PERGAMON Computers and Mathematics with Applications 40 (2000) 405–411

An International Journal
computers & mathematics with applications

www.elsevier.nl/locate/camwa

Comments on "Kalman-Filtering Methods for Computing Information Matrices for Time-Invariant, Periodic, and Generally Time-Varying VARMA Models and Samples"

J. TERCEIRO
Departamento de Economía Cuantitativa, Universidad Complutense de Madrid
Campus de Somosaguas, 28223 Madrid, Spain

(Received December 1994; revised and accepted July 1999)

Abstract—This paper discusses some results contained in a paper by Zadrozny and Mittnik [1] in relation to the work developed by Terceiro [2]. It also considers several issues concerning the efficiency of the algorithm and the state-space formulation of VARMA models.

Keywords—Exact information matrix, Kalman filter, VARMA models.

1. INTRODUCTION

This paper analyzes some aspects of the method for computing exact information matrices of VARMA models proposed by Zadrozny and Mittnik [1], in comparison with that of Terceiro [2].

One of the most frequent approaches to the computation of the information matrix, see [3, p. 140–142], consists of replacing the expression $\mathbf{\Gamma}_{ij}(t) = E[\partial_i \boldsymbol{\xi}(t) \partial_j \boldsymbol{\xi}(t)^\top]$ by the approximation $\mathbf{\Gamma}_{ij}(t) \simeq \partial_i \boldsymbol{\xi}(t) \partial_j \boldsymbol{\xi}(t)^\top$, where E is the expectation operator, $\boldsymbol{\xi}(t)$ the Kalman filter innovations, and $i, j = 1, \ldots, \alpha$ with $t = 1, \ldots, N$, α being the number of parameters to be estimated and N the sample size. This approximation is equivalent to using a particular realization of the random term instead of its expected value. Thus, it is valid only to the extent that its standard deviation is much smaller than its expected value. In this work, we will deal with the exact computational procedures for $\mathbf{\Gamma}_{ij}(t)$ proposed in [1] and [2].

Section 2 discusses some statements contained in [1] in view of the more general state-space formulation used in [2]. Special emphasis is laid on the computational efficiency of the proposed methods. The advantages of using the state-space formulation for VARMA models proposed by Terceiro, compared with the alternative of Zadrozny and Mittnik, are described in Section 3. This formulation is extended to include exogenous variables and observation errors. Section 4 concludes with remarks. From now on we will use, whenever possible, the notation of Zadrozny and Mittnik [1].

0898-1221/00/$ - see front matter Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
PII: S0898-1221(00)00168-1

2. EFFICIENCY CONSIDERATIONS

Zadrozny and Mittnik [1, p. 108] claim that "The present paper derives a more efficient version of the recursive Kalman-filtering method for computing sample and asymptotic information matrices for general VARMA models." This statement is debatable.

When Zadrozny and Mittnik compare the computational burden of their method with that of Terceiro, they should recognize that the formulation of the problem in [2, Chapters 2 and 3] is more general than theirs, see [1, p. 108–109]. Indeed, using the notation of Zadrozny and Mittnik, the formulation of Terceiro includes the terms that here will be denoted by $\mathbf{E}\boldsymbol{\nu}(t)$ and $\mathbf{L}\boldsymbol{\nu}(t)$, corresponding to the exogenous variables $\boldsymbol{\nu}(t)$ in equations (2.2) and (2.3) of [1], respectively. Note that improper state-space models, i.e., those that allow for a contemporaneous relationship between inputs and outputs, are common in economics, and in these formulations, the exogenous variables affect both the state and observation equations (see next section). Besides, Terceiro considers the possibility of errors in the variables. This implies that matrix $\mathbf{D}$ in the observation equation (2.3) does not have a simple structure of zeros and ones, but also depends on the model parameters, see [2, Chapters 3 and 6] and the next section.

