Pergamon

Computers Math. Applic. Vol. 28, No. 4, pp. 107–119, 1994

0898-1221/94 $7.00 + 0.00

0898-1221(94)00132-4

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

P. A. ZADROZNY
GTE Laboratories
40 Sylvan Road, Waltham, MA 02254, U.S.A.
and
S. MITTNIK
Institute of Statistics and Econometrics
Christian-Albrechts-Universität zu Kiel
Olshausenstr. 40, D-24098 Kiel, Germany

(Received April 1993; accepted May 1993)

Abstract—Under general conditions, the inverse sample information matrix can be used to establish a Cramer-Rao lower bound of the covariance matrix of parameter estimates of a model, and the inverse asymptotic information matrix is the asymptotic covariance matrix of the parameter estimates. The paper does two things. First, it derives a recursive Kalman-filtering method for computing exact sample and asymptotic information matrices for time-invariant, periodic, or generally time-varying Gaussian vector autoregressive moving-average (VARMA) models and samples. Second, it specializes the recursive method to a nonrecursive method for computing exact asymptotic information matrices for time-invariant or periodic VARMA models and samples

Keywords—Periodic VARMA models, Information matrix computation, Kalman filtering.

1. INTRODUCTION

Under general conditions, maximum likelihood parameter estimates of Gaussian, stationary, and invertible, vector autoregressive moving-average (VARMA) models are known to be consistent and asymptotically normally distributed, with asymptotic covariance matrix equal to the inverse asymptotic information matrix [1,2, Chapter 7]. Until recently, the asymptotic information matrix of the general Gaussian VARMA model has been stated only in terms of formulas involving integration over the frequency domain [2, pp. 428–430, 441–442]. More concise time-domain formulae, which do not require frequency-domain integration, have been derived only for univariate ARMA models [3, pp. 240–242]. Because the frequency-domain integrals can generally be computed only numerically, the information matrix has usually been approximated with analytical (if possible) or numerical derivatives of the likelihood function [4].

Recently, Zadrozny [5,6] derived a recursive Kalman-filtering method for computing exact sample and asymptotic information matrices of Gaussian VARMA models. The method involves forming the Gaussian likelihood function with the Kalman filter and recursively computing expected values of second-partial derivatives of the likelihood function. Iterating over the sample with the method produces the sample information matrix. Iterating until convergence produces

the asymptotic information matrix. The sample information matrix is of interest because it can be used to establish a Cramer-Rao lower bound of the sample covariance matrix of parameter estimates [7, p. 128]. More recently, Terceiro Lomba [8] stated a computationally less efficient version of this method.

The present paper derives a more efficient version of the recursive Kalman-filtering method for computing sample and asymptotic information matrices for general Gaussian VARMA models. The method is applicable to time-invariant, periodic, or generally time-varying models and samples. If the model and sample are time invariant or periodic and some common system-theoretic conditions hold, then the recursive computational equations and the matrices propagated by them converge to steady-state values or to periodic cycles. (Time invariant, time varying, and periodic models and samples are precisely defined at the end of Section 2.) If the model and sample are time invariant or periodic and the goal is to compute the asymptotic information matrix, then it is generally more efficient to directly solve the steady-state equations. Periodic models have recently begun to gain attention [9], in particular, as models of seasonal fluctuations [10]. The sample is periodic if, for example, different variables are observed at different frequencies [11]. The paper also extends the recursive method to a nonrecursive method for computing asymptotic information matrices for time-invariant or periodic models and samples.

To our knowledge, the only other similar results for computing exact sample or asymptotic information matrices are by Porat and Friedlander [7]. They describe a method, based on Levinson-Durbin filtering, for recursively computing the exact sample information matrix for univariate stationary linear dynamic models with possible exogenous variables. As examples, they work out details for some univariate ARMA models. Their method does not immediately extend to multivariate models. In effect, by replacing Levinson-Durbin filtering with Kalman filtering, the present method extends the method of Porat and Friedlander to multivariate models.

The paper proceeds as follows. Section 2 introduces the model and rewrites it in a state-space representation. In part as a means of establishing notation, Section 3 reviews the computation of the Gaussian likelihood function with the Kalman filter. Section 4 derives the recursive method. Section 5 derives the nonrecursive method for time-invariant models and samples. Section 6 extends the nonrecursive method to periodic models and samples. Section 7 concludes with remarks.

2. STATE-SPACE REPRESENTATION OF AN ARMA MODEL

Let $\mathbf{u}(t)$ be an $m \times 1$ vector generated by the stationary and invertible ARMA(p, q) model

$$\mathbf{A}(\mathbf{L})\ \mathbf{u}(t) = \mathbf{B}(\mathbf{L})\ \mathbf{e}(t) \qquad \text{for } t = 1, \ldots, N, \tag{2.1}$$

where $\mathbf{A}(\mathbf{L}) = \mathbf{I}_m - \sum_{k=1}^{p} \mathbf{A}_k\, \mathbf{L}^k$, $\mathbf{B}(\mathbf{L}) = \sum_{k=0}^{q} \mathbf{B}_k\, \mathbf{L}^k$, $\mathbf{L}$ is the lag operator, $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{e}(t)$ is an $m \times 1$ Gaussian white-noise disturbance vector with zero mean and constant covariance matrix. The distributional assumptions on $\mathbf{e}(t)$ are denoted by $\mathbf{e}(t) \sim \text{NIID}[\mathbf{0},\ \mathbf{\Sigma}_e]$, where $\mathbf{\Sigma}_e = E\ \mathbf{e}(t)\ \mathbf{e}(t)^\top$ (the superscript $\top$ denotes transposition). The model is, respectively, stationary and invertible if the characteristic equations $|\mathbf{A}(z)| = 0$ and $|\mathbf{B}(z)| = 0$ have all roots outside the unit circle. There is redundancy (an identification problem) between $\mathbf{B}_0$ and $\mathbf{\Sigma}_e$. Accordingly, we adopt the normalized parameterization, $\mathbf{B}_0$ = lower triangular and $\mathbf{\Sigma}_e = \mathbf{I}_m$. This parameterization is convenient for maximum likelihood estimation because, when all elements on the principal diagonal of $\mathbf{B}_0$ are nonzero, the model implies a nonsingular probability distribution for the observations.

Following Ansley and Kohn (see [12]), let $\mathbf{x}(t)$ be the $n \times 1$, state vector of the form $\mathbf{x}(t) = [\mathbf{x}_1(t)^\top, \ldots, \mathbf{x}_r(t)^\top]^\top$, where $r = \max(p, q+1)$ and $\mathbf{x}_k(t)$ for $k = 1, \ldots, r$, is $m \times 1$, so that $n = mr$. Let the law of motion of $\mathbf{x}(t)$ be

$$\mathbf{x}(t) = \mathbf{F}\ \mathbf{x}(t-1) + \mathbf{G}\ \mathbf{e}(t), \tag{2.2}$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \mathbf{0} & \dots & \mathbf{I} \\ \mathbf{A}_r & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{B}_0 \\ \vdots \\ \vdots \\ \mathbf{B}_{r-1} \end{bmatrix},$$

$\mathbf{A}_k = \mathbf{0}$ for $k > p$, $\mathbf{B}_k = \mathbf{0}$ for $k > q$, the zero and identity matrices in $\mathbf{F}$ are all $m \times m$, and $\mathbf{e}(t)$ is the $m \times 1$ disturbance vector from (2.1).

Associated with the state equation (2.2) is the observation equation

$$\mathbf{y}(t) = \mathbf{D}\,\mathbf{x}(t) + \boldsymbol{\zeta}(t), \tag{2.3}$$

where $\mathbf{D} = [\mathbf{I}, \mathbf{0}, \dots, \mathbf{0}]$ is the $m \times n$ selection matrix which picks $\mathbf{x}_1(t)$ out of $\mathbf{x}(t)$, and $\boldsymbol{\zeta}(t)$ is an optional vector of observation errors distributed as $\boldsymbol{\zeta}(t) \sim \text{NIID}[\mathbf{0}, \boldsymbol{\Sigma}_\zeta]$, where $\boldsymbol{\Sigma}_\zeta \geq \mathbf{0}$ (positive semidefinite). It is convenient to reparameterize $\boldsymbol{\Sigma}_\zeta$ to $\mathbf{R}$, where $\mathbf{R}$ is lower triangular and satisfies $\boldsymbol{\Sigma}_\zeta = \mathbf{R}\,\mathbf{R}^\top$, and, thereby, automatically impose $\boldsymbol{\Sigma}_\zeta \geq \mathbf{0}$. Observation errors are an integral part of the Kalman filter and specifying even a small amount of observation error can significantly improve the numerical stability of the computations [13].

As stated, the model, (2.2), and the sample, as characterized by the sampling scheme (2.3), are both time invariant. If the model and sample are time varying, then the system matrices in (2.2) and (2.3) vary over the sample and are written as $\mathbf{F}(t)$, $\mathbf{G}(t)$, $\mathbf{D}(t)$, and $\boldsymbol{\Sigma}_\zeta(t)$. If the model and sample are periodic, then, these matrices vary periodically with t, with some overall periodicity ρ.

