MAXIMUM LIKELIHOOD ESTIMATION
OF VECTOR AUTOREGRESSIVE MOVING
AVERAGE MODELS

by

Greg Reinsel

Technical Report No. 117
ONR Report No. 3
July, 1976

Research supported by the Office of Naval Research under
Contract No. N00014-76-C-0930 (NR 042-367).

Reproduction in whole or in part is permitted for any
purpose of the United States Government. Distribution
of this document is unlimited.

Department of Statistics
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213
Summary

A method is presented for the estimation of the parameters in the vector autoregressive moving average time series model. The estimation procedure is derived from the maximum likelihood approach and is based on Newton-Raphson techniques applied to the likelihood equations. The resulting two-step Newton-Raphson procedure is computationally simple, involving only generalized least squares estimation in the second step. This Newton-Raphson estimator is shown to be asymptotically efficient and to possess a limiting multivariate normal distribution.

Key Words: Vector autoregressive moving average process, Newton-Raphson method, maximum likelihood estimation, Generalized least squares.
1. INTRODUCTION

In the past several years, following the work of Box and Jenkins (1970) and Hannan (1969b, 1970), there has been a great deal of interest in the use and estimation of autoregressive moving average models (ARMA models) for a scalar time series of observations. The vector ARMA model has received much less attention. Hannan (1969b, 1970) has considered the estimation of the pure moving average model in the vector case using spectral methods. Nicholls (1976a) has extended this work to the case of estimation of the mixed vector autoregressive moving average model which contains exogenous variables, the so called VARMAX model. An iterative nonlinear least squares (maximum likelihood) estimation procedure for the stationary vector ARMA model has been presented by Tunnicliffe Wilson (1973). Others who have considered the estimation of vector ARMA time series models include Akaike (1973), Kashyap (1970), and Whittle (1963). However, to a large extent, the problem of estimation of mixed vector models has been considered too computationally complicated to be practically useful. Also, the estimation theory is much more difficult than in the scalar case. (See the recent paper by Dunsmuir and Hannan (1976) for a treatment of the estimation theory of more general vector linear processes which include the ARMA models).

In this paper a method of estimation for vector ARMA models will be presented which is asymptotically efficient yet computationally simple. The method to be proposed uses the maximum likelihood approach.
and is based on Newton-Raphson techniques applied to the likelihood equations (in the time domain). The resulting "Newton-Raphson" estimator is shown to be asymptotically equivalent to the maximum likelihood estimator and to possess a limiting multivariate normal distribution.

Thus we consider the estimation of the parameters in the model

\[ Y_t - \sum_{i=1}^{p} A_i Y_{t-i} = \epsilon_t + \sum_{i=1}^{q} G_i \epsilon_{t-i}, \quad (1.1) \]

where \( Y_t = (y_{t1}, \ldots, y_{tm})' \) and \( \epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tm})' \) are vectors of \( m \) components, and \( A_1, \ldots, A_p, G_1, \ldots, G_q \) are \((m \times m)\) matrices of the unknown parameters. With the use of the lag operator \( \ell \) such that \( \ell^i Y_t = Y_{t-i} \), this vector model may be written as

\[ A'(\ell)Y_t = G'(\ell)\epsilon_t, \quad (1.2) \]

where
\[
A(\ell) = I - A_1 \ell - \cdots - A_p \ell^p
\]
and
\[
G(\ell) = I + G_1 \ell + \cdots + G_q \ell^q.
\]

The \( \epsilon_t \) are assumed to be independent and identically distributed with mean vector 0, covariance matrix \( \Sigma \), and finite fourth order moments. We also assume that all roots of

\[ \det A(z) = 0 \quad \text{and} \quad \det G(z) = 0 \]

are greater than 1 in absolute value and there are no common roots to the two equations. This assumption implies that the matrix series

\[
A(z)^{-1} = \sum_{i=0}^{\infty} J_i z^i \quad \text{and} \quad G(z)^{-1} = \sum_{i=0}^{\infty} D_i z^i
\]
both converge for \( |z| < 1 + \Delta, \Delta > 0 \), and that \( Y_t \) can be
expressed as the stationary solution to the difference equation (1.1) as

\[ Y_t = A'(z)^{-1}G'(z) \varepsilon_t = \sum_{u=0}^{\infty} \sum_{v=0}^{q} J_u^r G_v^r \varepsilon_{t-u-v}, \quad G_0 = I. \]

The problem of identification for models of the form (1.1) has been discussed in Hannan (1969a, 1975), and we assume that the identifiability conditions given there are satisfied. In particular, we assume that the rank of \((A'_p : G'_q)\) is equal to \(m\).