Taking into account these extensions, expression (4.4) of [1] should be replaced by expression (4.17) of [2, p. 30]:

$$\partial_i \boldsymbol{\xi}(t) = -\mathbf{D}\partial_i \mathbf{x}(t \mid t-1) - \partial_i \mathbf{D}\mathbf{x}(t \mid t-1) - \partial_i \mathbf{L}\boldsymbol{\nu}(t), \tag{1}$$

and therefore,

$$\begin{aligned}\boldsymbol{\Gamma}_{ij}(t) &= E\left[\partial_i \boldsymbol{\xi}(t)\partial_j \boldsymbol{\xi}(t)^\top\right] \\ &= \mathbf{D}\mathbf{Z}_{ij}(t)\mathbf{D}^\top + \mathbf{D}\mathbf{U}_i(t)\partial_j\mathbf{D}^\top + \mathbf{D}E\left[\partial_i\mathbf{x}(t \mid t-1)\right]\boldsymbol{\nu}(t)^\top\partial_j\mathbf{L}^\top + \partial_i\mathbf{D}\mathbf{U}_j(t)^\top\mathbf{D}^\top \\ &\quad + \partial_i\mathbf{D}\mathbf{S}(t)\partial_j\mathbf{D}^\top + \partial_i\mathbf{D}E\left[\mathbf{x}(t \mid t-1)\right]\boldsymbol{\nu}(t)^\top\partial_j\mathbf{L}^\top + \partial_i\mathbf{L}\boldsymbol{\nu}(t)E\left[\partial_j\mathbf{x}(t \mid t-1)^\top\right]\mathbf{D}^\top \\ &\quad + \partial_i\mathbf{L}\boldsymbol{\nu}(t)E\left[\mathbf{x}(t \mid t-1)^\top\right]\partial_j\mathbf{D}^\top + \partial_i\mathbf{L}\boldsymbol{\nu}(t)\boldsymbol{\nu}(t)^\top\partial_j\mathbf{L}^\top,\end{aligned} \tag{2}$$

where $\mathbf{S}(t)$, $\mathbf{U}_i(t)$, and $\mathbf{Z}_{ij}(t)$ are defined in (4.9) of [1].

Also, under this general formulation [2, p. 87], equation (4.6) of [1] should be replaced by

$$\begin{aligned}\partial_i\mathbf{x}(t+1 \mid t) = {} & \left[\partial_i\mathbf{F} - \mathbf{K}(t)\partial_i\mathbf{D}\right]\mathbf{x}(t \mid t-1) + \boldsymbol{\Phi}(t)\partial_i\mathbf{x}(t \mid t-1) \\ & + \left[\partial_i\mathbf{E} - \mathbf{K}(t)\partial_i\mathbf{L}\right]\boldsymbol{\nu}(t) + \partial_i\mathbf{K}(t)\boldsymbol{\xi}(t),\end{aligned} \tag{3}$$

and it is easy to prove that the matrix of second-order moments, $\mathbf{W}^*_{ij}(t)$, of the augmented state vector $\mathbf{x}^*_i(t \mid t-1) = [\mathbf{x}(t \mid t-1)^\top,\ \partial_i\mathbf{x}(t \mid t-1)^\top]^\top$ can be written as

$$\begin{aligned}\mathbf{W}^*_{ij}(t+1) = {} & \mathbf{F}^*_i(t)\mathbf{W}^*_{ij}(t)\mathbf{F}^*_j(t)^\top + \mathbf{K}^*_i(t)\mathbf{M}(t)\mathbf{K}^*_j(t)^\top + \mathbf{A}^*_i(t)\boldsymbol{\nu}(t)E\left[\mathbf{x}^*_j(t \mid t-1)^\top\right]\mathbf{F}^*_j(t)^\top \\ & + \mathbf{A}^*_i(t)\boldsymbol{\nu}(t)\boldsymbol{\nu}(t)^\top\mathbf{A}^*_j(t)^\top + \mathbf{F}^*_i(t)E[\mathbf{x}^*_i(t \mid t-1)]\boldsymbol{\nu}(t)^\top\mathbf{A}^*_j(t)^\top,\end{aligned} \tag{4}$$

where

$$\mathbf{F}^*_i(t) = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \partial_i\mathbf{F} - \mathbf{K}(t)\partial_i\mathbf{D} & \boldsymbol{\Phi}(t)\end{bmatrix}, \qquad \mathbf{A}^*_i(t) = \begin{bmatrix} \mathbf{E} \\ \partial_i\mathbf{E} - \mathbf{K}(t)\partial_i\mathbf{L}\end{bmatrix}, \tag{5}$$

and $\mathbf{K}^*_i(t)$ coincides with the expression given in [1, p. 113].

