MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS OF AN AUTOREGRESSIVE PROCESS WITH MOVING AVERAGE RESIDUALS AND OTHER COVARIANCE MATRICES WITH LINEAR STRUCTURE

BY

T. W. ANDERSON

TECHNICAL REPORT NO. 15

DECEMBER 1973

PREPARED UNDER CONTRACT

N00014-67-A-0112-0030 (NR-042-034)

FOR THE OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS OF AN AUTOREGRESSIVE PROCESS WITH MOVING AVERAGE RESIDUALS AND OTHER COVARIANCE MATRICES WITH LINEAR STRUCTURE

by

T. W. ANDERSON

STANFORD UNIVERSITY

TECHNICAL REPORT NO. 15

DECEMBER, 1973

PREPARED UNDER THE AUSPICES OF

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030

T. W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

1. Introduction

A stationary stochastic process that serves as a useful model for time series analysis is the autoregressive process with moving average residuals $\{y_t\}$ which satisfies

$$\sum_{s=0}^{p} \beta_s y_{t-s} = \sum_{j=0}^{q} \alpha_j v_{t-j} , \tag{1.1}$$

$t = \ldots , -1, 0, 1, \ldots$, where the sequence $\{v_t\}$ consists of independently identically distributed random variables. [See Section 5.8 of T. W. Anderson (1971a) and Box and Jenkins (1970).] To avoid indeterminacy $\beta_0 = \alpha_0 = 1$. (An alternative of specifying the variance of v_t to be 1 and leaving α_0 as a free parameter is considered also.) The mean of v_t is independent of t and is taken to be 0 for convenience. (Modifications necessary to account for an arbitrary mean are also discussed.) When $\mathcal{E} y_t = 0$, the stationarity implies

$$\mathcal{E} y_t y_s = \sigma(t-s) , \tag{1.2}$$

dependent only on the difference of the indices.

We shall assume that the v_t's are normally distributed, that is, that the process is Gaussian. Then the model is completely specified by the coefficients in (1.1) and the variance of v_t , say σ^2 .

The statistical problem treated here is to estimate $\beta_1, \ldots , \beta_p$, $\alpha_1, \ldots , \alpha_q$, and σ^2 on the basis of a set of observations at T successive time points, $y_1, \ldots , y_T$.

If $\underset{\sim}{y} = (y_1, \ldots, y_T)'$, the density of the multivariate normal distribution $N(0, \underset{\sim}{\Sigma})$ of $\underset{\sim}{y}$ is

$$\frac{1}{(2\pi)^{\frac{1}{2}T}|\underset{\sim}{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y}} , \tag{1.3}$$

where

$$\mathcal{E} y_t\, y_s = \sigma_{ts} , \qquad t, s = 1, \ldots , T , \tag{1.4}$$

is the t,s-th element of $\underset{\sim}{\Sigma}$. If the distribution is that defined by (1.1), then (1.4) is (1.2); the covariances are functions of the parameters $\beta_1, \ldots , \beta_p$, $\alpha_1, \ldots , \alpha_q$, and σ^2.

The method of maximum likelihood can be considered, but in general an explicit solution cannot be found. The approach of this paper is to modify the model slightly so that the derivatives of the likelihood function set equal to 0 yield relatively simple equations. Since these equations are nonlinear, an iterative procedure is proposed that yields asymptotically efficient estimates at the first step (as $T \to \infty$).

The estimation problems for the pure autoregressive process and pure moving average process as well as the general mixed model are set up in terms of more general multivariate models. The case of N observations on the vector $\underset{\sim}{y}$ is included. This work is a continuation of earlier research on covariance matrices with linear structure by T. W. Anderson (1969), (1970), (1971b), and (1973). The iterative procedures are extensions of that presented in the last paper, which is essentially the method of scoring (as pointed out to me by J. N. K. Rao).

Durbin (1959), (1960) and A. M. Walker (1961), (1962) have proposed estimates, but they are not asymptotically efficient (as $T \to \infty$). Box and Jenkins (1970) have suggested maximizing the likelihood function by numerical means.

The covariance sequence (1.2) of a stationary process has a spectral representation. In the case of an absolutely continuous spectral distribution function

$$\sigma(h) = \int_{-\pi}^{\pi} f(\lambda) \cos \lambda h \, d\lambda \;, \qquad h = 0, \pm 1, \ldots . \tag{1.5}$$

The spectral density $f(\lambda)$ may be determined by

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sigma(h) \cos \lambda h \tag{1.6}$$

when the series on the right-hand side converges. In the case of model (1.1) the spectral density is

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{\left| \sum_{j=1}^{q} \alpha_j \, e^{i\lambda j} \right|^2}{\left| \sum_{r=0}^{p} \beta_r \, e^{i\lambda r} \right|^2} \; . \tag{1.7}$$

Clevenson (1970) and Parzen (1971) and Hannan (1969) have proposed estimation methods based on the sample spectral density (the so-called periodogram). The relationship between these methods and the ones presented in this paper will be explicated in a later paper.

If we let (1.1) be u_t, the spectral density of the stationary process $\{u_t\}$ is

$$(1.8)\qquad f_u(\lambda) = \frac{\sigma^2}{2\pi} \sum_{j=0}^{q} \alpha_j e^{i\lambda j} \sum_{j=0}^{q} \alpha_j e^{-i\lambda j}$$

$$= \frac{1}{2\pi} \sum_{h=-q}^{q} \sigma_u(h) e^{i\lambda h} ,$$

where

$$(1.9)\qquad \sigma_u(h) = \sigma^2 \sum_{k=0}^{q-|h|} \alpha_k \alpha_{k+|h|} , \; h = 0, \pm 1, \ldots , \pm q ,$$

are the nonzero covariances of $\{u_t\}$. The parameters $\alpha_1, \ldots , \alpha_q$, and σ^2 can be replaced by $\sigma_u(0), \sigma_u(1), \ldots , \sigma_u(q)$. We shall assume the roots of

$$(1.10)\qquad M(z) = \sum_{j=0}^{q} \alpha_j z^{q-j}$$

are less than 1 in absolute value. Then given $\sigma_u(0), \sigma_u(1), \ldots , \sigma_u(q) \neq 0$ $\sum_{h=-q}^{q} \sigma_u(h) z^h$ can be factored uniquely into $M(z)M(z^{-1})$, thus, defining $\alpha_1, \ldots , \alpha_q$, and σ^2.[See T. W. Anderson (1971a) and (1971b) for details.] Estimation of the pure moving average model in terms of $\sigma(0), \sigma(1), \ldots , \sigma(q)$ was treated by T. W. Anderson (1971b), (1973).

2. Estimation of Coefficients of Linear Transformations to Approximate Autoregressive Processes

2.1 A General Linear Transformation. Suppose $\underset{\sim}{y}$ is a T-component random vector defined by

$$\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{y} = \underset{\sim}{v} , \tag{2.1}$$

where $\underset{\sim}{K}_0, \underset{\sim}{K}_1, \ldots , \underset{\sim}{K}_p$ are $p + 1$ known linearly independent $T \times T$ matrices, $\beta_0 = 1$ and $\beta_1, \ldots , \beta_p$ are p parameters such that $\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell$ is nonsingular; we assume that there is at least one such set. Suppose $\underset{\sim}{v}$ is a T-component random variable with mean vector $\mathcal{E} v = 0$ and covariance matrix

$$\mathcal{C}(\underset{\sim}{v}) = \mathcal{E} \underset{\sim}{v}\underset{\sim}{v}' = \sigma^2 \underset{\sim}{I} . \tag{2.2}$$

Then

$$\underset{\sim}{y} = \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right)^{-1} \underset{\sim}{v} \tag{2.3}$$

has mean vector $\mathcal{E}\underset{\sim}{y} = 0$ and covariance matrix

$$\mathcal{C}(\underset{\sim}{y}) = \mathcal{E}\underset{\sim}{y}\underset{\sim}{y}' = \sigma^2 \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right)^{-1} \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \right)^{-1} = \sigma^2 \left(\sum_{k,\ell=0}^{p} \beta_k \beta_\ell \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \right)^{-1} \tag{2.4}$$

with inverse

$$\mathcal{C}^{-1}(\underset{\sim}{y}) = \frac{1}{\sigma^2} \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell = \frac{1}{\sigma^2} \sum_{k,\ell=0}^{p} \beta_k \beta_\ell \underset{\sim}{K}_k' \underset{\sim}{K}_\ell . \tag{2.5}$$

Let $\underset{\sim}{y}_1, \ldots , \underset{\sim}{y}_N$ be N observations on $\underset{\sim}{y}$, and let L denote the likelihood function when $\underset{\sim}{y}$ has a normal distribution. Then

$$(2.6)\qquad \frac{2}{N}\log L = -T\log 2\pi - T\log\sigma^2 + 2\log\left|\sum_{\ell=0}^{p}\beta_\ell \underset{\sim}{K}_\ell\right|$$
$$-\frac{1}{N\sigma^2}\sum_{\alpha=1}^{N}\left(\sum_{k=0}^{p}\beta_k \underset{\sim}{K}_k \underset{\sim}{y}_\alpha\right)'\left(\sum_{\ell=0}^{p}\beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{y}_\alpha\right)$$
$$= -T\log 2\pi - T\log\sigma^2 + 2\log\left|\sum_{\ell=0}^{p}\beta_\ell \underset{\sim}{K}_\ell\right|$$
$$-\frac{1}{\sigma^2}\operatorname{tr}\sum_{k,\ell=0}^{p}\beta_k\beta_\ell \underset{\sim}{K}'_k \underset{\sim}{K}_\ell \underset{\sim}{C},$$

where

$$(2.7)\qquad \underset{\sim}{C} = \frac{1}{N}\sum_{\alpha=1}^{N}\underset{\sim}{y}_\alpha \underset{\sim}{y}'_\alpha,$$

and tr denotes the trace of the matrix that follows. To find the partial derivatives of (2.6) with respect to $\beta_1, \ldots, \beta_p$ we use the results

$$(2.8)\qquad \frac{\partial\log|\underset{\sim}{A}|}{\partial\theta} = \frac{1}{|\underset{\sim}{A}|}\frac{\partial|\underset{\sim}{A}|}{\partial\theta}$$
$$= \frac{1}{|\underset{\sim}{A}|}\sum_{i,j=1}^{p}\frac{\partial|\underset{\sim}{A}|}{\partial a_{ij}}\frac{\partial a_{ij}}{\partial\theta}$$
$$= \frac{1}{|\underset{\sim}{A}|}\sum_{i,j=1}^{p}\operatorname{cof} a_{ij}\frac{\partial a_{ij}}{\partial\theta}$$
$$= \sum_{i,j=1}^{p}a^{ji}\frac{\partial a_{ij}}{\partial\theta}$$
$$= \operatorname{tr}\underset{\sim}{A}^{-1}\frac{\partial}{\partial\theta}\underset{\sim}{A}.$$

(The cofactor of a_{ij} in $\underset{\sim}{A}$ is denoted by $\operatorname{cof} a_{ij}$.) Then

$$(2.9)\qquad \frac{\partial}{\partial\beta_\ell}\frac{2}{N}\log L = 2\operatorname{tr}\left(\sum_{k=0}^{p}\beta_k \underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_\ell$$
$$-\frac{2}{N\sigma^2}\sum_{\alpha=1}^{N}\underset{\sim}{y}'_\alpha\sum_{k=0}^{p}\beta_k \underset{\sim}{K}'_k \underset{\sim}{K}_\ell \underset{\sim}{y}_\alpha$$

$$= 2 \operatorname{tr} \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k \right)^{-1} \underset{\sim}{K}_\ell - \frac{2}{\sigma^2} \operatorname{tr} \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{C} \,, \qquad \ell = 1, \ldots, p \,.$$

(2.10) $$\frac{\partial}{\partial \sigma^2} \frac{2}{N} \log L = - \frac{T}{\sigma^2} + \frac{1}{\sigma^4} \operatorname{tr} \sum_{k,\ell=0}^{p} \beta_k \beta_\ell \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{C} \,.$$

If $N = 1$ and $\underset{\sim}{y}_1 = \underset{\sim}{y}$, the derivatives (2.9) are

(2.11) $$2 \operatorname{tr} \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k \right)^{-1} \underset{\sim}{K}_\ell - \frac{2}{\sigma^2} \sum_{k=0}^{p} \beta_k \underset{\sim}{y}' \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{y} \,, \quad \ell = 1, \ldots, p \,,$$

and (2.10) is

(2.12) $$- \frac{T}{\sigma^2} + \frac{1}{\sigma^4} \sum_{k,\ell=0}^{p} \beta_k \beta_\ell \underset{\sim}{y}' \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{y} \,.$$

The maximum likelihood estimates may be defined by setting the derivatives equal to 0. [By the argument used in T. W. Anderson (1970) it follows that there is at least one relative maximum defined by the derivative equations.] The derivative equations are

(2.13) $$\operatorname{tr} \left(\sum_{k=0}^{p} \hat{\beta}_k \underset{\sim}{K}_k \right)^{-1} \underset{\sim}{K}_\ell = \frac{1}{\hat{\sigma}^2} \sum_{k=0}^{p} \hat{\beta}_k \operatorname{tr} \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{C} \,,$$

(2.14) $$\hat{\sigma}^2 = \frac{1}{T} \sum_{k,\ell=0}^{p} \hat{\beta}_k \hat{\beta}_\ell \operatorname{tr} \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{C} \,.$$

We can develop these equations in an alternative way by letting

(2.15) $$\underset{\sim}{K}_k \underset{\sim}{y}_\alpha = \underset{\sim}{y}_\alpha^{(k)} \,, \qquad k = 0, 1, \ldots, p, \ \alpha = 1, \ldots, N \,.$$

Then

(2.16) $$\begin{aligned} \frac{2}{N} \log L &= - T \log 2\pi - T \log \sigma^2 + 2 \log \left| \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right| \\ &\quad - \frac{1}{N\sigma^2} \sum_{\alpha=1}^{N} \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{y}_\alpha^{(k)} \right)' \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{y}_\alpha^{(\ell)} \right) \\ &= - T \log 2\pi - T \log \sigma^2 + 2 \log \left| \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right| - \frac{1}{\sigma^2} \underset{\sim}{\beta}' \underset{\sim}{M} \underset{\sim}{\beta} \,, \end{aligned}$$

where

$$(2.17) \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} ,$$

$$(2.18) \quad \underset{\sim}{M} = \frac{1}{N} \sum_{\alpha=1}^{N} \begin{pmatrix} \underset{\sim}{y}_\alpha^{(0)'} \underset{\sim}{y}_\alpha^{(0)} & \underset{\sim}{y}_\alpha^{(0)'} \underset{\sim}{y}_\alpha^{(1)} & \cdots & \underset{\sim}{y}_\alpha^{(0)'} \underset{\sim}{y}_\alpha^{(p)} \\ \underset{\sim}{y}_\alpha^{(1)'} \underset{\sim}{y}_\alpha^{(0)} & \underset{\sim}{y}_\alpha^{(1)'} \underset{\sim}{y}_\alpha^{(1)} & \cdots & \underset{\sim}{y}_\alpha^{(1)'} \underset{\sim}{y}_\alpha^{(p)} \\ \vdots & & & \\ \underset{\sim}{y}_\alpha^{(p)'} \underset{\sim}{y}_\alpha^{(0)} & \underset{\sim}{y}_\alpha^{(p)'} \underset{\sim}{y}_\alpha^{(1)} & \cdots & \underset{\sim}{y}_\alpha^{(p)'} \underset{\sim}{y}_\alpha^{(p)} \end{pmatrix} .$$

The partial derivatives of $(2/N) \log L$ set equal to 0 can be written in terms of the elements of $\underset{\sim}{M}$ as

$$(2.19) \qquad \left[\operatorname{tr} \left(\sum_{k=0}^{p} \hat{\beta}_k \underset{\sim}{K}_k \right)^{-1} \underset{\sim}{K}_\ell \right] = \frac{1}{\sigma^2} \hat{\underset{\sim}{\beta}}' \underset{\sim}{M} ,$$

$$(2.20) \qquad \hat{\sigma}^2 = \frac{1}{T} \operatorname{tr} \hat{\underset{\sim}{\beta}}' \underset{\sim}{M} \hat{\underset{\sim}{\beta}} ;$$

the left-hand side of (2.19) denotes a row vector with the ℓ-th component given explicitly.

