

AD-A031 116

CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/2
MAXIMUM LIKELIHOOD ESTIMATION OF VECTOR AUTOREGRESSIVE MOVING A--ETC(U)
JUL 76 6 REINSEL N00014-76-C-0930
TR-117 NL

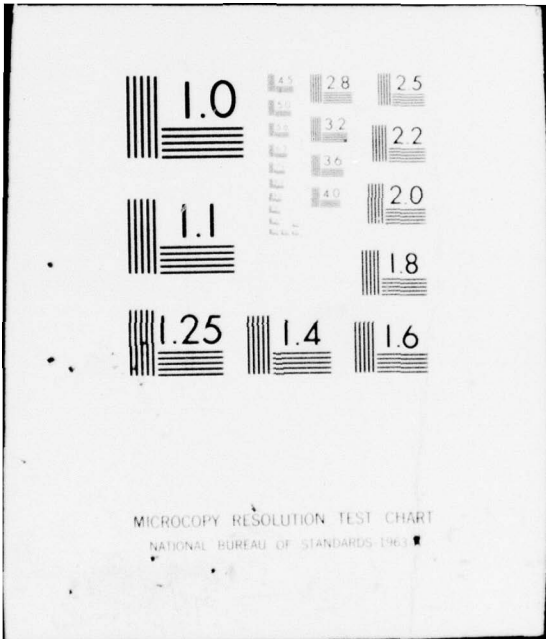
UNCLASSIFIED

1 of 1
AD
A031116



END

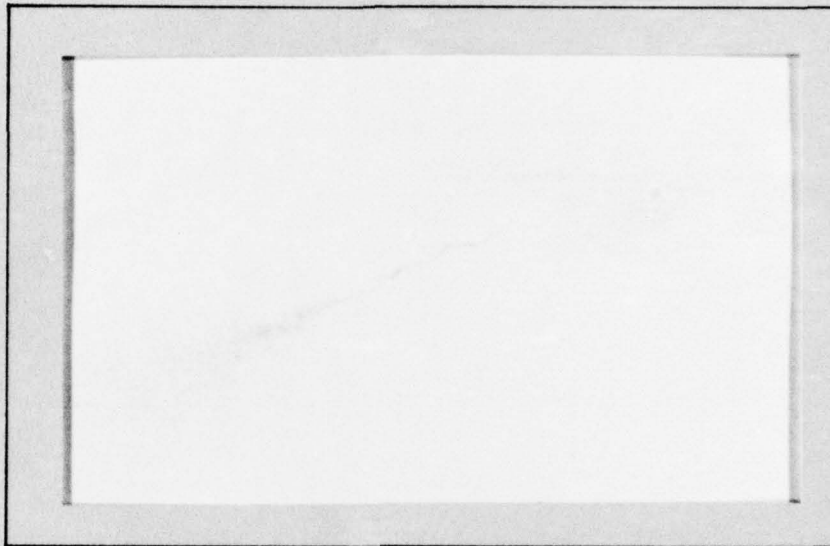
DATE
FILMED
11-76



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A031116

①



DEPARTMENT
OF
STATISTICS

DDC
RECEIVED
OCT 26 1976
A

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

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

see 1473

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.

ACCESSION FOR	
NTIS	Write Section <input checked="" type="checkbox"/>
DDI	Buy Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY	
DISSEMINATION/AVAILABILITY CODES	
ORIG.	2242L. NO. OF SPECIAL
A	

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

$$A'(\mathcal{L})Y_t = G'(\mathcal{L})\epsilon_t, \quad (1.2)$$

where

$$A(\mathcal{L}) = I - A_1 \mathcal{L} - \dots - A_p \mathcal{L}^p$$

and

$$G(\mathcal{L}) = I + G_1 \mathcal{L} + \dots + G_q \mathcal{L}^q.$$

The ϵ_t are assumed to be independent and identically distributed with mean vector 0, covariance matrix Σ , 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

$$Y_t = A'(\mathcal{L})^{-1} G'(\mathcal{L}) \epsilon_t = \sum_{u=0}^{\infty} \sum_{v=0}^q J'_u G'_v \epsilon_{t-u-v}, \quad G_0 = I.$$

