The Annals of Statistics
1998, Vol. 26, No. 4, 1636–1650

COMPUTATION OF THE EXACT INFORMATION MATRIX OF GAUSSIAN DYNAMIC REGRESSION TIME SERIES MODELS

BY ANDRÉ KLEIN, GUY MÉLARD AND TOUFIK ZAHAF

University of Amsterdam, Université Libre de Bruxelles and Université Libre de Bruxelles

In this paper, the computation of the exact Fisher information matrix of a large class of Gaussian time series models is considered. This class, which is often called the single-input–single-output (SISO) model, includes dynamic regression with autocorrelated errors and the transfer function model, with autoregressive moving average errors. The method is based on a combination of two computational procedures: recursions for the covariance matrix of the derivatives of the state vector with respect to the parameters, and the fast Kalman filter recursions used in the evaluation of the likelihood function. It is much faster than existing procedures. An expression for the asymptotic information matrix is also given.

1. Introduction. This paper is devoted to the exact Fisher information matrix of a time series $\{y_t;\, t = 1, \ldots, N\}$ generated by a single-input–single-output (SISO) process defined by the equation

$$\frac{\alpha(L)}{\beta(L)} y_t = \frac{\omega(L)}{\delta(L)} x_t + \frac{\theta(L)}{\phi(L)} e_t, \tag{1}$$

where x_t is an explanatory variable, and the e_t are normally and independently distributed random variables with mean zero and constant variance σ^2. Let L be the lag operator. The model depends on σ^2 and on $d = n+m+s+r+q+p+1$ parameters, which are the coefficients of the polynomials $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_n L^n$, $\beta(L) = 1 - \beta_1 L - \cdots - \beta_m L^m$, $\omega(L) = \omega_0 - \omega_1 L - \cdots - \omega_s L^s$, $\delta(L) = 1 - \delta_1 L - \cdots - \delta_r L^r$, $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$ and $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, stored in a $d \times 1$ vector λ, in the specified order. Denoting transposition by superscript T, $\lambda^T = (\alpha^T, \beta^T, \omega^T, \delta^T, \theta^T, \phi^T)$, where $\alpha^T = (\alpha_1, \ldots, \alpha_n)$, $\beta^T = (\beta_1, \ldots, \beta_m)$, $\omega^T = (\omega_0, \omega_1, \ldots, \omega_s)$, $\delta^T = (\delta_1, \ldots, \delta_r)$, $\theta^T = (\theta_1, \ldots, \theta_q)$, and $\phi^T = (\phi_1, \ldots, \phi_p)$. The model does not need to be used with all the polynomials, but it includes large classes of models: the autoregressive moving average (ARMA) model ($n = m = r = s = 0$), the ARMAX model ($m = r = p = 0$), the regression model with autocorrelated errors ($m = r = 0$) and the transfer function model ($n = m = 0$). The first order derivative of a scalar with respect to a column vector such as λ will be represented as a column vector.

Let $l(\lambda)$ be the likelihood function of the sample. The information matrix

$$J = -E\left(\frac{\partial^2 \log l}{\partial \lambda\, \partial \lambda^T}\right),$$

Received February 1997; revised December 1997.
AMS 1991 *subject classifications*. Primary 62M10; secondary 62F15.
Key words and phrases. Information matrix, SISO models, Chandrasekhar equations.

evaluated at the true unknown value of λ, is useful for obtaining the Cramér–Rao (lower) bound (CRB) of the estimated parameter vector $\hat{\lambda}$. Hence, a good estimate of its asymptotic covariance matrix is J^{-1}, assuming that the estimation method yields asymptotically efficient estimators. Most practitioners rely on the observed asymptotic information

$$J = -\frac{\partial^2 \log l}{\partial\lambda\, \partial\lambda^T},$$

obtained numerically within the optimization procedure that gives the estimates. The usefulness of the covariance matrix of the estimated parameters can be (a) experimental design, for example, determination of the number of observations for achieving a given accuracy [Dharan(1985)] or (b) Wald tests for the parameters, including tests for restrictions and the problem of zero pole cancellation [Klein and Spreij (1996)].

Obtaining algorithms for computing J has attracted attention in the statistical literature [e.g., Godolphin and Unwin (1983); Klein and Mélard (1989, 1990)], in engineering [e.g., Friedlander (1984); Pham (1989)], mainly for ARMA models, and also recently in econometrics. There have been some extensions to wider classes of models such as the vector ARMA or VARMA model [Newton (1978)], the SISO model [Klein and Mélard (1994a)], the multiple-input–single-output (MISO) model [Klein and Mélard (1994b)] and even VARMA models with complex linear restrictions [Mittnik and Zadrozny (1993)]. Many authors have used an approximation of the Gaussian log-likelihood based on the innovations sum of squares. This contrasts with the current practice of using the exact Gaussian likelihood, motivated by experiments of Ansley and Newbold (1980) and others.

It is well known that the *asymptotic* information matrix can be obtained using second order properties of the process. After the pioneering paper of Whittle (1953), there were attempts to obtain closed forms. The common point of most of these algorithms is to rely on the evaluation of covariances between two processes built on the same white noise process, either using a direct approach (more or less equivalent to an Euclidean algorithm) or using the evaluation of integrals of a rational function over the unit circle of the complex plane. In Klein and Mélard (1994a), the SISO model is considered, but the explanatory variable x_t is random, which forces us to select an ARMA specification for it. In this paper, an alternative and much simpler expression is given, conditionally with respect to x_t, in accordance with the general practice for inference in regression models.

We are mainly interested in this paper in the *exact* information matrix J. Porat and Friedlander (1986) have given a procedure for a model defined by

$$y_t = m_t + \frac{\theta(L)}{\phi(L)} e_t,$$

where m_t is a deterministic sequence. Their algorithm needs, however, a number of operations proportional to N^2. This can be quite expensive when N is large.

The problem has also been solved for a larger variety of models by Zadrozny (1989, 1992) and Terceiro (1990), using a state space formulation, but the closed form recurrences are given for each element of the information matrix, not for the information matrix as a whole, and the algorithms used are not the most efficient. For the SISO model, our approach is better because the specific parametrization is taken into account, the algorithm implies a smaller number of operations, and the given matrix recurrences produce the whole information matrix. These three points are discussed more thoroughly in the conclusion.

