Linear Algebra and its Applications 430 (2009) 674–691

Contents lists available at ScienceDirect

Linear Algebra and its Applications

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

Matrix differential calculus applied to multiple stationary time series and an extended Whittle formula for information matrices

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

ARTICLE INFO

Article history:
Received 4 February 2008
Accepted 5 September 2008

Submitted by V. Mehrmann

AMS classification:
15A57
15A69
26B12
62B10
62H12

Keywords:
Matrix differential calculus
Matrix polynomial
Fisher information matrix
VARMAX process
Whittle formula

ABSTRACT

The purpose of this paper is to set forth easily implementable expressions for the Fisher information matrix (FIM) of a Gaussian stationary vector autoregressive and moving average process with exogenous or input variables, a vector ARMAX or VARMAX process. The entries of the FIM are represented as circular integral expressions and can be computed by applying Cauchy's residue theorem. An extension of the Whittle formula for the FIM of multiple time series processes is developed for VARMAX processes. It will be shown that the extended Whittle formula yields the FIM when a bivariate structure, consisting of the VARMAX process and the exogenous-input process, is considered. Consequently, the equivalence between a frequency and time domain representation of the FIM of VARMAX processes is established. In order to obtain the results presented in this paper, the differentiation techniques developed and used in [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] are applied.

1. Introduction

The purpose of this paper is to set forth easily implementable expressions for the Fisher information matrix of a Gaussian stationary vector autoregressive and moving average process with exogenous or

* 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
doi:10.1016/j.laa.2008.09.019

input variables, a vector ARMAX or VARMAX process. A difficult computational problem involved in the statistical inference from time series is that of determining the asymptotic covariance matrix of the maximum likelihood estimators. The asymptotic covariance matrix is obtained by inverting the Fisher information matrix. The Fisher information plays a vital role in estimation theory and statistical signal processing and control, see e.g. [3,7] and more recently in physics, see e.g. [5,6]. VARMAX processes are of common use in signal processing, control and system theory and statistics, see e.g. [18,9]. The entries of the Fisher information matrix represented as circular integral expressions and computed by applying Cauchy's residue theorem. It concerns the evaluation of integrals of a rational function over the unit circle. These integrals can be computed by recursions with respect to the degree of the polynomials, see e.g. [23]. However, a more efficient and faster method consists of transforming the problem to the evaluation of the autocovariances of an ARMA process by means of the algorithm developed in [4].

Until recently the asymptotic covariance matrix of the Gaussian VARMA model has been stated only in terms of formulas involving integration over the frequency domain. For stationary time series models without inputs, the Whittle formula, developed in the pioneering paper [24], for the Fisher information matrix of multiple time series processes, is a frequency-domain integral representation and used in [21] to derive closed form expressions for the VARMA case. In [14], an equivalence between a frequency and time domain representation of the Fisher information matrix of VARMA processes is established. It is worth emphasizing that many results on the asymptotic Fisher information matrix of multiple time series processes are limited to these processes. In most of these cases, except in [15], Whittle's formula is used, see e.g. [21].

In the present paper, the Whittle formula is extended to VARMAX processes. Since a straightforward application of the Whittle formula in a VARMAX context does not yield the Fisher information matrix, an alternative approach will be developed. It will be shown that the original Whittle formula can be used to obtain the Fisher information matrix of a VARMAX model, when the process is rewritten in a bivariate form. The latter combines both the VARMAX process as well as the exogenous stationary input process. The Fisher information matrix of VARMAX processes as well as the corresponding extended Whittle formula developed in this paper are set forth both at the full matrix and block matrix level. We will establish equivalence between a frequency and time domain representation of the Fisher information matrix.

In order to obtain the results presented in this paper, we apply the methods developed in [15] to obtain the Fisher information matrix of a VARMA process. These methods involve differentiation of the error process with respect to the parameter matrices. In many studies, e.g. [14,16,19] when differentiation is applied in statistics and matrix calculus, the differentiation methods use the vectorization of matrices and matrix products in order to obtain the desired representations. In the present paper the differentiation is such that the structure of the matrix is left unchanged.

For other applications of matrix derivatives in statistics, see e.g. [19]. More recent developments on matrix derivatives are well covered in the survey paper by Wong [25]. Matrix calculus finds also applications in other areas of interest. The mathematical methods of quantum statistical inference are based on matrix derivatives, see e.g. [1,2]. For applications in econometrics, see e.g. [20].

Consider the vector difference equation representation of a linear system $\{y(t), t \in \mathbb{Z}\}$, $\mathbb{Z}$ the set of integers, of order (p, r, q)

$$\sum_{j=0}^{p} \alpha_j y(t-j) = \sum_{j=0}^{r} \gamma_j x(t-j) + \sum_{j=0}^{q} \beta_j \varepsilon(t-j), \quad t \in \mathbb{Z}, \tag{1}$$

where $y(t)$, $x(t)$ and $\varepsilon(t)$ are the outputs, the observed inputs $x(t)$ also named the exogenous or control variable depending on the field of application (econometrics, signal processing and systems and control), and the errors, respectively, and where $\alpha_j \in \mathbb{R}^{n\times n}$, $\gamma_j \in \mathbb{R}^{n\times m}$, and $\beta_j \in \mathbb{R}^{n\times n}$ are the associate parameter matrices. We impose $\alpha_0 \equiv \beta_0 \equiv I_n$.

Eq. (1) can compactly be written as

$$\alpha(L)y(t) = \gamma(L)x(t) + \beta(L)\varepsilon(t), \tag{2}$$

where

$$\alpha(L) = \sum_{j=0}^{p} \alpha_j L^j; \quad \gamma(L) = \sum_{j=0}^{r} \gamma_j L^j; \quad \beta(L) = \sum_{j=0}^{q} \beta_j L^j,$$

where L denotes the backward shift operator, e.g. $Lx(t) = x(t-1)$. The estimation of the matrices $\alpha_1, \alpha_2, \ldots, \alpha_p, \gamma_1, \gamma_2, \ldots, \gamma_r, \beta_1, \beta_2, \ldots, \beta_q$ has received considerable attention in the time series and statistical signal processing literature, see e.g. [3,8,10,18].

The left-hand side of (1) is the autoregressive part, the second term on the right-hand side the moving average part and process $x(t)$ are the input variables. The error $\{\varepsilon(t), t \in \mathbb{N}\}$ is a collection of uncorrelated zero mean n-dimensional random variables each having positive definite covariance matrix Σ and we assume, for all s, t, $\mathbb{E}_\vartheta\{x(s)\varepsilon^\top(t)\} = 0$, where $\top$ denotes the transposition and $\mathbb{E}_\vartheta$ represents the expected value under the parameter ϑ. The matrix ϑ represents all the VARMAX parameters, with the total number of parameters being $n^2(p+q) + mn(r+1)$. The choice for the $n \times \{n(p+q) + m(r+1)\}$ parameter matrix is

$$\vartheta = (\vartheta_1 \ \vartheta_2 \ \cdots \ \vartheta_p \ \vartheta_{p+1} \ \vartheta_{p+2} \ \cdots \ \vartheta_{p+r} \ \vartheta_{p+r+1} \ \vartheta_{p+r+2} \ \vartheta_{p+r+3} \ \cdots \ \vartheta_{p+r+q+1}), \tag{3}$$

$$= (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_p \ \gamma_0 \ \gamma_1 \ \cdots \ \gamma_{r-1} \ \gamma_r \ \beta_1 \ \beta_2 \ \cdots \ \beta_q). \tag{4}$$

The observed input variable $x(t)$ is assumed to be a stationary process with spectral density $R_x(\cdot)/2\pi$. If $x(t)$ is a m-dimensional VARMA process satisfying

$$a(L)x(t) = b(L)\eta(t), \tag{5}$$

then

$$R_x(\mathrm{e}^{\mathrm{i}\omega}) = a^{-1}(\mathrm{e}^{\mathrm{i}\omega})b(\mathrm{e}^{\mathrm{i}\omega})\Omega b^*(\mathrm{e}^{\mathrm{i}\omega})a^{-*}(\mathrm{e}^{\mathrm{i}\omega}), \quad \omega \in [-\pi, \pi], \tag{6}$$

where $\mathbb{E}\{\eta(t)\eta^\top(t)\} = \Omega$.

The assumption $\det(\alpha(z)) \neq 0, |z| \leqslant 1$ and $\det(\beta(z)) \neq 0, |z| \leqslant 1$ ensures that $\varepsilon(t)$ are the linear innovations, in the linear prediction of $y(t)$ from $x(s), y(s)$ when $s < t$. The elements of $\alpha^{-1}(z)$ and $\beta^{-1}(z)$ can be written in power series in z. In [13], the scalar version of (2) is considered. The authors proved that the asymptotic Fisher information matrix is singular if and only if the scalar polynomials $\alpha(z), \beta(z)$ and $\gamma(z)$ have at least one common root. In [16], the same property is considered for the asymptotic Fisher information matrix of a VARMA process. The authors show that the Fisher information matrix becomes singular if and only if the VARMA matrix polynomials have at least one common eigenvalue. A similar result probably holds for the VARMAX case as well.

The $\{n^2(p+q) + mn(r+1)\} \times \{n^2(p+q) + mn(r+1)\}$ asymptotic Fisher information matrix of the VARMAX process is given by

$$\mathscr{F}(\vartheta) = \mathbb{E}_\vartheta\left\{\left(\frac{\partial\varepsilon}{\partial\vartheta}\right)^\top \Sigma^{-1}\left(\frac{\partial\varepsilon}{\partial\vartheta}\right)\right\}. \tag{7}$$

The paper is organized as follows. The technique of differentiation with respect to a matrix applied to the VARMAX process is described in Section 2. A convenient representation for the derivative $\partial\varepsilon/\partial\vartheta$ will be constructed in order to evaluate (7). In Section 3, the results developed in Section 2 are used to construct the entries of the Fisher information matrix of a VARMAX process and closed form expressions are derived. In Section 4, we emphasize the fact that Whittle's formula for the Fisher information matrix of a VARMA process cannot be directly used for a VARMAX process. We therefore present an extended Whittle formula which corresponds to the Fisher information matrix of a VARMAX process. In the appendix explicit expressions of the block matrix representations of the Whittle formula are presented.

2. Differentiation of the error process

In this section, the approach used in [15] is extended to a VARMAX process since a new component, the input process $x(t)$, is introduced. This will lead to an appropriate representation for $\partial\varepsilon/\partial\vartheta$. Let us first briefly outline the differentiation rules used in the present paper.