If $N > 1$ and $\mathcal{E}\underset{\sim}{y} = \underset{\sim}{\mu}$, where $\underset{\sim}{\mu}$ is an arbitrary vector, then the sample mean

$$(2.21) \qquad \bar{\underset{\sim}{x}} = \frac{1}{N} \sum_{y=1}^{N} \underset{\sim}{y}_\alpha$$

is the maximum likelihood estimate of $\underset{\sim}{\mu}$, and in the likelihood equations (2.13) and (2.14), $\underset{\sim}{C}$ should be replaced by

$$(2.22) \qquad \hat{\underset{\sim}{C}} = \frac{1}{N} \sum_{\alpha=1}^{N} (\underset{\sim}{y}_\alpha - \bar{\underset{\sim}{y}})(\underset{\sim}{y}_\alpha - \bar{\underset{\sim}{y}})' .$$

In some models one wants $\mathcal{E}y_j = \mu$; that is, $\mathcal{E}\underset{\sim}{y} = \mu\underset{\sim}{\varepsilon}$, where

$\underset{\sim}{\varepsilon} = (1, 1, \ldots, 1)'$. Then $2/N$ times the logarithm of the likelihood function is (2.6) with $\underset{\sim}{C}$ replaced by

$$\underset{\sim}{C}^* = \frac{1}{N} \sum_{\alpha=1}^{N} (\underset{\sim}{y}_\alpha - \mu \underset{\sim}{\varepsilon}) (\underset{\sim}{y}_\alpha - \mu \underset{\sim}{\varepsilon})' . \tag{2.23}$$

The derivative of $2/N$ times the logarithm of the likelihood with respect to μ is

$$\frac{\partial}{\partial\mu} \frac{2}{N} \log L = \frac{2}{N\sigma^2} \underset{\sim}{\varepsilon}' \sum_{k,\ell=0}^{N} \beta_k \beta_\ell \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \sum_{\alpha=1}^{N} (\underset{\sim}{y}_\alpha - \mu \underset{\sim}{\varepsilon}) . \tag{2.24}$$

If $\underset{\sim}{\varepsilon}$ is a characteristic vector of $\underset{\sim}{K}_0, \underset{\sim}{K}_1, \ldots, \underset{\sim}{K}_p$, then

$$\hat{\mu} = \frac{1}{NT} \sum_{\alpha=1}^{N} \underset{\sim}{\varepsilon}' \underset{\sim}{y}_\alpha ; \tag{2.25}$$

and in the other derivative equations $\underset{\sim}{C}$ is replaced by

$$\frac{1}{N} \sum_{\alpha=1}^{N} (\underset{\sim}{y}_\alpha - \hat{\mu} \underset{\sim}{\varepsilon}) (\underset{\sim}{y}_\alpha - \hat{\mu} \underset{\sim}{\varepsilon})' . \tag{2.26}$$

If $\underset{\sim}{\varepsilon}$ is not a characteristic vector of $\underset{\sim}{K}_0, \underset{\sim}{K}_1, \ldots, \underset{\sim}{K}_p$, then usually (2.25) will not be the maximum likelihood estimate of μ.

The second derivatives of $(2/N) \log L$ defined by (2.6) are

$$\frac{\partial^2}{\partial\beta_j \, \partial\beta_\ell} \frac{2}{N} \log L = -2 \operatorname{tr} \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k \right)^{-1} \underset{\sim}{K}_j \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k \right)^{-1} \underset{\sim}{K}_\ell - \frac{2}{\sigma^2} \operatorname{tr} \underset{\sim}{K}_j' \underset{\sim}{K}_\ell \underset{\sim}{C} , \qquad j, \ell = 1, \ldots, p , \tag{2.27}$$

$$\frac{\partial^2}{\partial\beta_j \, \partial\sigma^2} \frac{2}{N} \log L = \frac{2}{\sigma^4} \operatorname{tr} \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_j' \underset{\sim}{K}_\ell \underset{\sim}{C} , \qquad j = 1, \ldots, p , \tag{2.28}$$

$$\frac{\partial^2}{(\partial\sigma^2)^2} \frac{2}{N} \log L = \frac{T}{\sigma^4} - \frac{2}{\sigma^6} \operatorname{tr} \sum_{k,\ell=0}^{p} \beta_k \beta_\ell \underset{\sim}{K}_k' \underset{\sim}{K}_\ell \underset{\sim}{C} . \tag{2.29}$$

The elements of the information matrix are N times

$$(2.30)\quad -\mathscr{E}\frac{\partial^2}{\partial\beta_j\,\partial\beta_\ell}\,\frac{1}{N}\log L = \mathrm{tr}\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_j\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_\ell$$

$$+\ \mathrm{tr}\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k'\right)^{-1}\underset{\sim}{K}_j'\,\underset{\sim}{K}_\ell\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1},$$

$$j,\ell = 1, \ldots, p,$$

$$(2.31)\quad -\mathscr{E}\frac{\partial^2}{\partial\beta_j\,\partial\sigma^2}\,\frac{1}{N}\log L = -\frac{1}{\sigma^2}\,\mathrm{tr}\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_j\,, \qquad j = 1, \ldots, p\,,$$

$$(2.32)\quad -\mathscr{E}\frac{\partial^2}{(\partial\sigma^2)}\,\frac{1}{N}\log L = \frac{T}{2\sigma^4}\,.$$

As $N \to \infty$, the normalized maximum likelihood estimates have a limiting normal distribution with covariance matrix whose inverse has elements given by (2.30), (2.31), and (2.32).

2.2 Autoregressive Process Approximated by a Linear Transformation.

The autoregressive process $\{y_t\}$ is (1.1) for $\alpha_1 = \ldots = \alpha_q = 0$, that is,

$$(2.33)\qquad \sum_{s=0}^{p}\beta_s\,y_{t-s} = v_t\,,$$

$t = \ldots, -1, 0, 1, \ldots$. Let $\underset{\sim}{y} = (\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_T)'$. Then the distribution of $y_1, \ldots, y_T$ is approximated by the distribution of $\underset{\sim}{y}$ defined by (2.1) when $\underset{\sim}{K}_g = \underset{\sim}{L}^g$, $g=0, 1, \ldots, p$, where

$$(2.34)\quad \underset{\sim}{L} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then

$$(2.35)\qquad \underset{\sim}{L}^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \end{pmatrix}.$$

In general $\underset{\sim}{L}^g$ has all 0's except for 1's g units below the main diagonal. We suppose $p + 1 \leq T$. Note that

$$(2.36)\qquad \underset{\sim}{L}^g \underset{\sim}{L}^h = \underset{\sim}{L}^{g+h}, \qquad g, h=0, 1, \ldots ,$$

$$(2.37)\qquad \underset{\sim}{L}^g = \underset{\sim}{0}, \qquad g = T, T+1, \ldots .$$

In this case

$$(2.38)\qquad \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \beta_1 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \beta_2 & \beta_1 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \beta_p & \beta_{p-1} & \beta_{p-2} & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \beta_p & \beta_{p-1} & \cdots & \beta_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix},$$

which is triangular with 0's above the main diagonal and has determinant 1.

The components of (2.1) are

(2.40) $$\sum_{s=0}^{t-1} \beta_s y_{t-s} = v_t , \qquad t=1, \ldots , p ,$$

(2.41) $$\sum_{s=0}^{p} \beta_s y_{t-s} = v_t, \qquad t=p+1, \ldots , T .$$

The equation (2.41) agrees with the autoregressive process (2.33), but the equation (2.40) is such that the sequence $y_1, \ldots , y_T$ does not start out as a stationary process. An alternative way of considering the equation (2.40) is that (2.41) holds with $y_0 = y_{-1} = \ldots = y_{-(p-1)} = 0$.

In this model we are often interested in $N = 1$ and $\underset{\sim}{y}_1 = \underset{\sim}{y}$. Then

(2.42) $$\underset{\sim}{y}^{(k)} = \underset{\sim}{K}_k \underset{\sim}{y} = \underset{\sim}{L}^k \underset{\sim}{y} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ y_1 \\ \vdots \\ y_{T-k} \end{pmatrix}, \qquad k=0, 1, \ldots , T-1 ,$$

where there are k 0's, and

(2.43) $$\underset{\sim}{L}^k \underset{\sim}{y} = \underset{\sim}{0} , \ k=T, T+1, \ldots .$$

Since $\sum_{k=0}^{p} \beta_k \underset{\sim}{L}_k$ is triangular with 0's above the main diagonal, then $(\sum_{k=0}^{p} \beta_k \underset{\sim}{L}^k)^{-1}$ is triangular with 0's above the main diagonal, and the determinant of $\sum_{k=0}^{p} \beta_k \underset{\sim}{L}^k$ is 1. [The diagonal terms of $(\sum_{k=0}^{p} \beta_k \underset{\sim}{L}^k)^{-1} \underset{\sim}{L}^\ell$ are 0, $\ell=1, \ldots , p$.] Then the derivative of $2/N$ times the logarithm of the determinant with respect to β_ℓ is

(2.44) $$\frac{\partial}{\partial \beta_\ell} \log \left| \sum_{k=0}^{p} \beta_k \underset{\sim}{L}^k \right| = \operatorname{tr} \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{L}^k \right)^{-1} \underset{\sim}{L}^\ell = 0 \qquad \ell = 1, \ldots , p .$$

The derivative equations (2.13) can be written in this case as

$$\sum_{k=1}^{p} \hat{\beta}_k \, \underset{\sim}{y}^{(k)'} \underset{\sim}{y}^{(\ell)} = - \underset{\sim}{y}^{(0)'} \underset{\sim}{y}^{(\ell)} , \qquad \ell=1, \ldots , p . \tag{2.45}$$

In components these are

$$\sum_{k=1}^{p} \hat{\beta}_k \sum_{t=1}^{T} y_{t-k} \, y_{t-\ell} = - \sum_{t=1}^{T} y_t \, y_{t-\ell} , \qquad \ell=1, \ldots , p , \tag{2.46}$$

where $y_0 = y_{-1} = \ldots = y_{-(p-1)} = 0$. These are the usual maximum likelihood estimates of $\beta_1, \ldots , \beta_p$ for initial values $y_0 = y_{-1} = \ldots = y_{-(p-1)} = 0$ or the "least squares estimates" since they minimize

$$\sum_{t=1}^{T} \left(\sum_{k=0}^{p} \beta_k \, y_{t-k} \right)^2 . \tag{2.47}$$

[See T. W. Anderson (1971), Sections 2.2 and 5.4, for example.]

Let

$$c_h = \frac{1}{T} \sum_{i=1}^{T-h} y_i \, y_{i+h} , \qquad h=0, 1, \ldots , T-1 , \tag{2.48}$$

The right-hand side of (2.46) is $-Tc_\ell$. The sum

$$\sum_{t=1}^{T} y_{t-k} \, y_{t-\ell} \tag{2.49}$$

differs from $Tc_{|k-\ell|}$ by omission of

$$\sum_{t=T-\max(k,\ell)+1}^{T-|k-\ell|} y_t \, y_{t+|k-\ell|} . \tag{2.50}$$

These terms can be added to the coefficients so as to make the equations agree with

$$\sum_{g=1}^{p} \hat{\beta}_g \, c_{g-f} = - c_f , \quad f=1, \ldots , p . \tag{2.51}$$

[See T. W. Anderson (1971a), Sec. 5.6, for example.] Then the estimates derived from (2.51) are the coefficients of a stationary process. [See Anderson (1971c), for example.] If we let

$$(2.52) \qquad \tilde{\underset{\sim}{y}}^{(k)} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ y_1 \\ \vdots \\ y_T \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad k=0, 1, \ldots, p,$$

where the first k components are 0 and the last $p-k$ components are 0, then (2.51) can be written

$$(2.53) \qquad \sum_{k=1}^{p} \hat{\beta}_k \tilde{\underset{\sim}{y}}^{(k)'} \tilde{\underset{\sim}{y}}^{(\ell)} = -\tilde{\underset{\sim}{y}}^{(0)'} \tilde{\underset{\sim}{y}}^{(\ell)}, \quad \ell=1, \ldots, p.$$

In this case of $\underset{\sim}{K}_k = \underset{\sim}{L}^k$ the elements of the information matrix are N times

$$(2.54) \qquad -\mathscr{E} \frac{\partial^2}{\partial\beta_j \, \partial\beta_\ell} \frac{1}{N} \log L = \operatorname{tr}\left(\sum_{k=0}^{p} \beta_k \underset{\sim}{L}'^k\right)^{-1} \underset{\sim}{L}'^j \underset{\sim}{L}^\ell \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{L}^k\right)^{-1}, \quad j, \ell=1, \ldots, p,$$

$$(2.55) \qquad -\mathscr{E} \frac{\partial^2}{\partial\beta_j \, \partial\sigma^2} \frac{1}{N} \log L = 0, \qquad j = 1, \ldots, p,$$

and (2.32).

