EXACT MAXIMUM LIKELIHOOD ESTIMATION OF STATIONARY VECTOR ARMA MODELS*

José Alberto Mauricio
Instituto Complutense de Análisis Económico
Universidad Complutense
Campus de Somosaguas
28223 Madrid

ABSTRACT

The problems of evaluating and maximizing the exact likelihood function of vector ARMA models are considered separately. A new and efficient procedure for evaluating the exact likelihood function is presented. This method puts together a set of useful features which can only be found separately in currently available algorithms. A procedure for maximizing the exact likelihood function, which takes full advantage of the properties offered by the evaluation algorithm, is also considered. Combining these two procedures, a new algorithm for exact maximum likelihood estimation of vector ARMA models is obtained. Comparisons with existing procedures, in terms of both analytical arguments and a numerical example , are given in order to show that the new estimation algorithm performs at least as well as existing ones, and that relevant real situations occur in which it does better.

RESUMEN

En este trabajo se estudian por separado los problemas asociados con la evaluación y la maximización de la función de verosimilitud exacta de modelos ARMA multivariantes. Por un lado, se diseña un nuevo algoritmo para evaluar dicha función eficientemente, que reúne una serie de propiedades que se encuentran por separado en los procedimientos actualmente disponibles. Por otro lado, se propone un mecanismo para maximizar la función de verosimilitud, que aprovecha las ventajas ofrecidas por el algoritmo de evaluación. Combinando ambos procedimientos, se obtiene un nuevo algoritmo de estimación por máxima verosimilitud exacta de modelos ARMA multivariantes, cuyo funcionamiento en la práctica resulta superior, en muchos aspectos, al de otros procedimientos alternativos.

* This paper summarizes the author's doctoral dissertation Evaluación y Maximización de la Función de Verosimilitud de Procesos ARMA Multivariantes (1993, Madrid: Universidad Complutense de Madrid), and was partially financed under the project for a Forecast and Monitor Service for the Spanish Economy, sponsored by Caja de Madrid (Spain), and by the DGICYT project PB89-0129 (Spain). The author is grateful to Arthur B. Treadway, under whose supervision this research was carried out, and to Emilio Cerdá, José Luis Gallego, Víctor M. Gonzalo, Miguel Jerez, Alfonso Novales, Daniel Peña and Fernando Tussel for helpful suggestions.

1. INTRODUCTION

This article proposes a new procedure for exact maximum likelihood estimation of vector ARMA models. A sharp distinction between evaluating and maximizing the likelihood function is made. This permits a detailed analysis of all problems that arise in the estimation process. The solutions obtained through this analysis can then be integrated into a complete estimation procedure that takes advantage of some properties of the likelihood function which have not been fully exploited in previous papers.

Although there has been abundant research on the evaluation of the likelihood function of vector ARMA models (e.g. Hillmer and Tiao 1979; Nicholls and Hall 1979; Hall and Nicholls 1980; Shea 1987), only a few authors have paid attention to the problem of its subsequent maximization. Furthermore, such attention has usually been restricted to suggesting, in a few lines, the use of a standard optimization algorithm to maximize the likelihood function, evaluated as extensively described in the cited papers. An interesting exception can be found in Shea (1984, pp. 99-100).

With regard to the computation of the likelihood function, none of the existing methods can be taken as fully satisfactory. This is due to the fact that many of the properties a method of evaluating the likelihood function should have, are scattered among the existing procedures. Thus, although each of many existing algorithms has some useful properties, it also lacks others that can be found in alternative procedures.

For instance, the algorithm of Shea (1984, 1987) can be considered, from a computational viewpoint, the most efficient among the existing ones. However, its use does not permit the automatic detection of non-invertible models. This task is easily handled with the algorithm of Hall and Nicholls (1980) and with an extension to the multivariate context of the algorithm of Ljung and Box (1979). However, the former involves a high computational cost (in many cases) and the latter both some computational inefficiency and a loss of numerical precision, due to the requirement for an explicit matrix inversion. Finally, the algorithm of Hillmer and Tiao (1979) permits the computation of neither the exact likelihood function nor an appropriate residual vector, except in the case of pure moving average models; this fact may become an important drawback when the model considered has an autoregressive part and the sample contains extreme values among the initial observations.

A thorough analysis of currently available procedures allows one to discover and fully exploit new possibilities ignored in the papers cited above. Thus, in Section 2, a new method of evaluating the exact likelihood function of vector ARMA models is described in detail. The new algorithm puts

together the advantages which can be found separately in existing procedures and does not suffer from any of their drawbacks. In Section 3, computational techniques for maximizing the likelihood function are considered. These techniques take full advantage of the properties offered by the evaluation algorithm. Section 4 gives an illustrative example of an actual situation in which the new estimation procedure performs better than one of the most frequently used procedures. Finally, in Section 5, conclusions are summarized.

2. EVALUATION OF THE EXACT LIKELIHOOD FUNCTION

Let $\mathbf{w}_t$ be an m-dimensional vector-valued time series. It is supposed that $\mathbf{w}_t$ follows the vector ARMA(p,q) model

$$\boldsymbol{\Phi}(B)\tilde{\mathbf{w}}_t = \boldsymbol{\Theta}(B)\mathbf{a}_t \,, \tag{2.1}$$

where $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$, $\boldsymbol{\Theta}(B) = \mathbf{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$, B is the back shift operator, $\tilde{\mathbf{w}}_t = \mathbf{w}_t - \boldsymbol{\mu}$, $\boldsymbol{\Phi}_i$ ($i = 1, \ldots, p$), $\boldsymbol{\Theta}_i$ ($i = 1, \ldots, q$) and $\boldsymbol{\mu}$ are $m\times m$, $m\times m$ and $m\times 1$ parameter matrices, respectively, and the $\mathbf{a}_t$'s are $m\times 1$ random vectors identically and independently distributed as $N(\mathbf{0}, \sigma^2\mathbf{Q})$, with $\sigma^2 > 0$ and $\mathbf{Q}$ ($m\times m$) symmetric and positive definite. This decomposition of $E[\mathbf{a}_t\mathbf{a}_t^T]$, although not unique, is useful for obtaining maximum likelihood estimates by maximizing a concentrated log likelihood as a function of $\boldsymbol{\Phi}_i$ ($i = 1, \ldots, p$), $\boldsymbol{\Theta}_i$ ($i = 1, \ldots, q$) and $\mathbf{Q}$ only (Section 3). For stationarity, it is required that the zeros of $|\boldsymbol{\Phi}(B)|$ lie outside the unit circle. Furthermore, (2.1) is assumed to satisfy the conditions derived by Hannan (1969) in order that the model be identified.

Consider a sample of size n and let $\tilde{\mathbf{w}} = (\tilde{\mathbf{w}}_1^T, \ldots, \tilde{\mathbf{w}}_n^T)^T$ (mean-corrected observations), $\mathbf{a} = (\mathbf{a}_1^T, \ldots, \mathbf{a}_n^T)^T$ (white noise perturbations), and $\mathbf{u}_* = (\tilde{\mathbf{w}}_{1-p}^T, \ldots, \tilde{\mathbf{w}}_0^T, \mathbf{a}_{1-q}^T, \ldots, \mathbf{a}_0^T)^T$ (unknown presample values). Then, equation (2.1) may be written as

$$\mathbf{D}_{\boldsymbol{\Phi},n}\tilde{\mathbf{w}} = \mathbf{D}_{\boldsymbol{\Theta},n}\mathbf{a} + \mathbf{V}\mathbf{u}_* \,, \tag{2.2}$$

where $\mathbf{D}_{\boldsymbol{\Phi},n}$ and $\mathbf{D}_{\boldsymbol{\Theta},n}$ are $nm\times nm$ block-matrices with identity matrices on the main diagonal, and $-\boldsymbol{\Phi}_k$ and $-\boldsymbol{\Theta}_k$, respectively, down the k-*th* subdiagonal. Further $\mathbf{V}$ is the $nm\times(p+q)m$ block-matrix $\mathbf{V} = (\mathbf{G}_{\boldsymbol{\Phi},n}, \mathbf{G}_{\boldsymbol{\Theta},n})$, where $\mathbf{G}_{\boldsymbol{\Phi},n}$ and $\mathbf{G}_{\boldsymbol{\Theta},n}$ are the following $nm\times pm$ and $nm\times qm$ block-matrices:

$$\mathbf{G}_{\boldsymbol{\Phi},n} = \begin{bmatrix} \boldsymbol{\Phi}_p & \boldsymbol{\Phi}_{p-1} & \cdots & \boldsymbol{\Phi}_1 \\ \mathbf{0} & \boldsymbol{\Phi}_p & \cdots & \boldsymbol{\Phi}_2 \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Phi}_p \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} ; \qquad \mathbf{G}_{\boldsymbol{\Theta},n} = \begin{bmatrix} -\boldsymbol{\Theta}_q & -\boldsymbol{\Theta}_{q-1} & \cdots & -\boldsymbol{\Theta}_1 \\ \mathbf{0} & -\boldsymbol{\Theta}_q & \cdots & -\boldsymbol{\Theta}_2 \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\boldsymbol{\Theta}_q \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} .$$

