Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Linear Algebra and its Applications 417 (2006) 140–149
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

An explicit expression for the Fisher information matrix of a multiple time series process

André Klein [a,*], Peter Spreij [b]

[a] *Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*
[b] *Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands*

Received 28 April 2004; accepted 25 April 2005
Available online 27 June 2005
Submitted by A. Mavriewicz

Abstract

The principal result in this paper is concerned with the derivative of a vector with respect to a block vector or matrix. This is applied to the asymptotic Fisher information matrix (FIM) of a stationary vector autoregressive and moving average time series process (VARMA). Representations which can be used for computing the components of the FIM are then obtained. In a related paper [A. Klein, A generalization of Whittle's formula for the information matrix of vector mixed time series, Linear Algebra Appl. 321 (2000) 197–208], the derivative is taken with respect to a vector. This is obtained by vectorizing the appropriate matrix products whereas in this paper the corresponding matrix products are left unchanged.

AMS classification: 15A69; 62H12

Keywords: Matrix differential rules; Matrix polynomial; Fisher information matrix; VARMA process

* Corresponding author. Tel.: +31 20 5254245; fax: +31 20 5254349.
E-mail address: a.a.b.klein@uva.nl (A. Klein).

0024-3795/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2005.04.024

1. Introduction

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

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

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

In this paper the derivative of a vector with respect to a matrix is considered. In [1] a different approach is used, the derivative is taken with respect to a vector, this is obtained by vectorizing the appropriate matrix products. This is implemented by the vec operator which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. The approach developed in this paper leaves the matrix products unchanged. The obtained results are applied to the Fisher information matrix of a VARMA process and lead to representations which can be used for computing the corresponding components of the information matrix.

2. The Fisher information matrix

Assume that time series $\{y(t), t \in \mathbb{N}\}$ satisfying Eq. (1) is a zero mean Gaussian time series. Then its stationary distribution depends on parameters $\vartheta = (\vartheta_1, \ldots, \vartheta_\ell)^\top$, where ℓ is the number of matrix parameters of the vector autoregressive moving average model. When the entries of $\vartheta_1, \ldots, \vartheta_\ell$ are considered, the number of parameters is equal to $n^2(p+q)$. The choice for the $n(p+q) \times n$ parameter matrix is

$$\vartheta = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \\ B_1 \\ B_2 \\ \vdots \\ B_q \end{pmatrix} \quad \text{or} \quad \vartheta = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_p \\ \vartheta_{p+1} \\ \vartheta_{p+2} \\ \vdots \\ \vartheta_{p+q} \end{pmatrix}. \tag{2}$$

When the representation of the parameter vector ϑ as defined in (2) is considered, the following equality holds for the $n^2(p+q) \times n^2(p+q)$ asymptotic Fisher information matrix

$$\mathscr{F}(\vartheta) = \mathbb{E}_\vartheta \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta} \right)^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta} \right) \right\} \tag{3}$$

and for simplicity t is omitted from $\varepsilon(t)$ in the right-hand side of (3).

3. Main result

In this section the results developed in this paper shall be presented. The derivative of a vector with respect to a block vector or matrix is first considered and the obtained representation of $\partial\varepsilon/\partial\vartheta$ is inserted in (3). Consequently, explicit expressions for the entries of $\mathscr{F}(\vartheta)$ are derived.

3.1. The derivative of a vector with respect to a block vector

In this section we introduce the derivative which is taken with respect to a block vector or a matrix. The approach set forth in this paper will allow us to give a representation for each element of the Fisher information matrix. We therefore rewrite the VARMA process as

$$y(t) = A^{-1}(L)B(L)\varepsilon(t) \tag{4}$$

and set forth a form for $\partial\varepsilon/\partial\vartheta$.

