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

André Klein^{a,*}, Guy Mélard^b

^a Formerly with Department of Quantitative Economics, University of Amsterdam, The Netherlands, now Rothschild Blv. 123 Apt.7, 65271 Tel-Aviv, Israel. ^b ECARES CP114/4, Université libre de Bruxelles, Avenue Franklin Roosevelt 50, B-1050 Bruxelles, Belgium.

Abstract

In this paper an algorithm is developed for the exact Fisher information matrix of a Gaussian vector ARMAX or VARMAX process. The algorithm proposed in this paper is composed by Chandrasekhar recursion equations at a vector-matrix level, and some of these recursions consist of derivatives based on appropriate differential rules applied to a state space model for a vector process. The chosen representation is such that the recursions extracted from the state space model are given in terms of expectations of derivatives of innovations, and not the process and observation disturbances. The algorithm will be illustrated by an example. On that example, a comparison is made with results from E4, a toolbox for Matlab, and with the asymptotic information matrix.

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

2010 MSC: 62B10, 62M10, 60G15, 37M10.

1. Introduction

This paper is devoted to the computation of the exact Fisher information matrix of an *m*-dimensional time series $\{y_1, \ldots, y_N\}$ of length N, generated by a Gaussian vector ARMAX, or VARMAX, process of order $(p, e, s), \{y_t, t \in \mathbb{Z}\},\$

^{*}Corresponding author

Email addresses: a.a.b.klein@uva.nl (André Klein), gmelard@ulb.ac.be (Guy Mélard)

 \mathbbm{Z} the set of integers. More precisely, consider the equation representation of a dynamic linear system,

$$\sum_{j=0}^{p} \alpha_j \ y_{t-j} = \gamma_0 + \sum_{j=1}^{e} \gamma_j \ u_{t-j} + \sum_{j=0}^{s} \beta_j \ w_{t-j}, \ t \in \mathbb{Z},$$
(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 $\alpha_j \in \mathbb{R}^{m \times m}$, $\gamma_0 \in \mathbb{R}^{m \times 1}$, $\gamma_j \in \mathbb{R}^{m \times r}$, and $\beta_j \in \mathbb{R}^{m \times m}$ are the associated parameter matrices. We additionally assume $\alpha_0 \equiv \beta_0 \equiv I_m$, the $m \times m$ identity matrix. The error $\{w_t, t \in \mathbb{Z}\}$ is a collection of Gaussian independent zero mean *m*-dimensional random variables with a positive definite covariance matrix Σ . In the following, we denote transposition by \top and the mathematical expectation by \mathbb{E} . We assume either that u_t is non stochastic or that u_t is stochastic. In the latter case, we assume $\mathbb{E}\{u_t \ w_{t'}^{\top}\} = 0$, for all t, t', and that statistical inference is performed conditionally on the values taken by u_t . Note that observations for u_t should be available for $t \geq 1 - e$.

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, where $z \in \mathbb{C}$ (with a duplicate use of z as an operator and as a complex variable which is usual in the signal and time series literature, e.g. [2], [8]). The assumptions $\det(\alpha(z)) \neq 0$ 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.

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 [2], [8], [25] and [32]. Let us store the coefficients in an $(\ell \times 1)$ vector $\theta = \operatorname{vec}(\alpha, \beta, \gamma)$, where $\ell = m^2(p+s) + m(re+1), \alpha = \operatorname{vec}(\alpha_1, \alpha_2, ..., \alpha_p), \beta =$ $\operatorname{vec}(\beta_1, \beta_2, ..., \beta_s), \gamma = \operatorname{vec}(\gamma_0, \gamma_1, \gamma_2, ..., \gamma_e)$, where $\operatorname{vec} X = \operatorname{col}(\operatorname{col}(X_{ij})_{i=1}^n)_{j=1}^n$, see e.g. [23]. In [9], the authors study the asymptotic properties of the maximum likelihood estimator of the coefficients of a VARMAX process based on Nobservations, $\hat{\theta}_N$ or more simply $\hat{\theta}$.

For that purpose, we need the Fisher information matrix. In time series models, we can distinguish the asymptotic Fisher information matrix and the exact Fisher information matrix. Most of the literature is devoted to the former but the exact Fisher information matrix is worth to be considered. On the one hand, using (2) for all $t \in \mathbb{Z}$ to determine the residual $w_t(\theta)$, and assuming that u_t is stochastic and that (y_t, u_t) is a Gaussian stationary process, the asymptotic Fisher information matrix $\mathcal{F}(\theta)$ is defined by the following $(\ell \times \ell)$ matrix which does not depend on t

$$\mathcal{F}(\theta) = \mathbb{E}\left\{ \left(\partial_{\theta} w_t(\theta)\right)^\top \Sigma^{-1} \left(\partial_{\theta} w_t(\theta)\right) \right\},\,$$

where $\partial_{\theta}(\cdot)$ denotes the $(v \times \ell)$ matrix $\partial(\cdot)/\partial \theta^{\top}$ for any $(v \times 1)$ column vector (\cdot) . See [22] for more details and Appendix B for examples. On the other hand, the exact Fisher information matrix $J_N(\theta)$ is based on the exact Gaussian likelihood $L(\theta)$, more precisely on

$$l(\theta) = -\log L(\theta) = \sum_{t=1}^{N} \left\{ \frac{m}{2} \log(2\pi) + \frac{1}{2} \log \det(B_t) + \frac{1}{2} \tilde{y}_t^\top B_t^{-1} \tilde{y}_t \right\}, \quad (3)$$

where \tilde{y}_t , the sample innovation, is defined below by (8) and $B_t = \mathbb{E}[\tilde{y}_t \tilde{y}_t^\top]$ is its covariance matrix. The exact information matrix is given by $J_N(\theta) = (1/N)\mathbb{E}(\partial^2 l(\theta)/\partial\theta\partial\theta^\top)$. It is shown in [21], where a formal proof is given at the matrix level that the following holds true

$$J_{N}(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left[\frac{1}{2} \left(\partial_{\theta} \operatorname{vec} B_{t} \right)^{\top} \left(B_{t} \otimes B_{t} \right)^{-1} \left(\partial_{\theta} \operatorname{vec} B_{t} \right) + \mathbb{E} \left\{ \left(\partial_{\theta} \widetilde{y}_{t} \right)^{\top} B_{t}^{-1} \left(\partial_{\theta} \widetilde{y}_{t} \right) \right\} \right]$$

$$\tag{4}$$

A proof for the scalar version of (4) is given in [31]. Note that, strictly speaking, the Fisher information matrix $\mathcal{I}(\theta)$ is defined as a limit for $N \to \infty$ and is such that $\sqrt{N}(\hat{\theta}_N - \theta) \to \mathcal{N}(0, \mathcal{I}^{-1}(\theta))$, in distribution, as $N \to \infty$ where \mathcal{N} is the normal distribution. This is true under suitable regularity assumptions. Here we are interested in the case of finite N, either $\mathcal{F}(\theta)$ or $J_N(\theta)$, and more specifically the latter.

In standard statistical theory (e.g. [24, Chapters 2 and 6]), assuming that the estimator is unbiased, the inverse of $\mathcal{I}(\theta)$ yields the Cramér-Rao bound and, provided that the estimators are asymptotically efficient, the asymptotic covariance matrix. For most purposes, the Fisher information matrix should be evaluated at the unknown true value θ but, more practically, at the maximum likelihood estimate $\hat{\theta}$, obtained for the series of observations, generally using an optimization algorithm. Then, tests on coefficients can be derived but the reverse problem can also be solved: how long should the series be in order to obtain a given statistical significance, see [6]. In [19], the authors considered the asymptotic Fisher information matrix of a VARMA process. They show that the Fisher information matrix is singular if and only if the matrix polynomials $\alpha(z)$ and $\beta(z)$ have at least one common root. Let us now present algorithms for computing the asymptotic and then the exact Fisher information matrices.