The algorithm stated in this paper is an expanded and improved version of a procedure sketched by Mélard and Klein (1994), where it was given for a stationary Gaussian (ARMA) process z_t of order (p, q) defined by

$$\phi(L)z_t = \theta(L)e_t.$$

A much more general model is used with complete and detailed closed form recurrences and their initial values are presented in a more elegant way. Furthermore, both the asymptotic and exact information matrices are considered.

The algorithm for the exact information matrix needs only a number of operations proportional to N, that is, an order of magnitude less than Porat and Friedlander (1986), by relying on a state space representation, recursions for the covariance matrix of the derivatives of the state vector with respect to the parameters and the fast Kalman filter recursions used in the evaluation of the likelihood function of a Gaussian ARMA time series [e.g., Mélard (1984)] and its derivatives [Mélard (1985); Kohn and Ansley (1985)]. Of course, there are six polynomials instead of only two.

This article is organized as follows. In the next section we formulate the model. In Section 3, a general expression for the information matrix for SISO models is given. The asymptotic version of the information matrix is described in Section 4, whereas in Section 5 the recurrence equations necessary for computing the exact information matrix for SISO models are provided. The conclusion is found in Section 6.

2. The model The SISO model defined in (1) can be written as

$$\alpha(L)\delta(L)\phi(L)y_t = \omega(L)\beta(L)\phi(L)x_t + \theta(L)\beta(L)\delta(L)e_t. \tag{2}$$

Let us denote m_t by

$$m_t = \frac{\beta(L)\omega(L)}{\alpha(L)\delta(L)}x_t \tag{3}$$

and w_t by

$$w_t = y_t - m_t = \frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}e_t. \tag{4}$$

Let us first suppose that data are available for y_t and x_t for $t = 1, \dots, N$. Therefore, m_t can be computed for $t > \max(m + s, n + r)$. In order to simplify notation, we shall rather suppose that y_t, m_t and x_t are available, respectively, for $t > 0$, $t > -(n+r)$ and $t > -(m+s)$.

3. The general expression for the exact information matrix. Let us denote by y, m and w the vectors composed, respectively, of y_t, m_t and w_t, $t = 1, \dots, N$. The log-likelihood from time 1 to time N can be written in the form

$$\log l = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(\det\Gamma) - \frac{1}{2}w^T\Gamma^{-1}w, \tag{5}$$

where Γ is the covariance matrix of the zero mean vector w. The element (i, j) of the exact information matrix J can be written as [see Porat and Friedlander (1986)]

$$J_{ij} = \frac{1}{2}\mathrm{Tr}\left\{\Gamma^{-1}\frac{\partial\Gamma}{\partial\lambda_i}\Gamma^{-1}\frac{\partial\Gamma}{\partial\lambda_j}\right\} + \left[\frac{\partial m}{\partial\lambda_i}\right]^T\Gamma^{-1}\left[\frac{\partial m}{\partial\lambda_j}\right]. \tag{6}$$

The algorithm of Porat and Friedlander (1986) makes use of the Levinson–Durbin algorithm for computing the orthogonal polynomials of a Toeplitz matrix.

We consider the sequence of Euclidean spaces $\mathscr{W}_t$ spanned by $\{w_1, w_2, \dots, w_t\}$, $t = 1, \dots, N$, with the covariance as the scalar product. The orthogonal projection of w_t in the subspace $\mathscr{W}_{t-1}$ is denoted by $\hat{w}_t$. Let $\hat{a}_t$ be the difference between w_t and $\hat{w}_t$, which is orthogonal to $\mathscr{W}_{t-1}$. It is called the sample innovation at time t. Let $h_t\sigma$ be the standard deviation of $\hat{a}_t$ and let the normalized sample innovation be $\hat{e}_t = \hat{a}_t/h_t$, with mean zero and variance σ^2. The $\hat{e}_t$ and the h_t can be obtained by the Gram–Schmidt orthogonalization procedure or any procedure which yields equivalent results. For example, the Cholesky factorization can be used instead, especially in the case where the covariance matrix is a band matrix (which corresponds to a pure moving average process). For a suitably specified model, the Kalman filter is also well adapted. These algorithms are computationally more efficient than the Gram–Schmidt procedure.

The likelihood function is built as the density of w. Equivalently, it can be written as the density of the vector $\hat{e}$ with general element $\hat{e}_t$, multiplied by the Jacobian of the transformation, which is $\prod_{t=1}^N h_t$. Hence the log-likelihood from time 1 to time N can be written in the form

$$\log l = -\frac{N}{2}\log(2\pi) - N\log\sigma - \sum_{t=1}^N \log h_t - \frac{1}{2}\sum_{t=1}^N \frac{\hat{e}_t^2}{\sigma^2}.$$

The information matrix is equal to minus the mathematical expectation of the matrix of second derivatives of the log-likelihood,

$$\frac{\partial^2 \log l}{\partial\lambda\,\partial\lambda^T} = -\sum_{t=1}^N \frac{1}{h_t}\frac{\partial^2 h_t}{\partial\lambda\,\partial\lambda^T} + \sum_{t=1}^N \frac{1}{h_t^2}\frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda^T} - \frac{1}{\sigma^2}\sum_{t=1}^N \frac{\partial\hat{e}_t}{\partial\lambda}\frac{\partial\hat{e}_t}{\partial\lambda^T} - \frac{1}{\sigma^2}\sum_{t=1}^N \hat{e}_t\frac{\partial^2\hat{e}_t}{\partial\lambda\,\partial\lambda^T}.$$

It can be written as

$$
\begin{aligned}
-E\left(\frac{\partial^2 \log l}{\partial\lambda\,\partial\lambda^T}\right) &= \sum_{t=1}^{N}\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\lambda\,\partial\lambda^T} - \sum_{t=1}^{N}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda^T} \\
&\quad + \frac{1}{\sigma^2}\sum_{t=1}^{N}E\left(\frac{\partial\hat{e}_t}{\partial\lambda}\frac{\partial\hat{e}_t}{\partial\lambda^T}\right) + \frac{1}{\sigma^2}\sum_{t=1}^{N}E\left(\hat{e}_t\frac{\partial^2\hat{e}_t}{\partial\lambda\,\partial\lambda^T}\right).
\end{aligned} \tag{7}
$$

The sequel of this section is devoted to evaluating the last term of the right-hand side of (7) and to obtaining a closed expression for the third term.