Consider a real, differential $(m \times n)$ matrix function $X(\vartheta)$ of a real $(1 \times \ell)$ vector $\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_\ell)$, where m, n and ℓ are positive integers. Let the $(m \times n)$ matrices $\partial_r X = (\partial X_{ij}/\partial\vartheta_r)$ with $r = 1, 2, \ldots, \ell$ be the first order partial derivatives of $X(\vartheta)$ with X_{ij} the (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 $r = 1, 2, \ldots, \ell$.

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

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

Consider now the VARMAX Eq. (2). From (2) one obtains

$$y(t) = \alpha^{-1}(L)\gamma(L)x(t) + \alpha^{-1}(L)\beta(L)\varepsilon(t). \tag{8}$$

Before applying the appropriate differentiation technique to Eq. (8), we note the properties $\mathrm{d}y(t) = 0$ and $\mathrm{d}x(t) = 0$, where $\mathrm{d}w(t)$ is the total differential of a process $w(t)$ with respect to the parameters $\alpha_1\ \alpha_2\ \cdots\ \alpha_p\ \gamma_0\ \gamma_1\ \cdots\ \gamma_{r-1}\ \gamma_r\ \beta_1\ \beta_2\ \cdots\ \beta_q$, because the realizations of the processes $y(t)$ and $x(t)$ do not depend on these parameters. Below we also use the differential rule, see e.g. [17]

$$\mathrm{d}\alpha^{-1}(L) = -\alpha^{-1}(L)\mathrm{d}\alpha(L)\alpha^{-1}(L). \tag{9}$$

When for typographical brevity the time argument for $x(t)$ and $\varepsilon(t)$ is omitted, we have for the VARMAX process given in (8)

$$\begin{aligned}\mathrm{d}\varepsilon = {} & \beta^{-1}(L)\mathrm{d}\alpha(L)\alpha^{-1}(L)\gamma(L)x - \beta^{-1}(L)\mathrm{d}\gamma(L)x \\ & + \beta^{-1}(L)\mathrm{d}\alpha(L)\alpha^{-1}(L)\beta(L)\varepsilon - \beta^{-1}(L)\mathrm{d}\beta(L)\varepsilon.\end{aligned} \tag{10}$$

We can now develop (10), using (3) and (4) according to

$$\begin{aligned}\frac{\partial\varepsilon}{\partial\vartheta}\Delta\vartheta = {} & \beta^{-1}(L)\{L\Delta\vartheta_1 + L^2\Delta\vartheta_2 + \cdots + L^p\Delta\vartheta_p\}\alpha^{-1}(L)\gamma(L)x \\ & - \beta^{-1}(L)\{\Delta\vartheta_{p+1} + L\Delta\vartheta_{p+2} + L^2\Delta\vartheta_{p+3} + \cdots + L^r\Delta\vartheta_{p+r+1}\}x \\ & + \beta^{-1}(L)\{L\Delta\vartheta_1 + L^2\Delta\vartheta_2 + \cdots + L^p\Delta\vartheta_p\}\alpha^{-1}(L)\beta(L)\varepsilon \\ & - \beta^{-1}(L)\{L\Delta\vartheta_{p+r+2} + L^2\Delta\vartheta_{p+r+3} + \cdots + L^q\Delta\vartheta_{p+r+q+1}\}\varepsilon,\end{aligned} \tag{11}$$

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

The last two terms in (11) are given in [15] and since the method used in this section has been introduced in [15], we will give a short description of the results, emphasizing the differences caused by the presence of the input process.

To construct the first n^2 columns of the matrix $\partial\varepsilon/\partial\vartheta$, we define the $n \times n$ matrix $E_{ij} = e_i e_j^\top$, where e_i and e_j are respectively the ith and jth standard basis vectors in $\mathbb{R}^n$, see [15]. The first n columns will be set forth by considering 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 matrices or block vectors necessary for deriving the first n columns of $\partial\varepsilon/\partial\vartheta$, corresponding to differentiation with respect to ϑ_1, are

$$\begin{pmatrix} E_{i1} & 0_{n\times n} & \cdots & 0_{n\times n} & 0_{n\times m} & \cdots & 0_{n\times m} & 0_{n\times n} & \cdots & 0_{n\times n}\end{pmatrix}.$$

In (11) $\Delta\vartheta_1$ shall consist of the first $n \times n$ matrices E_{i1} with $i = 1, 2, \ldots, n$, whereas all remaining $\Delta\vartheta_j$, where $j = 2, \ldots, p + r + q + 1$, are zero. Consequently, the first n columns of $\partial\varepsilon/\partial\vartheta$ are given by

$$L\beta^{-1}(L)E_{i1}\alpha^{-1}(L)\gamma(L)x + L\beta^{-1}(L)E_{i1}\alpha^{-1}(L)\beta(L)\varepsilon.$$

A similar method is applied to the $n^2 - n$ remaining columns associated with ϑ_1.

We proceed in a similar manner for the remaining columns associated with $\Delta\vartheta_2, \ldots, \Delta\vartheta_{p-1}$ in (11). For ϑ_p the standard basis block vectors are then given by

$$\begin{pmatrix} 0_{n\times n} & \ldots & 0_{n\times n} & \underset{\substack{\uparrow \\ p\text{th } n\times n \text{ block}}}{E_{ij}} & 0_{n\times m} & \ldots & 0_{n\times m} & 0_{n\times n} & \ldots & 0_{n\times n}\end{pmatrix}.$$

The corresponding n^2 columns are given by

$$L^p\beta^{-1}(L)E_{ij}\alpha^{-1}(L)\gamma(L)x + L^p\beta^{-1}(L)E_{ij}\alpha^{-1}(L)\beta(L)\varepsilon.$$

We now consider the construction of the $nm(r+1)$ columns associated with $\vartheta_{p+1}, \vartheta_{p+2}, \ldots, \vartheta_{p+r}, \vartheta_{p+r+1}$. For that purpose we introduce the $n \times m$ matrix $\mathscr{E}_{ij} = e_i^n(e_j^m)^\top$ where e_i^n is the ith standard basis vector in $\mathbb{R}^n$ and e_j^m is the jth standard basis vector in $\mathbb{R}^m$. For a column associated with ϑ_{p+k} where $k = 1, 2, \ldots, r+1$, the standard basis matrix–vector is then

$$\begin{array}{c} \qquad\qquad\qquad\qquad k\text{th } n\times m \text{ block} \\ \qquad\qquad\qquad\qquad\downarrow \\ (0_{n\times n} \quad \ldots \quad 0_{n\times n} \quad 0_{n\times m} \quad \mathscr{E}_{ij} \quad \ldots \quad 0_{n\times m} \quad 0_{n\times n} \quad \ldots \quad 0_{n\times n}). \end{array}$$

The corresponding columns of $\partial\varepsilon/\partial\vartheta$ are then

$$-L^k\beta^{-1}(L)\mathscr{E}_{ij}x.$$

Similarily, as for the first n^2p columns, the n^2q columns associated with $\vartheta_{p+r+2}, \vartheta_{p+r+3}, \ldots, \vartheta_{p+r+q+1}$ have the representation

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

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^2p columns.

We shall summarize the obtained results in Proposition 2.1, an extension of Proposition 3.1 in [15]. For that purpose we define

$$\phi_{ij}^x(L) = \beta^{-1}(L)E_{ij}\alpha^{-1}(L)\gamma(L) \quad \text{and} \quad \phi_{ij}^\varepsilon(L) = \beta^{-1}(L)E_{ij}\alpha^{-1}(L)\beta(L), \tag{12}$$

$$\psi_{ij}^x(L) = -\beta^{-1}(L)\mathscr{E}_{ij} \quad \text{and} \quad \psi_{ij}^\varepsilon(L) = -\beta^{-1}(L)E_{ij}. \tag{13}$$

Put

$$\begin{aligned} \Phi(L) = {} & (\phi_{11}^x(L)x + \phi_{11}^\varepsilon(L)\varepsilon, \phi_{21}^x(L)x + \phi_{21}^\varepsilon(L)\varepsilon, \ldots, \phi_{n1}^x(L)x \\ & + \phi_{n1}^\varepsilon(L)\varepsilon, \phi_{12}^x(L)x + \phi_{12}^\varepsilon(L)\varepsilon, \phi_{22}^x(L)x + \phi_{22}^\varepsilon(L)\varepsilon, \ldots, \phi_{n2}^x(L)x \\ & + \phi_{n2}^\varepsilon(L)\varepsilon, \ldots, \phi_{1n}^x(L)x + \phi_{1n}^\varepsilon(L)\varepsilon, \phi_{2n}^x(L)x + \phi_{2n}^\varepsilon(L)\varepsilon, \ldots, \phi_{nn}^x(L)x \\ & + \phi_{nn}^\varepsilon(L)\varepsilon), \\ \Pi(L) = {} & (\psi_{11}^x(L)x, \psi_{21}^x(L)x, \ldots, \psi_{n1}^x(L)x, \psi_{12}^x(L)x, \psi_{22}^x(L)x, \ldots, \psi_{n2}^x(L)x, \ldots, \\ & \psi_{1m}^x(L)x, \psi_{2m}^x(L)x, \ldots, \psi_{nm}^x(L)x), \\ \Psi(L) = {} & (\psi_{11}^\varepsilon(L)\varepsilon, \psi_{21}^\varepsilon(L)\varepsilon, \ldots, \psi_{n1}^\varepsilon(L)\varepsilon, \psi_{12}^\varepsilon(L)\varepsilon, \psi_{22}^\varepsilon(L)\varepsilon, \ldots, \psi_{n2}^\varepsilon(L)\varepsilon, \ldots, \\ & \psi_{1n}^\varepsilon(L)\varepsilon, \psi_{2n}^\varepsilon(L)\varepsilon, \ldots, \psi_{nn}^\varepsilon(L)\varepsilon), \end{aligned}$$

where the matrices $\Phi(L), \Pi(L)$ and $\Psi(L)$ have dimension $n \times n^2$, $n \times nm$ and $n \times n^2$, respectively.

Proposition 2.1. *The following representation of the* $\{n \times ((n^2(p+q) + mn(r+1))\}$ *matrix* $\partial\varepsilon/\partial\vartheta$ *holds true when the parameter matrix* ϑ *given in* (3) *is considered*:

$$\begin{aligned} \frac{\partial\varepsilon}{\partial\vartheta} & = (L\Phi(L), L^2\Phi(L), \ldots, L^p\Phi(L), \Pi(L), L\Pi(L), L^2\Pi(L), \ldots, L^r\Pi(L), \\ & \quad L\Psi(L), L^2\Psi(L), \ldots, L^q\Psi(L) \\ & = L\{(1, L, L^2, \ldots, L^{p-1}) \otimes \Phi(L), (L^{-1}, 1, L, \ldots, L^{r-1}) \otimes \Pi(L), \\ & \quad (1, L, L^2, \ldots, L^{q-1}) \otimes \Psi(L)\} \\ & = L\{u_p^\top(L) \otimes \Phi(L), L^{-1}u_{r+1}^\top(L) \otimes \Pi(L), u_q^\top(L) \otimes \Psi(L)\}, \end{aligned} \tag{14}$$

where $u_k^\top(L) = (1, L, L^2, \ldots, L^{k-1})$ *for positive integers* k *and* $\otimes$ *is the Kronecker product of two matrices*.