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 , $t=1, \dots, T$. We define the $(T-p) \times m$ matrices

$$Y = \begin{pmatrix} Y'_{p+1} \\ \vdots \\ Y'_T \end{pmatrix}, \quad \mathcal{L}^i Y = \begin{pmatrix} Y'_{p+1-i} \\ \vdots \\ Y'_{T-i} \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon'_{p+1} \\ \vdots \\ \epsilon'_T \end{pmatrix}, \quad \mathcal{L}^i \epsilon = \begin{pmatrix} \epsilon'_{p+1-i} \\ \vdots \\ \epsilon'_{T-i} \end{pmatrix}.$$

Then we can write the entire model in matrix form as

$$Y - \sum_{i=1}^p \mathcal{L}^i Y A_i = \epsilon + \sum_{i=1}^q \mathcal{L}^i \epsilon G_i. \quad (2.1)$$

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 $y = \text{vec}(Y)$, $e = \text{vec}(\epsilon)$, $\mathcal{L}^i y = \text{vec}(\mathcal{L}^i Y)$, $\mathcal{L}^i e = \text{vec}(\mathcal{L}^i \epsilon)$, $\alpha_i = \text{vec}(A_i)$, $i=1, \dots, p$, and $\gamma_i = \text{vec}(G_i)$, $i=1, \dots, q$, we can write the entire model in vector form as

$$y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i = e + \sum_{i=1}^q (I_m \otimes \mathcal{L}^i \epsilon) \gamma_i \quad (2.2)$$

To motivate the estimation procedure we suppose the ϵ_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, \dots, Y_p as fixed, and estimate from the likelihood function conditional on these values. Second, we assume that the initial disturbances $\epsilon_{p+1-q}, \dots, \epsilon_p$, 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

$$G = I_{m(T-p)} + \sum_{i=1}^q (G_i' \otimes L^i),$$

with $\det G = 1$. (Note that $G^{-1} = \sum_{i=0}^{T-p-1} (D_i' \otimes L^i)$). Thus we can

express the entire (modified) model in vector notation as

$$\begin{aligned}
 y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i &= e + \sum_{i=1}^q (I_m \otimes L^i \epsilon) \gamma_i \\
 &= e + \sum_{i=1}^q (G_i' \otimes L^i) e \\
 &= Ge.
 \end{aligned} \tag{2.3}$$

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

$$\begin{aligned}
 F = \frac{1}{(2\pi)^{\frac{1}{2}mN} |\Sigma|^{\frac{1}{2}N}} \exp \left[-\frac{1}{2} \left\{ y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i \right\}' G'^{-1} \right. \\
 \left. \cdot (\Sigma^{-1} \otimes I_N) G^{-1} \left\{ y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i \right\} \right],
 \end{aligned}$$

where $N = T-p$. Thus we obtain the partial derivatives

$$\begin{aligned}
 \frac{\partial \log F}{\partial \alpha_j} &= (I_m \otimes \mathcal{L}^j Y)' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} \left\{ y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i \right\}, \\
 &\hspace{20em} (j=1, \dots, p), \\
 \frac{\partial \log F}{\partial \gamma_j} &= (I_m \otimes L^j \epsilon)' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} \left\{ y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i \right\}, \\
 &\hspace{20em} (j=1, \dots, q),
 \end{aligned}$$

where ϵ is expressible in terms of the observable quantities Y_t through equation (2.3) as

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

Defining the vector

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

$$= \text{vec}(A_1 | \dots | A_p | G_1 | \dots | G_q)'$$

and the matrix

$$W = ((I_m \otimes Z^i Y), \dots, (I_m \otimes Z^p Y), (I_m \otimes L\epsilon), \dots, (I_m \otimes L^q \epsilon)),$$

we can express these derivatives collectively as

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

The maximum likelihood equations which result when (2.4) is set equal to 0 are nonlinear in the parameters θ (unless $q=0$, in which case the solution is just the least squares estimate $\hat{\theta} = (W'W)^{-1}W'y$ 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

