UvA ECONOMETRICS

Discussion Paper: 2011/05

An algorithm for the exact Fisher information matrix of vector ARMAX time series processes

André Klein and Guy Mélard

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

An algorithm for the exact Fisher information matrix of vector ARMAX time series processes

André Klein[a], Guy Mélard[b]
[a]Department of Quantitative Economics,
University of Amsterdam, Valckenierstraat 65-67
1018 XE Amsterdam, The Netherlands.
[b]ECARES CP114/4,
Universite libre de Bruxelles,
Avenue Franklin Roosevelt 50,
B-1050 Bruxelles, Belgium.

16/11/2011

Abstract

In this paper an algorithm is developed for the exact Fisher information matrix of a vector ARMAX Gaussian process, VARMAX. The algorithm developed in this paper is composed by recursion equations at a vector-matrix level and some of these recursions consist of derivatives. For that purpose appropriate differential rules are applied. The derivatives are derived from a state space model for a vector process. The chosen representation is such that the recursions extracted from the proposed state space model are given in terms of expectations of derivatives of innovations and not the process and observation disturbances. This enables us to produce an implementable algorithm for the VARMAX process. The algorithm will be illustrated by an example.

AMS 1991 *subject classifications.* Primary 62B10, 62M10, 62M12*;* secondary 15A57.

Keywords and phrases. Fisher information matrix, matrix differentiation, vector ARMAX process, E4 Toolbox.

Introduction

This paper is devoted to the computation of the exact Fisher information matrix of a m-dimensional time series $\{y_1, \ldots, y_N\}$ of length N, generated by a vector ARMAX, or VARMAX, Gaussian process of order (p, e, s), $\{y_t, t \in \mathbb{Z}\}$, $\mathbb{Z}$ the set of integers. More precisely, consider the vector difference equation representation of a dynamic linear system

$$\sum_{j=0}^{p} \alpha_j \; y_{t-j} = \gamma_0 + \sum_{j=1}^{e} \gamma_j \; u_{t-j} + \sum_{j=0}^{s} \beta_j \; w_{t-j}, \; t \in \mathbb{Z} \tag{1}$$

where y_t, u_t and w_t are, respectively, the observed output, the r-dimensional observed input, and the unobserved errors, and where $\alpha_j \in \mathbb{R}^{m\times m}$, $\gamma_j \in \mathbb{R}^{m\times r}$, $\gamma_0 \in \mathbb{R}^{m\times 1}$, and $\beta_j \in \mathbb{R}^{m\times m}$ are the associate parameter matrices. We additionally assume $\alpha_0 \equiv \beta_0 \equiv I_m$. Starting the summation in the first sum of the right-hand side with 1 rather than with 0 turns out to be more convenient. Furthermore, there is no loss in generality in the sense that u_t can always be redefined as u_{t-1}. The error $\{w_t, t \in \mathbb{Z}\}$ is a collection of independent zero mean m-dimensional random variables with a positive definite covariance matrix Σ and we assume, for all t, t', $\mathbb{E}\{\ u_t\ w_{t'}^\top\} = 0$, where $^\top$ denotes transposition. We assume either that u_t is non stochastic or that u_t is stochastic but that statistical inference is performed conditionally on the values taken by u_t. Note that observations for u_t should be available for $t \geq 1 - e$. If $e = 0$, we assume $u_0 = 0$.

We use z to denote the backward shift operator, for example $z\ u_t = u_{t-1}$, then (1) can be written as

$$\alpha(z)\ y_t = \gamma_0 + \gamma(z)\ u_t + \beta(z)\ w_t \tag{2}$$

where

$$\alpha(z) = \sum_{j=0}^{p} \alpha_j\ z^j;\ \gamma(z) = \sum_{j=1}^{e} \gamma_j\ z^j;\ \beta(z) = \sum_{j=0}^{s} \beta_j\ z^j$$

are the associated polynomial matrices . The assumption $\det(\alpha(z)) \neq 0$, and $\det(\beta(z)) \neq 0$ will be imposed so that the eigenvalues of the matrix polynomials $\alpha(z)$ and $\beta(z)$ will be outside the unit circle. The elements of $\alpha^{-1}(z)$ and $\beta^{-1}(z)$ can then be written as power series in z. An eigenvalue of a matrix polynomial $A(z)$ is a root of the scalar equation $\det A(z) = 0$, where $\det A(z)$ is the determinant of $A(z)$. Consequently, the characteristic polynomial of the matrix polynomial $A(z)$ is the polynomial $\det A(z) = 0$. In [24], the scalar version of (2) is considered. The authors have proved that the corresponding 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 [20], 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. Although the approach considered in this paper is the finite sample case and the property proved in [20] has not been studied yet for the VARMAX case, we assume that the matrix polynomials $\alpha(z)$ and $\beta(z)$ in (2) have no common eigenvalues. This is to ensure stability of the system considered.

Estimation of the matrices $\alpha_1, \alpha_2, ..., \alpha_p,\ \beta_1, \beta_2, ..., \beta_s,\ \gamma_0, \gamma_1, \gamma_2, ..., \gamma_e$, and Σ has received considerable attention in the time series and filtering theory literature [9], [10]. In [11], the authors study the asymptotic properties of maximum likelihood estimates of the coefficients of VARMAX processes, stored in a $(\ell \times 1)$ vector θ. In this paper

$$\theta = \text{vec}(\alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_s, \gamma_0, \gamma_1, \gamma_2, ..., \gamma_e). \tag{3}$$

Under mild assumptions but assuming that the estimators are asymptotically unbiased, the inverse of the asymptotic information matrix yields the Cramér-Rao bound, and provided that the estimators are asymptotically efficient, the asymptotic covariance matrix. Therefore, tests on coefficients can be derived but also the reverse problem can be solved: how long should be the series in order to obtain a given statistical significance, see [7]. In [30], an algorithm for the asymptotic Fisher information matrix of a VARMA process is developed at the scalar-level. It is based on a frequency domain representation of the Fisher information matrix, known as Whittle's formula, see [40]. See also [12], [8]. In [40], a scalar-level formula is developed for the asymptotic Fisher information matrix of a VARMA process. Whereas in [14], the equivalence between a time and frequency domain representation of the asymptotic Fisher information matrix of a VARMA process is displayed. This is done at the vector-matrix level.

The procedures used to evaluate the asymptotic information matrix rely on evaluating integrals of a rational function over the unit circle. These integrals can be computed by recurrences with respect to the degrees of the polynomials (e.g. [34], [33]). However, the most efficient method consists in transforming the problem to the evaluation of the autocovariances of an ARMA model. Klein and Mélard have used this approach in [15] by means of the algorithms developed in [38] or [6], the last one being slightly faster. It is true that until recently the asymptotic covariance matrix of the general Gaussian VARMA model has been stated only in terms of formulas involving integration over the frequency domain, e.g. [30]. In [16], Klein and Mélard have been mainly concerned with the single input single output (SISO) model but have also shown that their method can be used for the VARMA model. For recent references about the asymptotic information matrix, see [35].

More recently, the sample or exact information matrix has been studied. It is defined as minus the mathematical expectation of the Hessian of the exact likelihood function, evaluated at the final estimated value of the parameters. In [32] Porat and Friedlander have described an algorithm for a univariate ARMA model with a deterministic additive component. The method is both complex and computationally intensive. The number of scalar operations is indeed of the order of N^2, where N is the length of the sample. Independently, in [41], [42] (based on [28]) and [36] the respective authors have given a much more efficient algorithm, since the number of operations is proportional to N. The method is based on the Kalman filter, and has been applied to the VARMA model by [41] and [42], and to the general state space form by [36].

Although the algorithms in [41], [42] and [43], and Teircero in [36] need a number of operations which is proportional to N, these are not very efficient because the number of operations at each time is roughly proportional to the square of the size of the model. That number is generally smaller than N but not much so that the improvement with respect to the Porat and Friedlander method [32] can only be apparent. The reason for keeping the computational burden as low as possible is that the information matrix can be evaluated inside some optimization procedures. Mélard and Klein in [27] have given a method for computing the exact information matrix of a univariate ARMA model. That method is based on the alternative expression of the Gaussian exact likelihood in terms of the Chandrasekhar equations outlined in [29] instead of the Kalman filter equations. It is surprising that Terceiro in [36] has described the whole estimation procedure using the more computationally efficient Chandrasekhar equations instead of the better known Kalman filter recursions but that he has not mentioned at all that the Chandrasekhar equations can also be used for deriving the information matrix. This was done in [22] with an application to VARMA models but working with the prediction error of the state vector made it difficult to handle correctly initial conditions (see also [18]) and to generalize the approach to e.g. VARMAX models.

Meanwhile a software [37] called E4 has been developed on the basis of [36] but also of more recent contributions (see [13]). Under the form of a Matlab toolbox it offers various methods of estimation, signal extraction and decomposition for models represented in state space form. E4 can handle seasonal polynomials and does accept missing data. There is no problem to apply it to ARMA, ARMAX, SISO, VARMA, or VARMAX models. However, there is no detailed exposition of the computation of the exact Fisher information matrix beyond [36] and no detailed documentation of the various options related to the initial state vector (maximum likelihood, exogenous first value, exogenous mean value, zero) and the initial covariance matrix of the state vector (zero, Lyapunov or De Jong), except the last one which refers to [4].

We consider the exact Fisher information matrix $J(\theta)$ of dimension of VARMAX processes, as a generalization with some improvements of the method proposed in [27], [21], and [22]. Its main contributions are the use of recursions at a vector-matrix level, derivation of exact and explicit initial conditions, and computational performance. Indeed, instead of writing recursions for each element of the information matrix, we write recursions as consise as possible at the vector-matrix level. For that purpose, the differential rules used in [22] are applied. This implies the evaluation of the derivatives of w_t (given in (1)) with respect to the parameter vector θ whose form is defined in (3). Contrarily to [22], the approach is based on derivatives of the estimated state vector, not on the error of estimation of the state vector. A substantial complexity reduction is obtained. Computational performance also follows, partly because Chandrasekhar equations are used. A practical comparison with E4 is performed. The results are very close, although not identical, depending on the model and the E4 options used. This is a confirmation of the high quality of the relatively little known package. We suppose that relations similar to ours were used but they are not documented.

Contrary to the statements of Zadrozny and Mittnik in [43, p. 107], the inverse of the sample information matrix cannot be used to establish a Cramér-Rao bound because the estimators are biased and the bias has a very complex structure [32, p. 128]. It is also untrue that the covariance matrix of the parameter estimates can be obtained by inverting the exact information matrix, because the estimators of such dynamic models are efficient only asymptotically. However, the study of the exact information matrix is justified by the fact that it differs from the asymptotic information matrix. In [32, p. 123], the authors argued that only when $N \approx 500$ does the asymptotic information matrix come near the exact information matrix. Their numerical examples seem to indicate that the exact information leads to larger standard errors than the asymptotic information. Simulation studies of the univariate ARMA model in [2] also conclude that standard errors are slightly underestimated for small and moderate samples. That doesn't mean, however, that the exact information matrix is better than its asymptotic version for deriving standard errors of parameter estimates. There is no result about that, as far as we know.

The article is organized as follows. In Section II, we present the model. In Appendix A, we merely substitute the Kalman filter equations by the Chandrasekhar equations within Terceiro Lomba's approach. In section III, we give a closed form expression for the recursions needed to evaluate the information matrix as a whole. In section IV, we examine the special case of the VARMAX model for which the two approaches are compared in terms of the number of operations. In section V, we present some Monte Carlo results.