3. Closed form expressions for the entries of the Fisher information matrix

Easily implementable representations of the entries of $\mathscr{F}(\vartheta)$ shall be set forth by applying expression (14)–(7). We proceed with the block representation of $\mathscr{F}(\vartheta)$ which is given by

$$\mathscr{F}(\vartheta)=\begin{pmatrix}\mathscr{F}_{\alpha\alpha}(\vartheta) & \mathscr{F}_{\alpha\gamma}(\vartheta) & \mathscr{F}_{\alpha\beta}(\vartheta)\\ \mathscr{F}_{\gamma\alpha}(\vartheta) & \mathscr{F}_{\gamma\gamma}(\vartheta) & \mathscr{F}_{\gamma\beta}(\vartheta)\\ \mathscr{F}_{\beta\alpha}(\vartheta) & \mathscr{F}_{\beta\gamma}(\vartheta) & \mathscr{F}_{\beta\beta}(\vartheta)\end{pmatrix}. \tag{15}$$

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

$$\mathscr{F}(\vartheta)=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\begin{pmatrix}\mathscr{I}_{\alpha\alpha}(\vartheta) & \mathscr{I}_{\alpha\gamma}(\vartheta) & \mathscr{I}_{\alpha\beta}(\vartheta)\\ \mathscr{I}_{\gamma\alpha}(\vartheta) & \mathscr{I}_{\gamma\gamma}(\vartheta) & \mathscr{I}_{\gamma\beta}(\vartheta)\\ \mathscr{I}_{\beta\alpha}(\vartheta) & \mathscr{I}_{\beta\gamma}(\vartheta) & \mathscr{I}_{\beta\beta}(\vartheta)\end{pmatrix}\frac{\mathrm{d}z}{z} \tag{16}$$

is considered, where the integration in (16) and everywhere below is counterclockwise around the unit circle. We shall first consider an arbitrary entry of block $\mathscr{I}_{\alpha\alpha}(\vartheta)$. The other blocks can be treated similarly. For that purpose a useful equality is introduced. Consider the discrete-time stationary process $w(t)$ where $w(t)=H(L)u(t)$ 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 $w(t)$ and the input $u(t)$, the equality

$$\mathbb{E}_{\vartheta}\{w(t)u^{\top}(t)\}=\int_{-\pi}^{\pi}\Omega_{wu}(\omega)\mathrm{d}\omega,\quad \omega\in[-\pi,\pi] \tag{17}$$

holds true, where $\Omega_{wu}(\omega)$ is the cross spectral density of the proceses $w(t)$ and $u(t)$. It is defined as $\Omega_{wu}(\omega)=H(\mathrm{e}^{\mathrm{i}\omega})\Omega_u(\omega)$ where $\Omega_u(\omega)$ is the spectral density of the input process $u(t)$ which is given by

$$\Omega_u(\omega)=G(\mathrm{e}^{\mathrm{i}\omega})\Omega_v(\omega)G^{*}(\mathrm{e}^{\mathrm{i}\omega}). \tag{18}$$

Expression (18), which is a Hermitian matrix, is the definition of the spectral density of the stationary process $u(t)$. Here 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 (14) is inserted in (7), an arbitrary element of submatrix $\mathscr{F}_{\alpha\alpha}(\vartheta)$ then takes the form

$$\begin{aligned}&\mathbb{E}_{\vartheta}\{\mathrm{Tr}(L^{k+1}\phi^{x}_{ij}(L)x(L^{\nu+1}\phi^{x}_{lf}(L)x)^{\top}\Sigma^{-1})\}\\ &\quad+\mathbb{E}_{\vartheta}\{\mathrm{Tr}(L^{k+1}\phi^{\varepsilon}_{ij}(L)\varepsilon(L^{\nu+1}\phi^{\varepsilon}_{lf}(L)\varepsilon)^{\top}\Sigma^{-1})\},\end{aligned} \tag{19}$$

where $\mathrm{Tr}(M)$ is the trace of a square matrix M and $\nu,k=0,1,2,\ldots,p-1$ and $i,j,l,f=1,2,\ldots,n$. The indices (i,j) and (l,f) are associated with the non-zero elements of the matrices E_{ij} and E_{lf}, respectively, whereas the indices k and ν are associated with the corresponding coefficients α_k and α_ν of the matrix polynomial $\alpha(z)$. We consider the first part of (19) and using formula (17) yields

$$w(t)=L^{k+1}\phi^{x}_{ij}(L)x\quad\text{and}\quad u(t)=\Sigma^{-1}L^{\nu+1}\phi^{x}_{lf}(L)x.$$

The Hermitian positive definite spectral density of the process $u(t)$ is, by virtue of (18), equal to

$$\frac{1}{2\pi}\{\Sigma^{-1}\phi^{x}_{lf}(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})(\phi^{x}_{lf}(\mathrm{e}^{\mathrm{i}\omega}))^{*}\Sigma^{-1}\}.$$

Interchanging expectation $\mathbb{E}_{\vartheta}$ and trace in the first part of (19) and application of (17) leads to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\{\mathrm{e}^{\mathrm{i}\omega(k-\nu)}\phi^{x}_{ij}(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})(\phi^{x}_{lf}(\mathrm{e}^{\mathrm{i}\omega}))^{*}\Sigma^{-1}\}\mathrm{d}\omega.$$

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

$$(\mathscr{F}^{x}_{\alpha\alpha}(\vartheta))^{k,\nu}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{k-\nu}\mathrm{Tr}(\phi^{x}_{ij}(z)R_x(z)(\phi^{x}_{lf}(z))^{*}\Sigma^{-1})\frac{\mathrm{d}z}{z}. \tag{20}$$

The second part of (19) is now considered. We choose

$$w(t)=L^{k+1}\phi^{\varepsilon}_{ij}(L)\varepsilon \quad \text{and} \quad u(t)=\Sigma^{-1}L^{v+1}\phi^{\varepsilon}_{lf}(L)\varepsilon.$$

The connection between the processes $u(t)$ and $w(t)$ is

$$w(t)=L^{k-v}\phi^{\varepsilon}_{ij}(L)(\phi^{\varepsilon}_{lf}(L))^{-1}\Sigma u(t).$$

Since the process ε is white noise, it has a constant spectral density equal to $(1/2\pi)\Sigma$. The spectral density of the process $u(t)$ is then by virtue of (18) equal to

$$\frac{1}{2\pi}\{\Sigma^{-1}\phi^{\varepsilon}_{ij}(\mathrm{e}^{\mathrm{i}\omega})\Sigma(\phi^{\varepsilon}_{lf}(\mathrm{e}^{\mathrm{i}\omega}))^{*}\Sigma^{-1}\}.$$

Interchanging expectation $\mathbb{E}_{\vartheta}$ and trace in the second part of (19) and application of (17) leads to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\{\mathrm{e}^{\mathrm{i}\omega(k-v)}\phi^{\varepsilon}_{ij}(\mathrm{e}^{\mathrm{i}\omega})\Sigma(\phi^{\varepsilon}_{lf}(\mathrm{e}^{\mathrm{i}\omega}))^{*}\Sigma^{-1}\}\mathrm{d}\omega.$$

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

$$(\mathscr{F}^{\varepsilon}_{\alpha\alpha}(\vartheta))^{k,v}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{k-v}\mathrm{Tr}(\phi^{\varepsilon}_{ij}(z)\Sigma(\phi^{\varepsilon}_{lf}(z))^{*}\Sigma^{-1})\frac{\mathrm{d}z}{z}. \tag{21}$$

Both for (20) and (21) we have, $i,j,l,f=1,2,\ldots,n$. Consequently, in agreement with (19)–(21) we have

$$(\mathscr{F}_{\alpha\alpha}(\vartheta))^{k,v}_{i,j,l,f}=(\mathscr{F}^{x}_{\alpha\alpha}(\vartheta))^{k,v}_{i,j,l,f}+(\mathscr{F}^{\varepsilon}_{\alpha\alpha}(\vartheta))^{k,v}_{i,j,l,f}.$$

A similar approach is used for the remaining components of the Fisher information matrix $\mathscr{F}(\vartheta)$. An elementwise representation of $\mathscr{F}(\vartheta)$ in (16) then becomes

$$(\mathscr{F}_{\alpha\gamma}(\vartheta))^{k,g}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{g-k}\mathrm{Tr}(\psi^{x}_{lf}(z)R_x(z)(\phi^{x}_{ij}(z))^{*}\Sigma^{-1})\frac{\mathrm{d}z}{z}, \tag{22}$$

where $i,j,l=1,2,\ldots,n$ and $f=1,2,\ldots,m$. The fact that $R_x(z)$ is Hermitian, yields

$$(\mathscr{F}_{\gamma\alpha}(\vartheta))^{d,v}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{v-d}\mathrm{Tr}(\Sigma^{-1}\phi^{x}_{lf}(z)R_x(z)(\psi^{x}_{ij}(z))^{*})\frac{\mathrm{d}z}{z}, \tag{23}$$

where $g,d=0,1,\ldots,r$ and $i,f,l=1,2,\ldots,n$ and $j=1,2,\ldots,m$.
Furthermore, we have

$$(\mathscr{F}_{\alpha\beta}(\vartheta))^{k,s}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{s-k}\mathrm{Tr}(\psi^{\varepsilon}_{lf}(z)\Sigma(\phi^{\varepsilon}_{ij}(z))^{*}\Sigma^{-1})\frac{\mathrm{d}z}{z} \tag{24}$$

and

$$(\mathscr{F}_{\beta\alpha}(\vartheta))^{c,v}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{v-c}\mathrm{Tr}(\Sigma^{-1}\phi^{\varepsilon}_{lf}(z)\Sigma(\psi^{\varepsilon}_{ij}(z))^{*})\frac{\mathrm{d}z}{z}, \tag{25}$$

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

Finally

$$(\mathscr{F}_{\gamma\gamma}(\vartheta))^{d,g}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{d-g}\mathrm{Tr}(\psi^{x}_{ij}(z)R_x(z)(\psi^{x}_{lf}(z))^{*}\Sigma^{-1})\frac{\mathrm{d}z}{z}, \tag{26}$$

where $i,l=1,2,\ldots,n$ and $j,f=1,2,\ldots,m$. Similarly

$$(\mathscr{F}_{\beta\beta}(\vartheta))^{c,s}_{i,j,l,f}=\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{c-s}\mathrm{Tr}(\psi^{\varepsilon}_{ij}(z)\Sigma(\psi^{\varepsilon}_{lf}(z))^{*}\Sigma^{-1})\frac{\mathrm{d}z}{z}, \tag{27}$$

where $i,j,l,f=1,2,\ldots,n$.