In [28], 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 [39]. That approach can be generalized to VARMAX processes and put in matrix-level form, see [22]. 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. [30]). However, the most efficient method consists in transforming the problem to the evaluation of the autocovariances of an ARMA model, see e.g. [14]. In [15], the authors have been mainly concerned with the single input single output (SISO) model but have also indicated that their method can be used for the VARMA model. For recent references about the asymptotic information matrix, see [35].

More recently, the exact information matrix has been studied. In [31] Porat and Friedlander have described an algorithm for a univariate ARMA model with a deterministic additive component. The method is both complex and also computationally intensive since the number of scalar operations is of order N^2 . Independently, in [40], [41] and [36] the respective authors have given a much more efficient algorithm based on the Kalman filter. This has been applied to the VARMA model in [40] and [41], and to the general state space form by [36]. The latter general case has also been treated by [33] and [34] but in an approximate way. Although the algorithms in [40], [41], [42], and [36] need a number of operations which is proportional to N, these algorithms 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 so much, so that the improvement with respect to the Porat and Friedlander method [31] can only be apparent. Generalizing [26], 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 he has not mentioned at all that the Chandrasekhar equations can also be used to derive the information matrix. This was done in [20] with an application to VARMA models. Working with the prediction error of the state vector made it difficult to handle correctly the initial conditions (see also [16]) and impossible to generalize the approach to VARMAX models.

Meanwhile a software called E4 [37] has been developed on the basis of [36] but also of more recent contributions (see the references in [11]). 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 allow for the treatment of 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]. In particular, there is 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]).

In this paper, we consider the exact Fisher information matrix $J_N(\theta)$ of

VARMAX processes, as a generalization with some improvements of the method proposed in [26] and [20]. The main contributions are (a) the use of recursions at a vector-matrix level, (b) derivation of exact and explicit initial conditions, and (c) computational performance. Indeed, (a) instead of writing recursions for each element of the information matrix, we write recursions as concisely as possible at the vector-matrix level. For that purpose, the differential rules used in [20] are applied. Then (b), contrarily to [20], 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. Moreover explicit and exact initial conditions are deduced, as illustrated in Section 3. Finally, (c) 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 this relatively little known package. It may be that the relations used are similar to ours but this cannot be confirmed since they are not documented.

The article is organized as follows. In Section 2, we present the model as well as a closed form expression for the Kalman and the Chandrasekhar recursions needed to evaluate the information matrix at a matrix level and not componentwise. In Section 3, we examine the special case of the VARMAX model. In Section 4, we compare the exact information obtained by our method with the results given by E4. This is done by using a specific model and some data, and we confront the exact information with the asymptotic information.

2. State space model and exact information

2.1. The state space model

As will be seen in Section 3, our vector linear times series model can be put under a more general state space form. Let $x_t \in \mathbb{R}^n$ be the vector of the state variables, and ϕ , Γ , F, γ_0 , H be matrices of dimensions, respectively, $n \times n$, $n \times r$, $n \times m$, $m \times 1$, $m \times n$. The state space structure is:

$$x_{t+1} = \phi x_t + \Gamma u_t + F w_t \tag{5}$$

$$y_t = \gamma_0 + Hx_t + w_t. \tag{6}$$

2.2. The Kalman equations

There are several ways to compute the exact likelihood (3) of a VARMAX time series. Except for the closed form expression of a normal multivariate density, a simple representation is based on the Kalman filter equations, due to [12] and [13]. We use the traditional notation e.g. $\hat{y}_{t|t-1}$ to define the linear prediction of y_t conditionally on the information at time t - 1. It is given by

$$\widehat{y}_{t|t-1} = \gamma_0 + H\widehat{x}_{t|t-1},\tag{7}$$

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

$$\widetilde{y}_t = y_t - \widehat{y}_{t|t-1} = y_t - \gamma_0 - H\widehat{x}_{t|t-1}.$$
(8)

Now $\hat{x}_{t|t-1}$ is based on the recurrence

$$\widehat{x}_{t+1|t} = \phi \widehat{x}_{t|t-1} + \Gamma u_t + K_t \widetilde{y}_t, \tag{9}$$

where the Kalman filter gain K_t is given below and the initial condition is discussed at the end of Section 3.

The Kalman filter consists of a collection of recursions, one of them giving $P_{t+1|t}$, the covariance matrix of the prediction error of the state vector

$$\widetilde{x}_t = x_t - \widehat{x}_{t|t-1}.\tag{10}$$

Recall that $B_t = \mathbb{E}[\tilde{y}_t \tilde{y}_t^{\top}]$. Note that the same noise is used in (5) and (6). Hence the standard recursions (e.g. [1, Chapter 5]) are slightly simplified under the form of

$$B_{t} = HP_{t+1|t}H^{\top} + \Sigma,$$

$$K_{t} = (\phi P_{t|t-1}H^{\top} + F\Sigma)B_{t}^{-1}$$

$$P_{t+1|t} = \phi P_{t|t-1}\phi^{\top} + F\Sigma F^{\top} - K_{t}B_{t}K_{t}^{\top},$$
(11)

giving, respectively, the covariance matrix B_t needed in (4), the Kalman filter gain K_t used in (11) and the so-called Riccati equation. The initial condition for the latter is provided in the present stationary case by the linear system of equations $P_{1|0} = \phi P_{1|0} \phi^{\top} + F \Sigma F^{\top}$. For other approaches on the Kalman filter in time series, see [7] and [10].

2.3. The Chandrasekhar equations

Given time-invariance of the state space model, an alternative to the Kalman filter equations is provided by the so-called Chandrasekhar recursion equations [27], see also [1, Chapter 6] and [2]. These equations are the most computationally efficient, even with respect to the Kalman filter.

In our context the Chandrasekhar equations make use of matrices X_t and Y_t with respective dimensions $m \times m$, $n \times m$. Besides (7-9), the recurrences are

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

$$K_{t} = [K_{t-1}B_{t-1} + \phi Y_{t-1}X_{t-1}Y_{t-1}^{\top}H^{\top}]B_{t}^{-1},$$

$$Y_{t} = [\phi - K_{t-1}H]Y_{t-1},$$

$$X_{t} = X_{t-1} - X_{t-1}Y_{t-1}^{\top}H^{\top}B_{t}^{-1}HY_{t-1}X_{t-1}.$$
(12)

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 operation count is reduced because the $n \times n$ matrix $P_{t+1|t}$ does not need to be computed, except a part of it, $P_{1|0}H^{\top}$, for t = 0.

2.4. The exact Fisher Information Matrix

In this subsection, we use the differential rules of [20] recalled in Appendix A in order to compute the exact Fisher information matrix at the vector-matrix level. The technique for evaluating the necessary derivatives of the recursion equations is equivalent to [20], where the authors have used the prediction error of the state vector (10). In this paper recursions are expressed in terms of expectations of derivatives of the $\hat{x}_{t|t-1}$ and this leads to an explicit or implementable algorithm at the general state space level. This implies that 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 a complete set of recursions is provided in Section 3 for the VARMAX process. The derived algorithm is then implementable. The suggested differential rules are displayed using the notations recalled in Appendix A.