1 The Model

1.1 The state space model

A state space model is introduced which describes a vector linear times series process under its most general form. The following state space structure is considered.

$$x_{t+1} = \phi x_t + \Gamma u_t + F w_t \tag{4}$$
$$y_t = \gamma_0 + H x_t + w_t, \tag{5}$$

where $y_t \in \mathbb{R}^m$ is the vector of observations, $x_t \in \mathbb{R}^n$ is the vector of the state variables, $u_t \in \mathbb{R}^r$ is the vector of exogenous variables, $t \in \mathbb{N}$, $w_t \in \mathbb{R}^m$ is a Gaussian white noise process with $\mathbb{E}(w_t) = 0$, $\mathbb{E}(w_t w_t^\top) = \Sigma \geq 0$, and ϕ, Γ, F, γ_0, H are matrices of dimensions, respectively, $n \times n$, $n \times r$, $n \times m$, $m \times 1$, $m \times n$, and $\mathbb{E}$ denotes the mathematical expectation.

1.2 The Chandrasekhar equations

We now introduce the exact likelihood function of a time series $\{y_1, \ldots, y_N\}$ of length N. There are several ways to express it. Except for the closed form expression of a normal multivariate density, the simplest representation is based on the Chandrasekhar recursion equations, these equations were introduced in [29] and are also the most computationally efficient, even with respect to the Kalman filter. The Kalman filter consists of a collection of recursions, one of them giving the covariance matrix of the prediction error $P_{t+1|t}$, see e.g. [1].

The Chandrasekhar equations make use of smaller matrices given by

$$B_t = B_{t-1} + HY_{t-1}X_{t-1}Y_{t-1}^\top H^\top \tag{6}$$

$$K_t = [K_{t-1}B_{t-1} + \phi Y_{t-1}X_{t-1}Y_{t-1}^\top H^\top]B_t^{-1} \tag{7}$$

$$Y_t = [\phi - K_{t-1}H]Y_{t-1} \tag{8}$$

$$X_t = X_{t-1} - X_{t-1}Y_{t-1}^\top H^\top B_t^{-1}HY_{t-1}X_{t-1} \tag{9}$$

$$\widehat{y}_{t|t-1} = \gamma_0 + H\widehat{x}_{t|t-1} \tag{10}$$

$$\widehat{x}_{t+1|t} = \phi\widehat{x}_{t|t-1} + \Gamma u_t + K_t\widetilde{y}_t \tag{11}$$

where $\widehat{x}_{t|t-1}$ is the one-step-ahead prediction of the state vector,

$$\begin{aligned}\widetilde{y}_t &= y_t - \widehat{y}_{t|t-1} \\ &= y_t - \gamma_0 - H\widehat{x}_{t|t-1},\end{aligned} \tag{12}$$

$B_t = \mathbb{E}[\widetilde{y}_t\widetilde{y}_t^\top]$, and the matrices X_t, Y_t and K_t have dimension $m \times m$, $n \times m$ and $n \times m$, respectively. The initial conditions are: $B_1 = HP_{1|0}H^\top + \Sigma$, $Y_1 = \phi P_{1|0}H^\top + F\Sigma$, $K_1 = Y_1B_1^{-1}$, $X_1 = B_1^{-1}$. Note that the Riccati equation is stated by

$$P_{t+1|t} = \bar{\phi}_t P_{t|t-1}\bar{\phi}_t^\top + FQF^\top - K_tB_tK_t^\top, \tag{13}$$

where we denote $\bar{\phi}_t = (\phi - K_tH)$.

Given a time series of length N, minus the logarithm of the likelihood of the system described by (1) and (2) is

$$l(\theta) = -\log L(\theta) = \sum_{t=1}^{N}\left\{\frac{m}{2}\log(2\pi) + \frac{1}{2}\log|B_t| + \frac{1}{2}\widetilde{y}_t^\top B_t^{-1}\widetilde{y}_t\right\}, \tag{14}$$

where we denote the parameters by the $(\ell \times 1)$ vector θ.

2 The exact Fisher Information Matrix at the vector-matrix level

The exact information matrix J is given by the following $(\ell \times \ell)$ matrix

$$J(\theta) = \mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta^\top}\right), \tag{15}$$

where the parameter vector θ as defined in (3) is used, and $\ell = m^2(p+s) + m(re+1)$. It is shown in [27] and [23] (in the latter a formal proof is given, at the matrix level) that the following holds true

$$J(\theta) = \sum_{t=1}^{N}\left[\frac{1}{2}\left(\frac{\partial \mathrm{vec}B_t}{\partial\theta^\top}\right)^\top (B_t\otimes B_t)^{-1}\left(\frac{\partial \mathrm{vec}B_t}{\partial\theta^\top}\right) + \mathbb{E}\left\{\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\right)^\top B_t^{-1}\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\right)\right\}\right], \tag{16}$$

where vec$X$ =col(col$(X_{ij})_{i=1}^n)_{j=1}^n$, see e.g. [26]. An alternative proof for (16) is given in [32] at the scalar-level.

In this section we set forth the differential rules needed for computing the exact Fisher information matrix. The technique for evaluating the necessary derivatives of the recursion equations is equivalent with [22] where the authors set forth state recursions with forcing terms involving derivatives of an unobserved process and observation disturbances. They lead to recursions whose forcing terms involve second moments of these derivatives. At the general state space level it is not clear how the obtained recursions should be evaluated. In [22] the recursions have been successfully derived at the vector-matrix level for the VARMA model where the process disturbance, the observation disturbance, and the innovation are identical. Consequently, the recursions are implementable for VARMA models but not beyond. In this paper the approach of [42] and [36] is used. It consists of expressing recursions in terms of expectations of derivatives of the $\widehat{x}_{t|t-1}$, not the prediction error of the state vector

$$\widetilde{x}_t = x_t - \widehat{x}_{t|t-1}. \tag{17}$$

This leads to an explicit or implementable algorithm at the general state space level so the VARMA or VARMAX version can be obtained by substituting the appropriate parameters of the corresponding state space form. We shall only illustrate the main recursion for the general case but in Section 3 a complete set of recursions is provided for the vector ARMAX process. The derived algorithm is then implementable. The suggested differential rules are now displayed.

Consider a real, differentiable $(m \times n)$ matrix function $X(\theta)$ of real $(\ell \times 1)$ vector $\theta = (\theta_1, \cdots, \theta_\ell)^\top$, where m, n and ℓ are positive integers. Let $(m \times n)$ matrices $\partial_r X = (\partial X_{ij}/\partial\theta_r)$ with $r = 1, \cdots, \ell$ be the first order derivatives of $X(\theta)$ in partial derivative form with X_{ij} being the first element (i, j) of X. Write $dX_{ij} = \sum_{r=1}^{\ell}(\partial X_{ij}/\partial\theta_r)d\theta_r$, where $d\theta_r$ is an arbitrary perturbation of θ_r. The $(m \times n)$ matrix $dX = (dX_{ij})$ is the differential form of the first order derivative $X(\theta)$. An expression in differential form can instantaneously be put into a partial derivative form by replacing d with ∂_r for $r = 1, \ldots, \ell$. Let $X(\theta)$ and $Y(\theta)$ be real $(m \times n)$ and $(n \times p)$ differentiable matrix functions of the real vector $\theta(\ell \times 1)$, where m, n, p, and ℓ are positive integers. The usual scalar product rule of differentiation yields

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

Taking into account the property $\left(\partial y_t/\partial\theta^\top\right) = 0$ and $\left(\partial u_t/\partial\theta^\top\right) = 0$ (justified because the realisation does not depend on the parameters $\alpha_1, \alpha_2, ..., \alpha_p,\ \gamma_0, \gamma_1, \gamma_2, ..., \gamma_e, \beta_1, \beta_2, ..., \beta_s$) and $d\Sigma = 0$, (11) yields

$$d\hat{x}_{t+1|t} = d\phi\ \hat{x}_{t|t-1} + \phi\ d\hat{x}_{t|t-1} + d\Gamma\ u_t + dK_t\ \widetilde{y}_t + K_t\ d\widetilde{y}_t.$$

Let us vectorize the matrix $X(\theta)$ defined above according to the following rule

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

then the $(m\,n \times \ell)$ matrix $\partial\mathrm{vec}X(\theta)\,/\,\partial\theta^\top$ is the gradient form of first order derivatives of $X(\theta)$ and can be defined as $\mathrm{vec}d\,X(\theta) = (\partial(\mathrm{vec}X(\theta))\,/\,\partial\theta^\top)d\,\theta = d\mathrm{vec}X(\theta)$. Componentwise application of this rule to (11) gives

$$\begin{aligned} d\widehat{x}_{t+1|t} &= \left(\widehat{x}_{t|t-1}^\top \otimes I_n\right)\mathrm{vec}d\phi + \phi d\widehat{x}_{t|t-1} + \left(u_t^\top \otimes I_n\right)\mathrm{vec}d\Gamma + \left(\widetilde{y}_t^\top \otimes I_n\right)\mathrm{vec}dK_t + K_t d\widetilde{y}_t \\ &= \left(\widehat{x}_{t|t-1}^\top \otimes I_n\right)\frac{\partial\mathrm{vec}\phi}{\partial\theta^\top}d\theta + \phi\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}d\theta + \left(u_t^\top \otimes I_n\right)\frac{\partial\mathrm{vec}\Gamma}{\partial\theta^\top}d\theta \\ &\quad + \left(\widetilde{y}_t^\top \otimes I_n\right)\frac{\partial\mathrm{vec}K_t}{\partial\theta^\top}d\theta + K_t\frac{\partial\widetilde{y}_t}{\partial\theta^\top}d\theta. \end{aligned}$$

We can now formulate the appropriate derivative of $\hat{x}_{t+1|t}$ with respect to the vector θ

$$\frac{\partial \widehat{x}_{t+1|t}}{\partial \theta^\top} = \left(\widehat{x}_{t|t-1}^\top \otimes I_n\right) \frac{\partial \text{vec}\phi}{\partial \theta^\top} + \phi \frac{\partial \widehat{x}_{t|t-1}}{\partial \theta^\top} + \left(u_t^\top \otimes I_n\right) \frac{\partial \text{vec}\Gamma}{\partial \theta^\top} + \left(\widetilde{y}_t^\top \otimes I_n\right) \frac{\partial \text{vec}K_t}{\partial \theta^\top} + K_t \frac{\partial \widetilde{y}_t}{\partial \theta^\top}. \tag{18}$$

Similarly for the derivative of $\widetilde{y}_t$, to obtain from (12)

$$\frac{\partial \widetilde{y}_t}{\partial \theta^\top} = -\left\{ \frac{\partial \gamma_0}{\partial \theta^\top} + (\widehat{x}_{t\ |t-1}^\top \otimes I_m) \frac{\partial \text{vec}H}{\partial \theta^\top} + H \ \frac{\partial \widehat{x}_{t|t-1}}{\partial \theta^\top} \right\}. \tag{19}$$

We shall derive explicit recursions for $\mathbb{E}\left\{(\partial \widetilde{y}_t/\partial \theta^\top) \otimes (\partial \widetilde{y}_t/\partial \theta^\top)\right\}^\top$ accordingly by using the following rules [26].

Rule 1. $(A \otimes B)(C \otimes D) = AC \otimes BD$, where $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}, C \in \mathbb{R}^{n\times k}$, and $D \in \mathbb{R}^{q\times l}$.

Rule 2. $(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$

Rule 3. $(A \otimes B)^\top = A^\top \otimes B^\top$.

Rule 4. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if A^{-1} and B^{-1} exist.

Rule 5. Let $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$, then $M_{p,m}(A \otimes B)M_{n,q} = B \otimes A$,

where the commutation matrix $M_{m,n} \in \mathbb{R}^{mn\times mn}$ is defined by $M_{m,n} = \sum_{i=1}^{m}\sum_{j=1}^{n}(H_{i\ j} \otimes H_{i\ j}^\top)$, where $H_{i\ j}^\top = e_j^n\,(e_i^m)^\top$, and e_i^m is the unit column vector of order m. Note also the properties $M_{n,m}^\top = M_{m,n}$ and $M_{1,n} = M_{n,1} = I_n$ and taking the orthogonality into account yields $M_{n,m}M_{m,n} = I_{mn}$.

