Invertibility Condition of the Fisher Information Matrix of a VARMAX Process and the Tensor Sylvester Matrix

André Klein
University of Amsterdam

Guy Mélard
ECARES, Université libre de Bruxelles

April 2020

ECARES working paper 2020-11

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Invertibility Condition of the Fisher Information Matrix of a VARMAX Process and the Tensor Sylvester Matrix

André Klein* Guy Mélard†

Abstract

In this paper the invertibility condition of the asymptotic Fisher information matrix of a controlled vector autoregressive moving average stationary process, VARMAX, is displayed in a theorem. It is shown that the Fisher information matrix of a VARMAX process becomes invertible if the VARMAX matrix polynomials have no common eigenvalue. Contrarily to what was mentioned previously in a VARMA framework, the reciprocal property is untrue. We make use of tensor Sylvester matrices since checking equality of the eigenvalues of matrix polynomials is most easily done in that way. A tensor Sylvester matrix is a block Sylvester matrix with blocks obtained by Kronecker products of the polynomial coefficients by an identity matrix, on the left for one polynomial and on the right for the other one. The results are illustrated by numerical computations.

MSC Classification: 15A23, 15A24, 60G10, 62B10.

Keywords: Tensor Sylvester matrix; Matrix polynomial; Common eigenvalues; Fisher information matrix; Stationary VARMAX process.

Declaration of interest: none

*University of Amsterdam, The Netherlands, Rothschild Blv. 123 Apt.7, 6527123 Tel-Aviv, Israel (e-mail: A.A.B.Klein@contact.uva.nl).

†Université libre de Bruxelles, Solvay Brussels School of Economics and Management, ECARES, avenue Franklin Roosevelt 50 CP 114/04, B-1050 Brussels, Belgium (e-mail: gmelard@ulb.ac.be). **Corresponding author**

1 Introduction

This paper investigates the invertibility condition of the Fisher information matrix of a Gaussian vector ARMAX or VARMAX process. Controlled vector autoregressive moving average stationary processes, VARMAX processes are general-purpose representations in order to describe dynamic systems in engineering and in econometrics. At first, in order to set forth an invertibility condition of the Fisher information matrix, we worked on a factorization of the information matrix as derived in [13], like we did for VARMA models in [12], making use of tensor Sylvester resultant matrices. Such a tensor Sylvester matrix is associated to two monic matrix polynomials and it becomes singular if and only if the two matrix polynomials have at least one common eigenvalue, see [6]. In [12], it is said that the Fisher information matrix of a VARMA process, a VARMAX process without the control or exogenous variable, becomes invertible if and only if the tensor Sylvester matrix is invertible, in other words, if and only if the autoregressive and moving average matrix polynomials of the VARMA process have no common eigenvalue. This is called the Sylvester resultant property. As will be shown, the "'only if"' part of that result is, however, wrong when the dimension of the process is larger than 1. In [11], the Sylvester resultant property is shown for a scalar ARMAX process but it is no longer true, in general, for other than scalar ARMA or ARMAX processes. In the present paper, we correct the assertion stated in [12] for VARMA processes, and we extend the "'if"' part to a class of VARMAX processes. Although the results are less powerful, they are what is needed in practice: a necessary condition of invertibility of the Fisher information matrix, indicating a possible lack of identifiability so that parameter estimation is risky. A sufficient condition that should involve more information about the matrix polynomials is not as useful.

Consider the vector stochastic difference equation representation of a linear system of order (p, r, q) of the Gaussian process $\{y_t, t \in \mathbb{Z}\}$, $\mathbb{Z}$ is the set of integers. To be more specific, consider the equation representation of a dynamic linear system,

$$\sum_{j=0}^{p} A_j y_{t-j} = \sum_{j=0}^{r-1} C_j x_{t-j} + \sum_{j=0}^{q} B_j \varepsilon_{t-j}, \quad t \in \mathbb{Z} \tag{1}$$

where y_t, x_t and ε_t are respectively, the n-dimensional stochastic observed output, the n-dimensional observed input, and the n-dimensional unobserved errors, and where $A_j \in \mathbb{R}^{n\times n}$, $j = 1, ..., p$, $C_j \in \mathbb{R}^{n\times n}$, $j = 0, ..., r-1$, and $B_j \in \mathbb{R}^{n\times n}$, $j = 1, ..., q$, are associated parameter matrices. We additionally assume $A_0 = B_0 = I_n$, the $n \times n$ identity matrix. We suppose that C_0 is an invertible matrix. In the examples we will take $C_0 = I_n$ fixed instead of being a matrix of parameters so that the maximum lag $r-1$ is replaced by r. Note that the absence of lag induced by (1) is purely conventional. For example, a lagged effect of x'_t on y_t can be produced by defining $x_t = x'_{t-1}$. The error $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of independent zero mean n-dimensional Gaussian random variables with a strictly positive definite covariance matrix Σ. We denote this by $\Sigma \succ 0$. We shall denote transposition by ^{T} and the mathematical expectation by $\mathbb{E}$. We assume $\mathbb{E}\{x_t \varepsilon_{t'}^T\} = 0$, for all t, t'. We denote z the backward shift operator, for example