On the basis of the previous definitions, Nicholls and Hall (1979) have shown that the exact likelihood function of the parameters $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$, $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q)$, $\boldsymbol{\mu}$, σ^2 and $\mathbf{Q}$ is given by

$$L(\boldsymbol{\Phi},\boldsymbol{\Theta},\boldsymbol{\mu},\sigma^2,\mathbf{Q}|\mathbf{w}) = (2\pi\sigma^2)^{-\frac{nm}{2}}|\mathbf{Q}|^{-\frac{n}{2}}|\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}S(\boldsymbol{\Phi},\boldsymbol{\Theta},\boldsymbol{\mu},\mathbf{Q}|\mathbf{w})\right\} . \tag{2.3}$$

The quadratic form in the exponential is given by $S(\boldsymbol{\Phi},\boldsymbol{\Theta},\boldsymbol{\mu},\mathbf{Q}|\mathbf{w}) = (\boldsymbol{\Upsilon}\tilde{\mathbf{w}}+\boldsymbol{\Lambda}\hat{\mathbf{e}}_*)^T(\boldsymbol{\Upsilon}\tilde{\mathbf{w}}+\boldsymbol{\Lambda}\hat{\mathbf{e}}_*)$, where $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Lambda}$ are the following $(p+q+n)m\times nm$ and $(p+q+n)m\times(p+q)m$ matrices:

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \mathbf{0} \\ (\mathbf{I}\otimes\mathbf{R})\mathbf{K} \end{bmatrix} , \quad \boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{I} \\ (\mathbf{I}\otimes\mathbf{R})\mathbf{Z}\mathbf{T}^{-1} \end{bmatrix} , \tag{2.4}$$

and

$$\hat{\mathbf{e}}_* = \mathbf{T}\hat{\mathbf{u}}_* = -(\boldsymbol{\Lambda}^T\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Upsilon}\tilde{\mathbf{w}} \,. \tag{2.5}$$

In (2.4), the $nm\times nm$ matrix $\mathbf{K}$ is given by $\mathbf{K} = \mathbf{D}_{\boldsymbol{\Theta},n}^{-1}\mathbf{D}_{\boldsymbol{\Phi},n}$, the $nm\times(p+q)m$ matrix $\mathbf{Z}$ is given by $\mathbf{Z} = -\mathbf{D}_{\boldsymbol{\Theta},n}^{-1}\mathbf{V}$, and, if $E[\mathbf{a}_t\mathbf{a}_t^T] = \sigma^2\mathbf{Q}$ and $E[\mathbf{u}_*\mathbf{u}_*^T] = \sigma^2\boldsymbol{\Omega}$, then the $m\times m$ and $(p+q)m\times(p+q)m$ matrices $\mathbf{R}$ and $\mathbf{T}$ are such that $\mathbf{R}\mathbf{Q}\mathbf{R}^T = \mathbf{I}$ (i.e. $\mathbf{Q}^{-1} = \mathbf{R}^T\mathbf{R}$) and $\mathbf{T}\boldsymbol{\Omega}\mathbf{T}^T = \mathbf{I}$ (i.e. $\boldsymbol{\Omega}^{-1} = \mathbf{T}^T\mathbf{T}$). Further, the matrix $\boldsymbol{\Omega}$ can be partitioned as follows:

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix} .$$

The (i, j)-*th* block of $\mathbf{A}$ is given by $\mathbf{A}_{ij} = \sigma^{-2}E[\tilde{\mathbf{w}}_{i-p}\tilde{\mathbf{w}}_{j-p}^T] = \boldsymbol{\Gamma}(j-i)$ ($i, j = 1, \ldots, p$), and the (i, j)-*th* block of matrix $\mathbf{B}$ is $\mathbf{B}_{ij} = \sigma^{-2}E[\tilde{\mathbf{w}}_{i-p}\mathbf{a}_{j-q}^T] = \boldsymbol{\Gamma}_{\mathrm{wa}}(j-i-q+p)$ ($i = 1, \ldots, p$; $j = 1, \ldots, q$). Since $\mathbf{A}$ is symmetric and $E[\tilde{\mathbf{w}}_{t-i}\mathbf{a}_t^T] = \mathbf{0}$ for $i > 0$, in order to compute $\mathbf{A}$ and $\mathbf{B}$ only the theoretical autocovariance and cross-covariance matrices $\boldsymbol{\Gamma}(k)$, $k = 0, \ldots, p-1$, and $\boldsymbol{\Gamma}_{\mathrm{wa}}(k)$, $k = -q+1, \ldots, 0$, are needed. Finally, $\mathbf{C}$ is a block-diagonal $qm\times qm$ matrix with $\mathbf{Q}$'s along the main diagonal.

Thus, to evaluate (2.3) one must compute the determinant $|\mathbf{\Lambda}^T\mathbf{\Lambda}|$ and the quadratic form $S(\mathbf{\Phi},\mathbf{\Theta},\mu,\mathbf{Q}|\mathbf{w})$. Hall and Nicholls (1980) have suggested computing the latter as follows:

$$S(\mathbf{\Phi},\mathbf{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = (\mathbf{T}\bar{\mathbf{w}})^T[\mathbf{I}-\mathbf{\Lambda}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T](\mathbf{T}\bar{\mathbf{w}}) , \tag{2.6}$$

which may be regarded as the residual sum of squares of the regression of $(\mathbf{T}\bar{\mathbf{w}})$ on $\mathbf{\Lambda}$. To evaluate $(\mathbf{T}\bar{\mathbf{w}})$ and $\mathbf{\Lambda}$, they compute $(\mathbf{I}\otimes\mathbf{R})\mathbf{K}\bar{\mathbf{w}}$ and $(\mathbf{I}\otimes\mathbf{R})\mathbf{Z}$ [see (2.4) above] recursively.

It is shown next, by exploring in further detail the elements of (2.3), how to compute the determinant $|\mathbf{\Lambda}^T\mathbf{\Lambda}|$ and the quadratic form (2.6) in a computationally more efficient manner. It is also shown: (1) how to compute an approximation to the exact likelihood function to any desired degree of accuracy, (2) how to detect non-invertible and/or non-stationary models, and (3) how to calculate the residual vector for a given set of data and parameter values.

2.1. A New Method of Computing the Exact Likelihood Function

First, it may be noted, from (2.4), that

$$\mathbf{\Lambda}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T = \begin{bmatrix} (\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1} & (\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{T}^{-1T}\mathbf{Z}^T(\mathbf{I}\otimes\mathbf{R}^T) \\ (\mathbf{I}\otimes\mathbf{R})\mathbf{Z}\mathbf{T}^{-1}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1} & (\mathbf{I}\otimes\mathbf{R})\mathbf{Z}\mathbf{T}^{-1}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{T}^{-1T}\mathbf{Z}^T(\mathbf{I}\otimes\mathbf{R}^T) \end{bmatrix} ,$$

and, since $\mathbf{T}\bar{\mathbf{w}} = (\mathbf{0}^T, \eta^T)^T$, where

$$\eta = (\mathbf{I}\otimes\mathbf{R})\hat{\mathbf{a}}_0 , \tag{2.7}$$

$$\hat{\mathbf{a}}_0 = E[\mathbf{a}|\mathbf{w}, \mathbf{u}_*=\mathbf{0}] = \mathbf{K}\bar{\mathbf{w}} , \tag{2.8}$$

the quadratic form (2.6) can be written as

$$S(\mathbf{\Phi},\mathbf{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = \eta^T\eta - \eta^T(\mathbf{I}\otimes\mathbf{R})\mathbf{Z}\mathbf{T}^{-1}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{T}^{-1T}\mathbf{Z}^T(\mathbf{I}\otimes\mathbf{R}^T)\eta . \tag{2.9}$$

Also from (2.4), it may be noted that $\mathbf{\Lambda}^T\mathbf{\Lambda} = \mathbf{I} + \mathbf{T}^{-1T}\mathbf{X}\mathbf{T}^{-1}$, where the $(p+q)m\times(p+q)m$ matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \mathbf{V}_1^T\mathbf{H}^T\mathbf{H}\mathbf{V}_1 , \tag{2.10}$$

and, if $g = \max(p, q)$, then the $gm\times(p+q)m$ matrix $\mathbf{V}_1$ consists of the first gm rows of $\mathbf{V}$, and the