The equations corresponding to (2) and (4) from [1, p. 113] are

$$\boldsymbol{\Gamma}_{ij}(t) = \mathbf{D}\mathbf{Z}_{ij}\mathbf{D}^\top, \tag{6}$$

$$\mathbf{W}^*_{ij}(t+1) = \mathbf{F}^*_i(t)\mathbf{W}^*_{ij}(t)\mathbf{F}^*_j(t)^\top + \mathbf{K}^*_i(t)\mathbf{M}(t)\mathbf{K}^*_j(t)^\top, \qquad \text{with } \partial_i\mathbf{D} = \mathbf{0}. \tag{7}$$

That is, when there are no exogenous variables ($\mathbf{A}^*_i(t) = 0$) and no observation errors ($\partial_i\mathbf{D} = 0$), expressions (2) and (4) coincide with (6) and (7). However, the comparison of (2) and (4) with (6) and (7) suggests the following considerations.

First, although it is an unimportant issue and contrary to what is stated in [1, p. 113], the inclusion of exogenous variables makes it convenient to decompose $\mathbf{\Gamma}_{ij}(t)$ as $\mathbf{C}^*_{ij}(t)+\overline{\partial_i\boldsymbol{\xi}}(t)\overline{\partial_j\boldsymbol{\xi}}(t)^\top$, as is done in [2, p. 34], where $\overline{\partial_i\boldsymbol{\xi}}(t) = E[\partial_j\boldsymbol{\xi}(t)]$. This makes necessary the propagation of

$$\bar{\mathbf{x}}(t+1) = \mathbf{F}\bar{\mathbf{x}}(t) + \mathbf{E}\boldsymbol{\nu}(t),$$
$$\overline{\partial_i\mathbf{x}}\,(t+1 \mid t) = \mathbf{\Phi}(t)\overline{\partial_i\mathbf{x}}\,(t \mid t-1) + [\partial_i\mathbf{F} - \mathbf{K}(t)\partial_i\mathbf{D}]\,\bar{\mathbf{x}}(t) + [\partial_i\mathbf{E} - \mathbf{K}(t)\partial_i\mathbf{L}]\,\boldsymbol{\nu}(t)$$

to compute

$$\overline{\partial_i\boldsymbol{\xi}}(t) = \partial_i\mathbf{D}\bar{\mathbf{x}}(t) + \mathbf{D}\overline{\partial_i\mathbf{x}}(t \mid t-1) + \partial_i\mathbf{L}\boldsymbol{\nu}(t).$$

The first equation is propagated just once, despite the number of parameters. The other two are propagated once for each parameter. This decomposition, therefore, enhances the computation speed. In fact, the calculation of $\mathbf{C}^*_{ij}(t)$ has a lower computational burden than the direct calculation of $\mathbf{\Gamma}_{ij}(t)$, given that the first of these expressions is simpler, because it is a moment centered on the mean value, so the exogenous variables can be canceled. If this decomposition is not done, one has to run $\alpha(\alpha+1)/2$ propagations of the means instead of α.

Second, to avoid the complexity of expressions (2) and (4), it is convenient to define a linear system output by means of the $2m$ vector given by $[\partial_i\boldsymbol{\xi}(t)^\top,\, \partial_j\boldsymbol{\xi}(t)^\top]^\top$. To do that, taking into account (1) and (3), Terceiro [2, p. 87] defines the $3n$ augmented state vector

$$\mathbf{x}^*(t \mid t-1) = \left[\mathbf{x}(t \mid t-1)^\top,\ \partial_i\mathbf{x}(t \mid t-1)^\top,\ \partial_j\mathbf{x}(t \mid t-1)^\top\right]^\top.$$

With this definition, expressions (2) and (4) have the same structure as (6) and (7) and coincide with (D.11) and (D.10) of [2, p. 88]. Also, the matrices analogous to $\mathbf{\Gamma}_{ij}(t)$ and $\mathbf{W}^*_{ij}(t)$ are $\mathbf{B}^{ij}_t$ and $\mathbf{P}^c_t$ from [2, p. 85–89].