From (4), w_t is an ARMA$(n+p, m+q)$ process. We consider the decomposition of w_t onto $\mathscr{W}_{t-1}$ using the $\hat{e}_s$, $s \le t$, as an orthogonal basis. For time $t \ge 2$ we have

$$h_t\hat{e}_t + \hat{w}_t = w_t, \tag{8}$$

whereas $w_1 = h_1\hat{e}_1$. Hence, differentiation of (8) yields

$$\frac{\partial h_t}{\partial\lambda}\hat{e}_t + h_t\frac{\partial\hat{e}_t}{\partial\lambda} + \frac{\partial\hat{w}_t}{\partial\lambda} = \frac{\partial w_t}{\partial\lambda} \tag{9}$$

$$= -\frac{\partial m_t}{\partial\lambda}, \tag{10}$$

because of (4). Differentiating a second time gives

$$\frac{\partial^2 h_t}{\partial\lambda\,\partial\lambda^T}\hat{e}_t + h_t\frac{\partial^2\hat{e}_t}{\partial\lambda\,\partial\lambda^T} + 2\frac{\partial h_t}{\partial\lambda}\frac{\partial\hat{e}_t}{\partial\lambda^T} + \frac{\partial^2\hat{w}_t}{\partial\lambda\,\partial\lambda^T} = \frac{\partial^2 w_t}{\partial\lambda\,\partial\lambda^T}. \tag{11}$$

The inference being conditional on the regressor variable, m_t defined by (3) is not considered as a random variable, and since the normalized sample innovations have zero mean, we deduce from (10) that

$$\frac{\partial h_t}{\partial\lambda}E(\hat{e}_t^2) + h_tE\left(\frac{\partial\hat{e}_t}{\partial\lambda}\hat{e}_t\right) + E\left(\frac{\partial\hat{w}_t}{\partial\lambda}\hat{e}_t\right) = 0.$$

Since $\partial\hat{w}_t/\partial\lambda \in \mathscr{W}_{t-1}$,

$$h_tE\left(\frac{\partial\hat{e}_t}{\partial\lambda}\hat{e}_t\right) = -\frac{\partial h_t}{\partial\lambda}\sigma^2 \tag{12}$$

and, similarly from (11) and (12),

$$
\begin{aligned}
E\left(\hat{e}_t\frac{\partial^2\hat{e}_t}{\partial\lambda\,\partial\lambda^T}\right) &= -\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\lambda\,\partial\lambda^T}\sigma^2 - \frac{2}{h_t}\frac{\partial h_t}{\partial\lambda}E\left(\frac{\partial\hat{e}_t}{\partial\lambda^T}\hat{e}_t\right) \\
&= \sigma^2\left(-\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\lambda\,\partial\lambda^T} + 2\frac{1}{h_t^2}\frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda^T}\right).
\end{aligned} \tag{13}
$$

Summarizing (7) and (13), we have

$$-E\left(\frac{\partial^2 \log l}{\partial\lambda\,\partial\lambda^T}\right) = \sum_{t=1}^{N}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda^T} + \frac{1}{\sigma^2}\sum_{t=1}^{N}E\left(\frac{\partial\hat{e}_t}{\partial\lambda}\frac{\partial\hat{e}_t}{\partial\lambda^T}\right). \tag{14}$$

Note that $\partial h_t/\partial\delta$ and $\partial h_t/\partial\omega$ are equal to zero since the parameters of the transfer function $\omega(L)/\delta(L)$ play no role in the error model, and that second order derivatives do not appear in the final expression.

We are now left to evaluate the expectation of the product $(\partial\hat{e}_t/\partial\lambda)(\partial\hat{e}_t/\partial\lambda^T)$ as a function of $(\partial\hat{w}_t/\partial\lambda)(\partial\hat{w}_t/\partial\lambda^T)$ for each t using an equation which can be deduced from (9). Indeed, since (10) and the fact that the projections $\partial\hat{w}_t/\partial\lambda$ have zero mean,

$$(15)\qquad \begin{aligned} h_t^2 E\left(\frac{\partial\hat{e}_t}{\partial\lambda}\frac{\partial\hat{e}_t}{\partial\lambda^T}\right) &= E\left(\left[\frac{\partial\hat{w}_t}{\partial\lambda} + \frac{\partial h_t}{\partial\lambda}\hat{e}_t + \frac{\partial m_t}{\partial\lambda}\right]\left[\frac{\partial\hat{w}_t}{\partial\lambda^T} + \frac{\partial h_t}{\partial\lambda^T}\hat{e}_t + \frac{\partial m_t}{\partial\lambda^T}\right]\right) \\ &= E\left(\frac{\partial\hat{w}_t}{\partial\lambda}\frac{\partial\hat{w}_t}{\partial\lambda^T}\right) + \frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda^T}\sigma^2 + \frac{\partial m_t}{\partial\lambda}\frac{\partial m_t}{\partial\lambda^T}. \end{aligned}$$

There remains to obtain recurrence equations for $\partial m_t/\partial\lambda$ and for $E((\partial\hat{w}_t/\partial\lambda)(\partial\hat{w}_t/\partial\lambda^T))$. This is done in the Appendix for the former and in Section 5 for the latter. In Section 4 we obtain a simpler expression for the asymptotic case.

4. The asymptotic information matrix. As noted earlier, there is no published procedure for the asymptotic information matrix for the general SISO model except in Klein and Mélard (1994a), where the regressor is assumed to be stochastic. For the sake of comparison with the exact information matrix, we shall briefly adapt the technique described there to the case of inference conditional to the regressor.

Let us differentiate (4) with respect to λ, giving

$$(16)\qquad -\frac{\partial m_t}{\partial\lambda} - \left(\partial\left(\frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}\right)\Big/\partial\lambda\right)e_t - \frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}\frac{\partial e_t}{\partial\lambda} = 0;$$

hence

$$(17)\qquad \frac{\partial e_t}{\partial\lambda} = -\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)}\frac{\partial m_t}{\partial\lambda} - \left(\partial\left(\frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}\right)\Big/\partial\lambda\right)\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)}e_t$$

and, noting that $\widehat{e_t}$ and e_t are asymptotically equivalent,

$$(18)\qquad \begin{aligned} E\left(\frac{\partial\hat{e}_t}{\partial\lambda}\frac{\partial\hat{e}_t}{\partial\lambda^T}\right) = E\Bigg[&\left\{\left(\partial\left(\frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}\right)\Big/\partial\lambda\right)\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)}e_t\right\} \\ &\times\left\{e_t^T\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)}\left(\partial\left(\frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}\right)\Big/\partial\lambda\right)^T\right\}\Bigg] \\ &+\left(\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)}\frac{\partial m_t}{\partial\lambda}\right)\left(\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)}\frac{\partial m_t}{\partial\lambda}\right)^T, \end{aligned}$$

which should be introduced in (14).