$nm\times gm$ matrix $\mathbf{H}$ consists of the first gm columns of $(\mathbf{I}\otimes\mathbf{R})\mathbf{D}_{\mathbf{\Theta},n}^{-1}$. Thus, equation (2.9) can be rewritten as:

$$S(\mathbf{\Phi},\mathbf{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = \eta^T\eta - \eta^T(\mathbf{I}\otimes\mathbf{R})\mathbf{Z}(\mathbf{\Omega}^{-1}+\mathbf{V}_1^T\mathbf{H}^T\mathbf{H}\mathbf{V}_1)^{-1}\mathbf{Z}^T(\mathbf{I}\otimes\mathbf{R}^T)\eta . \tag{2.11}$$

Now, define the following $nm\times 1$ vector:

$$\mathbf{h} = \mathbf{D}_{\mathbf{\Theta},n}^{-1T}(\mathbf{I}\otimes\mathbf{R}^T)\eta , \tag{2.12}$$

and let $\tilde{\mathbf{h}}$ contain the first gm elements of $\mathbf{h}$. Then, equation (2.11) can be rewritten as

$$S(\mathbf{\Phi},\mathbf{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = \eta^T\eta - \tilde{\mathbf{h}}^T\mathbf{V}_1(\mathbf{\Omega}^{-1}+\mathbf{V}_1^T\mathbf{H}^T\mathbf{H}\mathbf{V}_1)^{-1}\mathbf{V}_1^T\tilde{\mathbf{h}} ,$$

where it may be verified that $\mathbf{V}_1(\mathbf{\Omega}^{-1} + \mathbf{V}_1^T\mathbf{H}^T\mathbf{H}\mathbf{V}_1)^{-1}\mathbf{V}_1^T = [(\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T)^{-1} + \mathbf{H}^T\mathbf{H}]^{-1}$. Thus, the quadratic form (2.11) is

$$S(\mathbf{\Phi},\mathbf{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = \eta^T\eta - \tilde{\mathbf{h}}^T[(\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T)^{-1}+\mathbf{H}^T\mathbf{H}]^{-1}\tilde{\mathbf{h}} . \tag{2.13}$$

Then, noting from (2.4) that $\mathbf{\Lambda}^T\mathbf{\Lambda} = \mathbf{I} + \mathbf{T}^{-1T}\mathbf{X}\mathbf{T}^{-1}$, it is clear that $\mathbf{T}^T\mathbf{\Lambda}^T\mathbf{\Lambda}\mathbf{T} = \mathbf{\Omega}^{-1} + \mathbf{X}$, so [see (2.10)] the determinant $|\mathbf{\Lambda}^T\mathbf{\Lambda}|$ can be calculated as

$$|\mathbf{\Lambda}^T\mathbf{\Lambda}| = |\mathbf{\Omega}|\,|\mathbf{\Omega}^{-1}+\mathbf{V}_1^T\mathbf{H}^T\mathbf{H}\mathbf{V}_1| = |\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T|\,|(\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T)^{-1}+\mathbf{H}^T\mathbf{H}| , \tag{2.14}$$

which is readily available as a byproduct of the evaluation of the second term on the right hand side of (2.13). Note that, when $m = 1$ (i.e. when dealing with univariate models), expressions (2.11) and (2.14) reduce to equations (2.6) and (2.4) respectively in Ljung and Box (1979). Also, note that the evaluation of these expressions requires the explicit inversion of the matrix $\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T$, which results in a loss of computational efficiency and, occasionally, in a loss of numerical precision too. This matrix inversion can be avoided as follows. Let $\mathbf{M}$ denote the Cholesky factor of $\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T$, so that $\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T = \mathbf{M}\mathbf{M}^T$ and $\mathbf{M}^T(\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T)^{-1}\mathbf{M} = \mathbf{I}$. Then, $[(\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T)^{-1} + \mathbf{H}^T\mathbf{H}]^{-1} = \mathbf{M}(\mathbf{I} + \mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M})^{-1}\mathbf{M}^T$, so that the quadratic form (2.13) can be finally expressed as

$$S(\Phi,\Theta,\mu,\Omega|\mathbf{w}) = \eta^T\eta - (\mathbf{M}^T\tilde{\mathbf{h}})^T(\mathbf{I}+\mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M})^{-1}(\mathbf{M}^T\tilde{\mathbf{h}}) \ . \tag{2.15}$$

The computation of the second term on the right hand side of (2.15) gives as a byproduct the components of the determinant (2.14), which can be written in the form

$$|\Lambda^T\Lambda| = |\mathbf{I}+\mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M}| \ . \tag{2.16}$$

The exact likelihood function (2.3) is then computed using (2.15) and (2.16). In order to evaluate these expressions, one needs: (1) the $gm\times gm$ symmetric matrix $\mathbf{V}_1\Omega\mathbf{V}_1^T$, (2) the $gm\times 1$ vector $\tilde{\mathbf{h}}$, (3) the $gm\times gm$ symmetric matrix $\mathbf{H}^T\mathbf{H}$, and (4) the $nm\times 1$ vector η.

Once the components of Ω are available (Hall and Nicholls 1980, pp. 254-256; Kohn and Ansley 1982), the (i, j)-*th* block ($i = 1, \ldots, g; j = 1, \ldots, i$) of $\mathbf{V}_1\Omega\mathbf{V}_1^T$ is given by

$$(\mathbf{V}_1\Omega V_1^T)_{ij} = \sum_{k=0}^{p-i}\Phi_{p-k}\mathbf{E}_{k+i,j} - \sum_{k=0}^{q-i}\Theta_{q-k}\mathbf{E}_{k+p+i,j} \ , \tag{2.17}$$

where, for $j = 1, \ldots, g$:

$$\mathbf{E}_{ij} = \sum_{k=j-i}^{p-i}\Gamma(k)\Phi_{p-k-i+j}^T - \sum_{k=j-i}^{q-i}\Gamma_{wa}(-q+p+k)\Theta_{q-k-i+j}^T \quad (i=1,\ldots,p) \ ,$$

$$\mathbf{E}_{ij} = \sum_{k=p+j-i}^{2p-i}\Gamma_{wa}(-q+p-k)^T\Phi_{2p-k-i+j}^T - \mathbf{Q}\Theta_{q+p-i+j}^T \quad (i=p+1,\ldots,p+q) \ ,$$

with $\Gamma(k) = \Gamma(-k)^T$ for $k < 0$, $\Gamma_{wa}(k) = \mathbf{0}$ for $k > 0$, and $\Theta_i = \mathbf{0}$ for $i > q$. Now, since $\tilde{\mathbf{h}}$ contains the first gm components of $\mathbf{h}$ [see (2.12)] and $\mathbf{H}$ consists of the first gm columns of $(\mathbf{I}\otimes\mathbf{R})\mathbf{D}_{\Theta,n}^{-1}$, in order to evaluate $\tilde{\mathbf{h}}$ and $\mathbf{H}^T\mathbf{H}$ the matrix $\mathbf{D}_{\Theta,n}^{-1}$ is needed. It can be shown (Hillmer and Tiao 1979, pp. 652-653) that $\mathbf{D}_{\Theta,n}^{-1}$ is a lower triangular block-matrix with identity matrices along the main diagonal, Ξ_1 down the subdiagonal, and so on, where the Ξ_k's are evaluated recursively as

$$\Xi_k = \sum_{j=1}^{q}\Theta_j\Xi_{k-j} \quad (k=1,\ldots,n-1) \ , \tag{2.18}$$

with $\Xi_0 = \mathbf{I}$ and $\Xi_k = \mathbf{0}$ for $k < 0$. Then, the *j-th* block of vector $\tilde{\mathbf{h}}$ is given by

$$\tilde{\mathbf{h}}_j = \sum_{i=0}^{n-j}\Xi_i^T\mathbf{R}^T\eta_{i+j} \quad (j=1,\ldots,g) \ . \tag{2.19}$$

