TV Amsterdam, The Netherlands*

Received 28 February 2004; accepted 8 February 2005

Available online 22 March 2005

Submitted by P. Van Dooren

---

### Abstract

A matrix is called a multiple resultant matrix associated to two matrix polynomials when it becomes singular if and only if the two matrix polynomials have at least one common eigenvalue. In this paper a new multiple resultant matrix is introduced. It concerns the Fisher information matrix (FIM) of a stationary vector autoregressive and moving average time series process (VARMA). The two matrix polynomials are the autoregressive and the moving average matrix polynomials of the VARMA process. In order to show that the FIM is a multiple resultant matrix two new representations of the FIM are derived. To construct such representations appropriate matrix differential rules are applied. The newly obtained representations are expressed in terms of the multiple Sylvester matrix and the tensor Sylvester matrix. The representation of the FIM expressed by the tensor Sylvester matrix is used to prove that the FIM becomes singular if and only if the autoregressive and moving average matrix polynomials have at least one common eigenvalue. It then follows that the FIM and the tensor Sylvester matrix have equivalent singularity conditions. In a simple numerical example it is

---

\* Corresponding author. Tel.: +31 20 5254245; fax: +31 20 5254349.

E-mail addresses: [a.a.klein@uva.nl](mailto:a.a.klein@uva.nl) (A. Klein), [gmelard@ulb.ac.be](mailto:gmelard@ulb.ac.be) (G. Mélard), [spreij@science.uva.nl](mailto:spreij@science.uva.nl) (P. Spreij).

shown however that the FIM fails to detect common eigenvalues due to some kind of numerical instability. Whereas the tensor Sylvester matrix reveals it clearly, proving the usefulness of the results derived in this paper.

© 2005 Elsevier Inc. All rights reserved.

*AMS classification:* 15A23; 15A57; 15A69; 62B10; 62H12

*Keywords:* Matrix differential rules; Multiple resultant matrix; Tensor Sylvester matrix; Matrix polynomial; Singularity; Common eigenvalues; Fisher information matrix; VARMA process

---

## 1. Introduction

A multiple resultant matrix associated to two matrix polynomials is singular if and only if the two matrix polynomials have at least one common eigenvalue. Vector ARMA (autoregressive-moving average) or VARMA( $p, q$ ) stochastic processes are general-purpose representations in order to describe dynamic systems in engineering and in econometrics. From the formal definition given below, it will be clear that they depend on two matrix polynomials of degrees  $p$  and  $q$  which are called the orders and characterize the complexity of the representation. Statistical inference on the parameters of such models, the matrix coefficients, is largely based on the asymptotic Fisher information matrix. Indeed, the (asymptotic) covariance matrix of the parameter estimators is the inverse of that Fisher information matrix. Reliable bounds for the coefficients can therefore only be found if the Fisher information matrix is nonsingular.

In this paper it is proven that the asymptotic Fisher information matrix of a VARMA process possesses the multiple resultant property associated with the coefficient matrix polynomials.

For the purpose of that proof, new compact representations of the Fisher information matrix are derived in terms of structured matrices: the multiple Sylvester matrix and the tensor Sylvester matrix. Gohberg and Lerer [2] have shown that the tensor Sylvester matrix has the multiple resultant property but not the multiple Sylvester matrix. Using that property, it is shown that the Fisher information matrix becomes singular if and only if the tensor Sylvester matrix is singular, in other words, if and only if the autoregressive and moving average matrix polynomials of the VARMA process have at least one common eigenvalue. Therefore, by checking the singularity of the tensor Sylvester matrix, it can quickly be checked whether the Fisher information matrix is singular. In that case, the model orders  $p$  and/or  $q$  need to be adapted. That check can also be used before generating artificial time series from VARMA processes [22,1], in particular in Monte Carlo studies or application of the bootstrap.

Before going on, let us introduce the statistical model more formally and explain the general context of its application. Consider the  $n$ -dimensional mixed autoregressive moving average stationary time stochastic process  $\{y(t), t \in \mathbb{N}\}$  or VARMA process, of order  $(p, q)$  that satisfies,

$$\sum_{j=0}^p A_j y(t - j) = \sum_{k=0}^q B_k \varepsilon(t - k), \quad t \in \mathbb{N}, \tag{1}$$

where  $A_0 \equiv B_0 \equiv I_n$ , the  $n$ -dimensional identity matrix, and the white noise process  $\{\varepsilon(t), t \in \mathbb{N}\}$  is a  $n$ -dimensional vector random variable, such that

$$\mathbb{E}_\vartheta \{\varepsilon(t)\} = 0, \quad \mathbb{E}_\vartheta \{\varepsilon(s)\varepsilon^\top(t)\} = \delta_{st}\Sigma.$$

The symbol  $\mathbb{E}_\vartheta$  is the expected value under the parameter  $\vartheta$ , an appropriate representation of  $\vartheta$  which consists of the VARMA parameters is given in the next section,  $\top$  denotes transposition,  $\delta_{st}$  is the usual Kronecker delta and the covariance matrix  $\Sigma$  is positive definite.

The VARMA process can also be summarized as follows

$$A(L)y(t) = B(L)\varepsilon(t),$$

where the matrix polynomials  $A(\cdot)$  and  $B(\cdot)$  are given by  $A(L) = \sum_{j=0}^p A_j L^j$ ,  $B(L) = \sum_{k=0}^q B_k L^k$  and  $L$  is the backward-shift operator  $L^k y(t) = y(t - k)$ . We further assume that the eigenvalues of the matrix polynomials  $A(L)$  and  $B(L)$  lie outside the unit disc so the elements of  $A^{-1}(L)$  and  $B^{-1}(L)$  can be written as power series in  $L$  with convergence radius one. These eigenvalues are obtained by solving the scalar polynomials  $\det A(L) = 0$  and  $\det B(L) = 0$  of degree  $pn$  and  $qn$  respectively,  $\det X$  is the determinant of  $X$ .

The estimation of the  $(n \times n)$  matrices  $A_1, \dots, A_p, B_1, \dots, B_q$ , and  $\Sigma$  have received considerable attention in the time series and filtering of multiple time series literature [5,6]. The Fisher information matrix is of fundamental importance for describing the asymptotic covariance structure of the estimated parameters since this covariance matrix is obtained by inverting the Fisher information matrix. Consequently, only a nonsingular Fisher information matrix can produce reliable covariances of the estimated VARMA parameters. Algorithms for the asymptotic Fisher information matrix have been developed by several authors. Newton [18] has constructed an algorithm for the case of a VARMA process at the scalar-level, when one element of the matrix is considered, and is based on Whittle’s formula, see [25]. In [8,9], algorithms are presented for a wider class of scalar time series processes like the SISO (single-input–single-output) and MISO (multiple-input–single-output) structures. The Fisher information matrix is also extensively studied in the statistical signal processing literature, see for example Weiss and Friedlander [24], Scharf and McWhorther [21], Karlsson et al. [7].

To obtain the appropriate representations of the Fisher information matrix of a VARMA process, matrix differential rules applied in [13–15] are used. These rules will be recalled in Section 2 before describing the compact representations which are the main contributions of this paper. We introduce also two simple examples: Example 1, with common eigenvalues, and Example 2, without a common eigenvalue. Numerical experiments on these examples are discussed in Section 3. They first show that the multiple Sylvester matrix gives a bad answer in both cases and that the tensor

Sylvester matrix leads, of course, to the right conclusion. Furthermore, the Fisher information matrix computed numerically for Example 1 wrongly appears invertible, stressing the usefulness of the criterion based on the tensor Sylvester matrix.

## 2. Compact representations of the Fisher information matrix

### 2.1. Block matrix representations

Assume that  $\{y(t), t \in \mathbb{N}\}$  is a zero mean Gaussian time series. Then its stationary distribution depends on parameters  $\vartheta = (\vartheta_1, \dots, \vartheta_\ell)^\top$ , where  $\ell$  is the number of parameters of the vector autoregressive moving average model and is equal to  $n^2(p+q)$ . The choice for the parameter vector is  $\vartheta = \text{vec}\{A_1, \dots, A_p, B_1, \dots, B_q\}$ . The vec operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other,  $\text{vec } X = \text{col}(\text{col}(X_{ij})_{i=1}^n)_{j=1}^n$ , where  $\text{col}(X_{ij})_{i=1}^n$  refers to the  $j$ th column of the matrix  $X$  with elements  $X_{1j}, \dots, X_{nj}$ . When the representation of the parameter vector  $\vartheta$  as defined above is considered, the following equality holds for the  $n^2(p+q) \times n^2(p+q)$  asymptotic Fisher information matrix

$$\mathcal{F}(\vartheta) = \mathbb{E}_\vartheta \left\{ \left( \frac{\partial \varepsilon}{\partial \vartheta} \right)^\top \Sigma^{-1} \left( \frac{\partial \varepsilon}{\partial \vartheta} \right) \right\}, \quad (2)$$

where  $\partial \varepsilon / \partial \vartheta$  is with dimension  $n \times n^2(p+q)$ , and for simplicity  $t$  is omitted from  $\varepsilon(t)$  in the right-hand side of (2).