Consider a real, differentiable $(m \times n)$ matrix function $X(\vartheta)$ of a real $(\ell \times 1)$ vector $\vartheta = (\vartheta_1, \ldots, \vartheta_\ell)^\top$, where m, n and ℓ are positive integers. Let $(m \times n)$ matrices $\partial_r X = (\partial X_{ij}/\partial\vartheta_r)$ with $r = 1, \ldots, \ell$ be the first order derivatives of $X(\vartheta)$ in partial derivative form with X_{ij} being the first (i, j) element of X. Write $\mathrm{d}X_{ij} = \sum_{r=1}^{\ell} (\partial X_{ij}/\partial\vartheta_r)\mathrm{d}\vartheta_r$, where $\mathrm{d}\vartheta_r$ is an arbitrary perturbation of ϑ_r. The $(m \times n)$ matrix $\mathrm{d}X = (\mathrm{d}X_{ij})$ 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 ∂_r for $r = 1, \ldots, \ell$. 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 ℓ are positive integers. The usual scalar product rule of differentiation yields

$$\mathrm{d}(XY) = (\mathrm{d}X)Y + X(\mathrm{d}Y).$$

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 ϑ, and as a second property the next differential rule is used

$$\mathrm{d}A^{-1}(L) = -A^{-1}(L)\mathrm{d}A(L)A^{-1}(L).$$

This enables us to formulate the following equation for the VARMA process given in (4)

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

We now set forth the representation

$$\frac{\partial\varepsilon}{\partial\vartheta}\Delta\vartheta = B^{-1}(L)\left\{L\Delta\vartheta_1 + L^2\Delta\vartheta_2 + \cdots + L^p\Delta\vartheta_p\right\}A^{-1}(L)B(L)\varepsilon - B^{-1}(L)\left\{L\Delta\vartheta_{p+1} + L^2\Delta\vartheta_{p+2} + \cdots + L^q\Delta\vartheta_{p+q}\right\}\varepsilon,$$

where $\Delta\vartheta_i$ is an arbitrary perturbation.

The next step consists of choosing an appropriate $\Delta\vartheta_i$. To construct the first n^2 columns of the matrix $\partial\varepsilon/\partial\vartheta$, the following approach is applied. We define the $n \times n$ matrix E_{ij} with the (i, j) th entry equal to 1 and 0 elsewhere. The first n columns will be constructed by means of the n standard basis vectors $e_1, e_2, \ldots, e_n$ in $\mathbb{R}^n$ belonging to E_{i1}, for $i = 1, 2, \ldots, n$. The standard basis block vectors necessary for deriving the first n columns of $\partial\varepsilon/\partial\vartheta$ and associated with ϑ_1 are then

$$\begin{pmatrix} E_{11} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}, \begin{pmatrix} E_{21} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}, \ldots, \begin{pmatrix} E_{n1} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}.$$

We see that $\Delta\vartheta_1$ shall consist of the $n \times n$ upper matrices E_{i1} with $i = 1, 2, \ldots, n$, whereas $\Delta\vartheta_2, \Delta\vartheta_3, \ldots, \Delta\vartheta_p, \Delta\vartheta_{p+1}, \Delta\vartheta_{p+2}, \ldots, \Delta\vartheta_{p+q}$ are zero. Consequently, the first n columns of $\partial\varepsilon/\partial\vartheta$ are given by

$$\begin{array}{c} LB^{-1}(L)E_{11}A^{-1}(L)B(L)\varepsilon \\ LB^{-1}(L)E_{21}A^{-1}(L)B(L)\varepsilon \\ LB^{-1}(L)E_{31}A^{-1}(L)B(L)\varepsilon \\ \vdots \\ LB^{-1}(L)E_{n1}A^{-1}(L)B(L)\varepsilon. \end{array}$$

The next n columns are constructed by considering the first, second up to the nth standard basis vector belonging to E_{i2}, $i = 1, 2, \ldots, n$. The standard basis block vectors associated with ϑ_1 are then

$$\begin{pmatrix} E_{12} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}, \begin{pmatrix} E_{22} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}, \ldots, \begin{pmatrix} E_{n2} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}.$$