From the special structure of matrix $\mathbf{H}$, the first block-column of $\mathbf{H}^T\mathbf{H}$ is given by

$$(\mathbf{H}^T\mathbf{H})_{i1} = \sum_{k=0}^{n-i}\Xi_k^T\mathbf{R}^T\mathbf{R}\Xi_{k+i-1} \quad (i=1,\ldots,g) \ , \tag{2.20}$$

and the remaining diagonal and subdiagonal blocks of matrix $\mathbf{H}^T\mathbf{H}$ are evaluated in the following recursive manner:

$$(\mathbf{H}^T\mathbf{H})_{ij} = (\mathbf{H}^T\mathbf{H})_{i-1,j-1} - \Xi_{n-i+1}^T\mathbf{R}^T\mathbf{R}\Xi_{n-j+1} \ , \tag{2.21}$$

with $i = 2, \ldots, g$ and $j = 2, \ldots, i$. Finally, from (2.7) and (2.8), the n blocks that make up vector $\hat{\mathbf{a}}_0$ can be computed recursively as follows:

$$\hat{\mathbf{a}}_{0i} = \tilde{\mathbf{w}}_i - \sum_{j=1}^{p}\Phi_j\tilde{\mathbf{w}}_{i-j} + \sum_{j=1}^{q}\Theta_j\hat{\mathbf{a}}_{0,i-j} \quad (i=1,\ldots,n) \ , \tag{2.22}$$

with $\tilde{\mathbf{w}}_i = \mathbf{0}$ for $i < 1$ and $\hat{\mathbf{a}}_{0i} = \mathbf{0}$ for $i < 1$. Then, the *i-th* block of vector η is given by $\eta_i = \mathbf{R}\hat{\mathbf{a}}_{0i}$ ($i = 1, \ldots, n$). The calculation of η through (2.7), (2.8) and (2.22) can also be found in Ljung and Box (1979, p. 267) and Hall and Nicholls (1980, p. 256). In summary, the following procedure is suggested to evaluate the exact likelihood function of a vector ARMA model (note that, except for step **(1)** below, no explicit matrix inversion is required):

(1) Compute the Cholesky factor of matrix $\mathbf{Q}$ (say $\mathbf{Q}_1$), its determinant ($|\mathbf{Q}| = |\mathbf{Q}_1|^2$), and a matrix $\mathbf{R}$ such that $\mathbf{R}\mathbf{Q}\mathbf{R}^T = \mathbf{I}$ ($\mathbf{R} = \mathbf{Q}_1^{-1}$).

(2) Evaluate the theoretical autocovariance and cross-covariance matrices $\Gamma(k)$ ($k = 0, \ldots, p-1$) and $\Gamma_{wa}(k)$ ($k = -q+1, \ldots, 0$).

(3) Compute matrix $\mathbf{V}_1\boldsymbol{\Omega}\mathbf{V}_1^T$ from (2.17), and its Cholesky factor $\mathbf{M}$.

(4) Evaluate the sequence $\boldsymbol{\Xi}_k$ ($k = 1, \ldots, n-1$) from (2.18).

(5) Calculate vector $\boldsymbol{\eta}$ using (2.22).

(6) Compute vector $\tilde{\mathbf{h}}$ from (2.19), and evaluate vector $\mathbf{M}^T\tilde{\mathbf{h}}$.

(7) Evaluate matrix $\mathbf{H}^T\mathbf{H}$ from (2.20) and (2.21).

(8) Compute matrix $\mathbf{I} + \mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M}$, its Cholesky factor (say $\mathbf{L}$), and its determinant ($|\mathbf{I} + \mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M}| = |\mathbf{L}|^2$), which is, in turn, the determinant (2.16).

(9) Use forward substitution to solve for $\boldsymbol{\lambda}$ in the triangular system $\mathbf{L}\boldsymbol{\lambda} = (\mathbf{M}^T\tilde{\mathbf{h}})$.

(10) Compute the quadratic form (2.15) as $S(\boldsymbol{\Phi},\boldsymbol{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = \boldsymbol{\eta}^T\boldsymbol{\eta} - \boldsymbol{\lambda}^T\boldsymbol{\lambda}$.

Following the guidelines in Hillmer and Tiao (1979, pp. 653-654) and Ljung and Box (1979, p. 269), this procedure can also take advantage from the special structure of some specific models, such as multiplicative pure MA models and pure AR models. The details are straightforward and have been omitted.

2.2. Properties of the New Algorithm

The procedure outlined above is, basically, the result of extending and taking one step beyond the method of Ljung and Box (1979) for the scalar ARMA model. Since the explicit inversion of matrix $\mathbf{V}_1\boldsymbol{\Omega}\mathbf{V}_1^T$ is avoided through the use of its Cholesky factorization, a more computationally efficient and numerically stable method is obtained. Further, since the new algorithm operates with $gm \times gm$ matrices [see (2.15) and (2.16)] instead of $(p+q)m \times (p+q)m$ matrices [see (2.6)], it is also preferable from a computational viewpoint to that of Hall and Nicholls (1980). To illustrate this, computer programs were written for the algorithm of Hall and Nicholls (1980) and the new algorithm developed in this paper. The exact likelihood function was evaluated for a variety of vector ARMA models suitable for annual, quarterly and monthly data, under the assumption that 25 years of data were

available. In Table 1 the ratio between the number of time-consuming operations (multiplications, divisions and square roots) required by the algorithm of Hall and Nicholls (1987) and those required by the new algorithm, is presented for each of the models considered. This ratio is always greater than or equal to one, and reaches its highest value for models with both p and q large.

[INSERT TABLE 1]

The comparison from a computational viewpoint between the new algorithm and that of Shea (1987) is summarized in Table 2, which contains the same kind of information as Table 1. Apart from minor change through refinements in coding, it can be seen that the relative efficiency of the new algorithm increases with m (except if p is high and much larger than q), and that the new algorithm is clearly preferable for low to medium order models, whereas the method of Shea (1987) is more efficient for some higher order models.

[INSERT TABLE 2]

Finally, note that the method of Hillmer and Tiao (1979) does not allow for an exact evaluation of the likelihood function when the model contains an autoregressive part; furthermore, in the case of pure moving average models, the expression of the exact likelihood function obtained in that paper is equivalent to that of Nicholls and Hall (1979).

With regard to other interesting properties, note first that when the model considered is invertible, the matrix sequence (2.18) converges to $\mathbf{0}$, the more quickly the larger the moduli of the zeros of $|\boldsymbol{\Theta}(B)|$ are (obviously, when $q = 0$, $\boldsymbol{\Xi}_k = \mathbf{0}$ for $k \geq 1$). This may be exploited in the subsequent computation of (2.19), (2.20) and (2.21), since if $\boldsymbol{\Xi}_k = \mathbf{0}$ for, say, $k \geq r^*$, then not all of the operations involved in those expressions need to be carried out. The sequence (2.18) may be

considered to have converged when

$$\left[\sum_{i=1}^{m}\sum_{j=1}^{m}|\mathbf{\Xi}_{r^*}(i,j)|\right] < \delta \ ,$$

where the parameter $\delta > 0$ can be used to control the desired degree of approximation to the exact computation of the whole sequence (2.18). It is possible to make the convergence criterion sufficiently rigid (i.e. δ sufficiently small) that the error implied by considering $\mathbf{\Xi}_k = \mathbf{0}$ for $k \geq r^*$, becomes negligible and hence so too does the difference between the exact (calculated with $\mathbf{\Xi}_k$ from $k = 1$ to $k = n-1$) and the 'approximate' (calculated with $\mathbf{\Xi}_k = \mathbf{0}$ for $k \geq r^*$) likelihood. Note that this property, which may save much computing time, is analogous to the 'quick recursions' property offered by the Chandrasekhar equations that form the basis of the method of Shea (1989, pp. 169-170). Furthermore, using the convergence property of (2.18) for invertible models, it is straightforward to detect the presence of any root of the moving average operator lying inside the unit circle, since in such a case the sequence (2.18) will be explosive. In practice, it has been observed that the following inequality holds for strictly non-invertible models:

$$\left[\sum_{i=1}^{m}\sum_{j=1}^{m}|\mathbf{\Xi}_{h}(i,j)|\right] > \sum_{k=1}^{\min(h,q)}\left[\sum_{i=1}^{m}\sum_{j=1}^{m}|\mathbf{\Xi}_{k}(i,j)|\right] \ ,$$

for at least one $h < n-1$. In general, when the MA operator has at least one root inside the unit circle, this condition will be true for h slightly larger than q, allowing detection of strict non-invertibility at the beginning of the computation of the sequence (2.18). In such a case, the algorithm flags a warning and stops, in order to avoid overflow problems in the subsequent computation of (2.20) and (2.21). However, there is no problem in evaluating these expressions when any root of the MA operator lies on the unit circle, provided the other roots have moduli larger than unity.

Also as a byproduct, the new algorithm provides a necessary (though not sufficient, except for pure autoregressive models) check on the stationarity of the model. This is due to the fact that the Cholesky

decomposition of $\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T$ exists if and only if $\mathbf{\Omega}$ is positive-definite, which is in turn a necessary condition for stationarity (it is also sufficient when $q = 0$; see, for example, Ansley 1979). Since the computation of the Cholesky factor of $\mathbf{V}_1\mathbf{\Omega}\mathbf{V}_1^T$ is a key step in the new algorithm, the impossibility of carrying out this task indicates that the model considered is not stationary. Note, however, that existence of the Cholesky decomposition does not guarantee stationarity for a mixed model.