However, as Terceiro [2, p. 32] points out explicitly, it is possible to take advantage of the special structure of the system matrices. There is no need to multiply terms that are known *a priori* to be $\mathbf{0}$. In fact, from the structure of the observation matrix $\mathbf{H}^c$ in [2, p. 88], corresponding to $\mathbf{D}$ from [1], it is clear that the computation of $\mathbf{B}^{ij}_t$ in [2, p. 88], corresponding to $\mathbf{\Gamma}_{ij}(t)$ in [1], only requires the propagation of blocks (1.1), (1.3), (2.1), and (2.3) of the $3n\times 3n$-dimension matrix $\mathbf{P}^c_t$ from [2, p. 88], with the starting value $\mathbf{P}^c_1 = \mathbf{0}$. Thus, the computational burden is equivalent to the propagation of $4n^2$ equations or to a $2n$ system like that proposed by Zadrozny and Mittnik. It is easy to see that if errors in variables and exogenous variables are not considered, the blocks (1.1), (1.3), (2.1), and (2.3) of $\mathbf{P}^c_t$ from [2] coincide with the matrices $\mathbf{S}(t)$, $\mathbf{U}_j(t)^\top$, $\mathbf{U}_i(t)$, and $\mathbf{Z}_{ij}(t)$ defined in (4.9) from [1]. In particular, the expressions obtained developing the blocks (2.1) and (2.3) of $\mathbf{P}^c_t$ in (D.10) of [2, p. 88] coincide with the expressions (4.12) and (4.13) of [1].

Third, in nonstationary models, which are considered in [2], it is not possible to use the expression $\mathbf{S}(t) = \mathbf{C}-\mathbf{V}(t)$ to avoid the propagation of the (1.1) block of $\mathbf{W}^*_{ij}(t)$ in equation (4.11) of [1], which corresponds to the block (1.1) of equation (D.10) in [2]. In this situation, $\mathbf{C}$ cannot be previously computed as the initial value of the matrix to be propagated by means of the Kalman filter, as is done in (3.8) of [1]. When the model is nonstationary, the computation of initial conditions is done by other methods, see [3, Chapter 3; 4] and the references therein. Terceiro [2, Appendix E] proposes a simple procedure to compute the profile likelihood that provides satisfactory results when the model is far from noninvertibility. Apart from that, it should be noted that avoiding the propagation of the block (1.1) in expression (4.11) of [1] in stationary models does not provide substantial gains because, unlike $\mathbf{U}_i(t)$ and $\mathbf{Z}_{ij}(t)$ matrices, this matrix is symmetric and has to be propagated only once, independently of the number of parameters. Therefore, the propagation for $t = 1,\dots,N$ of $[n(n+1)/2 + n^2\alpha + n^2\alpha(\alpha+1)/2]$ equations is strictly required, by the methods of both Zadrozny and Mittnik and Terceiro.

Note too, that in stationary models the general expression $\mathbf{S}(t) = \mathbf{C}(t) - \mathbf{V}(t)$ can only be reduced to $\mathbf{C}(t) = \mathbf{C}$ under certain assumptions on the initial conditions and when the model does not include exogenous variables. These assumptions are not made in the general framework considered by Terceiro. Besides, from the point of view of computational efficiency, and specifically, the algorithm's reliability and accuracy, it is not sensible to obtain the matrix $\mathbf{S}(t)$ as a difference of two positive definite, matrices $\mathbf{C}$ and $\mathbf{V}(t)$, as is done in [1 p. 113].

In summary, the Zadrozny-Mittnik procedure for computing exact information matrices of VARMA models is not more efficient, but just a particular case of the method developed by Terceiro. None of the three reasons given in [1, p. 113] for the inefficiency in the calculation of $\mathbf{\Gamma}_{ij}(t)$ proposed in [2] are valid, given the more general state-space formulation used.

3. STATE-SPACE REPRESENTATION OF VARMA MODELS

If such great stress is laid on computational efficiency as in [1], emphasizing other relevant aspects of the state-space formulation of the VARMA model would be reasonable, especially when they affect not only the efficiency of the whole estimation process but also its convergence properties and the computation of the exact information matrix.

Using the notation of Zadrozny and Mittnik, consider the m-dimensional stochastic process $\mathbf{u}(t)$, generated by the VARMA model

$$\mathbf{A}(L)\mathbf{u}(t) = \mathbf{B}(L)\mathbf{e}(t),$$
$$\mathbf{A}(L) = \mathbf{I}_m - \sum_{i=0}^{p} \mathbf{A}_i L^i, \qquad \mathbf{B}(L) = \sum_{i=0}^{q} \mathbf{B}_i L^i, \qquad \mathbf{e}(t) \sim \mathrm{NID}(\mathbf{0}, \mathbf{I}),$$

defining $\mathbf{e}^*(t) = \mathbf{B}_0\mathbf{e}(t)$, and following [2, Chapter 2], the state-space equivalent formulation is

$$\mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{G}\mathbf{e}^*(t), \tag{8}$$
$$\mathbf{u}(t) = \mathbf{D}\mathbf{x}(t) + \mathbf{e}^*(t), \tag{9}$$