Taking into account the property $\partial_{\theta} y_t = 0$ and $\partial_{\theta} u_t = 0$ (justified because the realization does not depend on the parameters α , β , and γ) and $d\Sigma = 0$, (9) yields $d\hat{x}_{t+1|t} = d\phi \ \hat{x}_{t|t-1} + \phi \ d\hat{x}_{t|t-1} + d\Gamma \ u_t + dK_t \ \tilde{y}_t + K_t \ d\tilde{y}_t$. Component-wise application of Rule 7 in Appendix A to (9) gives

$$d\widehat{x}_{t+1|t} = \left(\widehat{x}_{t|t-1}^{\top} \otimes I_n\right) \operatorname{vecd}\phi + \phi d\widehat{x}_{t|t-1} + \left(u_t^{\top} \otimes I_n\right) \operatorname{vecd}\Gamma + \left(\widetilde{y}_t^{\top} \otimes I_n\right) \operatorname{vecd}K_t + K_t d\widetilde{y}_t$$

We can now formulate the appropriate derivative of $\hat{x}_{t+1|t}$ with respect to θ by applying the approach described in [20], and recalled in Appendix A, to obtain

$$\partial_{\theta} \widehat{x}_{t+1|t} = \left(\widehat{x}_{t|t-1}^{\top} \otimes I_n \right) \partial_{\theta} \operatorname{vec} \phi + \phi \partial_{\theta} \widehat{x}_{t+1|t} + \left(u_t^{\top} \otimes I_n \right) \partial_{\theta} \operatorname{vec} \Gamma + \left(\widetilde{y}_t^{\top} \otimes I_n \right) \partial_{\theta} \operatorname{vec} K_t + K_t \partial_{\theta} \widetilde{y}_t.$$
(13)

Similarly for the derivative of \tilde{y}_t , we obtain from (8)

$$\partial_{\theta} \widetilde{y}_{t} = -\left\{ \partial_{\theta} \gamma_{0} + (\widehat{x}_{t \mid t-1}^{\top} \otimes I_{m}) \partial_{\theta} \operatorname{vec} H + H \partial_{\theta} \widehat{x}_{t \mid t-1} \right\}.$$
(14)

For computing the first term of (4) the derivatives of the Chandrasekhar equations are needed. The second term of (4) consists of the expected value of a stochastic component. We therefore vectorize $J_N(\theta)$ according to Rule 7, see Appendix A:

$$\operatorname{vec} J_{N}(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left\{ \frac{1}{2} \left[\left(\partial_{\theta} \operatorname{vec} B_{t} \right) \otimes \left(\partial_{\theta} \operatorname{vec} B_{t} \right) \right]^{\top} \operatorname{vec} \left(B_{t} \otimes B_{t} \right)^{-1} + \mathbb{E} \left\{ \partial_{\theta} \widetilde{y}_{t} \otimes \partial_{\theta} \widetilde{y}_{t} \right\}^{\top} \operatorname{vec} B_{t}^{-1} \right\}.$$
(15)

Equations (13) and (14) allow the right-hand side of (15) to be written in an appropriate way. This is fully done for the VARMAX case in Section 3.

3. An algorithm for the vector ARMAX model

An appropriate choice for a parametrization of (5) and (6) is given by

$$\phi = \begin{pmatrix} -\alpha_1 & I_m & 0_m \\ -\alpha_2 & 0_m & \ddots & \\ \vdots & \ddots & I_m \\ -\alpha_h & 0_m & \cdots & 0_m \end{pmatrix}, \quad F = \begin{pmatrix} \beta_1 - \alpha_1 \\ \beta_2 - \alpha_2 \\ \vdots \\ \beta_h - \alpha_h \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_h \end{pmatrix}, \quad (16)$$
and $H = \begin{pmatrix} I & 0 & \dots & 0_n \end{pmatrix}$

and
$$h = \max(p, s, e), 0_m$$
 is the $m \times m$ zero matrix, $\alpha_i = 0, i > p, \beta_i = 0, i > s,$

 $\gamma_i = 0, i > e$, 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

$$(x_t)_i = -\sum_{j=i}^p \alpha_j \ y_{t-j+i-1} + \sum_{j=i}^e \gamma_j \ u_{t-j+i-1} + \sum_{j=i}^s \beta_j \ w_{t-j+i-1}, \ t = 1, \dots, N.$$
(18)

Note that $\partial_{\theta} \operatorname{vec} H = 0$. Hence (14) simplifies to $\partial_{\theta} \tilde{y}_t = -\partial_{\theta} \gamma_0 - H \partial_{\theta} \hat{x}_{t|t-1}$ and we obtain a main recurrence equation analogous to (32) of [20]:

$$\mathbb{E}\left(\partial_{\theta}\widetilde{y}_{t}\otimes\partial_{\theta}\widetilde{y}_{t}\right) = (H\otimes H)\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\otimes\partial_{\theta}\widehat{x}_{t|t-1}\right) + \partial_{\theta}\gamma_{0}\otimes\partial_{\theta}\gamma_{0} + \partial_{\theta}\gamma_{0}\otimes\left(H\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\right)\right) + \left(H\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\right)\right)\otimes\partial_{\theta}\gamma_{0}, \quad (19)$$

but much shorter. Of course it is necessary to update the expectations in the right hand side of (19) by using, from (13), with the notation for commutation

matrices $M_{m,r}$ given in Appendix A,

$$\begin{split} \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t+1|t} \otimes \partial_{\theta} \widehat{x}_{t+1|t} \right) &= \left(\phi \otimes \phi \right) \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \partial_{\theta} \widehat{x}_{t|t-1} \right) + \left(K_{t} \otimes K_{t} \right) \mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \otimes \partial_{\theta} \widetilde{y}_{t} \right) \\ &+ \left\{ \left[\left\{ \mathbb{E} (\widehat{x}_{t|t-1}^{\top} \otimes \widehat{x}_{t|t-1}^{\top}) \otimes I_{n} \right\} M_{n,n^{2}} \right] \otimes I_{n} \right\} (\partial_{\theta} \operatorname{vec} \phi \otimes \partial_{\theta} \operatorname{vec} \phi) \\ &+ \left\{ M_{n,n} \left[\left\{ \phi \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \widehat{x}_{t|t-1}^{\top} \right) \right\} \otimes I_{n} \right] M_{\ell,n^{2}} \right\} (\partial_{\theta} \operatorname{vec} \phi \otimes I_{\ell}) \\ &+ \left((\mathbb{E} (\widehat{x}_{t|t-1}^{\top}) \otimes I_{n}) \partial_{\theta} \operatorname{vec} \phi \right) \otimes \left((u_{t}^{\top} \otimes I_{n}) \partial_{\theta} \operatorname{vec} \phi \right) \\ &+ \left((\mathbb{E} (\widehat{x}_{t|t-1}^{\top}) \otimes I_{n}) \partial_{\theta} \operatorname{vec} \phi \right) \otimes \left((u_{t}^{\top} \otimes I_{n}) \partial_{\theta} \operatorname{vec} \phi \right) \\ &+ \left(K_{t} \otimes I_{n} \right) \left\{ \mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \otimes \widehat{x}_{t|t-1}^{\top} \right) \otimes I_{n} \right\} (I_{\ell} \otimes \partial_{\theta} \operatorname{vec} \phi) \\ &+ \left(\phi \otimes K_{t} \right) \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \widehat{x}_{t|t-1}^{\top} \right) \otimes \partial_{\theta} \operatorname{vec} \Gamma \right\} \\ &+ \left(\phi \otimes K_{t} \right) \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \partial_{\theta} \widetilde{y}_{t} \right) + \left(K_{t} \otimes \phi \right) \mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \otimes \partial_{\theta} \widehat{x}_{t|t-1} \right) \\ &+ \left((u_{t}^{\top} \otimes I_{n}) \partial_{\theta} \operatorname{vec} \Gamma \right) \otimes \left(\left(\mathbb{E} \left(\widehat{x}_{t|t-1}^{\top} \right) \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \phi \right) \\ &+ \left((u_{t}^{\top} \otimes I_{n}) \partial_{\theta} \operatorname{vec} \Gamma \right) \otimes \left(\left(u_{t}^{\top} \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \Gamma \right) \\ &+ \left((u_{t}^{\top} \otimes I_{n}) \partial_{\theta} \operatorname{vec} \Gamma \right) \otimes \left(\left(u_{t}^{\top} \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \Gamma \right) \\ &+ \left((u_{t}^{\top} \otimes I_{n}) \partial_{\theta} \operatorname{vec} \Gamma \right) \otimes \left(\left(u_{t}^{\top} \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \Gamma \right) \\ &+ \left(\left(u_{t}^{\top} \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \Gamma \right) \otimes \left\{ K_{t} \mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \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(\partial_{\theta} \operatorname{vec} K_{t} \otimes \partial_{\theta} \operatorname{vec} K_{t} \right) \\ &+ \left\{ K_{t} \mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \right) \right\} \otimes \left(\left(u_{t}^{\top} \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \Gamma \right) , \\ &+ \left\{ K_{t} \mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \right) \right\} \otimes \left(\left(u_{t}^{\top} \otimes I_{n} \right) \partial_{\theta} \operatorname{vec} \Gamma \right) , \\ &+ \left\{ (U_{t}^{\top} \otimes U_{t} \right\} \otimes \left(U_{t}^{\top} \otimes U_{t} \right) \right\} \\ &+ \left\{ (U_{t}^{\top} \otimes U_{t} \right) \otimes \left(U_{t}^{\top} \otimes U_{t} \right) \otimes \left(U_{t}^{\top} \otimes U_$$