Finally, it is shown how to calculate the residuals for a given set of observations and parameter values, using some of the computations carried so far to evaluate the exact likelihood function. From (2.2) and the definition of $\mathbf{K}$, $\mathbf{Z}$, it is clear that $\hat{\mathbf{a}} = \mathbf{K}\tilde{\mathbf{w}} + \mathbf{Z}\hat{\mathbf{u}}_*$, with $\mathbf{K}\tilde{\mathbf{w}} = \hat{\mathbf{a}}_0$ [see (2.8)], and, from (2.5), it follows that $\hat{\mathbf{u}}_* = -\mathbf{T}^{-1}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T\mathbf{\Upsilon}\tilde{\mathbf{w}}$. Then, noting (2.4), (2.7), (2.8) and (2.12), it is possible to show that

$$\hat{\mathbf{a}} = \hat{\mathbf{a}}_0 - \mathbf{D}_{\boldsymbol{\theta},n}^{-1}\begin{bmatrix}\mathbf{M}(\mathbf{I} + \mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M})^{-1}\mathbf{M}^T\tilde{\mathbf{h}} \\ \mathbf{0}\end{bmatrix} . \qquad (2.23)$$

Thus, using previous computations, the exact residual vector $\hat{\mathbf{a}}$ can be evaluated as follows: (1) use backward substitution to solve for $\mathbf{c}$ in the triangular system $\mathbf{L}^T\mathbf{c} = \lambda$, (2) compute the $gm\times 1$ vector $\mathbf{d} = \mathbf{M}\mathbf{c}$, and (3) evaluate $\hat{\mathbf{a}} = \hat{\mathbf{a}}_0 - \mathbf{r}$, where the i-th block of the $nm\times 1$ vector $\mathbf{r} = \mathbf{D}_{\boldsymbol{\theta},n}^{-1}(\mathbf{d}^T, \mathbf{0}^T)^T$ is given by

$$\mathbf{r}_i = \sum_{j=1}^{i}\mathbf{\Xi}_{i-j}\mathbf{d}_j \quad (i = 1,\ldots,n) \ ,$$

with $\mathbf{d}_j = \mathbf{0}$ for $j > g$. Conditional on maximum likelihood estimates being equal to the true parameter values, the residuals thus calculated can be shown to be normally distributed with $E[\hat{\mathbf{a}}_t] = \mathbf{0}$. Further, it can be shown that, as t tends to n, $\hat{\mathbf{a}}_t$ converges in quadratic mean to $\mathbf{a}_t$ and the $\hat{\mathbf{a}}_t$'s tend to be uncorrelated, with $E[\hat{\mathbf{a}}_t\hat{\mathbf{a}}_t^T]$ converging to $\sigma^2\mathbf{Q}$ (when $q = 0$, this convergence occurs exactly for $t > p$). These properties are shared with the residuals obtained by using the Kalman filter to evaluate the exact likelihood function (Shea 1984, p.93; 1989, p. 162).

To conclude, Table 3 shows a comparative summary of the evaluation algorithms considered in this section, in terms of the following features: (1) exact evaluation of the likelihood function; (2) 'approximate' evaluation of the likelihood function, as accurate as desired and, in most cases, significantly faster than the exact evaluation; (3) computational efficiency, in terms of the number of time-consuming operations required; (4) numerical accuracy and stability; (5) detection, as a byproduct, of non-stationary and/or non-invertible models; and (6) evaluation of an appropriate residual vector using some of the computations carried to evaluate the likelihood function.

[INSERT TABLE 3]

In summary, the new evaluation algorithm provides a set of useful features, not found together in any of the other existing methods, which can be put to work effectively in the context of maximum likelihood estimation of vector ARMA models.

3. MAXIMIZATION OF THE EXACT LIKELIHOOD FUNCTION

Having devised a method of computing the exact likelihood function, we now seek how to maximize it with respect to the parameters $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$, $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q)$, μ, σ^2 and $\mathbf{Q}$. The parameter σ^2 may be differentiated out of equation (2.3) to yield the following concentrated log likelihood:

$$l_*(\boldsymbol{\Phi},\boldsymbol{\Theta},\mu,\mathbf{Q}|\mathbf{w}) = -\frac{mn}{2}\left[\log\left(\frac{2\pi}{mn}\right) + 1\right] - \frac{n}{2}\log(\Pi_1\,\Pi_2)\ , \tag{3.1}$$

where

$$\Pi_1 = (\eta^T\eta - \lambda^T\lambda)^m\ , \tag{3.2}$$

$$\Pi_2 = |\mathbf{Q}|\ |\mathbf{D}|^{\frac{1}{n}}\ , \tag{3.3}$$

and the $gm \times gm$ matrix $\mathbf{D}$ is given by $\mathbf{D} = \mathbf{I} + \mathbf{M}^T\mathbf{H}^T\mathbf{H}\mathbf{M}$ [see (2.15) and (2.16)]. Thus, maximizing (3.1) is equivalent to minimizing

$$\Pi = \Pi_1\,\Pi_2\ . \tag{3.4}$$

Let Π_0 be the value of (3.4) at the initial estimates of the parameters ($\Pi_0 = \Pi_{10}\Pi_{20}$). Thus, if we minimize, instead of (3.4), the function

$$F = \frac{\Pi}{\Pi_0} = \frac{\Pi_1}{\Pi_{10}}\,\frac{\Pi_2}{\Pi_{20}}\ , \tag{3.5}$$

using a routine that generates descent search directions in every iteration, then the objective function F always lies in the interval (0, 1). This fact has two advantages. On the one hand, it improves the overall accuracy and numerical stability of the minimization routine, especially in the computation of the gradient vector through finite differences. On the other hand, it provides a simple means of handling situations in which the algorithm generates new estimates that imply non-stationarity, non-invertibility and/or non-positive definiteness of the matrix $\mathbf{Q}$. In such instances, which can be detected as described in the previous section, the scaled objective function (3.5) is set to one. Thus, the minimization routine will reject these points and continue the search for an acceptable local optimum. Note that this strategy is basically the one proposed in Shea (1984, pp. 99-100), although we do not solve the determinantal polynomials $|\boldsymbol{\Phi}(B)| = 0$ and $|\boldsymbol{\Theta}(B)| = 0$ in order to check for non-stationarity and non-invertibility, since the new evaluation algorithm provides simpler means for carrying out those checks.

In order to generate improving search directions, we use a quasi-Newton method based on the factorized version of the BFGS formula (see, for example, Dennis and Schnabel 1983, ch. 9). Besides

computational efficiency, this method provides, as a byproduct, a means of estimating the covariance matrix of the parameter estimates, since the relevant information on the curvature of the objective function (3.5) is updated at every iteration along with the computation of the search direction. From (3.1), the information matrix is given by

$$\mathbf{I} = E\left[-\frac{n}{2\Pi^2}\nabla\Pi\nabla\Pi^T + \frac{n}{2\Pi}\nabla^2\Pi\right],$$

where $\nabla\Pi$ is the gradient vector of (3.1) and $\nabla^2\Pi$ the hessian matrix. Since $\nabla\Pi = \mathbf{0}$ at any local optimum, a sample estimate of the covariance matrix is given by $2F(n\nabla^2 F)^{-1}$, where F and $(\nabla^2 F)^{-1}$ are evaluated at the final estimates. If we use a quasi-Newton method based on the factorized version of the BFGS formula to minimize (3.5), we will have at the end of the iterative process an approximation to the Cholesky factor of $\nabla^2 F$ (see Dennis and Schnabel 1983, pp. 206-207), which makes the computation of the covariance matrix estimate straightforward.

In summary, we suggest the use of the following procedure to maximize the exact likelihood function of a vector ARMA model:

(1) Choose a suitable set of initial estimates of the parameters, and compute Π_{10} and Π_{20} from (3.2) and (3.3) respectively.

(2) Minimize the scaled objective function (3.5) using a quasi-Newton method based on the factorized version of the BFGS formula.

(3) On convergence, use the accumulated information on the Cholesky factor of $\nabla^2 F$ to evaluate a sample estimate of the covariance matrix as $2F(n\nabla^2 F)^{-1}$.