Let us consider the first term, since an equation for $\partial m_t/\partial\lambda$ has been obtained in the Appendix. The derivatives are easily simplified. For instance, the ith element of

$$\frac{\partial}{\partial \phi^T}\left(\frac{\beta(L)\theta(L)}{\alpha(L)\phi(L)}\right)\frac{\alpha(L)\phi(L)}{\beta(L)\theta(L)} = -\frac{1}{\phi(L)}\frac{\partial \phi(L)}{\partial \phi^T}$$

is $\phi_i L^i/\phi(L)$ and the jth element of the derivative with respect to β^T is $-\beta_j L^j/\beta(L)$. The derivatives with respect to ω^T and δ^T are identically equal to zero. The element of the first term of the right-hand side of (18) corresponding to ϕ_i and β_j is given by

$$-\phi_i\beta_j E\big[\phi^{-1}(L)e_{t-i}\beta^{-1}(L)e_{t-j}\big], \tag{19}$$

which can be computed as the covariance at lag $(i-j)$ between two autoregressive processes with respective autoregressive polynomials $\phi(L)$ and $\beta(L)$ built on the same white noise process e_t with variance σ^2. See Klein and Mélard (1994a) for more details and Tunnicliffe Wilson (1979) for an algorithm.

All the elements of the first term of the asymptotic information matrix, established in (18), can be summarized in a general form, with the notation given in the Appendix, as

$$\frac{1}{2\pi i}\oint_\gamma M^{(\lambda)}(z)\frac{dz}{z},$$

where γ is the positively oriented unit circle and

$$M^{(\lambda)}(z)=\begin{bmatrix}
\frac{M_n(z^{-1})M_n^T(z)}{\alpha(z^{-1})\alpha(z)} & -\frac{M_n(z^{-1})M_m^T(z)}{\alpha(z^{-1})\beta(z)} & 0 & 0 & -\frac{M_n(z^{-1})M_q^T(z)}{\alpha(z^{-1})\theta(z)} & \frac{M_n(z^{-1})M_p^T(z)}{\alpha(z^{-1})\phi(z)}\\
-\frac{M_m(z^{-1})M_n^T(z)}{\beta(z^{-1})\alpha(z)} & \frac{M_m(z^{-1})M_m^T(z)}{\beta(z^{-1})\beta(z)} & 0 & 0 & -\frac{M_m(z^{-1})M_q^T(z)}{\beta(z^{-1})\theta(z)} & -\frac{M_m(z^{-1})M_p^T(z)}{\beta(z^{-1})\phi(z)}\\
0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0\\
-\frac{M_q(z^{-1})M_n^T(z)}{\theta(z^{-1})\alpha(z)} & \frac{M_q(z^{-1})M_m^T(z)}{\theta(z^{-1})\beta(z)} & 0 & 0 & \frac{M_q(z^{-1})M_q^T(z)}{\theta(z^{-1})\theta(z)} & -\frac{M_q(z^{-1})M_p^T(z)}{\theta(z^{-1})\phi(z)}\\
\frac{M_p(z^{-1})M_n^T(z)}{\phi(z^{-1})\alpha(z)} & -\frac{M_p(z^{-1})M_m^T(z)}{\phi(z^{-1})\beta(z)} & 0 & 0 & -\frac{M_p(z^{-1})M_q^T(z)}{\phi(z^{-1})\theta(z)} & \frac{M_p(z^{-1})M_p^T(z)}{\phi(z^{-1})\phi(z)}
\end{bmatrix}.$$

5. The state space approach. This section consists of an expanded and improved version of a procedure sketched in Mélard and Klein (1994). The objective is to obtain a recurrence for $E[(\partial\hat{w}_t/\partial\lambda)(\partial\hat{w}_t/\partial\lambda^T)]$. This will be done using a fast version of the Kalman filter [named after Chandrasekhar; see Morf, Sidhu and Kailath (1974)], which is well adapted to the time invariance of the model in state space form. The approach is generally used in fast algorithms for evaluating the exact likelihood function of an ARMA process. Since all the recurrences bear on vectors instead of on matrices in the usual Kalman filter, they are suited to differentiation. As a side-product, the $\partial h_t/\partial\lambda$, which are also needed in (15), will be produced. In the present case, w_t is an ARMA$(n+p, m+q)$ process with AR and MA respective polynomials $\rho(L)=\alpha(L)\phi(L)$ and $\mu(L)=\beta(L)\theta(L)$.

There are several nearly equivalent state space representations. We use the same representation as Pearlman (1980) and Mélard (1984), using a state vector W_t of dimension $g = \max\{n+p, m+q+1\}$ written as

$$w_t = HW_t,$$

$$W_{t+1} = FW_t + Ge_{t+1},$$

where $H = (1, 0, \dots, 0)$, $G^T = (1, -\mu_1, \dots, -\mu_g)$ and $F = (F_{i,j})$ is a $g \times g$ matrix such that $F_{i,1} = \rho_i$ and $F_{i,j} = \delta_{i,j-1}$ ($j = 2, \dots, g$) for $i = 1, \dots, g$, using Kronecker's δ and the convention that $\rho_i = 0$, $i > n+p$, $\mu_0 = -1$ and $\mu_i = 0$, $i > m+q$.