since $\mathbb{E}\left[\widetilde{y}_{t}\right] = 0$, $\mathbb{E}\left[\widehat{x}_{t|t-1} \otimes \widetilde{y}_{t}\right] = 0$, $\mathbb{E}\left[\left(\partial_{\theta}\widehat{x}_{t|t-1}\right) \otimes \widetilde{y}_{t}\right] = 0$, $\mathbb{E}\left[\left(\partial_{\theta}\widetilde{y}_{t}\right) \otimes \widetilde{y}_{t}\right] = 0$.

Indeed sample innovations \tilde{y}_t are zero mean uncorrelated random variables. Also $\mathbb{E}(\hat{x}_{t|t-1} \otimes \tilde{y}_t) = 0$ because $\hat{x}_{t|t-1}$ is in the space spanned by the observations till time t-1 included, whereas \tilde{y}_t is orthogonal to that space. The explanation is similar for $\partial_{\theta} \hat{x}_{t|t-1}$ and $\partial_{\theta} \tilde{y}_t$. Note that there are several ways to write terms in (20). We have made sure to reduce computations for large n and ℓ . Note that

$$\mathbb{E}\left(\partial_{\theta}\widetilde{y}_{t}\otimes\hat{x}_{t|t-1}^{\top}\right) = -\partial_{\theta}\gamma_{0}\mathbb{E}(\hat{x}_{t|t-1}^{\top}) - H\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\otimes\hat{x}_{t|t-1}^{\top}\right).$$

For the implementation of the fundamental recurrence equation (20), we

need four additional recursions as follows:

1.
$$\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t+1|t}\otimes\widehat{x}_{t+1|t}^{\top}\right) = \left\{ \left[\mathbb{E}(\widehat{x}_{t|t-1}^{\top}\otimes\widehat{x}_{t|t-1}^{\top})\otimes I_{n}\right]M_{n,n^{2}}\right\} \left(\partial_{\theta}\operatorname{vec}\phi\otimes\phi^{\top}\right) \\ + \left[\mathbb{E}(\widehat{x}_{t|t-1}^{\top})\otimes I_{n}\right] \left(\partial_{\theta}\operatorname{vec}\phi\right)\otimes\left(u_{t}^{\top}\Gamma^{\top}\right) \\ + \left\{\phi\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\otimes\widehat{x}_{t|t-1}^{\top}\right)\right\} \left(I_{\ell}\otimes\phi^{\top}\right) \\ + \left(\phi\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\right)\right)\otimes\left(u_{t}^{\top}\Gamma^{\top}\right) + K_{t}\mathbb{E}\left(\partial_{\theta}\widetilde{y}_{t}\otimes\widehat{x}_{t|t-1}^{\top}\right) \left(I_{\ell}\otimes\phi^{\top}\right) \\ + \left(\left(u_{t}^{\top}\otimes I_{n}\right)\partial_{\theta}\operatorname{vec}\Gamma\right)\otimes\left(\mathbb{E}(\widehat{x}_{t|t-1}^{\top})\phi^{\top}\right) \\ + \left(\left(u_{t}^{\top}\otimes I_{n}\right)\partial_{\theta}\operatorname{vec}\Gamma\right)\otimes\left(u_{t}^{\top}\Gamma^{\top}\right) + \left(K_{t}\mathbb{E}\left(\partial_{\theta}\widetilde{y}_{t}\right)\right)\otimes\left(u_{t}^{\top}\Gamma^{\top}\right) \\ + \left[\left\{\left(\operatorname{vec}B_{t}\right)^{\top}\otimes I_{n}\right\}M_{m,mn}\right] \left(\partial_{\theta}\operatorname{vec}K_{t}\otimes K_{t}^{\top}\right).$$
(21)

2.
$$\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t+1|t}\right) = \left[\left(\mathbb{E}\widehat{x}_{t|t-1}\right)^{\top} \otimes I_{n}\right] \partial_{\theta} \operatorname{vec}\phi + (\phi - K_{t}H)\mathbb{E}\left(\partial_{\theta}\widehat{x}_{t|t-1}\right) + (u_{t}^{\top} \otimes I_{n})\partial_{\theta}\operatorname{vec}\Gamma.$$
 (22)

3.
$$\mathbb{E}\left(\hat{x}_{t+1|t} \otimes \hat{x}_{t+1|t}\right) = (\phi \otimes \phi) \mathbb{E}(\hat{x}_{t|t-1} \otimes \hat{x}_{t|t-1}) + (\phi \otimes \Gamma) \left[\mathbb{E}(\hat{x}_{t|t-1}) \otimes u_t\right] + (\Gamma \otimes \phi) \left[u_t \otimes \mathbb{E}(\hat{x}_{t|t-1})\right] + (\Gamma \otimes \Gamma) \left(u_t \otimes u_t\right) + (K_t \otimes K_t) \operatorname{vec} B_t$$

$$(23)$$

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 [20]. Of course the derivatives of the Chandrasekhar equations, equations (46) to (49) in [20], recalled in Appendix A, are also needed.

To be complete we also need to state $P_{1|0}H^{\top}$ and its derivatives. This was done in [20, Section 5, pp. 225-228] and doesn't need to be repeated here to save space. Note that the already complex initializations for $\mathbb{E}((\partial_{\theta} \tilde{x}_1) \otimes (\partial_{\theta} \tilde{x}_1))$, with \tilde{x}_1 defined by (10) and other expressions (most of p. 229) were wrong and replaced by a still more complex initialization procedure described in [16]. Fortunately, things are much simpler here. Besides B_1 , K_1 , Y_1 , X_1 , and $P_{1|0}H^{\top}$, and their derivatives with respect to θ , the following initial values are needed. Because of (18) for t = 1, after projection in the initial state space we have for each subvector of dimension m

$$(\hat{x}_{1|0})_i = \sum_{j=i}^e \gamma_j \ u_{i-j}, \tag{24}$$

for i = 1, ..., h, if e > 0 and 0, otherwise, hence

$$\mathbb{E}(\hat{x}_{1|0})_{i} = \sum_{j=i}^{e} \gamma_{j} u_{i-j}, \qquad \mathbb{E}(\hat{x}_{1|0} \otimes \hat{x}_{1|0})_{i,g} = \sum_{j=i}^{e} \sum_{k=g}^{e} (\gamma_{j} \otimes \gamma_{k})(u_{i-j} \otimes u_{g-k}),$$
$$\mathbb{E}\left(\partial_{\theta} \hat{x}_{1|0}\right)_{i} = \sum_{j=i}^{e} (u_{i-j}^{\top} \otimes I_{m})\partial_{\theta} \operatorname{vec} \gamma_{j},$$
$$\mathbb{E}\left[\left(\partial_{\theta} \hat{x}_{1|0}\right) \otimes \hat{x}_{1|0}^{\top}\right]_{i,g} = \sum_{j=i}^{e} \sum_{k=g}^{e} (u_{i-j}^{\top} \otimes I_{m} \otimes u_{g-k}^{\top}) \left(\partial_{\theta} \operatorname{vec} \gamma_{j} \otimes \gamma_{k}^{\top}\right),$$
$$\mathbb{E}\left[\left(\partial_{\theta} \hat{x}_{1|0}\right) \otimes \left(\partial_{\theta} \hat{x}_{1|0}\right)\right]_{i,g} = \sum_{j=i}^{e} \sum_{k=g}^{e} (u_{i-j}^{\top} \otimes I_{m} \otimes u_{g-k}^{\top} \otimes I_{m}) \left(\partial_{\theta} \operatorname{vec} \gamma_{j} \otimes \partial_{\theta} \operatorname{vec} \gamma_{k}\right)\right]$$