Each time we compute Π_1 and Π_2 from (3.2) and (3.3), we make use of the evaluation algorithm of the previous section, and set the scaled objective function (3.5) to one whenever the algorithm

detects non-stationarity, non-invertibility and/or non-positive definiteness of $\mathbf{Q}$. It must be noted that, for a mixed model, the algorithm may converge to a non-stationary point. Although this has never happened in practice, the computation (on convergence) of the roots of $|\Phi(B)| = 0$, should be performed to ensure that the final estimates are admissible. The residual vector is evaluated only after the minimization routine has converged, since it is not used during the iterative process. Note also that all of these computations can be speeded up using the approximation to the exact likelihood function discussed in the previous section. Finally, initial estimates may be conditional maximum likelihood estimates or those obtained with other fast linear estimation methods (e.g. Shea 1987; Koreisha and Pukkila 1989), although care must be taken to ensure that they are admissible.

4. AN EXAMPLE

It is well known (e.g. Hillmer and Tiao 1979; Ansley and Newbold 1980) that exact maximum likelihood estimation is usually preferable to other approximate estimation criteria, especially in the case of small to moderate sized samples and/or parameters close to the boundaries of the admissible regions. This issue is not pursued further here. Instead, it will be illustrated with an example of how actual situations may occur in which the estimation method proposed in this paper performs better than one of the most frequently used in practice, the 'exact' version of the procedure of Hillmer and Tiao (1979) as implemented in The PC SCA Statistical System, release 4.1 (see Liu and Hudak 1992, pp. 5.15-5.16).

A series of 120 monthly observations on the Energy component of the Spanish Industrial Production Index, covering the period January 1982 through December 1991, has been considered. The data were obtained from the Boletín Estadístico del Banco de España (Banco de España. Alcalá, 50. 28014-Madrid, Spain) and are available on request from the author.

Some alternative patterns of differentiation were tried on the original series, and it seems clear that $w_t = \nabla\nabla_{12}z_t$, where z_t denotes the natural logarithm of the original series, can be considered to be stationary (see Figure 1). The autocorrelation and partial autocorrelation functions for w_t (see Figure 1) suggest that this time series might be described by an MA(1)×MA(1)$_{12}$ model.

[INSERT FIGURE 1]

However, non-sample information suggests the inclusion of both an AR(2)$_{12}$ factor, expected to have imaginary roots and a period of about four years, which is a very usual structure for Spanish industrial production series, and two deterministic variables, ξ_{t1} and ξ_{t2}, representing a unit impulse effect in February 1990 and the occurrence of the Easter holiday, respectively. Thus, the following intervention model was specified:

$$z_t = \omega_1\xi_{t1} + \omega_2\xi_{t2} + N_t\ , \tag{4.1}$$

$$(1 - \Phi_1B^{12} - \Phi_2B^{24})\nabla\nabla_{12}N_t = (1 - \theta_1B)(1 - \Theta_1B^{12})a_t\ . \tag{4.2}$$

The estimates of (4.1) and (4.2) obtained with both the exact and the approximate versions of the new algorithm, and those obtained with the algorithm of Hillmer and Tiao (1979), are summarized in Table 4.

[INSERT TABLE 4]

From these results, it can be seen that the estimates obtained with the exact and the approximate (with $\delta = .01$) versions of the new algorithm are almost identical. Further, these estimates are close to those obtained with the procedure of Hillmer and Tiao (1979), except for the seasonal moving average parameter, which is estimated to be non-invertible using that method, whereas it is invertible if we rely on the new algorithm. This substantial difference can be explained by examining the

residuals depicted in Figures 2 and 3.

[INSERT FIGURES 2 AND 3]

Figure 2 presents the (standardized) residuals, evaluated from (2.23), corresponding to the estimates obtained with the new algorithm, along with its autocorrelation and partial autocorrelation functions. The same information is presented in Figure 3, using the output generated by the procedure of Hillmer and Tiao (1979).

With regard to Figure 3, it may be noted that the first 24 (order of the autoregressive operator) residuals are not available, since the first p observations of w_t are used by the method of Hillmer and Tiao (1979) as starting values to compute a sequence supposedly generated by the moving average part of the model. Further, the autocorrelation and partial autocorrelation functions of Figure 3 suggest the need for a first order seasonal moving average operator, which, in fact, is already included but estimated to be non-invertible.

The problem lies in the presence of the observation corresponding to January 1985, which shows a residual only slightly larger than two standard deviations in Figure 2, though it pushes the seasonal moving average operator out of the invertibility region when the method of Hillmer and Tiao (1979) is used. To see this, the following intervention model was estimated:

$$z_t = \omega_1\xi_{t1} + \omega_2\xi_{t2} + \omega_3\xi_{t3} + N_t\ , \tag{4.3}$$

$$(1 - \Phi_1B^{12} - \Phi_2B^{24})\nabla\nabla_{12}N_t = (1 - \theta_1B)(1 - \Theta_1B^{12})a_t\ , \tag{4.4}$$

where ξ_{t3} is a unit impulse variable in January 1985. The estimates of (4.2) and (4.3) are summarized in Table 5.

[INSERT TABLE 5]

The seasonal moving average operator now lies within the invertibility region using either the new algorithm or the procedure of Hillmer and Tiao (1979). This was, in fact, the case in Table 4 when the new algorithm was used. Finally, note that a situation of this kind would be hard to detect if the method of Hillmer and Tiao (1979) were used, since the residual corresponding to January 1985 does not appear in Figure 3. Of course, if the specified model did not contain any AR operator, the estimates obtained with either method would be almost identical.

5. CONCLUSIONS

Both the theoretical development and the illustration of the performance of the estimation algorithm proposed in this paper, have shown the following important points:

1. It is possible to improve, as in Section 2, existing methods of evaluating the exact likelihood function of vector ARMA models, in order to put together a set of useful features which can only be found separately in currently available methods.

2. The adaptation of a computationally efficient minimization routine to those features yields an estimation procedure, which not only provides true maximum likelihood estimates, but also provides useful instruments for diagnostic checking of the fitted models.

3. Actual situations may occur in which the new estimation algorithm performs better than the ones frequently used.

The procedures outlined in this paper can be taken as a starting point in the development of new methods of estimating some generalizations of the vector ARMA model, such as the joint estimation of both the ARMA parameters and the coefficients of common non-stationary factors in multivariate models with series containing such factors. Other applications of those procedures, including the joint

estimation of both the ARMA structure and the deterministic components associated with a vector of time series, are straightforward.

REFERENCES

Ansley, C. F. (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process," Biometrika, 64, 59-65.

Ansley, C. F., and Newbold, P. (1980), "Finite Sample Properties of Estimators for Autoregressive-Moving Average Models," Journal of Econometrics, 13, 159-183.

Dennis, J. E., and Schnabel, R. B. (1983), Numerical Methods for Unconstrained Optimization and Nonlinear Equations, New Jersey: Prentice-Hall.

Hannan, E. J. (1969), "The Identification of Vector Mixed Autoregressive-Moving Average Systems," Biometrika, 56, 223-225.

Hall, A. D., and Nicholls, D. F. (1980), "The Evaluation of Exact Maximum Likelihood Estimates for VARMA Models," Journal of Statistical Computation and Simulation, 10, 251-262.

Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," Journal of the American StatisticalAssociation, 74, 652-660.

Kohn, R., and Ansley, C. F. (1982), "A Note on Obtaining the Theoretical Autocovariances of an ARMA Process," Journal of Statistical Computation and Simulation, 15, 273-283.

Koreisha, S., and Pukkila, T. (1989), "Fast Linear Estimation Methods for Vector Autoregressive Moving Average Models," Journal of Time Series Analysis, 10, 325-339.

Liu, L. M., and Hudak, G. B. (1992), Forecasting and Time Series Analysis Using the SCA Statistical System (Vol. 1), DeKalb: Scientific Computing Associates.

Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," Biometrika, 65, 297-303.

Ljung, G. M., and Box, G. E. P. (1979), "The Likelihood Function of Stationary Autoregressive-Moving Average Models," Biometrika, 66, 265-270.

Nicholls, D. F., and Hall, A. D. (1979), "The Exact Likelihood Function of Multivariate Autoregressive-Moving Average Models," Biometrika, 66, 259-264.

Shea, B. L. (1984), "Maximum Likelihood Estimation of Multivariate ARMA Processes via the Kalman Filter," in Time Series Analysis: Theory and Practice (ed. O. D. Anderson), vol. 5, pp. 91-101, Amsterdam: North-Holland.

Shea, B. L. (1987), "Estimation of Multivariate Time Series," Journal of Time Series Analysis, 8, 95-109.

Shea, B. L. (1989), "Algorithm AS 242. The Exact Likelihood of a Vector Autoregressive-Moving Average Model," Applied Statistics, 38, 161-204.