where

$$F = \begin{bmatrix} \mathbf{A}_1 & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{k-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ \mathbf{A}_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{A}_1 + \mathbf{B}_1\mathbf{B}_0^{-1} \\ \mathbf{A}_2 + \mathbf{B}_2\mathbf{B}_0^{-1} \\ \vdots \\ \mathbf{A}_{k-1} + \mathbf{B}_{k-1}\mathbf{B}_0^{-1} \\ \mathbf{A}_k + \mathbf{B}_k\mathbf{B}_0^{-1} \end{bmatrix}, \quad \mathbf{D} = [\mathbf{I} \;\; \mathbf{0} \;\; \dots \;\; \mathbf{0}],$$

with $\mathbf{e}^*(t) \sim \mathrm{NID}(\mathbf{0}, \mathbf{B}_0\mathbf{B}_0^{\top})$. The dimension of the state vector $\mathbf{x}(t)$ is $n = k \times m$, with $k = \max(p, q)$.

Therefore, when $p < q + 1$, the formulation in [1, p. 108–109] yields a state vector whose dimension is greater than the one defined above. Note that in (8),(9), the system error and the measurement error are correlated, so that the Kalman filter equations would need to be modified to account for this, see [3, p. 112–113; 2, Appendix A]. This form also guarantees the nonsingularity of $\mathbf{M}(t)$ even with zero initial conditions, $\mathbf{V}(1) = \mathbf{0}$.

The structure of $\mathbf{F}$ and $\mathbf{D}$ defined in (8),(9) corresponds to the so-called canonical observable form, see [5, Chapter 6], where the transition matrix $\mathbf{F}$ is left companion, so that the nonzero eigenvalues, $\lambda_i(\mathbf{F}) \neq 0$, are the reciprocal of the roots of $|\mathbf{A}(L)| = 0$. This representation is minimal among all those having a transition matrix in $(m \times m)$ block companion form. However, strict minimality is not assured in general because system (8),(9) is observable, but not controllable.

It is convenient to discuss briefly the convergence properties of the Kalman filter corresponding to (8),(9), which is given by

$$\mathbf{x}(t+1 \mid t) = \mathbf{F}\mathbf{x}(t \mid t) + \mathbf{K}(t)\boldsymbol{\xi}(t), \tag{10}$$
$$\mathbf{u}(t) = \mathbf{D}\mathbf{x}(t \mid t) + \boldsymbol{\xi}(t), \tag{11}$$

where the filter gain, $\mathbf{K}(t)$, is computed by means of the recursions

$$\mathbf{M}(t) = \mathbf{D}\mathbf{V}(t)\mathbf{D}^\top + \mathbf{B}_0\mathbf{B}_0^\top, \tag{12}$$

$$\mathbf{K}(t) = \left[\mathbf{F}\mathbf{V}(t)\mathbf{D}^\top + \mathbf{G}\mathbf{B}_0\mathbf{B}_0^\top\right]\mathbf{M}(t)^{-1}, \tag{13}$$

$$\mathbf{V}(t+1) = \mathbf{F}\mathbf{V}(t)\mathbf{F}^\top + \mathbf{G}\mathbf{B}_0\mathbf{B}_0^\top\mathbf{G}^\top - \mathbf{K}(t)\mathbf{M}(t)\mathbf{K}(t)^\top. \tag{14}$$

Equation (14) can also be written as

$$\mathbf{V}(t+1) = \mathbf{\Phi}(t)\mathbf{V}(t)\mathbf{\Phi}(t)^\top + [\mathbf{G} - \mathbf{K}(t)]\begin{bmatrix}\mathbf{B}_0\mathbf{B}_0^\top & \mathbf{B}_0\mathbf{B}_0^\top \\ \mathbf{B}_0\mathbf{B}_0^\top & \mathbf{B}_0\mathbf{B}_0^\top\end{bmatrix}\begin{bmatrix}\mathbf{G}^\top \\ -\mathbf{K}(t)^\top\end{bmatrix}, \tag{14'}$$

with $\mathbf{\Phi}(t) = \mathbf{F} - \mathbf{K}(t)\mathbf{D}$.