2. THE METHOD OF ESTIMATION

For the estimation of the parameters in the model (1.2), we assume that we have the vectors of observations \(Y_t, t=1,\ldots,T\). We define the \((T-p) \times m\) matrices

\[
Y = \begin{pmatrix}
Y'_{p+1} \\
\vdots \\
Y'_{T-1} \\
Y_T
\end{pmatrix}, \quad Z^iY = \begin{pmatrix}
Y'_{p+1-i} \\
\vdots \\
Y'_{T-1-i} \\
Y_T-i
\end{pmatrix}, \quad \varepsilon = \begin{pmatrix}
\varepsilon'_{p+1} \\
\vdots \\
\varepsilon'_{T-1} \\
\varepsilon_T
\end{pmatrix}, \quad Z^i \varepsilon = \begin{pmatrix}
\varepsilon'_{p+1-i} \\
\vdots \\
\varepsilon'_{T-1-i} \\
\varepsilon_T-i
\end{pmatrix}.
\]

Then we can write the entire model in matrix form as

\[ Y = \sum_{i=1}^{p} Z^i Y A_i + \sum_{i=1}^{q} Z^i \varepsilon G_i. \quad (2.1) \]

Alternatively, we may express the model more conveniently in vector form using the theory of Kronecker products of matrices. For any \((m \times n)\) matrix \(C\), let \(\text{vec}(C)\) denote the \((mn \times 1)\) column vector
formed from $C$ by placing the columns of $C$ below each other starting from left to right. Then a result which we shall use to write the model in vector form is that

$$\text{vec}(ABC) = (C' \otimes A) \text{vec}(B),$$

where $A$, $B$, and $C$ are any matrices such that the matrix product $ABC$ is defined. (See Neudecker (1969) for a proof of this and other results for Kronecker products.) Defining the vectors $y = \text{vec}(Y)$, $e = \text{vec}(\varepsilon)$, $z_i = \text{vec}(z_i Y)$, $z_i e = \text{vec}(z_i \varepsilon)$, $\alpha_i = \text{vec}(A_i)$, $i=1,\ldots,p$, and $y_i = \text{vec}(G_i)$, $i=1,\ldots,q$, we can write the entire model in vector form as

$$y - \sum_{i=1}^{p} (I_m \otimes z_i Y) \alpha_i = e + \sum_{i=1}^{q} (I_m \otimes z_i \varepsilon) y_i$$

(2.2)

To motivate the estimation procedure we suppose the $\varepsilon_t$ are normally distributed and use the maximum likelihood approach. To simplify the form of the likelihood function certain assumptions will be made concerning the initial observations and disturbances. First, we consider the initial observations $Y_1,\ldots,Y_p$ as fixed, and estimate from the likelihood function conditional on these values. Second, we assume that the initial disturbances $\varepsilon_{p+1-q},\ldots,\varepsilon_p$ are equal to their unconditional expectations, which are 0. Then introducing the $(T-p) \times (T-p)$ lag matrix $L$ which has 1's on the diagonal directly below the main diagonal and 0's elsewhere, we define the $(m(T-p)) \times (m(T-p))$ matrix

$$G = I_m(T-p) + \sum_{i=1}^{q} (G_i^t \otimes L^i),$$
with \( \det G = 1 \). (Note that \( G^{-1} = \sum_{i=0}^{T-p-1} (D_i \otimes L_i^t) \)). Thus we can express the entire (modified) model in vector notation as

\[
\begin{align*}
    y - \sum_{i=1}^{p} (I_m \otimes \mathcal{Z}_i^t \mathbf{Y}) \mathbf{a}_i &= e + \sum_{i=1}^{q} (I_m \otimes L_i^t \mathbf{e}) \mathbf{v}_i \\
    &= e + \sum_{i=1}^{q} (G_i \otimes L_i^t) e \\
    &= Ge .
\end{align*}
\]

(2.3)

Based on the assumption of normality of the \( \mathbf{e}_t \), the (modified) likelihood function of the observations \( Y_{p+1}, \ldots, Y_T \), conditional on \( Y_1, \ldots, Y_p \), is

\[
F = \frac{1}{(2\pi)^{\frac{mN}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left\{ \sum_{i=1}^{p} (I_m \otimes \mathcal{Z}_i^t \mathbf{Y}) \mathbf{a}_i \right\} ' (\Sigma^{-1} \otimes I_N) G^{-1} \left\{ y - \sum_{i=1}^{p} (I_m \otimes \mathcal{Z}_i^t \mathbf{a}_i) \right\} \right\} ,
\]

where \( N = T-p \). Thus we obtain the partial derivatives

\[
\frac{\partial \log F}{\partial \mathbf{a}_j} = (I_m \otimes \mathcal{Z}_j^t \mathbf{Y}) ' G^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} \left\{ y - \sum_{i=1}^{p} (I_m \otimes \mathcal{Z}_i^t \mathbf{a}_i) \right\} ,
\]