Table 1. Ratio between the number of time-consuming operations required by the algorithm of Hall and Nicholls (1980) and those required by the new algorithm, to evaluate the exact likelihood function for various models

MODELS	Models for annual data ($n = 25$)	
	$m = 2$	$m = 4$
AR(1)	1.00	1.00
AR(2)	1.00	1.00
MA(1)	1.00	1.00
MA(2)	1.00	1.00
ARMA(1,1)	1.05	1.04
ARMA(2,1)	1.09	1.10
ARMA(1,2)	1.09	1.10
ARMA(2,2)	1.19	1.21
MODELS	**Models for quarterly data ($n = 100$)**	
	$m = 2$	$m = 4$
$AR(1)_4$	1.03	1.05
$MA(1)_4$	1.02	1.02
$ARMA(1,1)_4$	1.20	1.23
$AR(1) \times AR(1)_4$	1.04	1.07
$AR(1) \times MA(1)_4$	1.07	1.08
$AR(1) \times ARMA(1,1)_4$	1.24	1.28
$MA(1) \times AR(1)_4$	1.06	1.07
$MA(1) \times MA(1)_4$	1.03	1.04
$MA(1) \times ARMA(1,1)_4$	1.24	1.28
$ARMA(1,1) \times AR(1)_4$	1.07	1.09
$ARMA(1,1) \times MA(1)_4$	1.09	1.11
$ARMA(1,1) \times ARMA(1,1)_4$	1.30	1.34
MODELS	**Models for monthly data ($n = 300$)**	
	$m = 2$	$m = 4$
$AR(1)_{12}$	1.09	1.11
$MA(1)_{12}$	1.07	1.09
$ARMA(1,1)_{12}$	1.50	1.59
$AR(1) \times AR(1)_{12}$	1.10	1.12
$AR(1) \times MA(1)_{12}$	1.13	1.15
$AR(1) \times ARMA(1,1)_{12}$	1.52	1.61
$MA(1) \times AR(1)_{12}$	1.08	1.10
$MA(1) \times MA(1)_{12}$	1.09	1.10
$MA(1) \times ARMA(1,1)_{12}$	1.53	1.61
$ARMA(1,1) \times AR(1)_{12}$	1.09	1.11
$ARMA(1,1) \times MA(1)_{12}$	1.15	1.75
$ARMA(1,1) \times ARMA(1,1)_{12}$	1.57	1.65

NOTE: The operations required to compute the first $p-1$ autocovariance and the first $q-1$ cross-covariance matrices have been excluded, since they are required by both algorithms.

Table 2. Ratio between the number of time-consuming operations required by the algorithm of Shea (1989) and those required by the new algorithm, to evaluate the exact likelihood function for various models

MODELS	Models for annual data ($n = 25$)	
	$m = 2$	$m = 4$
AR(1)	1.06 *	1.20 *
AR(2)	1.00 *	1.07 *
MA(1)	2.37 *	2.69 *
MA(2)	1.75 *	1.92 *
ARMA(1,1)	2.37 *	2.75 *
ARMA(2,1)	1.83 *	2.00 *
ARMA(1,2)	1.76 *	1.95 *
ARMA(2,2)	1.74 *	1.94 *
MODELS	**Models for quarterly data ($n = 100$)**	
	$m = 2$	$m = 4$
$AR(1)_4$	0.84	0.77
$MA(1)_4$	1.48 *	1.60 *
$ARMA(1,1)_4$	1.56 *	1.72 *
$AR(1) \times AR(1)_4$	0.74	0.65
$AR(1) \times MA(1)_4$	1.51 *	1.65 *
$AR(1) \times ARMA(1,1)_4$	1.49 *	1.57 *
$MA(1) \times AR(1)_4$	1.53 *	1.60 *
$MA(1) \times MA(1)_4$	1.36 *	1.46 *
$MA(1) \times ARMA(1,1)_4$	1.41 *	1.54 *
$ARMA(1,1) \times AR(1)_4$	1.34 *	1.37 *
$ARMA(1,1) \times MA(1)_4$	1.38 *	1.49 *
$ARMA(1,1) \times ARMA(1,1)_4$	1.41 *	1.53 *
MODELS	**Models for monthly data ($n = 300$)**	
	$m = 2$	$m = 4$
$AR(1)_{12}$	0.48	0.36
$MA(1)_{12}$	1.09 *	1.16 *
$ARMA(1,1)_{12}$	1.12 *	1.19 *
$AR(1) \times AR(1)_{12}$	0.44	0.33
$AR(1) \times MA(1)_{12}$	1.10 *	1.17 *
$AR(1) \times ARMA(1,1)_{12}$	1.15 *	1.17 *
$MA(1) \times AR(1)_{12}$	0.93	0.89
$MA(1) \times MA(1)_{12}$	1.05 *	1.11 *
$MA(1) \times ARMA(1,1)_{12}$	1.08 *	1.13 *
$ARMA(1,1) \times AR(1)_{12}$	0.87	0.82
$ARMA(1,1) \times MA(1)_{12}$	1.06 *	1.12 *
$ARMA(1,1) \times ARMA(1,1)_{12}$	1.07 *	1.12 *

NOTES: An asterisk indicates that the new algorithm is preferred. The operations required to compute the first $p-1$ autocovariance and the first $q-1$ cross-covariance matrices have been excluded, since they are required by both algorithms.

Table 3. Comparison of the various algorithms for evaluating the likelihood function

	Hall and Nicholls (1980)	Ljung and Box (1979)	Hillmer and Tiao (1979)	Shea (1987, 1989)	New algorithm
Exact evaluation	Yes	Yes	No[a]	Yes	Yes
Approximate evaluation	Yes[b]	Yes[b]	Yes[c]	Yes	Yes
Computational efficiency	No	No	Yes[d]	Yes[e]	Yes[e]
Numerical precision and stability	Yes	No	No[a]	Yes	Yes
Detection of non-stationarity	Yes	Yes	No	Yes	Yes
Detection of non-invertibility	Yes	Yes	Yes	No	Yes
Calculation of appropriate residuals	Yes[b]	Yes	No[a]	Yes	Yes

[a] Except in the case of pure moving average models ($p = 0$).
[b] Though this feature is not mentioned nor exploited in the papers cited.
[c] In fact, this is the only possibility offered by this method, which is based on questionable assumptions and does not permit control of the accuracy of the approximation.
[d] Only when the model contains an autoregressive part ($p \neq 0$).
[e] See the comparison in Table 2.

Table 4. Estimation of the model (4.1)-(4.2)[a,b]

	Exact[c]	Approximate[c]	Hillmer and Tiao (1979)[d]
ω_1	−.04 (.01)	−.04 (.01)	−.04 (.01)
ω_2	−.09 (.02)	−.09 (.02)	−.09 (.02)
Φ_1	−.04 (.14)	−.04 (.14)	−.11 (.10)
Φ_2	−.23 (.12)	−.24 (.12)	−.23 (.10)
θ_1	.54 (.16)	.54 (.16)	.63 (.09)
Θ_1	.75 (.15)	.76 (.15)	1.09 (.08)
σ_a	.0293	.0293	.0240

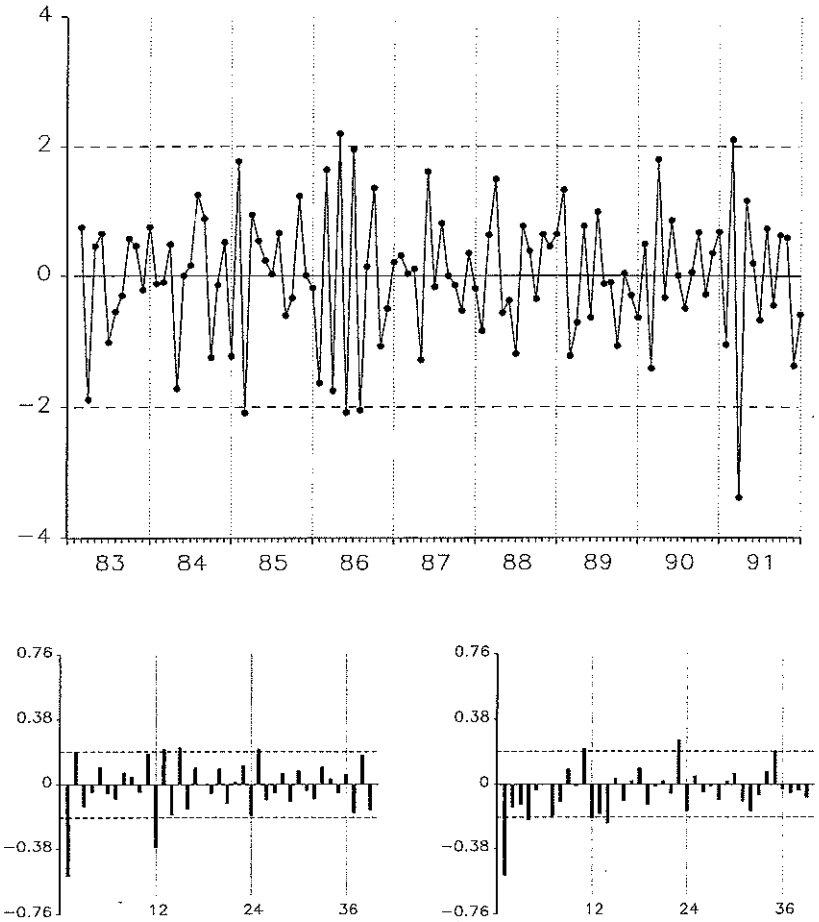
a Initial estimates: $\omega_1 = -.03$, $\omega_2 = -.09$, $\Phi_1 = .1$, $\Phi_2 = -.1$, $\theta_1 = .5$, $\Theta_1 = .5$.
b Estimated standard errors in parentheses.
c Convergence obtained in 36 iterations.
d Convergence not obtained within 100 iterations. The procedure was restarted using as initial estimates the final ones from the first column, but it did not converge.