It is of interest to compare the covariance matrix of $\underset{\sim}{y}$ defined by (2.1) with that of T terms from the stationary process defined by (2.33). For $p=1$ and $\beta_1=\beta$ the covariance matrix of the stationary process is

$$(2.56) \qquad \frac{\sigma^2}{1-\beta^2}\begin{pmatrix} 1 & -\beta & & \cdots & (-\beta)^{T-1} \\ -\beta & 1 & -\beta & \cdots & (-\beta)^{T-2} \\ \beta^2 & -\beta & 1 & \cdots & (-\beta)^{T-3} \\ \vdots & \vdots & \vdots & & \vdots \\ (-\beta)^{T-1} & (-\beta)^{T-2} & (-\beta)^{T-3} & \cdots & 1 \end{pmatrix}$$

and

$$(2.57) \qquad \mathcal{C}(\underset{\sim}{y}) = \frac{\sigma^2}{1-\beta^2}\begin{pmatrix} 1-\beta^2 & -\beta(1-\beta^2) & \beta^2(1-\beta^2) & \cdots & (-\beta)^{T-1}(1-\beta^2) \\ -\beta(1-\beta^2) & 1-\beta^4 & -\beta(1-\beta^4) & \cdots & (-\beta)^{T-2}(1-\beta^4) \\ \beta^2(1-\beta^2) & -\beta(1-\beta^4) & 1-\beta^6 & \cdots & (-\beta)^{T-3}(1-\beta^6) \\ \vdots & \vdots & \vdots & & \vdots \\ (-\beta)^{T-1}(1-\beta^2) & (-\beta)^{T-2}(1-\beta^4) & (-\beta)^{T-3}(1-\beta^6) & \cdots & 1-\beta^{2T} \end{pmatrix}.$$

For a stationary process $|\beta| < 1$, and hence the i,j-th element of $\mathcal{C}(\underset{\sim}{y})$ is close to the i,j-th element of (2.56) if i and j are large.

3. Estimation of Coefficients of Linear Transformations to Approximate Moving Average Processes

3.1 A General Linear Transformation. Another model is defined by

$$\underset{\sim}{y} = \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \underset{\sim}{v} , \tag{3.1}$$

where $\underset{\sim}{J}_0, \underset{\sim}{J}_1, \ldots, \underset{\sim}{J}_q$ are $q+1$ known linearly independent $T \times T$ matrices, $\alpha_0 = 1$, and $\alpha_1, \ldots, \alpha_q$ are q parameters such that $\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{J}_\ell$ is nonsingular; we assume that there is at least one such set. Suppose $\underset{\sim}{v}$ is a random vector with mean vector $\mathcal{E}\underset{\sim}{v} = \underset{\sim}{0}$ and covariance matrix $\mathcal{E}\underset{\sim}{v}\underset{\sim}{v}' = \sigma^2 \underset{\sim}{I}$. Then the mean vector of $\underset{\sim}{y}$ is $\mathcal{E}\underset{\sim}{y} = \underset{\sim}{0}$ and the covariance matrix is

$$\mathcal{C}(\underset{\sim}{y}) = \mathcal{E}\underset{\sim}{y}\underset{\sim}{y}' = \sigma^2 \sum_{k,\ell=0}^{q} \alpha_k \alpha_\ell \underset{\sim}{J}_k \underset{\sim}{J}_\ell' = \sigma^2 \left(\sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \right) \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{J}_\ell' \right). \tag{3.2}$$

If L denotes the likelihood function, then

$$\begin{aligned} \frac{2}{N} \log L &= -T \log 2\pi - T \log \sigma^2 - 2 \log \left| \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \right| \\ &\quad - \frac{1}{N\sigma^2} \sum_{\alpha=1}^{N} \underset{\sim}{y}_\alpha' \left(\sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k' \right)^{-1} \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{J}_\ell \right)^{-1} \underset{\sim}{y}_\alpha \\ &= -T \log 2\pi - T \log \sigma^2 - 2 \log \left| \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \right| \\ &\quad - \frac{1}{\sigma^2} \operatorname{tr} \left(\sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k' \right)^{-1} \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{J}_\ell \right)^{-1} \underset{\sim}{C} . \end{aligned} \tag{3.3}$$

We use the result that

$$\frac{\partial}{\partial\theta} \underset{\sim}{A}^{-1} = - \underset{\sim}{A}^{-1} \left(\frac{\partial}{\partial\theta} \underset{\sim}{A} \right) \underset{\sim}{A}^{-1} , \tag{3.4}$$

which follows from differentiating $\underset{\sim}{A}\underset{\sim}{A}^{-1} = \underset{\sim}{I}$. The partial derivatives of $(2/N) \log L$ are

$$(3.5)\quad \frac{\partial}{\partial\alpha_j}\ \frac{2}{N}\log L = -2\ \mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{J}_j$$

$$+\frac{2}{\sigma^2}\ \mathrm{tr}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1},$$

$$j=1,\ \ldots,\ q\ ,$$

$$(3.6)\quad \frac{\partial}{\partial\sigma^2}\ \frac{2}{N}\log L = -\frac{T}{\sigma^2}+\frac{1}{\sigma^4}\ \mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}'_k\right)^{-1}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\ .$$

The likelihood equations can be written [with the second term on the right-hand side of (3.5) transposed]

$$(3.7)\quad \mathrm{tr}\left(\sum_{k=0}^{q}\hat{\alpha}_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{J}_j = \frac{1}{\hat{\sigma}^2}\ \mathrm{tr}\left(\sum_{\ell=0}^{q}\hat{\alpha}_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{J}_j\left(\sum_{\ell=0}^{q}\hat{\alpha}_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\left(\sum_{\ell=0}^{q}\hat{\alpha}_\ell J'_\ell\right)^{-1},$$

$$j=1,\ \ldots\ ,\ q\ ,$$

$$(3.8)\quad \hat{\sigma}^2 = \frac{1}{T}\ \mathrm{tr}\left(\sum_{k=0}^{q}\hat{\alpha}_k \underset{\sim}{J}'_k\right)^{-1}\left(\sum_{\ell=0}^{q}\hat{\alpha}_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\ .$$

The second partial derivatives of $(2/N)\log L$ are

$$(3.9)\quad \frac{\partial^2}{\partial\alpha_i\,\partial\alpha_j}\ \frac{2}{N}\log L = 2\ \mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{J}_i\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{J}_j$$

$$-\frac{2}{\sigma^2}\ \mathrm{tr}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{J}_i\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}$$

$$-\frac{2}{\sigma^2}\ \mathrm{tr}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_i\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}$$

$$-\frac{2}{\sigma^2}\ \mathrm{tr}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_i\left(\sum_{\ell=0}^{q}\alpha_\ell \underset{\sim}{J}'_\ell\right)^{-1},$$

$$i,\ j = 1,\ \ldots\ ,\ q\ ,$$

$$(3.10)\quad \frac{\partial^2}{\partial\alpha_j\,\partial\sigma^2}\,\frac{2}{N}\log L = -\frac{2}{\sigma^4}\,\mathrm{tr}\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}'_\ell\right)^{-1}\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}'_\ell\right)^{-1},$$

$$j = 1, \dots , q,$$

$$(3.11)\quad \frac{\partial^2}{\partial(\sigma^2)^2}\,\frac{2}{N}\log L = \frac{T}{\sigma^4} - \frac{2}{\sigma^6}\,\mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k\,\underset{\sim}{J}'_k\right)^{-1}\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{C}\ .$$

The information matrix has elements which are N times

$$(3.12)\quad -\mathcal{E}\,\frac{\partial^2}{\partial\alpha_i\,\partial\alpha_j}\,\frac{1}{N}\log L = \mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k\,\underset{\sim}{J}_k\right)^{-1}\underset{\sim}{J}_i\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}_\ell\right)^{-1}\underset{\sim}{J}_j$$

$$+\ \mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k\,\underset{\sim}{J}_k\right)^{-1}\underset{\sim}{J}_i\,\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}'_\ell\right)^{-1},$$

$$i, j = 1, \dots , q ,$$

$$(3.13)\quad -\mathcal{E}\,\frac{\partial^2}{\partial\alpha_j\,\partial\sigma^2}\,\frac{1}{N}\log L = \frac{1}{\sigma^2}\,\mathrm{tr}\,\underset{\sim}{J}'_j\left(\sum_{\ell=0}^{q}\alpha_\ell\,\underset{\sim}{J}'_\ell\right)^{-1},\quad j = 1, \dots , q ,$$

$$(3.14)\quad -\mathcal{E}\,\frac{\partial^2}{\partial(\sigma^2)^2}\,\frac{1}{N}\log L = \frac{T}{2\,\sigma^4}\ .$$

As $N \to \infty$, the maximum likelihood estimates have a limiting normal distribution with covariance matrix whose inverse has elements given by (3.12), (3.13), and (3.14).

The likelihood equations (3.7) and (3.8) cannot in general be solved explicitly. However, the method of scoring can be used. If $L(\underset{\sim}{y}|\underset{\sim}{\theta})$ is the likelihood function of a vector parameter $\underset{\sim}{\theta}$, the Taylor's expansion of the (vector) derivative is

$$(3.15)\quad \frac{\partial}{\partial\underset{\sim}{\theta}}\log L(\underset{\sim}{y}|\underset{\sim}{\theta}) = \frac{\partial}{\partial\underset{\sim}{\theta}}\log L(\underset{\sim}{y}|\underset{\sim}{\theta})\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}^*} + \frac{\partial^2\log L(\underset{\sim}{y}|\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}\,\partial\underset{\sim}{\theta}'}\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}^*}(\underset{\sim}{\theta}-\underset{\sim}{\theta}^*) + R(\underset{\sim}{y}|\underset{\sim}{\theta},\underset{\sim}{\theta}^*)\ .$$

The matrix $(\partial^2/\partial\underset{\sim}{\theta}\ \partial\underset{\sim}{\theta}') \log L(\underset{\sim}{y}|\underset{\sim}{\theta})$ will be close to its expected value, which is a function of $\underset{\sim}{\theta}$, taken to be the "true" value of the parameter vector. Under certain conditions if $\underset{\sim}{\theta}^*$ is a consistent estimate of the "true" value, the solution to

$$\left[-\left.\frac{\partial^2 \log L(\underset{\sim}{y}|\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}\ \partial\underset{\sim}{\theta}'}\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}^*}\right](\underset{\sim}{\tilde{\theta}}-\underset{\sim}{\theta}^*) = \left.\frac{\partial}{\partial\underset{\sim}{\theta}} \log L(\underset{\sim}{y}|\underset{\sim}{\theta})\right|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}^*} \tag{3.16}$$

is a consistent, asymptotically efficient and asymptotically normal estimate of $\underset{\sim}{\theta}$. The procedure can be iterated; in suitable circumstances the sequence of vectors will converge to the maximum likelihood estimate, that is, a solution to the left-hand side of (3.15) set equal to $\underset{\sim}{0}$.

In the present case let $\hat{\alpha}_1^{(0)}, \ldots, \hat{\alpha}_q^{(0)}, \hat{\sigma}_0^2$ be a set of initial estimates, and let $\hat{\alpha}_1^{(i)}, \ldots, \hat{\alpha}_q^{(i)}, \hat{\sigma}_i^2$ be the solution to the i-th set of equations. It will be convenient to let

$$\underset{\sim}{\hat{A}}_{i-1} = \sum_{k=0}^{q} \hat{\alpha}_k^{(i-1)} \underset{\sim}{J}_k . \tag{3.17}$$

Then the i-th iteration involves the equations

$$\begin{aligned}\sum_{j=1}^{q} \left[\operatorname{tr} \underset{\sim}{\hat{A}}_{i-1}^{-1} \underset{\sim}{J}_g \underset{\sim}{\hat{A}}_{i-1}^{-1} \underset{\sim}{J}_j + \operatorname{tr} \underset{\sim}{\hat{A}}_{i-1}^{-1} \underset{\sim}{J}_g \underset{\sim}{J}_j' \underset{\sim}{\hat{A}}_{i-1}'^{-1}\right]\left(\hat{\alpha}_j^{(i)} - \hat{\alpha}_j^{(i-1)}\right) \\ + \frac{1}{\hat{\sigma}_{i-1}^2} \operatorname{tr} \underset{\sim}{A}_{i-1}^{-1} \underset{\sim}{J}_g \left(\hat{\sigma}_i^2 - \hat{\sigma}_{i-1}^2\right) \\ = -\operatorname{tr} \underset{\sim}{\hat{A}}_{i-1}^{-1} \underset{\sim}{J}_g + \frac{1}{\hat{\sigma}_{i-1}^2} \operatorname{tr} \underset{\sim}{\hat{A}}_{i-1}^{-1} \underset{\sim}{C} \underset{\sim}{\hat{A}}_{i-1}'^{-1} \underset{\sim}{J}_g' \underset{\sim}{\hat{A}}_{i-1}'^{-1} , \\ g = 1, \ldots, q ,\end{aligned} \tag{3.18}$$

$$(3.19)\quad \frac{1}{\hat{\sigma}^2_{i-1}} \sum_{j=1}^{q} \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_j \left(\hat{\alpha}^{(i)}_j - \hat{\alpha}^{(i-1)}_j \right) + \frac{T}{2\hat{\sigma}^4_{i-1}} \hat{\sigma}^2_i - \hat{\sigma}^2_{i-1}$$

$$= - \frac{T}{2\hat{\sigma}^2_{i-1}} + \frac{1}{2\hat{\sigma}^4_{i-1}} \operatorname{tr} \hat{\underset{\sim}{A}}'^{-1}_{i-1} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{C} .$$

These reduce to

$$(3.20)\quad \sum_{j=1}^{q} \left[\operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_j + \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \underset{\sim}{J}'_j \hat{\underset{\sim}{A}}'^{-1}_{i-1} \right] \hat{\alpha}^{(i)}_j + \frac{1}{\hat{\sigma}^2_{i-1}} \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \hat{\sigma}^2_i$$

$$= 2 \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g + \frac{1}{\hat{\sigma}^2_{i-1}} \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{C} \hat{\underset{\sim}{A}}'^{-1}_{i-1} \underset{\sim}{J}'_g \hat{\underset{\sim}{A}}'^{-1}_{i-1} - \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_0$$

$$- \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \underset{\sim}{J}'_0 \hat{\underset{\sim}{A}}'^{-1}_{i-1} , \qquad g = 1, \ldots , q ,$$

$$(3.21)\quad \sum_{j=1}^{q} \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_j \hat{\alpha}^{(i)}_j + \frac{T}{2\hat{\sigma}^2_{i-1}} \hat{\sigma}^2_i$$

$$= T + \frac{1}{2\hat{\sigma}^2_{i-1}} \operatorname{tr} \hat{\underset{\sim}{A}}'^{-1}_{i-1} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{C} - \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_0$$