$zx_t = x_{t-1}$. Then (1) can be written as

$$A(z)y_t = C(z)x_t + B(z)\varepsilon_t, \tag{2}$$

where

$$A(z) = \sum_{j=0}^{p} A_j z^j, \quad C(z) = \sum_{j=0}^{r-1} C_j z^j, \quad B(z) = \sum_{j=0}^{q} B_j z^j \tag{3}$$

are the associated matrix polynomials, where $z \in \mathbb{C}$ (the duplicate use of z as an operator and as a complex variable is usual in the signal and time series literature, e.g., [4], [2] and [10]). The assumptions $\det(A(z)) \neq 0$ far all $|z| \leq 1$ (causality) and $\det(B(z)) \neq 0$ for all $|z| \leq 1$ (invertibility) are imposed so that the eigenvalues of the matrix polynomials $A(z)$ and $B(z)$ will be outside the unit circle. Remind that the eigenvalues of a square matrix polynomial $A(z)$ are the roots of the equation $\det(A(z)) = 0$.

Remark 1.1. Note that there are restrictions in the VARMAX model being considered. First, the dimensions of the unobserved errors ϵ_t and of the observed output y_t are the same. This is often the case in the literature, although [3], for example, consider a VARMA model where the dimension of the unobserved errors is smaller than that of the observed output. Second, the dimensions of the observed input x_t and of the observed output y_t are the same. This is more restrictive. For example, it prevents to have a one-dimension input, which is a frequent case, or an input with a higher dimension than the output. The paper [14] that will be mentioned later does not require this. At this time it does not seem possible to avoid these constraints.

We store the VARMAX (p, r, q) coefficients in an $l = n^2(p + q + r) \times 1$ vector ϑ defined as follows

$$\vartheta = \mathrm{vec}\{A_1, A_2, \ldots, A_p, C_0, \ldots, C_{r-1}, B_1, B_2, \ldots B_q\}.$$

The vec operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other, according to vec $X = \mathrm{col}(\mathrm{col}(X_{ij})_{i=1}^{n})_{j=1}^{n}$, see e.g. [7], [16].
The observed input variable x_t is assumed to be a stationary n-dimensional Gaussian VARMA process with white noise process η_t satisfying $\mathbb{E}\{\eta_t \eta_t^T\} = \Omega$ and

$$a(z)x_t = b(z)\eta_t,$$

where $a(z)$ and $b(z)$ are respectively the autoregressive and moving average matrix polynomials such that $a(0) = b(0) = I_n$ and $\det(a(z)) \neq 0$ far all $|z| \leq 1$ and $\det(b(z)) \neq 0$ for all $|z| \leq 1$. The spectral density of process x_t is defined as, see e.g. [4]

$$R_x(e^{i\omega}) = a^{-1}(e^{i\omega})b(e^{i\omega})\Omega\, b^*(e^{i\omega})a^{*-1}(e^{i\omega})/(2\pi), \quad \omega \in [-\pi, \pi], \tag{4}$$

where i is the standard imaginary unit, ω is the frequency, the spectral density matrix $R_x(e^{i\omega})$ is Hermitian, and we further have, $R_x(e^{i\omega}) \succ 0$ and $\int_{-\pi}^{\pi} R_x(e^{i\omega})d\omega < \infty$. X^* is the complex conjugate transpose of matrix X. Therefore there is at least one solution of (1) which is stationary.

Before we display the Fisher information matrix we present the tensor Sylvester matrix.

2 The Tensor Sylvester Matrix

Consider the matrix polynomials $A(z) = \sum_{i=0}^{p} A_i z^i$ and $B(z) = \sum_{j=0}^{q} B_j z^j$, where A_p and B_q are invertible matrices, the $n(p+q) \times n(p+q)$ Sylvester matrix is defined as

$$S_{pq}(B,A) =: \begin{pmatrix} B_0 & B_1 & \dots & B_q & O_{n\times n} & \dots & O_{n\times n} \\ O_{n\times n} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & O_{n\times n} \\ O_{n\times n} & \dots & O_{n\times n} & B_0 & B_1 & \dots & B_q \\ A_0 & A_1 & \dots & A_p & O_{n\times n} & \dots & O_{n\times n} \\ O_{n\times n} & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & O_{n\times n} \\ O_{n\times n} & \dots & O_{n\times n} & A_0 & A_1 & \dots & A_p \end{pmatrix}. \tag{5}$$

Let us partition $S_{pq}(B,A)$ in two blocks with $n(p+q)$ columns and of numbers of rows respectively np and nq, and define $S_p(B)$ the upper matrix and $S_q(A)$ the lower matrix. Hence

$$S_{pq}(B,A) = \begin{pmatrix} S_p(B) \\ S_q(A) \end{pmatrix}.$$

If $n = 1$, then it is well known (e.g. [5]) that two monic scalar polynomials A(z) and B(z) have at least a common root if and only if $\det S(-B,A) = 0$ or when the matrix $S(-B,A)$ is singular. In [6], Gohberg and Lerer show that that property is no longer true if $n > 1$, i.e. for matrix polynomials. Let the operator $\otimes$ which represents the Kronecker product of two matrices. Following [6], we introduce the so-called tensor Sylvester matrix on the basis of (5)