for i, g = 1, ..., h, also if e > 0 and 0 otherwise. Note that i and g are block indices and that the elements of $((\partial_{\theta} \operatorname{vec} \gamma_1), ..., (\partial_{\theta} \operatorname{vec} \gamma_e))^{\top}$ are related to $\partial_{\theta} \operatorname{vec} \Gamma$ through a commutation matrix $((\partial_{\theta} \operatorname{vec} \gamma_1)^{\top}, ..., (\partial_{\theta} \operatorname{vec} \gamma_e)^{\top})^{\top} = M_{hmr,hmr} \partial_{\theta} \operatorname{vec} \Gamma$.

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 to those of E4, a toolbox for Matlab ([37], [11]), 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 [29] and [17], which were developed for VARMA models without exogenous variables.

4.1. The example

The results obtained through the algorithm described in this paper will be compared with the values of the entries of the asymptotic Fisher information matrix of a VARMAX process. First the asymptotic case is handled on the basis of [22]. The VARMAX process considered in this example is such that m = 2, r = 3 and p = q = s = e = 1. We further assume,

$$\Sigma = I_2 \text{ and } \Omega = I_3, \tag{25}$$

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 $\theta = \text{vec}(\alpha_1, \beta_1, \gamma_1, \gamma_2)$, where

$$\alpha_1 = \begin{pmatrix} \alpha_1^{11} & \alpha_1^{12} \\ \alpha_1^{21} & \alpha_1^{22} \end{pmatrix}, \ \beta_1 = \begin{pmatrix} \beta_1^{11} & \beta_1^{12} \\ \beta_1^{21} & \beta_1^{22} \end{pmatrix}, \ \gamma_j = \begin{pmatrix} \gamma_j^{11} & \gamma_j^{12} & \gamma_j^{13} \\ \gamma_j^{21} & \gamma_j^{22} & \gamma_j^{23} \end{pmatrix}, \ j = 1, 2$$

For the numerical illustration, like in [22], we assume

$$\alpha_1 = 0, \gamma_1 = 0, \gamma_2 = 0, \tag{26}$$

and specific entries of the matrix polynomial $\beta(z)$ with

$$\beta_1^{11} = 6/5, \beta_1^{12} = 1/2, \beta_1^{21} = -(7/5) \text{ and } \beta_1^{22} = -(1/5).$$
 (27)

4.2. The asymptotic Fisher information matrix

A partitioned form of the asymptotic Fisher information matrix is considered in Appendix B. It is partially based on the theory in [21] and on the example limited there to the case of the γ 's. As a by-product of the present paper, it appears that the formulas for the mixed blocks involving α and β are wrong. Corrected versions of these blocks are displayed in Appendix B.

The results for the exact information matrix at $\theta = (\alpha_1, \beta_1, \gamma_1, \gamma_2)$ 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 rather at the maximum likelihood estimate $\hat{\theta}$. In that case, different blocks for α and β will be obtained for another series.

For N = 1000000, the results for $J_N(\theta)$ are given in Appendix C. That suggests the conjecture of convergence of the exact Fisher information matrix to the asymptotic Fisher information matrix. That property is established by [18] for VARMA models and 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], [11]) 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 exogenous 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. 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 [38]. 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 (i.e. the square roots of the diagonal elements of $J_N^{-1}(\theta)/N$) 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 therefore omitted from the table) 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 (24). 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 variables u by a VAR process with a mean vector 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 for 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 [3] treats stochastic but uncorrelated exogenous variables. Apparently an equation like (24) is not mentioned. Nevertheless, our analysis is largely 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	lyapunov	$\neq 0$	$\neq 0$	0.4217	0.4454	0.14579
		zero	$\neq 0$	0	0.4198	0.4431	0.14579
	iu	lyapunov	0	0	0.4278	0.4517	0.14490
		zero	0	0	0.4265	0.4501	0.14576
	au	lyapunov	0	0	0.4278	0.4517	0.14574
		zero	0	0	0.4265	0.4501	0.14576
	zero	lyapunov	0	0	0.4278	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
	\mathbf{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

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 [20] 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 [11]. Our results are close to those obtained using E4 but not identical. We have pointed out the cause of discrepancy, more specifically that (24), the exact initialization when e > 1, is not supported by E4. For long series we have compared our results with the asymptotic information matrix, as proposed and illustrated by

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	•	
$\mathbf{K}\mathbf{M}$			0.6108	0.7244	0.5625	0.6671		
			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		
$\mathbf{K}\mathbf{M}$			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	0.0877	0.1053	0.0843	<u>0.1193</u>	0.0594	0.1073
	au	lyapunov	0.0893	0.1072	0.0849	0.1180	0.0405	0.0573
	zero	lyapunov	0.0889	0.1052	0.0857	0.1193	0.0569	0.0936
$\mathbf{K}\mathbf{M}$			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	0.1179	0.1554	0.1039	0.1439	0.0547	0.0948
	au	lyapunov	0.1201	0.1556	0.1021	0.1400	0.0423	0.0599
	zero	lyapunov	0.1201	0.1544	0.1045	0.1424	0.0529	0.0853
KM			0.1177	0.1532	0.1061	0.1443	0.0554	0.0958

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, γ_1^{ij} 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.

[22]. That comparison leads to the suggestion of a conjecture generalizing [18] 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. We are grateful to the referee and the associate editor for their very useful comments on the two versions of this paper which significantly improved the quality of the manuscript.

References

- B. D. O. Anderson, J. B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs N. J., 1979.
- [2] P. E. Caines, Linear Stochastic Systems, Wiley, New York, 1988.

- [3] J. Casals, S. Sotoca, Exact initial conditions for maximum likelihood estimation of state space models with stochastic inputs, Economics Letters, 57 (1997) 261-267.
- [4] P. de Jong, Smoothing and interpolation with the state-space model, J. Amer Statist. Assoc., 84 (1989) 1085-1088.
- [5] P. de Jong, Chu-Chun-Lin, Stationary and non-stationary state space models, J. Time Series Analysis, 15 (1994) 151-166.
- [6] B. G. Dharan, A priori sample size evaluation and information matrix computation for time series models, J. Statist. Comput. Simul., 21 (1985) 171-177.
- [7] J. Durbin, S. J. Koopman, Time Series Analysis by State Space Methods, Oxford University Press, New York, 2001.
- [8] E. J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, Wiley, New York, 1988.
- [9] E. J. Hannan, W. T. M. Dunsmuir, M. Deistler, Estimation of vector Armax models, J. Multivariate Analysis, 10 (1980) 275-295.
- [10] A. C. Harvey, Forecasting, Structural Time Series and the Kalman Filter, Cambridge University Press, Cambridge, 1989.
- [11] M. Jerez, J. Casals, 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).
- [12] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering, Transactions ASME. Series D 42 (1960) 34-45.
- [13] R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, Journal of Basic Engineering, Transactions ASME. Series D 83 (1961) 95-108.