The matrix $\mathbf{V}(t+1)$ in (14′) is the result of the sum of two positive semidefinite matrices. Consequently, numerical computations based on (14′) will be better conditioned than those based on (14). These equations are equivalent to (3.2)–(3.7) in [1]. Note that the optional vector of observation errors, $\boldsymbol{\zeta}(t)$ in [1, p. 109], is not considered because it is not present in the original VARMA formulation. From (13),(14), it is easy to show that

$$\mathbf{V}(t+1) = \bar{\mathbf{\Phi}}\mathbf{V}(t)\bar{\mathbf{\Phi}}^\top - \bar{\mathbf{\Phi}}\mathbf{V}(t)\mathbf{D}^\top\left[\mathbf{D}\mathbf{V}(t)\mathbf{D}^\top + \mathbf{B}_0\mathbf{B}_0^\top\right]^{-1}\mathbf{D}\mathbf{V}(t)\bar{\mathbf{\Phi}}^\top, \tag{15}$$

with $\bar{\mathbf{\Phi}} = \mathbf{F} - \mathbf{G}\mathbf{D}$.

If $\mathbf{V}(t)$ converges to an equilibrium solution $\mathbf{V}$, it must satisfy the algebraic Riccati equation corresponding to (15), which is

$$\mathbf{V} = \bar{\mathbf{\Phi}}\mathbf{V}\bar{\mathbf{\Phi}}^\top - \bar{\mathbf{\Phi}}\mathbf{V}\mathbf{D}^\top\left[\mathbf{D}\mathbf{V}\mathbf{D}^\top + \mathbf{B}_0\mathbf{B}_0^\top\right]^{-1}\mathbf{D}\mathbf{V}\bar{\mathbf{\Phi}}^\top, \tag{16}$$

where $\mathbf{V} = \mathbf{0}$ is an obvious solution.

Particularizing the results of [6, Theorems 4.1 and 4.2] to the formulation (8),(9), it can be stated, subjected to $\mathbf{V}(1) \geq 0$ that $\lim \mathbf{V}(t) = \mathbf{0}$ as $t \to \infty$ if and only if the pair $(\mathbf{D}, \mathbf{F})$ is detectable. When $|\lambda_i(\bar{\mathbf{\Phi}})| \leq 1$, this solution, $\mathbf{V} = \mathbf{0}$, is called the strong solution of (14), and if $|\lambda_i(\bar{\mathbf{\Phi}})| < 1$, it is called the stabilizing solution. In this last case, the convergence to $\mathbf{V} = \mathbf{0}$ is exponential.

It is obvious that the definition of $\mathbf{D}$ and $\mathbf{F}$ given in (8),(9) assures observability, and therefore, the weaker requirement of detectability. Besides, $\mathbf{V}(1) \geq 0$ is satisfied not only for the solution of the Lyapunov equation (3.8) in [1], but also for the more general initialization criteria of (14) in nonstationary situations, proposed in [2; 3, Chapter 3; 4].

From definitions in (8),(9), it is easy to see that $(\mathbf{F} - \mathbf{G}\mathbf{D})$ is a left companion matrix, whose nonzero eigenvalues coincide with the inverse of the roots of $|\mathbf{B}(L)| = 0$. Therefore, invertibility of the VARMA model assures the exponential convergence of (14) to $\mathbf{V} = \mathbf{0}$.

Under these conditions, equations (12)–(14) converge to

$$\mathbf{M} = \mathbf{B}_0\mathbf{B}_0^\top, \tag{17}$$

$$\mathbf{K} = \mathbf{G}, \tag{18}$$

$$\mathbf{V} = \mathbf{0}, \tag{19}$$

and model (8),(9) coincides with the steady-state innovations model given by

$$\mathbf{x}(t+1 \mid t) = \mathbf{F}\mathbf{x}(t \mid t-1) + \mathbf{K}\boldsymbol{\xi}(t), \tag{20}$$

$$\mathbf{u}(t) = \mathbf{D}\mathbf{x}(t \mid t-1) + \boldsymbol{\xi}(t), \tag{21}$$

with $\boldsymbol{\xi}(t) = \mathbf{B}_0\mathbf{e}(t)$.

Therefore, the state estimate converges to $\mathbf{x}(t)$ and the Kalman filter can estimate the states with a null error.

The simplicity of (17)–(19) has obvious implications for the maximum likelihood estimation of VARMA models. When using the state-space formulation of Zadrozny and Mittnik, the convergence values of the Kalman filter do not have such a simple and explicit interpretation in terms of the original VARMA model parameters. This is due to the fact that

(a) the state vector does not converge to its estimate, and
(b) the formulation includes the optional vector $\boldsymbol{\zeta}(t)$.