The fact that the input $x(t)$ and the noise $\varepsilon(t)$ are orthogonal processes implies that

$$\mathscr{F}_{\gamma\beta}(\vartheta)=0. \tag{28}$$

The above matrix representations (20)–(27) are multivariate extensions of their scalar couterparts, see [13]. The algorithm developed in [4] can be used for computing (20)–(27) as well as the computer program displayed in [23] and based on [22] algorithm.

For an appropriate computation of $(\alpha(z))^{-1}$, $(\beta(z))^{-1}$, $(\alpha(z^{-1}))^{-1}$ and $(\beta(z^{-1}))^{-1}$ which appear in (20)-(27), we refer to procedures proved in [11] and used in [16]. In [21] closed form expressions for the Fisher information matrix of a VARMA process are set forth. These representations are derived from the Whittle formula where the derivatives of the spectral density of the output process $y(t)$ are considered. The closed form expressions given in [21] as well as (20)–(27) are easily implementable. In [14] the equivalence between the VARMA version of (7) and a matrix-level representation of Whittle's formula is established. In the next section a similar interconnection is set forth for a VARMAX process.

4. An extension of the Whittle formula to VARMAX processes

In this section an extension of the Whittle formula is set forth. An equivalence between two representations is then established. It concerns a time and frequency-domain representation given by (7) and the latter is a circular integral with a Hermitian integrand. First we show that the Whittle formula for VARMA processes derived in the pioneering paper [24] does not yield the Fisher information matrix of VARMAX processes when the spectral density of the observations $y(t)$ given in (2) is considered. For that purpose we recall the known representation of the Fisher information matrix of multiple stationary time series processes or VARMA process developed in [24]

$$\mathscr{F}^{e,h}_{ijkl}(\vartheta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\left(\frac{\partial f_y(\mathrm{e}^{\mathrm{i}\omega})}{\partial \vartheta^e_{ij}} f_y^{-1}(\mathrm{e}^{\mathrm{i}\omega}) \frac{\partial f_y(\mathrm{e}^{\mathrm{i}\omega})}{\partial \vartheta^h_{lk}} f_y^{-1}(\mathrm{e}^{\mathrm{i}\omega})\right) \mathrm{d}\omega, \tag{29}$$

where $f_y(\mathrm{e}^{\mathrm{i}\omega})$ is the spectral density matrix of the process $y(t)$. The spectral density of the process given by (2) with $z = \mathrm{e}^{\mathrm{i}\omega}$ is

$$f_y(z) = \frac{1}{2\pi}(\alpha^{-1}(z)\gamma(z)R_x(z)\gamma^*(z)\alpha^{-*}(z) + \alpha^{-1}(z)\beta(z)\Sigma\beta^*(z)\alpha^{-*}(z)). \tag{30}$$

It can be verified that the representations of the VARMA components of the Fisher information matrix of the VARMAX process (2), as derived in the previous section, are equivalent with the corresponding schemes developed through (29), see [21,15]. It concerns the representations (21), (24), (25) and (27). However, such an equivalence does not hold for the exogenous components of (2) when (29) is applied. This will be illustrated in the next section by means of an example where representation (29) is applied to a parameter associated with an exogenous component of process (2).

4.1. An example of the Whittle formula for a VARMAX process

Consider the following VARMAX process with $n = 2$, $m = 3$ and $p = q = r = 1$. The appropriate matrix polynomials are then

$$\alpha(L) = \begin{pmatrix} 1+\alpha_1^{11}L & \alpha_1^{12}L \\ \alpha_1^{21}L & 1+\alpha_1^{22}L \end{pmatrix}, \quad \beta(L) = \begin{pmatrix} 1+\beta_1^{11}L & \beta_1^{12}L \\ \beta_1^{21}L & 1+\beta_1^{22}L \end{pmatrix},$$
$$\gamma(L) = \begin{pmatrix} \gamma_0^{11}+\gamma_1^{11}L & \gamma_0^{12}+\gamma_1^{12}L & \gamma_0^{13}+\gamma_1^{13}L \\ \gamma_0^{21}+\gamma_1^{21}L & \gamma_0^{22}+\gamma_1^{22}L & \gamma_0^{23}+\gamma_1^{23}L \end{pmatrix}. \tag{31}$$

Since $m = 3$ the input process $x(t) \in \mathbb{R}^3$, the corresponding matrix polynomials are

$$a(L) = \begin{pmatrix} 1+a_1^{11}L & a_1^{12}L & a_1^{13}L \\ a_1^{21}L & 1+a_1^{22}L & a_1^{23}L \\ a_1^{31}L & a_1^{32}L & 1+a_1^{33}L \end{pmatrix} \quad \text{and}$$
$$b(L) = \begin{pmatrix} 1+b_1^{11}L & b_1^{12}L & b_1^{13}L \\ b_1^{21}L & 1+b_1^{22}L & b_1^{23}L \\ b_1^{31}L & b_1^{32}L & 1+b_1^{33}L \end{pmatrix}.$$

We further assume

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{32}$$

The spectral densities are then

$$f_y(\mathrm{e}^{\mathrm{i}\omega}) = \frac{1}{2\pi}(\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})\{\gamma(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})\gamma^*(\mathrm{e}^{\mathrm{i}\omega}) + \beta(\mathrm{e}^{\mathrm{i}\omega})\beta^*(\mathrm{e}^{\mathrm{i}\omega})\}\alpha^{-*}(\mathrm{e}^{\mathrm{i}\omega})),$$
$$R_x(\mathrm{e}^{\mathrm{i}\omega}) = a^{-1}(\mathrm{e}^{\mathrm{i}\omega})b(\mathrm{e}^{\mathrm{i}\omega})b^*(\mathrm{e}^{\mathrm{i}\omega})a^{-*}(\mathrm{e}^{\mathrm{i}\omega}).$$

We shall consider the (γ,γ) block consisting of the parameters of the matrix polynomial $\gamma(z)$ given in (31). Therefore, the appropriate derivatives in (29) with respect to an element of the (γ,γ) block are given by

$$\frac{\partial f_y(\mathrm{e}^{\mathrm{i}\omega})}{\partial \gamma_1^{ij}} = \frac{\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{2\pi}\{\mathrm{e}^{\mathrm{i}\omega}\mathscr{E}_{ij}R_x(\mathrm{e}^{\mathrm{i}\omega})\gamma^*(\mathrm{e}^{\mathrm{i}\omega}) + \mathrm{e}^{-\mathrm{i}\omega}\gamma(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{ij}^\top\}\alpha^{-*}(\mathrm{e}^{\mathrm{i}\omega}) \tag{33}$$

and

$$\frac{\partial f_y(\mathrm{e}^{\mathrm{i}\omega})}{\partial \gamma_0^{lk}} = \frac{\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{2\pi}\{\mathscr{E}_{lk}R_x(\mathrm{e}^{\mathrm{i}\omega})\gamma^*(\mathrm{e}^{\mathrm{i}\omega}) + \gamma(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{lk}^\top\}\alpha^{-*}(\mathrm{e}^{\mathrm{i}\omega}), \tag{34}$$

where $i,l = 1,2$ and $j,k = 1,2,3$.

Recall that $\mathscr{E}_{ij} = e_ie_j^\top$, where e_i and e_j are standard basis vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively. As can be seen, when $f_y^{-1}(\mathrm{e}^{\mathrm{i}\omega})$ and the derivatives (33) and (34) are used in the Whittle formula (29), the matrix polynomial $\gamma(z)$ contributes to the pole location when Cauchy's residue theorem is applied. A similar situation occurs for the remaining parameter blocks that are associated with exogenous components. This is not the case in expression (26), see also [13] for a scalar equivalent of (26). For that purpose we consider (26) for the (γ,γ) block with the matrix polynomials given in (31). First we use

$$\Pi(L) = (\psi_{11}^x(L)x, \psi_{21}^x(L)x, \psi_{12}^x(L)x, \psi_{22}^x(L)x, \psi_{13}^x(L)x, \psi_{23}^x(L)x) \tag{35}$$

as given in Proposition 2.1. We use (31), (32) and (35) for the computation of the 12×12 submatrix $\mathscr{F}_{\gamma\gamma}(\vartheta)$ in (15). This yields for the Fisher information matrix (7)

$$\mathbb{E}_\vartheta\left\{\left(\frac{\partial\varepsilon}{\partial\gamma}\right)^\top\left(\frac{\partial\varepsilon}{\partial\gamma}\right)\right\} = \mathscr{F}_{\gamma\gamma}(\vartheta), \tag{36}$$

where $\partial\varepsilon/\partial\gamma$ is according to Proposition 2.1 given by the 2×12 matrix $\frac{\partial\varepsilon}{\partial\gamma} = (1,L) \otimes \Pi(L)$ or, more explicitely

$$\begin{aligned}\frac{\partial\varepsilon}{\partial\gamma} = -\{&\beta^{-1}(L)\mathscr{E}_{11}x, \beta^{-1}(L)\mathscr{E}_{21}x, \beta^{-1}(L)\\ &\mathscr{E}_{12}x, \beta^{-1}(L)\mathscr{E}_{22}x, \beta^{-1}(L)\mathscr{E}_{13}x, \beta^{-1}(L)\mathscr{E}_{23}x,\\ &L\beta^{-1}(L)\mathscr{E}_{11}x, L\beta^{-1}(L)\mathscr{E}_{21}x, L\beta^{-1}(L)\mathscr{E}_{12}x, L\beta^{-1}(L)\mathscr{E}_{22}x,\\ &L\beta^{-1}(L)\mathscr{E}_{13}x, L\beta^{-1}(L)\mathscr{E}_{23}x\}.\end{aligned} \tag{37}$$