The corresponding n columns are

$$\begin{gathered} LB^{-1}(L)E_{12}A^{-1}(L)B(L)\varepsilon \\ LB^{-1}(L)E_{22}A^{-1}(L)B(L)\varepsilon \\ LB^{-1}(L)E_{32}A^{-1}(L)B(L)\varepsilon \\ \vdots \\ LB^{-1}(L)E_{n2}A^{-1}(L)B(L)\varepsilon. \end{gathered}$$

We proceed in this way to obtain the last n columns associated with ϑ_1. The appropriate standard basis block vectors are given by

$$\begin{pmatrix} E_{1n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}, \begin{pmatrix} E_{2n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}, \ldots, \begin{pmatrix} E_{nn} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}.$$

The corresponding n columns are

$$\begin{gathered} LB^{-1}(L)E_{1n}A^{-1}(L)B(L)\varepsilon \\ LB^{-1}(L)E_{2n}A^{-1}(L)B(L)\varepsilon \\ LB^{-1}(L)E_{3n}A^{-1}(L)B(L)\varepsilon \\ \vdots \\ LB^{-1}(L)E_{nn}A^{-1}(L)B(L)\varepsilon. \end{gathered}$$

Similarily for the next n^2 columns associated with ϑ_2. In this case the standard basis block vectors have the following representation

$$\begin{pmatrix} 0_{n\times n} \\ E_{ij} \\ \vdots \\ 0_{n\times n} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix}$$

with $i, j = 1, 2, \ldots, n$. The matrix E_{ij} is associated with $\Delta\vartheta_2$ and $\Delta\vartheta_1, \Delta\vartheta_3, \ldots, \Delta\vartheta_p, \Delta\vartheta_{p+1}, \Delta\vartheta_{p+2}, \ldots, \Delta\vartheta_{p+q}$ are zero. The corresponding n^2 columns are

$$L^2 B^{-1}(L) E_{ij} A^{-1}(L) B(L)\varepsilon,$$

for each $j = 1, 2, \ldots, n$ we have $i = 1, 2, \ldots, n$. A similar approach is applied for the remaining columns associated with ϑ_p, the standard basis block vectors are

$$\begin{pmatrix} 0_{n\times n} \\ \vdots \\ 0_{n\times n} \\ E_{ij} \\ 0_{n\times n} \\ \vdots \\ 0_{n\times n} \end{pmatrix} \to p\text{th } n \times n \text{ block}$$

with $i, j = 1, 2, \ldots, n$. The corresponding n^2 columns are given by

$$L^p B^{-1}(L) E_{ij} A^{-1}(L) B(L)\varepsilon.$$

The $n^2 q$ columns associated with $\vartheta_{p+1}, \vartheta_{p+2}, \ldots, \vartheta_{p+q}$, have the representation

$$-L^k B^{-1}(L) E_{ij}\varepsilon,$$

where $k = 1, 2, \ldots, q$ and for each k we have the same specification for the matrices E_{ij} as for the first $n^2 p$ columns.

We shall summarize the results in a proposition. For that purpose we define

$$\phi_{ij}(L) = B^{-1}(L) E_{ij} A^{-1}(L) B(L) \quad \text{and} \quad \psi_{ij}(L) = -B^{-1}(L) E_{ij}.$$

The following representations are now introduced

$$\begin{aligned} \Phi(L) = (&\phi_{11}(L)\varepsilon, \phi_{21}(L)\varepsilon, \ldots, \phi_{n1}(L)\varepsilon, \phi_{12}(L)\varepsilon, \phi_{22}(L)\varepsilon, \ldots, \\ &\phi_{n2}(L)\varepsilon, \ldots, \phi_{1n}(L)\varepsilon, \phi_{2n}(L)\varepsilon, \ldots, \phi_{nn}(L)\varepsilon) \end{aligned}$$

and

$$\begin{aligned} \Psi(L) = (&\psi_{11}(L)\varepsilon, \psi_{21}(L)\varepsilon, \ldots, \psi_{n1}(L)\varepsilon, \psi_{12}(L)\varepsilon, \psi_{22}(L)\varepsilon, \\ &\ldots, \psi_{n2}(L)\varepsilon, \ldots, \psi_{1n}(L)\varepsilon, \psi_{2n}(L)\varepsilon, \ldots, \psi_{nn}(L)\varepsilon)\,. \end{aligned}$$