In this section we will derive two compact expressions for the Fisher information matrix of VARMA processes. Contrarily to Whittle's formula [25] and the algorithm developed in [18] which are both at the scalar-level, the Fisher information matrix developed in this paper is at the vector-matrix level, meaning that the matrix is considered as a whole, which is the only way to exhibit algebraic properties. For an efficient description of the blocks constituting  $\mathcal{F}(\vartheta)$  we decompose the vector parameter  $\vartheta$  accordingly to obtain,  $\vartheta = (\vartheta_a^\top \quad \vartheta_b^\top)^\top$  where  $\vartheta_a = \text{vec}\{A_1, \dots, A_p\}$ , and  $\vartheta_b = \text{vec}\{B_1, \dots, B_q\}$ . We shall proceed with the block representation of  $\mathcal{F}(\vartheta)$  which is given by

$$\mathcal{F}(\vartheta) = \begin{pmatrix} \mathcal{F}_{aa}(\vartheta) & \mathcal{F}_{ab}(\vartheta) \\ \mathcal{F}_{ba}(\vartheta) & \mathcal{F}_{bb}(\vartheta) \end{pmatrix}.$$

In a dynamic stationary stochastic context it has long been shown useful to use Fourier transform representations or, alternatively, circular integral representations, also called  $z$ -transform. We want to express  $\mathcal{F}(\vartheta)$  by integral representations like

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \mathcal{I}(z) \frac{dz}{z} = \frac{1}{2\pi i} \oint_{|z|=1} \begin{pmatrix} \mathcal{I}_{aa}(z) & \mathcal{I}_{ab}(z) \\ \mathcal{I}_{ba}(z) & \mathcal{I}_{bb}(z) \end{pmatrix} \frac{dz}{z}, \quad (3)$$

where the integration is counterclockwise around the unit circle. Explicit expressions for  $\mathcal{I}(z)$  or the blocks  $\mathcal{I}_{aa}(z)$ ,  $\mathcal{I}_{ab}(z)$ ,  $\mathcal{I}_{ba}(z)$  and  $\mathcal{I}_{bb}(z)$  will be given in our new representations (9), (15), (17). Representation (17) will allow us to prove the

resultant property of the Fisher information matrix of the VARMA process. In order to evaluate the blocks  $\mathcal{F}_{aa}(\vartheta)$ ,  $\mathcal{F}_{ab}(\vartheta) = \mathcal{F}_{ba}^\top(\vartheta)$  and  $\mathcal{F}_{bb}(\vartheta)$  matrix differential rules shall be applied, see [17].

We now evaluate  $\partial\varepsilon/\partial\vartheta_a$  and  $\partial\varepsilon/\partial\vartheta_b$  with dimension  $n \times n^2p$  and  $n \times n^2q$  respectively. For that purpose we rewrite the VARMA process as  $y(t) = A^{-1}(L)B(L)\varepsilon(t)$  and we derive a form for the  $n \times n^2(p + q)$  matrix  $\partial\varepsilon/\partial\vartheta$  which can be decomposed in two terms, one term is associated with the vector autoregressive part and the second term with the vector moving average part. This will allow appropriate expressions for the different blocks to be set forth.

### 2.2. Some differential rules

Consider a real, differentiable  $(m \times n)$  matrix function  $X(\vartheta)$  of a real  $(\ell \times 1)$  vector  $\vartheta = (\vartheta_1, \dots, \vartheta_\ell)^\top$ , where  $m, n$  and  $\ell$  are positive integers. Let  $(m \times n)$  matrices  $\partial_r X = (\partial X_{ij}/\partial\vartheta_r)$  with  $r = 1, \dots, \ell$  be the first order derivatives of  $X(\vartheta)$  in partial derivative form with  $X_{ij}$  being the first  $(i, j)$  element of  $X$ . Write  $dX_{ij} = \sum_{r=1}^{\ell} (\partial X_{ij}/\partial\vartheta_r) d\vartheta_r$ , where  $d\vartheta_r$  is an arbitrary perturbation of  $\vartheta_r$ . The  $(m \times n)$  matrix  $dX = (dX_{ij})$  is the differential form of the first order derivative  $X(\vartheta)$ . An expression in differential form can instantaneously be put into a partial derivative form by replacing  $d$  with  $\partial_r$  for  $r = 1, \dots, \ell$ . Let us vectorize  $X(\vartheta)$  then the  $(mn \times \ell)$  matrix  $\partial \text{vec } X(\vartheta)/\partial\vartheta$  is the gradient form of first order derivatives of  $X(\vartheta)$  and can be defined as  $\text{vec } dX(\vartheta) = (\partial(\text{vec } X(\vartheta))/\partial\vartheta) d\vartheta = \text{dvec } X(\vartheta)$ .

Let  $X(\vartheta)$  and  $Y(\vartheta)$  be real  $(m \times n)$  and  $(n \times p)$  differentiable matrix functions of the real vector  $\vartheta(\ell \times 1)$ , where  $m, n, p$ , and  $\ell$  are positive integers. The usual scalar product rule of differentiation yields

$$d(XY) = (dX)Y + X(dY).$$

The following properties are taken into account. The first property to be considered is  $\partial y(t)/\partial\vartheta = 0$ , this holds because the given realization of  $y(t)$  is independent of variations in  $\vartheta$ , and as a second property the next differential rule is used

$$dA^{-1}(L) = -A^{-1}(L)dA(L)A^{-1}(L).$$

This enables us to formulate the following equation for the VARMA process

$$d\varepsilon = B^{-1}(L)dA(L)A^{-1}(L)B(L)\varepsilon - B^{-1}(L)dB(L)\varepsilon.$$

Recall the rule

$$\text{vec } ABC = (C^\top \otimes A)\text{vec } B,$$

where  $\otimes$  denotes the Kronecker product,  $A, B$  and  $C$  have appropriate dimensions. Componentwise application of this rule yields for  $d\varepsilon$

$$\begin{aligned} d\varepsilon &= ((A^{-1}(L)B(L)\varepsilon)^\top \otimes B^{-1}(L))\text{vec } dA(L) - (\varepsilon^\top \otimes B^{-1}(L))\text{vec } dB(L) \\ &= \{(A^{-1}(L)B(L)\varepsilon)^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial\vartheta} d\vartheta - (\varepsilon^\top \otimes B^{-1}(L)) \frac{\partial \text{vec } B(L)}{\partial\vartheta} d\vartheta. \end{aligned}$$

Consequently, we obtain

$$\frac{\partial \varepsilon}{\partial \vartheta} = \{(A^{-1}(L)B(L)\varepsilon)^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial \vartheta} - (\varepsilon^\top \otimes B^{-1}(L)) \frac{\partial \text{vec } B(L)}{\partial \vartheta}.$$

This representation of  $\partial \varepsilon / \partial \vartheta$  can be summarized as

$$\frac{\partial \varepsilon}{\partial \vartheta} = \left( \frac{\partial \varepsilon}{\partial \vartheta_a} : 0_{n \times n^2 q} \right) + \left( 0_{n \times n^2 p} : \frac{\partial \varepsilon}{\partial \vartheta_b} \right) = \begin{pmatrix} \frac{\partial \varepsilon}{\partial \vartheta_a} & \frac{\partial \varepsilon}{\partial \vartheta_b} \end{pmatrix}.$$

Note that

$$\begin{aligned} \frac{\partial \varepsilon}{\partial \vartheta_a} &= \{(A^{-1}(L)B(L)\varepsilon)^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial \vartheta_a} \quad \text{and} \\ \frac{\partial \varepsilon}{\partial \vartheta_b} &= -(\varepsilon^\top \otimes B^{-1}(L)) \frac{\partial \text{vec } B(L)}{\partial \vartheta_b}. \end{aligned}$$

### 2.3. Representation with reordered factors

An appropriate representation for the four blocks which compose  $\mathcal{F}(\vartheta)$  can then be set forth by applying formula (2). We shall use block  $\mathcal{F}_{aa}(\vartheta)$  to illustrate how the representations of the blocks are obtained. For that purpose a useful equality is introduced. Consider the discrete-time stationary process  $x(t)$  where  $x(t) = H(L)u(t)$  and  $H(L)$  is an asymptotically stable filter. For evaluating the covariance matrix of  $x(t)$ , the following equation holds true

$$\mathbb{E}_\vartheta \{x(t)x^\top(t)\} = \int_{-\pi}^\pi \Phi_x(\omega) d\omega, \tag{4}$$

where  $\Phi_x(\omega)$  is the spectral density of the process  $x(t)$  and is defined as  $\Phi_x(\omega) = H(e^{i\omega})\Phi_u(\omega)H(e^{-i\omega})^\top$ , with  $\Phi_u(\omega)$  being the spectral density of the input process  $u(t)$ . In order to apply equality (4) to block  $\mathcal{F}_{aa}(\vartheta)$  that is given by

$$\mathcal{F}_{aa}(\vartheta) = \mathbb{E}_\vartheta \left\{ \left( \frac{\partial \varepsilon}{\partial \vartheta_a} \right)^\top \Sigma^{-1} \left( \frac{\partial \varepsilon}{\partial \vartheta_a} \right) \right\}, \tag{5}$$

we rearrange the elements of the right-hand side of (5) so that representation  $x(t)x^\top(t)$  is obtained. For that purpose the rule

$$(A_1 \otimes B_1)(A_2 \otimes B_2) \cdots (A_m \otimes B_m) = (A_1 A_2 \cdots A_m) \otimes (B_1 B_2 \cdots B_m) \tag{6}$$

is used, where the matrices  $A_1, A_2, \dots, A_m$  and  $B_1, B_2, \dots, B_m$  have appropriate dimensions, see e.g. [16]. Note that we will use precedence of the Kronecker product over the matrix product and, consequently, omit the parentheses in the right-hand side of (6). We therefore rewrite  $\partial \varepsilon / \partial \vartheta_a$  accordingly, to obtain

$$(\varepsilon^\top \otimes I_n) \{(A^{-1}(L)B(L))^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial \vartheta_a}.$$