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

where θ_0 is an initial estimate of θ . It can be shown that on neglecting terms which, when divided by T , converge to 0 in probability as $T \rightarrow \infty$, we will have the approximation

$$-\frac{\partial^2 \log F}{\partial \theta \partial \theta'} = W' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} W \quad (2.6)$$

To obtain the Newton-Raphson estimator of θ , we assume that we have initial estimates, $\theta_0 = (\tilde{\alpha}'_1, \dots, \tilde{\alpha}'_p, \tilde{\gamma}'_1, \dots, \tilde{\gamma}'_q)'$ and $\tilde{\Sigma}$, which are consistent estimates of θ and Σ to the order $T^{-\frac{1}{2}}$ in probability. These estimates may be obtained as follows:

(a) Compute the sample autocovariances

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

where $\bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$. We then estimate the parameters $\alpha_i = \text{vec}(A_i)$

consistently by solving the vector Yule-Walker type equations

$$C(-s) - \sum_{i=1}^p \tilde{A}'_i C(i-s) = 0, \quad (s=q+1, \dots, q+p).$$

(b) Having obtained the estimates \tilde{A}_i , we form

$$C_V(s) = \sum_{j=0}^p \sum_{l=0}^p \tilde{A}'_j C(s+j-l) \tilde{A}_l = C'_V(-s), \quad (s=0, 1, \dots, q),$$

where $\tilde{A}_0 = -I_m$, and

$$\tilde{f}_V(\lambda) = \frac{1}{2\pi} \sum_{s=-q}^q C_V(s) e^{-is\lambda}, \quad -\pi \leq \lambda \leq \pi.$$

These are consistent estimates of the autocovariances and spectral density matrix, respectively, of the moving average process

$V_t = \epsilon_t + \sum_{i=1}^q G'_i \epsilon_{t-i}$. 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

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

to obtain consistent estimates $\tilde{\gamma}_i = \text{vec}(\tilde{G}_i)$, $i=1, \dots, q$, and $\tilde{\Sigma}$. 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

$$\tilde{G} = I_{mN} + \sum_{i=1}^q (\tilde{G}'_i \otimes L^i),$$

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

$$\tilde{W} = ((I_m \otimes \mathcal{L}^i Y), \dots, (I_m \otimes \mathcal{L}^p Y), (I_m \otimes L\tilde{\epsilon}), \dots, (I_m \otimes L^q \tilde{\epsilon})).$$

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

$$\begin{aligned} & \tilde{W}' \tilde{G}'^{-1} (\tilde{\Sigma}^{-1} \otimes I_N) \tilde{G}^{-1} \tilde{W} (\hat{\theta} - \theta_0) \\ & = \tilde{W}' \tilde{G}'^{-1} (\tilde{\Sigma}^{-1} \otimes I_N) \tilde{G}^{-1} \left\{ y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \tilde{\alpha}_i \right\}. \end{aligned}$$

Since $\tilde{W} \theta_0 + y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \tilde{\alpha}_i$

$$= y + \sum_{i=1}^q (I_m \otimes L^i \tilde{\epsilon}) \tilde{\gamma}_i,$$

the Newton-Raphson equations can be written in the form

$$\bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \bar{W} \hat{\theta} = \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \left\{ y + \sum_{i=1}^q (I_m \otimes L^i \bar{\epsilon}) \bar{y}_i \right\}. \quad (2.7)$$

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

$$\begin{aligned} y + \sum_{i=1}^q (I_m \otimes L^i \bar{\epsilon}) \bar{y}_i \\ = \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i + \sum_{i=1}^q (I_m \otimes L^i \bar{\epsilon}) \gamma_i + \bar{G} e \\ + \sum_{i=1}^q \left\{ I_m \otimes (L^i \bar{\epsilon} - L^i \epsilon) \right\} (\bar{y}_i - \gamma_i), \end{aligned} \quad (2.8)$$

where the last term on the right hand side of the equation is to be neglected and the error term $\bar{G} e$ 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 θ_0 has been obtained, the vector