3. COMPUTATION OF THE LIKELIHOOD FUNCTION WITH THE KALMAN FILTER

Let

$$\boldsymbol{\theta} = [\text{vec}(\mathbf{A}_1)^\top, \dots, \text{vec}(\mathbf{A}_p)^\top, \text{vech}(\mathbf{B}_0)^\top, \text{vec}(\mathbf{B}_1)^\top, \dots, \text{vec}(\mathbf{B}_q)^\top]^\top$$

be the $\alpha \times 1$ vector which collects the unnormalized parameters of (2.1), where $\text{vec}(\cdot)$ vectorizes a matrix columnwise and $\text{vech}(\cdot)$ vectorizes the lower-triangular part of a matrix columnwise (including elements on the principal diagonal). $\boldsymbol{\Sigma}_\zeta$ (or its reparameterization $\mathbf{R}$) is not included in $\boldsymbol{\theta}$ because $\mathbf{B}_0$ and $\boldsymbol{\Sigma}_\zeta$ are not both identifiable. We assume that $\boldsymbol{\Sigma}_\zeta$ is set according to relative precision of data and that $\mathbf{B}_0$ is estimated.

Let $Y(t) = \{\mathbf{y}(\tau)\}_{\tau=1}^{t}$, such that $Y(N) = \{\mathbf{y}(t)\}_{t=1}^{N}$ is the full sample of observations. Then $\boldsymbol{\xi}(t) = \mathbf{y}(t) - E[\mathbf{y}(t) \mid y(t-1)]$ is the innovation of $\mathbf{y}(t)$ and $\mathbf{M}(t) = E\,\boldsymbol{\xi}(t)\,\boldsymbol{\xi}(t)^\top$ is its covariance matrix. Ignoring terms which are independent of $\boldsymbol{\theta}$, -2 times the Gaussian log-likelihood function can be expressed as

$$L(N) = \sum_{t=1}^{N} \left\{ \ln |\mathbf{M}(t)| + \boldsymbol{\xi}(t)^\top\, \mathbf{M}(t)^{-1}\, \boldsymbol{\xi}(t) \right\}. \tag{3.1}$$

Given the data, the model in state-space form, and values of its parameters, the discrete-time Kalman filter provides a way of recursively computing $\boldsymbol{\xi}(t)$ and $\mathbf{M}(t)$, for $t = 1, \dots, N$ [12,14,15].

Let $\mathbf{x}(t \mid t-1) = E[\mathbf{x}(t) \mid Y(t-1)]$ be the prediction of the state vector one period ahead, with associated error, $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{x}(t \mid t-1)$, and error covariance matrix, $\mathbf{V}(t) = E[\tilde{\mathbf{x}}(t)\,\tilde{\mathbf{x}}(t)^\top]$. Then, given initial values $\mathbf{x}(1 \mid 0)$ and $\mathbf{V}(1)$, the following Kalman-filtering equations recursively compute $\boldsymbol{\xi}(t)$ and $\mathbf{M}(t)$, for $t = 2, \dots, N$,

$$\mathbf{M}(t) = \boldsymbol{\Sigma}_\zeta + \mathbf{D}\,\mathbf{V}(t)\,\mathbf{D}^\top, \tag{3.2}$$

$$\boldsymbol{\xi}(t) = \mathbf{y}(t) - \mathbf{D}\,\mathbf{x}(t \mid t-1), \tag{3.3}$$

$$\mathbf{K}(t) = \mathbf{F}\,\mathbf{V}(t)\,\mathbf{D}^\top\,\mathbf{M}(t)^{-1}, \tag{3.4}$$

$$\mathbf{x}(t+1 \mid t) = \mathbf{F}\,\mathbf{x}(t \mid t-1) + \mathbf{K}(t)\,\boldsymbol{\xi}(t), \tag{3.5}$$

$$\boldsymbol{\Phi}(t) = \mathbf{F} - \mathbf{K}(t)\,\mathbf{D}, \tag{3.6}$$

$$\mathbf{V}(t+1) = \boldsymbol{\Phi}(t)\,\mathbf{V}(t)\,\boldsymbol{\Phi}(t)^\top + \mathbf{K}(t)\,\boldsymbol{\Sigma}_\zeta\,\mathbf{K}(t)^\top + \mathbf{G}\,\mathbf{G}^\top. \tag{3.7}$$

$\mathbf{K}(t)$ is called the Kalman gain matrix; $\boldsymbol{\Phi}(t)$ is called the closed loop matrix. Equations (3.2)–(3.7) are one of many theoretically—but not numerically—equivalent ways of writing the Kalman filter [14]. In particular, equation (3.7) is a numerically stable version of the discrete-time recursive Riccati equation [13].

If the model is stationary, then the exact likelihood function is obtained by setting $\mathbf{x}(1 \mid 0)$ to the unconditional mean of $\mathbf{x}(t)$ and setting $\mathbf{V}(1)$ to the unconditional covariance of $\mathbf{x}(t)$. Accordingly, when the data have been adjusted for means and fixed effects (as they should be for a VARMA model), $\mathbf{x}(1 \mid 0) = \mathbf{0}$ and $\mathbf{V}(1) = \mathbf{C}$, where $\mathbf{C}$ solves the discrete-time algebraic Lyapunov equation

$$\mathbf{C} - \mathbf{F}\,\mathbf{C}\,\mathbf{F}^\top = \mathbf{G}\,\mathbf{G}^\top. \tag{3.8}$$

If the model is nonstationary, then, other methods must be used for setting $\mathbf{x}(1 \mid 0)$ and $\mathbf{V}(1)$ (e.g., [16]).

There are various ways of solving (3.8). A straightforward and generally effective way is as follows. First, apply the vectorization rule [17, p. 30]

$$\mathrm{vec}(\mathbf{A}\,\mathbf{B}\,\mathbf{C}) = [\mathbf{C}^\top \otimes \mathbf{A}]\,\mathrm{vec}(\mathbf{B}), \tag{3.9}$$

where $\otimes$ denotes the Kronecker product, to obtain the $n^2 \times n^2$ linear system

$$[\mathbf{I}_{n^2} - (\mathbf{F} \otimes \mathbf{F})]\,\mathrm{vec}(\mathbf{C}) = \mathrm{vec}(\mathbf{G}\,\mathbf{G}^\top), \tag{3.10}$$

where $\mathbf{I}_{n^2}$ is the $n^2 \times n^2$ identity matrix. Then, exploit the symmetry of $\mathbf{C}$ and $\mathbf{G}\,\mathbf{G}^\top$ to reduce (3.9) to the $n(n+1)/2 \times n(n+1)/2$ linear system $\mathbf{A}\,\mathbf{z} = \mathbf{b}$, where $\mathbf{z} = \mathrm{vech}(\mathbf{C})$ and $\mathbf{b} = \mathrm{vech}(\mathbf{G}\,\mathbf{G}^\top)$. Finally, solve $\mathbf{A}\,\mathbf{z} = \mathbf{b}$ for $\mathbf{z}$ by some standard method (e.g., Gaussian elimination with partial pivoting; [18, pp. 52–69]).

The n^2 eigenvalues of $[\mathbf{I}_{n^2} - (\mathbf{F} \otimes \mathbf{F})]$ are $1 - \lambda_i\,\lambda_j$, for i and $j = 1, \ldots, n$, where the λ's are eigenvalues of $\mathbf{F}$. Stationarity means that $|\lambda_i| < 1$, for $i = 1, \ldots, n$, hence, that $1 - \lambda_i\,\lambda_j \neq 0$, for i and $j = 1, \ldots, n$. Therefore, stationarity implies that $[\mathbf{I}_{n^2} - (\mathbf{F} \otimes \mathbf{F})]$ is invertible, so that (3.8) should yield a unique positive semidefinite value of $\mathbf{C}$.

A more efficient method for computing $\mathbf{C}$ is as follows. First, following Mittnik [19], write $\mathbf{C} = \boldsymbol{\Theta}\,\boldsymbol{\Omega}\,\boldsymbol{\Theta}^\top$, where $\boldsymbol{\Theta}$ is an $n \times 2n$ matrix of zeroes and rearranged elements of $\boldsymbol{\theta}$, and $\boldsymbol{\Omega}$ is the $2n \times 2n$ matrix of contemporaneous and serial covariances of the underlying process, $\mathbf{u}(t)$, and the disturbance vector, $\mathbf{e}(t)$. The (2,2)-block of $\boldsymbol{\Omega}$ is $\mathbf{I}_n$, the $n \times n$ identity matrix. The off-diagonal blocks of $\boldsymbol{\Omega}$ comprise zeroes and Wold MA coefficients of the model. The $(1,1)$-block of $\boldsymbol{\Omega}$ can be computed with algorithms proposed by Mittnik [20,21]. Relative efficiency of the algorithms depends, in part, on the autoregressive order p. Finally, given $\boldsymbol{\Omega}$, recursively compute $\boldsymbol{\Theta}\,\boldsymbol{\Omega}\,\boldsymbol{\Theta}^\top$ [19].

Computations are efficient when they are reliable (or numerically stable), accurate, and speedy. In a maximum likelihood estimation, speed is important because the likelihood function is computed repeatedly. Speed is less important in the present computations, because they will be executed only once. However, in computing $\mathbf{C}$, reliability and accuracy decreases as the size of

the linear system to be solved increases (cf. [18, pp. 24–28]). In this respect, it might also be desirable to reduce equations (4.8) and (5.12) to (5.14), (6.12), (6.16), and (6.18), derived in Sections 4–6, with analogues of Mittnik's results.