$$S^{\otimes}_{pq}(B,A) = \begin{pmatrix} S_p(B \otimes I_n) \\ S_q(I_n \otimes A) \end{pmatrix},$$

where $\otimes$ is Kronecker's product. Similarly, for two matrix polynomials $A(z) = \sum_{i=0}^{p} A_i z^i$ and $C(z) = \sum_{j=0}^{r} C_j z^j$, with C_r invertible, we will need the $n(p+r) \times n(p+r)$ Sylvester matrix defined as

$$S_{pr}(C,A) = \begin{pmatrix} S_p(C) \\ S_q(A) \end{pmatrix},$$

and

$$S^{\otimes}_{pr}(C,A) = \begin{pmatrix} S_p(C \otimes I_n) \\ S_r(I_n \otimes A) \end{pmatrix}.$$

It is said in [6] with a sketch of proof that the resultant property holds for tensor Sylvester matrices, i.e. the two matrix polynomials $A(z)$ and $B(z)$ have at least a common eigenvalue if and only if $\det S^{\otimes}(-B,A) = 0$ or when the matrix $S^{\otimes}(-B,A)$ is singular. Similarly, to assess if the two matrix polynomials $A(z)$ and $C(z)$ have at least a common eigenvalue, we will investigate singularity of the matrix $S^{\otimes}(-C,A)$.

Example 2.1. We will treat in Section 5 three examples with $n = 2$ and $p = r-1 = q = 1$. Let A_{ik}, C_{ik}, B_{ik}, $i,k = 1,2$ denote the respective first degree coefficients A_1, C_1, and B_1, and $C_0 = I_2$ instead of being a matrix of parameters, in addition to $A_0 = B_0 = I_2$, so that $r-1$ is replaced by r. Then $S^{\otimes}_{pq}(-B,A)$ and $S^{\otimes}_{pr}(-C,A)$ take the form of 8×8 matrices

$$\begin{aligned} S^{\otimes}_{pq}(-B,A) &= \begin{pmatrix} S_p(-B \otimes I_n) \\ S_q(I_n \otimes A) \end{pmatrix} = \begin{pmatrix} -I_4 & -B_1 \otimes I_2 \\ I_4 & I_2 \otimes A_1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 0 & 0 & 0 & -B_{11} & 0 & -B_{12} & 0 \\ 0 & -1 & 0 & 0 & 0 & -B_{11} & 0 & -B_{12} \\ 0 & 0 & -1 & 0 & -B_{21} & 0 & -B_{22} & 0 \\ 0 & 0 & 0 & -1 & 0 & -B_{21} & 0 & -B_{22} \\ 1 & 0 & 0 & 0 & A_{11} & A_{12} & 0 & 0 \\ 0 & 1 & 0 & 0 & A_{21} & A_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & A_{11} & A_{12} \\ 0 & 0 & 0 & 1 & 0 & 0 & A_{21} & A_{22} \end{pmatrix}. \end{aligned} \tag{6}$$

and

$$S^{\otimes}_{pr}(-C,A) = \begin{pmatrix} S_p(-C \otimes I_n) \\ S_r(I_n \otimes A) \end{pmatrix} = \begin{pmatrix} -I_4 & -C_1 \otimes I_2 \\ I_4 & I_2 \otimes A_1 \end{pmatrix},$$

and a detailed expression similar to (6) with B_{ik} replaced by C_{ik}, $i,k = 1,2$.

In [12] it is said incorrectly that the Fisher information matrix of a VARMA process (i.e. the case without x_t) fulfills the resultant property. This is done by inserting the tensor Sylvester matrix in an appropriate factorized form of the Fisher information matrix. Then, another matrix $\mathcal{M}$ is introduced and it is shown in [12, Lemma 2.3] that it is singular if and only if the matrix polynomials have at least one common eigenvalue. This is correct. The mistake appears in the "‘if"' part of [12, Proposition 2.4] when it is said that the Fisher information matrix is singular if and only if $\mathcal{M}$ is singular. Indeed, using the notations for the matrices $\mathcal{J}(e^{i\omega})$ and $\Lambda(e^{i\omega})$ introduced, respectively, in [12, bottom of p. 303] and [12, top of p. 304], and y in the kernel of $\mathcal{M}$, the claim that the circular integral of $y^*\Lambda(e^{i\omega})\mathcal{J}(e^{i\omega})\Lambda^*(e^{i\omega})y$ in [12, bottom of p. 305] equals zero implies that $\Lambda(e^{i\omega})\mathcal{J}(e^{i\omega})\Lambda^*(e^{i\omega})y$ is identically zero is wrong, breaking the remaining of the "proof". See Remark 2.2 for a counterexample. This mistake was pinpointed by [17] with the help of the other co-authors of [12].

Remark 2.2. That the implication mentioned above is wrong can be checked on a simple univariate ARMA(1,1) model with $A(z) = 1 + \vartheta_1 z$ and $B(z) = 1 + \vartheta_2 z$, and a common root $\vartheta_1 = \vartheta_2$, and assume causality and invertibility so that $-1 < \vartheta_1 < 1$. Then the matrix $\mathcal{M}(\vartheta_\infty)$ defined in [12, Equation (18)] is a singular matrix and its kernel has rows $(-1-\vartheta_1)$ and $(1 \;\; \vartheta_1)$.