The vector $\hat{W}_t$ is defined by the projection of the elements of W_t onto $\mathscr{W}_{t-1}$. It can be computed using the Chandrasekhar recurrences

$$\hat{W}_t = F\hat{W}_{t-1} + (K_{t-1}/h_{t-1})\hat{e}_{t-1}, \tag{20}$$

$$\hat{w}_t = H\hat{W}_t, \tag{21}$$

$$h_t\hat{e}_t = w_t - \hat{w}_t, \tag{22}$$

$$\nu_{t-1} = HL_{t-1}/h_{t-1}^2, \tag{23}$$

$$K_t = K_{t-1} - \nu_{t-1}FL_{t-1}, \tag{24}$$

$$L_t = FL_{t-1} - \nu_{t-1}K_{t-1}, \tag{25}$$

$$h_t^2 = h_{t-1}^2(1 - \nu_{t-1}^2). \tag{26}$$

The initial conditions are the following. Denoting $\gamma_k = \operatorname{cov}(w_t, w_{t-k})/\sigma^2$ and $\tau_k = \operatorname{cov}(w_t, e_{t-k})/\sigma^2$ [see Mélard (1984) or Demeure and Mullis (1989) for algorithms], and

$$\psi_k = \sum_{j=k}^{g} (\rho_j \gamma_{j-k+1} - \mu_{j-1}\tau_{j-k})$$

and $\psi_{g+1} = 0$, then take $\hat{W}_1 = 0$, $h_1^2 = \gamma_0$ and $\rho_k\gamma_0 + \psi_{k+1}$ as the kth element of $K_1 = L_1$ [see Mélard (1984)].

We denote the derivative of a $g \times 1$ column vector z with respect to a $d \times 1$ column vector λ as the $gd \times 1$ column vector defined by

$$\left(\frac{\partial z}{\partial \lambda}\right)^T = \left(\left(\frac{\partial z_1^T}{\partial \lambda}\right)^T, \dots, \left(\frac{\partial z_g^T}{\partial \lambda}\right)^T\right).$$

That notation is not the one recommended in Magnus and Neudecker [(1988), Section 9.3-4], but is appropriate for the problem studied in this paper.

In order to be able to use recurrence equations, it is necessary to differentiate the initial conditions given above, including the covariances and taking care that $\hat{W}_1 = 0$ identically and thus $\partial\widehat{W}_1/\partial\lambda = 0$. For more details, see Mélard (1985).

The derivatives with respect to λ of the recurrences (20)–(26) are then written. The most delicate part is the derivative of the first term of (20), namely (omitting the subscript $t-1$ to simplify the notation),

$$\frac{\partial}{\partial\lambda}\begin{pmatrix} \rho_1\hat{w}+\hat{W}_2 \\ \rho_2\hat{w}+\hat{W}_3 \\ \vdots \\ \rho_{g-1}\hat{w}+\hat{W}_g \\ \rho_g\hat{w} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial\rho_1}{\partial\lambda}\hat{w}+\rho_1\dfrac{\partial\hat{w}}{\partial\lambda}+\dfrac{\partial\hat{W}_2}{\partial\lambda} \\ \dfrac{\partial\rho_2}{\partial\lambda}\hat{w}+\rho_2\dfrac{\partial\hat{w}}{\partial\lambda}+\dfrac{\partial\hat{W}_3}{\partial\lambda} \\ \vdots \\ \dfrac{\partial\rho_{g-1}}{\partial\lambda}\hat{w}+\rho_{g-1}\dfrac{\partial\hat{w}}{\partial\lambda}+\dfrac{\partial\hat{W}_g}{\partial\lambda} \\ \dfrac{\partial\rho_g}{\partial\lambda}\hat{w}+\rho_g\dfrac{\partial\hat{w}}{\partial\lambda} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{\partial\rho_1}{\partial\lambda} \\ \dfrac{\partial\rho_2}{\partial\lambda} \\ \vdots \\ \dfrac{\partial\rho_{g-1}}{\partial\lambda} \\ \dfrac{\partial\rho_g}{\partial\lambda} \end{pmatrix}\hat{w} + \begin{pmatrix} \rho_1\dfrac{\partial\hat{w}}{\partial\lambda}+\dfrac{\partial\hat{W}_2}{\partial\lambda} \\ \rho_2\dfrac{\partial\hat{w}}{\partial\lambda}+\dfrac{\partial\hat{W}_3}{\partial\lambda} \\ \vdots \\ \rho_{g-1}\dfrac{\partial\hat{w}}{\partial\lambda}+\dfrac{\partial\hat{W}_g}{\partial\lambda} \\ \rho_g\dfrac{\partial\hat{w}}{\partial\lambda} \end{pmatrix} = R\hat{w}+(F\otimes I_d)\frac{\partial\hat{W}}{\partial\lambda},$$

where R is the vector of derivatives of the first column of F with respect to λ, and $\otimes$ denotes the Kronecker product. Let $D = F\otimes I_d$ and $C_{t-1} = K_{t-1}/h_{t-1}^2$. Recall that $\hat{a}_t = h_t\hat{e}_t = w_t - \hat{w}_t$, so we obtain

$$\frac{\partial\hat{a}_t}{\partial\lambda} = \frac{\partial w_t}{\partial\lambda} - \frac{\partial\hat{w}_t}{\partial\lambda} = -\frac{\partial m_t}{\partial\lambda} - \frac{\partial\hat{w}_t}{\partial\lambda} = -\frac{\partial m_t}{\partial\lambda} - H\frac{\partial\hat{W}_t}{\partial\lambda}.$$

Hence, the derivative of (20) is

$$\frac{\partial\hat{W}_t}{\partial\lambda} = RH\hat{W}_{t-1} + D\frac{\partial\hat{W}_{t-1}}{\partial\lambda} + \frac{\partial C_{t-1}}{\partial\lambda}\hat{a}_{t-1} + C_{t-1}\otimes\frac{\partial\hat{a}_{t-1}}{\partial\lambda}. \tag{27}$$

The other derivatives are

$$\frac{\partial\nu_{t-1}}{\partial\lambda} = \frac{1}{h_{t-1}^2}\left(H\frac{\partial L_{t-1}}{\partial\lambda} - 2\frac{HL_{t-1}}{h_{t-1}^3}\frac{\partial h_{t-1}}{\partial\lambda}\right),$$

$$\frac{\partial K_t}{\partial\lambda} = \frac{\partial K_{t-1}}{\partial\lambda} - (FL_{t-1})\otimes\frac{\partial\nu_{t-1}}{\partial\lambda} - \nu_{t-1}\left\{RHL_{t-1} + D\frac{\partial L_{t-1}}{\partial\lambda}\right\},$$

$$\frac{\partial L_t}{\partial \lambda} = RHL_{t-1} + D\frac{\partial L_{t-1}}{\partial \lambda} - K_{t-1} \otimes \frac{\partial \nu_{t-1}}{\partial \lambda} - \nu_{t-1}\frac{\partial K_{t-1}}{\partial \lambda},$$

$$\frac{\partial h_t^2}{\partial \lambda} = \frac{\partial h_{t-1}^2}{\partial \lambda}(1-\nu_{t-1}^2) - 2h_{t-1}^2\nu_{t-1}\frac{\partial \nu_{t-1}}{\partial \lambda}.$$