If $\sigma^2 = 1$ and α_0 is a free parameter (not specified), the likelihood satisfies (3.3) with $\sigma^2 = 1$, the first partial derivatives are (3.5) for $j = 0, 1, \ldots , q$, the elements of the information matrix are N times (3.12) for $i, j = 0, 1, \ldots , q$, and the equations for scoring are

$$(3.22)\quad \sum_{j=0}^{q} \left[\operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_j + \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g \underset{\sim}{J}'_j \hat{\underset{\sim}{A}}'^{-1}_{i-1} \right] \left(\hat{\alpha}^{(i)}_j - \hat{\alpha}^{(i-1)}_j \right)$$

$$= - \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{J}_g + \operatorname{tr} \hat{\underset{\sim}{A}}^{-1}_{i-1} \underset{\sim}{C} \hat{\underset{\sim}{A}}'^{-1}_{i-1} \underset{\sim}{J}'_g \hat{\underset{\sim}{A}}'^{-1}_{i-1} ,$$

$$g = 0, 1, \ldots , q ,$$

These reduce to

$$(3.23)\qquad \sum_{j=0}^{q}\left[\operatorname{tr} \hat{\underset{\sim}{A}}_{i-1}^{-1} \underset{\sim}{J}_g \hat{\underset{\sim}{A}}_{i-1}^{-1} \underset{\sim}{J}_j + \operatorname{tr} \hat{\underset{\sim}{A}}_{i-1}^{-1} \underset{\sim}{J}_g \underset{\sim}{J}_j' \hat{\underset{\sim}{A}}_{i-1}'^{-1}\right] \hat{\alpha}_j^{(i)}$$

$$= \operatorname{tr} \hat{\underset{\sim}{A}}_{i-1}^{-1} \underset{\sim}{J}_g + \operatorname{tr} \hat{\underset{\sim}{A}}_{i-1}^{-1} \underset{\sim}{C} \hat{\underset{\sim}{A}}_{i-1}'^{-1} \underset{\sim}{J}_g' \hat{\underset{\sim}{A}}_{i-1}'^{-1} ,$$

$$g = 0, 1, \ldots , q .$$

3.2 Moving Average Process Approximated by a Linear Transformation.

The moving average process $\{y_t\}$ is (1.1) for $\beta_1 = \ldots = \beta_q = 0$, that is,

$$(3.24)\qquad y_t = \sum_{j=0}^{q} \alpha_j v_{t-j} ,$$

$t = \ldots , -1, 0, 1, \ldots$. Then the distribution of $y_1, \ldots , y_T$ is approximated by the distribution of $\underset{\sim}{y}$ defined by (3.1) when $\underset{\sim}{J}_g = \underset{\sim}{L}^g$, $g = 0, 1, \ldots , q$. The components of (3.1) are

$$(3.25)\qquad y_t = \sum_{j=0}^{t-1} \alpha_j v_{t-j} , \qquad t = 1 ,\ldots , q ,$$

$$(3.26)\qquad y_t = \sum_{j=0}^{q} \alpha_j v_{t-j} , \qquad t = q+1, \ldots , T .$$

The equations (3.26) correspond to a moving average process; the moving averages of the first q observations, represented by (3.25), are truncated.

The covariance matrix of the moving average process defined by (3.24) is

$$(3.27)\qquad \sum_{j=0}^{q} \sigma^2 \alpha_j^2 \underset{\sim}{I} + \sum_{i=1}^{q} \sum_{j=0}^{q-i} \sigma^2 \alpha_j \alpha_{j+i} (\underset{\sim}{L}^i + \underset{\sim}{L}'^i).$$

This is of the form considered in T. W. Anderson (1969), (1970), (1971b), and (1973), namely

(3.28) $$\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g ,$$

where $\underset{\sim}{G}_0 = \underset{\sim}{I}$ and

(3.29) $$\underset{\sim}{G}_g = (\underset{\sim}{L}^g + \underset{\sim}{L}'^g) , \qquad g = 1, \ldots , q .$$

The covariance matrix of $y_1, \ldots , y_T$ defined by (3.25) and (3.26) is for $q = 2$, for example,

(3.30) $$\sigma^2 \begin{pmatrix} \alpha_0^2 & \alpha_0\alpha_1 & \alpha_0\alpha_2 & 0 & \cdots & 0 \\ \alpha_0\alpha_1 & \alpha_0^2 + \alpha_1^2 & \alpha_0\alpha_1 + \alpha_1\alpha_2 & \alpha_0\alpha_2 & \cdots & 0 \\ \alpha_0\alpha_2 & \alpha_0\alpha_1 + \alpha_1\alpha_2 & \alpha_0^2 + \alpha_1^2 + \alpha_2^2 & \alpha_0\alpha_1 + \alpha_1\alpha_2 & \cdots & 0 \\ 0 & \alpha_0\alpha_2 & \alpha_0\alpha_1 + \alpha_1\alpha_2 & \alpha_0^2 + \alpha_1^2 + \alpha_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha_0^2 + \alpha_1^2 + \alpha_2^2 \end{pmatrix} .$$

This matrix differs from (3.27) for $q = 2$ in the upper left-hand 2×2 submatrix in (3.30). If T is large relative to q the difference between the two models will not be important; the model (3.1) with $\underset{\sim}{J}_j \doteq \underset{\sim}{L}^j$ can be considered as an approximation to the moving average process.

When $\underset{\sim}{J}_j = \underset{\sim}{L}^j$,

(3.31) $$\operatorname{tr}\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{J}_\ell\right)^{-1} \underset{\sim}{J}_j = \operatorname{tr}\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^j = 0 , \quad j = 1, \ldots , q ,$$

(3.32) $$\operatorname{tr}\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{J}_\ell\right)^{-1} \underset{\sim}{J}_0 = \operatorname{tr}\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} .$$

The likelihood equations (3.7) and (3.8) for $\hat{\alpha}_1, \ldots, \hat{\alpha}_q$ and $\hat{\sigma}^2$ (with $\hat{\alpha}_0 = 1$) are

$$\text{(3.33)} \qquad \operatorname{tr}\left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^g \left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{C} \left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell \underset{\sim}{L}'^\ell\right)^{-1} = 0 , \quad g = 1, \ldots , q ,$$

$$\text{(3.34)} \qquad \hat{\sigma}^2 = \frac{1}{T} \operatorname{tr}\left(\sum_{k=0}^{q} \hat{\alpha}_k \underset{\sim}{L}'^k\right)^{-1}\left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{C} .$$

The method of scoring leads to

$$\text{(3.35)} \qquad \sum_{j=1}^{q} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{L}^g \underset{\sim}{L}'^j \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} \left(\hat{\alpha}_j^{(i)} - \hat{\alpha}_j^{(i-1)}\right)$$

$$= \frac{1}{\hat{\sigma}_{i-1}^2} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{C} \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} \underset{\sim}{L}'^g \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} , \quad g = 1, \ldots , q ,$$

$$\text{(3.36)} \qquad \frac{T}{2\hat{\sigma}_{i-1}^4}\left(\hat{\sigma}_i^2 - \hat{\sigma}_{i-1}^2\right) = -\frac{T}{2\hat{\sigma}_{i-1}^2} + \frac{1}{2\hat{\sigma}_{i-1}^4} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{C} .$$

These can be written

$$\text{(3.37)} \qquad \sum_{j=1}^{q} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{L}^g \underset{\sim}{L}'^j \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} \hat{\alpha}_j^{(i)} = - \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{L}^g \hat{\underset{\sim}{A}}{}_{i-1}'^{-1}$$

$$+ \frac{1}{\hat{\sigma}_{i-1}^2} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{C} \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} \underset{\sim}{L}'^g \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} ,$$

$$g = 1, \ldots , q ,$$

$$\text{(3.38)} \qquad \hat{\sigma}_i^2 = \frac{1}{T} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} \hat{\underset{\sim}{A}}{}_{i-1}^{-1} \underset{\sim}{C} .$$

The set of linear equations (3.37) are solved for $\hat{\alpha}_1^{(i)}, \ldots , \hat{\alpha}_q^{(i)}$.

If the parameters are $\alpha_0, \alpha_1, \ldots , \alpha_q$ $(\sigma^2 = 1)$, then the likelihood equations are (3.33) for $g = 0, 1, \ldots , q$. The equations for scoring are

$$(3.39)\quad \left(\operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-2} + \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\right)\left(\hat{\alpha}_0^{(i)} - \hat{\alpha}_0^{(i-1)}\right) + \sum_{j=1}^{q} \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{\prime j}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\left(\hat{\alpha}_j^{(i)} - \hat{\alpha}_j^{(i-1)}\right) = -\operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1} + \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{C}\,\hat{\mathbf{A}}_{i-1}^{\prime -2}\,,$$

$$(3.40)\quad \sum_{j=0}^{q} \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{g}\,\mathbf{L}^{\prime j}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\left(\hat{\alpha}_j^{(i)} - \hat{\alpha}_j^{(i-1)}\right) = \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{C}\,\hat{\mathbf{A}}_{i-1}^{\prime}\,\mathbf{L}^{\prime g}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,,\qquad g = 1, \ldots, q\,.$$

These reduce to

$$(3.41)\quad \left(\operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-2} + \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\right)\hat{\alpha}_0^{(i)} + \sum_{j=1}^{q} \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{\prime j}\,\hat{\mathbf{A}}_{i-1}^{\prime}\,\hat{\alpha}_j^{(i)} = \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-2}\,\hat{\alpha}_0^{(i-1)} + \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{C}\,\hat{\mathbf{A}}_{i-1}^{\prime -2}\,,$$

$$(3.42)\quad \sum_{j=0}^{q} \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{g}\,\mathbf{L}^{\prime j}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,\hat{\alpha}_j^{(i)} = \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{C}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,\mathbf{L}^{\prime g}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,,\qquad g = 1, \ldots, q\,.$$

These form a set of $q + 1$ linear equations in $q + 1$ unknowns.

If $N = 1$ and $\mathbf{y}_1 = \mathbf{y}$, then $\mathbf{C} = \mathbf{y}\mathbf{y}'$ and

$$(3.43)\quad \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{C}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,\mathbf{L}^{\prime g}\,\hat{\mathbf{A}}_{i-1}^{\prime -1} = \mathbf{y}'\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,\mathbf{L}^{\prime g}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{y}\,,\qquad g = 0, 1, \ldots, q\,.$$

The equations (3.37) and (3.38) are then

$$(3.44)\quad \sum_{j=1}^{q} \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{g}\left(\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{j}\right)'\hat{\alpha}_j^{(i)} = \frac{\left(\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{y}\right)'\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{g}\left(\hat{\mathbf{A}}_{i-1}^{-1}\mathbf{y}\right)}{\hat{\sigma}_{i-1}^{2}} - \operatorname{tr}\hat{\mathbf{A}}_{i-1}^{-1}\,\mathbf{L}^{g}\,\hat{\mathbf{A}}_{i-1}^{\prime -1}\,,\qquad g = 1, \ldots, q\,,$$

$$\hat{\sigma}_i^2 = \frac{1}{T}\left(\hat{\underset{\sim}{A}}_{i-1}^{-1}\,\underset{\sim}{y}\right)'\ \hat{\underset{\sim}{A}}_{i-1}^{-1}\,\underset{\sim}{y} \quad . \tag{3.45}$$

The calculation of $\hat{\underset{\sim}{A}}_{i-1}^{-1}\,\underset{\sim}{y}$ can be done by solving

$$\sum_{\ell=0}^{q} \hat{\alpha}_\ell^{(i-1)}\,\underset{\sim}{L}^\ell\,\underset{\sim}{z} = \underset{\sim}{y} \ . \tag{3.46}$$

The matrix of coefficients has the form (2.38) (with β_ℓ replaced by $\hat{\alpha}_\ell^{(i-1)}$, $\ell = 1, \ldots , q$). The component equations are $z_1 = y_1$,

$$z_t + \sum_{s=1}^{t-1} \hat{\alpha}_s^{(i-1)}\, z_{t-s} = y_t \ , \qquad t = 2, \ldots , q \ , \tag{3.47}$$

$$z_t + \sum_{s=1}^{q} \hat{\alpha}_s^{(i-1)}\, z_{t-s} = y_t \ , \qquad t = q+1, \ldots , T \ . \tag{3.48}$$

These can be solved successively for $z_2, \ldots , z_T$. Each component z_t involves at most q multiplications and the entire solution less than qT multiplications.

The first column of $\hat{\underset{\sim}{A}}_{i-1}^{-1}$ can be obtained by solving (3.46) with $\underset{\sim}{y}$ replaced by the first column of $\underset{\sim}{I}$. Thus $z_1 = 1$ and the successive calculations are

$$z_t = -\sum_{s=1}^{t-1} \hat{\alpha}_s^{(i-1)}\, z_{t-s} \ , \qquad t = 2, \ldots , q \ , \tag{3.49}$$

$$z_t = -\sum_{s=1}^{q} \hat{\alpha}_s^{(i-1)}\, z_{t-s} \ , \qquad t = q+1, \ldots , T \ . \tag{3.50}$$

The $(j+1)$-th column of $\hat{\underset{\sim}{A}}_{i-1}^{-1}$ is simply $\underset{\sim}{L}^j$ times the first column; that is, it is the first column displaced by j units for

$$(3.51) \qquad \underset{\sim}{L}^j \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ z_1 \\ \vdots \\ z_{T-j} \end{pmatrix}, \qquad j = 1, \ldots, T-1,$$

$$(3.52) \qquad \underset{\sim}{L}^j \, \underset{\sim}{z} = \underset{\sim}{0} , \qquad j = T, T+1, \ldots .$$

Thus the calculation of $\hat{\underset{\sim}{A}}_{i-1}^{-1} \, \underset{\sim}{E}^g$ involves less than Tq multiplications.

Another way of looking at the calculation of $(\Sigma_{\ell=0}^{q} \alpha_\ell \, \underset{\sim}{L}^\ell)^{-1}$, where we drop the carat and superscript on $\hat{\alpha}_\ell^{(i-1)}$ for convenience is to see that

$$(3.53) \qquad \underset{\sim}{I} = \sum_{\ell=0}^{q} \alpha_\ell \, \underset{\sim}{L}^\ell \sum_{j=0}^{T-1} \delta_j \, \underset{\sim}{L}^j$$

$$= \sum_{\ell=0}^{q} \sum_{j=0}^{T-1} \alpha_\ell \, \delta_j \, \underset{\sim}{L}^{\ell+j}$$

$$= \sum_{i=0}^{T-1} \sum_{\ell+j=i} \alpha_\ell \, \delta_j \, \underset{\sim}{L}^i$$

because $\underset{\sim}{L}^i = \underset{\sim}{0}$ for $i = T, T+1, \ldots$ if $\delta_0 = 1$,

$$(3.54) \qquad \alpha_0 \, \delta_0 = 1 ,$$

$$(3.55) \qquad \sum_{\ell=0}^{i} \alpha_\ell \, \delta_{i-\ell} = 0 , \qquad i = 1, \ldots, q-1 ,$$

$$(3.56) \qquad \sum_{\ell=0}^{q} \alpha_\ell \, \delta_{i-\ell} = 0 , \qquad i = q, q+1, \ldots .$$

The coefficients $\delta_0, \delta_1, \ldots$ satisfy the homogeneous linear difference equation (3.56) with q boundary conditions (3.54) and (3.55). Therefore

$$(3.57) \qquad \delta_i = \sum_{\ell=1}^{q} k_\ell \, z_\ell^i , \qquad i = 0, 1, \ldots ,$$

where $z_1, \ldots, z_q$ are the roots of the associated polynomial equation

$$\sum_{\ell=0}^{q} \alpha_\ell z^{q-\ell} = 0 , \tag{3.58}$$

and $k_1, \ldots, k_q$ are determined so (3.57) satisfies the boundary conditions (3.54) and (3.55). [The form (3.58) is on the basis that the q roots are different.] Then the inverse is

$$\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} = \sum_{i=0}^{\infty} \delta_i \underset{\sim}{L}^i = \sum_{i=0}^{T-1} \delta_i \underset{\sim}{L}^i . \tag{3.59}$$

It may be observed that (3.54), (3.55), and (3.56) are identical to (39) and (40) of Section 5.2 of T. W. Anderson (1971a) with β_j replaced by α_j and p replaced by q. Thus the coefficients $\delta_0, \delta_1, \ldots$ correspond to the moving average representation of an autoregressive process with coefficients $1, \alpha_1, \ldots, \alpha_q$.