We assume that enough assumptions are in force so that the Kalman filter converges, as $t \to \infty$, and that the converged filter is asymptotically stable. Sufficient conditions for this to occur are that $\mathbf{D}\,\mathbf{G}\,\mathbf{G}^\top\,\mathbf{D}^\top > \mathbf{0}$ (positive definite) or $\mathbf{\Sigma}_\varsigma > \mathbf{0}$ (or some "combination" of these matrices is positive definite), and that system (2.2) and (2.3) is stabilizable and detectable [22, pp. 533–535]. In the time-invariant case, the Kalman filter is said to converge if $\mathbf{M}(t)$, $\mathbf{K}(t)$, $\mathbf{\Phi}(t)$, and $\mathbf{V}(t)$ converge to constant values. Equations (3.2)–(3.7) show that $\mathbf{M}(t)$, $\mathbf{K}(t)$, and $\mathbf{\Phi}(t)$ converge to constant values if and only if $\mathbf{V}(t)$ does. In the time-invariant case, the Kalman filter is asymptotically stable if the converged closed loop matrix, $\mathbf{\Phi} = \lim_{t\to\infty} \mathbf{\Phi}(t)$, is a stability matrix (has all eigenvalues inside the unit circle). In Section 6, these notions of convergence and asymptotic stability are extended to the periodic case.

4. RECURSIVE COMPUTATION OF SAMPLE AND ASYMPTOTIC INFORMATION MATRICES

Let $\mathbf{I}_\theta(N)$ and $\mathbf{I}_\theta^\infty(N)$, respectively, denote sample and asymptotic information matrices of $\boldsymbol{\theta}$. Then, under general conditions, $\mathbf{I}_\theta(N) = \frac{1}{2}E[\nabla^2 \mathcal{L}(N)]$ and $\mathbf{I}_\theta^\infty(N) = \frac{1}{2}\lim_{N\to\infty} E[\nabla^2 \mathcal{L}(N)/N]$, where $\nabla^2 \mathcal{L}(N)$ is the approximate $\alpha \times \alpha$ Hessian matrix of $L(N)$, with (i,j)-element

$$\partial_{ij}^2 \mathcal{L}(N) = \sum_{t=1}^{N} \operatorname{tr}\left\{\mathbf{M}(t)^{-1}\left[\partial_i \mathbf{M}(t)\;\mathbf{M}(t)^{-1}\,\partial_j \mathbf{M}(t) + 2\,\partial_i \boldsymbol{\xi}(t)\,\partial_j \boldsymbol{\xi}(t)^\top\right]\right\}, \tag{4.1}$$

α being the dimension of $\boldsymbol{\theta}$ [5, pp. 547–549].

In general, ∂_i denotes a first-partial derivative with respect to $\boldsymbol{\theta}_i$, element i of $\boldsymbol{\theta}$, and ∂_{ij}^2 denotes a second-partial derivative with respect to $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. Because $\nabla^2 \mathcal{L}$ is symmetric, it suffices to compute its lower-triangular part, i.e., to consider $i = 1,\ldots,\alpha$ and $j = 1,\ldots,i$.

Let $\mathbf{\Gamma}_{ij}(t) = E\,\partial_i \boldsymbol{\xi}(t)\,\partial_j \boldsymbol{\xi}(t)^\top$. Then $I_{\theta,ij}(N)$, element (i,j) of $\mathbf{I}_\theta(N)$, is given by

$$I_{\theta,ij}(N) = \sum_{t=1}^{N} \operatorname{tr}\left\{\mathbf{M}(t)^{-1}\left[\frac{1}{2}\,\partial_i \mathbf{M}(t)\;\mathbf{M}(t)^{-1}\,\partial_j \mathbf{M}(t) + \mathbf{\Gamma}_{ij}(t)\right]\right\}. \tag{4.2}$$

For notational simplicity, we shall sometimes collectively refer to partial derivatives in their more concise gradient (or Jacobian) form. For example, the gradient form of $\partial_i \mathbf{M}(t)$, for $i = 1,\ldots,\alpha$, is $\nabla \mathbf{M}(t) = [\operatorname{vec}(\partial_1 \mathbf{M}(t)),\ldots,\operatorname{vec}(\partial_\alpha \mathbf{M}(t))]$ [17, p. 175]. For the same reason, but without assigning any particular algebraic structure, we shall refer to $\mathbf{\Gamma}_{ij}(t)$, for $i = 1,\ldots,\alpha$ and $j = 1,\ldots,i$, as $\mathbf{\Gamma}(t)$.

To compute $\mathbf{I}_\theta(N)$, we need to compute $\mathbf{M}(t)$, $\nabla \mathbf{M}(t)$ and $\mathbf{\Gamma}(t)$, for $t = 1,\ldots,N$. $\mathbf{M}(t)$ is computed by iterating on (3.2), (3.4), (3.6), and (3.7), starting with $\mathbf{M}(1) = \mathbf{\Sigma}_\varsigma + \mathbf{D}\,\mathbf{C}\,\mathbf{D}^\top$, where $\mathbf{C}$ solves (3.8). We now describe the computation of:

(1) $\nabla \mathbf{M}(t)$, for $t = 2,\ldots,N$;
(2) $\nabla \mathbf{M}(1)$;
(3) $\mathbf{\Gamma}(t)$, for $t = 2,\ldots,N$; and
(4) $\mathbf{\Gamma}(1)$.

4.1. Computation of $\nabla \mathbf{M}(t)$

First, conditional on $\{\mathbf{y}(t)\}_{t=1}^{N}$, differentiate (3.2)–(3.5) and (3.7) with respect to $\boldsymbol{\theta}$, to obtain

$$\partial_i \mathbf{M}(t) = \mathbf{D}\ \partial_i \mathbf{V}(t)\ \mathbf{D}^\top, \tag{4.3}$$

$$\partial_i \boldsymbol{\xi}(t) = -\mathbf{D}\ \partial_i \mathbf{x}(t \mid t-1), \tag{4.4}$$

$$\partial_i \mathbf{K}(t) = \partial_i \mathbf{F}\ \mathbf{V}(t)\ \mathbf{D}^\top\ \mathbf{M}(t)^{-1} + \mathbf{F}\ \partial_i \mathbf{V}(t)\ \mathbf{D}^\top\ \mathbf{M}(t)^{-1} - \mathbf{K}(t)\ \partial_i \mathbf{M}(t)\ \mathbf{M}(t)^{-1}, \tag{4.5}$$

$$\partial_i \mathbf{x}(t+1 \mid t) = \partial_i \mathbf{F}\ \mathbf{x}(t \mid t-1) + \mathbf{F}\ \partial_i \mathbf{x}(t \mid t-1) + \partial_i \mathbf{K}(t)\ \boldsymbol{\xi}(t) + \mathbf{K}(t)\, \partial_i \boldsymbol{\xi}(t), \tag{4.6}$$

$$\partial_i \mathbf{V}(t+1) = \boldsymbol{\Phi}(t)\, \partial_i \mathbf{V}(t)\ \boldsymbol{\Phi}(t)^\top + \partial_i \mathbf{F}\ \mathbf{V}(t)\ \boldsymbol{\Phi}(t)^\top + \boldsymbol{\Phi}(t)\ \mathbf{V}(t)\, \partial_i \mathbf{F}^\top + \partial_i \mathbf{G}\ \mathbf{G}^\top + \mathbf{G}\ \partial_i \mathbf{G}^\top, \tag{4.7}$$

for $i = 1, \ldots, \alpha$ [5, p. 546)]. Equation (4.7) is obtained by differentiating (3.7) and using (3.4) to combine and simplify terms [23, p. 470]. Then, for $t = 2, \ldots, N$, compute $\nabla \mathbf{M}(t)$ by iterating with (4.3) and (4.7). (Equations (4.4)–(4.7) will be used to derive equations for computing $\boldsymbol{\Gamma}(t)$.)

For $i = 1, \ldots, \alpha$, $\partial_i \mathbf{F}$ and $\partial_i \mathbf{G}$ are selection matrices of zeroes and ones. For each i, the matrices are zero except for a single element equal to one. For example, because $\boldsymbol{\theta}_1$ is element (1,1) of $\mathbf{A}_1$ and $\mathbf{F}$, element (1,1) of $\partial_1 \mathbf{F}$ is one, remaining elements of $\partial_1 \mathbf{F}$ are zero, and $\partial_1 \mathbf{G}$ is zero.

$\nabla \mathbf{M}(t)$ can be expressed more directly in terms of

$$\nabla \mathbf{V}(t) = [\mathrm{vec}(\partial_1 \mathbf{V}(t)), \ldots, \mathrm{vec}(\partial_\alpha \mathbf{V}(t))].$$

Applying (3.9)–(4.3) yields $\mathrm{vec}(\partial_i \mathbf{M}(t)) = [\mathbf{D} \otimes \mathbf{D}]\ \mathrm{vec}(\partial_i \mathbf{V}(t))$, hence, $\nabla \mathbf{M}(t) = [\mathbf{D} \otimes \mathbf{D}]\, \nabla \mathbf{V}(t)$. Then, given that $\mathbf{D} = [\mathbf{I}, \mathbf{0}, \ldots, \mathbf{0}]$,

$$\nabla \mathbf{M}(t) = [\mathrm{vec}(\partial_1 \mathbf{V}_{11}(t)), \ldots, \mathrm{vec}(\partial_\alpha \mathbf{V}_{11}(t))],$$

where $\partial_i \mathbf{V}_{11}(t)$ is block (1,1) of size $m \times m$ of $\partial_i \mathbf{V}(t)$.