We need also those expectations

$$E[\hat{a}_t^2] = h_t^2\sigma^2$$

and

$$E\left[\frac{\partial \hat{W}_t}{\partial \lambda}\right] = DE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\right] - C_{t-1} \otimes E\left[\frac{\partial \hat{w}_{t-1}}{\partial \lambda}\right] + C_{t-1} \otimes \left[\frac{\partial m_{t-1}}{\partial \lambda}\right].$$

Denoting $S = H \otimes I_d$ and noticing that $\hat{a}_{t-1}$ is not correlated with $\hat{W}_{t-1}$, $\partial \hat{W}_{t-1}/\partial\lambda$ and $\partial\widehat{a}_{t-1}/\partial\lambda$, we deduce the recurrence equations which are needed:

$$E(\hat{W}_t\hat{W}_t^T) = FE(\hat{W}_{t-1}\hat{W}_{t-1}^T)F^T + C_{t-1}C_{t-1}^T h_t^2\sigma^2, \tag{28}$$

$$\begin{aligned} E\left[\hat{W}_t\frac{\partial \hat{W}_t^T}{\partial \lambda^T}\right] &= FE\left[\hat{W}_{t-1}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]D^T + FE[\hat{W}_{t-1}\hat{W}_{t-1}^T]H^T R^T \\ &\quad - F\left\{C_{t-1}^T \otimes E\left[\hat{W}_{t-1}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]S^T\right\} \\ &\quad - F\left\{C_{t-1}^T \otimes E[\hat{W}_{t-1}]\frac{\partial m_{t-1}}{\partial \lambda^T}\right\} + C_{t-1}\frac{\partial C_{t-1}^T}{\partial \lambda^T}h_t^2\sigma^2, \end{aligned} \tag{29}$$

$$\begin{aligned} E\left[\frac{\partial \hat{W}_t}{\partial \lambda}\frac{\partial \hat{W}_t^T}{\partial \lambda^T}\right] &= RHE[\hat{W}_{t-1}\hat{W}_{t-1}^T]H^T R^T + DE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]D^T \\ &\quad + RHE\left[\hat{W}_{t-1}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]D^T + DE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\hat{W}_{t-1}^T\right]H^T R^T \\ &\quad + \{C_{t-1}C_{t-1}^T\} \otimes \Bigg\{HE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]H^T + \left[\frac{\partial m_{t-1}}{\partial \lambda}\frac{\partial m_{t-1}}{\partial \lambda^T}\right] \\ &\qquad + HE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\right]\frac{\partial m_{t-1}}{\partial \lambda^T} \\ &\qquad + \left(\frac{\partial m_{t-1}}{\partial \lambda}HE\left[\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]\right)\Bigg\} \\ &\quad - R\left[C_{t-1}^T \otimes \left\{HE[\hat{W}_{t-1}]\frac{\partial m_{t-1}}{\partial \lambda^T} + HE\left[\hat{W}_{t-1}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]S^T\right\}\right] \\ &\quad - \left[C_{t-1} \otimes \left\{SE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\hat{W}_{t-1}^T\right]H^T + \frac{\partial m_{t-1}}{\partial \lambda}E[\hat{W}_{t-1}^T]H^T\right\}\right]R^T \end{aligned} \tag{30}$$

$$
\begin{aligned}
&- D\left[C_{t-1}^T \otimes \left\{E\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]S^T + E\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\right]\frac{\partial m_{t-1}}{\partial \lambda^T}\right\}\right] \\
&- \left[C_{t-1} \otimes \left\{SE\left[\frac{\partial \hat{W}_{t-1}}{\partial \lambda}\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right] + \frac{\partial m_{t-1}}{\partial \lambda}E\left[\frac{\partial \hat{W}_{t-1}^T}{\partial \lambda^T}\right]\right\}\right]D^T \\
&+ \frac{\partial C_{t-1}}{\partial \lambda}\frac{\partial C_{t-1}^T}{\partial \lambda^T}h_t^2\sigma^2.
\end{aligned}
$$

In spite of the improved notation used here, (28)–(30) are much more complex than the corresponding recurrences in Mélard and Klein (1994) because terms in $\partial m_{t-1}/\partial\lambda$ which do not exist there are introduced in the new last term of (27).

6. Conclusion. We have discussed the asymptotic information matrix and the exact information matrix of a large class of Gaussian time series models. It should be stressed that the asymptotic information matrix is only an approximation, being related to the conditional likelihood function, not the exact likelihood function. We have given algorithms for both the asymptotic and exact information matrix as a whole instead of element by element.

Although the algorithm may seem more complex than the direct use of (6), the number of operations is obviously $O(N)$ instead of $O(N^3)$, and $O(N^2)$ for the algorithm of Porat and Friedlander (1986). Furthermore many of the operations in (30) can be avoided because many matrices of derivatives will be composed of 0's and 1's [see Zadrozny (1989), page 547]. In any case, there are at most $O(d^2g^2)$ operations at each t.

The algorithms of Terceiro (1990) and Zadrozny (1989) are slightly less efficient for several reasons. First, each element (i, j) of the information matrix is computed separately. In the case of Terceiro (1990), each of them is even computed using a specific form of the state space model where equations of the original state vector and of its derivatives with respect to λ_i and λ_j are stacked.

Second, we have used the Chandrasekhar recurrences [Morf, Sidhu and Kailath (1974)] instead of the Kalman filter used by both Zadrozny (1989) and Terceiro (1990), which implies a reduction of the complexity of the algorithm. Indeed the Chandrasekhar recurrences are computationally more efficient by an order of magnitude with respect to the Kalman filter. Very often fast procedures are also much more complex, but it is not the case here. On the contrary, the Chandrasekhar equations are also slightly simpler than the corresponding Kalman filter equations. If we consider the univariate ARMA(p, q) model and take $q = p$ to simplify the comparisons, the number of operations (multiplications and divisions) of the algorithm of Terceiro is of order $72p^4N$, whereas ours is of order $32p^3N$. Except for very high N, where the asymptotic information matrix may be enough, the action can be in the constants, but $72p^4$ will be more easily close to N than $32p^3$ [thereby coming close to the number of operations of the method of Porat and Friedlander (1986)]. One may add that the reason for keeping the computational burden as low as possible is that the information matrix can be needed in an iterative score method and therefore be invoked a large number of times by an optimization algorithm.