Then

$$\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^k = \sum_{i=0}^{T-1} \delta_i \underset{\sim}{L}^{i+k} = \sum_{i=0}^{T-1-k} \delta_i \underset{\sim}{L}^{i+k} \tag{3.60}$$

because $\underset{\sim}{L}^{i+k} = 0$ if $i+k \geq T$.

The coefficient of $\hat{\alpha}_k^{(i)}$ in the j-th equation of (3.44) has the form

$$\operatorname{tr} \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^j \underset{\sim}{L}'^k \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}'^\ell\right)^{-1} \tag{3.61}$$

$$= \operatorname{tr} \sum_{g=0}^{T-1-j} \sum_{i=0}^{T-1-k} \delta_g \delta_i \underset{\sim}{L}^{g+j} \underset{\sim}{L}'^{i+k} , \qquad j, k = 1, \ldots, q .$$

A matrix $\underset{\sim}{L}^h \underset{\sim}{L}'^\ell$ has all elements 0 except along the diagonal $h - \ell$ entries below the main diagonal, which consists of 1's and 0's. In particular, $L^h L'^\ell$ has only 0's on the main diagonal if $h \neq \ell$, and $L^h L'^h$ has 1's on the main diagonal except for that first h entries being 0. Hence

$$\text{(3.62)} \qquad \operatorname{tr} \underset{\sim}{L}^h \underset{\sim}{L}'^{\ell} = 0 , \qquad h \neq \ell ,$$

$$\text{(3.63)} \qquad \operatorname{tr} \underset{\sim}{L}^h \underset{\sim}{L}'^h = T-h , \qquad h = 0, 1, \ldots , T-1 ,$$

$$\text{(3.64)} \qquad \operatorname{tr} \underset{\sim}{L}^h \underset{\sim}{L}'^h = 0. \qquad h = T, T+1, \ldots .$$

Thus (3.61) is

$$\text{(3.65)} \qquad \operatorname{tr}\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^j \underset{\sim}{L}'^k \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}'^\ell\right)^{-1}$$
$$= \sum_{i=0}^{T-1 - \max(j,k)} [T-i- \max(j,k)] \, \delta_{i+|k-j|} \delta_i .$$

Note that

$$\text{(3.66)} \qquad \sigma^2 \sum_{i=0}^{\infty} \delta_{i+|k-j|} \delta_i = \sigma_{AR}(k-j) ,$$

where $\sigma_{AR}(k-j)$ is the $(k-j)$-th covariance of the autoregressive process corresponding to the coefficients $1, \alpha_1, \ldots , \alpha_q$ and variance σ^2. Thus (3.65) is approximately $T \sigma_{AR}(k-j)/\sigma^2$, especially if the roots of (3.58) are small and thus the series (3.66) converges rapidly. In particular

$$\text{(3.67)} \qquad \lim_{T\to\infty} \frac{1}{T} \operatorname{tr}\left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^j \underset{\sim}{L}'^k \left(\sum_{\ell=0}^{q} \alpha_\ell \underset{\sim}{L}'^\ell\right)^{-1} = \frac{\sigma_{AR}(k-j)}{\sigma^2} .$$

There are various ways of calculating $\sigma_{AR}(h)$ given $\alpha_1, \ldots , \alpha_q$ and σ^2 [See Section 5.2 of T. W. Anderson (1971a), for example.]

The equations (3.44) are approximately

$$\text{(3.68)} \qquad \sum_{k=1}^{q} \hat{\sigma}_{AR}^{(i-1)}(k-j) \, \hat{\alpha}_k^{(i)} = d_j , \qquad j = 1, \ldots , q ,$$

where

$$(3.69)\quad d_j = \frac{1}{T}\left[\left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell^{(i-1)} \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{y}\right]' \left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell^{(i-1)} \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^j \left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell^{(i-1)} \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{y}$$
$$- \frac{\hat{\sigma}_{i-1}^2}{T} \operatorname{tr} \left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell^{(i-1)} \underset{\sim}{L}^\ell\right)^{-1} \underset{\sim}{L}^g \left(\sum_{\ell=0}^{q} \hat{\alpha}_\ell^{(i-1)} \underset{\sim}{L}'^\ell\right)^{-1}$$
$$j = 1, \ldots, q .$$

The $q \times q$ matrix whose elements are $\hat{\sigma}_{AR}^{(i-1)}(k-j)$ are the covariances of an autoregressive process of order q, whose coefficients are $1, \hat{\alpha}_1^{(i-1)}, \ldots, \hat{\alpha}_q^{(i-1)}$. Then the solution to (3.68) is

$$(3.70)\qquad \hat{\alpha}_k^{(i)} = \sum_{j=1}^{q} f_{kj}\, d_j ,$$

where $(f_{kj}) = [\hat{\sigma}_{AR}(k-j)]^{-1}$. The elements f_{kj} are the coefficients of the quadratic form of $v_1, \ldots, v_q$ having a normal distribution with covariance matrix $[\hat{\sigma}_{AR}^{(i-1)}(k-j)]$. The matrix is

$$(3.71)\quad \hat{\sigma}_{i-1}^2 \begin{pmatrix} 1 & \hat{\alpha}_1^{(i-1)} & \cdots \\ \hat{\alpha}_1^{(i-1)} & 1 + \hat{\alpha}_1^{(i-1)2} & \cdots \\ \hat{\alpha}_2^{(i-1)} & \hat{\alpha}_1^{(i-1)} + \hat{\alpha}_2^{(i-1)}\hat{\alpha}_1^{(i-1)} & \cdots \\ \vdots & \vdots & \\ \hat{\alpha}_q^{(i-1)} & & \end{pmatrix} .$$

The matrix is persymmetric; that is, it is symmetric about the transverse diagonal. If q is odd the middle term is

$$(3.72)\qquad \hat{\sigma}_{i-1}^2 \left[1 + \hat{\alpha}_1^{(i-1)2} + \ldots + \hat{\alpha}_{(q-1)/2}^{(i-1)2}\right] .$$

The matrix is essentially derived in Section 6.2 of T. W. Anderson (1971).

It can further be shown that

$$(3.73)\quad \lim \frac{1}{T} \operatorname{tr} \left(\sum_{\ell=0}^{q} \alpha_\ell L^\ell\right)^{-1} L^j L'^k \left(\sum_{\ell=0}^{q} \alpha_\ell L'^\ell\right)^{-1} = \int_{-\pi}^{\pi} \frac{\cos \lambda j \cos \lambda k}{f(x)}\, d\lambda .$$

4. Estimation of Coefficients of Linear Transformations When a Covariance Matrix Has Linear Structure; Autoregressive Processes with Moving Average Residuals.

Let

$$(4.1)\qquad \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{y} = \underset{\sim}{u} ,$$

where $\mathscr{E}\underset{\sim}{u} = 0$ and

$$(4.2)\qquad \mathscr{C}(\underset{\sim}{u}) = \mathscr{E}\underset{\sim}{u}\underset{\sim}{u}' = \sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g ,$$

$\underset{\sim}{G}_0, \underset{\sim}{G}_1, \dots , \underset{\sim}{G}_q$ are $q + 1$ known linearly independent symmetric $T \times T$ matrices and $\sigma_0, \sigma_1, \dots , \sigma_q$ are $q + 1$ parameters such that $\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g$ is positive definite. Then $\underset{\sim}{y}$ has mean vector $\mathscr{E}\underset{\sim}{y} = \underset{\sim}{0}$ and covariance matrix

$$(4.3)\qquad \mathscr{C}(\underset{\sim}{y}) = \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell\right)^{-1} \sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g \left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k'\right)^{-1}$$

with inverse

$$(4.4)\qquad \mathscr{C}^{-1}(\underset{\sim}{y}) = \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell$$

$$= \sum_{k,\ell=0}^{p} \beta_k \beta_\ell \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{K}_\ell .$$

We assume $\beta_0 = 1$.

If $\underset{\sim}{u}$ is normally distributed, then $2/N$ times the logarithm of the likelihood is

$$(4.5)\qquad \frac{2}{N} \log L = - T \log 2\pi + 2 \log \left| \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right| - \log \left| \sum_{g=0}^{q} \sigma_g G_g \right|$$

$$- \operatorname{tr} \sum_{k,\ell=0}^{p} \beta_k \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{C} .$$

The partial derivatives are

$$(4.6)\quad \frac{\partial}{\partial \sigma_f} \frac{2}{N} \log L = - \operatorname{tr}\left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f$$

$$+ \operatorname{tr} \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{C} ,$$

$$f = 0, 1, \ldots, q ,$$

$$(4.7)\quad \frac{\partial}{\partial \beta_\ell} \frac{2}{N} \log L = 2 \operatorname{tr}\left(\sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k\right)^{-1} \underset{\sim}{K}_\ell - 2 \operatorname{tr} \underset{\sim}{C} \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{K}_\ell ,$$

$$\ell = 1, \ldots, p .$$

In case $\underset{\sim}{K}_k = \underset{\sim}{L}^k$, $k = 0, 1, \ldots, p$, $N = 1$, and $\underset{\sim}{y}^{(k)} = \underset{\sim}{L}^k \underset{\sim}{y}$, the derivative equations are

$$(4.8)\quad \operatorname{tr}\left(\sum_{g=0}^{q} \hat{\sigma}_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f = \sum_{k,\ell=0}^{p} \hat{\beta}_k \hat{\beta}_\ell \underset{\sim}{y}^{(k)'} \left(\sum_{g=0}^{q} \hat{\sigma}_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f \left(\sum_{g=0}^{q} \hat{\sigma}_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{y}^{(\ell)} ,$$

$$f = 0, 1, \ldots, q ,$$

$$(4.9)\quad \sum_{k=1}^{p} \underset{\sim}{y}^{(k)'} \left(\sum_{g=0}^{q} \hat{\sigma}_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{y}^{(\ell)} \hat{\beta}_k = - \underset{\sim}{y}^{(0)'} \left(\sum_{g=0}^{q} \hat{\sigma}_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{y}^{(\ell)} ,$$

$$\ell = 1, \ldots, p .$$

The second partial derivatives of $(2/N) \log L$ are

$$(4.10)\quad \frac{\partial^2}{\partial \sigma_f\, \partial \sigma_h} \frac{2}{N} \log L = \operatorname{tr} \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_h$$

$$- 2 \operatorname{tr} \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1}$$

$$\underset{\sim}{G}_h \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{C} ,$$

$$f, h = 0, 1, \ldots, q ,$$

$$(4.11)\quad \frac{\partial^2}{\partial \sigma_f\, \partial \beta_\ell} \frac{2}{N} \log L = 2 \operatorname{tr} \sum_{k=0}^{p} \beta_k \underset{\sim}{K}_k' \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{G}_f \left(\sum_{g=0}^{q} \sigma_g \underset{\sim}{G}_g\right)^{-1} \underset{\sim}{K}_\ell \underset{\sim}{C} ,$$

$$f = 0, 1, \ldots, q , \quad \ell = 1, \ldots, p ,$$

$$(4.12)\qquad \frac{\partial^2}{\partial\beta_\ell\,\partial\beta_{\ell'}}\ \frac{2}{N}\log L = -2\,\mathrm{tr}\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k'\right)^{-1}\underset{\sim}{K}_\ell\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_{\ell'}$$

$$-\,2\,\mathrm{tr}\,\underset{\sim}{C}\,\underset{\sim}{K}_{\ell'}'\left(\sum_{g=0}^{q}\sigma_g\,\underset{\sim}{G}_g\right)^{-1}\underset{\sim}{K}_\ell\ ,$$

$$\ell,\ \ell' = 1,\ \ldots\ ,\ p\ .$$

The information matrix has elements that are N times

$$(4.13)\quad -\mathcal{E}\,\frac{\partial^2}{\partial\sigma_f\,\partial\sigma_h}\ \frac{1}{N}\log L = \frac{1}{2}\ \mathrm{tr}\left(\sum_{g=0}^{q}\sigma_g\,\underset{\sim}{G}_g\right)^{-1}\underset{\sim}{G}_f\left(\sum_{g=0}^{q}\sigma_g\,\underset{\sim}{G}_g\right)^{-1}\underset{\sim}{G}_h\ ,$$

$$f,\ h = 0,\ 1,\ \ldots\ ,\ q\ ,$$

$$(4.14)\quad -\mathcal{E}\,\frac{\partial^2}{\partial\sigma_f\,\partial\beta_\ell}\ \frac{1}{N}\log L = -\,\mathrm{tr}\,\underset{\sim}{G}_f\left(\sum_{g=0}^{q}\sigma_g\,\underset{\sim}{G}_g\right)^{-1}\underset{\sim}{K}_\ell\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\ ,$$

$$f = 0,\ 1,\ \ldots\ ,\ q\qquad \ell = 1,\ \ldots\ ,\ p\ ,$$

$$(4.15)\quad -\,\frac{\partial^2}{\partial\beta_\ell\,\partial\beta_{\ell'}}\ \frac{1}{N}\log L = \mathrm{tr}\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_\ell\left(\sum_{k=0}^{p}\beta_k\,\underset{\sim}{K}_k\right)^{-1}\underset{\sim}{K}_{\ell'}$$

$$+\,\mathrm{tr}\left(\sum_{\ell=0}^{p}\beta_\ell\,\underset{\sim}{K}_\ell\right)^{-1}\sum_{g=0}^{q}\sigma_g\,\underset{\sim}{G}_g\left(\sum_{\ell=0}^{p}\beta_\ell\,\underset{\sim}{K}_\ell'\right)^{-1}\underset{\sim}{K}_{\ell'}'\left(\sum_{g=0}^{q}\sigma_g\,\underset{\sim}{G}_g\right)^{-1}\underset{\sim}{K}_\ell\ ,$$

$$\ell,\ \ell' = 1,\ \ldots\ ,\ p\ .$$

Let

$$(4.16)\qquad \hat{\underset{\sim}{B}}_{i-1} = \sum_{\ell=0}^{p}\hat{\beta}_\ell^{(i-1)}\,\underset{\sim}{K}_\ell\ .$$