4.2. Computation of $\nabla \mathbf{M}(1)$

First, differentiate (3.8) and vectorize the result (with (3.9)), to obtain

$$[\mathbf{I}_{n^2} - (\mathbf{F} \otimes \mathbf{F})]\ \mathrm{vec}(\partial_i \mathbf{C}) = \mathrm{vec}(\partial_i \mathbf{F}\ \mathbf{C}\ \mathbf{F}^\top + \mathbf{F}\ \mathbf{C}\ \partial_i \mathbf{F}^\top + \partial_i \mathbf{G}\ \mathbf{G}^\top + \mathbf{G}\ \partial_i \mathbf{G}^\top), \tag{4.8}$$

for $i = 1, \ldots, \alpha$ [5, p. 547]. Then, solve (4.8) for $\mathrm{vec}(\partial_i \mathbf{C})$, for $i = 1, \ldots, \alpha$ and set

$$\nabla \mathbf{M}(1) = [\mathrm{vec}(\partial_1 \mathbf{C}_{11}), \ldots, \mathrm{vec}(\partial_\alpha \mathbf{C}_{11})],$$

where $\partial_i \mathbf{C}_{11}$ is block (1,1) of size $m \times m$ of $\partial_i \mathbf{C}$. Model stationarity ensures invertibility of $[\mathbf{I}_{n^2} - (\mathbf{F} \otimes \mathbf{F})]$.

4.3. Computation of $\boldsymbol{\Gamma}(t)$

Equation (4.4) shows that

$$\boldsymbol{\Gamma}_{ij}(t) = \mathbf{D}\ E[\partial_i \mathbf{x}(t \mid t-1)\, \partial_j \mathbf{x}(t \mid t-1)^\top]\ \mathbf{D}^\top.$$

Define the augmented predicted state vector, $\mathbf{x}_i^*(t \mid t-1) = [\mathbf{x}(t \mid t-1)^\top, \partial_i \mathbf{x}(t \mid t-1)^\top]^\top$, and the second-moment matrix, $\mathbf{W}_{ij}^*(t) = E[\mathbf{x}_i^*(t \mid t-1)\ \mathbf{x}_j^*(t \mid t-1)^\top]$, for $i = 1, \ldots, \alpha$ and $j = 1, \ldots, i$. Partition $\mathbf{W}_{ij}(t)$ into four $n \times n$ blocks,

$$\mathbf{W}_{ij}(t) = \begin{bmatrix} \mathbf{S}(t) & \mathbf{U}_j(t)^\top \\ \mathbf{U}_i(t) & \mathbf{Z}_{ij}(t) \end{bmatrix}, \tag{4.9}$$

so that $\mathbf{\Gamma}_{ij}(t) = \mathbf{D}\ \mathbf{Z}_{ij}(t)\ \mathbf{D}^\top$. Combine (3.5), (4.4), and (4.6) as

$$\mathbf{x}_i^*(t+1 \mid t) = \mathbf{F}_i^*(t)\ \mathbf{x}_i^*(t \mid t-1) + \mathbf{K}_i^*(t)\ \boldsymbol{\xi}(t), \tag{4.10}$$

where

$$\mathbf{F}_i^*(t) = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \partial_i \mathbf{F} & \mathbf{\Phi}(t) \end{bmatrix} \quad \text{and} \quad \mathbf{K}_i^* = \begin{bmatrix} \mathbf{K}(t) \\ \partial_i \mathbf{K}(t) \end{bmatrix}.$$

Because $E[\mathbf{x}_i^*(t \mid t-1)\ \boldsymbol{\xi}(t)^\top] = 0$, (4.10) implies that

$$\mathbf{W}_{ij}^*(t+1) = \mathbf{F}_i^*(t)\ \mathbf{W}_{ij}^*\ \mathbf{F}_j^*(t)^\top + \mathbf{K}_i^*(t)\ \mathbf{M}(t)\ \mathbf{K}_j^*(t)^\top. \tag{4.11}$$

Recall that, by definition, $\mathbf{V}(t) = E[\tilde{\mathbf{x}}(t)\ \tilde{\mathbf{x}}(t)^\top]$ and $\mathbf{S}(t) = E[\mathbf{x}(t \mid t-1)\ \mathbf{x}(t \mid t-1^\top]$ and, by stationarity, $\mathbf{C} = E[\mathbf{x}(t)\ \mathbf{x}(t)^\top]$, for $t = 1, \ldots, N$. Moreover, because $E[\mathbf{x}(t \mid t-1)\ \tilde{\mathbf{x}}(t)^\top] = \mathbf{0}$, $\mathbf{x}(t) = \mathbf{x}(t \mid t-1) + \tilde{\mathbf{x}}(t)$ implies that $\mathbf{C} = \mathbf{S}(t) + \mathbf{V}(t)$, for $t = 1, \ldots, N$. Therefore, $\mathbf{S}(t) = \mathbf{C} - \mathbf{V}(t)$, for $t = 1, \ldots, N$, where $\mathbf{C}$ and $\mathbf{V}(t)$ are previously computed, so that the (1,1)-block of (4.11) can be ignored.

$\mathbf{U}_i(t)$ and $\mathbf{Z}_{ij}(t)$ are updated with the (2,1)- and (2,2)-blocks of (4.11),

$$\mathbf{U}_i(t+1) = \partial_i \mathbf{F}\ \mathbf{S}(t)\ \mathbf{F}^\top + \mathbf{\Phi}(t)\ \mathbf{U}_i(t)\ \mathbf{F}^\top + \partial_i \mathbf{K}(t)\ \mathbf{M}(t)\ \mathbf{K}(t)^\top, \tag{4.12}$$

$$\begin{aligned} \mathbf{Z}_{ij}(t+1) = {} & \partial_i \mathbf{F}\ \mathbf{S}(t)\, \partial_j \mathbf{F}^\top + \mathbf{\Phi}(t)\ \mathbf{U}_i(t)\, \partial_j \mathbf{F}^\top + \partial_i \mathbf{F}\ \mathbf{U}_j(t)^\top\ \mathbf{\Phi}(t)^\top \\ & + \mathbf{\Phi}(t)\ \mathbf{Z}_{ij}(t)\ \mathbf{\Phi}(t)^\top + \partial_i \mathbf{K}(t)\ \mathbf{M}(t)\, \partial_j \mathbf{K}(T)^\top. \end{aligned} \tag{4.13}$$

In sum, to compute $\mathbf{\Gamma}(t)$, for $t = 2, \ldots, N$: set $\mathbf{S}(t) = \mathbf{C} - \mathbf{V}(t)$; compute $\partial_i \mathbf{K}(t)$ with (4.5); compute $\mathbf{U}_i(t)$ and $\mathbf{Z}_{ij}(t)$ with (4.12) and (4.13); and, pick $\mathbf{\Gamma}_{ij}(t) = \mathbf{D}\ \mathbf{Z}_{ij}(t)\ \mathbf{D}^\top$ out of $\mathbf{Z}_{ij}(t)$, for $i = 1, \ldots, \alpha$ and $j = 1, \ldots, i$.

4.4. Computation of $\mathbf{\Gamma}(1)$

Because $\mathbf{x}(1 \mid 0)$ is set identically equal to zero for all values of $\boldsymbol{\theta}$, so is $\mathbf{x}_i^*(1 \mid 0)$, for $i = 1, \ldots, \alpha$. Therefore, for all values of $\boldsymbol{\theta}$, $\mathbf{Z}_{ij}(1) = \mathbf{0}$, hence, $\mathbf{\Gamma}_{ij}(1) = \mathbf{0}$, for $i = 1, \ldots, \alpha$ and $j = 1, \ldots, i$.

To compute $\mathbf{I}_\theta^\infty(N)$ recursively, compute $\mathbf{I}_\theta(N)$ for successively higher values of N until $\mathbf{I}_\theta(N)/N$ has converged in some matrix norm [18, pp. 11–16]. If $\mathbf{D}\ \mathbf{G}\ \mathbf{G}^\top\ \mathbf{D}^\top > \mathbf{0}$ or $\mathbf{\Sigma}_\zeta > \mathbf{0}$ and the system (2.2) and (2.3) is stabilizable and detectable, then $\mathbf{V}(t)$ and, hence, $\mathbf{M}(t)$, $\mathbf{K}(t)$, and $\mathbf{\Phi}(t)$, converge to the same limiting values for any positive semidefinite value of $\mathbf{V}(1)$ [22, pp. 533–535]. Therefore, $\mathbf{I}_\theta^\infty(N)$ is independent of $\mathbf{V}(1)$ and $\nabla \mathbf{V}(1)$, so that, when computing $\mathbf{I}_\theta^\infty(N)$, it is simplest to set $\mathbf{V}(1) = \mathbf{0}$ and $\nabla \mathbf{V}(1) = \mathbf{0}$.