Table 5. Estimation of the model (4.2)-(4.3)[a,b]

	Exact[c]	Approximate[c]	Hillmer and Tiao (1979)[d]
ω_1	−.04 (.01)	−.04 (.01)	−.04 (.01)
ω_2	−.10 (.02)	−.10 (.02)	−.10 (.02)
ω_3	.08 (.02)	.08 (.02)	.07 (.02)
Φ_1	−.12 (.14)	−.12 (.14)	−.17 (.11)
Φ_2	−.34 (.12)	−.33 (.12)	−.31 (.10)
θ_1	.39 (.15)	.39 (.15)	.50 (.10)
Θ_1	.64 (.15)	.64 (.15)	.70 (.09)
σ_a	.0280	.0280	.0266

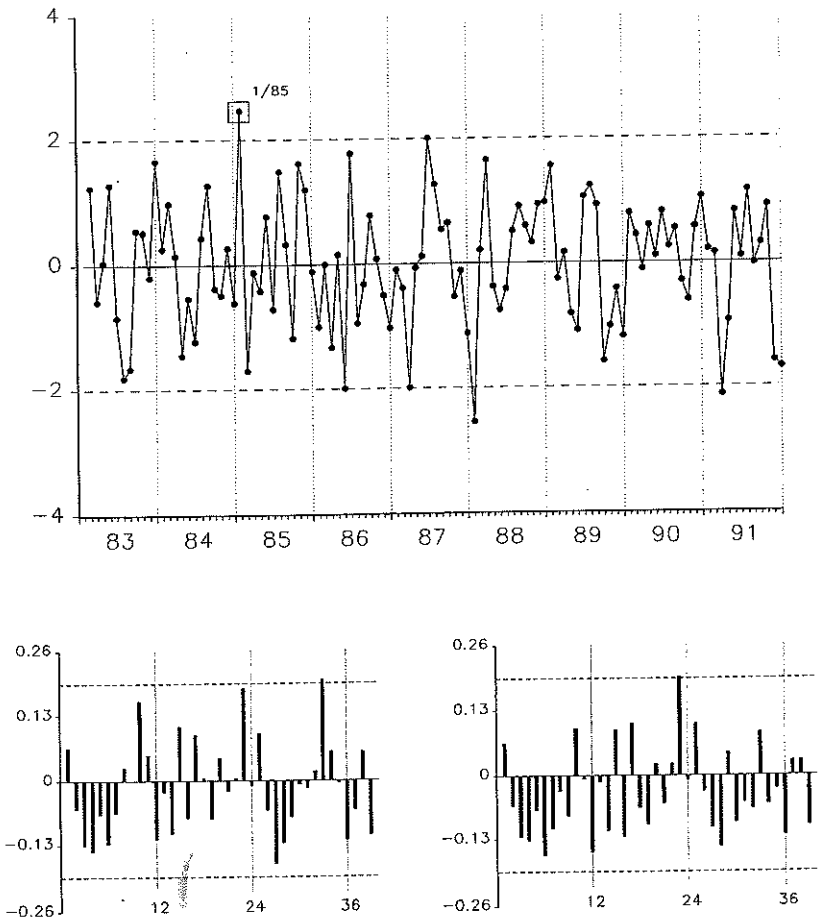
a Initial estimates: $\omega_1 = -.03$, $\omega_2 = -.09$, $\omega_3 = .08$, $\Phi_1 = .1$, $\Phi_2 = -.1$, $\theta_1 = .5$, $\Theta_1 = .5$.
b Esimated standard errors in parentheses.
c Convergence obtained in 41 iterations.
d Convergenge obtained in 50 iterations.

Figure 1. Series $w_t = \nabla\nabla_{12}z_t$: standardized series (top)[a], autocorrelation (bottom left)[b] and partial autocorrelation (bottom right) functions.



a $\bar{w} = -.0011$ (.0051); $\sigma_w = .0524$.
b Ljung-Box (1978) statistic: $Q(39) = 118.0$. This high value of the Q statistic merely reflects the so far unmodeled structure observed in the autocorrelation and partial autocorrelation functions.

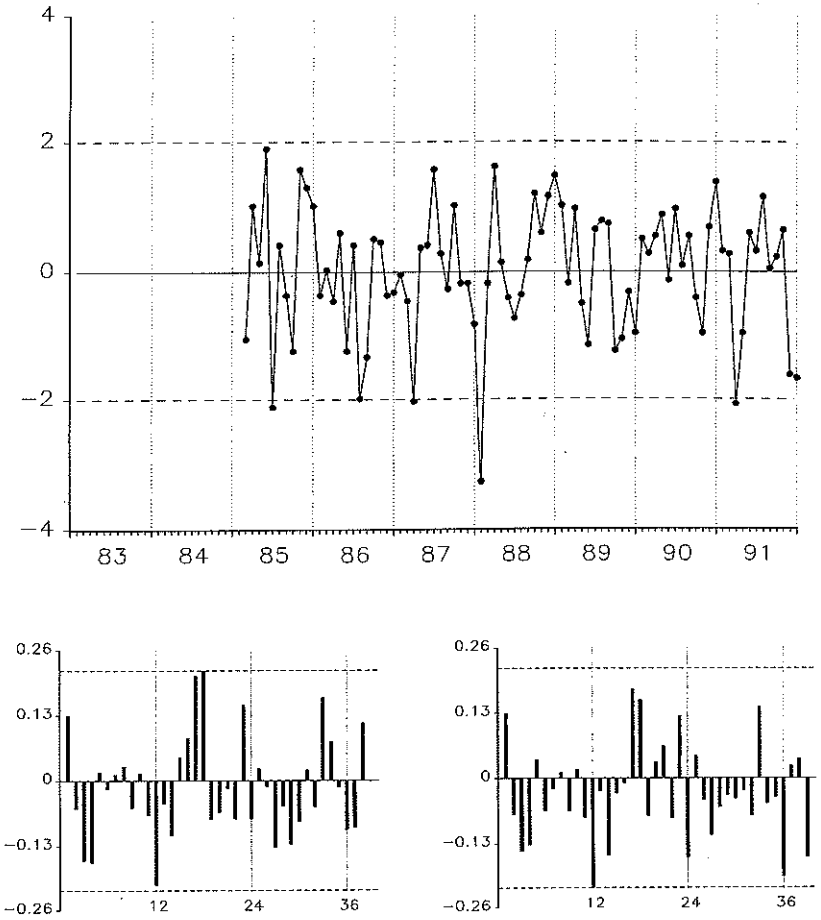
Figure 2. Estimation of (4.1)-(4.2) with the new algorithm: standardized residuals (top)[a], autocorrelation (bottom left)[b] and partial autocorrelation (bottom right) functions.



[a] $\bar{a} = -.0016$ (.0027); $\hat{\sigma}_a = .0293$.

[b] Ljung-Box (1978) statistic: $Q(33) = 45.8$. Although $Q(33)$ suggests misspecification, no structure is appreciated in the autocorrelation function (most likely, that value is due to outliers).

Figure 3. Estimation of (4.1)-(4.2) with the algorithm of Hillmer and Tiao (1979): standardized residuals (top)[a], autocorrelation (bottom left)[b] and partial autocorrelation (bottom right) functions.



[a] $\bar{a} = -.0001$ (.0024); $\hat{\sigma}_a = .0240$.

[b] Ljung-Box (1978) statistic: $Q(33) = 44.3$. The autocorrelation and partial autocorrelation functions suggest the need for a first order seasonal moving average operator.