- [14] A. Klein, G. Mélard, Fisher's information matrix for seasonal autoregressive-moving average models, Journal of Time Series Analysis Vol.11 (1990), 231-237.
- [15] A. Klein, G. Mélard, Computation of the Fisher information matrix for SISO models, IEEE Transactions on Signal Processing, 42 (1994) 684-688.
- [16] A. Klein, G. Mélard, 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., 420 (2007) 729-730.
- [17] A. Klein, G. Mélard, J. Niemczyk, 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.
- [18] A. Klein, G. Mélard, A. Saidi, The asymptotic and exact Fisher information matrices of a vector ARMA process, Statist. Prob. Letters, 78 (2008) 1430-1433.
- [19] A. Klein, G. Mélard, P. Spreij, On the resultant property of the Fisher information matrix of a vector ARMA process, Linear Algebra Appl., 403 (2005) 291-313.
- [20] A. Klein, G. Mélard, T. Zahaf, Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules, Linear Algebra Appl., **321** (2000) 209-232.
- [21] A. Klein, H. Neudecker, A direct derivation of the exact Fisher information matrix of Gaussian vector state space models, Linear Algebra Appl., 321 (2000) 233-238.
- [22] A. Klein, P. Spreij, Matrix differential calculus applied to multiple stationary time series and an extended Whittle formula for information matrices, Linear Algebra Appl., 430 (2009) 674-691.

- [23] P. Lancaster, M. Tismenetsky, The Theory of Matrices with Applications, (second edition), Academic Press, Orlando, 1985.
- [24] E. L. Lehmann, Theory of Point Estimation, Wiley, New York, 1983.
- [25] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2006.
- [26] G. Mélard, A. Klein, On a fast algorithm for the exact information matrix of a Gaussian ARMA time series, IEEE Trans. Signal Processing, 42 (1994) 2201-2203.
- [27] M. Morf, G. S. Sidhu, T. Kailath, Some new algorithms for recursive estimation on constant, linear, discrete-time systems, IEEE Trans. Automat. Contr., 19 (1974) 315-323.
- [28] H. J. Newton, The information matrices of the parameters of multiple mixed time series, J. Multivariate Analysis, 8 (1978) 317-323.
- [29] 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, 2004 1593-1600.
- [30] V. Peterka, P. Vidinčev, Rational-fraction approximation of transfer functions. First IFAC Symposium on Identification in Automatic Control Systems. Prague, 1967.
- [31] B. Porat, B. Friedlander, Computation of the exact information matrix of Gaussian time series with stationary random components, IEEE Trans. Acoust., Speech, Signal Processing, 14 (1986) 118-130.
- [32] G. C. Reinsel, Elements of Multivariate Time Series Analysis, Second Edition, Springer-Verlag, New York, 1997.

- [33] M. Segal, E. Weinstein, A new method for evaluating the log-likelihood gradient (score) of linear dynamic systems, IEEE Transactions on Automatic Control, **33** (1988) 763-766.
- [34] M. Segal, E. Weinstein, A new method for evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems, IEEE Transactions on Information Theory, 35 (1989) 682-687.
- [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, Berlin, 1990.
- [37] J. Terceiro, J. M. Casals, M. Jerez, G. R. Serano, S. Sotoca, Time Series Analysis using MATLAB, including a complete MATLAB Toolbox, http://www.ucm.es/info/icae/e4, 2000.
- [38] M. W. Watson, R. F. Engle, Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models, J. of Econometrics, 23 (1983) 385-400.
- [39] P. Whittle, The analysis of multiple stationary time series, J. Royal Statist. Soc. B, 15 (1953) 125-139.
- [40] P. A. Zadrozny, Analytic derivatives for estimation of linear dynamic models, Comp. & Math. With Appls., 18 (1989) 539-553.
- [41] P. A. Zadrozny, Errata to Analytical derivatives for estimation of linear dynamic models, Comp. & Math. With Appls., 24 (1992) 289-290.
- [42] P. A. Zadrozny, S. Mittnik, Kalman filtering methods for computing information matrices for time-invariant periodic and generally time-varying VARMA models and samples, Comp. & Math. With Appls., 28 (1994)

107-119.

A. Appendix

First let us recall the notations for the derivatives with respect to vectors and the needed related rules also used in [20].

Consider a real differentiable $(m \times n)$ matrix function $X(\theta)$ of real $(\ell \times 1)$ vector $\theta = (\theta_1, \ldots, \theta_\ell)^\top$, where m, n and ℓ are positive integers. Let $(m \times n)$ matrices $\partial_r X = \partial_\theta X_{ij}$ with $r = 1, \ldots, \ell$ be the first order derivatives of $X(\theta)$ in partial derivative form with X_{ij} being the element (i, j) of X. Then the $(m n \times \ell)$ matrix $\partial_\theta \operatorname{vec} X(\theta)$ is defined.

We further recall the rules also used in [20].

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 $M_{m,r}$ is defined by $M_{m,r} = \sum_{i=1}^{m} \sum_{j=1}^{r} (\mathcal{E}_{ij} \otimes \mathcal{E}_{ij}^{\top}) \in \mathbb{R}^{mr \times mr}$, where $\mathcal{E}_{ij} = e_i^m (e_j^r)^{\top}$, and e_i^m is the *i*-th unit standard basis column vector in \mathbb{R}^m and e_j^r is the *j*-th unit standard basis column vector in \mathbb{R}^r . Note also the properties $M_{r,m}^{\top} = M_{m,r}$ and $M_{1,m} = M_{m,1} = I_m$ and taking the orthogonality into account yields $M_{r,m}M_{m,r} = I_{mr}$.

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

We add

Rule 7. vec $ABC = (C^{\top} \otimes A)$ vecB, where $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$ and $C \in \mathbb{R}^{p \times s}$.

Computationally, the recursions of the paper are written in the less demanding form. Several times Rules 1 and 5 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 + mpnqklmultiplications, generally bigger than what is required by the right hand side mnk + pql + mkpl multiplications.

The equation for the general state space, which is a generalization of (19), is of the form:

$$\mathbb{E} \left(\partial_{\theta} \widetilde{y}_{t} \otimes \partial_{\theta} \widetilde{y}_{t} \right) = \left\{ \left\{ M_{m,1} (\mathbb{E} (\widehat{x}_{t|t-1}^{\top} \otimes \widehat{x}_{t|t-1}^{\top}) \otimes I_{m}) M_{mn,n} \right\} \otimes I_{m} \right\} (\partial_{\theta} \operatorname{vec} H \otimes \partial_{\theta} \operatorname{vec} H) \\
+ \left\{ M_{m,m} \left[\left\{ H\mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \widehat{x}_{t|t-1}^{\top} \right) \right\} \otimes I_{m} \right] M_{mn,\ell} \right\} (\partial_{\theta} \operatorname{vec} H \otimes I_{\ell}) \\
+ \left[\left\{ H\mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \widehat{x}_{t|t-1}^{\top} \right) \right\} \otimes I_{m} \right] (I_{\ell} \otimes \partial_{\theta} \operatorname{vec} H) \\
+ (H \otimes H) \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \otimes \partial_{\theta} \widehat{x}_{t|t-1} \right) \\
+ \partial_{\theta} \gamma_{0} \otimes \left((\mathbb{E} \left(\widehat{x}_{t|t-1}^{\top} \right) \otimes I_{m}) \partial_{\theta} \operatorname{vec} H + H \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \right) \right) \\
+ \left((\mathbb{E} \left(\widehat{x}_{t|t-1}^{\top} \right) \otimes I_{m}) \partial_{\theta} \operatorname{vec} H + H \mathbb{E} \left(\partial_{\theta} \widehat{x}_{t|t-1} \right) \right) \\
+ \partial_{\theta} \gamma_{0} \otimes \partial_{\theta} \gamma_{0}, \quad (A.1)$$

where the commutation matrix $M_{a,b}$ is defined in Rule 5.