The VARMA formulation can be extended to include a vector, $\boldsymbol{\nu}(t)$, of exogenous variables such that

$$\mathbf{A}(L)\mathbf{u}(t) = \mathbf{H}(L)\boldsymbol{\nu}(t) + \mathbf{B}(L)\mathbf{e}(t), \tag{22}$$

with $\mathbf{H}(L) = \sum_{i=0}^{s} \mathbf{H}_i L^i$.

In this case, the corresponding state-space formulation proposed by Terceiro 2, Chapter 2] is

$$\mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{E}\boldsymbol{\nu}(t) + \mathbf{G}\mathbf{e}^*(t), \tag{23}$$

$$\mathbf{u}(t) = \mathbf{D}\mathbf{x}(t) + \mathbf{L}\boldsymbol{\nu}(t) + \mathbf{e}^*(t), \tag{24}$$

with

$$\mathbf{E} = \begin{bmatrix} \mathbf{A}_1\mathbf{H}_0 + \mathbf{H}_1 \\ \mathbf{A}_2\mathbf{H}_0 + \mathbf{H}_2 \\ \vdots \\ \mathbf{A}_{k-1}\mathbf{H}_0 + \mathbf{H}_{k-1} \\ \mathbf{A}_k\mathbf{H}_0 + \mathbf{H}_k \end{bmatrix}, \qquad \mathbf{L} = \mathbf{H}_0, \tag{25}$$

and $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{D}$ are defined as in (8),(9).

The dimension of the state vector in (23),(24) is $n = k \times m$, with $k = \max(p, q, s)$. As we discussed in the previous section, the inclusion of the terms $\mathbf{E}\boldsymbol{\nu}(t)$ and $\mathbf{L}\boldsymbol{\nu}(t)$ in (23),(24) yields expressions (2) and (4), which generalize those of Zadrozny and Mittnik. The vector of exogenous variables $\boldsymbol{\nu}(t)$ can contain both stochastic and deterministic variables. Examples with deterministic exogenous variables include the representation of shock effects and the modelling of seasonal variations of $\mathbf{u}(t)$ by periodic functions of time, among others.

It is also possible to extend the formulation (23),(24) to allow for observation errors. Following [2, Chapter 3], we have

$$\mathbf{u}^*(t) = \mathbf{u}(t) + \boldsymbol{\varepsilon}_{\mathbf{u}}(t),$$

$$\boldsymbol{\nu}^*(t) = \boldsymbol{\nu}(t) + \boldsymbol{\varepsilon}_{\boldsymbol{\nu}}(t),$$

where $\boldsymbol{\varepsilon}_{\mathbf{u}}(t)$ and $\boldsymbol{\varepsilon}_{\boldsymbol{\nu}}(t)$ are the observation errors affecting $\mathbf{u}(t)$ and $\boldsymbol{\nu}(t)$, respectively.

If it is assumed that the vector of stochastic exogenous variables is generated by the VARMA model,

$$\bar{\mathbf{A}}(L)\boldsymbol{\nu}(t) = \overline{\mathbf{B}}(L)\mathbf{e}_{\boldsymbol{\nu}}(t)$$

can be written in state-space form as

$$\mathbf{x}_{\boldsymbol{\nu}}(t+1) = \overline{\mathbf{F}}\mathbf{x}_{\boldsymbol{\nu}}(t) + \overline{\mathbf{G}}\mathbf{e}^*_{\boldsymbol{\nu}}(t), \tag{26}$$

$$\boldsymbol{\nu}(t) = \overline{\mathbf{D}}\mathbf{x}_{\boldsymbol{\nu}}(t) + \mathbf{e}^*_{\boldsymbol{\nu}}(t), \tag{27}$$

with $\overline{\mathbf{F}}$, $\overline{\mathbf{G}}$, $\overline{\mathbf{D}}$, and $\mathbf{e}^*_{\boldsymbol{\nu}}(t)$ defined as $\mathbf{F}$, $\mathbf{G}$, $\mathbf{D}$, and $\mathbf{e}^*(t)$ in (8),(9).