The entries of $\mathscr{F}_{\gamma\gamma}(\vartheta)$ can be computed by inserting representation (37) in (36) and the partitioned form of $\mathscr{F}_{\gamma\gamma}(\vartheta)$ is given by

$$\mathscr{F}_{\gamma\gamma}(\vartheta) = \begin{pmatrix} \mathscr{F}_{\gamma_0\gamma_0}(\vartheta) & \mathscr{F}_{\gamma_0\gamma_1}(\vartheta) \\ \mathscr{F}_{\gamma_1\gamma_0}(\vartheta) & \mathscr{F}_{\gamma_1\gamma_1}(\vartheta) \end{pmatrix}. \tag{38}$$

Using (26) for computing the elements of (38) yields

$$(\mathscr{F}_{\gamma_0\gamma_0}(\vartheta))_{i,j,l,f}^{0,0} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}(\beta^{-1}(z)\mathscr{E}_{ij}R_x(z)\mathscr{E}_{lf}^\top\beta^{-*}(z))\frac{\mathrm{d}z}{z}, \tag{39}$$

$$(\mathscr{F}_{\gamma_0\gamma_1}(\vartheta))^{0,1}_{i,j,l,f} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}z^{-1}(\beta^{-1}(z)\mathscr{E}_{ij}R_x(z)\mathscr{E}_{lf}^{\top}\beta^{-*}(z))\frac{\mathrm{d}z}{z}, \tag{40}$$

$$(\mathscr{F}_{\gamma_1\gamma_0}(\vartheta))^{1,0}_{i,j,l,f} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}z(\beta^{-1}(z)\mathscr{E}_{lf}R_x(z)\mathscr{E}_{ij}^{\top}\beta^{-*}(z))\frac{\mathrm{d}z}{z}, \tag{41}$$

$$(\mathscr{F}_{\gamma_1\gamma_1}(\vartheta))^{1,1}_{i,j,l,f} = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}(\beta^{-1}(z)\mathscr{E}_{ij}R_x(z)\mathscr{E}_{lf}^{\top}\beta^{-*}(z))\frac{\mathrm{d}z}{z}, \tag{42}$$

where $i,l = 1,2$ and $j,f = 1,2,3$. It is straightforward to see that the matrix polynomial $\gamma(z)$ does not affect the pole location in (39)–(42), contrary to (29). This example clearly reveals that the Whittle formula (29) does not correspond to the Fisher information matrix of process (2) when the spectral density of the observations $y(t)$, given in (2), is used. However, as will be shown in the next section, when an appropriate bivariate representation, consisting of the VARMAX process and the exogenous variable, is used, equality of a matrix-level representation of the extended version of (29) and (7) holds true.

To illustrate the computation of (39)–(42) we consider a numerical example where the input process $x(t)$ is driven by a white noise process with covariance Ω given in (32). Additionaly we have for the entries of the matrix polynomial $\beta(z)$ with the setting given in (31), $\beta_1^{11} = 6/5$, $\beta_1^{12} = 1/2$, $\beta_1^{21} = -(7/5)$ and $\beta_1^{22} = -(1/5)$. The basic assumption that the eigenvalues of the matrix polynomial $\beta(z)$ lie outside the unit circle is fulfilled since the eigenvalues are: $(5/23)(-5 \pm \mathrm{i}\sqrt{21})$ with modulus equal to 1.47442. We first choose $(\mathscr{F}_{\gamma_0\gamma_0}(\vartheta))^{0,0}_{1,1,1,1}$, to obtain the following circular integral expression:

$$\begin{aligned}(\mathscr{F}_{\gamma_0\gamma_0}(\vartheta))^{0,0}_{1,1,1,1} &= \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}(\beta^{-1}(z)\mathscr{E}_{11}\mathscr{E}_{11}^{\top}\beta^{-*}(z))\frac{\mathrm{d}z}{z}\\ &= -\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\frac{500z(1-15z+z^2)}{(50+50z+23z^2)(23+50z+50z^2)}\frac{\mathrm{d}z}{z}.\end{aligned}$$

For applying Cauchy's residue theorem we have to consider the poles whithin the unit circle and these are given by the polynomial $(23 + 50z + 50z^2)$. For evaluating the integral, the algorithm developed in [4] or the computer program displayed in [23] and based on the Peterka–Vidinčev [22] algorithm can be implemented. This yields $(\mathscr{F}_{\gamma_0\gamma_0}(\vartheta))^{0,0}_{1,1,1,1} = (\mathscr{F}_{\gamma_1\gamma_1}(\vartheta))^{1,1}_{1,1,1,1} = 7.82242$. We proceed by computing an element of block $\mathscr{F}_{\gamma_0\gamma_1}(\vartheta)$ and $\mathscr{F}_{\gamma_1\gamma_0}(\vartheta)$ involving the parameters γ_1^{13} and γ_0^{23}, to obtain

$$\begin{aligned}(\mathscr{F}_{\gamma_1\gamma_0}(\vartheta))^{1,0}_{1,3,2,3} &= \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}z(\beta^{-1}(z)\mathscr{E}_{13}\mathscr{E}_{23}^{\top}\beta^{-*}(z))\frac{\mathrm{d}z}{z}\\ &= \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\frac{50z^2(-25+89z+70z^2)}{(50+50z+23z^2)(23+50z+50z^2)}\frac{\mathrm{d}z}{z}\\ &= -3.3552\end{aligned} \tag{43}$$

and

$$\begin{aligned}(\mathscr{F}_{\gamma_0\gamma_1}(\vartheta))^{0,1}_{2,3,1,3} &= \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\mathrm{Tr}z^{-1}(\beta^{-1}(z)\mathscr{E}_{23}\mathscr{E}_{13}^{\top}\beta^{-*}(z))\frac{\mathrm{d}z}{z}\\ &= -\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\frac{50(-70-89z+25z^2)}{(50+50z+23z^2)(23+50z+50z^2)}\frac{\mathrm{d}z}{z}.\\ &= -3.3552.\end{aligned} \tag{44}$$

Since the matrix $\mathscr{F}_{\gamma\gamma}(\vartheta)$ is symmetric we have $\mathscr{F}_{\gamma_0\gamma_1}(\vartheta) = \mathscr{F}^{\top}_{\gamma_1\gamma_0}(\vartheta)$ and property (55) proved in Lemma A.1 is numericaly confirmed through the computation of (43) and (44) that results in $(\mathscr{F}_{\gamma_1\gamma_0}(\vartheta))^{1,0}_{1,3,2,3} = (\mathscr{F}_{\gamma_0\gamma_1}(\vartheta))^{0,1}_{2,3,1,3}$.

4.2. The Whittle formula for a VARMAX process

We proceed by presenting a setting which makes Whittle's formula (29) appropriate for the Fisher information matrix of a VARMAX process. The setting proposed in this section has also been applied in [12].

Insert the exogenous variable $x(t) = a^{-1}(L)b(L)\eta(t)$, given by (5), in (2). This leads to

$$\alpha(L)y(t) = \gamma(L)a^{-1}(L)b(L)\eta(t) + \beta(L)\varepsilon(t). \tag{45}$$

Consider the bivariate representation of the VARMAX process (2), based on the exogenous variable $x(t)$ and representation (45),

$$\begin{pmatrix} \alpha(L) & 0_{n\times m} \\ 0_{m\times n} & a(L) \end{pmatrix} \begin{pmatrix} y(t) \\ x(t) \end{pmatrix} = \begin{pmatrix} \beta(L) & \gamma(L)a^{-1}(L)b(L) \\ 0_{m\times n} & b(L) \end{pmatrix} \begin{pmatrix} \varepsilon(t) \\ \eta(t) \end{pmatrix} \tag{46}$$

or equivalently

$$\begin{pmatrix} y(t) \\ x(t) \end{pmatrix} = \begin{pmatrix} \alpha^{-1}(L)\beta(L) & \alpha^{-1}(L)\gamma(L)a^{-1}(L)b(L) \\ 0_{m\times n} & a^{-1}(L)b(L) \end{pmatrix} \begin{pmatrix} \varepsilon(t) \\ \eta(t) \end{pmatrix}. \tag{47}$$

Let us denote the vectors $\begin{pmatrix} y(t) \\ x(t) \end{pmatrix}$ and $\begin{pmatrix} \varepsilon(t) \\ \eta(t) \end{pmatrix}$ by $\xi(t)$ and $\delta(t)$, respectively. Since the white noise processes $\{\varepsilon(t)\}$ and $\{\eta(t)\}$ are not correlated we have

$$\mathbb{E}\{\delta(t)\delta(t)^\top\} = \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix}$$

and the spectral density matrix of $\delta(t)$ is

$$f_\delta = \frac{1}{2\pi} \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix}. \tag{48}$$

The spectral density matrix $f_\xi(e^{i\omega})$, of the extended vector $\xi(t)$ displayed in (47), is obtained by combining (48) and (18)

$$\begin{aligned} f_\xi(e^{i\omega}) = \frac{1}{2\pi} &\begin{pmatrix} \alpha^{-1}(e^{i\omega})\beta(e^{i\omega}) & \alpha^{-1}(e^{i\omega})\gamma(e^{i\omega})a^{-1}(e^{i\omega})b(e^{i\omega}) \\ 0_{m\times n} & a^{-1}(e^{i\omega})b(e^{i\omega}) \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix} \\ &\times \begin{pmatrix} \alpha^{-1}(e^{i\omega})\beta(e^{i\omega}) & \alpha^{-1}(e^{i\omega})\gamma(e^{i\omega})a^{-1}(e^{i\omega})b(e^{i\omega}) \\ 0_{m\times n} & a^{-1}(e^{i\omega})b(e^{i\omega}) \end{pmatrix}^*. \end{aligned} \tag{49}$$

The inverse of (49) is then

$$\begin{aligned} f_\xi^{-1}(e^{i\omega}) = 2\pi &\begin{pmatrix} \beta^{-1}(e^{i\omega})\alpha(e^{i\omega}) & -\beta^{-1}(e^{i\omega})\gamma(e^{i\omega}) \\ 0_{m\times n} & b^{-1}(e^{i\omega})a(e^{i\omega}) \end{pmatrix}^* \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & \Omega^{-1} \end{pmatrix} \\ &\times \begin{pmatrix} \beta^{-1}(e^{i\omega})\alpha(e^{i\omega}) & -\beta^{-1}(e^{i\omega})\gamma(e^{i\omega}) \\ 0_{m\times n} & b^{-1}(e^{i\omega})a(e^{i\omega}) \end{pmatrix}. \end{aligned} \tag{50}$$

The main idea in this section consists of showing that when the spectral density matrix $f_\xi(e^{i\omega})$ is used instead of $f_y(e^{i\omega})$ in Whittle's formula (29), one obtains the Fisher information matrix of multiple time series with exogenous variables.