Before formulating the next rule, we consider the random vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, jointly distributed with $\mathbb{E}(x)=\mu_1$, $\mathbb{E}(y) = \mu_2$ and $\mathbb{E}\left\{(y - \mu_2)(x - \mu_1)^\top\right\} = \Omega$, leads to.

Rule 6. $\mathbb{E}(x \otimes y) = \text{vec}\Omega + \mu_1 \otimes \mu_2$

For solving the first term of (16) the derivatives of the Chandrasekhar equations are used, whereas the second term consists of the expected value of a stochastic component, we therefore vectorize $J(\theta)$ according to

$$\begin{aligned} \text{vec}J(\theta) = \sum_{t=1}^{N} \Bigg\{ &\frac{1}{2} \left[\left(\frac{\partial \text{vec}B_t}{\partial \theta^\top} \right) \otimes \left(\frac{\partial \text{vec}B_t}{\partial \theta^\top} \right) \right]^\top \text{vec}\left(B_t \otimes B_t\right)^{-1} \\ &+ \mathbb{E}\left\{ \frac{\partial \widetilde{y}_t}{\partial \theta^\top} \otimes \frac{\partial \widetilde{y}_t}{\partial \theta^\top} \right\}^\top \text{vec}B_t^{-1} \Bigg\}. \end{aligned} \tag{20}$$

The second term of (20) which is of interest, $\mathbb{E}\left\{(\partial \widetilde{y}_t/\partial \theta^\top) \otimes (\partial \widetilde{y}_t/\partial \theta^\top)\right\}^\top$, requires additional recursions that shall be constructed by the differential rules used in [22]. It will be illustrated for the derivatives of the Kalman one step ahead predictor $\hat{x}_{t+1|t}$ and the innovation $\widetilde{y}_t$ given in (11) and (12) respectively.

Equations (18) and (19) allow the right-hand side of (20) to be written in an appropriate way. This is fully done for the VARMAX case in Section 3. To successfully express $\mathbb{E}\left\{(\partial \widetilde{y}_t/\partial \theta^\top) \otimes (\partial \widetilde{y}_t/\partial \theta^\top)\right\}^\top$ for state space model (4) and (5), the following properties are taken into account,

$$\begin{aligned} &\mathbb{E}\left(\widehat{x}_{t+1|t} \otimes v_{t+1}\right) = 0,\ \mathbb{E}\left(u_{t+1} \otimes v_{t+1}\right) = 0, \\ &\mathbb{E}\left(x_{t+1} \otimes v_{t+1}\right) = 0,\ \mathbb{E}\left[\left(\partial \widehat{x}_{t|t-1}/\partial \theta^\top\right)^\top \otimes v_{t+1}\right] = 0, \end{aligned}$$

to obtain

$$\mathbb{E}\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\otimes\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\right) = \left\{\left\{M_{m,1}(\mathbb{E}(\widehat{x}_{t|t-1}^\top\otimes\widehat{x}_{t|t-1}^\top)\otimes I_m)M_{mn,n}\right\}\otimes I_m\right\}\left(\frac{\partial\mathrm{vec}H}{\partial\theta^\top}\otimes\frac{\partial\mathrm{vec}H}{\partial\theta^\top}\right)$$
$$+\left\{M_{mm}\left[\left\{H\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\otimes\widehat{x}_{t|t-1}^\top\right)\right\}\otimes I_m\right]M_{mn,\ell}\right\}\left(\frac{\partial\mathrm{vec}H}{\partial\theta^\top}\otimes I_\ell\right)$$
$$+\left[\left\{H\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\otimes\widehat{x}_{t|t-1}^\top\right)\right\}\otimes I_m\right]\left(I_\ell\otimes\frac{\partial\mathrm{vec}H}{\partial\theta^\top}\right)$$
$$+(H\otimes H)\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\otimes\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\right)$$
$$+\frac{\partial\gamma_0}{\partial\theta^\top}\otimes\left((\mathbb{E}\left(\widehat{x}_{t\;|t-1}^\top\right)\otimes I_m)\frac{\partial\mathrm{vec}H}{\partial\theta^\top}+H\;\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\right)\right)$$
$$+\left((\mathbb{E}\left(\widehat{x}_{t\;|t-1}^\top\right)\otimes I_m)\frac{\partial\mathrm{vec}H}{\partial\theta^\top}+H\;\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\right)\right)\otimes\frac{\partial\gamma_0}{\partial\theta^\top}$$
$$+\frac{\partial\gamma_0}{\partial\theta^\top}\otimes\frac{\partial\gamma_0}{\partial\theta^\top}, \tag{21}$$

where the the commutation matrix M_{ab} is defined in Rule 5. In the next section, the construction of the algorithm for $J(\theta)$ is displayed. This will be done for the state-space model (4) and (5) with an appropriate parametrization. The details are set forth in Appendix A.

3 An algorithm for the vector ARMAX model

An appropriate choice for a parametrization of (4) and (5) is given by

$$\phi=\begin{pmatrix}-\alpha_1 & I_m & & 0_m\\ -\alpha_2 & 0_m & \ddots & \\ \vdots & & \ddots & I_m\\ -\alpha_h & 0_m & \cdots & 0_m\end{pmatrix},\ F=\begin{pmatrix}\beta_1-\alpha_1\\ \beta_2-\alpha_2\\ \vdots\\ \beta_h-\alpha_h\end{pmatrix},\ \Gamma=\begin{pmatrix}\gamma_1\\ \gamma_2\\ \vdots\\ \gamma_h\end{pmatrix}, \tag{22}$$

$$\text{and } H=\begin{pmatrix} I_m & 0_m & . & . & . & 0_m\end{pmatrix} \tag{23}$$

and $h=\max(p,s,e)$, $\alpha_i=0_m, i>p$, $\beta_i=0_m, i>s$, $\gamma_i=0_{m\times r}, i>e$, and consequently $n=hm$. More precisely the i-th $m\times 1$ block, $i=1,...,h$, of the state vector x_t is composed of

$$-\sum_{j=i}^{p}\alpha_j\;y_{t-j+i-1}+\sum_{j=i}^{e}\gamma_j\;u_{t-j+i-1}+\sum_{j=i}^{s}\beta_j\;w_{t-j+i-1},\;t\in\mathbb{Z}. \tag{24}$$

Note that $\partial\mathrm{vec}H/\partial\theta^\top=0$. Hence (19) simplifies to

$$\frac{\partial\widetilde{y}_t}{\partial\theta^\top}=-\frac{\partial\gamma_0}{\partial\theta^\top}-H\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}.$$

respectively and we obtain a main recurrence equation analogous to (32) of [22]:

$$\mathbb{E}\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\otimes\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\right)=(H\otimes H)\mathbb{E}\left(\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}\right)$$
$$+\frac{\partial\gamma_0}{\partial\theta^\top}\otimes\left(H\;\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\right)\right)+\left(H\;\mathbb{E}\left(\frac{\partial\widehat{x}_{t|t-1}}{\partial\theta^\top}\right)\right)\otimes\frac{\partial\gamma_0}{\partial\theta^\top}$$
$$+\frac{\partial\gamma_0}{\partial\theta^\top}\otimes\frac{\partial\gamma_0}{\partial\theta^\top},$$

but much shorter. Of course it is necessary to update the expectation in the right hand side by using, from (18),

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial \hat{x}_{t+1|t}}{\partial\theta^\top}\otimes\frac{\partial \hat{x}_{t+1|t}}{\partial\theta^\top}\right) &= (\phi\otimes\phi)\,\mathbb{E}\left(\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\right)\\
&+\left\{\left[\left\{\mathbb{E}(\hat{x}_{t|t-1}^\top\otimes\hat{x}_{t|t-1}^\top)\otimes I_n\right\}M_{n,n^2}\right]\otimes I_n\right\}\left(\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\otimes\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\right)\\
&+\left\{M_{n,n}\left[\left\{\phi\mathbb{E}\left(\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)\right\}\otimes I_n\right]M_{\ell,n^2}\right\}\left(\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\otimes I_\ell\right)\\
&+\left((\mathbb{E}(\hat{x}_{t|t-1}^\top)\otimes I_n)\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\right)\otimes\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\\
&+\left[M_{n,n}\left\{K_t\mathbb{E}\left(\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)\otimes I_n\right\}M_{\ell,n^2}\right]\left(\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\otimes I_\ell\right)\\
&+(K_t\otimes I_n)\left\{\mathbb{E}\left(\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)\otimes I_n\right\}\left(I_\ell\otimes\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\right)\\
&+(\phi\otimes I_n)\left\{\mathbb{E}\left(\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)\otimes I_n\right\}\left(I_\ell\otimes\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\right)\\
&+\{\phi\otimes u_t^\top\otimes I_n\}\left\{\mathbb{E}\left(\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\right)\otimes\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right\}\\
&+(\phi\otimes K_t)\mathbb{E}\left(\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\right)\\
&+\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\otimes\left(\left(\mathbb{E}\left(\hat{x}_{t|t-1}^\top\right)\otimes I_n\right)\frac{\partial \mathrm{vec}\phi}{\partial\theta^\top}\right)\\
&+\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\otimes\left\{\phi\mathbb{E}\left(\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\right)\right\}\\
&+\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\otimes\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\\
&+\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\otimes\left\{K_t\mathbb{E}\left(\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\right)\right\}\\
&+\left\{\left[\left\{\mathbb{E}\left(\widetilde{y}_t^\top\otimes\widetilde{y}_t^\top\right)\otimes I_n\right\}M_{m,nm}\right]\otimes I_n\right\}\left(\frac{\partial \mathrm{vec}K_t}{\partial\theta^\top}\otimes\frac{\partial \mathrm{vec}K_t}{\partial\theta^\top}\right)\\
&+(K_t\otimes\phi)\mathbb{E}\left(\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\otimes\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\right)\\
&+\left\{K_t\mathbb{E}\left(\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\right)\right\}\otimes\left((u_t^\top\otimes I_n)\frac{\partial \mathrm{vec}\Gamma}{\partial\theta^\top}\right)\\
&+(K_t\otimes K_t)\mathbb{E}\left(\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\otimes\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\right),
\end{aligned}\tag{25}$$

because

$$\mathbb{E}\left[\widetilde{y}_t\right]=0,\ \mathbb{E}\left[\hat{x}_{t|t-1}\otimes\widetilde{y}_t\right]=0,\ \mathbb{E}\left[\frac{\partial \hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\widetilde{y}_t\right]=0,\ \mathbb{E}\left[\frac{\partial \widetilde{y}_t}{\partial\theta^\top}\otimes\widetilde{y}_t\right]=0.$$

Indeed sample innovations $\widetilde{y}_t$ are zero mean uncorrelated random variables. Also $\mathbb{E}(\hat{x}_{t|t-1}\otimes\widetilde{y}_t)=0$ because $\hat{x}_{t|t-1}$ is in the space spanned by the observations till time $t-1$ included, whereas $\widetilde{y}_t$ is orthogonal to that space. The explanation is similar for $\partial\hat{x}_{t|t-1}/\partial\theta^\top$ and

$\partial\widetilde{y}_t/\partial\theta^\top$. Note that there are several ways to write terms in (25). We have made sure to reduce computations for large n and ℓ. Note that

$$\mathbb{E}\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)=-\frac{\partial\gamma_0}{\partial\theta^\top}\mathbb{E}(\hat{x}_{t|t-1}^\top)-H\mathbb{E}\left(\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right).$$

For the implementation of the fundamental recurrence equation (25), we need four additional recursions as follows:

1.