We now can write

$$\begin{aligned} & \left(\frac{\partial \varepsilon}{\partial \vartheta_a}\right)^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta_a}\right) \\ &= \left(\frac{\partial \text{vec } A(L)}{\partial \vartheta_a}\right)^\top \{A^{-1}(L)B(L) \otimes B^{-\top}(L)\} (\varepsilon \otimes I_n) \Sigma^{-1} (\varepsilon^\top \otimes I_n) \\ & \quad \times \{(A^{-1}(L)B(L))^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial \vartheta_a} \\ &= \left(\frac{\partial \text{vec } A(L)}{\partial \vartheta_a}\right)^\top \{A^{-1}(L)B(L) \otimes B^{-\top}(L)\} (\varepsilon \otimes \Sigma^{-1}) (\varepsilon^\top \otimes I_n) \\ & \quad \times \{(A^{-1}(L)B(L))^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial \vartheta_a}. \end{aligned}$$

In order to obtain a symmetric expression we apply a Cholesky factorization to  $\Sigma^{-1}$ . Since  $\Sigma$  is positive definite we can write  $\Sigma^{-1} = \Gamma \Gamma^\top$ , where  $\Gamma$  is a unique lower triangular matrix with positive diagonal entries. To obtain

$$\begin{aligned} & \mathbb{E}_\vartheta \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta_a}\right)^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta_a}\right) \right\} \\ &= \mathbb{E}_\vartheta \left\{ \left(\frac{\partial \text{vec } A(L)}{\partial \vartheta_a}\right)^\top \{A^{-1}(L)B(L) \otimes B^{-\top}(L)\} (\varepsilon \otimes \Gamma) (\varepsilon \otimes \Gamma)^\top \right. \\ & \quad \left. \times \{(A^{-1}(L)B(L))^\top \otimes B^{-1}(L)\} \frac{\partial \text{vec } A(L)}{\partial \vartheta_a} \right\}. \end{aligned}$$

#### 2.4. First integral representation

We have now a similar representation to the left-hand side of (4) where

$$x(t) = \left(\frac{\partial \text{vec } A(L)}{\partial \vartheta_a}\right)^\top \{A^{-1}(L)B(L) \otimes B^{-\top}(L)\} (\varepsilon \otimes \Gamma).$$

The next step consists of formulating the spectral density of  $(\varepsilon \otimes \Gamma)$ . For that purpose the corresponding covariance matrix has to be computed, to obtain

$$\mathbb{E}_\vartheta \{(\varepsilon \otimes \Gamma)(\varepsilon \otimes \Gamma)^\top\} = \mathbb{E}_\vartheta \{\varepsilon \varepsilon^\top \otimes \Gamma \Gamma^\top\} = \Sigma \otimes \Sigma^{-1}.$$

Since the white noise process  $\varepsilon$  has a constant spectral density (independent of the frequency  $\omega$ ), then it is straightforward to conclude that in view of (4) the value of the spectral density of  $(\varepsilon \otimes \Gamma)$  is  $(1/2\pi)(\Sigma \otimes \Sigma^{-1})$ . As a consequence, in view of (4), the matrix block  $\mathcal{F}_{aa}(\vartheta)$  can now be given by the integral expression

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \text{vec } A(e^{i\omega})}{\partial \vartheta_a}\right)^\top \{A^{-1}(e^{i\omega})B(e^{i\omega}) \otimes B^{-\top}(e^{i\omega})\} (\Sigma \otimes \Sigma^{-1})$$

$$\begin{aligned}
& \times \left\{ \left( A^{-1}(e^{-i\omega})B(e^{-i\omega}) \right)^\top \otimes B^{-1}(e^{-i\omega}) \right\} \left( \frac{\partial \text{vec } A(e^{-i\omega})}{\partial \vartheta_a} \right) d\omega \\
& = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{\partial \text{vec } A(e^{i\omega})}{\partial \vartheta_a} \right)^\top \left\{ A^{-1}(e^{i\omega})B(e^{i\omega})\Sigma \otimes B^{-\top}(e^{i\omega})\Sigma^{-1} \right\} \\
& \quad \times \left\{ \left( A^{-1}(e^{-i\omega})B(e^{-i\omega}) \right)^\top \otimes B^{-1}(e^{-i\omega}) \right\} \left( \frac{\partial \text{vec } A(e^{-i\omega})}{\partial \vartheta_a} \right) d\omega,
\end{aligned}$$

to obtain

$$\begin{aligned}
& \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{\partial \text{vec } A(e^{i\omega})}{\partial \vartheta_a} \right)^\top \left\{ A^{-1}(e^{i\omega})B(e^{i\omega})\Sigma(A^{-1}(e^{-i\omega})B(e^{-i\omega}))^\top \right. \\
& \quad \left. \otimes B^{-\top}(e^{i\omega})\Sigma^{-1}B^{-1}(e^{-i\omega}) \right\} \left( \frac{\partial \text{vec } A(e^{-i\omega})}{\partial \vartheta_a} \right) d\omega.
\end{aligned}$$

It can be seen that the integrand, the spectral density of the derived representation of  $x(t)$ , is Hermitian. Equivalently for  $z = e^{i\omega}$  we have

$$\begin{aligned}
& \mathcal{F}_{aa}(\vartheta) \\
& = \frac{1}{2\pi i} \oint_{|z|=1} \left( \frac{\partial \text{vec } A(z)}{\partial \vartheta_a} \right)^\top \left( A^{-1}(z)B(z)\Sigma B^\top(z^{-1})A^{-\top}(z^{-1}) \right. \\
& \quad \left. \otimes B^{-\top}(z)\Sigma^{-1}B^{-1}(z^{-1}) \right) \left( \frac{\partial \text{vec } A(z^{-1})}{\partial \vartheta_a} \right) \frac{dz}{z}.
\end{aligned}$$

Analogously for the remaining blocks, to obtain

$$\begin{aligned}
& \mathcal{F}_{ab}(\vartheta) = \mathbb{E}_\vartheta \left\{ \left( \frac{\partial \varepsilon}{\partial \vartheta_a} \right)^\top \Sigma^{-1} \left( \frac{\partial \varepsilon}{\partial \vartheta_b} \right) \right\} \\
& = -\frac{1}{2\pi i} \oint_{|z|=1} \left( \frac{\partial \text{vec } A(z)}{\partial \vartheta_a} \right)^\top \\
& \quad \times \left( A^{-1}(z)B(z)\Sigma \otimes B^{-\top}(z)\Sigma^{-1}B^{-1}(z^{-1}) \right) \left( \frac{\partial \text{vec } B(z^{-1})}{\partial \vartheta_b} \right) \frac{dz}{z},
\end{aligned}$$

$$\begin{aligned}
& \overline{\mathcal{F}}_{ba}(\vartheta) \\
& = \mathbb{E}_\vartheta \left\{ \left( \frac{\partial \varepsilon}{\partial \vartheta_b} \right)^\top \Sigma^{-1} \left( \frac{\partial \varepsilon}{\partial \vartheta_a} \right) \right\} = -\frac{1}{2\pi i} \oint_{|z|=1} \left( \frac{\partial \text{vec } B(z)}{\partial \vartheta_b} \right)^\top \\
& \quad \times \left( \Sigma B^\top(z^{-1})A^{-\top}(z^{-1}) \otimes B^{-\top}(z)\Sigma^{-1}B^{-1}(z^{-1}) \right) \left( \frac{\partial \text{vec } A(z^{-1})}{\partial \vartheta_a} \right) \frac{dz}{z}
\end{aligned}$$



and

$$\begin{aligned} \mathcal{F}_{bb}(\vartheta) &= \mathbb{E}_{\vartheta} \left\{ \left( \frac{\partial \varepsilon}{\partial \vartheta_b} \right)^\top \Sigma^{-1} \left( \frac{\partial \varepsilon}{\partial \vartheta_b} \right) \right\} = \frac{1}{2\pi i} \oint_{|z|=1} \left( \frac{\partial \text{vec } B(z)}{\partial \vartheta_b} \right)^\top \\ &\quad \times \left( \Sigma \otimes B^{-\top}(z) \Sigma^{-1} B^{-1}(z^{-1}) \right) \left( \frac{\partial \text{vec } B(z^{-1})}{\partial \vartheta_b} \right) \frac{dz}{z}. \end{aligned}$$

The representation of the parameter vector  $\vartheta$  leads to the equalities

$$\frac{\partial \text{vec } A(z)}{\partial \vartheta_a} = zu_p^\top(z) \otimes I_{n^2} \quad \text{and} \quad \frac{\partial \text{vec } B(z)}{\partial \vartheta_b} = zu_q^\top(z) \otimes I_{n^2}, \tag{7}$$

where  $u_x^\top(z) = (1, z, z^2, \dots, z^{x-1})$  for positive integers  $x$ .