Proposition 3.1. *The following representation holds true*

$$\begin{aligned}\frac{\partial\varepsilon}{\partial\vartheta} &= \Big(L\Phi(L), L^2\Phi(L), \dots, L^p\Phi(L), L\Psi(L), L^2\Psi(L), \dots, L^q\Psi(L)\Big)\\ &= L\left\{(1, L, \dots, L^{p-1})\otimes\Phi(L),\ (1, L, \dots, L^{q-1})\otimes\Psi(L)\right\}\\ &= L\left\{u_p^\top(L)\otimes\Phi(L), u_q^\top(L)\otimes\Psi(L)\right\}, \end{aligned}\tag{5}$$

where $u_x^\top(L) = (1, L, L^2, \dots, L^{x-1})$ *for positive integers* x *and* $A\otimes B$ *is the Kronecker product of* A *and* B.

3.2. Representation of the entries of the Fisher information matrix

Representations which can be used for computing the entries of $\mathscr{F}(\vartheta)$ shall now be set forth by applying formula (3) when (5) is considered. We shall proceed with the block representation of $\mathscr{F}(\vartheta)$ which is given by

$$\mathscr{F}(\vartheta) = \begin{pmatrix}\mathscr{F}_{aa}(\vartheta) & \mathscr{F}_{ab}(\vartheta)\\ \mathscr{F}_{ba}(\vartheta) & \mathscr{F}_{bb}(\vartheta)\end{pmatrix}.\tag{6}$$

In a dynamic stationary stochastic context it has long been shown useful to use Fourier transform representations which provide alternatively circular integral representations. For evaluating $\mathscr{F}(\vartheta)$ the following integral representation

$$\mathscr{F}(\vartheta) = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\begin{pmatrix}\mathscr{I}_{aa}(z) & \mathscr{I}_{ab}(z)\\ \mathscr{I}_{ba}(z) & \mathscr{I}_{bb}(z)\end{pmatrix}\frac{\mathrm{d}z}{z}\tag{7}$$

is considered, the integration is counterclockwise around the unit circle. Below appropriate representations of the blocks composing (7) are derived. We shall use an arbitrary element of block $\mathscr{I}_{aa}(z)$ in (7) to illustate how the representations of the remaining components 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 the input process is described by $u(t) = G(L)v(t)$. $H(L)$ and $G(L)$ are asymptotically stable filters. For evaluating the cross covariance matrix of the output $x(t)$ and the input $u(t)$, the following equation holds true

$$\mathbb{E}_\vartheta\left\{x(t)u^\top(t)\right\} = \int_{-\pi}^{\pi}\Omega_{xu}(\omega)\mathrm{d}\omega \quad \omega\in[-\pi,\pi],\tag{8}$$

where $\Omega_{xu}(\omega)$ is the cross spectral density of the processes $x(t)$ and $u(t)$. It is defined as $\Omega_{xu}(\omega) = H(\mathrm{e}^{\mathrm{i}\omega})\Omega_u(\omega)$ with $\Omega_u(\omega)$ being the spectral density of the input process $u(t)$ which is given by $\Omega_u(\omega) = G(\mathrm{e}^{\mathrm{i}\omega})\Omega_v(\omega)G^*(\mathrm{e}^{\mathrm{i}\omega})$. Y^* denotes the complex conjugate transpose of the matrix Y and $\Omega_v(\omega)$ is the spectral density of the process $v(t)$. When representation (5) is inserted in (3), an arbitrary element of block $\mathscr{F}_{aa}(\vartheta)$ then takes the form

$$\mathbb{E}_\vartheta\left\{\mathrm{Tr}\left(L^{k+1}\phi_{ij}(L)\varepsilon\varepsilon^\top L^{r+1}\phi_{lm}^\top(L)\Sigma^{-1}\right)\right\},\tag{9}$$

where Tr(*M*) is the trace of the square matrix *M* and $k, r = 0, 1, \ldots, p-1$ and $i, j, l, m = 1, 2, \ldots, n$.