$$\begin{aligned}\mathbb{E}\left(\frac{\partial\hat{x}_{t+1|t}}{\partial\theta^\top}\otimes\hat{x}_{t+1|t}^\top\right)=&\left\{\left[\mathbb{E}(\hat{x}_{t|t-1}^\top\otimes\hat{x}_{t|t-1}^\top)\otimes I_n\right]M_{n,n^2}\right\}\left(\frac{\partial\mathrm{vec}\phi}{\partial\theta^\top}\otimes\phi^\top\right)\\&+\left[\mathbb{E}(\hat{x}_{t|t-1}^\top)\otimes I_n\right]\left(\frac{\partial\mathrm{vec}\phi}{\partial\theta^\top}\right)\otimes(u_t^\top\Gamma^\top)\\&+\left\{\phi\mathbb{E}\left(\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)\right\}(I_\ell\otimes\phi^\top)\\&+\left(\phi\mathbb{E}\left(\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}\right)\right)\otimes(u_t^\top\Gamma^\top)\\&+\left((u_t^\top\otimes I_n)\frac{\partial\mathrm{vec}\Gamma}{\partial\theta^\top}\right)\otimes(\mathbb{E}(\hat{x}_{t|t-1}^\top)\phi^\top)\\&+\left((u_t^\top\otimes I_n)\frac{\partial\mathrm{vec}\Gamma}{\partial\theta^\top}\right)\otimes(u_t^\top\Gamma^\top)\\&+[\{(\mathrm{vec}B_t)^\top\otimes I_n\}M_{m,mn}]\left(\frac{\partial\mathrm{vec}K_t}{\partial\theta^\top}\otimes K_t^\top\right)\\&+K_t\mathbb{E}\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\otimes\hat{x}_{t|t-1}^\top\right)(I_\ell\otimes\phi^\top)\\&+\left(K_t\mathbb{E}\left(\frac{\partial\widetilde{y}_t}{\partial\theta^\top}\right)\right)\otimes(u_t^\top\Gamma^\top).\end{aligned}\tag{26}$$

2.

$$\begin{aligned}\mathbb{E}\left(\frac{\partial\hat{x}_{t+1|t}}{\partial\theta^\top}\right)=&\left[(\mathbb{E}\hat{x}_{t|t-1})^\top\otimes I_n\right]\frac{\partial\mathrm{vec}\phi}{\partial\theta^\top}\\&+(\phi-K_tH)\mathbb{E}\left(\frac{\partial\hat{x}_{t|t-1}}{\partial\theta^\top}\right)+(u_t^\top\otimes I_n)\frac{\partial\mathrm{vec}\Gamma}{\partial\theta^\top}.\end{aligned}\tag{27}$$

3.

$$\begin{aligned}\mathbb{E}\left(\hat{x}_{t+1|t}\otimes\hat{x}_{t+1|t}\right)=&(\phi\otimes\phi)\,\mathbb{E}(\hat{x}_{t|t-1}\otimes\hat{x}_{t|t-1})+(\phi\otimes\Gamma)\left[\mathbb{E}(\hat{x}_{t|t-1})\otimes u_t\right]\\&+(\Gamma\otimes\phi)\left[u_t\otimes\mathbb{E}(\hat{x}_{t|t-1})\right]+(\Gamma\otimes\Gamma)(u_t\otimes u_t)+(K_t\otimes K_t)\,\mathrm{vec}B_t.\end{aligned}\tag{28}$$

4.

$$\mathbb{E}\left(\hat{x}_{t+1|t}\right)=\phi\mathbb{E}(\hat{x}_{t|t-1})+\Gamma u_t.$$

This set of recursions is nevertheless much lighter than equations (52) to (65) in [22]. Of course the derivatives of the Chandrasekhar equations, equations (46) to (49) in [22], recalled in Appendix A are also needed.

Computationally, the recursions are written in the less demanding form. Several times Rules 1 and 5 of Section 2 have been used to put random variables next to each other to set forth expectations whereas Rule 1 has been avoided when possible because otherwise the number of operations is increased without necessity. Indeed, the left hand side of Rule 1 requires $mpnq + nqkl + mpnqkl$ multiplications, generally bigger than what is required by the right hand side $mnk + pql + mkpl$ multiplications.

To be complete we also need to state the initial values of the matrices in the Chandrasekhar equations, B_1, Y_1, K_1, X_1, as well as $P_{1|0}H$, and their derivatives. This was done in [22, Section 5, pp. 225-228] and doesn't need to be repeated here to save space. Note that the already complex initializations for $\mathbb{E}\left((\partial\widetilde{x}_1/\partial\theta^\top)\otimes(\partial\widetilde{x}_1/\partial\theta^\top)\right)$, with $\widetilde{x}_1$ defined by (17) and other expressions (most of p. 229) were wrong and replaced by a still more complex initialisation procedure described in [18]. Fortunately, things are much simpler. Besides B_1, K_1, Y_1, X_1, and $P_{1|0}H$, and their derivatives with respect to θ, the following initial values are needed. Because of (24) for $t = 1$, after projection in the initial state space we have for each subvector of dimension m

$$(\hat{x}_{1|0})_i = \sum_{j=i}^{e} \gamma_j \; u_{i-j}, \tag{29}$$

for $i = 1, ..., h$, hence

$$\mathbb{E}(\hat{x}_{1|0})_i = \sum_{j=i}^{e}\gamma_j u_{i-j}, \qquad \mathbb{E}(\hat{x}_{1|0}\otimes\hat{x}_{1|0})_{i,g} = \sum_{j=i}^{e}\sum_{k=g}^{e}(\gamma_j\otimes\gamma_k)(u_{i-j}\otimes u_{g-k}),$$

$$\mathbb{E}\left(\frac{\partial\hat{x}_{1|0}}{\partial\theta^\top}\right)_i = \sum_{j=i}^{e}(u_{i-j}^\top\otimes I_m)\frac{\partial\text{vec}\gamma_j}{\partial\theta^\top},$$

$$\mathbb{E}\left[\left(\frac{\partial\hat{x}_{1|0}}{\partial\theta^\top}\right)\otimes\hat{x}_{1|0}^\top\right]_{i,g} = \sum_{j=i}^{e}\sum_{k=g}^{e}\left(u_{i-j}^\top\otimes I_m\otimes u_{g-k}^\top\right)\left(\frac{\partial\text{vec}\gamma_j}{\partial\theta^\top}\otimes\gamma_k^\top\right),$$

$$\mathbb{E}\left[\left(\frac{\partial\hat{x}_{1|0}}{\partial\theta^\top}\right)\otimes\left(\frac{\partial\hat{x}_{1|0}}{\partial\theta^\top}\right)\right]_{i,g} = \sum_{j=i}^{e}\sum_{k=g}^{e}(u_{i-j}^\top\otimes I_m\otimes u_{g-k}^\top\otimes I_m)\left(\frac{\partial\text{vec}\gamma_j}{\partial\theta^\top}\otimes\frac{\partial\text{vec}\gamma_k}{\partial\theta^\top}\right).$$

for $i, g = 1, ..., h$. Note that i and g are block indices and that the elements of $\left((\partial\text{vec}\gamma_1/\partial\theta^\top), ..., (\partial\text{vec}\gamma_e/\partial\theta^\top)\right)^\top$ are related to $\partial\text{vec}\Gamma/\partial\theta^\top$ through a commutation matrix

$$\left(\left(\frac{\partial\text{vec}\gamma_1}{\partial\theta^\top}\right)^\top, ..., \left(\frac{\partial\text{vec}\gamma_e}{\partial\theta^\top}\right)^\top\right)^\top = M_{hmr,hmr}\frac{\partial\text{vec}\Gamma}{\partial\theta^\top}. \tag{30}$$

These initializations are slightly simpler than the erroneous initializations in [22] and considerably much simpler than the iterative procedure in its corrected version [18]. Of course, they are also more general here because they hold for VARMAX models , not simply for VARMA models.

4 A numerical example and a comparison with the E4 Toolbox

In this section some numerical results are displayed for an example. Furthermore, the results are compared with those of E4, a toolbox for Matlab ([37], [13]) which can be used to evaluate the exact information matrix of general state space models. Our implementation is available at location http:\\homepages.ulb.ac.be\~gmelard\rech\km12prog.zip. It is heavily based on [31] and [19], which were developed for VARMA models without exogeneous variables.

4.1 The example

The results obtained through the algorithm set forth in this paper shall be compared with the values of the entries of the asymptotic Fisher information matrix of a VARMAX process. First the asymptotic case is set forth on the basis of [25]. The VARMAX process considered in this example is such that $m = 2$, $r = 3$ and $p = q = s = e = 1$. The appropriate matrix polynomials are

$$\alpha(L) = \begin{pmatrix} 1+\alpha_1^{11}L & \alpha_1^{12}L \\ \alpha_1^{21}L & 1+\alpha_1^{22}L \end{pmatrix}, \qquad \beta(L) = \begin{pmatrix} 1+\beta_1^{11}L & \beta_1^{12}L \\ \beta_1^{21}L & 1+\beta_1^{22}L \end{pmatrix},$$

and

$$\gamma(L) = \begin{pmatrix} \gamma_1^{11}+\gamma_2^{11}L & \gamma_1^{12}+\gamma_2^{12}L & \gamma_1^{13}+\gamma_2^{13}L \\ \gamma_1^{21}+\gamma_2^{21}L & \gamma_1^{22}+\gamma_2^{22}L & \gamma_1^{23}+\gamma_2^{23}L \end{pmatrix}. \tag{31}$$

We further assume,

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

where Ω denotes the instantaneous covariance matrix of the white noise process used to generate u, assumed to be independent from w.The parameter vector configuration is given by $\theta = \mathrm{vec}(\alpha_1, \beta_1, \gamma_1, \gamma_2)$, where

$$\alpha_1 = \begin{pmatrix} \alpha_1^{11} & \alpha_1^{12} \\ \alpha_1^{21} & \alpha_1^{22} \end{pmatrix}, \beta_1 = \begin{pmatrix} \beta_1^{11} & \beta_1^{12} \\ \beta_1^{21} & \beta_1^{22} \end{pmatrix}, \gamma_1 = \begin{pmatrix} \gamma_1^{11} & \gamma_1^{12} & \gamma_1^{13} \\ \gamma_1^{21} & \gamma_1^{22} & \gamma_1^{23} \end{pmatrix}, \gamma_2 = \begin{pmatrix} \gamma_2^{11} & \gamma_2^{12} & \gamma_2^{13} \\ \gamma_2^{21} & \gamma_2^{22} & \gamma_2^{23} \end{pmatrix}.$$

4.2 The asymptotic Fisher information matrix

A partitioned form of the asymptotic Fisher information matrix multiplied by N, the number of observations, is considered, this in order to display the interactions between the different parameters involved,

$$\mathcal{F}(\theta) = \begin{pmatrix} \mathcal{F}_{\alpha\alpha}(\theta) & \mathcal{F}_{\alpha\beta}(\theta) & \mathcal{F}_{\alpha\gamma}(\theta) \\ \mathcal{F}_{\beta\alpha}(\theta) & \mathcal{F}_{\beta\beta}(\theta) & \mathcal{F}_{\beta\gamma}(\theta) \\ \mathcal{F}_{\gamma\alpha}(\theta) & \mathcal{F}_{\gamma\beta}(\theta) & \mathcal{F}_{\gamma\gamma}(\theta) \end{pmatrix}. \tag{33}$$