### 2.5. Second integral representation

Before proceeding with additional developments, the next property is set forth. Consider the partitioned matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

then the Kronecker product  $A \otimes B$  takes the form

$$A \otimes B = \begin{pmatrix} A_{11} \otimes B & A_{12} \otimes B \\ A_{21} \otimes B & A_{22} \otimes B \end{pmatrix}. \tag{8}$$

The rules (6) and (8) as well as the properties given in the right-hand sides of the equalities in (7) are applied to the newly obtained representations of the blocks  $\mathcal{F}_{aa}(\vartheta)$ ,  $\mathcal{F}_{ab}(\vartheta)$ ,  $\mathcal{F}_{ba}(\vartheta)$  and  $\mathcal{F}_{bb}(\vartheta)$  which constitute  $\mathcal{F}(\vartheta)$ . Consequently, the Fisher information matrix can be written as

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \begin{pmatrix} F_{aa}(z) & F_{ab}(z) \\ F_{ba}(z) & F_{bb}(z) \end{pmatrix} \otimes B^{-\top}(z) \Sigma^{-1} B^{-1}(z^{-1}) \frac{dz}{z}, \tag{9}$$

where

$$\begin{aligned} F_{aa}(z) &= u_p(z) u_p^\top(z^{-1}) \otimes A^{-1}(z) B(z) \Sigma B^\top(z^{-1}) A^{-\top}(z^{-1}), \\ F_{ab}(z) &= -\{u_p(z) u_q^\top(z^{-1}) \otimes A^{-1}(z) B(z) \Sigma\}, \\ F_{ba}(z) &= -\{u_q(z) u_p^\top(z^{-1}) \otimes \Sigma B^\top(z^{-1}) A^{-\top}(z^{-1})\}, \\ F_{bb}(z) &= u_q(z) u_q^\top(z^{-1}) \otimes \Sigma. \end{aligned}$$

The matrix  $\begin{pmatrix} F_{aa}(z) & F_{ab}(z) \\ F_{ba}(z) & F_{bb}(z) \end{pmatrix}$  can then be set forth accordingly, to obtain

$$\begin{pmatrix} u_p(z) \otimes A^{-1}(z) (-B(z)) \\ u_q(z) \otimes A^{-1}(z) A(z) \end{pmatrix} \Sigma \begin{pmatrix} u_p(z) \otimes A^{-1}(z) (-B(z)) \\ u_q(z) \otimes A^{-1}(z) A(z) \end{pmatrix}^*, \tag{10}$$

where  $Y^*$  denotes the complex conjugate transpose of the matrix  $Y$ . It can be verified that representation (9) can also be obtained when  $\partial\varepsilon/\partial\vartheta$  is substituted in (2). The block matrices in (10) when multiplied with  $I_p$  and  $I_q$  can be rewritten as

$$I_p u_p(z) \otimes A^{-1}(z)(-B(z)) = (I_p \otimes A^{-1}(z))(u_p(z) \otimes (-B(z))), \tag{11}$$

$$I_q u_q(z) \otimes A^{-1}(z)A(z) = (I_q \otimes A^{-1}(z))(u_q(z) \otimes A(z)). \tag{12}$$

2.6. Representation based on the multiple Sylvester matrix

In order to write the Fisher information matrix  $\mathcal{F}(\vartheta)$  in a compact form and in terms of a structured matrix, we introduce the matrix block version of the Sylvester matrix which is given by the  $n(p + q) \times n(p + q)$  matrix

$$S(-B, A) = \begin{pmatrix} -I_n & -B_1 & \cdots & -B_q & 0_{n \times n} & \cdots & 0_{n \times n} \\ 0_{n \times n} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n \times n} \\ 0_{n \times n} & \cdots & 0_{n \times n} & -I_n & -B_1 & \cdots & -B_q \\ I_n & A_1 & \cdots & A_p & 0_{n \times n} & \cdots & 0_{n \times n} \\ 0_{n \times n} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0_{n \times n} \\ 0_{n \times n} & \cdots & 0_{n \times n} & I_n & A_1 & \cdots & A_p \end{pmatrix}.$$

In the scalar case the Sylvester matrix  $S(-b, a)$  associated with the polynomials  $a(z)$  and  $b(z)$  and having the form given above is called a resultant matrix. A resultant matrix of two scalar polynomials  $a(z)$  and  $b(z)$  becomes singular if and only if the polynomials  $a(z)$  and  $b(z)$  have at least one common root, see e.g. [16]. The number of common roots (counting multiplicities) of the polynomials  $a(z)$  and  $b(z)$  is equal to  $\dim \text{Ker } S(-b, a)$ .

However, in the matrix polynomial case the Sylvester matrix  $S(-B, A)$  does not have the same property as in the scalar case. We illustrate this with two examples. Consider the matrix polynomials  $A(z)$  and  $B(z)$  given by:

Example 1:  $A(z) = \begin{pmatrix} 1 + 0.6z & 0.2z \\ 0.4z & 1 - 0.6z \end{pmatrix}$  and  $B(z) = \begin{pmatrix} 1 + 0.5z & 0.76z \\ 0.25z & 1 - 0.5z \end{pmatrix}.$

This choice of  $A(z)$  and  $B(z)$  does not result in a singular matrix  $S(-B, A)$  despite the fact that the corresponding eigenvalues of both matrix polynomials  $A(z)$  and  $B(z)$  coincide and are equal to  $-1.50756$  and  $1.50756$ . These eigenvalues are obtained by solving the scalar polynomials  $\det A(z) = 0$  and  $\det B(z) = 0$  so that  $\det A(z) = \det B(z) = 1 - \frac{11}{25}z^2$ . Analogously, consider the matrix polynomials given by:

Example 2:  $A(z) = \begin{pmatrix} 1 - 0.8z & 0.2z \\ -1.2z & 1 - 0.2z \end{pmatrix}$  and  $B(z) = \begin{pmatrix} 1 & z \\ -0.5z & 1 + 0.5z \end{pmatrix}$ .

In this case the matrix  $S(-B, A)$  is not invertible although the matrix polynomials  $A(z)$  and  $B(z)$  do not have common eigenvalues.

It is clear that when the Fisher information matrix  $\mathcal{F}(\vartheta)$  is expressed in terms of the Sylvester matrix  $S(-B, A)$ , one will not get insight in the singularity condition of  $\mathcal{F}(\vartheta)$ . Despite this property we shall proceed constructing a representation of  $\mathcal{F}(\vartheta)$  in terms of  $S(-B, A)$ . This can be justified since the purpose of this paper also consists of developing new representations of the Fisher information matrix  $\mathcal{F}(\vartheta)$  that are expressed in terms of known structured matrices.

It can be shown through matrix multiplication that

$$S(-B, A) (u_{p+q}(z) \otimes I_n) = \begin{pmatrix} u_p(z) \otimes (-B(z)) \\ u_q(z) \otimes A(z) \end{pmatrix}. \tag{13}$$

Eqs. (11) and (12) combined with (13), results in the following form for the first matrix in the right-hand side of (10)

$$\begin{aligned} & \begin{pmatrix} u_p(z) \otimes A^{-1}(z)(-B(z)) \\ u_q(z) \otimes A^{-1}(z)A(z) \end{pmatrix} \\ &= \begin{pmatrix} I_p \otimes A^{-1}(z) & 0_{pn \times qn} \\ 0_{qn \times pn} & I_q \otimes A^{-1}(z) \end{pmatrix} \begin{pmatrix} u_p(z) \otimes (-B(z)) \\ u_q(z) \otimes A(z) \end{pmatrix} \\ &= (I_{p+q} \otimes A^{-1}(z)) S(-B, A) (u_{p+q}(z) \otimes I_n). \end{aligned} \tag{14}$$

Combining (9), (10) and (14) leads to a compact form of the Fisher information matrix in terms of the structured matrix  $S(-B, A)$  which is given in the following proposition.