Take $y = (y_1, y_2)^T$ as a vector of the kernel with, e.g. $y_2 = 1$. Then $y_1 = -\vartheta_1$. We follow the proof in [12, Proposition 2.4 p. 305, lines -7 and -6], we need to study $\Lambda(z)\mathcal{J}(z)\Lambda^*(z)y$ which is a vector with 2 components. Straightforward calculation for $\vartheta_1 = \vartheta_2$ leads to the following two components: $y_1 - y_2$ and $y_2 - y_1$. Replacing y_1 and y_2 yields respectively $-\vartheta_1 - 1 < 0$ and $1 + \vartheta_1 > 0$. The strict inequalities come from the causality condition. Clearly these two integrands are not identically 0, breaking down the conclusion of [12, p. 305, lines -6] that $\Lambda(e^{i\omega})\mathcal{J}(e^{i\omega})\Lambda^*(e^{i\omega})y \equiv 0$. Hence the determinant of $\Lambda(e^{i\omega})\mathcal{J}(e^{i\omega})\Lambda^*(e^{i\omega})$ is not equal to 0 and the remaining is false.

For that reason, we will use the results in [13], where tensor Sylvester matrices are part of a factorized form of the Fisher information matrix of a VARMAX process only for the description that it provides. We will rather use identifiability conditions stated in [8] and [9], see also [10]. Like in [11] for the case $n = 1$, the presence of the third polynomial C(z) has to be taken into account in the resultant property and this appears not at all trivial, as will be shown.

In the next section the Fisher information matrix $\mathcal{F}(\vartheta)$ of the VARMAX process is displayed according to [13].

3 The Fisher Information Matrix of VARMAX Processes

The Fisher information is an ingredient of the Cramér-Rao inequality. Some structured matrix properties of the Fisher information matrix of Gaussian stationary processes have been investigated, see e.g. [15].
When time series models are the subject, using (2) for all $t \in \mathbb{Z}$ to determine the residual $\varepsilon_t(\vartheta)$, to emphasize the dependency on the parameter vector, ϑ, and assuming that x_t is stochastic and that (y_t, x_t) is a Gaussian stationary process, the asymptotic Fisher information matrix, $\mathcal{F}(\vartheta)$, is defined by the following $(l \times l)$ matrix, which does not depend on t:

$$\mathcal{F}(\vartheta) = \mathbb{E}\left\{ \left(\frac{\partial \varepsilon_t(\vartheta)}{\partial \vartheta^T} \right)^T \Sigma^{-1} \left(\frac{\partial \varepsilon_t(\vartheta)}{\partial \vartheta^T} \right) \right\} \tag{7}$$

where the $(v \times l)$ matrix $\partial(\cdot)/\partial\vartheta^T$, the derivative with respect to ϑ^T, for any $(v \times 1)$ column vector $(\cdot)$ and l is the total number of parameters and ϑ is defined in (1). Equality (7) is used for computing the Fisher information matrix of the various time series processes, see [13] for more details. The derivative, $\partial\varepsilon_t/\partial\vartheta^T$, of size $n \times n^2(p+q+r)$, is computed in [13], to obtain

$$\begin{aligned}\frac{\partial\varepsilon_t}{\partial\vartheta^T} &= \{(A^{-1}(z)C(z)x_t)^T \otimes B^{-1}(z)\}\frac{\partial \operatorname{vec} A(z)}{\partial\vartheta^T} \\ &+\{(A^{-1}(z)B(z)\varepsilon_t)^T \otimes B^{-1}(z)\}\frac{\partial \operatorname{vec} A(z)}{\partial\vartheta^T} \\ &-\{x_t^T \otimes B^{-1}(z)\}\frac{\partial \operatorname{vec} C(z)}{\partial\vartheta^T} \\ &-(\varepsilon_t^T \otimes B^{-1}(z))\frac{\partial \operatorname{vec} B(z)}{\partial\vartheta^T}.\end{aligned} \tag{8}$$

By substituting $\partial\varepsilon_t/\partial\vartheta^T$ under form (8) in (7) yields the representation of the Fisher information matrix of a VARMAX process. For each positive integer k, denote $u_k(z) = (1, z, z^2, \ldots, z^{k-1})^T$. Let

$$\mathcal{L}(A(z)) = \begin{pmatrix} I_p \otimes A(z) \otimes I_n & O_{pn^2\times rn^2} & O_{pn^2\times qn^2} \\ O_{rn^2\times pn^2} & O_{rn^2\times rn^2} & O_{rn^2\times qn^2} \\ O_{qn^2\times pn^2} & O_{qn^2\times rn^2} & I_q \otimes I_n \otimes A(z) \end{pmatrix},$$

$$\mathcal{W}(A(z)) = \begin{pmatrix} I_p \otimes A(z) \otimes I_n & O_{pn^2\times rn^2} & O_{pn^2\times qn^2} \\ O_{rn^2\times pn^2} & I_r \otimes I_n \otimes A(z) & O_{rn^2\times qn^2} \\ O_{qn^2\times pn^2} & O_{qn^2\times rn^2} & O_{qn^2\times qn^2} \end{pmatrix},$$