Taking into consideration that the input $u(t)$ and the white noise $w(t)$ are orthogonal processes leads to the property

$$\mathcal{F}_{\gamma\beta}(\theta) = 0. \tag{34}$$

In [25], the authors focused on the γ parameters considering its crucial role in VARMAX processes. The computations were extended here in order to compute the whole asymptotic information matrix. The partitioned form of $\mathcal{F}_{\gamma\gamma}(\theta)$ is considered, to obtain

$$\mathcal{F}_{\gamma\gamma}(\theta) = \begin{pmatrix} \mathcal{F}_{\gamma_1\gamma_1}(\theta) & \mathcal{F}_{\gamma_1\gamma_2}(\theta) \\ \mathcal{F}_{\gamma_2\gamma_1}(\theta) & \mathcal{F}_{\gamma_2\gamma_2}(\theta) \end{pmatrix}. \tag{35}$$

The parametrization of input coefficient matrix $\gamma = \left((\mathrm{vec}\gamma_1)^\top, (\mathrm{vec}\gamma_2)^\top\right)^\top$ is given by

$$\mathrm{vec}\gamma_1 = (\gamma_1^{11}, \gamma_1^{21}, \gamma_1^{12}, \gamma_1^{22}, \gamma_1^{13}, \gamma_1^{23})^\top,\ \mathrm{vec}\gamma_2 = (\gamma_2^{11}, \gamma_2^{21}, \gamma_2^{12}, \gamma_2^{22}, \gamma_2^{13}, \gamma_2^{23})^\top.$$

The entries of $\mathcal{F}_{\gamma\gamma}(\theta)$, displayed in (35), are computed according to the expressions derived in [25], to obtain

$$(\mathcal{F}_{\gamma_1\gamma_1}(\theta_\gamma))_{i,j,l,f}^{1,1} = \frac{1}{2\pi i}\oint_{|z|=1} \mathrm{Tr}\left(\beta^{-1}(z)\mathcal{E}_{ij}R_u(z)\mathcal{E}_{lf}^\top\beta^{-*}(z)\right)\frac{dz}{z},$$

$$(\mathcal{F}_{\gamma_1\gamma_2}(\theta_\gamma))^{1,2}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} \text{Tr } z^{-1}\left(\beta^{-1}(z)\mathcal{E}_{ij}R_u(z)\mathcal{E}^\top_{lf}\beta^{-*}(z)\right)\frac{dz}{z},$$

$$(\mathcal{F}_{\gamma_2\gamma_1}(\theta_\gamma))^{2,1}_{l,f,i,j} = \frac{1}{2\pi i} \oint\limits_{|z|=1} \text{Tr } z\left(\beta^{-1}(z)\mathcal{E}_{lf}R_u(z)\mathcal{E}^\top_{ij}\beta^{-*}(z)\right)\frac{dz}{z},$$

$$(\mathcal{F}_{\gamma_2\gamma_2}(\theta_\gamma))^{2,2}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} \text{Tr }\left(\beta^{-1}(z)\mathcal{E}_{ij}R_u(z)\mathcal{E}^\top_{lf}\beta^{-*}(z)\right)\frac{dz}{z},$$

where $i, l = 1, 2$ and $j, f = 1, 2, 3$. The Cauchy integral is counterclockwise and the poles, eigenvalues of the appropriate matrix polynomials, which are inside the unit circle $|z| = 1$ are used for the computations. The $m \times r$ matrix $\mathcal{E}_{ij} = e^m_i \left(e^r_j\right)^\top$ where e^m_i is the i-th standard basis vector in $\mathbb{R}^m$ and e^r_j is the j-th standard basis vector in $\mathbb{R}^r$. The Hermitian matrix $R_u(z)$ is the spectral density of the input process u_t, in this example it is driven by a white noise process with covariance Ω given in (32), consequently, $R_u(z) = (1/2\pi)I_3$, where I_3 is the 3-dimensional identity matrix. The representation $\beta^{-*}(z)$ refers to the complex conjugate transpose of $\beta^{-1}(z)$.
For this numerical illustration, like in [25] we assume $\alpha_1 = 0$, $\gamma_1 = 0$, $\gamma_2 = 0$, and specific entries of the matrix polynomial $\beta(z)$ with $\beta^{11}_1 = 6/5$, $\beta^{12}_1 = 1/2$, $\beta^{21}_1 = -(7/5)$ and $\beta^{22}_1 = -(1/5)$. The basic assumption that the eigenvalues of the matrix polynomial $\beta(z)$ lie outside the unit circle is fulfilled since the eigenvalues are: $(5/23)\left(-5 \pm i\sqrt{21}\right)$ with modulus equal to 1.47442. We first choose $(\mathcal{F}_{\gamma_2\gamma_2}(\theta))^{1,1}_{1,1,1,1}$, to obtain the following circular integral expression

$$\begin{aligned}(\mathcal{F}_{\gamma_2\gamma_2}(\theta))^{1,1}_{1,1,1,1} &= \frac{1}{2\pi i} \oint\limits_{|z|=1} \text{Tr }\left(\beta^{-1}(z)\mathcal{E}_{11}\mathcal{E}^\top_{11}\beta^{-*}(z)\right)\frac{dz}{z} \\ &= -\frac{1}{2\pi i} \oint\limits_{|z|=1} \frac{500\ z\ (1 - 15z + z^2)}{(50 + 50z + 23z^2)\,(23 + 50z + 50z^2)}\frac{dz}{z}.\end{aligned}$$

For applying Cauchy's residue theorem we have to consider the poles whithin the unit circle which are provided by the polynomial $\left(23 + 50z + 50z^2\right)$. For evaluating the integral, the algorithm developed in [6] or the computer program displayed in [34] and based on the Peterka-Vidinčev [33] algorithm can be implemented. This yields $(\mathcal{F}_{\gamma_1\gamma_1}(\theta))^{1,1}_{1,1,1,1} = (\mathcal{F}_{\gamma_2\gamma_2}(\theta))^{2,2}_{1,1,1,1} = 7.82242$. We proceed by computing an element of block $\mathcal{F}_{\gamma_1\gamma_2}(\theta)$ and $\mathcal{F}_{\gamma_2\gamma_1}(\theta)$ involving the parameters γ^{13}_2 and γ^{23}_1, to obtain

$$\begin{aligned}(\mathcal{F}_{\gamma_2\gamma_1}(\theta))^{2,1}_{1,3,2,3} &= \frac{1}{2\pi i} \oint\limits_{|z|=1} \text{Tr } z\left(\beta^{-1}(z)\mathcal{E}_{13}\mathcal{E}^\top_{23}\beta^{-*}(z)\right)\frac{dz}{z} \\ &= \frac{1}{2\pi i} \oint\limits_{|z|=1} \frac{50\ z^2\ (-25 + 89z + 70z^2)}{(50 + 50z + 23z^2)\,(23 + 50z + 50z^2)}\frac{dz}{z} \\ &= -3.3552\end{aligned} \tag{36}$$

and

$$(\mathcal{F}_{\gamma_1\gamma_2}(\theta))^{1,2}_{2,3,1,3} = \frac{1}{2\pi i}\oint\limits_{|z|=1} \text{Tr } z^{-1}\left(\beta^{-1}(z)\mathcal{E}_{23}\mathcal{E}_{13}^{\top}\beta^{-*}(z)\right)\frac{dz}{z} \tag{37}$$
$$= -\frac{1}{2\pi i}\oint\limits_{|z|=1} \frac{50\left(-70-89z+25z^2\right)}{\left(50+50z+23z^2\right)\left(23+50z+50z^2\right)}\frac{dz}{z}.$$
$$= -3.3552.$$

A numerical confirmation of $\mathcal{F}_{\gamma_1\gamma_2}(\theta) = \mathcal{F}^{\top}_{\gamma_2\gamma_1}(\theta)$ is illustrated by the obtained equality $(\mathcal{F}_{\gamma_2\gamma_1}(\theta))^{2,1}_{1,3,2,3} = (\mathcal{F}_{\gamma_1\gamma_2}(\theta))^{1,2}_{2,3,1,3}$. We proceed analogously for the remaining γ parameters, it yields the following submatrices of (35) rounded to 3 decimal places

$$\mathcal{F}_{\gamma_1\gamma_1}(\theta) = \mathcal{F}_{\gamma_2\gamma_2}(\theta) = \begin{pmatrix} 7.822 & 2.780 & 0 & 0 & 0 & 0 \\ 2.780 & 2.500 & 0 & 0 & 0 & 0 \\ 0 & 0 & 7.822 & 2.780 & 0 & 0 \\ 0 & 0 & 2.780 & 2.500 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7.822 & 2.780 \\ 0 & 0 & 0 & 0 & 2.780 & 2.500 \end{pmatrix}$$

and

$$\mathcal{F}_{\gamma_1\gamma_2}(\theta) = \mathcal{F}^{\top}_{\gamma_2\gamma_1}(\theta) = \begin{pmatrix} -5.495 & 0.163 & 0 & 0 & 0 & 0 \\ -3.355 & -0.890 & 0 & 0 & 0 & 0 \\ 0 & 0 & -5.495 & 0.163 & 0 & 0 \\ 0 & 0 & -3.355 & -0.890 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5.495 & 0.163 \\ 0 & 0 & 0 & 0 & -3.355 & -0.890 \end{pmatrix}.$$

It can be seen that the submatrices of $\mathcal{F}_{\gamma\gamma}(\theta)$ are block Toeplitz matrices. The parametrization for the submatrix β is $\text{vec}\beta_1 = (\beta_1^{11}, \beta_1^{21}, \beta_1^{12}, \beta_1^{22})^{\top}$. We have according to [25]

$$(\mathcal{F}_{\beta\beta}(\theta))^{c,s}_{i,j,l,f} = \frac{1}{2\pi i}\oint\limits_{|z|=1} z^{c-s}\text{Tr }\left(\beta^{-1}(z)E_{ij}\Sigma E_{lf}^{\top}\beta^{-*}(z)\Sigma^{-1}\right)\frac{dz}{z},$$

the choice $\Sigma = I_2$, where I_2 is the two-dimensional identity matrix and $c, s = 1$, yields

$$(\mathcal{F}_{\beta\beta}(\theta))^{1,1}_{i,j,l,f} = \frac{1}{2\pi i}\oint\limits_{|z|=1} \text{Tr}\left(\beta^{-1}(z)E_{ij}E_{lf}^{\top}\beta^{-*}(z)\right)\frac{dz}{z},$$

where the $m \times m$ matrix $E_{ij} = e_i e_j^{\top}$, the vectors e_i and e_j are respectively the i-th and j-th standard basis vectors in $\mathbb{R}^m$, $m = 2$, and $i, j, l, f = 1, 2$, to obtain (to 3 decimal places)

$$\mathcal{F}_{\beta\beta}(\theta) = \begin{pmatrix} 7.822 & 2.780 & 0 & 0 \\ 2.780 & 2.500 & 0 & 0 \\ 0 & 0 & 7.822 & 2.780 \\ 0 & 0 & 2.780 & 2.500 \end{pmatrix}.$$

First we set forth the general representation of the entries of the asymptotic Fisher information submatrix $\mathcal{F}_{\alpha\alpha}(\theta)$ which are computed according to [25], to obtain

$$(\mathcal{F}_{\alpha\alpha}(\theta))^{k,v}_{i,j,l,f} = (\mathcal{F}^{u}_{\alpha\alpha}(\theta))^{k,v}_{i,j,l,f} + (\mathcal{F}^{w}_{\alpha\alpha}(\theta))^{k,v}_{i,j,l,f},$$

where

$$(\mathcal{F}^u_{\alpha\alpha}(\theta))^{k,v}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} z^{k-v}\mathrm{Tr}\left(\beta^{-1}(z)E_{ij}\alpha^{-1}(z)\gamma(z)R_u(z)\gamma^*(z)\alpha^{-*}(z)E^\top_{lf}\beta^{-*}(z)\Sigma^{-1}\right)\frac{dz}{z} \tag{38}$$

and

$$(\mathcal{F}^w_{\alpha\alpha}(\theta))^{k,v}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} z^{k-v}\mathrm{Tr}\left(\beta^{-1}(z)E_{ij}\alpha^{-1}(z)\beta(z)\Sigma\beta^*(z)\alpha^{-*}(z)E^\top_{lf}\beta^{-*}(z)\Sigma^{-1}\right)\frac{dz}{z}. \tag{39}$$