Finally, the derivatives of the Chandrasekhar equations are considered, using the rule $dA^{-1} = -A^{-1} (dA) A^{-1}$ to obtain

$$\partial_{\theta} \operatorname{vec} B_{t} = \partial_{\theta} \operatorname{vec} B_{t-1} + \left[\left(HY_{t-1}X_{t-1}^{\top} \right) \otimes H \right] \partial_{\theta} \operatorname{vec} Y_{t-1} + \left[\left(HY_{t-1} \right) \otimes \left(HY_{t-1} \right) \right] \partial_{\theta} \operatorname{vec} X_{t-1} + \left[H \otimes \left(HY_{t-1}X_{t-1} \right) \right] \partial_{\theta} \operatorname{vec} Y_{t-1}^{\top}, \tag{A.2}$$

$$\partial_{\theta} \operatorname{vec} K_{t} = \left[\left(B_{t}^{-1} B_{t-1} \right) \otimes I_{n} \right] \partial_{\theta} \operatorname{vec} K_{t-1} + \left[\left(B_{t}^{-1} H Y_{t-1} X_{t-1}^{\top} Y_{t-1}^{\top} \right) \otimes I_{n} \right] \partial_{\theta} \operatorname{vec} \phi \\ + \left[B_{t}^{-1} \otimes K_{t-1} \right] \partial_{\theta} \operatorname{vec} B_{t-1} + \left[\left(B_{t}^{-1} H Y_{t-1} X_{t-1}^{\top} \right) \otimes \phi \right] \partial_{\theta} \operatorname{vec} Y_{t-1} \\ - \left[B_{t}^{-1} \otimes \left(K_{t-1} B_{t-1} B_{t}^{-1} \right) \right] \partial_{\theta} \operatorname{vec} B_{t} \\ + \left[\left(B_{t}^{-1} H Y_{t-1} \right) \otimes \phi Y_{t-1} \right] \partial_{\theta} \operatorname{vec} X_{t-1} \\ + \left[\left(B_{t}^{-1} H \right) \otimes \left(\phi Y_{t-1} X_{t-1} \right) \right] \partial_{\theta} \operatorname{vec} Y_{t-1}^{\top} \\ - \left[B_{t}^{-1} \otimes \left(\phi Y_{t-1} X_{t-1} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} \right) \right] \partial_{\theta} \operatorname{vec} B_{t},$$
(A.3)

$$\partial_{\theta} \operatorname{vec} Y_{t} = \left[Y_{t-1}^{\top} \otimes I_{n}\right] \partial_{\theta} \operatorname{vec} \phi + \left[I_{k} \otimes \phi\right] \partial_{\theta} \operatorname{vec} Y_{t-1} - \left[\left(Y_{t-1}^{\top} H^{\top}\right) \otimes I_{n}\right] \partial_{\theta} \operatorname{vec} K_{t} - \left[I_{k} \otimes (K_{t} H)\right] \partial_{\theta} \operatorname{vec} Y_{t-1},$$
(A.4)

$$\partial_{\theta} \operatorname{vec} X_{t} = \partial_{\theta} \operatorname{vec} X_{t-1} - \left[\left(X_{t-1}^{\top} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} H Y_{t-1} \right) \otimes I_{k} \right] \partial_{\theta} \operatorname{vec} X_{t-1} - \left[\left(X_{t-1}^{\top} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} H \right) \otimes X_{t-1} \right] \partial_{\theta} \operatorname{vec} Y_{t-1}^{\top} + \left[\left(X_{t-1}^{\top} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} \right) \otimes \left(X_{t-1} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} \right) \right] \partial_{\theta} \operatorname{vec} B_{t} - \left[X_{t-1}^{\top} \otimes \left(X_{t-1} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} H \right) \right] \partial_{\theta} \operatorname{vec} Y_{t-1} - \left[I_{k} \otimes \left(X_{t-1} Y_{t-1}^{\top} H^{\top} B_{t}^{-1} H Y_{t-1} \right) \right] \partial_{\theta} \operatorname{vec} X_{t-1}.$$
(A.5)

B. Appendix

We derive the asymptotic information matrix for the example of Section 4.1. The appropriate matrix polynomials are

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

and

$$\gamma(z) = \begin{pmatrix} \gamma_1^{11} + \gamma_2^{11}z & \gamma_1^{12} + \gamma_2^{12}z & \gamma_1^{13} + \gamma_2^{13}z \\ \gamma_1^{21} + \gamma_2^{21}z & \gamma_1^{22} + \gamma_2^{22}z & \gamma_1^{23} + \gamma_2^{23}z \end{pmatrix}.$$
 (B.1)

Let us consider the partitioned form of the asymptotic information matrix

$$\mathcal{F}(\theta) = \begin{pmatrix} \mathcal{F}_{\alpha\alpha}(\theta) & \mathcal{F}_{\alpha\beta}(\theta) & \mathcal{F}_{\alpha\gamma}(\theta) \\ \mathcal{F}_{\beta\alpha}(\theta) & \mathcal{F}_{\beta\beta}(\theta) & \mathcal{F}_{\beta\gamma}(\theta) \\ \mathcal{F}_{\gamma\alpha}(\theta) & \mathcal{F}_{\gamma\beta}(\theta) & \mathcal{F}_{\gamma\gamma}(\theta) \end{pmatrix}.$$
 (B.2)

Taking into consideration that the input u_t and the white noise w_t are orthogonal processes leads to the property

$$\mathcal{F}_{\gamma\beta}(\theta) = 0. \tag{B.3}$$

The numerical example displayed in [22] is such that the authors focus on some entries of the submatrix $\mathcal{F}_{\gamma\gamma}(\theta)$, considering the crucial role of the γ

parameters in VARMAX processes. The computations are extended here in order to compute the whole asymptotic information matrix. The partitioned form of $\mathcal{F}_{\gamma\gamma}(\theta)$ is considered, to obtain

$$\mathcal{F}_{\gamma\gamma}(\theta) = \begin{pmatrix} \mathcal{F}_{\gamma_1\gamma_1}(\theta) & \mathcal{F}_{\gamma_1\gamma_2}(\theta) \\ \mathcal{F}_{\gamma_2\gamma_1}(\theta) & \mathcal{F}_{\gamma_2\gamma_2}(\theta) \end{pmatrix}.$$
 (B.4)

The parametrization of input coefficient matrix $\gamma = ((\operatorname{vec}\gamma_1)^{\top}, (\operatorname{vec}\gamma_2)^{\top})^{\top}$ is given by $\operatorname{vec}\gamma_j = (\gamma_j^{11}, \gamma_j^{21}, \gamma_j^{12}, \gamma_j^{22}, \gamma_j^{13}, \gamma_j^{23})^{\top}$, j = 1, 2. Like in the numerical illustrations proposed in [22], we assume the specific values given by (26) and (27). The basic assumption that the eigenvalues of the matrix polynomial $\beta(z)$ lie outside the unit circle is fulfilled since the eigenvalues are $(5/23) (-5 \pm i\sqrt{21})$ with modulus equal to 1.47442. This assumption is fundamental for evaluating the appropriate integrals displayed in this paper and in [22], so that the Peterka and Vidinčev [30] algorithm can be implemented. According to [22], with \mathcal{E}_{ij} as defined in Appendix A, Rule 5, the elements of (B.4) can be written

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

where $\operatorname{Tr}(M)$ is the trace of a square matrix M, and the subscripts are $i, l = 1, 2, \ldots, m$ and $j, f = 1, 2, \ldots, r$ and the superscripts are $d, g = 1, \ldots, e$. The Cauchy integral is counterclockwise, X^* is the complex conjugate transpose of complex matrix X and X^{-*} is its inverse. The spectral density $R_u(z)$ of the input process u_t is an $r \times r$ Hermitian matrix. For a definition, see e.g. [2] and in [22] it is given by the equations (17) and (18) page 679. Given (25), $R_u(z) = I_3$, and we obtain in the example

$$(\mathcal{F}_{\gamma_1\gamma_1}(\theta))_{1,1,1,1}^{1,1} = \frac{1}{2\pi i} \oint_{\substack{|z|=1}} \operatorname{Tr} \left(\beta^{-1}(z)\mathcal{E}_{11}\mathcal{E}_{11}^\top \beta^{-*}(z)\right) \frac{dz}{z}$$
$$= -\frac{1}{2\pi i} \oint_{\substack{|z|=1}} \frac{500 \ z \ (1 - 15z + z^2)}{(50 + 50z + 23z^2) \ (23 + 50z + 50z^2)} \frac{dz}{z}$$
$$= 7.82242.$$

We proceed accordingly for the remaining γ parameters. It yields the following submatrices of (B.4) 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

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

It can be seen that the submatrices of $\mathcal{F}_{\gamma\gamma}(\theta)$ are block Toeplitz matrices.