Property (9) for the derivative of $f_\xi^{-1}(e^{i\omega})$ leads to the following alternative representation of (29) when the spectral density $f_\xi(e^{i\omega})$ is used. It holds that

$$\mathscr{F}^{e,h}_{ijkl}(\vartheta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\left(\frac{\partial f_\xi(e^{i\omega})}{\partial \vartheta^e_{ij}} \frac{\partial f_\xi^{-1}(e^{i\omega})}{\partial \vartheta^h_{kl}} \right) d\omega. \tag{51}$$

Representation (51) of Whittle's formula shall be subsequently used.

In [14], the following interconnection between a time and frequency-domain representation of the Fisher information matrix of a VARMA process has been established:

$$\mathbb{E}_\vartheta\left\{\left(\frac{\partial\varepsilon}{\partial\zeta}\right)^\top \Lambda^{-1}\left(\frac{\partial\varepsilon}{\partial\zeta}\right)\right\} = \frac{1}{4\pi\mathrm{i}}\oint_{|z|=1}\left(\frac{\partial \mathrm{vec} f(z)}{\partial\zeta}\right)^* (f^\top(z)\otimes f(z))^{-1} \times \left(\frac{\partial \mathrm{vec} f(z)}{\partial\zeta}\right)\frac{\mathrm{d}z}{z}, \tag{52}$$

where the right-hand side is a matrix level representation of Whittle's formula for the Fisher information matrix of a VARMA process. The vec operator is defined as $\mathrm{vec}\,X = \mathrm{col}(\mathrm{col}(X_{ij})_{i=1}^n)_{j=1}^n$ and $\mathrm{col}(X_{ij})_{i=1}^n$ refers to the jth column of the matrix X with elements $X_{1j},\ldots,X_{nj}$. The spectral density of the VARMA process is given by $f(z)$, ζ is the parameter vector with representation $\zeta = \mathrm{vec}\{\alpha_1,\alpha_2,\ldots,\alpha_p,\beta_1,\beta_2,\ldots,\beta_q\}$ and Λ is the positive definite covariance matrix of the VARMA white noise process. The left-hand side of (52) is the VARMA equivalent to (7). However, the derivatives in [14] are defined differently than in this paper. The approach used for the VARMA equivalent of (10) consists of vectorizing the $(m\times n)$ matrix function $X(\vartheta)$ introduced in Section 2 according to the following rule:

$$\mathrm{vec}(ABC) = (C^\top\otimes A)\mathrm{vec}\,B \quad \text{where } A\in\mathbb{R}^{m\times n},\ B\in\mathbb{R}^{n\times p} \text{ and } C\in\mathbb{R}^{p\times s}.$$

The $(mn\times\ell)$ matrix $\partial\mathrm{vec}\,X(\vartheta)/\partial\vartheta$, the gradient form of first order derivatives of $X(\vartheta)$, can be defined as $\mathrm{vec}(\mathrm{d}X(\vartheta)) = (\partial(\mathrm{vec}\,X(\vartheta))/\partial\vartheta)\mathrm{d}\vartheta = \mathrm{dvec}\,X(\vartheta)$. Componentwise application of this rule to the VARMA equivalent of (10) results in a different representation of $\partial\varepsilon/\partial\vartheta$ than the one displayed in Proposition 2.1. In Section 2, $\partial\varepsilon/\partial\vartheta$ is set forth at a component-level and involves all the entries of ϑ. As a consequence, the Fisher information matrix displayed in Section 3 has an elementwise representation that involves all the entries of the parameter ϑ. The computation of these entries rely on evaluating integrals of a rational function over the unit circle. This is illustrated in Section 4.1. In [14], as a result of a different method, $\partial\varepsilon/\partial\vartheta$ is given at the vector–matrix level so that the scalar entries of the parameter ϑ can not be directly located in the corresponding Fisher information matrix.

In Theorem 4.1 below, the main result of this section, it is shown that a VARMAX equivalent to (52) can be established when the spectral density of the extended vector $\xi(t)$, based on the setting (46) and (47), is considered.

Theorem 4.1. *The equality*

$$\mathbb{E}_\vartheta\left\{\left(\frac{\partial\varepsilon}{\partial\vartheta}\right)^\top \Sigma^{-1}\left(\frac{\partial\varepsilon}{\partial\vartheta}\right)\right\} = \frac{1}{4\pi\mathrm{i}}\oint_{|z|=1}\left(\frac{\partial \mathrm{vec} f_\xi(z)}{\partial\vartheta}\right)^* (f_\xi^\top(z)\otimes f_\xi(z))^{-1} \times \left(\frac{\partial \mathrm{vec} f_\xi(z)}{\partial\vartheta}\right)\frac{\mathrm{d}z}{z} \tag{53}$$

holds true for the bivariate form (46). *The left-hand side is given by* (7) *and the right-hand side is a representation of Whittle's formula in matrix form applied to the process* $(y(t), x(t))$ *given by* (47).

Proof. We use the principal result of [14], in this paper presented as (52), but now applied to the process $\{\xi(t)\}$, driven by the white noise process $\{\delta(t)\}$ (see (46)) and with parameter vector ϑ. The essential thing to do first is to compute the derivative process $\{\frac{\mathrm{d}\delta}{\mathrm{d}\vartheta}\}$. It immediately follows from (46), that

$$\frac{\mathrm{d}\delta}{\mathrm{d}\vartheta} = \begin{pmatrix}\frac{\mathrm{d}\varepsilon}{\mathrm{d}\vartheta}\\ 0\end{pmatrix}.$$

As a consequence we get that

$$\mathbb{E}_\vartheta\left\{\left(\frac{\partial\delta}{\partial\vartheta}\right)^\top \begin{pmatrix}\Sigma & 0\\ 0 & \Omega\end{pmatrix}^{-1}\left(\frac{\partial\delta}{\partial\vartheta}\right)\right\} = \mathbb{E}_\vartheta\left\{\left(\frac{\partial\varepsilon}{\partial\vartheta}\right)^\top \Sigma^{-1}\left(\frac{\partial\varepsilon}{\partial\vartheta}\right)\right\}. \tag{54}$$

This gives the left-hand side of (53). The right-hand side of this equation is just the right-hand side of (52) with $f = f_\xi$ and identifying ζ with ϑ. $\square$

Alternatively, we can also prove the equality of each submatrix in the left-hand side of (53) and the corresponding submatrices in the right-hand side. For that purpose we use the extended Whittle formula (51) for each submatrix of (15). This is outlined in the appendix. The approach in the appendix can be applied to the VARMAX example (31) to illustrate the correspondence between the extended Whittle formula and the Fisher information matrix of a VARMAX process. Eqs. (56)–(60) are sufficient to illustrate this correspondence. It can also be seen that the computation of (20)–(27) requires much less numerical operations than using the Whittle formula (53).

5. Conclusion

In this paper easily implementable formulas of the Fisher information matrix of a VARMAX process have been derived. The Fisher information matrix set forth in this paper consists of an elementwise representation. The entries are closed form expressions described by circular integrals and can be computed by applying Cauchy's residue theorem. An appropriate extension of the Whittle formula leads to a correspondence with the Fisher information matrix of a VARMAX process. This implies an equality between a time and frequency-domain representation of the Fisher information matrix of VARMAX processes. These results are obtained by using appropriate matrix differential rules. From the numerical point of view it can also be concluded that using representations (20)–(27) is far less computationally expensive than applying the Whittle formula given in the right-hand side of (53).

Acknowledgements

The authors would like to thank an anonymous referee for his valuable comments and suggestions that have improved the quality of the paper and Guy Mélard for his comments on multiple time series.

Appendix A

In this appendix, we present the block matrix representation of the Whittle formula and prove the equivalence to the results in Section 3. First we present a lemma that we repeatedly use in the computations to follow.

Lemma A.1. *Consider a matrix polynomial $A(z) \in \mathbb{C}^{n\times n}$ with real coefficients, the property*

$$\frac{1}{2\pi \mathrm{i}} \oint_{|z|=1} \mathrm{Tr}\, A(z) \frac{\mathrm{d}z}{z} = \frac{1}{2\pi \mathrm{i}} \oint_{|z|=1} \mathrm{Tr}\, A^*(z) \frac{\mathrm{d}z}{z} \tag{55}$$

holds true.

Proof. We shall first prove the following property. If $f(z)$ is analytic inside the unit circle then

$$\frac{1}{2\pi \mathrm{i}} \oint_{\substack{|z|=1 \\ \curvearrowleft}} f(z) \frac{\mathrm{d}z}{z} = \frac{1}{2\pi \mathrm{i}} \oint_{\substack{|z|=1 \\ \curvearrowleft}} f(z^{-1}) \frac{\mathrm{d}z}{z},$$

the integral being taken counter-clockwise.

Set $v = z^{-1}$, then $\frac{\mathrm{d}v}{v} = -\frac{\mathrm{d}z}{z}$, the integral can now be written according to

$$\frac{1}{2\pi \mathrm{i}} \oint_{\substack{|z|=1 \\ \curvearrowleft}} f(z^{-1}) \frac{\mathrm{d}z}{z} = -\frac{1}{2\pi \mathrm{i}} \oint_{\substack{|v|=1 \\ \curvearrowright}} f(v) \frac{\mathrm{d}v}{v} = \frac{1}{2\pi \mathrm{i}} \oint_{\substack{|z|=1 \\ \curvearrowleft}} f(z) \frac{\mathrm{d}z}{z}.$$

We now consider a matrix polynomial $A(z) \in \mathbb{C}^{n\times n}$ with real coefficients. Define $\mathrm{Tr}\, A(z) = f(z)$ where $f(z)$ is a scalar holomorfic function. Then it is straightforward to observe that $\mathrm{Tr}\, A^*(z) = f(z^{-1})$. Consequently, the property above yields equality (55). $\square$