In this example the submatrix associated with the (α, α) block is considered when $\alpha = 0$ and we further have for the input matrix polynomial $\gamma(z) = 0$, the α parameters in the Fisher information matrix are ordered accordingly, to obtain $\mathrm{vec}\alpha_1 = (\alpha_1^{11}, \alpha_1^{21}, \alpha_1^{12}, \alpha_1^{22})^\top$. Additionally, when $\Sigma = \alpha(z) = I_2$, and the indices $k, v = 1$, $m = 2$ combined with $i, j, l, f = 1, 2$, yield $(\mathcal{F}^u_{\alpha\alpha}(\theta))^{k,v}_{i,j,l,f} = 0$, since expression (38) depends on the input process $u(t)$ through the spectral density $R_u(z)$ and the matrix polynomial $\gamma(z)$. The Fisher information submatrix associated with block $(\alpha,\ \alpha)$ is therefore given by (39) according to

$$(\mathcal{F}_{\alpha\alpha}(\theta))^{1,1}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} \mathrm{Tr}\left(\beta^{-1}(z)E_{ij}\beta(z)\beta^*(z)E^\top_{lf}\beta^{-*}(z)\right)\frac{dz}{z},$$

when summarized for all $i, j, l, f = 1, 2$, we obtain

$$\mathcal{F}_{\alpha\alpha}(\theta) = \begin{pmatrix} 7.855 & 3.648 & -8.979 & -6.855 \\ 3.648 & 4.588 & -0.170 & -3.648 \\ -8.979 & -0.170 & 25.665 & 8.979 \\ -6.855 & -3.648 & 8.979 & 7.855 \end{pmatrix}.$$

The submatrix associated with $\alpha\beta$ is considered. The entries of the appropriate Fisher information submatrix are computed according to

$$(\mathcal{F}_{\alpha\beta}(\theta))^{k,s}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} z^{k-s}\mathrm{Tr}\left(-\beta^{-1}(z)E_{ij}\alpha^{-1}(z)\beta(z)\Sigma E^\top_{lf}\beta^{-*}(z)\Sigma^{-1}\right)\frac{dz}{z}, \tag{40}$$

where the superscripts $k = 0, \dots, p-1$ and $s = 0, \dots, q-1$ and the subscripts $i, j, l, f = 1, \dots, m$. The choice, $\Sigma = \alpha(z) = I_2$, combined with $k, s = 1$, $m = 2$ and $i, j, l, f = 1, 2$, yields

$$(\mathcal{F}_{\alpha\beta}(\theta))^{1,1}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} \mathrm{Tr}\left(-\beta^{-1}(z)E_{ij}\beta(z)E^\top_{lf}\beta^{-*}(z)\right)\frac{dz}{z},$$

when summarized, to obtain

$$\mathcal{F}_{\alpha\beta}(\theta) = \mathcal{F}^\top_{\beta\alpha}(\theta) = \begin{pmatrix} -1.229 & 1.246 & 2.747 & 1.678 \\ -2.976 & -1.431 & -0.082 & 0.445 \\ -7.693 & -4.697 & -8.921 & -3.451 \\ 0.229 & -1.246 & -2.747 & -2.678 \end{pmatrix}.$$

Note that the entries of $\mathcal{F}_{\beta\alpha}(\theta)$ are computed according to

$$(\mathcal{F}_{\beta\alpha}(\theta))^{s,k}_{l,f,i,j} = \frac{1}{2\pi i} \oint\limits_{|z|=1} z^{s-k}\mathrm{Tr}\left(-\Sigma^{-1}\beta^{-1}(z)E_{lf}\Sigma\beta^*(z)\alpha^{-*}(z)E^\top_{ij}\beta^{-*}(z)\right)\frac{dz}{z}. \tag{41}$$

The entries of the submatrices $\mathcal{F}_{\alpha\gamma}(\theta)$ and $\mathcal{F}_{\gamma\alpha}(\theta)$ are given by the following equations when expressed by the Cauchy integrals

$$(\mathcal{F}_{\alpha\gamma}(\theta))^{k,g}_{i,j,l,f} = \frac{1}{2\pi i} \oint\limits_{|z|=1} z^{k-g} \mathrm{Tr}\left(-\beta^{-1}(z) E_{ij} \alpha^{-1}(z)\gamma(z) R_u(z) \mathcal{E}_{lf}^{\top} \beta^{-*}(z) \Sigma^{-1}\right) \frac{dz}{z} \quad (42)$$

and

$$(\mathcal{F}_{\gamma\alpha}(\theta))^{g,k}_{l,f,i,j} = \frac{1}{2\pi i} \oint\limits_{|z|=1} z^{g-k} \mathrm{Tr}\left(-\Sigma^{-1} \beta^{-1}(z) \mathcal{E}_{lf} R_u(z) \gamma^{*}(z) \alpha^{-*}(z) E_{ij}^{\top} \beta^{-*}(z)\right) \frac{dz}{z}, \quad (43)$$

where the superscripts $k = 0, \ldots, p-1$ and $g = 0, \ldots, e$, the subscripts $i, j, l = 1, \ldots, m$ and $f = 1, \ldots, r$. Note that the above equations (41), (40), (42) and (43) are corrected with respect to [25]. The property $\mathcal{F}_{\alpha\gamma}(\theta) = \mathcal{F}^{\top}_{\gamma\alpha}(\theta)$ holds. When the input matrix polynomial $\gamma(z) = 0$ then $\mathcal{F}_{\alpha\gamma}(\theta) = 0$.

We have generated 1000 observations of u and w, using Gaussian deviates with mean 0 and variance 1 and then obtained y using (1) with $\alpha_1 = 0$, $\gamma_1 = 0$, $\gamma_2 = 0$, and β_1 as given above. The data are available at http:\\homepages.ulb.ac.be\~gmelard\rech\km12data.zip. Now using the method in this paper we have computed the exact information matrix. This model (and also simpler models) allowed us to check (and sometimes correct) the Matlab program based on the theory. For these single simulated series of $N = 1000$ observations, we obtained for the exact Fisher information multiplied by N

$$\mathcal{F}_{\gamma_1\gamma_1}(\theta) = \begin{pmatrix} 7.678 & 2.735 & 0.450 & 0.084 & -0.107 & 0.014 \\ 2.735 & 2.459 & 0.205 & 0.099 & -0.098 & -0.055 \\ 0.450 & 0.205 & 6.608 & 2.438 & 0.000 & -0.114 \\ 0.084 & 0.099 & 2.438 & 2.324 & 0.131 & 0.012 \\ -0.107 & -0.098 & 0.000 & 0.131 & 7.481 & 2.661 \\ -0.014 & -0.055 & -0.114 & 0.012 & 2.661 & 2.383 \end{pmatrix},$$

$$\mathcal{F}_{\gamma_2\gamma_2}(\theta) = \begin{pmatrix} 7.704 & 2.743 & 0.479 & 0.099 & -0.100 & 0.003 \\ 2.743 & 2.462 & 0.213 & 0.104 & -0.094 & -0.057 \\ 0.479 & 0.213 & 6.634 & 2.450 & 0.015 & -0.119 \\ 0.099 & 0.104 & 2.450 & 2.330 & 0.140 & 0.011 \\ -0.100 & -0.094 & 0.015 & 0.140 & 7.472 & 2.647 \\ 0.003 & -0.057 & -0.119 & 0.011 & 2.647 & 2.376 \end{pmatrix},$$

$$\mathcal{F}_{\gamma_1\gamma_2}(\theta) = \mathcal{F}^{\top}_{\gamma_2\gamma_1}(\theta) = \begin{pmatrix} -5.405 & 0.147 & -0.274 & 0.057 & -0.049 & -0.107 \\ -3.286 & -0.864 & -0.288 & -0.117 & 0.095 & 0.012 \\ -0.480 & -0.110 & -4.401 & 0.342 & 0.219 & 0.160 \\ -0.221 & -0.133 & -2.883 & -0.687 & 0.013 & 0.094 \\ 0.123 & 0.021 & -0.212 & -0.146 & -5.274 & 0.126 \\ 0.076 & 0.064 & -0.055 & -0.115 & -3.140 & -0.793 \end{pmatrix},$$

$$\mathcal{F}_{\alpha\alpha}(\theta) = \begin{pmatrix} 7.834 & 3.639 & -8.952 & -6.835 \\ 3.639 & 4.580 & -0.167 & -3.639 \\ -8.952 & -0.167 & 25.593 & 8.951 \\ -6.835 & -3.639 & 8.951 & 7.834 \end{pmatrix}, \quad \mathcal{F}_{\beta\beta}(\theta) = \begin{pmatrix} 7.799 & 2.772 & 0.005 & 0.001 \\ 2.772 & 2.493 & 0.005 & 0.003 \\ 0.005 & 0.005 & 7.790 & 2.766 \\ 0.001 & 0.003 & 2.766 & 2.489 \end{pmatrix},$$

$$\mathcal{F}_{\alpha\beta}(\theta) = \mathcal{F}^{\top}_{\beta\alpha}(\theta) = \begin{pmatrix} -1.227 & 1.241 & 2.739 & 1.672 \\ -2.970 & -1.431 & -0.083 & 0.443 \\ -7.671 & -4.685 & -8.896 & -3.440 \\ 0.229 & -1.242 & -2.739 & -2.670 \end{pmatrix}.$$

The results for the exact information matrix at $\theta = (\alpha_1, \beta_1, \gamma_1, \gamma_2)$, where $\alpha_1 = 0$, $\gamma_1 = 0$, $\gamma_2 = 0$, and $\beta_1^{11} = 6/5$, $\beta_1^{12} = 1/2$, $\beta_1^{21} = -(7/5)$, $\beta_1^{22} = -(1/5)$ are close to those of the asymptotic information matrix. Note that the results are dependent on the simulated values for u, since our exact information matrix is conditional on u. More precisely, the blocks for γ would not be the same for another set of simulations whereas those for α and β would be the same. Of course, in practice, the Fisher information matrix is evaluated not at the unknown true value θ but, for example, at the maximum likelihood estimate $\widehat{\theta}$, obtained for the series of observations. In that case, different blocks for α and β will be obtained for another series. For $N = 1000000$, the results are given in the Appendix B. That suggests the following conjecture that the convergence of the exact Fisher information matrix to the asymptotic Fisher information matrix, established by [17] for VARMA models should be extended to VARMAX models, at least under some assumptions. Note also that the asymptotic information matrix considered here is not conditional, which means that an alternative conditional definition should be used.

4.3 Comparison with the E4 Toolbox

We have mentioned in the introduction E4, a toolbox for Matlab ([37], [13]) which can evaluate the exact information matrix of general state space models, and can be specialized to VARMAX models. Note that E4 can be used to estimate the parameters of the models by themselves or in composite formulation, unconstrained or subject to linear and/or nonlinear constraints on the parameters, under standard conditions or in an extended framework that allows for observation errors, missing data or vector GARCH errors.

For a comparison with E4, we have used the same simulated series as in the previous subsection but with $N = 100$. We have then derived the exact information by using E4 with the several options for econd (ml or maximum likelihood, iu or exogenous first value, au or exogeneous mean, or zero) and vcond (idejong or based on [4], lyapunov or zero). It appears that for our model (and perhaps because of the particular configuration of the coefficients), the econd=auto option is identical to econd = ml, the maximum likelihood estimation of the initial state vector, and that the results for vcond = idejong and vcond = lyapunov are identical.