The parametrization for the submatrix β is $\operatorname{vec}\beta_1 = (\beta_1^{11}, \beta_1^{21}, \beta_1^{12}, \beta_1^{22})^\top$. We have according to [22]

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

where c, s = 0, 1, ..., q - 1 and i, j, l, f = 1, ..., m and when applied to the case $\Sigma = I_2$, and taking c, s = 1, it yields

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

where the $m \times m$ matrix $E_{ij} = e_i^m (e_j^m)^{\top}$, where e_i^m and e_j^m are defined in Appendix A, Rule 5. The case m = 2, and i, j, l, f = 1, 2, leads to (to 3 decimal

places)

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

Let $\operatorname{vec}\alpha_1 = (\alpha_1^{11}, \alpha_1^{21}, \alpha_1^{12}, \alpha_1^{22})^\top$. Now 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 [22], to obtain

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

where

$$\left(\mathcal{F}_{\alpha\alpha}^{u}(\theta)\right)_{i,j,l,f}^{k,v} = \frac{1}{2\pi i} \oint_{|z|=1} z^{k-v} \operatorname{Tr}\left(\beta^{-1}(z) E_{ij} \alpha^{-1}(z) \gamma(z) R_{u}(z) \gamma^{*}(z) \alpha^{-*}(z) E_{lf}^{\top} \beta^{-*}(z) \Sigma^{-1}\right) \frac{dz}{z}$$
(B.5)

and

$$\left(\mathcal{F}^{w}_{\alpha\alpha}(\theta)\right)^{k,v}_{i,j,l,f} = \frac{1}{2\pi i} \oint_{|z|=1} z^{k-v} \operatorname{Tr}\left(\beta^{-1}(z) E_{ij} \alpha^{-1}(z) \beta(z) \Sigma \beta^{*}(z) \alpha^{-*}(z) E_{lf}^{\top} \beta^{-*}(z) \Sigma^{-1}\right) \frac{dz}{z},$$
(B.6)

where k, v = 0, 1, ..., p - 1 and i, j, l, f = 1, ..., m. Since in the example $\gamma(z) = 0$, (B.5) vanishes, and since $\alpha = 0$, $\alpha(z) = I_2$, the block (α, α) of the Fisher information matrix given by (B.6) becomes

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

giving after computation for i,j,l,f=1,2

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

The submatrix associated with $\alpha\beta$ is now considered. The entries of the

appropriate Fisher information submatrix are computed according to

$$\left(\mathcal{F}_{\alpha\beta}(\theta)\right)_{i,j,l,f}^{k,s} = \frac{1}{2\pi i} \oint_{|z|=1} z^{k-s} \operatorname{Tr}\left(-\beta^{-1}(z) E_{ij} \alpha^{-1}(z) \beta(z) \Sigma E_{lf}^{\top} \beta^{-*}(z) \Sigma^{-1}\right) \frac{dz}{z}.$$
(B.7)

The choice, $\Sigma = \alpha(z) = I_2$, combined with k, s = 1, m = 2 and i, j, l, f = 1, 2, yields

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

giving after computation

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

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

$$\left(\mathcal{F}_{\beta\alpha}(\theta)\right)_{l,f,i,j}^{s,k} = \frac{1}{2\pi i} \oint_{|z|=1} z^{s-k} \operatorname{Tr}\left(-\Sigma^{-1}\beta^{-1}(z) E_{lf} \Sigma \beta^*(z) \alpha^{-*}(z) E_{ij}^{\top} \beta^{-*}(z)\right) \frac{dz}{z}.$$
(B.8)

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 Cauchy integrals

$$\left(\mathcal{F}_{\alpha\gamma}(\theta)\right)_{i,j,l,f}^{k,g} = \frac{1}{2\pi i} \oint_{|z|=1} z^{k-g} \operatorname{Tr}\left(-\beta^{-1}(z) E_{ij} \alpha^{-1}(z) \gamma(z) R_u(z) \mathcal{E}_{lf}^{\top} \beta^{-*}(z) \Sigma^{-1}\right) \frac{dz}{z}$$
(B.9)

and

$$\left(\mathcal{F}_{\gamma\alpha}(\theta)\right)_{l,f,i,j}^{g,k} = \frac{1}{2\pi i} \oint_{|z|=1} z^{g-k} \operatorname{Tr}\left(-\Sigma^{-1}\beta^{-1}(z)\mathcal{E}_{lf}R_u(z)\gamma^*(z)\alpha^{-*}(z)E_{ij}^{\top}\beta^{-*}(z)\right) \frac{dz}{z},$$
(B.10)

where the subscripts are k = 0, ..., p - 1, g = 1, ..., e, i, j, l = 1, ..., m and f = 1, ..., r. Note that the aforementioned equations (B.7), (B.8), (B.9) and (B.10) are corrected with respect to [22]. The property $\mathcal{F}_{\alpha\gamma}(\theta) = \mathcal{F}_{\gamma\alpha}^{\top}(\theta)$ holds. When the input matrix polynomial $\gamma(z) = 0$ then $\mathcal{F}_{\alpha\gamma}(\theta) = 0$.

We have generated 1000 observations of u_t and w_t , using Gaussian deviates with mean 0 and variance 1 and then obtained y_t 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 $\gmelard/rech/km12data.zip$. By using now the method developed 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

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

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

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

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

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

C. Appendix

For ${\cal N}=1000000$ observations, we obtained for the exact information matrix

$$(J_N(\theta))_{\gamma_1\gamma_1} = \begin{pmatrix} 7.810 & 2.775 & 0.003 & 0.004 & -0.009 & -0.012 \\ 2.775 & 2.495 & -0.002 & 0.001 & 0.005 & -0.004 \\ 0.003 & -0.002 & 7.788 & 2.770 & 0.007 & 0.004 \\ 0.004 & 0.001 & 2.770 & 2.494 & 0.000 & 0.001 \\ -0.009 & 0.005 & 0.007 & 0.000 & 7.832 & 2.784 \\ -0.012 & -0.004 & 0.004 & 0.001 & 2.784 & 2.504 \end{pmatrix},$$

$$(J_N(\theta))_{\gamma_2\gamma_2} = \begin{pmatrix} 7.810 & 2.775 & 0.003 & 0.004 & -0.009 & -0.012 \\ 2.775 & 2.495 & -0.002 & 0.001 & 0.005 & -0.004 \\ 0.003 & -0.002 & 7.788 & 2.770 & 0.007 & 0.004 \\ 0.004 & 0.001 & 2.770 & 2.494 & 0.000 & 0.001 \\ -0.009 & 0.005 & 0.007 & 0.000 & 7.832 & 2.784 \\ -0.012 & -0.004 & 0.004 & 0.001 & 2.784 & 2.504 \end{pmatrix},$$

and

$$(J_N(\theta))_{\gamma_1\gamma_2} = (J_N(\theta))_{\gamma_2\gamma_1}^{\top} = \begin{pmatrix} -5.487 & 0.162 & -0.007 & -0.003 & 0.019 & 0.010 \\ -3.349 & -0.888 & -0.003 & -0.003 & 0.004 & 0.006 \\ 0.003 & 0.004 & -5.465 & 0.168 & -0.010 & -0.003 \\ -0.001 & 0.002 & -3.342 & -0.885 & -0.005 & -0.004 \\ -0.010 & -0.013 & -0.004 & 0.000 & -5.500 & 0.165 \\ 0.007 & -0.002 & -0.001 & 0.002 & -3.362 & -0.893 \end{pmatrix},$$

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

,

and finally

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