$$\Phi(z) = \mathcal{L}(A^{-1}(z)) \begin{pmatrix} S_p^{\otimes}(-B) \\ O_{rn^2\times n^2(p+q)} \\ S_q^{\otimes}(A) \end{pmatrix} (u_{p+q}(z) \otimes I_{n^2}),$$

$$\Lambda(z) = \mathcal{W}(A^{-1}(z)) \begin{pmatrix} S_p^{\otimes}(-C) \\ S_r^{\otimes}(A) \\ O_{qn^2\times n^2(p+r)} \end{pmatrix} (u_{p+r}(z) \otimes I_{n^2}),$$

$$S_{p,q}^{\otimes}(-B, A) = \begin{pmatrix} S_p^{\otimes}(-B) \\ S_q^{\otimes}(A) \end{pmatrix}, S_{p,r}^{\otimes}(-C, A) = \begin{pmatrix} S_p^{\otimes}(-C) \\ S_r^{\otimes}(A) \end{pmatrix},$$

$$\Psi(z) = R_x(z) \otimes B^{-T}(z)\Sigma^{-1}B^{-1}(z^{-1}) \quad \text{and} \quad \Theta(z) = \Sigma \otimes B^{-T}(z)\Sigma^{-1}B^{-1}(z^{-1}),$$

where the Hermitian spectral density matrix, $R_x(z)$, is defined in (4) . Let

$$\mathcal{A}(z) := \Phi(z)\Theta(z)\Phi^*(z), \quad \mathcal{B}(z) := \Lambda(z)\Psi(z)\Lambda^*(z).$$

Then, as shown in [13], the Fisher information matrix is given by

$$\mathcal{F}(\vartheta) = \frac{1}{2\pi i}\oint_{|z|=1} \mathcal{A}(z)\frac{dz}{z} + \frac{1}{2\pi i}\oint_{|z|=1} \mathcal{B}(z)\frac{dz}{z}.$$

The integrals are counterclockwise.

There are other ways to write the information matrix but the way that was selected makes use of tensor Sylvester matrices.

4 Main Result

First, we need to remind a few concepts, those of monic polynomials, reciprocal polynomials, eigenvalues of matrix polynomials, unimodular polynomial matrix, and common left divisor of matrix polynomials.

A monic square matrix polynomial $A(z)$ of dimension n and degree p is such that the coefficient of degree p is I_n.

The reciprocal polynomials of the polynomials $A(z)$, $B(z)$ and $C(z)$ of respective degrees p, q and $r-1$ are, respectively, $A^*(z) = z^p A(z^{-1})$, $B^*(z) = z^q B(z^{-1})$ and $C^*(z) = z^{r-1} C(z^{-1})$. $A^*(z)$ and $B^*(z)$ are monic polynomials since $A(0) = B(0) = I_n$. $C^*(z)$ is not necessarily a monic polynomial but, since C_0 is invertible, $C_0^{-1} C^*(z)$ is a monic polynomial and it is easy to see that [11] is applicable.

The eigenvalues of a $n \times n$ matrix polynomial $D(z)$ of degree s are the roots of the equation $\det(D(z)) = 0$ which is a polynomial equation of degree less or equal to ns. If the degree of $\det(D(z))$ is smaller than ns, the number s_n, say, of roots of the equation $\det(D(z)) = 0$ is smaller than ns. We can speak of missing roots for $\det(D(z)) = 0$ or missing eigenvalues for $D(z)$. Note however that the degree of $\det(D^*(z))$ is equal to ns and that the number of roots of the equation $\det(D^*(z)) = 0$ is equal to ns. The s_n roots of $\det(D(z)) = 0$ are the inverse of the roots of $\det(D^*(z)) = 0$ that are strictly different from zero, and, on the other hand, the missing roots of $\det(D(z)) = 0$ correspond to the roots of $\det(D^*(z)) = 0$ equal to 0. We will convene for the sequel of this paper that the possible missing roots of $\det(D(z)) = 0$ or missing eigenvalues of $D(z)$ are equal to infinity, so that all the eigenvalues of $D^*(z)$ are the inverse of those of $D(z)$. We have supposed that the eigenvalues of $A(z)$ and $B(z)$ in (3) are all strictly larger than one in modulus. The eigenvalues of the reciprocal polynomials $A^*(z)$ and $B^*(z)$ are therefore all strictly smaller than one in modulus.

A unimodular polynomial matrix $U(z)$ is a polynomial square matrix such that its determinant is a non-zero constant, instead of being a polynomial in z. Consequently $U^{-1}(z)$ is also a polynomial matrix. In general, the inverse of a square polynomial matrix is a matrix with rational elements, not polynomial elements.