We have first examined the blocks (α, γ) and (β, γ) of the exact information matrix which were exactly 0 above. For some combinations of the options of E4, these blocks are not exactly 0. This is the case for econd = ml or the maximum likelihood estimation of the initial state vector. Note however that when econd = ml but vcond = zero (zero initial covariance matrix of the state vector), the block (β, γ) is exactly 0 but not the block (α, γ). E4 can also provide an approximation of the information matrix, the Watson and Engle approximation [39].

We have looked further in Table 1 at the other option combinations of E4 for which the blocks (α, γ) and (β, γ) are exactly zero, by comparing the E4 estimated standard errors for the 20 parameters to those obtained by our exact method. To save space, Table 1 contains only the results for a subset of parameters, i.e. α_1^{11}, β_1^{11} and γ_2^{23}. It appears that the results are identical for the parameters α and β for these option combinations vcond = lyapunov (or vcond = idejong) and econd = iu or econd = au or econd = zero. For these parameters, they are not identical to our exact results (denoted by KM in the tables) when vcond = zero or when econd = ml. The results are not identical for the parameters γ. On the contrary, the Watson-Engle approximation is bad for the parameters α and β but is nearly as good as the other E4 results for the parameters γ.

These results lead to the suggestion that, at least when $e > 1$, which is the case here, none of the E4 state vector initializations corresponds to (29). In order to illustrate the differences between the E4 options in a case where they are more sensitive than in the previous example, we have changed the generation of the exogenous variables so that the first value is more different from zero and also from the mean value, in order to increase the difference between the initial state vector options. As a matter a fact, we have generated the three variable u by a VAR process with mean different from 0. To emphasize the differences, we have also reduced the length of the series from 100 to 50. For the reasons mentioned above, only vcond = lyapunov was considered. As shown in Table 2,.the results are different for the γ's. None of the three options is uniformly better pour the 12 parameters γ but econd = iu has the smallest standard deviation than econd = au or econd = zero. A closer look at the E4 Toolbox manual [37] and at [3] reveals that they refer to [5] for the deterministic case whereas the latter paper treats stochastic but uncorrelated exogeneous variables. Apparently an equation like (29) is not mentioned. Nevertheless, our analysis is confirmed by the E4 results, and, likewise, the power of E4, which can handle a larger variety of state space models (including the case of nonstationary roots) is also emphasized.

Table 1. For different option combinations of E4 and the method of the paper (KM), results for the information matrix for the blocks (α, γ) and (β, γ), and for the standard errors of the parameters α_1^{11}, β_1^{11}, and γ_2^{23}. These results were obtained for simulated time series of 100 observations. Underlined E4 standard errors are identical to our KM results.

Method	econd	vcond	(α, γ)	(β, γ)	α_1^{11}	β_1^{11}	γ_2^{23}
E4	ml	idejong	$\neq 0$	$\neq 0$	0.4217	0.4454	0.14579
		lyapunov	$\neq 0$	$\neq 0$	0.4217	0.4454	0.14579
		zero	$\neq 0$	0	0.4198	0.4431	0.14579
	iu	idejong	0	0	$\underline{0.4278}$	$\underline{0.4517}$	0.14490
		lyapunov	0	0	$\underline{0.4278}$	$\underline{0.4517}$	0.14490
		zero	0	0	0.4265	0.4501	0.14576
	au	idejong	0	0	$\underline{0.4278}$	$\underline{0.4517}$	0.14576
		lyapunov	0	0	$\underline{0.4278}$	$\underline{0.4517}$	0.14574
		zero	0	0	0.4265	0.4501	0.14576
	zero	idejong	0	0	$\underline{0.4278}$	$\underline{0.4517}$	0.14576
		lyapunov	0	0	$\underline{0.4278}$	$\underline{0.4517}$	0.14576
		zero	0	0	0.4265	0.4501	0.14576
E4 Watson-Engle	ml	lyapunov	$\neq 0$	$\neq 0$	0.3448	0.3611	0.1467
	ml	zero	$\neq 0$	$\neq 0$	0.3457	0.3616	0.1468
	iu	lyapunov	$\neq 0$	$\neq 0$	0.3506	0.3682	0.1459
	au/zero	lyapunov	$\neq 0$	$\neq 0$	0.3741	0.3626	0.1470
	not ml	zero	$\neq 0$	$\neq 0$	0.3507	0.3683	0.1468
KM			0	0	0.4278	0.4517	0.14565

Table 2. For some option combinations of E4 and the method of the paper (KM), results for the standard errors of the parameters α_1^{ij}, β_1^{ij}, $i, j = 1, 2$, and γ_2^{ij}, $i = 1, 2$, $j = 1, 2, 3$. These results were obtained for simulated time series of 50 observations. Underlined E4 standard errors are the closest from KM results.

Method	econd	vcond	α_1^{11}	α_1^{21}	α_1^{12}	α_1^{22}		
E4	iu/au/zero	lyapunov	0.6108	0.7244	0.5625	0.6672		
KM			0.6108	0.7244	0.5625	0.6672		
			β_1^{11}	β_1^{21}	β_1^{12}	β_1^{22}		
E4	iu/au/zero	lyapunov	0.6452	0.7356	0.4555	0.6991		
KM			0.6452	0.7356	0.4555	0.6991		
			γ_1^{11}	γ_1^{21}	γ_1^{12}	γ_1^{22}	γ_1^{13}	γ_1^{23}
E4	iu	lyapunov	<u>0.0877</u>	0.1053	0.0843	<u>0.1193</u>	<u>0.0594</u>	<u>0.1073</u>
	au	lyapunov	0.0893	<u>0.1072</u>	0.0849	0.1180	0.0405	0.0573
	zero	lyapunov	0.0889	0.1052	<u>0.0857</u>	<u>0.1193</u>	0.0569	0.0936
KM			0.0875	0.1085	0.0861	0.1194	0.0613	0.1092
			γ_2^{11}	γ_2^{21}	γ_2^{12}	γ_2^{22}	γ_2^{13}	γ_2^{23}
E4	iu	lyapunov	<u>0.1179</u>	0.1554	0.1039	<u>0.1439</u>	<u>0.0547</u>	<u>0.0948</u>
	au	lyapunov	0.1201	0.1556	0.1021	0.1400	0.0423	0.0599
	zero	lyapunov	0.1201	<u>0.1544</u>	<u>0.1045</u>	0.1424	0.0529	0.0853
KM			0.1177	0.1532	0.1061	0.1443	0.0554	0.0958

5 Conclusion

This paper has established recursions at the matrix level for the exact Fisher information matrix of a VARMAX stochastic process, conditionally with respect to exogenous (deterministic or stochastic) variables. It can be seen as a generalization of [22] which was restricted to VARMA processes but the approach is more useful and also simpler. We could compare our results with E4, a Matlab Toolbox, which is aimed at estimation of a more general state space model, including the evaluation of the gradient and the exact information matrix. Note that, although the general principle stated by [36] is the same, the expressions there are not given at the matrix level but at the scalar level, and we could not find the detailed expressions in the literature, e.g. the papers cited in [13]. Our results are close to those obtained using E4 but not identical. We could pinpoint the cause of discrepancy, more specifically that (29) is not supported by E4. For long series we could also compare our results with the asymptotic information matrix, as proposed and illustrated by [25]. That comparison leads to the suggestion of a conjecture generalizing [17] from VARMA to VARMAX models. A first investigation of that conjecture indicates that it will not be true without additional assumptions.

Acknowledgements

We thank Toufik Zahaf and Jurek Niemczyk for their contribution at the start of this project when it was devoted to VARMA models.

Appendix A

In this Appendix, the derivatives of the Chandrasekhar equations are considered, using the rule

$$dA^{-1} = -A^{-1}\,(dA)\;A^{-1}$$

to obtain

$$\frac{\partial\left(\mathrm{vec}B_t\right)}{\partial\theta^\top} = \frac{\partial\left(\mathrm{vec}B_{t-1}\right)}{\partial\theta^\top} + \left[\left(HY_{t-1}X_{t-1}^\top\right)\otimes H\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}\right)}{\partial\theta^\top} + \left[\left(HY_{t-1}\right)\otimes\left(HY_{t-1}\right)\right]\frac{\partial\left(\mathrm{vec}X_{t-1}\right)}{\partial\theta^\top}$$
$$+ \left[H\otimes\left(HY_{t-1}X_{t-1}\right)\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}^\top\right)}{\partial\theta^\top}, \tag{44}$$

$$\begin{aligned}\frac{\partial\left(\mathrm{vec}K_t\right)}{\partial\theta^\top} &= \left[\left(B_t^{-1}B_{t-1}\right)\otimes I_n\right]\frac{\partial\left(\mathrm{vec}K_{t-1}\right)}{\partial\theta^\top} + \left[\left(B_t^{-1}HY_{t-1}X_{t-1}^\top Y_{t-1}^\top\right)\otimes I_n\right]\frac{\partial\left(\mathrm{vec}\phi\right)}{\partial\theta^\top}\\
&+ \left[B_t^{-1}\otimes K_{t-1}\right]\frac{\partial\left(\mathrm{vec}B_{t-1}\right)}{\partial\theta^\top} + \left[\left(B_t^{-1}HY_{t-1}X_{t-1}^\top\right)\otimes\phi\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}\right)}{\partial\theta^\top}\\
&- \left[B_t^{-1}\otimes\left(K_{t-1}B_{t-1}B_t^{-1}\right)\right]\frac{\partial\left(\mathrm{vec}B_t\right)}{\partial\theta^\top}\\
&+ \left[\left(B_t^{-1}HY_{t-1}\right)\otimes\phi Y_{t-1}\right]\frac{\partial\left(\mathrm{vec}X_{t-1}\right)}{\partial\theta^\top}\\
&+ \left[\left(B_t^{-1}H\right)\otimes\left(\phi Y_{t-1}X_{t-1}\right)\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}^\top\right)}{\partial\theta^\top}\\
&- \left[B_t^{-1}\otimes\left(\phi Y_{t-1}X_{t-1}Y_{t-1}^\top H^\top B_t^{-1}\right)\right]\frac{\partial\left(\mathrm{vec}B_t\right)}{\partial\theta^\top},\end{aligned} \tag{45}$$

$$\begin{aligned}\frac{\partial\left(\mathrm{vec}Y_t\right)}{\partial\theta^\top} &= \left[Y_{t-1}^\top\otimes I_n\right]\frac{\partial\left(\mathrm{vec}\phi\right)}{\partial\theta^\top} + \left[I_k\otimes\phi\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}\right)}{\partial\theta^\top}\\
&- \left[\left(Y_{t-1}^\top H^\top\right)\otimes I_n\right]\frac{\partial\left(\mathrm{vec}K_t\right)}{\partial\theta^\top}\\
&- \left[I_k\otimes\left(K_tH\right)\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}\right)}{\partial\theta^\top},\end{aligned} \tag{46}$$

$$\begin{aligned}\frac{\partial\left(\mathrm{vec}X_t\right)}{\partial\theta^\top} &= \frac{\partial\left(\mathrm{vec}X_{t-1}\right)}{\partial\theta^\top} - \left[\left(X_{t-1}^\top Y_{t-1}^\top H^\top B_t^{-1}HY_{t-1}\right)\otimes I_k\right]\frac{\partial\left(\mathrm{vec}X_{t-1}\right)}{\partial\theta^\top}\\
&- \left[\left(X_{t-1}^\top Y_{t-1}^\top H^\top B_t^{-1}H\right)\otimes X_{t-1}\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}^\top\right)}{\partial\theta^\top}\\
&+ \left[\left(X_{t-1}^\top Y_{t-1}^\top H^\top B_t^{-1}\right)\otimes\left(X_{t-1}Y_{t-1}^\top H^\top B_t^{-1}\right)\right]\frac{\partial\left(\mathrm{vec}B_t\right)}{\partial\theta^\top}\\
&- \left[X_{t-1}^\top\otimes\left(X_{t-1}Y_{t-1}^\top H^\top B_t^{-1}H\right)\right]\frac{\partial\left(\mathrm{vec}Y_{t-1}\right)}{\partial\theta^\top}\\
&- \left[I_k\otimes\left(X_{t-1}Y_{t-1}^\top H^\top B_t^{-1}HY_{t-1}\right)\right]\frac{\partial\left(\mathrm{vec}X_{t-1}\right)}{\partial\theta^\top}.\end{aligned} \tag{47}$$