Combining (23),(24) with (26),(27) gives

$$\begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{x}_{\boldsymbol{\nu}}(t+1) \end{bmatrix} = \begin{bmatrix} \mathbf{F} & \mathbf{L}\overline{\mathbf{D}} \\ \mathbf{0} & \overline{\mathbf{F}} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}_{\boldsymbol{\nu}}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{E}\mathbf{e}^*_{\boldsymbol{\nu}}(t) + \mathbf{G}\mathbf{e}^*(t) \\ \overline{\mathbf{G}}\mathbf{e}^*_{\boldsymbol{\nu}}(t) \end{bmatrix}, \tag{28}$$

$$\begin{bmatrix} \mathbf{u}^*(t) \\ \boldsymbol{\nu}^*(t) \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{L}\overline{\mathbf{D}} \\ \mathbf{0} & \overline{\mathbf{D}} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}_{\boldsymbol{\nu}}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{L}\mathbf{e}^*_{\boldsymbol{\nu}}(t) + \mathbf{e}^*(t) + \boldsymbol{\varepsilon}_{\mathbf{u}}(t) \\ \mathbf{e}^*_{\boldsymbol{\nu}}(t) + \boldsymbol{\varepsilon}_{\boldsymbol{\nu}}(t) \end{bmatrix}, \tag{29}$$

which is the state-space formulation of a VARMA model with errors in both exogenous and endogenous variables. Note that the observation matrix $\mathbf{D}$ depends in general on the parameters to be estimated. Consequently, expressions (2) and (4) allow for $\partial_i \mathbf{D} \neq \mathbf{0}$.

When some exogenous variables are observed without error, including them in the observations vector is not convenient, because it unnecessarily augments its dimension and the Kalman filter requires the inversion of matrices whose order is precisely the dimension of this vector. Therefore, a VARMA model with observation errors in some exogenous variables can be written as

$$\mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{E}\nu(t) + \mathbf{G}\mathbf{w}(t), \tag{30}$$

$$\mathbf{u}(t) = \mathbf{D}\mathbf{x}(t) + \mathbf{L}\nu(t) + \mathbf{C}\varepsilon(t), \tag{31}$$

where the perturbations are such that

$$E\left[\mathbf{w}(t)\right] = \mathbf{0}, \qquad E\left[\varepsilon(t)\right] = \mathbf{0},$$

$$E\left\{ \begin{bmatrix} \mathbf{w}(t_1) \\ \varepsilon(t_1) \end{bmatrix} \begin{bmatrix} \mathbf{w}(t_2)^\top & \varepsilon(t_2)^\top \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^\top & \mathbf{R} \end{bmatrix} \delta_{t_1 t_2},$$

with

$$\delta_{t_1 t_2} = \begin{cases} 1, & t_1 \neq t_2, \\ 0, & t_1 = t_2. \end{cases}$$

Note that equations (30),(31) are not in the canonical observable form defined in (8),(9) and, particularly, the matrix $\mathbf{D}$ depends on the parameters to be estimated. This is the general formulation used in [2] for the computation of the exact information matrix.

4. CONCLUSIONS

The Zadrozny and Mittnik results on the computation of the exact information matrix are a special case of those obtained in [2]. Both coincide when the model is stationary and has no exogenous variables and no observation errors. In fact, taking into account the structure of a certain matrix, they are computationally equivalent.

The state-space formulation used in [2] has some advantages over the alternative used by Zadrozny and Mittnik [1] and other authors, see [3; 7, Chapter 13]. It assures that convergence values of the Kalman filter have a transparent interpretation in terms of the VARMA model parameters and, in some cases, has a smaller dimension for the state vector, thus reducing computational burden and simplifying the estimation of general initial conditions. Although not relevant in this context, advocated state-space form also has convenient properties in problems of smoothing and signal extraction.

REFERENCES

1. P.A. Zadrozny and S. Mittnik, Kalman-filtering methods for computing information matrices for time-invariant, periodic, and generally time-varying VARMA models and samples, *Computers Math. Applic.* **28** (4), 107–119, (1994).
2. J. Terceiro Lomba, *Estimation of Dynamic Econometric Models with Errors in Variables*, Lecture Notes in Economics and Mathematical Systems, N. 339, Springer-Verlag, New York, (1990).
3. A.C. Harvey, *Forecasting Structural Time Series and Kalman Filter*, Cambridge University Press, Cambridge, (1989).
4. P. Marshall, State space models with diffuse initial conditions, *Journal of Time Series Analysis* **13** (5), 411–414, (1992).
5. T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, (1980).
6. C.E. de Souza, M.R. Gevers and G.C. Goodwin, Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices, *IEEE Transactions on Automatic Control* **31** (9), 831–838, (1986).
7. J.D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, (1994).