\( (j=1, \ldots, p) \),

\[
\frac{\partial \log F}{\partial \mathbf{Y}_j} = (I_m \otimes \mathcal{L}_j^t \mathbf{e}) ' G^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} \left\{ y - \sum_{i=1}^{p} (I_m \otimes \mathcal{Z}_i^t \mathbf{a}_i) \right\} ,
\]

\( (j=1, \ldots, q) \),

- 5 -
where $\epsilon$ is expressible in terms of the observable quantities $Y_t$ through equation (2.3) as

$$e = \text{vec}(\epsilon) = G^{-1}\left\{y - \sum_{i=1}^{P} (I_m \otimes \mathcal{L}^i Y)\alpha_i\right\}.$$  

Defining the vector

$$\theta = (\alpha'_1, \ldots, \alpha'_p, \gamma'_1, \ldots, \gamma'_q)'$$

and the matrix

$$W = ((I_m \otimes Y), \ldots, (I_m \otimes Y^2), (I_m \otimes Y), \ldots, (I_m \otimes Y^q)),$$

we can express these derivatives collectively as

$$\frac{\partial \log F}{\partial \theta} = W'G^{-1}(\Sigma^{-1} \otimes I_N)G^{-1}\left\{y - \sum_{i=1}^{P} (I_m \otimes \mathcal{L}^i Y)\alpha_i\right\}. \quad (2.4)$$

The maximum likelihood equations which result when (2.4) is set equal to 0 are nonlinear in the parameters $\theta$ (unless $q = 0$, in which case the solution is just the least squares estimate $\hat{\beta} = (W'W)^{-1}W'y$ which is nearly identical to the Yule-Walker estimate obtained in (a) below). Thus for $q > 0$ these equations can only be solved by numerical procedures such as the Newton-Raphson method. The Newton-Raphson equations for an approximate maximum likelihood estimator $\hat{\theta}$ are

$$- \frac{\partial^2 \log F}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0) = \left. \frac{\partial \log F}{\partial \theta} \right|_{\theta = \hat{\theta}} \quad (2.5)$$

- 6 -
where \( \theta_0 \) is an initial estimate of \( \theta \). It can be shown that on neglecting terms which, when divided by \( T \), converge to 0 in probability as \( T \to \infty \), we will have the approximation

\[
\frac{\partial^2 \log F}{\partial \theta_0} = W'G^{-1}(\Sigma^{-1} \otimes I_N)G^{-1}W.
\] (2.6)

To obtain the Newton-Raphson estimator of \( \theta \), we assume that we have initial estimates, \( \theta_0 = (\bar{\alpha}', \ldots, \bar{\alpha}_p', \bar{\nu}_1', \ldots, \bar{\nu}_q')' \) and \( \Sigma \), which are consistent estimates of \( \theta \) and \( \Sigma \) to the order \( T^{-1} \) in probability. These estimates may be obtained as follows:

(a) Compute the sample autocovariances

\[
C(s) = \frac{1}{T} \sum_{t=1}^{T-S} (Y_t - \bar{Y})(Y_{t+S} - \bar{Y})' = C'(-s), \quad (s=0,1,\ldots,p+q),
\]

where \( \bar{Y} = \frac{1}{T} \sum_{t=1}^{T} Y_t \). We then estimate the parameters \( \alpha_1 = \text{vec}(A_1) \) consistently by solving the vector Yule-Walker type equations

\[
C(-s) - \sum_{i=1}^{P} \bar{A}_i'C(i-s) = 0, \quad (s=q+1,\ldots,q+p).
\]

(b) Having obtained the estimates \( \bar{A}_1 \), we form

\[
C_v(s) = \sum_{j=0}^{P} \sum_{l=0}^{P} \bar{A}_j' C(s+j-l) \bar{A}_l = C'_v(-s), \quad (s=0,1,\ldots,q),
\]

where \( \bar{A}_0 = -I_m \), and

\[
\bar{\varphi}_v(\lambda) = \frac{1}{2\pi} \sum_{s=-q}^{q} C_v(s)e^{-i\lambda s}, \quad -\pi \leq \lambda \leq \pi.
\]