**Proposition 2.1.** *The Fisher information matrix of a VARMA process when expressed in terms of  $S(-B, A)$  has the following representation*

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \Psi(z) \Sigma \Psi^*(z) \otimes B^{-\top}(z) \Sigma^{-1} B^{-1}(z^{-1}) \frac{dz}{z}, \tag{15}$$

where

$$\Psi(z) = (I_{p+q} \otimes A^{-1}(z)) S(-B, A) (u_{p+q}(z) \otimes I_n).$$

The components of  $\mathcal{F}(\vartheta)$  can be computed from (15) by applying Cauchy’s residue theorem to a Hermitian matrix polynomial. For each element of this matrix polynomial it consists of evaluating integrals of a rational function over the unit circle in the complex plane, the algorithm of Peterka and Vidinčev [19] can be

applied. However, contrariwise to the scalar ARMA case [12], from (15) we cannot infer that the Fisher information matrix becomes singular if the Sylvester matrix  $S(-B, A)$  becomes singular or vice versa. However, we will show below that when the matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue the Fisher information matrix  $\mathcal{F}(\vartheta)$  becomes singular irrespective of the nonsingularity of the matrix  $S(-B, A)$ . It can be concluded that the singularity condition of  $\mathcal{F}(\vartheta)$  is hidden when representation (15) is considered. To characterize singularity of  $\mathcal{F}(\vartheta)$  a new resultant matrix will be used.

2.7. Representation based on the tensor Sylvester matrix

Gohberg and Lerer [2] have set forth the tensor resultant  $S^{\otimes}(-B, A) \triangleq S(-B \otimes I_n, I_n \otimes A)$  and proved that the matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue if and only if  $\det S^{\otimes}(-B, A) = 0$  or when the matrix  $S^{\otimes}(-B, A)$  is singular. In other words, the tensor resultant  $S^{\otimes}(-B, A)$  becomes singular if and only if the scalar polynomials  $\det A(z) = 0$  and  $\det B(z) = 0$  have at least one common root. The  $n^2(p + q) \times n^2(p + q)$  tensor Sylvester matrix is given by

$$S^{\otimes}(-B, A) = \begin{pmatrix} (-I_n) \otimes I_n & (-B_1) \otimes I_n & \cdots & (-B_q) \otimes I_n & 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} \\ 0_{n^2 \times n^2} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0_{n^2 \times n^2} \\ 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} & (-I_n) \otimes I_n & (-B_1) \otimes I_n & \cdots & (-B_q) \otimes I_n \\ I_n \otimes I_n & I_n \otimes A_1 & \cdots & I_n \otimes A_p & 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} \\ 0_{n^2 \times n^2} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0_{n^2 \times n^2} \\ 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} & I_n \otimes I_n & I_n \otimes A_1 & \cdots & I_n \otimes A_p \end{pmatrix}.$$

For inserting  $S^{\otimes}(-B, A)$  in (3) we rewrite the integrand of (9) accordingly, to obtain

$$\begin{pmatrix} \mathcal{J}_{aa}(z) & \mathcal{J}_{ab}(z) \\ \mathcal{J}_{ba}(z) & \mathcal{J}_{bb}(z) \end{pmatrix} = \begin{pmatrix} u_p(z) \otimes A^{-1}(z)(-B(z)) \otimes I_n \\ u_q(z) \otimes A^{-1}(z)A(z) \otimes I_n \end{pmatrix} \left( \Sigma \otimes B^{-\top}(z) \Sigma^{-1} B^{-1}(z^{-1}) \right) \times \begin{pmatrix} u_p(z) \otimes A^{-1}(z)(-B(z)) \otimes I_n \\ u_q(z) \otimes A^{-1}(z)A(z) \otimes I_n \end{pmatrix}^* \tag{16}$$

Next some property of the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  is given. It is straightforward to verify that (13) can be extended to the representation

$$S^{\otimes}(-B, A)(u_{p+q}(z) \otimes I_{n^2}) = \begin{pmatrix} u_p(z) \otimes (-B(z)) \otimes I_n \\ u_q(z) \otimes I_n \otimes A(z) \end{pmatrix}.$$

We now proceed with the first matrix term in the right-hand side of (16) which can be rewritten when the rule

$$A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C)$$

is used, as

$$\begin{pmatrix} I_p u_p(z) \otimes (A^{-1}(z) \otimes I_n)(-B(z) \otimes I_n) \\ I_q u_q(z) \otimes (A^{-1}(z) \otimes I_n)(A(z) \otimes I_n) \end{pmatrix}.$$

Taking the property

$$(A^{-1}(z) \otimes I_n)(A(z) \otimes I_n) = I_n \otimes I_n = (I_n \otimes A^{-1}(z))(I_n \otimes A(z))$$

into account, the first matrix block of (16) can be represented as

$$\begin{aligned} & \begin{pmatrix} (I_p \otimes A^{-1}(z) \otimes I_n) & (u_p(z) \otimes (-B(z) \otimes I_n)) \\ (I_q \otimes I_n \otimes A^{-1}(z)) & (u_q(z) \otimes I_n \otimes A(z)) \end{pmatrix} \\ &= \begin{pmatrix} I_p \otimes A^{-1}(z) \otimes I_n & 0_{pn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & I_q \otimes I_n \otimes A^{-1}(z) \end{pmatrix} \begin{pmatrix} u_p(z) \otimes (-B(z) \otimes I_n) \\ u_q(z) \otimes I_n \otimes A(z) \end{pmatrix} \\ &= \begin{pmatrix} I_p \otimes A^{-1}(z) \otimes I_n & 0_{pn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & I_q \otimes I_n \otimes A^{-1}(z) \end{pmatrix} S^{\otimes}(-B, A) (u_{p+q}(z) \otimes I_{n^2}). \end{aligned}$$

This derivation leads to an alternative representation for the Fisher information matrix  $\mathcal{F}(\vartheta)$  that will be given in the next proposition.

**Proposition 2.2.** *The Fisher information matrix of a VARMA process when expressed in terms of the tensor Sylvester matrix has the following representation*

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} \Phi(z)\Theta(z)\Phi^*(z) \frac{dz}{z}, \tag{17}$$

where

$$\Phi(z) = \begin{pmatrix} I_p \otimes A^{-1}(z) \otimes I_n & 0_{pn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & I_q \otimes I_n \otimes A^{-1}(z) \end{pmatrix} S^{\otimes}(-B, A) (u_{p+q}(z) \otimes I_{n^2})$$

and

$$\Theta(z) = \Sigma \otimes B^{-\top}(z)\Sigma^{-1}B^{-1}(z^{-1}).$$

### 2.8. New multiple resultant matrices

The representation of the Fisher information matrix  $\mathcal{F}(\vartheta)$  in (17) shall be used for proving that  $\mathcal{F}(\vartheta)$  will become singular if and only if the autoregressive and moving average matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue. In other words, the matrix  $\mathcal{F}(\vartheta)$  will become singular if and only if the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  is singular. We proceed as follows. Set

$$\Phi(z)\Theta(z)\Phi^*(z) = \mathcal{I}(z)$$

and

$$A(z) = \begin{pmatrix} I_p \otimes A(z) \otimes I_n & 0_{pn^2 \times qn^2} \\ 0_{qn^2 \times pn^2} & I_q \otimes I_n \otimes A(z) \end{pmatrix},$$

to obtain

$$A(z)\mathcal{F}(z)A^*(z) = S^\otimes(-B, A) (u_{p+q}(z) \otimes I_{n^2}) \Theta(z) (u_{p+q}(z) \otimes I_{n^2})^* \\ \times [S^\otimes(-B, A)]^\top.$$

Then

$$\mathcal{M}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} A(z)\mathcal{F}(z)A^*(z) \frac{dz}{z} \\ = S^\otimes(-B, A) \mathcal{P}(\vartheta) [S^\otimes(-B, A)]^\top, \quad (18)$$

where

$$\mathcal{P}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} (u_{p+q}(z) \otimes I_{n^2}) \Theta(z) (u_{p+q}(z) \otimes I_{n^2})^* \frac{dz}{z}.$$

In [12], a representation similar to (18) is derived to prove the resultant property of the Fisher information matrix of a scalar ARMA time series process. For establishing the resultant property of the Fisher information matrix  $\mathcal{F}(\vartheta)$  of a VARMA process, we first establish the resultant property of  $\mathcal{M}(\vartheta)$ . This will be summarized in the next lemma.

**Lemma 2.3.** *The matrix  $\mathcal{M}(\vartheta)$  as formulated in (18) becomes singular iff the matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue.*

**Proof.** Clearly the matrix  $\mathcal{M}(\vartheta)$  becomes singular if the matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue in view of Eq. (18) and the resultant property of  $S^\otimes(-B, A)$ . In order to prove the converse, it suffices to prove that  $\mathcal{P}(\vartheta)$  is strictly positive definite or  $\mathcal{P}(\vartheta) > 0$ . This can be shown via the following computation. Suppose that there is a fixed vector  $x$  such that  $\mathcal{P}(\vartheta)x = 0$ . So,

$$0 = \frac{1}{2\pi i} \oint_{|z|=1} (u_{p+q}(z) \otimes I_{n^2}) \Theta(z) (u_{p+q}(z) \otimes I_{n^2})^* x \frac{dz}{z}.$$

Take  $z = e^{i\omega}$ , we then get

$$0 = \frac{1}{2\pi} \int_0^{2\pi} x^* (u_{p+q}(e^{i\omega}) \otimes I_{n^2}) \Theta(e^{i\omega}) (u_{p+q}(e^{i\omega}) \otimes I_{n^2})^* x d\omega.$$

Note that  $(u_{p+q}(e^{i\omega}) \otimes I_{n^2}) \Theta(e^{i\omega}) (u_{p+q}(e^{i\omega}) \otimes I_{n^2})^* \geq 0$ .