The method of scoring leads to the following iterative procedure:

$$(4.17)\qquad \sum_{h=0}^{q}\mathrm{tr}\,(\hat{\underset{\sim}{\Sigma}}{}^{u}_{i-1})^{-1}\,\underset{\sim}{G}_f\,(\hat{\underset{\sim}{\Sigma}}{}^{u}_{i-1})^{-1}\,\underset{\sim}{G}_h\left(\hat{\sigma}_h^{(i)} - \hat{\sigma}_h^{(i-1)}\right)$$

$$-\,2\sum_{\ell=1}^{p}\mathrm{tr}\,\underset{\sim}{G}_f\,(\hat{\underset{\sim}{\Sigma}}{}^{u}_{i-1})^{-1}\,\underset{\sim}{K}_\ell\,\hat{\underset{\sim}{B}}{}^{-1}_{i-1}\left(\hat{\beta}_\ell^{(i)} - \hat{\beta}_\ell^{(i-1)}\right)$$

$$= -\,\mathrm{tr}\,(\hat{\underset{\sim}{\Sigma}}{}^{u}_{i-1})^{-1}\,\underset{\sim}{G}_f + \mathrm{tr}\,\hat{\underset{\sim}{B}}{}'_{i-1}\,(\hat{\underset{\sim}{\Sigma}}{}^{u}_{i-1})^{-1}\,\underset{\sim}{G}_f\,(\hat{\underset{\sim}{\Sigma}}{}^{u}_{i-1})^{-1}\,\hat{\underset{\sim}{B}}_{i-1}\,\underset{\sim}{C}\ ,$$

$$f = 0,\ 1,\ \ldots\ ,\ q\ ,$$

(4.18)

$$- 2 \sum_{h=0}^{q} \operatorname{tr} \underset{\sim}{G}_h (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_j \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \left(\hat{\sigma}_h^{(i)} - \hat{\sigma}_h^{(i-1)}\right)$$

$$+ 2 \sum_{\ell=1}^{p} \left[\operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_j \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_\ell + \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\Sigma}^u_{i-1} \hat{\underset{\sim}{B}}{}'^{-1}_{i-1} \underset{\sim}{K}'_\ell (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_j\right] \left(\hat{\beta}_\ell^{(i)} - \hat{\beta}_\ell^{(i-1)}\right)$$

$$= 2 \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_j - 2 \operatorname{tr} \underset{\sim}{C} \hat{\underset{\sim}{B}}{}'_{i-1} (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_j , \qquad j = 1, \ldots , p .$$

These equations are equivalent to

$$\sum_{h=0}^{q} \operatorname{tr} (\hat{\Sigma}^u_{i-1})^{-1} G_f (\hat{\Sigma}^u_{i-1})^{-1} G_h \hat{\sigma}_h^{(i)} - 2 \sum_{\ell=1}^{p} \operatorname{tr} \underset{\sim}{G}_f (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_\ell \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\underset{\sim}{\beta}}{}_\ell^{(i)} \tag{4.19}$$

$$= -2 \operatorname{tr} \underset{\sim}{G}_f (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} + \operatorname{tr} \hat{\underset{\sim}{B}}{}'_{i-1} \quad (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{G}_f (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \hat{\underset{\sim}{B}}_{i-1} \underset{\sim}{C}$$

$$+ 2 \operatorname{tr} \underset{\sim}{G}_f (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_0 \hat{\underset{\sim}{B}}{}^{-1}_{i-1} , \qquad f = 0, 1, \ldots , q ,$$

$$- 2 \sum_{h=0}^{q} \operatorname{tr} \underset{\sim}{G}_h (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_j \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\underset{\sim}{\sigma}}{}_h^{(i)} \tag{4.20}$$

$$+ 2 \sum_{\ell=1}^{p} \left[\operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_j \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_\ell + \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\underset{\sim}{\Sigma}}{}^u_{i-1} \hat{\underset{\sim}{B}}{}'^{-1}_{i-1} \underset{\sim}{K}'_\ell (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_j\right] \hat{\beta}_\ell^{(i)}$$

$$= 4 \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_j - 2 \operatorname{tr} \underset{\sim}{C} \hat{\underset{\sim}{B}}{}'_{i-1} (\hat{\Sigma}^u_{i-1})^{-1} \underset{\sim}{K}_j - 2 \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_j \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_0$$

$$- 2 \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\underset{\sim}{\Sigma}}{}^u_{i-1} \hat{\underset{\sim}{B}}{}'^{-1}_{i-1} \underset{\sim}{K}'_0 (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{K}_j , \qquad j = 1, \ldots , p .$$

If $\underset{\sim}{K}_j = \underset{\sim}{L}^j$, then $\operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{K}_j = \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \underset{\sim}{L}^j = 0$. Then (4.20) is

$$- 2 \sum_{h=0}^{q} \operatorname{tr} \underset{\sim}{G}_h (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{L}^j \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\sigma}_h^{(i)} \tag{4.21}$$

$$+ 2 \sum_{\ell=1}^{p} \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\underset{\sim}{\Sigma}}{}^u_{i-1} \quad \hat{\underset{\sim}{B}}{}'^{-1}_{i-1} \underset{\sim}{L}'^\ell (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{L}^j \hat{\beta}_\ell^{(i)}$$

$$= - 2 \operatorname{tr} \underset{\sim}{C} \hat{\underset{\sim}{B}}{}'_{i-1} (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{L}^j - 2 \operatorname{tr} \hat{\underset{\sim}{B}}{}^{-1}_{i-1} \hat{\underset{\sim}{\Sigma}}{}^u_{i-1} \hat{\underset{\sim}{B}}{}'^{-1}_{i-1} (\hat{\underset{\sim}{\Sigma}}{}^u_{i-1})^{-1} \underset{\sim}{L}^j ,$$

$$j = 1, \ldots , p ,$$

The matrix of coefficients of $\hat{\sigma}_0^{(i)}, \hat{\sigma}_1^{(i)}, \dots, \hat{\sigma}_q^{(i)}, \hat{\beta}_1^{(i)}, \dots, \hat{\beta}_p^{(i)}$ is

$$(4.22)\qquad \begin{pmatrix} \operatorname{tr} (\hat{\Sigma}_{i-1}^u)^{-1} G_f (\hat{\Sigma}_{i-1}^u)^{-1} G_h & -2 \operatorname{tr} G_f (\hat{\Sigma}_{i-1}^u)^{-1} L^\ell \hat{B}_{i-1}^{-1} \\ -2 \operatorname{tr} G_h (\hat{\Sigma}_{i-1}^u)^{-1} L^j \hat{B}_{i-1}^{-1} & 2 \operatorname{tr} \hat{B}_{i-1}^{-1} \hat{\Sigma}_{i-1}^u \hat{B}_{i-1}'^{-1} L'^\ell (\hat{\Sigma}_{i-1}^u)^{-1} L^j \end{pmatrix} .$$

If $C = yy'$, the right-hand side of (4.21) is

$$(4.23)\qquad -2 (\hat{B}_{i-1} y)' (\hat{\Sigma}_{i-1}^u)^{-1} L^j y ,$$

and the quadratic form on the right-hand side of (4.19) is

$$(4.24)\qquad (\hat{B}_{i-1} y)' (\hat{\Sigma}_{i-1}^u)^{-1} G_f (\hat{\Sigma}_{i-1}^u)^{-1} \hat{B}_{i-1} y .$$

When Σ^u is to represent the covariance matrix of a moving average process, $G_0 = I$,

$$(4.25)\qquad G_g = L^g + L'^g , \qquad g = 1, \dots, q ,$$

and

$$(4.26)\qquad \sigma_g = \sigma^2 \sum_{j=1}^{q-g} \alpha_j \alpha_{j+g} , \qquad g = 1, \dots, q .$$

Since $L^g L'^h$ is L^{g-h}, $h \leq g$, except for at most h 1's being replaced by 0's, $\hat{\Sigma}_{i-1}^u$ and $\hat{B}_{i-1}'^{-1} L'^\ell$ almost commute and the lower right-hand corner of (4.22) is approximately

$$(4.27)\qquad 2 \operatorname{tr} \hat{B}_{i-1}'^{-1} L'^\ell L^j \hat{B}_{i-1}^{-1} .$$

5. Estimation of Coefficients of Linear Transformation; Autoregressive Processes with Moving Average Residuals

Here we combine Sections 2 and 3. Let

$$\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \underset{\sim}{y} = \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \underset{\sim}{v} , \tag{5.1}$$

where $\underset{\sim}{K}_0, \underset{\sim}{K}_1, \ldots , \underset{\sim}{K}_p$ are $p + 1$ known linearly independent $T \times T$ matrices, $\underset{\sim}{J}_0, \underset{\sim}{J}_1, \ldots, \underset{\sim}{J}_q$ are $q + 1$ known linearly independent matrices, $\beta_0 = \alpha_0 = 1$, $\beta_1, \ldots , \beta_p , \alpha_1, \ldots , \alpha_q$ are $p + q$ parameters, and $\underset{\sim}{v}$ is a T-component random vector with mean vector $\mathscr{E} \underset{\sim}{v} = \underset{\sim}{0}$ and covariance matrix $\mathscr{C}(\underset{\sim}{v}) = \sigma^2 \underset{\sim}{I}$. Then

$$\underset{\sim}{y} = \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right)^{-1} \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \underset{\sim}{v} \tag{5.2}$$

has mean vector $\underset{\sim}{0}$ and covariance matrix

$$\mathscr{C}(y) = \sigma^2 \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right)^{-1} \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}'_k \left(\sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}'_\ell \right)^{-1} = \sigma^2 \underset{\sim}{B}^{-1} \underset{\sim}{A} \underset{\sim}{A}' \underset{\sim}{B}'^{-1} , \tag{5.3}$$

where $\underset{\sim}{A} = \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k$ and $\underset{\sim}{B} = \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell$.

If $\underset{\sim}{y}_1, \ldots , \underset{\sim}{y}_N$ are N observations on $\underset{\sim}{y}$ with a normal distribution, $2/N$ times the logarithm of the likelihood function $\underset{\sim}{L}$ is

$$\begin{aligned} \frac{2}{N} \log L = {} & - T \log 2\pi - T \log \sigma^2 + 2 \log \left| \sum_{\ell=0}^{p} \beta_\ell \underset{\sim}{K}_\ell \right| - 2 \log \left| \sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \right| \\ & - \operatorname{tr} \frac{1}{\sigma^2} \sum_{\ell', \ell=0}^{p} \beta_{\ell'} \beta_\ell \underset{\sim}{K}'_{\ell'} \left(\sum_{k=0}^{q} \alpha_k \underset{\sim}{J}'_k \right)^{-1} \left(\sum_{k=0}^{q} \alpha_k \underset{\sim}{J}_k \right)^{-1} \underset{\sim}{K}_\ell \underset{\sim}{C} . \end{aligned} \tag{5.4}$$

The partial derivatives are

$$(5.5)\quad \frac{\partial}{\partial\alpha_g}\frac{2}{N}\log L = -2\,\mathrm{tr}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{J}_g$$

$$+\frac{2}{\sigma^2}\,\mathrm{tr}\sum_{\ell,\ell'=0}^{p}\beta_{\ell'}\beta_\ell\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{K}_\ell \underset{\sim}{C}\underset{\sim}{K}'_{\ell'}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}'_k\right)^{-1}\underset{\sim}{J}'_g\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}'_k\right)^{-1}$$

$$= -2\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{J}_g + \frac{2}{\sigma^2}\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{B}'\underset{\sim}{A}'^{-1}\underset{\sim}{J}'_g\underset{\sim}{A}'^{-1},\quad g = 1,\ldots,q,$$

$$(5.6)\quad \frac{\partial}{\partial\beta_h}\frac{2}{N}\log L = 2\,\mathrm{tr}\left(\sum_{\ell=0}^{p}\beta_\ell \underset{\sim}{K}_\ell\right)^{-1}\underset{\sim}{K}_h$$

$$-\frac{2}{\sigma^2}\sum_{\ell=0}^{p}\beta_\ell \underset{\sim}{K}'_\ell\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}'_k\right)^{-1}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{K}_h \underset{\sim}{C}$$

$$= 2\,\mathrm{tr}\,\underset{\sim}{B}^{-1}\underset{\sim}{K}_h - \frac{2}{\sigma^2}\,\mathrm{tr}\,\underset{\sim}{B}'\underset{\sim}{A}'^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{C},\quad h = 1,\ldots,p,$$

$$(5.7)\quad \frac{\partial}{\partial\sigma^2}\frac{2}{N}\log L = -\frac{T}{\sigma^2}+\frac{1}{\sigma^4}\,\mathrm{tr}\sum_{\ell',\ell=0}^{p}\beta_{\ell'}\beta_\ell \underset{\sim}{K}'_{\ell'}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}'_k\right)^{-1}\left(\sum_{k=0}^{q}\alpha_k \underset{\sim}{J}_k\right)^{-1}\underset{\sim}{K}_\ell \underset{\sim}{C}$$

$$= -\frac{T}{\sigma^2}+\frac{1}{\sigma^4}\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{B}'\underset{\sim}{A}'^{-1}.$$

The maximum likelihood estimates are defined by setting the derivatives equal to 0.