These are consistent estimates of the autocovariances and spectral density matrix, respectively, of the moving average process.
$V_t = \epsilon_t + \sum_{i=1}^{q} G_i' \epsilon_{t-1}$. If $\tilde{F}_v(\lambda)$ is an Hermitian nonnegative definite matrix, i.e., $\tilde{F}_v(\lambda) \geq 0$ for $-\pi \leq \lambda \leq \pi$, it may be factorized in the form

$$\tilde{F}_v(\lambda) = \frac{1}{2\pi} \tilde{G}'(e^{i\lambda}) \tilde{G}(e^{-i\lambda}),$$

to obtain consistent estimates $\tilde{v}_i = \text{vec}(\tilde{G}_i)$, $i=1, \ldots, q$, and $\Sigma$. Robinson (1967, pp. 190-200) describes a computer program for such a factorization. Consistent estimates of the $G_i$ may also be obtained by a method which does not require the factorizing of the spectral density matrix $\tilde{F}_v(\lambda)$ (see Hannan (1975), page 977, equation (5)).

Now we let

$$\tilde{G} = \text{Im}N + \sum_{i=1}^{q} (\tilde{G}_i' \otimes L_i^1),$$

$$\tilde{e} = \tilde{G}^{-1} \left\{ y - \sum_{i=1}^{p} (\text{Im} \otimes L^1Y) \tilde{G}_i \right\}, \quad \tilde{e} = \text{vec}(\tilde{G}),$$

$$\tilde{W} = ((\text{Im} \otimes L^1Y), \ldots, (\text{Im} \otimes L^pY), (\text{Im} \otimes L\tilde{e}), \ldots, (\text{Im} \otimes L^q\tilde{e})).$$

Then substituting (2.4) and (2.6) into equation (2.5), we obtain the following Newton-Raphson equations for $\theta$,

$$\tilde{W}' \tilde{G}^{-1} (\tilde{G}^{-1} \otimes \text{Im}N) \tilde{G}^{-1} \tilde{W}(\theta - \theta_0)$$

$$= \tilde{W}' \tilde{G}^{-1} (\tilde{G}^{-1} \otimes \text{Im}N) \tilde{G}^{-1} \left\{ y - \sum_{i=1}^{p} (\text{Im} \otimes L^1Y) \tilde{G}_i \right\}.$$

Since $\tilde{W} \theta_0 + y - \sum_{i=1}^{p} (\text{Im} \otimes L^1Y) \tilde{G}_i$

$$= y + \sum_{i=1}^{q} (\text{Im} \otimes L^1\tilde{e}) \tilde{v}_i.$$
the Newton-Raphson equations can be written in the form

$$W'G^{-1}(Z'Z)^{-1}W\hat{\theta} = W'G^{-1}(Z'Z)^{-1}W\hat{\theta} = \left\{ y + \sum_{i=1}^{q} (I_m \otimes L^i \bar{\xi})\bar{\gamma}_i \right\}, \quad (2.7)$$

The Newton-Raphson solution \( \hat{\theta} \) to equation (2.7) can be interpreted as the generalized least squares solution to the identity

$$y + \sum_{i=1}^{q} (I_m \otimes L^i \bar{\xi})\bar{\gamma}_i$$

$$= \sum_{i=1}^{p} (I_m \otimes L^i \bar{\xi})\alpha_i + \sum_{i=1}^{q} (I_m \otimes L^i \bar{\xi})\gamma_i + \bar{\theta}e$$

$$+ \sum_{i=1}^{q} \left( I_m \otimes (L^i \bar{\xi} - L^i \bar{\xi}) \right) (\bar{\gamma}_i - \gamma_i), \quad (2.8)$$

where the last term on the right hand side of the equation is to be neglected and the error term \( \bar{\theta}e \) is treated as having covariance matrix \( \bar{G}(\Sigma \otimes I_N)\bar{G}' \).

The asymptotic properties of the estimator \( \hat{\theta} \) will be given in the next section. Before proceeding to these, we conclude this section by briefly discussing the computations that are needed to complete the estimation procedure and obtain \( \hat{\theta} \) as the solution to equation (2.7). Once the initial estimate \( \theta_0 \) has been obtained, the vector

$$\bar{e} = \text{vec}(\bar{\xi}) = \bar{G}^{-1} \left\{ y - \sum_{i=1}^{p} (I_m \otimes L^i \bar{\xi})\bar{\alpha}_i \right\}$$

may be computed recursively from \( \bar{G} \bar{e} = y - \sum_{i=1}^{p} (I_m \otimes L^i \bar{\xi})\bar{\alpha}_i \) as

$$\bar{\xi}_t = Y_t - \sum_{i=1}^{p} \bar{K}_i Y_{t-i} - \sum_{i=1}^{q} \bar{G}_i \bar{\xi}_{t-i}, \quad (t=p+1, \ldots, T),$$
where \( \xi_{p+1-q} = \ldots = \xi_p = 0 \). Similar calculations may be used to obtain the columns of the matrix of "independent" variables 

\[
\bar{W} = \bar{G}^{-1}\bar{W} \text{ recursively from } \bar{G}\bar{W} = \bar{W},
\]

and the vector of the "dependent" variable 

\[
\bar{y} = \bar{G}^{-1}\left\{ y + \sum_{i=1}^{q} (I_m \otimes L_i^e)\bar{y}_i \right\} \text{ recursively from } 
\]

\[
\bar{G}\bar{y} = y + \sum_{i=1}^{q} (I_m \otimes L_i^e)\bar{y}_i.
\]