The (γ,γ) block of the Fisher information matrix of the VARMAX process (2) is first considered. The appropriate matrices to be inserted in (51) and which are based on (49) and (50) respectively are introduced, to obtain

$$\begin{aligned}\frac{\partial f_\xi(e^{i\omega})}{\partial \gamma_{ij}^{(d)}} &= \frac{1}{2\pi}\begin{pmatrix}0 & \alpha^{-1}(e^{i\omega})\mathscr{E}_{ij}e^{i\omega d}a^{-1}(e^{i\omega})b(e^{i\omega})\\ 0 & 0\end{pmatrix}\begin{pmatrix}\Sigma & 0\\ 0 & \Omega\end{pmatrix}\\ &\quad\times\begin{pmatrix}\beta^\top(e^{-i\omega})\alpha^{-\top}(e^{-i\omega}) & 0\\ b^\top(e^{-i\omega})a^{-\top}(e^{-i\omega})\gamma^\top(e^{-i\omega})\alpha^{-\top}(e^{-i\omega}) & b^\top(e^{-i\omega})a^{-\top}(e^{-i\omega})\end{pmatrix}\\ &\quad+\frac{1}{2\pi}\begin{pmatrix}\alpha^{-1}(e^{i\omega})\beta(e^{i\omega}) & \alpha^{-1}(e^{i\omega})\gamma(e^{i\omega})a^{-1}(e^{i\omega})b(e^{i\omega})\\ 0 & a^{-1}(e^{i\omega})b(e^{i\omega})\end{pmatrix}\begin{pmatrix}\Sigma & 0\\ 0 & \Omega\end{pmatrix}\\ &\quad\times\begin{pmatrix}0 & 0\\ b^\top(e^{-i\omega})a^{-\top}(e^{-i\omega})\mathscr{E}_{ij}^\top e^{-i\omega d}\alpha^{-\top}(e^{-i\omega}) & 0\end{pmatrix}\\ &= \mathscr{H}(e^{i\omega})+\mathscr{H}^*(e^{i\omega}),\end{aligned}\tag{56}$$

where

$$\mathscr{H}(e^{i\omega}) = \frac{1}{2\pi}\begin{pmatrix}\alpha^{-1}(e^{i\omega})\mathscr{E}_{ij}e^{i\omega d}R_x(e^{i\omega})\gamma^\top(e^{-i\omega})\alpha^{-\top}(e^{-i\omega}) & \alpha^{-1}(e^{i\omega})\mathscr{E}_{ij}e^{i\omega d}R_x(e^{i\omega})\\ 0 & 0\end{pmatrix}$$

and $R_x(e^{i\omega})$ is the spectral density of the process $x(t)$ given in (6).

It is followed by

$$\begin{aligned}\frac{\partial f_\xi^{-1}(e^{i\omega})}{\partial \gamma_{lf}^{(g)}} &= 2\pi\begin{pmatrix}0 & 0\\ -\mathscr{E}_{lf}^\top e^{-i\omega g}\beta^{-\top}(e^{-i\omega}) & 0\end{pmatrix}\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & \Omega^{-1}\end{pmatrix}\\ &\quad\times\begin{pmatrix}\beta^{-1}(e^{i\omega})\alpha(e^{i\omega}) & -\beta^{-1}(e^{i\omega})\gamma(e^{i\omega})\\ 0 & b^{-1}(e^{i\omega})a(e^{i\omega})\end{pmatrix}\\ &\quad+2\pi\begin{pmatrix}\alpha^\top(e^{-i\omega})\beta^{-\top}(e^{-i\omega}) & 0\\ -\gamma^\top(e^{-i\omega})\beta^{-\top}(e^{-i\omega}) & a^\top(e^{-i\omega})b^{-\top}(e^{-i\omega})\end{pmatrix}\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & \Omega^{-1}\end{pmatrix}\\ &\quad\times\begin{pmatrix}0 & -\beta^{-1}(e^{i\omega})\mathscr{E}_{lf}e^{i\omega g}\\ 0 & 0\end{pmatrix}\\ &= \mathscr{P}(e^{i\omega})+\mathscr{P}^*(e^{i\omega}),\end{aligned}\tag{57}$$

where

$$\mathscr{P}(e^{i\omega}) = 2\pi\begin{pmatrix}0 & 0\\ -\mathscr{E}_{lf}^\top e^{-i\omega g}\beta^{-\top}(e^{-i\omega})\Sigma^{-1}\beta^{-1}(e^{i\omega})\alpha(e^{i\omega}) & \mathscr{E}_{lf}^\top e^{-i\omega g}\beta^{-\top}(e^{-i\omega})\Sigma^{-1}\beta^{-1}(e^{i\omega})\gamma(e^{i\omega})\end{pmatrix}.$$

Insertion of (56) and (57) in (51) is the next step. We remind the property that when the square matrices $A_1, A_2, \ldots, A_n$ which do not necessarily have the same dimensions, constitute the main diagonal of a square matrix, then

$$\operatorname{Tr}\begin{pmatrix}A_1 & \ddots & \ddots & \ddots\\ \ddots & A_2 & \ddots & \ddots\\ \ddots & \ddots & \ddots & \ddots\\ \ddots & \ddots & \ddots & A_n\end{pmatrix} = \operatorname{Tr}A_1+\operatorname{Tr}A_2+\cdots+\operatorname{Tr}A_n.$$

Representation (51) for the (γ,γ) block is given by

$$(\mathscr{F}_{\gamma\gamma}(\vartheta))^{d,g}_{i,j,l,f} = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\left(\frac{\partial f_\xi(\mathrm{e}^{\mathrm{i}\omega})}{\partial\gamma_{ij}^{(d)}}\frac{\partial f_\xi^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\gamma_{lf}^{(g)}}\right)\mathrm{d}\omega \tag{58}$$

$$= \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}(\mathscr{E}_{ij}\mathrm{e}^{\mathrm{i}\omega d}R_x(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{lf}^\top \mathrm{e}^{-\mathrm{i}\omega g}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega}))\mathrm{d}\omega \tag{59}$$

$$+ \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}(R_x(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{ij}^\top \mathrm{e}^{-\mathrm{i}\omega d}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{lf}\mathrm{e}^{\mathrm{i}\omega g})\mathrm{d}\omega. \tag{60}$$

Since expression (59) is the complex conjugate transpose of (60) we have by virtue of (55) that (58) becomes for $z = \mathrm{e}^{\mathrm{i}\omega}$

$$\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{d-g}\mathrm{Tr}\{\beta^{-1}(z)\mathscr{E}_{ij}R_x(z)(\beta^{-1}(z)\mathscr{E}_{lf})^*\Sigma^{-1}\}\frac{\mathrm{d}z}{z},$$

which is equal to (26).

A similar approach is applied to the remaining submatrices of the Fisher information matrix. A summary is therefore given. The (α,γ) block is

$$(\mathscr{F}_{\alpha\gamma}(\vartheta))^{k,g}_{i,j,l,f} = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\left(\frac{\partial f_\xi(\mathrm{e}^{\mathrm{i}\omega})}{\partial\alpha_{ij}^{(k)}}\frac{\partial f_\xi^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\gamma_{lf}^{(g)}}\right)\mathrm{d}\omega \tag{61}$$

$$= -\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}(\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{ij}\mathrm{e}^{\mathrm{i}\omega k}\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})\gamma(\mathrm{e}^{\mathrm{i}\omega}) \times R_x(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{lf}^\top\mathrm{e}^{-\mathrm{i}\omega g}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1})\mathrm{d}\omega \tag{62}$$

$$- \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}(R_x(\mathrm{e}^{\mathrm{i}\omega})\gamma^\top(\mathrm{e}^{-\mathrm{i}\omega})\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{ij}^\top\mathrm{e}^{-\mathrm{i}\omega k} \times \beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})\mathscr{E}_{lf}\mathrm{e}^{\mathrm{i}\omega g})\mathrm{d}\omega. \tag{63}$$

Since expression (62) is the complex conjugate transpose of (63), we have by virtue of (55) that (61) becomes for $z = \mathrm{e}^{\mathrm{i}\omega}$

$$-\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{g-k}\mathrm{Tr}\{\beta^{-1}(z)\mathscr{E}_{lf}R_x(z)(\beta^{-1}(z)E_{ij}\alpha^{-1}(z)\gamma(z))^*\Sigma^{-1}\}\frac{\mathrm{d}z}{z}, \tag{64}$$

which is equal to (22).

The integrand in the (γ,α) block integral expression, given by

$$(\mathscr{F}_{\gamma\alpha}(\vartheta))^{d,v}_{i,j,l,f} = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\left(\frac{\partial f_\xi(\mathrm{e}^{\mathrm{i}\omega})}{\partial\gamma_{ij}^{(d)}}\frac{\partial f_\xi^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\alpha_{lf}^{(v)}}\right)\mathrm{d}\omega \tag{65}$$

is the complex conjugate transpose of the integrand in (64), we then have

$$-\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{v-d}\mathrm{Tr}\{\alpha^{-1}(z)\gamma(z)R_x(z)(\beta^{-1}(z)\mathscr{E}_{ij})^*\Sigma^{-1}\beta^{-1}(z)E_{lf}\}\frac{\mathrm{d}z}{z},$$

which is equal to (23).

We now proceed with the (α,α) block which can be written as

$$(\mathscr{F}_{\alpha\alpha}(\vartheta))^{k,v}_{i,j,l,f} = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\left(\frac{\partial f_\xi(\mathrm{e}^{\mathrm{i}\omega})}{\partial\alpha_{ij}^{(k)}}\frac{\partial f_\xi^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\alpha_{lf}^{(v)}}\right)\mathrm{d}\omega \tag{66}$$

$$= \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{Tr}(E_{ij}\mathrm{e}^{\mathrm{i}\omega k}\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega}) \times \beta(\mathrm{e}^{\mathrm{i}\omega})\Sigma\beta^\top(\mathrm{e}^{-\mathrm{i}\omega})\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{lf}^\top\mathrm{e}^{-\mathrm{i}\omega v}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega}))\mathrm{d}\omega \tag{67}$$

$$
\begin{aligned}
&+\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(E_{ij}\mathrm{e}^{\mathrm{i}\omega k}\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})\gamma(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})\gamma^{\top}(\mathrm{e}^{-\mathrm{i}\omega})\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\\
&\times E_{lf}^{\top}\mathrm{e}^{-\mathrm{i}\omega v}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\\
&\times\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega}))\mathrm{d}\omega \qquad (68)\\
&+\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{ij}^{\top}\mathrm{e}^{-\mathrm{i}\omega k}\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{lf}^{\top}\mathrm{e}^{-\mathrm{i}\omega v})\mathrm{d}\omega \qquad (69)\\
&+\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{ij}\mathrm{e}^{\mathrm{i}\omega k}\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega v})\mathrm{d}\omega \qquad (70)\\
&+\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})\beta(\mathrm{e}^{\mathrm{i}\omega})\Sigma\beta^{\top}(\mathrm{e}^{-\mathrm{i}\omega})\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\\
&\times E_{ij}^{\top}\mathrm{e}^{-\mathrm{i}\omega k}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega v})\mathrm{d}\omega \qquad (71)\\
&+\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\Big(\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})\gamma(\mathrm{e}^{\mathrm{i}\omega})R_x(\mathrm{e}^{\mathrm{i}\omega})\gamma^{\top}(\mathrm{e}^{-\mathrm{i}\omega})\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\\
&\times\ E_{ij}^{\top}\mathrm{e}^{-\mathrm{i}\omega k}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega v}\Big)\,\mathrm{d}\omega. \qquad (72)
\end{aligned}
$$