The inefficiency in [8] is as follows. First, corresponding to $\mathbf{x}_i^*(t \mid t-1)$, Terceiro Lomba, defines

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

so that his analogue of (4.11) is 50% larger than necessary, i.e., has dimension $3n^2 \times 3n^2$ instead of $2n^2 \times 2n^2$. Second, he decomposes $\mathbf{W}_{ij}^*(t)$ as $\mu_i^*(t)\ \mu_j^*(t)^\top + \mathbf{C}_{ij}^*(t)$, where $\mu_i^*(t) = E[\mathbf{x}_i^*(t)]$ and $\mathbf{C}_{ij}^*(t) = E[(\mathbf{x}_i^*(t) - \mu_i^*(t))\,(\mathbf{x}_j^*(t) - \mu_j^*(t))^\top]$ and propagates the $\mu_i^*(t)$'s and $\mathbf{C}_{ij}^*(t)$'s separately. This is unnecessary even when, as in his analysis, the model includes exogenous variables. Third, Terceiro Lomba does not reduce (4.11) to (4.12) and (4.13). In particular, he does not use $\mathbf{S}(t) = \mathbf{C} - \mathbf{V}(t)$, which continues to hold when exogenous variables are present, and, therefore, retains the (1,1)-block of (4.11).

5. NONRECURSIVE COMPUTATION OF THE ASYMPTOTIC INFORMATION MATRIX FOR TIME-INVARIANT MODELS AND SAMPLES

If, as has been assumed so far, the model is time invariant and stationary, the sample is time invariant, and the Kalman filter converges and is asymptotically stable, then the (i,j)-element of $\mathbf{I}_\theta^\infty(N)$ can be expressed as

$$I_{\theta,ij}^\infty(N) = N \operatorname{tr}\left\{\mathbf{M}^{-1}\left[\frac{1}{2}\,\partial_i\mathbf{M}\;\mathbf{M}^{-1}\,\partial_j\mathbf{M} + \mathbf{\Gamma}_{ij}\right]\right\}, \tag{5.1}$$

for $i = 1,\ldots,\alpha$ and $j = 1,\ldots,i$, where $\partial_i\mathbf{M} = \lim_{t\to\infty} \partial_i\mathbf{M}(t)$ and $\mathbf{\Gamma}_{ij} = \lim_{t\to\infty} \mathbf{\Gamma}_{ij}(t)$. This section derives a nonrecursive method for computing $\mathbf{M}$, $\nabla\mathbf{M}$, and $\mathbf{\Gamma}$.

Under the aforementioned conditions, as $t \to \infty$, equations (3.2), (3.4), (3.6), and (3.7) converge to

$$\mathbf{M} = \mathbf{\Sigma}_\zeta + \mathbf{D}\,\mathbf{V}\,\mathbf{D}^\top, \tag{5.2}$$

$$\mathbf{K} = \mathbf{F}\,\mathbf{V}\,\mathbf{D}^\top\,\mathbf{M}^{-1}, \tag{5.3}$$

$$\mathbf{\Phi} = \mathbf{F} - \mathbf{K}\,\mathbf{D}, \tag{5.4}$$

$$\mathbf{V} = \mathbf{\Phi}\,\mathbf{V}\,\mathbf{\Phi}^\top + \mathbf{K}\,\mathbf{\Sigma}_\zeta\,\mathbf{K}^\top + \mathbf{G}\,\mathbf{G}^\top; \tag{5.5}$$

equations (4.3), (4.5), and (4.7) converge to

$$\partial_i\mathbf{M} = \mathbf{D}\,\partial_i\mathbf{V}\,\mathbf{D}^\top, \tag{5.6}$$

$$\partial_i\mathbf{K} = \partial_i\mathbf{F}\,\mathbf{V}\,\mathbf{D}^\top\,\mathbf{M}^{-1} + \mathbf{F}\,\partial_i\mathbf{V}\,\mathbf{D}^\top\,\mathbf{M}^{-1} - \mathbf{K}\,\partial_i\mathbf{M}\,\mathbf{M}^{-1}, \tag{5.7}$$

$$\partial_i\mathbf{V} = \mathbf{\Phi}\,\partial_i\mathbf{V}\,\mathbf{\Phi}^\top + \partial_i\mathbf{F}\,\mathbf{V}\,\mathbf{\Phi}^\top + \mathbf{\Phi}\,\mathbf{V}\,\partial_i\mathbf{F}^\top + \partial_i\mathbf{G}\,\mathbf{G}^\top + \mathbf{G}\,\partial_i\mathbf{G}^\top; \tag{5.8}$$

and equations (4.12) and (4.13) converge to

$$\mathbf{U}_i = \mathbf{\Phi}\,\mathbf{U}_i\,\mathbf{F}^\top + \partial_i\mathbf{F}\,\mathbf{S}\,\mathbf{F}^\top + \partial_i\mathbf{K}\,\mathbf{M}\,\mathbf{K}^\top, \tag{5.9}$$

$$\mathbf{Z}_{ij} = \mathbf{\Phi}\,\mathbf{Z}_{ij}\,\mathbf{\Phi}^\top + \partial_i\mathbf{F}\,\mathbf{S}\,\partial_j\mathbf{F}^\top + \mathbf{\Phi}\,\mathbf{U}_i\,\partial_j\mathbf{F}^\top + \partial_i\mathbf{F}\,\mathbf{U}_j^\top\,\mathbf{\Phi}^\top + \partial_i\mathbf{K}\,\mathbf{M}\,\partial_j\mathbf{K}^\top, \tag{5.10}$$

where the absence of the time argument, t, in a derived matrix denotes its steady-state value.

5.1. Computation of M

First, use (5.2)–(5.4) to successively eliminate $\mathbf{M}$, $\mathbf{K}$, and $\mathbf{\Phi}$ to write (5.5) in the classical form of the algebraic Riccati equation,

$$\mathbf{V} = \mathbf{G}\,\mathbf{G}^\top + \mathbf{F}\,\mathbf{V}\,\mathbf{F}^\top - \mathbf{F}\,\mathbf{V}\,\mathbf{D}^\top\,[\mathbf{\Sigma}_\zeta + \mathbf{D}\,\mathbf{V}\,\mathbf{D}^\top]^{-1}\,\mathbf{D}\,\mathbf{V}\,\mathbf{F}^\top. \tag{5.11}$$

Then, solve (5.11) for $\mathbf{V} \geq \mathbf{0}$. Many recursive and nonrecursive methods have been proposed for doing this [22, pp. 243–253]. A generally reliable, accurate, and quick Schur-decomposition method is described by [24]. Finally, given $\mathbf{V}$, set $\mathbf{M}$ with (5.2).

5.2. Computation of ∇M

First, vectorize (5.8), to obtain

$$[\mathbf{I}_{n^2} - (\mathbf{\Phi}\otimes\mathbf{\Phi})]\,\operatorname{vec}(\partial_i\mathbf{V}) = \operatorname{vec}(\partial_i\mathbf{F}\,\mathbf{V}\,\mathbf{\Phi}^\top + \mathbf{\Phi}\,\mathbf{V}\,\partial_i\mathbf{F}^\top + \partial_i\mathbf{G}\,\mathbf{G}^\top + \mathbf{G}\,\partial_i\mathbf{G}^\top). \tag{5.12}$$

Then, solve (5.12) for $\operatorname{vec}(\partial_i\mathbf{V})$, for $i = 1,\ldots,\alpha$. Finally, set

$$\nabla\mathbf{M} = [\operatorname{vec}(\partial_1\mathbf{V}_{11}),\ldots,\operatorname{vec}(\partial_\alpha\mathbf{V}_{11})],$$

where $\partial_i\mathbf{V}_{11}$ is block (1,1) of size $n \times n$ of $\partial_i\mathbf{V}$. Asymptotic stability of the Kalman filter ensures that $\mathbf{\Phi}$ is a stability matrix, hence, that $[\mathbf{I}_{n^2} - (\mathbf{\Phi}\otimes\mathbf{\Phi})]$ is invertible.

5.3. Computation of Γ

Set $\mathbf{S} = \mathbf{C} - \mathbf{V}$. Compute $\partial_i \mathbf{K}$ with (5.7). Vectorize (5.9) and (5.10), to obtain

$$[\mathbf{I}_{n^2} - (\mathbf{F} \otimes \mathbf{\Phi})] \operatorname{vec}(\mathbf{U}_i) = \operatorname{vec}(\partial_j \mathbf{F}\ \mathbf{S}\ \mathbf{F}^\top + \partial_i \mathbf{K}\ \mathbf{M}\ \mathbf{K}^\top), \tag{5.13}$$

$$\begin{aligned}[\mathbf{I}_{n^2} - (\mathbf{\Phi} \otimes \mathbf{\Phi})] \operatorname{vec}(\mathbf{Z}_{ij}) = \operatorname{vec}(&\partial_i \mathbf{F}\ \mathbf{S}\ \partial_j \mathbf{F}^\top + \mathbf{\Phi}\ \mathbf{U}_i\ \partial_j \mathbf{F}^\top \\ &+ \partial_i \mathbf{F}\ \mathbf{U}_j^\top \mathbf{\Phi}^\top + \partial_i \mathbf{K}\ \mathbf{M}\ \partial_j \mathbf{K}^\top).\end{aligned} \tag{5.14}$$

Solve (5.13) for $\operatorname{vec}(\mathbf{U}_i)$, for $i = 1, \ldots, \alpha$. Solve (5.14) for $\operatorname{vec}(\mathbf{Z}_{ij})$, for $i = 1, \ldots, \alpha$ and $j = 1, \ldots, i$. Finally, set $\mathbf{\Gamma}_{ij} = \mathbf{D}\ \mathbf{Z}_{ij}\ \mathbf{D}^\top$. Stationarity of the model and asymptotic stability of the Kalman filter ensure invertibility of the system matrices of (5.13) and (5.14).