Then the estimator \( \hat{\theta} \) is simply the generalized least squares solution to the regression of \( \bar{y} \) on \( \bar{W} \), i.e.,

\[
\hat{\theta} = \left\{ \bar{W}'(\Sigma^{-1} \otimes I_N)\bar{W} \right\}^{-1} \bar{W}'(\Sigma^{-1} \otimes I_N)\bar{y}.
\]

3. THE ASYMPTOTIC DISTRIBUTION OF THE ESTIMATOR

To describe the asymptotic distribution of the estimator \( \hat{\theta} \) we introduce the matrices \( \Psi, \Phi, \) and \( \Omega \), whose \((g,h)\)th submatrices will be defined below. In expressing the following integrals, for notational convenience we will omit the argument variable \( e^{i\lambda} \) from the matrix polynomials \( A(e^{i\lambda}) \) and \( G(e^{i\lambda}) \) defined in Section 1, so that these will appear below simply as \( A \) and \( G \). Thus we define the submatrices

\[
\Psi_{g-h} = \lim_{T \to \infty} \frac{1}{T} E \left\{ (I_m \otimes \Sigma Y)'G^{-1}(\Sigma^{-1} \otimes I_N)G^{-1}(I_m \otimes \Sigma^h Y) \right\} 
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(g-h)\lambda} G^{-1}\Sigma^{-1}G^{-1} \otimes (A^{-1}G^{-1} \Sigma^{-1}G^* \hat{A}^{-1})' \, d\lambda,
\]

\((g,h=1, \ldots, p).\)
\[ \phi_{g-h} = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left\{ (I_m \otimes L^g \xi)' G^{-1}(\Sigma^{-1} \otimes I_N)G^{-1}(I_m \otimes L^h \xi) \right\} \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{(g-h)\lambda} G^{-1}_\Sigma^{-1} G^{-1}_* \otimes \Sigma \, d\lambda, \quad (g, h = 1, \ldots, q). \]

\[ \Omega_{g-h} = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left\{ (I_m \otimes L^g Y)' G^{-1}(\Sigma^{-1} \otimes I_N)G^{-1}(I_m \otimes L^h \xi) \right\} \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{(g-h)\lambda} G^{-1}_\Sigma^{-1} G^{-1}_* \otimes (\Sigma G^{-1}_A)' \, d\lambda, \quad (g = 1, \ldots, p; \ h = 1, \ldots, q). \]

Then we have the following

THEOREM. For the model (1.1) under the assumptions stated in Section 1, let \( \hat{\Theta} \) denote the estimator of \( \Theta = \text{vec}(A_1 | \cdots | A_p | G_1 | \cdots | G_q)' \) as obtained from equation (2.7). Then the distribution of \( \sqrt{T}(\hat{\Theta} - \Theta) \) converges to a multivariate normal distribution as \( T \to \infty \), with mean vector 0 and covariance matrix equal to

\[ V^{-1} = \left(\begin{array}{cc} \Psi & \Omega \\ \Omega' & \Phi \end{array}\right)^{-1}, \quad (3.1) \]

Before considering the proof of the theorem, we make the following comments:

(1) The matrix \( V \) defined in (3.1) is equal to
the "asymptotic information" matrix of $\theta$. Thus the asymptotic distribution of $\hat{\theta}$ is identical to that of the maximum likelihood estimator based on the assumption of normality of the disturbances $\varepsilon_t$, so that $\hat{\theta}$ is asymptotically efficient in that sense. Also, $\hat{\theta}$ converges to $\theta$ in probability as $T \to \infty$.