Then we must have  $x^* (u_{p+q}(e^{i\omega}) \otimes I_{n^2}) \Theta(e^{i\omega}) \equiv 0$ , but since  $\Theta(e^{i\omega}) > 0$  we have  $x^* (u_{p+q}(e^{i\omega}) \otimes I_{n^2}) \equiv 0$ . Fully written as

$$x^* (I_{n^2}, e^{i\omega} I_{n^2}, e^{2i\omega} I_{n^2}, \dots, e^{(p+q-1)i\omega} I_{n^2})^\top \equiv 0$$

or

$$(x_1^*, x_2^*, \dots, x_{p+q}^*)(I_{n^2}, zI_{n^2}, z^2I_{n^2}, \dots, z^{(p+q-1)}I_{n^2})^\top = 0 \quad \text{with } z \in \mathbb{C}.$$

It is straightforward to see that for  $z = 0$  we have  $x_1^* = 0$ , to obtain

$$zx_2^*I_{n^2} + z^2x_3^*I_{n^2} + \dots + z^{(p+q-1)}x_{p+q}^*I_{n^2} = 0,$$

divide by  $z$  and take then  $z = 0$  results in  $x_2^* = 0$ . A similar approach is done for the remaining components of  $x^*$  to conclude that  $x_j^* = 0$  for  $j = 1, 2, \dots, p + q$ . As a consequence,  $x = 0$  and hence  $\mathcal{P}(\vartheta) > 0$ .  $\square$

The equivalence of the singularity conditions of the Fisher information matrix  $\mathcal{F}(\vartheta)$  and the matrix  $\mathcal{M}(\vartheta)$  is shown in the next proposition.

**Proposition 2.4.** *The Fisher information matrix  $\mathcal{F}(\vartheta)$  becomes singular iff the matrix  $\mathcal{M}(\vartheta)$  is singular.*

**Proof.** If  $\mathcal{F}(\vartheta)$  is singular, there exists a fixed vector  $x \neq 0$ , such that  $\mathcal{F}(\vartheta)x = 0$ . Representation (17) yields

$$0 = \frac{1}{2\pi i} \oint_{|z|=1} x^* \mathcal{J}(z)x \frac{dz}{z}.$$

But  $\mathcal{J}(z) \geq 0$  for all  $|z| = 1$  yields  $\mathcal{J}(z)x \equiv 0$ , since  $x \neq 0$  it can be concluded that singularity of  $\mathcal{F}(\vartheta)$  implies  $\det \mathcal{J}(z) = 0$ .

It is straightforward to verify that  $\det \mathcal{J}(z) \equiv 0$  results in a singular matrix  $\mathcal{F}(\vartheta)$ . The proof follows directly from the approach just applied.

We shall now establish the singularity condition for the matrix  $\mathcal{M}(\vartheta)$ .

If  $\mathcal{M}(\vartheta)$  is singular there exists a fixed vector  $y \neq 0$ , such that  $\mathcal{M}(\vartheta)y = 0$ . From Lemma 2.3 it can be deduced that this implies a singular tensor Sylvester matrix  $S^\otimes(-B, A)$ . From Eq. (18) it then follows that

$$0 = \frac{1}{2\pi i} \oint_{|z|=1} A(z)\mathcal{J}(z)A^*(z)y \frac{dz}{z}.$$

Taking  $z = e^{i\omega}$ , we get

$$0 = \frac{1}{2\pi} \int_0^{2\pi} y^* A(e^{i\omega})\mathcal{J}(e^{i\omega})A^*(e^{i\omega})y \, d\omega.$$

As before we conclude from this  $A(e^{i\omega})\mathcal{J}(e^{i\omega})A^*(e^{i\omega})y \equiv 0$  and hence  $\det A(e^{i\omega})\mathcal{J}(e^{i\omega})A^*(e^{i\omega}) = 0$ . But since  $\det A^*(e^{i\omega}) \neq 0$  and  $\det A(e^{i\omega}) \neq 0$ , we must have  $\det \mathcal{J}(e^{i\omega}) \equiv 0$  or  $\det \mathcal{J}(z) \equiv 0$ . But then the Fisher information matrix  $\mathcal{F}(\vartheta)$  becomes singular. Conversely, if  $\det \mathcal{J}(z) \equiv 0$  it leads to a singular matrix  $\mathcal{M}(\vartheta)$ , this can be directly shown by virtue of the proof just done. If  $\mathcal{F}(\vartheta)$  is singular then  $\det \mathcal{J}(z) \equiv 0$  and hence trivially  $\mathcal{M}(\vartheta)$  is singular in view of (18).  $\square$

### 2.9. Main conclusions

By combining Lemma 2.3 and Proposition 2.4 one concludes that the Fisher information matrix  $\mathcal{F}(\vartheta)$  becomes singular if and only if the tensor Sylvester resultant matrix  $S^{\otimes}(-B, A)$  is singular. By virtue of Gohberg and Lerer [2], this will happen if and only if the matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue. This explains the aspect of singularity of the Fisher information matrix  $\mathcal{F}(\vartheta)$ . This also allows us to introduce a new resultant matrix, namely the matrix  $\mathcal{F}(\vartheta)$ . In other words, the Fisher information matrix  $\mathcal{F}(\vartheta)$  of a VARMA process has the same fundamental algebraic property as the tensor Sylvester resultant matrix  $S^{\otimes}(-B, A)$ . In [12,11] it is proved that the Fisher information matrix of scalar ARMA and ARMAX time series processes have the resultant property. Apparently the class of matrices consisting of the Fisher information matrices associated with various stationary time series processes (the scalar ARMA, ARMAX and vector ARMA processes) represents a new class of resultant matrices. However, the question of singularity of  $\mathcal{F}(\vartheta)$  is also interesting from a statistical point of view. In [10] a Wald test is formulated for testing common roots between the autoregressive and moving average polynomials of a scalar ARMA process. Such a test is equivalent with testing the singularity of the corresponding Fisher information matrix. Consequently, the results obtained in this paper should allow a similar test to be formulated in the multiple time series case.

For the evaluation of the matrix polynomials  $(A(z))^{-1}$ ,  $(B(z))^{-1}$ ,  $(A(z^{-1}))^{-1}$  and  $(B(z^{-1}))^{-1}$  which appear in (15) and (17), a property proved in Gohberg et al. [3] is considered.

### 2.10. Inverting matrix polynomials

Let  $\tilde{A}(z) = z^p A(z^{-1})$  and  $\tilde{B}(z) = z^q B(z^{-1})$ . The companion matrices associated with the matrix polynomials  $\tilde{A}(z)$  and  $\tilde{B}(z)$  are defined by the  $np \times np$  and  $nq \times nq$  matrices

$$C_A = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & & & & I \\ -A_p & -A_{p-1} & \dots & \dots & -A_1 \end{pmatrix} \quad \text{and}$$

$$C_B = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & & & & I \\ -B_q & -B_{q-1} & \dots & \dots & -B_1 \end{pmatrix},$$



respectively. As in the scalar case, the properties

$$\det(Iz - C_A) = \det\tilde{A}(z) \quad \text{and} \quad \det(Iz - C_B) = \det\tilde{B}(z)$$

and

$$\det(I - zC_A) = \det A(z) \quad \text{and} \quad \det(I - zC_B) = \det B(z)$$

hold, see [3].

The following equalities hold for every  $z \in \mathbb{C}$  which is not an eigenvalue of the matrix polynomials  $\tilde{A}(z)$  and  $\tilde{B}(z)$ ,

$$(\tilde{A}(z))^{-1} = P_A(Iz - C_A)^{-1}R_A \quad \text{and} \quad (\tilde{B}(z))^{-1} = P_B(Iz - C_B)^{-1}R_B$$

with

$$\begin{aligned} \text{the } n \times np \text{ matrix } P_A &= \begin{pmatrix} I & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \\ np \times n \text{ matrix } R_A &= \begin{pmatrix} 0 & \cdots & 0 & I \end{pmatrix}^\top, \end{aligned}$$

$$\begin{aligned} \text{the } n \times nq \text{ matrix } P_B &= \begin{pmatrix} I & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \\ nq \times n \text{ matrix } R_B &= \begin{pmatrix} 0 & \cdots & 0 & I \end{pmatrix}^\top. \end{aligned}$$

Since  $(A(z^{-1}))^{-1} = z^p (\tilde{A}(z))^{-1}$  and  $(B(z^{-1}))^{-1} = z^q (\tilde{B}(z))^{-1}$  we then have

$$\begin{aligned} (A(z^{-1}))^{-1} &= z^p P_A(Iz - C_A)^{-1}R_A \quad \text{and} \\ (B(z^{-1}))^{-1} &= z^q P_B(Iz - C_B)^{-1}R_B, \end{aligned}$$

and

$$\begin{aligned} (A(z))^{-1} &= z^{-p+1} P_A(I - zC_A)^{-1}R_A \quad \text{and} \\ (B(z))^{-1} &= z^{-q+1} P_B(I - zC_B)^{-1}R_B. \end{aligned}$$

These properties will enable us to evaluate the Fisher information matrix in a more efficient way. In the next section some numerical illustration is provided.