Third, both the algorithms of Terceiro (1990) and Zadrozny (1989) are described for a general state space model, in vector form and with a multivariate input variable, but the parametrization is not specified. The SISO model can of course be written under the state space form using (2), but the parameters which appear in the coefficients are products of three polynomials. In our approach, the whole matrix is computed in one run and takes the specific form of the model and parametrization into account. The proposed method can be easily implemented, using the purely matrix recurrences given in the paper, which are appropriate for a modern computer language with matrix support. Generalization to multiple input is straightforward.

Our matrix presentation can probably be improved in the future by computing Cholesky factors of the information matrix, thereby making it possible to check that it is strictly positive definite (as a check of model identification). Computation element by element makes this check impossible or at least unreliable because of rounding errors. Note that factorization of the *asymptotic* information matrix has already been exploited by Klein and Spreij (1996, 1997) for ARMAX and ARMA models, respectively.

APPENDIX

We assume that $\partial m_t/\partial\lambda = 0$ for $t \le 0$. Given the initial values for x_t and m_t for $t \le 0$, the subsequent values of $\partial m_t/\partial\lambda$ can be obtained by recurrence. By virtue of (3) we shall derive a recurrence relationship for $\partial m_t/\partial\lambda$ which yields

$$\begin{aligned}\frac{\partial m_t}{\partial\lambda} &= \sum_{i=1}^{n}\alpha_i\frac{\partial m_{t-i}}{\partial\lambda} + \sum_{j=1}^{r}\delta_j\frac{\partial m_{t-j}}{\partial\lambda} + \sum_{i=1}^{n}\sum_{j=1}^{r}\delta_j\alpha_i\frac{\partial m_{t-i-j}}{\partial\lambda}\\ &\quad - \sum_{i=1}^{n}m_{t-i}\frac{\partial\alpha_i}{\partial\lambda} - \sum_{j=1}^{r}m_{t-j}\frac{\partial\delta_j}{\partial\lambda} + \sum_{i=1}^{n}\sum_{j=1}^{r}\left(\frac{\partial\alpha_i}{\partial\lambda}\delta_j + \alpha_i\frac{\partial\delta_j}{\partial\lambda}\right)m_{t-i-j}\\ &\quad + \frac{\partial\omega_0}{\partial\lambda}x_t - \frac{\partial\omega_0}{\partial\lambda}\sum_{k=1}^{m}\beta_k x_{t-k} - \omega_0\sum_{k=1}^{m}\frac{\partial\beta_k}{\partial\lambda}x_{t-k} - \sum_{l=1}^{s}\frac{\partial\omega_l}{\partial\lambda}x_{t-l}\\ &\quad + \sum_{l=1}^{s}\sum_{k=1}^{m}\left(\frac{\partial\omega_l}{\partial\lambda}\beta_k + \omega_l\frac{\partial\beta_k}{\partial\lambda}\right)x_{t-l-k}.\end{aligned}$$

We give now the general form of each term. Let $\Lambda_{k,t}$ be an $(d\times k)$ matrix with $(\partial m_{t-j}/\partial\lambda_i)$ as element (i, j), $i = 1, \ldots, d$, $j = 1, \ldots, k$ and $M_k(L) = (1L\ \cdots\ L^{k-1})^T$. We denote

$$\begin{aligned}\tilde{M}_n^{(\alpha)}(L) &= \left((M_n(L))^T\ 0_m^T\ 0_{s+1}^T\ 0_r^T\ 0_q^T\ 0_p^T\right)^T,\\ \tilde{M}_r^{(\delta)}(L) &= \left(0_n^T\ 0_m^T\ 0_{s+1}^T\ (M_r(L))^T\ 0_q^T\ 0_p^T\right)^T,\\ \tilde{M}_s^{(\omega)}(L) &= \left(0_n^T\ 0_{m+1}^T\ (M_s(L))^T\ 0_r^T\ 0_q^T\ 0_p^T\right)^T,\\ \tilde{M}_m^{(\beta)}(L) &= \left(0_n^T\ (M_m(L))^T\ 0_{s+1}^T\ 0_r^T\ 0_q^T\ 0_p^T\right)^T,\end{aligned}$$

so we can write:

$$\sum_{i=1}^{n} \alpha_i \frac{\partial m_{t-i}}{\partial \lambda} = \Lambda_{n,t}\, \alpha,$$

$$\sum_{j=1}^{r} \delta_j \frac{\partial m_{t-j}}{\partial \lambda} = \Lambda_{r,t}\, \delta,$$

$$\sum_{i=1}^{n} \sum_{j=1}^{r} \alpha_i \delta_j \frac{\partial m_{t-i-j}}{\partial \lambda} = \alpha^T M_n(L)\, \Lambda_{r,t-1}\, \delta,$$

$$\sum_{i=1}^{n} m_{t-i} \frac{\partial \alpha_i}{\partial \lambda} = \tilde{M}_n^{(\alpha)}(L) m_{t-1},$$

$$\sum_{j=1}^{r} m_{t-j} \frac{\partial \delta_j}{\partial \lambda} = \tilde{M}_r^{(\delta)}(L) m_{t-1},$$

$$\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{\partial \alpha_i}{\partial \lambda} \delta_j m_{t-i-j} = \tilde{M}_n^{(\alpha)}(L)\, \delta^T M_r(L) m_{t-2},$$

$$\sum_{i=1}^{n} \sum_{j=1}^{r} \alpha_i \frac{\partial \delta_j}{\partial \lambda} m_{t-i-j} = \tilde{M}_r^{(\delta)}(L)\, \alpha^T M_n(L) m_{t-2},$$

and,

$$\frac{\partial \omega_0}{\partial \lambda} x_t = \left(0_n^T \; 0_m^T \; x_t \; 0_s^T \; 0_r^T \; 0_q^T \; 0_p^T\right)^T,$$

$$\omega_0 \sum_{k=1}^{m} \frac{\partial \beta_k}{\partial \lambda} x_{t-k} = \omega_0 \tilde{M}_m^{(\beta)}(L) x_{t-1} = \left(0_n^T \; \omega_0 M_m^T(L)\, x_{t-1}\, 0_{s+1}^T \; 0_r^T \; 0_q^T \; 0_p^T\right)^T,$$