(2) If we let $\sigma_{ij}$ denote the $(i,j)$th element of $\Sigma^{-1}$, then we have

$$\frac{\partial \log F}{\partial \sigma_{ij}} = \begin{cases} \frac{N}{2} \sigma_{ii} - \frac{1}{2} e'(E_{ii} \otimes I_N)e & \text{for } i=j, \\ N \sigma_{ij} - \frac{1}{2} e' \left\{ (E_{ij} + E_{ji}) \otimes I_N \right\} e & \text{for } i \neq j, \end{cases}$$

where $e = G^{-1} \left\{ y - \sum_{i=1}^{p} (I_m \otimes x_i')a_i \right\}$ and $E_{ij}$ denotes the $(m \times m)$ matrix with 1 in the $(i,j)$th position and zeroes elsewhere. We have used the fact that for a nonsingular matrix $\Sigma$,

$$\frac{\partial \log |\Sigma^{-1}|}{\partial \sigma_{ij}} = \begin{cases} \sigma_{ii} & \text{for } i=j, \\ 2\sigma_{ij} & \text{for } i \neq j. \end{cases}$$

(see Golderberger (1964), Chapter 1)

Thus an asymptotically efficient estimator of $\Sigma$ is given by

$$\hat{\Sigma} = \frac{1}{N} \hat{\varepsilon}' \hat{\varepsilon},$$

where $\hat{\varepsilon} = \text{vec}(\hat{\varepsilon})$ denotes the vector $\varepsilon$ above with estimates $\hat{\theta}$. 

- 12 -
used in place of $\theta$. This estimator $\hat{\theta}$ is asymptotically uncorrelated with $\theta$. Similarly, the covariance matrix of $\hat{\theta}$ may be estimated by

$$
\left\{ \hat{\theta}' \hat{\Omega}^{-1} (\hat{\Sigma}^{-1} \otimes I_N) \hat{\Omega}^{-1} \hat{\theta} \right\}^{-1}.
$$

(3) The estimation procedure for $\hat{\theta}$ can easily be modified to include the case where elements of the autoregressive moving average coefficient matrices $A_1, \ldots, A_p$, $G_1, \ldots, G_q$ are specified to be zero. We simply modify the definitions of the matrix $\tilde{G}$ and the vector $\tilde{\epsilon}$ according to these restrictions (although this is not necessary for the results of the theorem to hold), and in the matrix $\tilde{W}$ we delete the columns of $(I_m \otimes L_{i-1}^t Y)$, $i=1, \ldots, p$, and $(I_m \otimes L_{i-1}^t \tilde{\epsilon})$, $i=1, \ldots, q$, corresponding to the zero elements of $a_i = \text{vec}(A_i)$ and $\gamma_i = \text{vec}(G_i)$, respectively. The resulting estimator obtained from equation (2.7) will again be asymptotically efficient, consistent, and will possess an asymptotic normal distribution with covariance matrix similar to $V^{-1}$ (but with the appropriate rows and columns of $V$ deleted).

PROOF OF THEOREM: The proof of the theorem is much the same as that of a similar theorem for the scalar case given in Reinsel (1976a), and so we shall omit most of the details. We ignore the effect of the modification of the initial disturbances $\epsilon_t$, and hence the use of the matrix $G$ in equation (2.8), since the modification has a negligible effect as $T \to \infty$ and does not affect the asymptotic properties of the estimator. Then using the (modified) identity (2.8) and equation (2.7), we have
\[
\sqrt{T} (\hat{\theta} - \theta) = \left\{ \frac{1}{T} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) \mathbf{G}^{-1} \mathbf{W} \right\}^{-1} \cdot \frac{1}{\sqrt{T}} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) e
\]

\[
+ \left\{ \frac{1}{T} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) \mathbf{G}^{-1} \mathbf{W} \right\}^{-1} \sum_{i=1}^{q} \frac{1}{T} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) \mathbf{G}^{-1} \left\{ \mathbf{I}_m \otimes (\mathbf{L}^\top \mathbf{\varepsilon} - \mathbf{L}^\top \mathbf{\varepsilon}) \right\} \sqrt{T} (\overline{\gamma}_1 - \gamma_1).
\]

(3.2)

Now it can be shown that

\[
\text{plim} \lim_{T \to \infty} \frac{1}{T} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) \mathbf{G}^{-1} \mathbf{W} = \lim_{T \to \infty} \frac{1}{T} \mathbf{E} \left\{ \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) \mathbf{G}^{-1} \mathbf{W} \right\} = \mathbf{V}.
\]

Also, as in Reinsel(1976a), it can be shown that each of the summands in the second term on the right side of (3.2) has a probability limit equal to 0 by the consistency of the initial estimates (and the boundedness in probability of the vectors \( \sqrt{T} (\overline{\gamma}_1 - \gamma_1) \)). And finally, again by the consistency of the initial estimates, from (3.2) it is not hard to see that the asymptotic distribution of \( \sqrt{T} (\hat{\theta} - \theta) \) will be the same as that of \( \mathbf{V}^{-1} \cdot \frac{1}{\sqrt{T}} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) e \).