Three $n \times n$ matrix polynomials $A(z)$, $B(z)$, and $C(z)$ have a common left divisor if there exist $n \times n$ polynomial matrices $F(z)$, $A'(z)$, $B'(z)$ and $C'(z)$ such that $A(z) = F(z)A'(z)$, $B(z) = F(z)B'(z)$, and $C(z) = F(z)C'(z)$. In matrix form we can write $(A(z)\ B(z)\ C(z)) = F(z)(A'(z)\ B'(z)\ C'(z))$. Then $(A(z)\ B(z)\ C(z))$ is called a right multiple of $F(z)$. A left divisor $F(z)$ is called the greatest common left divisor of $(A(z)\ B(z)\ C(z))$ if it is a right multiple of all left divisors. Multiplying a greatest common left divisor to the right by any unimodular polynomial matrix yields another greatest common left divisor. As shown by [10], a greatest common left divisor can be constructed by elementary column operations: interchange any two columns, multiply any column by a real number different from 0, add a polynomial multiple of any column to any column. Also, it corresponds to right multiplication of $(A(z)\ B(z)\ C(z))$ by an appropriate unimodular matrix. The concept can also be defined for rectangular matrices with the same number of rows although we will not consider that generalization.

Lemma 4.1. *Assume that $A(z)$, $B(z)$ and $C(z)$ have the prescribed degrees, respectively p, q and $r-1$, with $\det(A(z)) \neq 0$ and $\det(B(z)) \neq 0$, for z in the unit circle of $\mathbb{C}$, and Σ is non-singular. Assume also that x_t has an absolutely continuous spectrum with spectral-density non-zero on a set of positive measure on $(-\pi, \pi]$. Then a necessary and sufficient condition of identifiability of a stationary VARMAX process satisfying* (1) *is (i) $A(z)$, $B(z)$ and $C(z)$ have I_n as greatest common left divisor, (ii) the matrix $(A_p \; B_q \; C_{r-1})$ has rank n.*

Proof. Since Σ is non-singular, and given the assumptions on $A(z)$ and $B(z)$, the conditions (8a), (8b) and (8c) of [9] are satisfied for $A^{-1}(z)B(z)$, and the lemma is a special case of [9, Theorem 2] in the case $(\mathrm{iii})_2$. □

Since identifiability is equivalent to the inversibility of the Fisher information matrix, we are interested in finding a sufficient condition for identifiability. Equivalently, we are looking for a simple necessary condition for the lack of identifiability or the singularity of the Fisher information matrix.

If the lack of identifiability occurs because (i) in Lemma 4.1 is not satisfied, that means that there exists a non-unimodular polynomial matrix $F(z)$, i.e. with $\det(F(z))$ different from a constant and matrices $A'(z)$, $B'(z)$ and $C'(z)$ such that $(A(z) \; B(z) \; C(z)) = F(z)(A'(z) \; B'(z) \; C'(z))$. We have

$$(\det(A(z)) \; \det(B(z)) \; \det(C(z))) = \det(F(z))(\det(A'(z)) \; \det(B'(z)) \; \det(C'(z))),$$

where $\det(F(z))$ is a polynomial different from a constant. Hence, the equations $\det(A(z)) = 0$, $\det(B(z)) = 0$ and $\det(C(z)) = 0$ have at least one common root. Consequently, the matrix polynomials $A(z)$, $B(z)$ and $C(z)$ have at least one common eigenvalue. The same is also true for the reciprocal matrix polynomials $A^*(z)$, $B^*(z)$ and $C^*(z)$. Therefore the two tensor Sylvester matrices $S^{\otimes}_{pq}(-B, A)$ and $S^{\otimes}_{pr}(-C, A)$ introduced in Section 2 do not have full rank.

Now, if the lack of identifiability occurs because (ii) in Lamma 4.1 is not satisfied, that means that the matrix $(A_p \; B_q \; C_{r-1})$ has rank strictly smaller than n. Consequently, row n, say, of that matrix is a linear combination of the other $n-1$ rows and the matrices A_p, B_q, and C_{r-1} do not have full rank. Hence the determinants $\det(A(z))$, $\det(B(z))$ and $\det(C(z))$ are polynomials of degree strictly smaller than, respectively, np, nq, and $n(r-1)$. Hence the reciprocal matrix polynomials $A^*(z)$, $B^*(z)$ and $C^*(z)$ have at least one zero eigenvalue, hence at least one common eigenvalue. Therefore the two tensor Sylvester matrices $S^{\otimes}_{pq}(-B, A)$ and $S^{\otimes}_{pr}(-C, A)$ introduced in Section 2 do not have full rank. Note that [12] has assumed that the determinants of $\det(A(z))$ and $\det(B(z))$ have their maximum degrees.

We can conclude the following theorem.

Theorem 4.2. *Under the assumptions of Lemma 4.1 a necessary condition of invertibility of the Fisher information matrix, $\mathcal{F}(\vartheta)$, associated to the VARMAX model with the matrix polynomials $A(z), C(z)$ and $B(z)$ of degree $p, r-1, q$ respectively, is that the reciprocal matrix polynomials $A^*(z), B^*(z)$ and $C^*(z)$ have no common eigenvalue, and thus the two tensor Sylvester matrices $S^{\otimes}_{pq}(-B, A)$ and $S^{\otimes}_{pr}(-C, A)$ introduced in Section 2 have full rank.*

For a discussion of identifiability without coprimeness, see [18].