Appendix B

For $N = 1000000$ observations, we obtained

$$\mathcal{F}_{\gamma_1\gamma_1}(\theta) = \begin{pmatrix} 7.833 & 2.784 & 0.007 & -0.002 & -0.018 & -0.003\\ 2.784 & 2.503 & 0.004 & -0.001 & -0.009 & -0.004\\ 0.007 & 0.004 & 7.830 & 2.781 & 0.011 & -0.004\\ -0.002 & -0.001 & 2.781 & 2.497 & -0.005 & -0.006\\ -0.018 & -0.009 & -0.011 & -0.005 & 7.820 & 2.778\\ -0.003 & -0.004 & -0.005 & -0.006 & 2.778 & 2.497 \end{pmatrix},$$

$$
\mathcal{F}_{\gamma_2\gamma_2}(\theta) = \begin{pmatrix} 7.833 & 2.784 & 0.007 & -0.002 & -0.018 & -0.003 \\ 2.784 & 2.503 & 0.004 & -0.001 & -0.009 & -0.004 \\ 0.007 & 0.004 & 7.830 & 2.781 & -0.011 & -0.005 \\ -0.002 & -0.001 & 2.781 & 2.497 & -0.005 & -0.006 \\ -0.018 & -0.009 & -0.011 & -0.005 & 7.820 & 2.778 \\ -0.003 & -0.004 & -0.005 & -0.006 & 2.778 & 2.497 \end{pmatrix},
$$

and

$$
\mathcal{F}_{\gamma_1\gamma_2}(\theta) = \mathcal{F}^{\top}_{\gamma_2\gamma_1}(\theta) = \begin{pmatrix} -5.502 & 0.164 & -0.003 & 0.001 & 0.011 & -0.001 \\ -3.362 & -0.893 & -0.004 & -0.001 & 0.006 & 0.001 \\ -0.015 & -0.009 & -5.507 & 0.158 & 0.005 & -0.003 \\ 0.001 & -0.002 & -3.355 & -0.891 & 0.007 & 0.003 \\ 0.020 & 0.005 & 0.004 & -0.003 & -5.496 & 0.161 \\ 0.008 & 0.005 & 0.006 & 0.001 & -3.352 & -0.889 \end{pmatrix},
$$

$$
\mathcal{F}_{\alpha\alpha}(\theta) = \begin{pmatrix} 7.855 & 3.648 & -8.979 & -6.855 \\ 3.648 & 4.588 & -0.170 & -3.648 \\ -8.979 & -0.170 & 25.665 & 8.979 \\ -6.855 & -3.648 & 8.979 & 7.855 \end{pmatrix}, \qquad \mathcal{F}_{\beta\beta}(\theta) = \begin{pmatrix} 7.822 & 2.780 & 0.000 & 0.000 \\ 2.780 & 2.500 & 0.000 & 0.000 \\ 0.000 & 0.000 & 7.822 & 2.780 \\ 0.000 & 0.000 & 2.780 & 2.500 \end{pmatrix},
$$

and finally

$$
\mathcal{F}_{\alpha\beta}(\theta) = \mathcal{F}^{\top}_{\beta\alpha}(\theta) = \begin{pmatrix} -1.229 & 1.246 & 2.747 & 1.678 \\ -2.976 & -1.431 & -0.082 & 0.445 \\ -7.697 & -4.697 & -8.921 & -3.451 \\ 0.229 & -1.246 & -2.747 & -2.678 \end{pmatrix}.
$$

References

[1] B.D.O. Anderson and J.B. Moore , "Optimal Filtering," Prentice-Hall, Englewood Cliffs, 1979, N.J.

[2] C.F. Ansley and P. Newbold, "Finite sample properties of estimation for autoregressive moving average models," *J. Econometrics*, vol.**13**, pp.159-183, 1980.

[3] J.Casals and S. Sotoca, "Exact initial conditions for maximum likelihood estimation of state space models with stochastic inputs," *Economics Letters*, vol.**57**, pp.261-267, 1997.

[4] P. de Jong, "Smoothing and interpolation with the state-space model," *J. Amer Statist. Assoc.*, vol. **84**, pp. 1085-1088, 1989.

[5] P. de Jong and Chu-Chun-Lin, "Stationary and non-stationary state space models," *J. Time Series Analysis*, vol. **15**, pp. 151-166, 1994.

[6] C. J. Demeure and C. T. Mullis, "The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.**37** pp. 545-552, 1989.

[7] B. G. Dharan, "A priori sample size evaluation and information matrix computation for time series models," *J. Statist. Comput. Simul.*, vol. **21**, pp.171-177, 1985.

[8] B. Friedlander , "On the computation of the Cramér-Rao bound for ARMA parameter estimation," *IEEE Trans. Acoust. Speech, Signal Processing,* vol. **32**, pp. 721-727, 1984.

[9] E.J. Hannan, "Multiple Time Series," John Wiley, 1970, New York.

[10] E.J. Hannan and M. Deistler, "The Statistical Theory of Linear Systems," John Wiley and Sons, 1988,New York.

[11] E.J. Hannan., W.T.M. Dunsmuir and M. Deistler , "Estimation of Vector Armax Models," *J.Multivariate Analysis*, vol.**10**, pp.275-295, 1980.

[12] E. J. Godolphin and J. M. Unwin , "Evaluation of the covariance matrix for the maximum likelihood estimator of a Gaussian autoregressive-moving average process," *Biometrika*, vol. **70**, pp. 279-284, 1983.

[13] M. Jerez, J. M Casals and S. Sotoca, "Signal Extraction for Linear State-Space Models including a free MATLAB Toolbox for time series modeling and decomposition", Lambert Academic Publishing, 2011, Saarbrücken (Germany).

[14] A.Klein, "A generalization of Whittle's formula for the information matrix of vector mixed time series," *Linear Algebra Appl. vol.* **321** 197-208, 2000.

[15] A. Klein and G. Mélard, "Fisher's information matrix for seasonal autoregressive-moving average models," *J. Time Series Analysis*, vol. **11**, pp. 231-237, 1990.

[16] A. Klein and G. Mélard, "Computation of the Fisher information matrix for SISO models," *IEEE Transactions on Signal Processing,* vol. **42**, pp. 684-688, 1994.

[17] A. Klein, G. Mélard and A. Saidi, "The asymptotic and exact Fisher information matrices of a vector ARMA process," *Statist. Prob. Letters*, vol. **78**, pp.1430-1433, 2008.

[18] A. Klein, G. Mélard and J. Niemczyk, Corrections to "Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules," *Linear Algebra Appl.*, vol. **420**, pp. 729-730, 2007.

[19] A. Klein, G. Mélard, J. Niemczyk and T. Zahaf, "A program for computing the exact Fisher information matrix of a Gaussian VARMA model," University of Amsterdam Econometrics Dept. discussion paper 2004/15 (http://www1.fee.uva.nl/pp/bin/486fulltext.pdf), 2004.

[20] A. Klein, G. Mélard and P. Spreij, "On the resultant property of the Fisher information matrix of a vector ARMA process," *Linear Algebra Appl.*, vol.**403** 291-313, 2005.

[21] A. Klein, G. Mélard and T. Zahaf, "Computation of the exact information matrix of Gaussian dynamic regression time series models," *Ann. Statist.*, vol. **26**, pp. 1636-1650, 1998.

[22] A. Klein, G. Mélard and T. Zahaf, "Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules," *Linear Algebra Appl.*, vol. **321**, pp. 209-232, 2000.

[23] A. Klein and H. Neudecker, "A direct derivation of the exact Fisher information matrix of Gaussian vector state space models," *Linear Algebra Appl.*, vol. **321**, pp. 233-238, 2000.

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

[25] A. Klein and P. Spreij, "Matrix differential calculus applied to multiple stationary time series and an extended Whittle formula for information matrices," *Linear Algebra Appl.*, vol. **430** 674-691, 2009.

[26] P. Lancaster and M. Tismenetsky, "The Theory of Matrices with Applications," (second edition), Academic Press, 1985, Orlando.

[27] G. Mélard and A. Klein, "On a fast algorithm for the exact information matrix of a Gaussian ARMA time series,"*IEEE Trans. Signal Processing*, vol. **42**, pp. 2201-2203, 1994.

[28] S. Mittnik and P.A. Zadrozny, "Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic model," *Econometrica*, vol. **61**, pp. 857-870, 1993.

[29] M. Morf, G. S. Sidhu and T. Kailath, "Some new algorithms for recursive estimation on constant, linear, discrete-time systems," *IEEE Trans. Automat. Contr,* vol. **19**, pp. 315-323, 1974.

[30] H. J. Newton, "The information matrices of the parameters of multiple mixed time series," *J.Multivariate Analysis*, vol. **8**, pp. 317-323, 1978.

[31] J. Niemczyk, "Computing the derivatives of the autocovariances of a VARMA process, COMPSTAT'2004 16th Symposium held in Prague, *Proceedings in Computational Statistics*, Jaromir Antoch (Ed.), Physica-Verlag, Heidelberg, pp. 1593-1600, 2004.

[32] B. Porat and B. Friedlander, "Computation of the exact information matrix of Gaussian time series with stationary random components," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. **14**, pp.118-130, 1986.

[33] V. Peterka and P. Vidinčev, Rational-fraction approximation of transfer functions. First IFAC Symposium on Identification in Automatic Control Systems. Prague, 1967.

[34] T.Söderström, "Description of a program for integrating rational functions around the unit circle," Technical Report 8467R, Department of Technology. Uppsala University, 1984.

[35] T.Söderström, "On computing the Cramér-Rao bound and covariance matrices for PEM estimates in linear state space models," 14th IFAC Symposium on System Identification, Newcastle, Australia, March, 29-31, 2006.

[36] J. Terceiro Lomba, "Estimation of Dynamic Econometric Models with Errors in Variables," Springer-Verlag, 1990,Berlin.

[37] J. Terceiro, J. M Casals, M. Jerez, G. R. Serano and S. Sotoca, Time Series Analysis using MATLAB, including a complete MATLAB Toolbox, http://www.ucm.es/info/icae/e4, 2000.

[38] G. T. TUNNICLIFFE WILSON, "Some efficient computational procedures for high order ARMA models," *J. Statist. Comput. Simul.*, vol. **8**, pp.303-309, 1979.

[39] M. W. WATSON AND R. F. ENGLE, "Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models," *J. of Econometrics.*, vol. **23**, pp.385-400, 1983.

[40] P.WHITTLE, "The analysis of multiple stationary time series," *J.Royal Statist. Soc. B.*, vol. **15**, pp.125-139, 1953.

[41] P. A. ZADROZNY, "Analytic Derivatives for Estimation of Linear Dynamic Models," *Comp. & Math. With Appls.*, vol. **18**, pp.539-553, 1989.

[42] P. A. ZADROZNY, Errata to "Analytical derivatives for estimation of linear dynamic models," *Comp. & Math. With Appls.*, vol. **24**, pp.289-290, 1992.

[43] P.A. ZADROZNY AND S. MITTNIK (1994), "Kalman filtering methods for computing information matrices for time-invariant periodic and generally time-varying VARMA models and samples," *Comp. & Math. With Appls.*, vol. **28**, pp.107-119, 1994.