Thus it follows that the results of the theorem will be established once it is shown that the vector \( \frac{1}{\sqrt{T}} \mathbf{W}' \mathbf{G}'^{-1} \left( \Sigma^{-1} \otimes \mathbf{I}_N \right) e \) has a limiting normal distribution as \( T \to \infty \) with mean vector 0 and covariance matrix equal to \( \mathbf{V} \).

To establish the asymptotic normality of this vector, we consider in detail the asymptotic behavior of a subvector of
\[
\frac{1}{\sqrt{T}} W' G^{-1} (\Sigma^{-1} \otimes I_N) e. \text{ For example, we have}
\]
\[
\frac{1}{\sqrt{T}} (I_m \otimes L^4 Y)' G^{-1} (\Sigma^{-1} \otimes I_N) e
\]
\[
= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-p-1} \left\{ D_j \Sigma^{-1} \otimes (L^4 Y)' \right\} e
\]
\[
= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-p-1} \text{vec} \left\{ (L^4 Y)' \in \Sigma^{-1} D_j' \right\}
\]
\[
= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-p-1} \sum_{t=p+1}^{T} \text{vec}(Y_{t-i-j} e_t \Sigma^{-1} D_j')
\]
\[
= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{j=0}^{T-p-1} \text{vec}(Y_{t-i-j} e_t \Sigma^{-1} D_j')
\]
\[
= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{j=0}^{T-p-1} \sum_{u=0}^{\min(t-p-1,n)} \sum_{v=0}^{q} \text{vec}(J' \Sigma^{-1} e_t \Sigma^{-1} D_j')
\]
\[
= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{j=0}^{T-p-1} \sum_{u=0}^{\min(t-p-1,n)} \sum_{v=0}^{q} \text{vec}(J' \Sigma^{-1} e_t \Sigma^{-1} D_j') + R_{Tn}
\]
\[
= Z_{Tn} + R_{Tn},
\]
for \( n < T-p-1, \)
where
\[ Z_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} W_{tn}, \quad (3.4) \]

\[ W_{tn} = \sum_{j=0}^{\min(t-p-1,n)} \sum_{u=0}^{n} \sum_{v=0}^{q} \text{vec}(J'_u G_v \xi_{t-1-j-u-v} \xi_t^{-1} D'_j), \quad (3.5) \]

\[ (t=p+1, \ldots, T), \]

and
\[ R_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{j=0}^{\min(t-p-1,n)} \sum_{u=n+1}^{\infty} \sum_{v=0}^{q} \text{vec}(J'_u G_v \xi_{t-1-j-u-v} \xi_t^{-1} D'_j) \]
\[ + \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{j=0}^{\min(t-p-1,n)} \sum_{u=n+1}^{\infty} \sum_{v=0}^{q} \text{vec}(J'_u G_v \xi_{t-1-j-u-v} \xi_t^{-1} D'_j). \quad (3.6) \]

Now since the elements of the matrices \( J_u \) and \( D_u \) converge to 0 exponentially as \( u \to \infty \), it is not hard to show that the remainder term \( R_{Tn} \) converges to 0 in probability as \( n \to \infty \), uniformly in \( T \) (e.g., it can be shown that \( E(R'_{Tn} R_{Tn}) \to 0 \) as \( n \to \infty \), uniformly in \( T \)). For fixed \( n \), the vectors \( W_{tn} \) have 0 means, are uncorrelated, and the covariance matrix of \( W_{tn} \) is equal to
\[ E(W_{tn} W'_{tn}) = V_{tn} \]
\[ = \sum_{j,k=0}^{\min(t-p-1,n)} \sum_{u,m=0}^{n} \sum_{v,s=0}^{q} E \left\{ \text{vec}(J'_u G_v \xi_{t-1-j-u-v} \xi_t^{-1} D'_j) \cdot \text{vec}(J'_m G_s \xi_{t-1-k-m-s} \xi_t^{-1} D'_k) \right\} \]
\[
\min(t-p-1,n) = \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} e_t \otimes J_u' G_v') e_{t-1-j-u-v} e_{t-1-k-m-s} \right\}
\]

\[
= \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} e_t \otimes J_u' G_v') \Sigma (D_k^{\Sigma^{-1}} e_t \otimes J_m' G_s') \right\}
\]

\[
\min(t-p-1,n) = \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} e_t \otimes J_u' G_v') \Sigma (D_k^{\Sigma^{-1}} e_t \otimes J_m' G_s') \right\}
\]

\[
= \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} e_t \otimes J_u' G_v' \Sigma G_s J_m) \right\}
\]