### 3. Numerical experiments

In this section numerical experiments are carried out in order to illustrate that an unnecessary computation of the Fisher information matrix of a VARMA process can be avoided when the results obtained in this paper are taken into consideration. The experiments were performed using MATLAB and Mathematica. The two examples of Section 2 that correspond with a VARMA process when  $n = 2$  and  $p = q = 1$ , are considered. As a first example we consider the tensor Sylvester matrix in the presence of common eigenvalues and establish its singularity.

### 3.1. Tensor Sylvester matrix for Example 1

For Example 1, the tensor Sylvester matrix is

$$S^{\otimes}(-B, A) = \begin{pmatrix} -1 & 0 & 0 & 0 & -0.5 & 0 & -0.76 & 0 \\ 0 & -1 & 0 & 0 & 0 & -0.5 & 0 & -0.76 \\ 0 & 0 & -1 & 0 & -0.25 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & -1 & 0 & -0.25 & 0 & 0.5 \\ 1 & 0 & 0 & 0 & 0.6 & 0.2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.4 & -0.6 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0.6 & 0.2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0.4 & -0.6 \end{pmatrix}.$$

It can easily be checked that the determinant of that matrix is equal to zero and even that its rank is equal to 6. This illustrates that, contrariwise to the Sylvester matrix  $S(-B, A)$ , the tensor Sylvester matrix becomes singular when the matrix polynomials  $A(z)$  and  $B(z)$  have at least one common eigenvalue.

### 3.2. Tensor Sylvester matrix for Example 2

For that example, the tensor Sylvester matrix is

$$S^{\otimes}(-B, A) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0.5 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0.5 & 0 & -0.5 \\ 1 & 0 & 0 & 0 & -0.8 & 0.2 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1.2 & -0.2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -0.8 & 0.2 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1.2 & -0.2 \end{pmatrix}.$$

The determinant of that matrix is different from zero. This confirms that the matrix polynomials  $A(z)$  and  $B(z)$  have no common eigenvalue in spite of the fact that the Sylvester matrix  $S(-B, A)$  is singular, with rank 3. Indeed the two sets of eigenvalues are complex, with a modulus respectively equal to 1.5811 for  $A(z)$  and 1.4142 for  $B(z)$ .

Let us now look at the Fisher information matrix and the resultant matrix  $\mathcal{M}(\vartheta)$  for the first example of Section 2. The second example is not interesting as far as the illustration of this paper is concerned because we expect a nonsingular Fisher information matrix.

### 3.3. The Fisher information matrix for Example 1

The Fisher information matrix is

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} R(z) dz,$$

where  $R(z)$  can be deduced from (15) or (17). Even in this simple example, it is much too cumbersome to give all the rational elements of  $R(z)$  in detail. Therefore we confine ourselves to give for example, element (1,1) which is equal to

$$R_{11}(z) = \frac{25(8 - 21z + 8z^2)(87,500 + 66,090z - 393,209z^2 + 66,090z^3 + 87,500z^4)}{256(-25 + 11z^2)^2(-11 + 25z^2)^2}.$$

Evaluating the integral using e.g. the Peterka and Vidinčev [19] algorithm implemented by Södertröm [23], yields the following matrix:

$$\mathcal{F}(\vartheta) = \begin{pmatrix} 4.20240 & 1.66919 & -0.17265 & -0.34627 & -1.81825 & -0.46554 & 0.93619 & 0.81064 \\ 1.66919 & 3.35066 & -0.02293 & 0.20263 & -0.42162 & -2.26694 & 0.56469 & -0.17981 \\ -0.17265 & -0.02293 & 1.33498 & 0.10744 & -0.12168 & -0.19206 & -1.37124 & -0.15414 \\ -0.34627 & 0.20263 & 0.10744 & 2.30449 & -0.12618 & 0.22791 & -0.19806 & -2.17429 \\ -1.81825 & -0.42162 & -0.12168 & -0.12618 & 1.62760 & 0.31622 & 0.0 & 0.0 \\ -0.46554 & -2.26694 & -0.19206 & 0.22791 & 0.31622 & 2.26637 & 0.0 & 0.0 \\ 0.93619 & 0.56469 & -1.37124 & -0.19806 & 0.0 & 0.0 & 1.62760 & 0.31622 \\ 0.81064 & -0.17981 & -0.15414 & -2.17429 & 0.0 & 0.0 & 0.31622 & 2.26637 \end{pmatrix}.$$

(shown with 5 decimals but the computations were done in double precision). The same results were obtained in Mathematica using the procedure deduced from Cauchy’s residue theorem which is simple here since there are only two poles to consider:  $\pm\sqrt{11}/5$ . It can be verified that the determinant of that matrix is numerically not equal to zero, although the value is close to zero namely 0.01176 and that its rank is equal to 8 so it has full rank. The eigenvalues of the matrix evaluated by Mathematica are, in decreasing order: 7.1174, 4.7505, 3.4032, 2.7438, 0.7789, 0.1213, 0.0588, and 0.0067. For understanding this result we locate close roots in the numerators and the denominators, the following elements possess close roots:

- $R_{51}(z), R_{61}(z), R_{52}(z), R_{62}(z)$ . Numerator : 0.643717, denominator : 0.663325
- $R_{15}(z), R_{25}(z), R_{16}(z), R_{26}(z)$ . Numerator : 1.55348, denominator : 1.50756
- $R_{71}(z), R_{81}(z), R_{72}(z), R_{82}(z)$ . Numerator : 0.635714, denominator : 0.663325
- $R_{17}(z), R_{27}(z), R_{18}(z), R_{28}(z)$ . Numerator : 1.57303, denominator : 1.50756

with a total of 16 elements with close roots. Note, the close root in the denominators is either the common eigenvalue or its inverse. It has already been checked in other contexts that evaluation of integrals in these circumstances is not at all accurate which explains the fact that the theoretical result is not confirmed numerically. It leads to the superiority of the tensor Sylvester matrix for confirming the presence of common eigenvalues between the autoregressive and moving average matrix polynomials. There are no numerical problems occurring since all the elements of the tensor Sylvester matrix are directly available. Therefore, before computing the Fisher information matrix it is recommended to check the rank of the tensor Sylvester matrix and improve the ratio of the largest to the smallest eigenvalue of that matrix e.g. [4]. In this simple numerical example we see that, simply for numerical reasons, the Fisher information matrix fails to detect common eigenvalues whereas the tensor

Sylvester matrix reveals it clearly, proving the usefulness of the results derived in this paper.

Since the matrix is numerically invertible, we have inverted it in order to obtain the asymptotic covariance matrix of the estimators. There, two variances (those of the third and seventh parameters) are abnormally high (respectively equal to 65.8 and 69.0) which may suggest an identification problem.

### 3.4. New matrix resultant for Example 1

The resultant matrix  $\mathcal{M}(\vartheta) = S^{\otimes}(-B, A)\mathcal{P}(\vartheta)[S^{\otimes}(-B, A)]^T$  has the form

$$\mathcal{M}(\vartheta) = \begin{pmatrix} 2.08175 & -0.32386 & -1.31682 & -1.08175 & -1.02042 & -0.08512 & -0.11168 & 0.46935 \\ -0.32386 & 5.03487 & -0.90113 & 0.32386 & 0.19338 & -1.12887 & 0.06209 & 0.25876 \\ -1.31682 & -0.90113 & 3.02909 & 1.31682 & -0.03674 & 0.15439 & -0.87348 & -0.70268 \\ -1.08175 & 0.32386 & 1.31682 & 2.08175 & 0.02042 & 0.08512 & 0.11168 & -1.46935 \\ -1.02042 & 0.19338 & -0.03674 & 0.02042 & 1.01878 & -0.04473 & 0.0 & 0.0 \\ -0.08512 & -1.12887 & 0.15439 & 0.08512 & -0.04473 & 1.25637 & 0.0 & 0.0 \\ -0.11168 & 0.06209 & -0.87348 & 0.11168 & 0.0 & 0.0 & 1.01878 & -0.04473 \\ 0.46935 & 0.25876 & -0.70268 & -1.46935 & 0.0 & 0.0 & -0.04473 & 1.25637 \end{pmatrix}.$$

The matrix  $\mathcal{M}(\vartheta)$  has rank equal to 6, this is equivalent with the rank of the corresponding tensor Sylvester matrix (Section 3.1). The resultant property of  $\mathcal{M}(\vartheta)$  is confirmed in this numerical example in contrast to the case of the Fisher information matrix.