5 Numerical experiments

To save space, we will use examples based on the simplest case, i.e. $n = 2$ and $p = r = q = 1$. For all our examples we will have

$$R_x = \begin{pmatrix} 2.0 & 0.0 \\ 0.0 & 3.0 \end{pmatrix}, \quad \Sigma = I_2.$$

Example 5.1. Let A, B and C be defined by

$$A = \begin{pmatrix} -0.8 & 0.0 \\ -0.5 & -a \end{pmatrix}, \quad B = \begin{pmatrix} -b & 0.0 \\ -0.5 & -0.6 \end{pmatrix}, \quad C = \begin{pmatrix} -a & 0.0 \\ -0.5 & -0.7 \end{pmatrix},$$

where a and b are constants. The eigenvalues of $A(z)$, $B(z)$ and $C(z)$ are, respectively the pairs $(0.8, a)$, $(0.6, b)$ and $(0.7, a)$ so that, whatever a and b, there is a common eigenvalue for $A(z)$ and $C(z)$. Clearly the model is identifiable except if $a = b = 0.8$ because then the factor $1-0.8z$ can be simplified on the first row of the system equation. The tensor Sylvester matrices are respectively

$$S^{\otimes}_{pq}(-B, A) = \begin{pmatrix} -1 & 0 & 0 & 0 & b & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & b & 0 & 0 \\ 0 & 0 & -1 & 0 & 0.5 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0.5 & 0 & 0.6 \\ 1 & 0 & 0 & 0 & -0.8 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -0.5 & -a & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -0.8 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -0.5 & -a \end{pmatrix},$$

$$S^{\otimes}_{pr}(-C, A) = \begin{pmatrix} -1 & 0 & 0 & 0 & a & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & a & 0 & 0 \\ 0 & 0 & -1 & 0 & 0.5 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0.5 & 0 & 0.7 \\ 1 & 0 & 0 & 0 & -0.8 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -0.5 & -a & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -0.8 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -0.5 & -a \end{pmatrix}.$$

The first one has rank 8 except if $b = 0.8$ whereas the second one has always a determinant equal to zero, as expected given the eigenvalues.

Take $b = 0.8$ to simplify the discussion. Then, using Mathematica, it can be seen that the Fisher information matrix has a determinant proportional to $(4 - 5a)$ with a strictly positive factor and indeed it is 0 if and only if $a = 0.8$. If $a = 0.8$, the 2nd, 6th and 10th rows of the Fisher information matrix contain the fractions

$$\text{Row2} = \left(\frac{1125}{416}, \frac{75}{16}, 0, 0, -\frac{375}{208}, -\frac{25}{8}, 0, 0, -\frac{375}{416}, -\frac{25}{16}, 0, 0\right),$$

$$\text{Row6} = \left(-\frac{375}{208}, -\frac{25}{8}, 0, 0, \frac{375}{208}, \frac{25}{8}, 0, 0, 0, 0, 0, 0\right),$$

$$\text{Row10} = \left(-\frac{375}{416}, -\frac{25}{16}, 0, 0, 0, 0, 0, 0, \frac{375}{416}, \frac{25}{16}, 0, 0\right),$$

and it is easy to check that $-\text{ow2} = -\text{Row6} - \text{Row10}$.

Example 5.2. Let A, B and C be defined by

$$A = \begin{pmatrix} 0.6 & 0.2 \\ 0.0 & 0.0 \end{pmatrix}, \quad B = \begin{pmatrix} 0.5 & 0.76 \\ 0.0 & 0.0 \end{pmatrix}, \quad C = \begin{pmatrix} 0.8 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}.$$

This example is a generalization of a VARMA model considered by [17] based on an example in [1]. The determinants of $A(z)$, $B(z)$ and $C(z)$ have degree 1 so that we need to consider the reciprocal matrix polynomials $A^*(z)$, $B^*(z)$ and $C^*(z)$ which have respective roots $(-0.6, 0)$, $(-0.5, 0)$, and $(-0.8, 0)$. According to the common eigenvalue 0, the two 8×8 tensor Sylvester matrices $S^{\otimes}(-B, A)$ and $S^{\otimes}(-C, A)$ have rank $7 < 8$. Hence, we can consider singularity of the Fisher information matrix. This is confirmed by computation with Mathematica which shows that the matrix has rank 10. Moreover, rows 3, 4, 7, 8, 11 and 12 of the Fisher information matrix contain the fractions

$$\text{Row3} = \left(\frac{4}{105}, \frac{38}{2625}, \frac{16}{3}, \frac{152}{75}, 0, 0, -4, -\frac{38}{25}, 0, 0, -\frac{4}{3}, -\frac{38}{75}\right)$$

$$\text{Row4} = \left(\frac{152}{2625}, \frac{1444}{6}5625, \frac{152}{75}, \frac{13276}{1875}, 0, 0, -\frac{38}{25}, -\frac{3319}{625}, 0, 0, -\frac{38}{75}, -\frac{3319}{1875}\right)$$

$$\text{Row7} = \left(-\frac{4}{7}, -\frac{38}{175}, -4, -\frac{38}{25}, 0, 0, 4, \frac{38}{25}, 0, 0, 0, 0\right)$$