6. NONRECURSIVE COMPUTATION OF THE ASYMPTOTIC INFORMATION MATRIX FOR PERIODIC MODELS AND SAMPLES

Consider, for example, a bivariate periodic model of monthly seasonality which is estimated with a sample in which the first variable is observed monthly, the second variable is observed quarterly, at the end of each quarter, and the sample starts in the first month of a quarter. Then, the model has periodicity 12, so that $\mathbf{F}(t)$ and $\mathbf{G}(t)$ may assume up to 12 different values in a year and $\mathbf{D}(t)$ and $\mathbf{\Sigma}_\zeta(t)$ vary as follows. For $t = 1, 2, 4, 5, \ldots$, $\mathbf{D}(t)$ is the n-dimensional row vector $[1, 0, \ldots, 0]$, and $\Sigma_\zeta(t)$ is a constant nonnegative scalar; for $t = 3, 6, 9, \ldots$, $\mathbf{D}(t)$ is the $2 \times n$ dimensional matrix $[\mathbf{I}, \mathbf{0}, \ldots, \mathbf{0}]$, and $\mathbf{\Sigma}_\zeta(t)$ is a constant 2×2 positive semidefinite matrix. Because the second variable is only observed every third month, the sample has periodicity 3, and, because 12 is evenly divisible by 3, the model and sample have an overall periodicity $\rho = 12$. Jones [15] first suggested handling such missing data by varying $\mathbf{D}(t)$ and $\mathbf{\Sigma}_\zeta(t)$ in the Kalman filter; Ansley and Kohn [12] extended the idea to multivariate models.

In general, then, we may consider a model and sample with overall periodicity ρ. In such case, the Kalman filter converges if $\mathbf{M}(t)$, $\mathbf{K}(t)$, $\mathbf{\Phi}(t)$, and $\mathbf{V}(t)$ converge to limit cycles with periodicity ρ, as $t \to \infty$. Again, $\mathbf{M}(t)$, $\mathbf{K}(t)$, and $\mathbf{\Phi}(t)$ converge in this sense if and only if $\mathbf{V}(t)$ does. Formally, $\{\mathbf{V}(t)\}_{t=1}^{\infty}$ converges to a limit cycle with periodicity ρ, $\{\mathbf{V}_k\}_{k=1}^{\rho}$, if for any $\varepsilon > 0$, there is a T such that $t \geq T$ implies $||\mathbf{V}(t) - \mathbf{V}(t - \rho)|| < \varepsilon$, where $||\cdot||$ is some matrix norm. Accordingly, $\{\mathbf{M}(t)\}_{t=1}^{\infty}$, $\{\mathbf{K}(t)\}_{t=1}^{\infty}$, and $\{\mathbf{\Phi}(t)\}_{t=1}^{\infty}$ converge to limit cycles $\{\mathbf{M}_k\}_{k=1}^{\rho}$, $\{\mathbf{K}_k\}_{k=1}^{\rho}$, and $\{\mathbf{\Phi}_k\}_{k=1}^{\rho}$, with periodicity ρ.

Suppose, in fact, that the model and sample are periodic, with overall periodicity ρ, the model is stationary, and the Kalman filter converges and is asymptotically stable. Then, element (i, j) of $\mathbf{I}_\theta^\infty(N)$ can be expressed as

$$I_{\theta,ij}^{\infty}(N) = \frac{N}{\rho} \sum_{k=1}^{\rho} \operatorname{tr}\left\{\mathbf{M}_k^{-1}\left[\frac{1}{2}\partial_i \mathbf{M}_k\ \mathbf{M}_k^{-1}\partial_j \mathbf{M}_k + \mathbf{M}_k^{-1}\ \mathbf{\Gamma}_{k,ij}\right]\right\}, \tag{6.1}$$

for $i = 1, \ldots, \alpha$ and $j = 1, \ldots, i$, where $\{\partial_i \mathbf{M}_k\}_{k=1}^{\rho}$ and $\{\mathbf{\Gamma}_{k,ij}\}_{k=1}^{\rho}$ are the limit cycles of $\{\partial_i \mathbf{M}(t)\}_{t=1}^{\infty}$ and $\{\mathbf{\Gamma}_{ij}(t)\}_{t=1}^{\infty}$. This section describes how to compute $\mathbf{M}_k$, $\nabla \mathbf{M}_k$, and $\mathbf{\Gamma}_k$ with an extension of the method of the previous section. When $\rho = 1$, the computations reduce to those for the time-invariant case in the previous section.

To simplify notation, $\mathbf{F}(t)$, $\mathbf{G}(t)$, $\mathbf{D}(t)$, and $\mathbf{\Sigma}_\zeta(t)$ are henceforth written as $\mathbf{F}_k$, $\mathbf{G}_k$, $\mathbf{D}_k$, and $\mathbf{\Sigma}_{\zeta,k}$, for $k = 1, \ldots, \rho$. The different proofs of convergence and asymptotic stability of the linear optimal filtering problem and the dual linear optimal control problem, respectively, given by Zadrozny [11] and Todd [10], can be extended to prove that, under appropriate extensions of $\mathbf{D}\ \mathbf{G}\ \mathbf{G}^\top\ \mathbf{D}^\top > \mathbf{0}$ or $\mathbf{\Sigma}_\zeta > \mathbf{0}$ and stabilizability and detectability, the Kalman filter of a periodic model and sample converges and is asymptotically stable.

Define $\bar{\mathbf{F}}_k$ as the cyclical product $\bar{\mathbf{F}}_k = \mathbf{F}_{k-1} \cdots \mathbf{F}_1 \, \mathbf{F}_\rho \cdots \mathbf{F}_k$, for $k = 1, \ldots, \rho$. A periodic model is stationary if $\bar{\mathbf{F}}_k$ is a stability matrix, for $k = 1, \ldots, \rho$. A sufficient but generally not necessary condition for this to occur is that $\mathbf{F}_k$ is a stability matrix, for $k = 1, \ldots, \rho$. Similarly, define $\bar{\mathbf{\Phi}}_k = \mathbf{\Phi}_{k-1} \cdots \mathbf{\Phi}_1 \, \mathbf{\Phi}_\rho \cdots \mathbf{\Phi}_k$, for $k = 1, \ldots, \rho$. Then, the Kalman filter is asymptotically stable if $\bar{\mathbf{\Phi}}_k$ is a stability matrix, for $k = 1, \ldots, \rho$.

If the model and sample are periodic, with periodicity ρ, the model is stationary, and the Kalman filter converges and is asymptotically stable, then (5.2)–(5.11) become

$$\mathbf{V}_{k+1} = \mathbf{G}_k \, \mathbf{G}_k^\top + \mathbf{F}_k \, \mathbf{V}_k \, \mathbf{F}_k^\top - \mathbf{F}_k \, \mathbf{V}_k \, \mathbf{D}_k^\top [\mathbf{\Sigma}_{\zeta,k} + \mathbf{D}_k \, \mathbf{V}_k \, \mathbf{D}_k^\top]^{-1} \times \mathbf{D}_k \, \mathbf{V}_k \, \mathbf{F}_k^\top, \tag{6.2}$$

$$\mathbf{M}_k = \mathbf{\Sigma}_{\zeta,k} + \mathbf{D}_k \, \mathbf{V}_k \, \mathbf{D}_k^\top, \tag{6.3}$$

$$\mathbf{K}_k = \mathbf{F}_k \, \mathbf{V}_k \, \mathbf{D}_k^\top \, \mathbf{M}_k^{-1}, \tag{6.4}$$

$$\mathbf{\Phi}_k = \mathbf{F}_k - \mathbf{K}_k \, \mathbf{D}_k, \tag{6.5}$$

$$\partial_i \mathbf{V}_{k+1} = \mathbf{\Phi}_k \, \partial_i \mathbf{V}_k \, \mathbf{\Phi}_k^\top + \partial_i \mathbf{F}_k \, \mathbf{V}_k \, \mathbf{\Phi}_k^\top + \mathbf{\Phi}_k \, \mathbf{V}_k \, \partial_i \mathbf{F}_k^\top + \partial_i \mathbf{G}_k \, \mathbf{G}_k^\top + \mathbf{G}_k \, \partial_i \mathbf{G}_k^\top, \tag{6.6}$$

$$\partial_i \mathbf{M}_k = \mathbf{D}_k \, \partial_i \mathbf{V}_k \, \mathbf{D}_k^\top, \tag{6.7}$$

$$\partial_i \mathbf{K}_k = \partial_i \mathbf{F}_k \, \mathbf{V}_k \, \mathbf{D}_k^\top \, \mathbf{M}_k^{-1} + \mathbf{F}_k \, \partial_i \mathbf{V}_k \, \mathbf{D}_k^\top \, \mathbf{M}_k^{-1} - \mathbf{K}_k \, \partial_i \mathbf{M}_k \, \mathbf{M}_k^{-1}, \tag{6.8}$$

$$\mathbf{U}_{k+1,i} = \mathbf{\Phi}_k \, \mathbf{U}_{k,i} \, \mathbf{F}_k^\top + \partial_i \mathbf{F}_k \, \mathbf{S}_k \, \mathbf{F}_k^\top + \partial_i \mathbf{K}_k \, \mathbf{M}_k \, \mathbf{K}_k^\top, \tag{6.9}$$

$$\mathbf{Z}_{k+1,ij} = \mathbf{\Phi}_k \, \mathbf{Z}_{k,ij} \, \mathbf{\Phi}_k^\top + \partial_i \mathbf{F}_k \, \mathbf{S}_k \, \partial_j \mathbf{F}_k^\top + \mathbf{\Phi}_k \, \mathbf{U}_{k,i} \, \partial_j \mathbf{F}_k^\top + \partial_i \mathbf{F}_k \, \mathbf{U}_{k,j}^\top \, \mathbf{\Phi}_k^\top + \partial_i \mathbf{K}_k \, \mathbf{M}_k \, \partial_j \mathbf{K}_k^\top, \tag{6.10}$$

for $k = 1, \ldots, \rho$, $i = 1, \ldots, \alpha$, and $j = 1, \ldots, i$, where $\{\mathbf{S}_k\}_{k=1}^{\rho}$, $\{\mathbf{U}_{k,i}\}_{k=1}^{\rho}$, and $\{\mathbf{Z}_{k,ij}\}_{k=1}^{\rho}$ are the limit cycles of $\{\mathbf{S}(t)\}_{t=1}^{\infty}$, $\{\mathbf{U}_i(t)\}_{t=1}^{\infty}$, and $\{\mathbf{Z}_{ij}(t)\}_{t=1}^{\infty}$.