Expression (67) is the complex conjugate transpose of (71), (68) is the complex conjugate transpose of (72) and expression (69) is the complex conjugate transpose of (70). Expression (70) can be represented for $z=\mathrm{e}^{\mathrm{i}\omega}$ as

$$\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{v+k-1}\mathrm{Tr}(\alpha^{-1}(z)E_{ij}\alpha^{-1}(z)E_{lf})\mathrm{d}z. \qquad (73)$$

The scalar equation $\det(\alpha(z))=0$ has all its roots outside the unit circle and the smallest values of the integers v and k is one. Consequently, there are no poles within the unit circle and so the integral (73) is equal to zero. This implies that the integral expressions (69) and (70) vanish, by virtue of (55). The remaining integral expressions summarizing (66) are by virtue of (55) and for $z=\mathrm{e}^{\mathrm{i}\omega}$, given by

$$
\begin{aligned}
&\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{k-v}\mathrm{Tr}\{E_{ij}\alpha^{-1}(z)\beta(z)\Sigma(\beta^{-1}(z)E_{lf}\alpha^{-1}(z)\beta(z))^{*}\Sigma^{-1}\beta^{-1}(z)\}\frac{\mathrm{d}z}{z} \qquad (74)\\
&\quad+\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}z^{k-v}\mathrm{Tr}\{\beta^{-1}(z)E_{ij}\alpha^{-1}(z)\gamma(z)R_x(z)\\
&\quad\times(\beta^{-1}(z)E_{lf}\alpha^{-1}(z)\gamma(z))^{*}\Sigma^{-1}\}\frac{\mathrm{d}z}{z}. \qquad (75)
\end{aligned}
$$

Representations (74) and (75) are equal to (21) and (20) respectively.

The (α,β) block can now be represented according to (51), to obtain

$$
\begin{aligned}
(\mathscr{F}_{\alpha\beta}(\vartheta))_{i,j,l,f}^{k,s}&=-\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\left(\frac{\partial f_{\xi}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\alpha_{ij}^{(k)}}\frac{\partial f_{\xi}^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\beta_{lf}^{(s)}}\right)\mathrm{d}\omega \qquad (76)\\
&=-\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(E_{ij}\mathrm{e}^{\mathrm{i}\omega k}\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})\\
&\quad\times\beta(\mathrm{e}^{\mathrm{i}\omega})\Sigma E_{lf}^{\top}\mathrm{e}^{-\mathrm{i}\omega s}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega}))\mathrm{d}\omega \qquad (77)\\
&\quad-\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{ij}\mathrm{e}^{\mathrm{i}\omega k}\alpha^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega s})\mathrm{d}\omega \qquad (78)\\
&\quad-\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{ij}^{\top}\mathrm{e}^{-\mathrm{i}\omega s}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\mathrm{e}^{-\mathrm{i}\omega k}E_{lf}^{\top})\mathrm{d}\omega \qquad (79)\\
&\quad-\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\Sigma\beta^{\top}(\mathrm{e}^{-\mathrm{i}\omega})\alpha^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{ij}^{\top}\mathrm{e}^{-\mathrm{i}\omega k}\\
&\quad\times\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega s})\mathrm{d}\omega. \qquad (80)
\end{aligned}
$$

Expression (78) is the complex conjugate transpose of (79) and expression (77) is the complex conjugate transpose of (80). Expression (78) can be represented for $z = \mathrm{e}^{\mathrm{i}\omega}$ according to

$$-\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{s+k-1}\mathrm{Tr}(\beta^{-1}(z)E_{ij}\alpha^{-1}(z)E_{lf})\mathrm{d}z. \tag{81}$$

The scalar equations $\det(\alpha(z)) = 0$ and $\det(\beta(z)) = 0$ have all their roots outside the unit circle and the smallest values of the integers s and k is one. Consequently, there are no poles within the unit circle so integral (81) is equal to zero, this implies that the integral expressions (78) and (79) vanish, by virtue of (55). The remaining integral expressions summarizing (76) are by virtue of (55) and for $z = \mathrm{e}^{\mathrm{i}\omega}$ given by

$$-\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{s-k}\mathrm{Tr}\{\Sigma(\beta^{-1}(z)E_{ij}\alpha^{-1}(z)\beta(z))^*\Sigma^{-1}\beta^{-1}(z)E_{lf}\}\frac{\mathrm{d}z}{z}. \tag{82}$$

Considering the representations displayed in (12) and (13) when inserted in (82) yield (24).

The integrand in the (β, α) block integral expression, given by

$$(\mathscr{F}_{\beta\alpha}(\vartheta))^{c,v}_{i,j,l,f} = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\left(\frac{\partial f_\xi(\mathrm{e}^{\mathrm{i}\omega})}{\partial\beta^{(c)}_{ij}}\frac{\partial f_\xi^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\alpha^{(v)}_{lf}}\right)\mathrm{d}\omega \tag{83}$$

is the complex conjugate transpose of (82), this yields

$$-\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{v-c}\mathrm{Tr}\{\alpha^{-1}(z)\beta(z)\Sigma(\beta^{-1}(z)E_{ij})^*\Sigma^{-1}\beta^{-1}(z)E_{lf}\}\frac{\mathrm{d}z}{z},$$

which is equal to (25).

The (β, β) block can be represented according to (51), to obtain

$$(\mathscr{F}_{\beta\beta}(\vartheta))^{c,s}_{i,j,l,f} = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\left(\frac{\partial f_\xi(\mathrm{e}^{\mathrm{i}\omega})}{\partial\beta^{(c)}_{ij}}\frac{\partial f_\xi^{-1}(\mathrm{e}^{\mathrm{i}\omega})}{\partial\beta^{(s)}_{lf}}\right)\mathrm{d}\omega \tag{84}$$

$$= \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(E_{ij}\mathrm{e}^{\mathrm{i}\omega c}\Sigma E_{lf}^\top\mathrm{e}^{-\mathrm{i}\omega s}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega}))\mathrm{d}\omega \tag{85}$$

$$+ \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(E_{ij}\mathrm{e}^{\mathrm{i}\omega c}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega s}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega}))\mathrm{d}\omega \tag{86}$$

$$+ \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(E_{ij}^\top\mathrm{e}^{-\mathrm{i}\omega c}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})E_{lf}^\top\mathrm{e}^{-\mathrm{i}\omega s}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega}))\mathrm{d}\omega \tag{87}$$

$$+ \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}(\Sigma E_{ij}^\top\mathrm{e}^{-\mathrm{i}\omega c}\beta^{-\top}(\mathrm{e}^{-\mathrm{i}\omega})\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf}\mathrm{e}^{\mathrm{i}\omega s})\mathrm{d}\omega. \tag{88}$$

Expression (86) is the complex conjugate transpose of (87) and (85) is the complex conjugate transpose of (88). Expression (86) can be represented for $z = \mathrm{e}^{\mathrm{i}\omega}$ accordingly, to obtain

$$\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{c+s-1}\mathrm{Tr}(E_{ij}\beta^{-1}(z)E_{lf}\beta^{-1}(z))\mathrm{d}z = 0.$$

Using the same arguments as for (81) justifies this conclusion. Consequently, the terms (86) and (87) vanish. The remaining integral expressions summarizing (84) are by virtue of (55) and for $z = \mathrm{e}^{\mathrm{i}\omega}$ given by

$$\frac{1}{2\pi\mathrm{i}}\oint_{|z|=1} z^{c-s}\mathrm{Tr}\{E_{ij}\Sigma(\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})E_{lf})^*\Sigma^{-1}\beta^{-1}(\mathrm{e}^{\mathrm{i}\omega})\}\frac{\mathrm{d}z}{z},$$

which is equal to (27).

References

[1] O.E. Barndorff-Nielsen, R.D. Gill, Fisher Information in quantum statistics, J. Phys. A, 30 (2000) 4481–4490.
[2] O.E. Barndorff-Nielsen, R.D. Gill, P.E. Jupp, On quantum statistical inference II, J. Roy. Statist. Soc. Ser. B 65 (2003) 775–816.

[3] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer-Verlag, Berlin, New York, 1991.
[4] C.J. Demeure, C.T. Mullis, The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences, IEEE Trans. Acoust. Speech Signal Process. 37 (1989) 545–552.
[5] B.R. Frieden, Physics from Fisher Information: A Unification, Cambridge University Press, New York, 1998.
[6] B.R. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, New York, 2004.
[7] B. Friedlander, On the computation of the Cramér–Rao bound for ARMA parameter estimation, IEEE Trans. Acoust. Speech Signal Process. 32 (4) (1984) 721–727.
[8] E.J. Hannan, Multiple Time Series, John Wiley, New York, 1970.
[9] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, John Wiley and Sons, New York, 1988.
[10] E.J. Hannan, W.T.M. Dunsmuir, M. Deistler, Estimation of vector Armax models, J. Multivariate Anal. 10 (1980) 275–295.
[11] I. Gohberg, P. Lancaster, P.L. Rodman, Matrix Polynomials, Academic Press, New York, 1982.
[12] A. Klein, G. Mélard, Computation of the Fisher information matrix for SISO models, IEEE Trans. Signal Process. 42 (1994) 684–688.
[13] 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.
[14] A. Klein, A generalization of Whittle's formula for the information matrix of vector mixed time series, Linear Algebra Appl. 321 (2000) 197–208.
[15] 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.
[16] A. Klein, G. Mélard, P. Spreij, On the resultant property of the Fisher information matrix of a vector ARMA process, Linear Algebra Appl. 403 (2005) 291–313.
[17] P. Lancaster, M. Tismenetsky, The Theory of Matrices with Applications, second ed., Academic Press, Orlando, 1985.
[18] L. Ljung, T. Söderström, Theory and Practice of Recursive Identification, MIT Press, Cambridge, Mass, 1983.
[19] E.C. MacRae, Matrix derivatives with an application to an adaptive linear decision problem, Ann. Statist. 2 (1974) 337–346.
[20] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, third ed., John Wiley and Sons, Chichester, New York, 2007.
[21] H.J. Newton, The information matrices of multiple time series, J. Multivariate Anal. 8 (1978) 317–323.
[22] V. Peterka, P. Vidinčev, Rational-fraction approximation of transfer functions, in: First IFAC Symposium on Identification in Automatic Control Systems, Prague, 1967.
[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] P. Whittle, The analysis of multiple stationary time series, J. Roy. Statist. Soc. Ser. B 15 (1953) 125–139.
[25] C.S. Wong, On the use of differentials in statistics, Linear Algebra Appl. 70 (1985) 285–299.