Note that the elements of

$$\mathcal{P}(\vartheta) = \frac{1}{2\pi i} \oint_{|z|=1} (u_{p+q}(z) \otimes I_{n^2}) \Theta(z) (u_{p+q}(z) \otimes I_{n^2})^* \frac{dz}{z}$$

are much easier to compute than those of the Fisher information matrix. For example, the integrand of element (1,1) is equal to

$$\frac{625(8 - 21z + 8z^2)}{16(-25 + 11z^2)(-11 + 25z^2)}.$$

The matrix  $\mathcal{P}(\vartheta)$  has the following form

$$\mathcal{P}(\vartheta) = \begin{pmatrix} 1.62760 & 0.31622 & 0 & 0 & -0.89286 & -0.72470 & 0 & 0 \\ 0.31622 & 2.26637 & 0 & 0 & -1.07887 & 0.89286 & 0 & 0 \\ 0 & 0 & 1.62760 & 0.31622 & 0 & 0 & -0.89286 & -0.72470 \\ 0 & 0 & 0.31622 & 2.26637 & 0 & 0 & -1.07887 & 0.89286 \\ -0.89286 & -1.07887 & 0 & 0 & 1.62760 & 0.31622 & 0 & 0 \\ -0.72470 & 0.89286 & 0 & 0 & 0.31622 & 2.26637 & 0 & 0 \\ 0 & 0 & -0.89286 & -1.07887 & 0 & 0 & 1.62760 & 0.31622 \\ 0 & 0 & -0.72470 & 0.89286 & 0 & 0 & 0.31622 & 2.26637 \end{pmatrix}.$$

In this numerical example it can be seen that the matrix  $\mathcal{P}(\vartheta)$  is symmetric block Toeplitz. However, a generalization of the block Toeplitz property of  $\mathcal{P}(\vartheta)$  should be

investigated. The matrix  $\mathcal{P}(\vartheta)$  is strictly positive definite with the numerical eigenvalues, 3.46132, 3.46132, 3.1278, 3.1278, 0.766171, 0.766171, 0.43265, 0.43265 and determinant equal to 12.8792.

#### 4. Conclusion

It is shown that the Fisher information matrix  $\mathcal{F}(\vartheta)$  of a VARMA process is a multiple resultant matrix with respect to the autoregressive and moving average matrix polynomials. For that purpose we have developed compact representations of  $\mathcal{F}(\vartheta)$  so they can be summarized in one single equation consisting of one term. Other representations of the Fisher information matrix of a VARMA process as outlined in [15] consist of four till 16 terms. In this paper it is reduced to just one term. This has allowed us to establish new elegant algebraic results about the Fisher information matrix that could not be obtained using the alternative representations. The representation of the Fisher information matrix  $\mathcal{F}(\vartheta)$  that is explained by the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  is used to prove that the Fisher information matrix  $\mathcal{F}(\vartheta)$  is singular if and only if the autoregressive and moving average matrix polynomials have at least one common eigenvalue. It then follows that the Fisher information matrix and the tensor Sylvester matrix have equivalent singularity conditions. In the case of scalar ARMA processes the Sylvester matrix  $S(-B, A)$  and the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  coincide. But in the multivariate case only the tensor Sylvester matrix has the resultant property.

This can have interesting applications when numerical aspects are considered. The results derived in this paper suggest that during the modeling procedure one should first consider the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  before computing the Fisher information matrix  $\mathcal{F}(\vartheta)$ . It is clear that the issue of singularity is much easier to check with the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  than with Fisher's information matrix. Since the components of the matrix  $S^{\otimes}(-B, A)$  are directly available in terms of the matrix coefficients of the VARMA process, no additional computation is necessary. Contrarily, the Fisher information matrix  $\mathcal{F}(\vartheta)$  is composed of elements that have to be computed by means of Cauchy's residue theorem applied to rational functions. These rational functions consist of high degree scalar polynomials. In the example considered in Section 3.3, a VARMA process with  $n = 2$  and  $p = q = 1$ , the degree of the scalar polynomials appearing in the numerator and denominator can sometimes be equal to 8.

Let us assume that an empirical researcher has identified a given model and estimated its parameters using the approach described in e.g. [20]. He could then check the obtained model by computing the determinant of the tensor Sylvester matrix  $S^{\otimes}(-B, A)$ . This is a straightforward exercise. Knowledge of the singularity condition of the Fisher information matrix  $\mathcal{F}(\vartheta)$  is then directly available. This knowledge is crucial because it informs us whether we should start computing the elements of the Fisher information matrix  $\mathcal{F}(\vartheta)$  or not. It is clearly illustrated in Section 3 that an

apparently nonsingular  $\mathcal{F}(\vartheta)$  can be obtained in a case where it should be singular according to  $S^{\otimes}(-B, A)$ .

When the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  is singular one should consider a different VARMA process, generally simpler, and constitute the entries of the corresponding tensor Sylvester matrix  $S^{\otimes}(-B, A)$ . One shall proceed in this way until the tensor Sylvester matrix  $S^{\otimes}(-B, A)$  of the new model has full rank. This will guarantee a nonsingular Fisher information matrix  $\mathcal{F}(\vartheta)$  and one can then compute its elements. This approach saves a lot of unnecessary computations and will eventually result in more reliable covariances of the estimated VARMA parameters. In the example of Section 3, these covariances would be nonsense.

The results obtained in this paper can be used to examine some additional algebraic or other mathematical properties of the Fisher information matrix of a VARMA process. Whereas in a statistical framework, the results derived in this paper can be applied to set forth a statistical test for testing possible common eigenvalues of the autoregressive and moving average matrix polynomials.

### Acknowledgments

G. Mélard has benefited from an IAP-network in Statistics, contract P5/24, financially supported by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC). The authors wish to thank a referee for some detailed suggestions.

### References

- [1] P. Barone, A method for generating independent realization of a multivariate normal stationary and invertible ARMA( $p, q$ ) process, *J. Time Ser. Anal.* 8 (1987) 125–130.
- [2] I. Gohberg, L. Lerer, Resultants of matrix polynomials, *Bull. Amer. Math. Soc.* 82 (1976) 565–567.
- [3] I. Gohberg, P. Lancaster, P.L. Rodman, *Matrix Polynomials*, Academic Press, New York, 1982.
- [4] G.H. Golub, C.F. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, 1996.
- [5] E.J. Hannan, *Multiple Time Series*, John Wiley, New York, 1970.
- [6] E.J. Hannan, M. Deistler, *The Statistical Theory of Linear Systems*, John Wiley and Sons, New York, 1988.
- [7] E. Karlsson, T. Söderström, P. Stoica, The Cramér–Rao lower bound for noisy input–output systems, *Signal Process.* 80 (2000) 2421–2447.
- [8] A. Klein, G. Mélard, Computation of the Fisher information matrix for SISO models, *IEEE Trans. Signal Process.* 42 (1994) 684–688.
- [9] A. Klein, G. Mélard, The Fisher information matrix of multiple input single output time series models, *J. Comput. Appl. Math.* 51 (1994) 349–356.
- [10] A. Klein, Hypothesis testing of common roots, *Statist. Probab. Lett.* 20 (1994) 163–167.
- [11] A. Klein, P. Spreij, On Fisher’s information matrix of an ARMAX process and Sylvester’s resultant matrices, *Linear Algebra Appl.* 237–238 (1996) 579–590.
- [12] A. Klein, P. Spreij, On Fisher’s information matrix of an ARMA process, in: I. Csiszar, Gy. Michaletzky (Eds.), *Stochastic Differential and Difference Equations*, Progress in Systems and Control Theory, vol. 23, Birkhauser, Boston, 1997, pp. 273–284.

- [13] A. Klein, G. Méléard, 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.
- [14] A. Klein, H. Neudecker, A direct derivation of the exact Fisher information matrix of Gaussian vector state space models, *Linear Algebra Appl.* 321 (2000) 233–238.
- [15] A. Klein, A generalization of Whittle's formula for the information matrix of vector mixed time series, *Linear Algebra Appl.* 321 (2000) 197–208.
- [16] P. Lancaster, M. Tismenetsky, *The Theory of Matrices with Applications*, second ed., Academic Press, Orlando, 1985.
- [17] J.R. Magnus, H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised ed., Wiley, New York, 1999.
- [18] H.J. Newton, The information matrices of the parameters of multiple mixed time series, *J. Multivariate Anal.* 8 (1978) 317–323.
- [19] V. Peterka, P. Vidinčev, Rational-fraction approximation of transfer functions, in: *First IFAC Symposium on Identification in Automatic Control Systems*, Prague, 1967.
- [20] G.C. Reinsel, *Elements of Multivariate Time Series Analysis*, second ed., Springer-Verlag, New York, 1997.
- [21] L.L. Scharf, L.T. McWhorther, Geometry of the Cramér–Rao bound, *Signal Process.* 31 (1993) 301–311.
- [22] B.L. Shea, A note on the generation of independent realisations of a vector autoregressive moving average process, *J. Time Ser. Anal.* 9 (1988) 403–410.
- [23] T. Söderström, Description of a program for integrating rational functions around the unit circle, Technical Report 8467R, Department of Technology, Uppsala University, 1984.
- [24] A. Weiss, B. Friedlander, On the Cramér–Rao bound for direction finding of correlated signals, *IEEE Trans. Signal Process.* 41 (1993) 495–499.
- [25] P. Whittle, The analysis of multiple stationary time series, *J. Roy. Statist. Soc. B* 15 (1953) 125–139.