6.1. Computation of $\mathbf{M}_k$

Equation (6.2) needs to be solved for $\mathbf{V}_k \geq \mathbf{0}$, for $k = 1, \ldots, \rho$. This can be done in at least two ways. First, following Todd [10], combine (6.2), for $k = 1, \ldots, \rho$, into a Riccati equation of the form (5.11), in terms of $\mathbf{V} = \text{diag}[\mathbf{V}_1, \ldots, \mathbf{V}_\rho]$. Then, use some standard method such as the Schur-decomposition method (see [24]) to solve the Riccati equation for $\mathbf{V} \geq \mathbf{0}$. Alternatively, for $k = 1$ (or any other $k = 2, \ldots, \rho$), iterate $\rho - 1$ times with (6.2) to obtain a nonlinear equation in terms of $\mathbf{V}_1$. Solve this equation for$\mathbf{V}_1 \geq \mathbf{0}$ with any general method for solving nonlinear equations (e.g., the trust-region method [25]). Then, given $\mathbf{V}_1$, iterate $\rho - 1$ times with (6.2) to obtain $\mathbf{V}_2, \ldots, \mathbf{V}_\rho$. Finally, given $\mathbf{V}_k$ with either method, compute $\mathbf{M}_k$, $\mathbf{K}_k$, and $\mathbf{\Phi}_k$ with (6.3)–(6.5), for $k = 1, \ldots, \rho$.

6.2. Computation if $\nabla \mathbf{M}_k$

For $k = 1, \ldots, \rho$ and $i = 1, \ldots, \alpha$, compute $\mathbf{\Pi}_{k+\rho,i}$ by iterating over ν with

$$\mathbf{\Pi}_{\nu+1,i} = \mathbf{\Phi}_{\nu+1} \, \mathbf{\Pi}_{\nu,i} \mathbf{\Phi}_{\nu+1}^\top + \partial_i \mathbf{F}_{\nu+1} \, \mathbf{V}_{\nu+1} \, \mathbf{\Phi}_{\nu+1}^\top + \mathbf{\Phi}_{\nu+1} \, \mathbf{V}_{\nu+1} \partial_i \mathbf{F}_{\nu+1}^\top + \partial_i \mathbf{G}_{\nu+1} \, \mathbf{G}_{\nu+1}^\top + \mathbf{G}_{\nu+1} \, \partial_i \mathbf{G}_{\nu+1}^\top, \tag{6.11}$$

for $\nu = k, \ldots, k + \rho - 1$, starting with

$$\mathbf{\Pi}_{k,i} = \partial_i \mathbf{F}_k \, \mathbf{V}_k \, \mathbf{\Phi}_k^\top + \mathbf{\Phi}_k \, \mathbf{V}_k \, \partial_i \mathbf{F}_k^\top + \partial_i \mathbf{G}_k \, \mathbf{G}_k^\top + \mathbf{G}_k \, \partial_i \mathbf{G}_k^\top.$$

Note that, because given matrices are periodic, $\partial_i \mathbf{V}_{k+\rho} = \partial_i \mathbf{V}_k$. Then, iterate $\rho - 1$ times with (6.6), set $\partial_i \mathbf{V}_{k+\rho} = \partial_i \mathbf{V}_k$, and vectorize the result, to obtain

$$[\mathbf{I}_{n^2} - (\bar{\mathbf{\Phi}}_k \otimes \bar{\mathbf{\Phi}}_k)] \, \text{vec}(\partial_i \mathbf{V}_k) = \text{vec}(\mathbf{\Pi}_{k+\rho,i}), \tag{6.12}$$

for $k = 1, \ldots, \rho$ and $i = 1, \ldots, \alpha$. Then, solve (6.12) for $\mathrm{vec}(\partial_i \mathbf{V}_k)$ and compute $\partial_i \mathbf{M}_k$ and $\partial_i \mathbf{K}_k$ with (6.7) and (6.8), for $k = 1, \ldots, \rho$ and $i = 1, \ldots, \alpha$. Finally, set $\nabla \mathbf{M}_k = [\mathrm{vec}(\partial_1 \mathbf{V}_{k,11}), \ldots, \mathrm{vec}(\partial_\alpha \mathbf{V}_{k,11})]$, for $k = 1, \ldots, \rho$, where $\partial_i \mathbf{V}_{k,11}$ is the (1,1)-block of size $m \times m$ of $\partial_i \mathbf{V}_k$. Asymptotic stability of the Kalman filter ensures that $\bar{\mathbf{\Phi}}_k$ is a stability matrix, hence, that $[\mathbf{I}_{n^2} - (\bar{\mathbf{\Phi}}_k \otimes \bar{\mathbf{\Phi}}_k)]$ is invertible, for $k = 1, \ldots, \rho$.

6.3. Computation of $\mathbf{\Gamma}_k$

First, the periodic counterpart of (3.8) is needed and is set up as follows. For $k = 1, \ldots, \rho$, compute $\mathbf{\Lambda}_{k+\rho}$ by iterating with

$$\mathbf{\Lambda}_{\nu+1} = \mathbf{F}_{\nu+1}\, \mathbf{\Lambda}_\nu\, \mathbf{F}_{\nu+1}^\top + \mathbf{G}_{\nu+1}\, \mathbf{G}_{\nu+1}^\top, \qquad \text{for } \nu = k, \ldots, k+\rho-1, \tag{6.13}$$

starting with $\mathbf{\Lambda}_k = \mathbf{G}_k\, \mathbf{G}_k^\top$. Then, the periodic counterpart of (3.8) is

$$\mathbf{C}_k - \bar{\mathbf{F}}_k\, \mathbf{C}_k\, \bar{\mathbf{F}}_k^\top = \mathbf{\Lambda}_{k+\rho}, \qquad \text{for } k = 1, \ldots, \rho. \tag{6.14}$$

Stationarity of the model ensures that (6.14) yields a unique positive semidefinite solution of $\mathbf{C}_k$.

Note that in the periodic case, in recursive computation of $L(N)$, $\mathbf{I}_\theta(N)$, and $\mathbf{I}_\theta^\infty(N)$ according to Sections 3 and 4, $\mathbf{V}(1)$ should be set to the value of $\mathbf{C}_k$ which solves (6.14) for the value of k corresponding to $t = 1$. Similarly, in this case, $\nabla \mathbf{V}(1)$ should be set to the value of $\nabla \mathbf{C}_k$ which solves the periodic analogue of (4.8), derived from (6.14), for the value of k corresponding to $t = 1$.

To compute $\mathbf{\Gamma}_k$, first, for $k = 1, \ldots, \rho$, solve (6.14) with any method for solving (3.8) and set $\mathbf{S}_k = \mathbf{C}_k - \mathbf{V}_k$. Then, obtain an equation for $\mathrm{vec}(\mathbf{U}_{k,i})$ analogous to (6.12). For $k = 1, \ldots, \rho$, and $i = 1, \ldots, \alpha$, compute $\mathbf{\Upsilon}_{k+\rho,i}$ by iterating with

$$\mathbf{\Upsilon}_{\nu+1,i} = \mathbf{\Phi}_{\nu+1}\, \mathbf{\Upsilon}_{\nu,i}\, \mathbf{F}_{\nu+1}^\top + \partial_i \mathbf{F}_{\nu+1}\, \mathbf{S}_{\nu+1}\, \mathbf{F}_{\nu+1}^\top + \partial_i \mathbf{K}_{\nu+1}\, \mathbf{M}_{\nu+1}\, \mathbf{K}_{\nu+1}^\top, \tag{6.15}$$

for $\nu = k, \ldots, k+\rho-1$, starting with $\mathbf{\Upsilon}_{k,i} = \partial_i \mathbf{F}_k\, \mathbf{S}_k\, \mathbf{F}_k^\top + \partial_i \mathbf{K}_k\, \mathbf{M}_k\, \mathbf{K}_k^\top$. Then, iterate $\rho - 1$ times with (6.9), set $\mathbf{U}_{k+\rho,i} = \mathbf{U}_{k,i}$, and vectorize the result, to obtain

$$[\mathbf{I}_{n^2} - (\bar{\mathbf{F}}_k \otimes \bar{\mathbf{\Phi}}_k)]\, \mathrm{vec}(\mathbf{U}_{k,i}) = \mathrm{vec}(\mathbf{\Upsilon}_{k+\rho,i}). \tag{6.16}$$

Then, for $k = 1, \ldots, \rho$ and $i = 1, \ldots, \alpha$, solve (6.16) for $\mathrm{vec}(\mathbf{U}_{k,i})$. Stationarity of the model and asymptotic stability of the Kalman filter ensure that $[\mathbf{I}_{n^2} - (\bar{\mathbf{F}}_k \otimes \bar{\mathbf{\Phi}}_k)]$ is invertible, for $k = 1, \ldots, \rho$.