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

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

$$\bar{\epsilon}_t = Y_t - \sum_{i=1}^p \bar{A}'_i Y_{t-1} - \sum_{i=1}^q \bar{G}'_i \bar{\epsilon}_{t-1}, \quad (t=p+1, \dots, T),$$

where $\tilde{\epsilon}_{p+1-q} = \dots = \tilde{\epsilon}_p = 0$. Similar calculations may be used to obtain the columns of the matrix of "independent" variables

$$\bar{W} = \bar{G}^{-1}\tilde{W} \text{ recursively from } \bar{G}\bar{W} = \tilde{W},$$

and the vector of the "dependent" variable

$$\bar{y} = \bar{G}^{-1}\left\{y + \sum_{i=1}^q (I_m \otimes L^i \tilde{\epsilon}) \tilde{y}_i\right\} \text{ recursively from}$$

$$\bar{G} \bar{y} = y + \sum_{i=1}^q (I_m \otimes L^i \tilde{\epsilon}) \tilde{y}_i.$$

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

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

3. THE ASYMPTOTIC DISTRIBUTION OF THE ESTIMATOR

To describe the asymptotic distribution of the estimator $\hat{\theta}$ we introduce the matrices $\Psi, \Phi,$ and Ω , 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

$$\Psi_{g-h} = \lim_{T \rightarrow \infty} \frac{1}{T} E \left\{ (I_m \otimes e^{gY})' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} (I_m \otimes e^{hY}) \right\}$$

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

$$(g, h=1, \dots, p).$$

$$\begin{aligned} \Phi_{g-h} &= \lim_{T \rightarrow \infty} \frac{1}{T} E \left\{ (I_m \otimes L^g \epsilon)' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} (I_m \otimes L^h \epsilon) \right\} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(g-h)\lambda} G^{-1} \Sigma^{-1} G'^{-1} \otimes \Sigma \, d\lambda, \quad (g, h=1, \dots, q). \end{aligned}$$

$$\begin{aligned} \Omega_{g-h} &= \lim_{T \rightarrow \infty} \frac{1}{T} E \left\{ (I_m \otimes L^g Y)' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} (I_m \otimes L^h \epsilon) \right\} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(g-h)\lambda} G^{-1} \Sigma^{-1} G'^{-1} \otimes (\Sigma G A^{-1})' \, d\lambda, \end{aligned}$$

$$(g=1, \dots, p; h=1, \dots, q).$$

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 $\theta = \text{vec}(A_1 | \dots | A_p | G_1 | \dots | G_q)'$ as obtained from equation (2.7). Then the distribution of $\sqrt{T}(\hat{\theta} - \theta)$ converges to a multivariate normal distribution as $T \rightarrow \infty$, with mean vector 0 and covariance matrix equal to

$$V^{-1} = \begin{pmatrix} \Psi & \Omega \\ \Omega' & \Phi \end{pmatrix}^{-1}. \quad (3.1)$$

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

(1) The matrix V defined in (3.1) is equal to

$$\lim_{T \rightarrow \infty} \frac{1}{T} E \left\{ W' G'^{-1} (\Sigma^{-1} \otimes I_N) G^{-1} W \right\} = \lim_{T \rightarrow \infty} - \frac{1}{T} E \left(\frac{\partial^2 \log F}{\partial \theta \partial \theta'} \right),$$

the "asymptotic information" matrix of θ . 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 ϵ_t , so that $\hat{\theta}$ is asymptotically efficient in that sense. Also, $\hat{\theta}$ converges to θ in probability as $T \rightarrow \infty$.

(2) If we let σ^{ij} denote the (i,j) th element of Σ^{-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 $e = G^{-1} \left\{ y - \sum_{i=1}^p (I_m \otimes \mathcal{L}^i Y) \alpha_i \right\}$ 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 Σ ,

$$\frac{\partial \log |\Sigma^{-1}|}{\partial \sigma^{ij}} = \begin{cases} \sigma_{ii} & \text{for } i=j \\ 2\sigma_{ij} & \text{for } i \neq j. \end{cases} \quad (\text{see Golderberger (1964), Chapter 1})$$