$$\frac{\partial \omega_0}{\partial \lambda} \sum_{k=1}^{m} \beta_k x_{t-k} = \left(0_n^T \; 0_m^T \; \beta^T M_m(L)\, x_{t-1}\, 0_s^T \; 0_r^T \; 0_q^T \; 0_p^T\right)^T,$$

$$\sum_{l=1}^{s} \frac{\partial \omega_l}{\partial \lambda} x_{t-l} = \tilde{M}_s^{(\omega)}(L) x_{t-1},$$

$$\sum_{l=1}^{s} \sum_{k=1}^{m} \frac{\partial \omega_l}{\partial \lambda} \beta_k x_{t-l-k} = \tilde{M}_s^{(\omega)}(L)\, \beta^T M_m(L) x_{t-2},$$

$$\sum_{l=1}^{s} \sum_{k=1}^{m} \omega_l \frac{\partial \beta_k}{\partial \lambda} x_{t-l-k} = \tilde{M}_m^{(\beta)}(L)\, \omega^T M_s(L) x_{t-2}.$$

Note that the derivatives with respect to ϕ^T and θ^T are equal to zero.

Acknowledgments. We are grateful to Editor John Rice and an anonymous Associate Editor for valuable comments that enabled us to improve the justification of this paper. We also thank the anonymous referees, particularly the one who pointed out that our method is simple to implement.

REFERENCES

ANSLEY, C. F. and NEWBOLD, P. (1980). Finite sample properties of estimators for autoregressive moving average models. *J. Econometrics* **13** 159–183.

DEMEURE, C. J. and MULLIS, C. T. (1989). The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences. *IEEE Trans. Acoust., Speech, Signal Processing* **37** 545–552.

DHARAN, B. G. (1985). A priori sample size evaluation and information matrix computation for time series models. *J. Statist. Comput. Simul.* **21** 171–177.

FRIEDLANDER, B. (1984). On the computation of the Cramér–Rao bound for ARMA parameter estimation. *IEEE Trans. Acoust., Speech, Signal Processing* **32** 721–727.

GODOLPHIN, E. J. and UNWIN, J. M. (1983). Evaluation of the covariance matrix for the maximum likelihood estimator of a Gaussian autoregressive-moving average process. *Biometrika* **70** 279–284.

KLEIN, A. and MÉLARD, G. (1989). On algorithms for computing the covariance matrix of estimates in autoregressive-moving average models. *Comput. Statist. Quarterly* **5** 1–9.

KLEIN, A. and MÉLARD, G. (1990). Fisher's information matrix for seasonal autoregressive-moving average models. *J. Time Ser. Anal.* **11** 231–237.

KLEIN, A. and MÉLARD, G. (1994a). Computation of the Fisher information matrix for SISO models. *IEEE Trans. Signal Processing* **42** 684–688.

KLEIN, A. and MÉLARD, G. (1994b). The information matrix of multiple input single output time series models. *J. Comput. Appl. Math.* **51** 349–356.

KLEIN, A. and SPREIJ, P. (1996). On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices. *Linear Algebra Appl.* **237/238** 579–590.

KLEIN, A. and SPREIJ, P. (1997). On Fisher's information matrix of an ARMA process. In *Stochastic Differential and Difference Equations. Progress in Systems and Control Theory* 273–284. Birkhäuser, Boston.

KOHN, R. and ANSLEY, C. F. (1985). Computing the likelihood and its derivatives for a Gaussian ARMA model. *J. Statist. Comput. Simul.* **22** 229–263.

MAGNUS, J. R. and NEUDECKER, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

MÉLARD, G. (1984). Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models. *J. Roy. Statist. Soc. Ser. C* **33** 104–114.

MÉLARD, G. (1985). Exact derivatives of the likelihood of ARMA processes. In *Proceedings of the Statistical Computing Section* 187–192. Amer. Statist. Assoc., Washington, DC.

MÉLARD, G. and KLEIN, A. (1994). On a fast algorithm for the exact information matrix of a Gaussian ARMA time series. *IEEE Trans. Signal Processing* **42** 2201–2203.

MITTNIK, S. and ZADROZNY, P. A. (1993). Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic model. *Econometrica* **61** 857–870.

MORF, M., SIDHU, G. S. and KAILATH, T. (1974). Some new algorithms for recursive estimation on constant, linear, discrete-time systems. *IEEE Trans. Automat. Control* **19** 315–323.

NEWTON, H. J. (1978). The information matrices of the parameters of multiple mixed time series. *J. Multivariate Anal.* **8** 317–323.

PEARLMAN, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika* **67** 232–233.

PHAM, D. T. (1989). Cramér–Rao bounds for AR parameter and reflection coefficient estimators. *IEEE Trans. Acoust., Speech, Signal Processing* **37** 769–772.

PORAT, B. and FRIEDLANDER, B. (1986). Computation of the exact information matrix of Gaussian time series with stationary random components. *IEEE Trans. Acoust., Speech, Signal Processing* **14** 118–130.

TERCEIRO, J. L. (1990). *Estimation of Dynamic Econometric Models with Errors in Variables.* Springer, Berlin.

TUNNICLIFFE WILSON, G. T. (1979). Some efficient computational procedures for high order ARMA models. *J. Statist. Comput. Simul.* **8** 303–309.

WHITTLE, P. (1953). Estimation and information in time series. *Ark. Math. Fys. Astr.* **2** 423–434.

ZADROZNY, P. A. (1989). Analytical derivatives for estimation of linear dynamic models. *Comput. Math. Appl.* **18** 539–553.

ZADROZNY, P. A. (1992). Errata to "Analytical derivatives for estimation of linear dynamic models." *Comput. Math. Appl.* **24** 289–290.

ANDRÉ KLEIN
DEPARTMENT OF ACTUARIAL SCIENCES
QUANTITATIVE METHODS AND ECONOMETRICS
UNIVERSITY OF AMSTERDAM
ROETERSSTRAAT 11, 1018 WB
AMSTERDAM
THE NETHERLANDS
E-MAIL: aklein@fee.uva.nl

GUY MÉLARD
TOUFIK ZAHAF
INSTITUT DE STATISTIQUE ET
DE RECHERCHE OPÉRATIONNELLE
UNIVERSITÉ LIBRE DE BRUXELLES
CP 210, BD DU TRIOMPHE
BRUXELLES
B-1050 BELGIQUE
E-MAIL: gmelard@ulb.ac.be
tzahaf@ulb.ac.be