\[
= \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} D_k' \otimes J_u' G_v' \Sigma G_s J_m) \right\}, \tag{3.7}
\]

where the summation extends over only those indices for which \(j+u+v = k+m+s\). Thus the covariance matrix of \(Z_{Tn}\) is

\[
E(Z_{Tn} Z'_{Tn}) = \frac{1}{T} \sum_{t=p+1}^{T} V_{tn}, \text{ and}
\]

\[
\lim_{T \to \infty} E(Z_{Tn} Z'_{Tn}) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=p+1}^{T} \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} D_k' \otimes J_u' G_v' \Sigma G_s J_m) \right\}
\]

\[
= \sum_{j,k=0}^{n} \sum_{u,m=0}^{q} \sum_{v,s=0}^{q} \left\{ (D_j^{\Sigma^{-1}} D_k' \otimes J_u' G_v' \Sigma G_s J_m) \right\}
\]

\[
= V_n. \tag{3.8}
\]
Also note that for fixed $n$, $\{W_{tn}\}_{t=p+1}^T$ forms a sequence of $m = (i+2n+q)$-dependent random vectors, and the $W_{tn}$ possess uniformly bounded fourth moments, since the $\varepsilon_t$ have finite fourth order moments. Thus it follows by Theorem 1 in Schönhed (1971) that

$$Z_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^T W_{tn}$$

has a limiting multivariate normal distribution as $T \to \infty$, with mean vector $0$ and covariance matrix $V_n$. Then using a vector analogue of Theorem 7.7.1 in Anderson (1971), as given by Lemma 2 in Schönhed (1971), we can conclude that

$$\frac{1}{\sqrt{T}} (I_m \otimes \mathbb{1}_Y) G^{-1}(\Sigma^{-1} \otimes I_N) e = Z_{Tn} + R_{Tn}$$

converges in distribution as $T \to \infty$ to a multivariate normal distribution with mean vector $0$ and covariance matrix equal to

$$\lim_{n \to \infty} V_n = \sum_{j,k=0}^\infty \sum_{u,m=0}^\infty \sum_{v,s=0}^q (D_j \Sigma^{-1} D_k \otimes J^t G^t \Sigma G J_m)$$

$$(j+u+v=k+m+s)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} G^{-1} \Sigma^{-1} G^{-1} \otimes (A^* \Sigma A^{-1} G^* G^{-1}) \, d\lambda$$

$$= V_0 .$$

(3.9)

A similar argument can be used to show that all the subvectors of

$$\frac{1}{\sqrt{T}} W^t G^{-1}(\Sigma^{-1} \otimes I_N) e$$

have a joint limiting distribution as $T \to \infty$ which is multivariate normal with mean vector $0$ and covariance matrix $V$, and hence the theorem is established.
4. EXTENSIONS

The theory and method of estimation in this paper can easily be extended to the case of the mixed vector autoregressive moving average model with exogenous variables, the so called VARMAX model, of the form

$$Y_t - \sum_{i=1}^{p} A_i'Y_{t-i} = \sum_{i=0}^{k} B_i'X_{t-1} + \epsilon_t + \sum_{i=1}^{q} G_i'\epsilon_{t-i},$$

where the observable exogenous variables $X_t$ are suitably prescribed. In fact, the method which has been used here for the estimation of the vector ARMA model has been applied by Reinsel (1976b) to the more general dynamic simultaneous equations model with autoregressive moving average disturbances,

$$(I - A_0)'Y_t - \sum_{i=1}^{r} A_i'Y_{t-i} = \sum_{i=0}^{k} B_i'X_{t-1} + U_t, \quad A_0 \neq 0,$$

where the disturbances $U_t$ satisfy

$$U_t - \sum_{i=1}^{p} R_i'U_{t-i} = \epsilon_t + \sum_{i=1}^{q} G_i'\epsilon_{t-i}.$$
REFERENCES


**Report Title:** Maximum likelihood estimation of vector autoregressive moving average models

**Authors:** Greg Reinsel

**Performing Organization:** Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pa. 15213

**Controlling Office:** Office of Naval Research, Statistics and Probability Program, Arlington, Virginia 22217

**Controlled Office:** Office of Naval Research

**Abstract:**
A method is presented for the estimation of the parameters in the vector autoregressive moving average time series model. The estimation procedure is derived from the maximum likelihood approach and is based on Newton-Raphson techniques applied to the likelihood equations. The resulting two-step Newton-Raphson procedure is computationally simple, involving only generalized least squares estimation in the second step. This Newton-Raphson estimator is shown to be consistent and asymptotically normal.
Block 20.

to be asymptotically efficient and to possess a limiting multivariate normal distribution.