Thus an asymptotically efficient estimator of Σ is given by

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

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

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

$$\{W'G^{-1}(\hat{\Sigma}^{-1} \otimes I_N)G^{-1}W\}^{-1} .$$

(3) The estimation procedure for $\hat{\theta}$ can easily be modified to include the case where elements of the autoregressive moving average coefficient matrices $A_1, \dots, A_p, G_1, \dots, G_q$ are specified to be zero. We simply modify the definitions of the matrix \tilde{G} and the vector \tilde{e} 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 $(I_m \otimes L^i Y)$, $i=1, \dots, p$, and $(I_m \otimes L^i \tilde{\epsilon})$, $i=1, \dots, q$, corresponding to the zero elements of $\alpha_i = \text{vec}(A_i)$ 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 ϵ_t , and hence the use of the matrix G in equation (2.8), since the modification has a negligible effect as $T \rightarrow \infty$ and does not affect the asymptotic properties of the estimator. Then using the (modified) identity (2.8) and equation (2.7), we have

$$\sqrt{T} (\hat{\theta} - \theta) = \left\{ \frac{1}{T} \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \bar{W} \right\}^{-1} \cdot \frac{1}{\sqrt{T}} \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) e$$

$$+ \left\{ \frac{1}{T} \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \bar{W} \right\}^{-1} \quad (3.2)$$

$$\cdot \sum_{i=1}^q \frac{1}{T} \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \left\{ I_m \otimes (L^i \bar{\epsilon} - L^i \epsilon) \right\} \cdot \sqrt{T} (\bar{\gamma}_i - \gamma_i).$$

Now it can be shown that

$$\text{plim}_{T \rightarrow \infty} \frac{1}{T} \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \bar{W} = \lim_{T \rightarrow \infty} \frac{1}{T} E \left\{ \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) \bar{G}^{-1} \bar{W} \right\} = V.$$

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 $\sqrt{T} (\bar{\gamma}_i - \gamma_i)$). 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 $V^{-1} \cdot \frac{1}{\sqrt{T}} \bar{W}' \bar{G}'^{-1} (\bar{\Sigma}^{-1} \otimes I_N) e$.

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

To establish the asymptotic normality of this vector, we consider in detail the asymptotic behavior of a subvector of

$\frac{1}{\sqrt{T}} W' G'^{-1} (\Sigma^{-1} \otimes I_N) e$. For example, we have

$$\begin{aligned}
 & \frac{1}{\sqrt{T}} (I_m \otimes \mathcal{L}^1 Y)' G'^{-1} (\Sigma^{-1} \otimes I_N) e \\
 &= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-p-1} \left\{ D_j \Sigma^{-1} \otimes (L^j \mathcal{L}^1 Y)' \right\} e \\
 &= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-p-1} \text{vec} \left\{ (L^j \mathcal{L}^1 Y)' \epsilon_{\Sigma^{-1} D'_j} \right\} \\
 &= \frac{1}{\sqrt{T}} \sum_{j=0}^{T-p-1} \sum_{t=p+j+1}^T \text{vec} (Y_{t-i-j} \epsilon'_t \Sigma^{-1} D'_j) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=p+1}^T \sum_{j=0}^{t-p-1} \text{vec} (Y_{t-i-j} \epsilon'_t \Sigma^{-1} D'_j) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=p+1}^T \sum_{j=0}^{t-p-1} \sum_{u=0}^{\infty} \sum_{v=0}^q \text{vec} (J'_u G'_v \epsilon_{t-i-j-u-v} \epsilon'_t \Sigma^{-1} D'_j) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=p+1}^T \left\{ \sum_{j=0}^{\min(t-p-1, n)} \sum_{u=0}^n \sum_{v=0}^q \text{vec} (J'_u G'_v \epsilon_{t-i-j-u-v} \epsilon'_t \Sigma^{-1} D'_j) \right\} \\
 & \quad + R_{Tn} \tag{3.3}
 \end{aligned}$$

$$= Z_{Tn} + R_{Tn}$$

for $n < T-p-1$,

where

$$Z_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^T W_{tn}, \quad (3.4)$$

$$W_{tn} = \sum_{j=0}^{\min(t-p-1, n)} \sum_{u=0}^n \sum_{v=0}^q \text{vec}(J'_u G'_v \epsilon_{t-i-j-u-v} \epsilon'_{t \Sigma^{-1} D'_j}), \quad (3.5)$$