We have now a similar representation to the left-hand side of (8) where

$$x(t) = L^{k+1}\phi_{ij}(L)\varepsilon \quad \text{and} \quad u(t) = \Sigma^{-1}L^{r+1}\phi_{lm}(L)\varepsilon.$$

The connection between the processes $x(t)$ and $u(t)$ is then given by

$$x(t) = L^{k-r}\phi_{ij}(L)\phi_{lm}^{-1}(L)\Sigma u(t).$$

Since the white noise process ε has a constant spectral density (independent of the frequency ω) then it is straightforward to conclude that in view of (8) the value of the spectral density of ε is $(1/2\pi)\,\Sigma$. The spectral density of the input process $u(t)$ is then

$$\frac{1}{2\pi}\left\{\Sigma^{-1}\phi_{lm}(\mathrm{e}^{\mathrm{i}\omega})\Sigma\phi_{lm}^{*}(\mathrm{e}^{\mathrm{i}\omega})\Sigma^{-1}\right\}.$$

Permutation of expectancy $\mathbb{E}_\vartheta$ and trace in (9) and application of (8) leads to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\left\{\mathrm{e}^{\mathrm{i}\omega(k-r)}\phi_{ij}(\mathrm{e}^{\mathrm{i}\omega})\Sigma\phi_{lm}^{*}(\mathrm{e}^{\mathrm{i}\omega})\Sigma^{-1}\right\}\mathrm{d}\omega.$$

Equivalently for $z = \mathrm{e}^{\mathrm{i}\omega}$ we have

$$(\mathscr{F}_{aa}(\vartheta))_{i,j,l,m}^{k,r} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{k-r}\mathrm{Tr}\left(\Sigma\phi_{lm}^{*}(z)\Sigma^{-1}\phi_{ij}(z)\right)\frac{\mathrm{d}z}{z}, \tag{10}$$

where $k, r = 0, 1, \ldots, p-1$ and $i, j, l, m = 1, 2, \ldots, n$.

A similar approach is used for the remaining components of the Fisher information matrix $\mathscr{F}(\vartheta)$. Representation (7) of $\mathscr{F}(\vartheta)$ can then be summarized accordingly, to obtain

$$(\mathscr{F}_{ab}(\vartheta))_{i,j,l,m}^{k,f} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{k-f}\mathrm{Tr}\left(\Sigma\phi_{ij}^{*}(z)\Sigma^{-1}\psi_{lm}(z)\right)\frac{\mathrm{d}z}{z}, \tag{11}$$

where $k = 0, 1, \ldots, p-1$ and $f = 0, 1, \ldots, q-1$ and $i, j, l, m = 1, 2, \ldots, n$.

$$(\mathscr{F}_{ba}(\vartheta))_{i,j,l,m}^{f,k} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{f-k}\mathrm{Tr}\left(\Sigma\psi_{ij}^{*}(z)\Sigma^{-1}\phi_{lm}(z)\right)\frac{\mathrm{d}z}{z}, \tag{12}$$

where $f = 0, 1, \ldots, q-1$ and $k = 0, 1, \ldots, p-1$ and $i, j, l, m = 1, 2, \ldots, n$.

$$(\mathscr{F}_{bb}(\vartheta))_{i,j,l,m}^{h,f} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{h-f}\mathrm{Tr}\left(\Sigma\psi_{ij}^{*}(z)\Sigma^{-1}\psi_{lm}(z)\right)\frac{\mathrm{d}z}{z}, \tag{13}$$

where $f, h = 0, 1, \ldots, q-1$ and $i, j, l, m = 1, 2, \ldots, n$.