The second partial derivatives of $(2/N)\log L$ are

$$(5.8)\quad \frac{\partial^2}{\partial\alpha_g\,\partial\alpha_f}\frac{2}{N}\log L = 2\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{J}_g\underset{\sim}{A}^{-1}\underset{\sim}{J}_f - \frac{2}{\sigma^2}\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{J}_f\underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{B}'\underset{\sim}{A}'^{-1}\underset{\sim}{J}'_g\underset{\sim}{A}'^{-1}$$

$$-\frac{2}{\sigma^2}\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{B}'\underset{\sim}{A}'^{-1}\underset{\sim}{J}'_f\underset{\sim}{A}'^{-1}\underset{\sim}{J}'_g\underset{\sim}{A}'^{-1}$$

$$-\frac{2}{\sigma^2}\,\mathrm{tr}\,\underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{C}\underset{\sim}{B}'\underset{\sim}{A}'^{-1}\underset{\sim}{J}'_g\underset{\sim}{A}'^{-1}\underset{\sim}{J}'_f\underset{\sim}{A}'^{-1},$$

$$g, f = 1,\ldots,q,$$

$$(5.9)\quad \frac{\partial^2}{\partial\alpha_g\,\partial\beta_h}\ \frac{2}{N}\log L = \frac{2}{\sigma^2}\operatorname{tr} \underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{C}\,\underset{\sim}{B}'\,\underset{\sim}{A}'^{-1}\underset{\sim}{J}_g'\,\underset{\sim}{A}'^{-1}$$

$$+\frac{2}{\sigma^2}\operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{J}_g\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{C}\,\underset{\sim}{B}'\,\underset{\sim}{A}'^{-1},$$

$$g = 1, \ldots, q,\ h = 1, \ldots, p,$$

$$(5.10)\quad \frac{\partial^2}{\partial\beta_h\,\partial\beta_j}\ \frac{2}{N}\log L = -\ 2\operatorname{tr}\underset{\sim}{B}^{-1}\underset{\sim}{K}_j\underset{\sim}{B}^{-1}\underset{\sim}{K}_h - \frac{2}{\sigma^2}\operatorname{tr}\underset{\sim}{K}_j'\,\underset{\sim}{A}'^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{C},$$

$$h, j = 1, \ldots, p,$$

$$(5.11)\quad \frac{\partial^2}{\partial\alpha_g\,\partial\sigma^2}\ \frac{2}{N}\log L = -\frac{2}{\sigma^4}\operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{B}\,\underset{\sim}{C}\,\underset{\sim}{B}'\,\underset{\sim}{A}'^{-1}\underset{\sim}{J}_g'\,\underset{\sim}{A}'^{-1},\quad g = 1, \ldots, q,$$

$$(5.12)\quad \frac{\partial^2}{\partial\beta_h\,\partial\sigma^2}\ \frac{2}{N}\log L = \frac{2}{\sigma^4}\operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{C}\,\underset{\sim}{B}'\,\underset{\sim}{A}'^{-1},\qquad h = 1, \ldots, p,$$

$$(5.13)\quad \frac{\partial^2}{\partial(\sigma^2)^2}\ \frac{2}{N}\log L = \frac{T}{\sigma^4} - \frac{2}{\sigma^6}\operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{B}\,\underset{\sim}{C}\,\underset{\sim}{B}'\,\underset{\sim}{A}'^{-1}.$$

The elements of the information matrix are N times

$$(5.14)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\,\partial\alpha_f}\ \frac{1}{N}\log L = \operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{J}_g\underset{\sim}{A}^{-1}\underset{\sim}{J}_f + \operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{J}_g\underset{\sim}{J}_f'\,\underset{\sim}{A}'^{-1},$$

$$g, f = 1, \ldots, q,$$

$$(5.15)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\,\partial\beta_h}\ \frac{1}{N}\log L = -\operatorname{tr}\underset{\sim}{J}_g'\,\underset{\sim}{A}'^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{B}^{-1}\underset{\sim}{A} - \operatorname{tr}\underset{\sim}{J}_g\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{B}^{-1},$$

$$g = 1, \ldots, q,\ h = 1, \ldots, p,$$

$$(5.16)\quad -\mathcal{E}\frac{\partial^2}{\partial\beta_h\,\partial\beta_j}\ \frac{1}{N}\log L = \operatorname{tr}\underset{\sim}{B}^{-1}\underset{\sim}{K}_j\underset{\sim}{B}^{-1}\underset{\sim}{K}_h + \operatorname{tr}\underset{\sim}{K}_j'\,\underset{\sim}{A}'^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{K}_h\underset{\sim}{B}^{-1}\underset{\sim}{A}\,\underset{\sim}{A}'\,\underset{\sim}{B}'^{-1},$$

$$h, j = 1, \ldots, p,$$

$$(5.17)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\ \partial\sigma^2}\ \frac{1}{N}\log L = \frac{1}{\sigma^2}\operatorname{tr} \mathbf{J}_g'\ \mathbf{A}'^{-1}\ , \qquad g = 1, \ldots , q\ ,$$

$$(5.18)\quad -\mathcal{E}\frac{\partial^2}{\partial\beta_h\ \partial\sigma^2}\ \frac{1}{N}\log L = -\frac{1}{\sigma^2}\operatorname{tr} \mathbf{K}_h\ \mathbf{B}^{-1}\ , \qquad h = 1, \ldots , p\ ,$$

$$(5.19)\quad -\mathcal{E}\frac{\partial^2}{\partial(\sigma^2)^2}\ \frac{1}{N}\log L = \frac{T}{2\sigma^4}\ .$$

The method of scoring can be developed from these results.

If $\mathbf{J}_g = \mathbf{K}_g = \mathbf{L}^g$, then the elements of the information matrix are N times

$$(5.20)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\ \partial\alpha_f}\ \frac{1}{N}\log L = \operatorname{tr} \mathbf{A}^{-1}\ \mathbf{L}^g\ \mathbf{L}'^f\ \mathbf{A}'^{-1}\ , \qquad g\ , f = 1, \ldots , q\ ,$$

$$(5.21)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\ \partial\beta_h}\ \frac{1}{N}\log L = -\operatorname{tr} \mathbf{A}^{-1}\ \mathbf{L}^h\ \mathbf{B}^{-1}\ \mathbf{A}\ \mathbf{L}'^g\ \mathbf{A}'^{-1}\ ,$$

$$g = 1, \ldots , q,\ h = 1, \ldots , p\ ,$$

$$(5.22)\quad -\mathcal{E}\frac{\partial^2}{\partial\beta_j\ \partial\beta_h}\ \frac{1}{N}\log L = \operatorname{tr} \mathbf{A}^{-1}\ \mathbf{L}^h\ \mathbf{B}^{-1}\ \mathbf{A}\ \mathbf{A}'\ \mathbf{B}'^{-1}\ \mathbf{L}'^j\ \mathbf{A}'^{-1}\ ,$$

$$h,\ j = 1, \ldots , p\ ,$$

$$(5.23)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\ \partial\sigma^2}\ \frac{1}{N}\log L = 0\ , \qquad g = 1, \ldots , q\ ,$$

$$(5.24)\quad -\mathcal{E}\frac{\partial}{\partial\beta_h\ \partial\sigma^2}\ \frac{1}{N}\log L = 0\ , \qquad h = 1, \ldots , p\ .$$

Note that $\mathbf{L}^g$, $g = 0, 1, \ldots$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}^{-1}$, and $\mathbf{B}^{-1}$ are polynomials in $\mathbf{L}$ and hence commute. Thus (5.21) and (5.22) are

$$(5.25)\quad -\mathcal{E}\frac{\partial^2}{\partial\alpha_g\ \partial\beta_h}\ \frac{1}{N}\log L = -\operatorname{tr} \underset{\sim}{L}^h\ \underset{\sim}{B}^{-1}\ \underset{\sim}{L}'^g\ \underset{\sim}{A}'^{-1}\ ,$$

$$g = 1, \ldots , q\ ,\quad h = 1, \ldots , p\ ,$$

$$(5.26)\quad -\mathcal{E}\frac{\partial^2}{\partial\beta_h\ \partial\beta_j}\ \frac{1}{N}\log L = \operatorname{tr} \underset{\sim}{B}'^{-1}\ \underset{\sim}{L}'^j\ \underset{\sim}{L}^h\ \underset{\sim}{B}^{-1}\ ,\quad h, j = 1, \ldots , p\ .$$

When $\underset{\sim}{J}_g = \underset{\sim}{K}_g = \underset{\sim}{L}^g$, then the method of scoring involves the solution of

$$(5.27)\quad \sum_{f=1}^{q} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1}\ \underset{\sim}{L}^g\ \underset{\sim}{L}'^f\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1}\ \hat{\alpha}_f^{(i)}$$

$$-\sum_{h=1}^{p} \operatorname{tr} \underset{\sim}{L}'^g\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1}\ \underset{\sim}{L}^h\ \hat{\underset{\sim}{B}}{}_{i-1}^{-1}\ \hat{\beta}_h^{(i)}$$

$$= \frac{1}{\hat{\sigma}_{i-1}^2} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1}\ \hat{\underset{\sim}{B}}_{i-1}\ \underset{\sim}{C}\ \hat{\underset{\sim}{B}}{}_{i-1}'\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1}\ \underset{\sim}{L}'^g\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} + \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1}\ \underset{\sim}{L}^g\ (\hat{\underset{\sim}{B}}{}_{i-1}'^{-1} - \hat{\underset{\sim}{A}}{}_{i-1}'^{-1})\ ,$$

$$g = 1, \ldots , q\ ,$$

$$(5.28)\quad -\sum_{f=1}^{q} \operatorname{tr} \underset{\sim}{L}^j\ \hat{\underset{\sim}{B}}{}_{i-1}^{-1}\ \underset{\sim}{L}'^f\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1}\ \hat{\alpha}_f^{(i)}$$

$$+\sum_{h=1}^{p} \operatorname{tr} \hat{\underset{\sim}{B}}{}_{i-1}'^{-1}\ \underset{\sim}{L}'^j\ \underset{\sim}{L}^h\ \hat{\underset{\sim}{B}}{}_{i-1}^{-1}\ \hat{\beta}_h^{(i)}$$

$$= -\frac{1}{\hat{\sigma}_{i-1}^2} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1}\ \underset{\sim}{L}^j\ \underset{\sim}{C}\ \hat{\underset{\sim}{B}}{}_{i-1}'\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1} + \operatorname{tr} (\hat{\underset{\sim}{A}}{}_{i-1}'^{-1} - \hat{\underset{\sim}{B}}{}_{i-1}'^{-1})\ \underset{\sim}{L}^j\ \hat{\underset{\sim}{B}}{}_{i-1}^{-1}\ ,$$

$$j = 1, \ldots , p\ ,$$

$$(5.29)\quad \hat{\sigma}_i^2 = \frac{1}{T} \operatorname{tr} \hat{\underset{\sim}{A}}{}_{i-1}^{-1}\ \hat{\underset{\sim}{B}}_{i-1}\ \underset{\sim}{C}\ \hat{\underset{\sim}{B}}{}_{i-1}'\ \hat{\underset{\sim}{A}}{}_{i-1}'^{-1}\ .$$

If $N = 1$, $\underset{\sim}{y}_1 = \underset{\sim}{y}$, and $\underset{\sim}{C} = \underset{\sim}{y}\underset{\sim}{y}$, the right-hand sides of (5.27), (5.28), and (5.29) are, respectively,

$$(5.30)\quad \frac{1}{\hat{\sigma}^2_{i-1}} \left(\hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \hat{\underset{\sim}{B}}_{i-1}\, \underset{\sim}{y}\right)' \hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \underset{\sim}{L}^{g} \left(\hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \hat{\underset{\sim}{B}}_{i-1}\, \underset{\sim}{y}\right) + \operatorname{tr} \hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \underset{\sim}{L}^{g} \left(\hat{\underset{\sim}{B}}{}'^{-1}_{i-1} - \hat{\underset{\sim}{A}}{}'^{-1}_{i-1}\right),$$

$$g = 1, \ldots, q,$$

$$(5.31)\quad -\frac{1}{\hat{\sigma}^2_{i-1}} \left(\hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \hat{\underset{\sim}{B}}_{i-1}\, \underset{\sim}{y}\right)' \underset{\sim}{L}_{j} \left(\hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \underset{\sim}{y}\right) + \operatorname{tr} \left(\hat{\underset{\sim}{A}}{}'^{-1}_{i-1} - \hat{\underset{\sim}{B}}{}'^{-1}_{i-1}\right) \underset{\sim}{L}^{j}\, \hat{\underset{\sim}{B}}{}^{-1}_{i-1},$$

$$j = 1, \ldots, p,$$

$$(5.32)\quad \frac{1}{T} \left(\hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \hat{\underset{\sim}{B}}_{i-1}\, \underset{\sim}{y}\right)' \left(\hat{\underset{\sim}{A}}{}^{-1}_{i-1}\, \hat{\underset{\sim}{B}}_{i-1}\, \underset{\sim}{y}\right).$$

6. Asymptotic Theory

The exact distributions of the maximum likelihood estimates developed in this paper cannot be obtained in closed form in general. However, asymptotic distributions can be found. If $N \to \infty$ we have the case of repeated observations on the random vector $\underset{\sim}{y}$; in the case of time series, however, N may be 1 and $T \to \infty$. In either case when consistent estimates of the parameters are used as initial estimates, the estimates obtained in the first step of the iteration procedure are consistent, asymptotically normal, and asymptotically efficient (when normalized by $\sqrt{N}$ or $\sqrt{T}$, as the case may be).

In the model of Section 2.1 no iteration is involved and the asymptotic properties are the usual ones as the number of observations N increases. The model of Section 2.2 is the autoregressive model with the first p observations treated as fixed $(y_{-p+1} = \dots = y_0 = 0)$; the asymptotic theory as $T \to \infty$ is well known. [See T. W. Anderson (1971), Section 5.5, for example.]

For each of the models in the other sections [as well as the model $\underset{\sim}{\Sigma} = \Sigma_{g=0}^{q} \sigma_g \underset{\sim}{G}_g$ treated in T. W. Anderson (1971b), (1973)] an iterative procedure was proposed. If the initial estimates are consistent, the matrix of coefficients of the linear equations is a consistent estimate of the information matrix of one observation. The asymptotic distribution of the right-hand sides is normal with covariance matrix equal to this matrix. It then follows that the estimates have the stated properties. We shall carry out the details of the proof only for the model of Section 3.2, which shows the pattern.