($t=p+1, \dots, T$),

and

$$\begin{aligned} R_{Tn} = & \frac{1}{\sqrt{T}} \sum_{t=p+n+2}^T \sum_{j=n+1}^{t-p-1} \sum_{u=0}^{\infty} \sum_{v=0}^q \text{vec}(J'_u G'_v \epsilon_{t-i-j-u-v} \epsilon'_{t \Sigma^{-1} D'_j}) \\ & + \frac{1}{\sqrt{T}} \sum_{t=p+1}^T \sum_{j=0}^{\min(t-p-1, n)} \sum_{u=n+1}^{\infty} \sum_{v=0}^q \text{vec}(J'_u G'_v \epsilon_{t-i-j-u-v} \epsilon'_{t \Sigma^{-1} D'_j}). \end{aligned} \quad (3.6)$$

Now since the elements of the matrices J_u and D_u converge to 0 exponentially as $u \rightarrow \infty$, it is not hard to show that the remainder term R_{Tn} converges to 0 in probability as $n \rightarrow \infty$, uniformly in T (e.g., it can be shown that $E(R'_{Tn} R_{Tn}) \rightarrow 0$ as $n \rightarrow \infty$, 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

$$\begin{aligned} E(W_{tn} W'_{tn}) &= V_{tn} \\ &= \sum_{j,k=0}^{\min(t-p-1, n)} \sum_{u,m=0}^n \sum_{v,s=0}^q E \left\{ \text{vec}(J'_u G'_v \epsilon_{t-i-j-u-v} \epsilon'_{t \Sigma^{-1} D'_j}) \right. \\ & \quad \left. \cdot \text{vec}(J'_m G'_s \epsilon_{t-i-k-m-s} \epsilon'_{t \Sigma^{-1} D'_k})' \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j,k=0}^{\min(t-p-1,n)} \sum_{u,m=0}^n \sum_{v,s=0}^q E \left\{ (D_j \Sigma^{-1} \epsilon_t \otimes J'_u G'_v) \epsilon_{t-i-j-u-v} \epsilon'_{t-i-k-m-s} \right. \\
&\quad \left. \cdot (D_k \Sigma^{-1} \epsilon_t \otimes J'_m G'_s) \right\} \\
&= \sum_{j,k=0}^{\min(t-p-1,n)} \sum_{u,m=0}^n \sum_{v,s=0}^q E \left\{ (D_j \Sigma^{-1} \epsilon_t \otimes J'_u G'_v) \Sigma (D_k \Sigma^{-1} \epsilon_t \otimes J'_m G'_s) \right\} \\
&= \sum_{j,k=0}^{\min(t-p-1,n)} \sum_{u,m=0}^n \sum_{v,s=0}^q E \left\{ (D_j \Sigma^{-1} \epsilon_t \epsilon'_{t \Sigma^{-1}} D'_k \otimes J'_u G'_v \Sigma G'_s J'_m) \right\} \\
&= \sum_{j,k=0}^{\min(t-p-1,n)} \sum_{u,m=0}^n \sum_{v,s=0}^q (D_j \Sigma^{-1} D'_k \otimes J'_u G'_v \Sigma G'_s J'_m), \quad (3.7)
\end{aligned}$$

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

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

$$\begin{aligned}
&\lim_{T \rightarrow \infty} E(Z_{Tn} Z'_{Tn}) \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=p+1}^T \sum_{j,k=0}^{\min(t-p-1,n)} \sum_{u,m=0}^n \sum_{v,s=0}^q (D_j \Sigma^{-1} D'_k \otimes J'_u G'_v \Sigma G'_s J'_m) \\
&\quad (j+u+v=k+m+s) \\
&= \sum_{j,k=0}^n \sum_{u,m=0}^n \sum_{v,s=0}^q (D_j \Sigma^{-1} D'_k \otimes J'_u G'_v \Sigma G'_s J'_m) \\
&\quad (j+u+v=k+m+s) \\
&= V_n. \quad (3.8)
\end{aligned}$$