Next, obtain an analogous equation for $\mathrm{vec}(\mathbf{Z}_{k,ij})$. For $k = 1, \ldots, \rho$, $i = 1, \ldots, \alpha$, and $j = 1, \ldots, i$, compute $\mathbf{\Psi}_{k+\rho,ij}$ by iterating with

$$\begin{aligned}\mathbf{\Psi}_{\nu+1,ij} = {} & \mathbf{\Phi}_{\nu+1}\, \mathbf{\Psi}_{\nu,ij}\, \mathbf{\Phi}_{\nu+1}^\top + \partial_i \mathbf{F}_{\nu+1}\, \mathbf{S}_{\nu+1}\, \partial_j \mathbf{F}_{\nu+1}^\top + \mathbf{\Phi}_{\nu+1}\, \mathbf{U}_{\nu+1,i}\, \partial_j \mathbf{F}_{\nu+1}^\top \\ & + \partial_i \mathbf{F}_{\nu+1}\, \mathbf{U}_{\nu+1,j}^\top\, \mathbf{\Phi}_{\nu+1}^\top + \partial_i \mathbf{K}_{\nu+1}\, \mathbf{M}_{\nu+1}\, \partial_j \mathbf{K}_{\nu+1}^\top,\end{aligned} \tag{6.17}$$

for $\nu = k, \ldots, k+\rho-1$, starting with

$$\mathbf{\Psi}_{k,ij} = \partial_i \mathbf{F}_k\, \mathbf{S}_k\, \partial_j \mathbf{F}_k^\top + \mathbf{\Phi}_k\, \mathbf{U}_{k,i}\, \partial_j \mathbf{F}_k^\top + \partial_i \mathbf{F}_k\, \mathbf{U}_{k,j}^\top\, \mathbf{\Phi}_k^\top + \partial_i \mathbf{K}_k\, \mathbf{M}_k\, \partial_j \mathbf{K}_k^\top.$$

Then, iterate $\rho - 1$ times with (6.10), set $\mathbf{Z}_{k+\rho,ij} = \mathbf{Z}_{k,ij}$, and vectorize the result, to obtain

$$[\mathbf{I}_{n^2} - (\bar{\mathbf{\Phi}}_k \otimes \bar{\mathbf{\Phi}}_k)]\, \mathrm{vec}(\mathbf{Z}_{k,ij}) = \mathrm{vec}(\mathbf{\Psi}_{k+\rho,ij}). \tag{6.18}$$

Then, for $k = 1, \ldots \rho$, $i = 1, \ldots, \alpha$, and $j = 1, \ldots, i$, solve (6.18) for $\mathrm{vec}(\mathbf{Z}_{k,ij})$. Kalman filter asymptotic stability ensures invertibility of $[\mathbf{I}_{n^2} - (\bar{\mathbf{\Phi}}_k \otimes \bar{\mathbf{\Phi}}_k)]$. Finally, for $k = 1, \ldots, \rho$, $i = 1, \ldots, \alpha$, and $j = 1, \ldots, i$, set $\mathbf{\Gamma}_{k,ij} = \mathbf{Z}_{k,ij,11}$, the (1,1)-block of size $m \times m$ of $\mathbf{Z}_{k,ij}$.

7. CONCLUDING REMARKS

Some simple extensions and modifications make the recursive method of Section 4, for computing sample and asymptotic information matrices, applicable to generally time-varying models and samples. First, the constant state-space coefficient matrices, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{D}$, and $\mathbf{\Sigma}_\zeta$, must be written everywhere as time varying, i.e., as $\mathbf{F}(t)$, $\mathbf{G}(t)$, $\mathbf{D}(t)$, and $\mathbf{\Sigma}_\zeta(t)$. In general, the observation-selection matrix, $\mathbf{D}(t)$, may also depend on $\boldsymbol{\theta}$. This will occur, for example, if the model has simultaneous-equations components. If so, then, the computational equations must be extended in the obvious way to include derivatives of $\mathbf{D}(t)$ [5,6,8]. Also, if the model varies nonperiodically, then, $\mathbf{V}(1)$ and $\nabla\mathbf{V}(1)$ must be set by means other than solutions of Lyapunov equations. For example, Ansley and Kohn's method [12] could be applied to obtain, in effect, $\mathbf{V}(1) = \lim_{c\to\infty} c\mathbf{I}_n$ and $\nabla\mathbf{V}(1) = \mathbf{0}$.

The nonrecursive methods of Sections 5 and 6, for computing asymptotic information matrices in time-invariant and periodic models and samples reveal that stationarity of the model and asymptotic stability of the Kalman filter are sufficient conditions for the existence of the asymptotic information matrix. Although, in theory, invertibility of the moving-average part of the model is necessary only to ensure asymptotic normality of parameter estimates, in practice, invertibility is also necessary in computation of the likelihood function and its gradient, in order to prevent computed innovations from becoming too large (overflowing) at some iteration. However, if one is only computing information matrices recursively or nonrecursively and, in particular, is not computing innovations, then, noninvertibility should not cause numerical difficulties.

The paper has derived results only in partial-derivative-form equations, i.e., at the level of ∂_i expressions of derivatives. All of the results can also be derived in gradient-form equations, i.e., at the level of ∇ expressions of derivatives. This was not done because, in the present context, gradient-form equations require many more steps to be derived, are more difficult to implement, are computationally much less efficient, and provide no compensating insights. Gradient-form equations are more difficult to implement because they require numerous permutations of matrix elements, which are absent from the partial-derivative-form equations, and are computationally much less efficient because they replace many ordinary matrix products with much sparser Kronecker matrix products.

REFERENCES

1. Y. Hosoya and M. Taniguchi, A central limit theorem for stationary processes and the parameter estimation of linear processes, *Annals of Statistics* **10**, 132–153 (1982).
2. P.E. Caines, *Linear Stochastic Systems*, Wiley, New York, (1988).
3. G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden-Day, San Francisco, (1976).
4. E.R. Berndt, B.H. Hall, R.E. Hall, and J.A. Hausman, Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* **3/4**, 653–665 (1974).
5. P.A. Zadrozny, Analytic derivatives for estimation of linear dynamic models, In *System-Theoretic Methods in Economic Modelling,* (Edited by S. Mittnik), Special Issue of *Computers Math. Applic.* **18** (6/7), 539–553 (1989).
6. P.A. Zadrozny, Errata to 'Analytic derivatives for estimation of linear dynamic models', *Computers Math. Applic.* **24** (8/9), 289–290 (1992).
7. B. Porat and B. Friedlander, Computation of the exact information matrix of gaussian time series with stationary random components, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-34**, 118–130 (1986).
8. J. Terceiro Lomba, Estimation of dynamic econometric models with errors in variables, *Lecture Notes in Economics and Mathematical Systems*, No. 339, Springer-Verlag, New York, (1990).
9. G.C. Tiao and M.R. Grupe, Hidden periodic autoregressive-moving average models in time series data, *Biometrika* **67**, 365–373 (1980).
10. R.M. Todd, Periodic linear-quadratic methods for modelling seasonality, *Journal of Economic Dynamics and Control* **14**, 763–795 (1990).
11. P.A. Zadrozny, Estimating a multivariate ARMA model with mixed-frequency data: An application to forecasting U.S. GNP at monthly intervals, Working Paper No. 90-6, Research Department, Federal Reserve Bank of Atlanta, (1990).

12. C.F. Ansley and R. Kohn, Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions, *Annals of Statistics* **13**, 1286–1316 (1985).
13. M. Verhaegen and P. Van Dooren, Numerical aspects of different kalman filter implementations, *IEEE Transactions on Automatic Control* **AC-31**, 907–917 (1986).
14. R.H. Jones, Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics* **22**, 389–395 (1980).
15. B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, (1979).
16. C.F.Ansley and R. Kohn, Exact likelihood of vector autoregressive moving-average process with missing or aggregated data, *Biometrika* **70**, 275–278 (1983).
17. J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York, (1988).
18. G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, (1983).
19. S. Mittnik, On efficient exact maximum likelihood estimation of higher-order ARMA models, In *Computational Methods in Economics and Econometrics*, pp. 31–42, (Edited by H.M. Amman, D.A. Belsley, and L.F. Pau), Kluwer, Dordrecht, (1992).
20. S. Mittnik, Computation of the theoretical autocovariance matrices of multivariate ARMA time series, *Journal of the Royal Statistical Society, Series B* **52**, 151–155 (1990).
21. S. Mittnik, Computing theoretical autocovariances of multivariate autoregressive moving-average models by using a block-Levinson method, *Journal of the Royal Statistical Society, Series B* **55**, 435–440 (1993).
22. H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, Wiley-Interscience, New York, (1972).
23. P.A. Zadrozny, Analytic derivatives for estimation of discrete-time linear-quadratic dynamic optimization models, *Econometrica* **56**, 467–472 (1988).
24. A.J. Laub, A Schur method for solving algebraic Riccati equations, *IEEE Transactons on Automatic Control* **AC-24**, 913–921 (1979).
25. J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, (1983).