$$\text{Row8} = \left(-\frac{152}{175}, -\frac{1444}{4375}, -\frac{38}{25}, -\frac{3319}{625}, 0, 0, \frac{38}{25}, \frac{3319}{625}, 0, 0, 0, 0\right)$$

$$\text{Row11} = \left(\frac{8}{15}, \frac{76}{375}, -\frac{4}{3}, -\frac{38}{75}, 0, 0, 0, 0, 0, 0, \frac{4}{3}, \frac{38}{75}\right)$$

$$\text{Row12} = \left(\frac{304}{375}, \frac{2888}{9375}, -\frac{38}{75}, -\frac{3319}{1875}, 0, 0, 0, 0, 0, 0, \frac{38}{75}, \frac{3319}{1875}\right)$$

and it appears that $\text{Row3} = -\text{Row11} - \text{Row7}$ and $\text{Row4} = -\text{Row8} - \text{Row12}$.

Example 5.3. Let A, B and C be defined by

$$A = \begin{pmatrix} 0.6 & 0.2 \\ 0.4 & \text{-}0.6 \end{pmatrix}, \quad B = \begin{pmatrix} 0.5 & 0.76 \\ 0.25 & -0.5 \end{pmatrix}, \quad C = \begin{pmatrix} 0.7 & 0.1 \\ -0.5 & -0.7 \end{pmatrix}. \tag{9}$$

This example is a generalization of Example 1 of the bivariate VARMA(1,1) model in [12]. There, the two matrix polynomials had exactly the same two eigenvalues. As a consequence, the 8×8 tensor Sylvester matrix $S^{\otimes}(-B, A)$ is singular. In [12], it was concluded, wrongly, that the Fisher information matrix should be singular although the numerical computations in Matlab lead to a smallest eigenvalue equal to 0.0067. This incoherency was explained by numerical inaccuracy. In [17] Mélard gave a second view to that example and discovered that the necessary and sufficiency of invertibility of the Fisher information matrix is only a necessary condition. For the VARMAX model related to (9), the third matrix polynomial is such that the three matrix polynomials of degree 1 have the same eigenvalues $\pm 5/\sqrt{11}$. Hence, the two 8×8 tensor Sylvester matrices $S^{\otimes}(-B, A)$ and $S^{\otimes}(-C, A)$ are singular and more precisely have rank $6 < 8$. Exact computations with Mathematica indicates, however, that the determinant of the 12×12 Fisher information matrix is strictly positive and that the smallest eigenvalue is 0.0919. This confirms that a sufficient condition for invertibility of that matrix is that the two tensor Sylvester matrices have maximum rank but it is not a necessary condition.

Acknowledgments

We thank Peter Spreij for his comments on the errors in [12].

References

[1] G. Athanasopoulos and F. Vahid, VARMA versus VAR for macroeconomic forecasting, Journal of Business and Economic Statistics 26:237–252, 2008.

[2] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer Verlag, Berlin, New York, NY, USA, 1991.

[3] P. J. Brockwell, A. Lindner, and B. Vollenbröker, Strictly stationary solutions of multivariate ARMA equations with i.i.d. noise Ann Inst Stat Math 64:1089–1119, 2012.

[4] P. Caines, *Linear Stochastic Systems*; John Wiley and Sons: New York, NY, USA 1988.

[5] H. Dym, *Linear Algebra in Action*; Amer. Math. Soc, Volume 78, Providence, Rhode Island, USA 2006.

[6] I. Gohberg and L. Lerer, *Resultants of matrix polynomials*; Bull.Amer.Math.Soc. **82** 565–567,1976.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*; John Hopkins University Press, Baltimore, 1996.

[8] E. J. Hannan, The identification of vector mixed autoregressive-moving average systems, *Biometrika* **56**(1) (1969), 223–225.

[9] E. J. Hannan, The identification problem for multiple equation systems with moving average errors, *Econometrica* **39**(5) (1971), 223–225.

[10] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*; John Wiley and Sons: New York, NY, USA 1988.

[11] A. Klein and P. Spreij, *On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices*, Linear Algebra Appl., 1996. **237/238**, 579– 590.

[12] A. Klein, G. Mélard and P. Spreij *On the Resultant Property of the Fisher Information Matrices of a Vector ARMA process*, Linear Algebra Appl, **403**, 2005, 291–313.

[13] A. Klein and P. Spreij, *Tensor Sylvester matrices and the Fisher information matrix of VARMAX processes*, Linear Algebra Appl, 2010, **432**, 1975–1989.

[14] A. Klein, G. Mélard, *An algorithm for the exact Fisher information matrix of vector ARMAX time series*, Linear algebra Appl 2014, **446**, 1–24.

[15] A. Klein , *Matrix Algebraic Properties of the Fisher Information Matrix of Stationary Processes*, Entropy 2014, **16**, 2023–2055.

[16] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Applications*, 2nd ed.; Academic Press: Orlando, FL, USA, 1985.

[17] G. Mélard, An indirect proof for the asymptotic properties of VARMA model estimators, 2020, submitted.

[18] L. L. Wegge, Armax(p,r,q) Parameter Identifiablity without Coprimeness, Department of Economics Working Paper 1217, University of California, Davis, 2012.