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 ϵ_t have finite fourth order moments. Thus it follows by Theorem 1 in Schönfeld (1971)

that

$Z_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^T W_{tn}$ has a limiting multivariate normal distribution

as $T \rightarrow \infty$, 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 Schönfeld (1971), we can conclude that

$$\frac{1}{\sqrt{T}} (I_m \otimes \mathcal{L}^i Y)' G'^{-1} (\Sigma^{-1} \otimes I_N) e = Z_{Tn} + R_{Tn}$$

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

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

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

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

$$= \Psi_0 .$$

(3.9)

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

$$\frac{1}{\sqrt{T}} W' G'^{-1} (\Sigma^{-1} \otimes I_N) e$$

have a joint limiting distribution as $T \rightarrow \infty$ 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

$$Y_t - \sum_{i=1}^p A'_i Y_{t-i} = \sum_{i=0}^k B'_i X_{t-i} + \epsilon_t + \sum_{i=1}^q G'_i \epsilon_{t-i},$$

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,

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

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

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60, 255-265.
- Anderson, T. W. (1971). The Statistical Analysis of Time Series. New York: Wiley.
- Box, G.E.P. and Jenkins, G. M. (1970). Time Series Analysis Forecasting and Control. San Francisco: Holden-Day, Inc.
- Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. To appear in Advances in Applied Probability.
- Goldberger, A. S. (1964). Econometric Theory. New York: Wiley.
- Hannan, E. J. (1969a). The identification of vector mixed autoregressive-moving average systems. Biometrika 56, 223-225.
- Hannan, E. J. (1969b). The estimation of mixed moving average autoregressive systems. Biometrika 56, 579-593.
- Hannan, E. J. (1970). Multiple Time Series. New York: Wiley.
- Hannan, E. J. (1975). The estimation of ARMA models. Ann. Statist. 3, 975-981.
- Kashyap, R. L. (1970). Maximum likelihood identification of stochastic linear systems. I.E.E.E. Trans. Auto. Control AC-15, 25-34.
- Neudecker, H. (1969). Some theorems of matrix differentiation with special reference to Kronecker matrix products. J. Amer. Statist. Assoc. 64, 953-963.

- Nicholls, D. F. (1976a). The efficient estimation of vector linear time series models. To appear in Biometrika.
- Nicholls, D. F. (1976b). A comparison of estimation methods for vector linear time series models.
- Reinsel, G. (1976a). Maximum likelihood estimation of stochastic linear difference equations with autoregressive moving average errors. Technical Report No. 112, Department of Statistics, Carnegie-Mellon University.
- Reinsel, G. (1976b). Estimation of linear time series models. Unpublished Ph.D. dissertation, Department of Mathematics and Statistics, University of Pittsburgh.
- Robinson, E. A. (1967). Multichannel Time Series Analysis with Digital Computer Programs. San Francisco: Holden-Day.
- Schonfeld, P. (1971). A useful central limit theorem for m -dependent variables. Metrika 17, 116-128.
- Tunnicliffe Wilson, G. (1972). The factorization of matricial spectral densities. SIAM J. Appl. Math. 23, 420-426.
- Tunnicliffe Wilson, G. (1973). The estimation of parameters in multivariate time series models. J. R. Statist. Soc. B 35, 76-85.
- Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika 50, 129-134.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ONR Report No. 3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MAXIMUM LIKELIHOOD ESTIMATION OF VECTOR AUTOREGRESSIVE MOVING AVERAGE MODELS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
7. AUTHOR(s) 10 Greg Reinsel		6. PERFORMING ORG. REPORT NUMBER Technocal Report No. 117
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Carnegie-Mellon University Pittsburgh, Pa. 15213		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-76-C-0930
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 NR-042-367
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 14 TR-117, 3-ONR		12. REPORT DATE 11 July 1976
		13. NUMBER OF PAGES 21 12 25p
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Vector autoregressive moving average process, Newton-Raphson method, Maximum likelihood estimation, Generalized least squares.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) 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 (please turn page)		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

391190

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Block 20.

to be asymptotically efficient and to possess a limiting multi-variate normal distribution.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)