We shall now present Whittle's formula for the VARMA process (1). It is given by the following equality, see [4]

$$\mathscr{F}_{jk}(\vartheta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\left(\frac{\partial\Omega(\omega)}{\partial\vartheta_j}\Omega^{-1}(\omega)\frac{\partial\Omega(\omega)}{\partial\vartheta_k}\Omega^{-1}(\omega)\right)\mathrm{d}\omega, \tag{14}$$

where the spectral density of the VARMA process (1) is

$$\Omega(\omega) = \left(\frac{1}{2\pi}\right) A^{-1}(\mathrm{e}^{\mathrm{i}\omega}) B(\mathrm{e}^{\mathrm{i}\omega}) \Sigma B^{\top}(\mathrm{e}^{-\mathrm{i}\omega}) A^{-\top}(\mathrm{e}^{-\mathrm{i}\omega}).$$

In [1] the equivalence between (3) and the matrix-level version of (14) is shown. Consequently, the entries of the matrix-level version of (14) are explicitly given by (10)–(13).

4. A numerical example

In this section a numerical example of Fisher's information matrix is given for a bivariate VARMA (1, 1). It is based on the expressions developed in this paper. Consider the vector process with the following autoregressive and moving average matrix polynomials:

$$A(z) = \begin{pmatrix} 1 - 0.8\,z & 0.2\,z \\ -1.2\,z & 1 - 0.2\,z \end{pmatrix} \quad \text{and} \quad B(z) = \begin{pmatrix} 1 & z \\ -0.5\,z & 1 + 0.5\,z \end{pmatrix}.$$

Applying Cauchy's residue theorem to expressions (10)–(13) leads to the Fisher information matrix

$$\begin{pmatrix} 3.11081 & -1.08243 & 1.30797 & -0.09511 & -1.27989 & 1.16848 & 0.47011 & 0.66848 \\ -1.08243 & 3.78382 & -1.12772 & 0.34058 & -0.36413 & -1.90217 & -0.86413 & 1.09783 \\ 1.30797 & -1.12772 & 5.03714 & -1.86141 & 0.57337 & -0.02717 & -1.17663 & 0.47283 \\ -0.09511 & 0.34058 & -1.86141 & 5.25725 & -0.28804 & 1.03261 & 0.21196 & -1.96739 \\ -1.27989 & -0.36413 & 0.57337 & -0.28804 & 1.75 & -0.5 & 0.0 & 0.0 \\ 1.16848 & -1.90217 & -0.02717 & 1.03261 & -0.5 & 3. & 0.0 & 0.0 \\ 0.47011 & -0.86413 & -1.17663 & 0.21196 & 0.0 & 0.0 & 1.75 & -0.5 \\ 0.66848 & 1.09783 & 0.47283 & -1.96739 & 0.0 & 0.0 & -0.5 & 3. \end{pmatrix}.$$

The eigenvalues of the matrix are: 8.20923, 6.85511, 4.05189, 3.51982, 2.27653, 1.37935, 0.290461, 0.106623, and the determinant is equal to 78.0513. However, it is worth mentioning that time series analysts for computational purposes frequently prefer to use observed information computed directly using numerical differentiation. Using the approach developed in this paper does not require numerical differentiation since an analytic procedure for the derivatives has been set forth. The obtained representations (10), (11), (12) and (13) can then be computed by using e.g. the Peterka and Vidinčev [2] algorithm implemented by Söderström [3] for circular integrals of the type derived in this paper.

Acknowledgments

The authors wish to thank B. de Pagter for valuable comments on the subject of derivation with respect to a matrix as well as one referee that has suggested to illustrate the results by a numerical example.

References

[1] A. Klein, A generalization of Whittle's formula for the information matrix of vector mixed time series, Linear Algebra Appl. 321 (2000) 197–208.
[2] V. Peterka, P. Vidinčev, Rational-fraction approximation of transfer functions, First IFAC Symposium on Identification in Automatic Control Systems, Prague, 1967.
[3] T. Söderström, Description of a program for integrating rational functions around the unit circle, Technical Report 8467R, Department of Technology, Uppsala University, 1984.
[4] P. Whittle, The analysis of multiple stationary time series, J. Royal Statist. Soc. B. 15 (1953) 125–139.