Let $\underset{\sim}{y} = (y_1, \dots, y_T)'$ be defined by

$$(6.1) \qquad \underset{\sim}{y} = \sum_{k=0}^{q} \alpha_k \underset{\sim}{L}^k \underset{\sim}{v} = \underset{\sim}{A} \underset{\sim}{v} .$$

We shall let $T \to \infty$. We assume that the roots of (3.58) are less than 1 in absolute value. Then (3.44) and (3.45) for $i = 1$ are

$$(6.2) \qquad \sum_{j=1}^{q} \operatorname{tr} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \underset{\sim}{L}'^j \hat{\underset{\sim}{A}}_0'^{-1} \hat{\alpha}_j^{(1)} = \frac{1}{\hat{\sigma}_0^2} \underset{\sim}{y}' \hat{\underset{\sim}{A}}_0'^{-1} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{y} - \operatorname{tr} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \hat{\underset{\sim}{A}}_0'^{-1} , \qquad g = 1, \ldots, q ,$$

$$(6.3) \qquad \hat{\sigma}_1^2 = \frac{1}{T} \underset{\sim}{y}' \hat{\underset{\sim}{A}}_0'^{-1} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{y} .$$

We shall show that

$$(6.4) \qquad \operatorname*{plim}_{T\to\infty} \frac{1}{T} \operatorname{tr} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \underset{\sim}{L}'^j \hat{\underset{\sim}{A}}_0'^{-1} = \lim_{T\to\infty} \frac{1}{T} \operatorname{tr} \underset{\sim}{A}^{-1} \underset{\sim}{L}^g \underset{\sim}{L}'^j \underset{\sim}{A}'^{-1} .$$

The right-hand side is given by (3.67). The left-hand side is

$$(6.5) \qquad \sum_{i=0}^{T-1-\max(g,j)} \left[1 - \frac{i + \max(g,j)}{T} \right] \hat{\delta}^0_{i+|j-g|} \hat{\delta}^0_i ,$$

where $\hat{\delta}_0^0 = 1$, $\hat{\delta}_i^0$, $i = 1, \ldots$, constitute the solutions to (3.55) and (3.56) with α_ℓ replaced by $\hat{\alpha}_\ell^0$, $\ell = 1, \ldots, q$. With arbitrarily high probability $\hat{\alpha}_1^0, \ldots, \hat{\alpha}_q^0$ are such that the roots of the polynomial equation with these coefficients are less than 1 in absolute value, in fact, are less than $\rho < 1$ for some ρ [greater than the largest root of (3.58)]. Then (6.5) converges in probability to

$$(6.6) \qquad \sum_{i=0}^{\infty} \delta_{i+|g-j|} \delta_i = \frac{\sigma_{AR}(g-j)}{\sigma^2} .$$

We can write (6.2) as

$$(6.7)\quad \sum_{j=1}^{q} \frac{1}{T} \operatorname{tr} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \underset{\sim}{L}'^j \hat{\underset{\sim}{A}}_0'^{-1} \sqrt{T}\,(\hat{\alpha}_j^{(1)} - \alpha_j)$$

$$= \frac{1}{\sqrt{T}} \left[\frac{1}{\hat{\sigma}_0^2} \underset{\sim}{y}' \hat{\underset{\sim}{A}}_0'^{-1} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{y} - \operatorname{tr} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \underset{\sim}{A}' \hat{\underset{\sim}{A}}_0'^{-1} \right], \qquad g = 1, \ldots, q .$$

We want to show that the right-hand sides have a limiting normal distribution with means 0 and covariance matrix (6.4).

Consider

$$(6.8)\quad \frac{1}{\sqrt{T}\sigma^2} \underset{\sim}{y}'\underset{\sim}{A}'^{-1} \underset{\sim}{A}^{-1} \underset{\sim}{L}^g \underset{\sim}{A}^{-1} \underset{\sim}{y} = \frac{1}{\sqrt{T}\sigma^2} \underset{\sim}{v}' \underset{\sim}{A}^{-1} \underset{\sim}{L}^g \underset{\sim}{v}$$

$$= \sum_{i=0}^{\infty} \delta_i \frac{1}{\sqrt{T}\sigma^2} \underset{\sim}{v}' \underset{\sim}{L}^{i+g} \underset{\sim}{v}$$

$$= \sum_{i=0}^{T-g-1} \delta_i \frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^{T-(i+g)} v_t v_{t+i+g} .$$

For any n the set $(1/\sqrt{T})\Sigma_{t=1}^{T} v_t v_{t+1}, \ldots, (1/\sqrt{T})\Sigma_{t=1}^{T} v_t v_{t+n}$ have a limiting normal distribution [Theorem 7.7.6 of T. W. Anderson (1971a), for example] with means 0 and covariances

$$(6.9)\quad \frac{1}{T} \mathcal{E} \sum_{t,s=1}^{T} v_t v_{t+j} v_s v_{s+h} = \frac{1}{T} \mathcal{E} \sum_{t=1}^{T} v_t^2 v_{t+j} v_{t+h}$$

$$= \sigma^4 , \qquad j = h = 1, \ldots ,$$

$$= 0 , \qquad j \neq h .$$

Then the set

$$(6.10)\quad \sum_{i=0}^{n-q} \delta_i \frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^{T} v_t v_{t+i+g} , \qquad g = 1, \ldots , q ,$$

has a limiting normal distribution with means 0 and covariances

$$(6.11) \quad \frac{1}{\sigma^4} \sum_{i,j=0}^{n-q} \delta_i \delta_j \frac{1}{T} \mathcal{E} \sum_{t,s=1}^{T} v_t v_{t+i+g} v_s v_{s+j+h} = \sum_{i=0}^{n-q-|g-h|} \delta_i \delta_{i+|g-h|} ,$$

which has the limit as $n \to \infty$ of (6.6). That the limiting distribution of (6.8) is the limit as $n \to \infty$ of the limiting distribution of (6.10) is justified by Corollary 7.7.1 of T. W. Anderson (1971a), for example. Note that

$$(6.12) \quad \mathcal{E} \left(\sum_{i=n-q+1}^{T-g-1} \delta_i \frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^{T-(i+g)} v_t v_{t+i+g} \right)^2 \leq \sum_{i=n-q+1}^{T-g-1} \delta_i^2 \leq \sum_{i=n-q+1}^{\infty} \delta_i^2 .$$

Now consider the difference of (6.8) and (6.7), which is

$$(6.13) \quad \frac{1}{\sqrt{T}} \left[\frac{1}{\sigma^2} \underset{\sim}{v}' \underset{\sim}{A}^{-1} \underset{\sim}{L}^g \underset{\sim}{v} - \frac{1}{\hat{\sigma}_0^2} \underset{\sim}{v}' \underset{\sim}{A}' \hat{\underset{\sim}{A}}_0'^{-1} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{A} \underset{\sim}{v} + \operatorname{tr} \hat{\underset{\sim}{A}}_0^{-1} \underset{\sim}{L}^g \underset{\sim}{A}' \hat{\underset{\sim}{A}}_0'^{-1} \right]$$

We write

$$(6.14) \quad \hat{\underset{\sim}{A}}_0^{-1} = \underset{\sim}{A}^{-1} - \hat{\underset{\sim}{A}}_0^{-1} (\hat{\underset{\sim}{A}}_0 - \underset{\sim}{A}) \underset{\sim}{A}^{-1} .$$

Then (6.13) is

$$(6.15)\quad \frac{1}{\sqrt{T}}\left\{\frac{1}{\sigma^2}\underset{\sim}{v}'\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\underset{\sim}{v}\right.$$

$$-\frac{1}{\hat{\sigma}_0^2}\left[\underset{\sim}{v}'\underset{\sim}{A}'(\underset{\sim}{A}'^{-1}-\underset{\sim}{A}'^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1})(\underset{\sim}{A}^{-1}-\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1})\underset{\sim}{L}^{g}(\underset{\sim}{A}^{-1}-\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1})\underset{\sim}{A}\underset{\sim}{v}\right]$$

$$\left.+\operatorname{tr}(\underset{\sim}{A}^{-1}-\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1})\underset{\sim}{L}^{g}\underset{\sim}{A}'(\underset{\sim}{A}'^{-1}-\underset{\sim}{A}'^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1})\right\}$$

$$=\frac{1}{\sqrt{T}}\left\{\left(\frac{1}{\sigma^2}-\frac{1}{\hat{\sigma}_0^2}\right)\underset{\sim}{v}'\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\underset{\sim}{v}+\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\underset{\sim}{v}\right.$$

$$+\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\underset{\sim}{v}+\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{v}-\operatorname{tr}\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1}$$

$$-\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\underset{\sim}{v}-\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{v}$$

$$-\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{v}$$

$$+\frac{1}{\hat{\sigma}_0^2}\underset{\sim}{v}'(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{v}$$

$$\left.+\operatorname{tr}\hat{\underset{\sim}{A}}_0^{-1}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1}\underset{\sim}{L}^{g}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})'\hat{\underset{\sim}{A}}_0'^{-1}\right\}.$$

The first term on the right-hand side of (6.15) has probability limit 0 because (6.8) has a limiting normal distribution and $\operatorname{p\,lim}_{T\to\infty}\hat{\sigma}_0^2=\sigma^2>0$. Each of the third and fourth terms are

$$(6.16)\quad \frac{1}{\hat{\sigma}_0^2}\frac{1}{\sqrt{T}}\underset{\sim}{v}'\hat{\underset{\sim}{A}}_0^{-1}\underset{\sim}{L}^{g}(\hat{\underset{\sim}{A}}_0-\underset{\sim}{A})\underset{\sim}{A}^{-1}\underset{\sim}{v}=\frac{1}{\hat{\sigma}_0^2}\sum_{k=1}^{q}(\hat{\alpha}_k^{(0)}-\alpha_k)\sum_{i,j=0}^{\infty}\hat{\delta}_i^0\delta_j\frac{1}{\sqrt{T}}\underset{\sim}{v}'\underset{\sim}{L}^{g+i+j+k}\underset{\sim}{v}\ .$$

Let

$$(6.17)\qquad W_{hT}=\sum_{j=0}^{\infty}\delta_j\frac{1}{\sqrt{T}}\underset{\sim}{v}'\underset{\sim}{L}^{h+j}\underset{\sim}{v}\ .$$

Then

$$(6.18)\qquad \mathcal{E}W_{hT}^2\le\sigma^4\sum_{j=0}^{\infty}\delta_j^2$$

We can write

$$(6.19) \qquad \sum_{i,j=0}^{\infty} \hat{\delta}_i^0 \, \delta_j \frac{1}{\sqrt{T}} \underset{\sim}{v}' \underset{\sim}{L}^{g+h+i+j} \underset{\sim}{v} = \sum_{i=0}^{\infty} \hat{\delta}_i^0 \, W_{g+k+i,T} \, .$$

With arbitrarily high probability $|\hat{\delta}_i^0| < \rho_0^i$ for some ρ_0 such that $0 < \rho_0 < \rho_1 < 1$. Then the square of (6.19) is less than

$$(6.20) \qquad \sum_{i=0}^{\infty} \left(\frac{\hat{\delta}_i^0}{\rho_1^i} \right)^2 \sum_{i=0}^{\infty} \rho_1^{2i} W_{g+k+i,T}^2 \, .$$

Since the expected value of the second sum is less than $\sigma^4 \sum_{j=0}^{\infty} \delta_j^2/(1-\rho_1^2)$, (6.20) is bounded in probability. Since $p \lim_{T\to\infty} \hat{\alpha}_k^{(0)} = \alpha_k$, (6.16) has probability limit 0. The second term and fifth term give

$$(6.21) \qquad \frac{1}{\hat{\sigma}_0^2} \frac{1}{\sqrt{T}} \underset{\sim}{v}' (\hat{\underset{\sim}{A}}_0 - \underset{\sim}{A})' \hat{\underset{\sim}{A}}_0'^{-1} \underset{\sim}{A}^{-1} \underset{\sim}{L}^g \underset{\sim}{v} - \frac{1}{\sqrt{T}} \operatorname{tr} (\hat{\underset{\sim}{A}}_0 - \underset{\sim}{A})' \hat{\underset{\sim}{A}}_0'^{-1} \underset{\sim}{A}^{-1} \underset{\sim}{L}^g$$

$$= \frac{1}{\hat{\sigma}_0^2} \sum_{k=1}^{q} (\hat{\alpha}_k^{(0)} - \alpha_k) \Bigg[\sum_{i,j=0}^{\infty} \hat{\delta}_i^0 \, \delta_j \frac{1}{\sqrt{T}} \left(\underset{\sim}{v}' \underset{\sim}{L}'^{k+i} \underset{\sim}{L}^{g+j} \underset{\sim}{v} - \sigma^2 \operatorname{tr} \underset{\sim}{L}'^{k+i} \underset{\sim}{L}^{g+j} \right)$$

$$+ \frac{1}{\sqrt{T}} \sum_{i,j=0}^{\infty} \hat{\delta}_i^0 \, \delta_j \, (\sigma^2 - \hat{\sigma}_0^2) \operatorname{tr} \underset{\sim}{L}'^{k+i} \underset{\sim}{L}^{g+j} \Bigg] \, .$$

The sum of δ_j times the first parenthesis is treated like (6.17); note that the parenthesis has mean 0 and (6.18) as a bound on the expected value of its square. The same argument carries through. If $\sqrt{T}(\hat{\sigma}^2 - \hat{\sigma}_0^2)$ is bounded in probability [or $\sqrt{T}(\hat{\alpha}_k^{(0)} - \alpha_k)$ is], then the second term converges to 0 in probability. The other terms in (6.15) are treated similarly.

It follows from these results that the solutions to (6.7), namely $\sqrt{T}(\hat{\alpha}_1^{(1)} - \alpha_1), \ldots, \sqrt{T}(\hat{\alpha}_q^{(1)} - \alpha_q)$ have a limiting normal distribution with means 0 and a covariance matrix that is the inverse of the information matrix.

The sample covariances c_h defined for (2.48) are consistent estimates of $\sigma(h)$, $h = 0, 1, \ldots, p+q$. From these can be obtained consistent estimates of $\beta_1, \ldots, \beta_p$, $\sigma_u(0), \ldots, \sigma_u(q)$ and of $\beta_1, \ldots, \beta_p$, $\alpha_1, \ldots, \alpha_q$, and σ^2 as described in Section 5.8.1.

REFERENCES

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., New York.

Anderson, T. W. (1969), "Statistical inference for covariance matrices with linear structure," Proceedings of the Second International Symposium on Multivariate Analysis (P. R. Krishnaiah, ed.), Academic Press, Inc., New York, 1969, pp. 55-56

Anderson, T. W. (1970), "Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices," Essays in Probability and Statistics, University of North Carolina Press, Chapel Hill, 1970, pp. 1-24. Also Technical Report No. 3, 1966, Contract AF 41(609) -2653, Teacher's College, Columbia University.

Anderson, T. W. (1971a), The Statistical Analysis of Time Series, John Wiley & Sons, Inc., New York.

Anderson, T. W. (1971b), "Estimation of covariance matrices with linear structure and moving average processes of finite order," Technical Report No. 6, October 29, 1971, Contract N00014-67-A-0112-0030, Stanford University, Stanford, California.

Anderson, T. W. (1971c), "The stationarity of an estimated autoregressive process," Technical Report No. 7, November 15, 1971, Contract N00014-67-A-0112-0030, Stanford University, Stanford, California.

Anderson, T. W. (1973), "Asymptotically efficient estimation of covariance matrices with linear structure," Annals of Statistics, **1**, 135-141.

Box, George E. P., and Gwilym M. Jenkins (1970), Time Series Analysis Forecasting and Control, Holden-Day Inc., San Francisco.

Clevenson, M. Lawrence (1970), "Asymptotically efficient estimates of the parameters of a moving average time series," Stanford University, Statistics Department.

Durbin, J. (1959), "Efficient estimation of parameters in moving-average models," J. Roy. Statist. Soc. Ser. B, **22**, pp. 139-153.

Durbin, J. (1960), "The fitting of time-series models," Review of the International Statistical Institute, **28**, 233-244.

Hannan, E. J. (1969), "The estimation of mixed moving average autoregressive systems," Biometrika, **56**, pp. 579-594.

Parzen, Emanuel (1971), "Efficient estimation of stationary time series mixed schemes," Technical Report No. 16, Contract NONR-225 80, Stanford University, Stanford, California.

Walker, A. M. (1961), "Large-sample estimation for moving-average models," Biometrika, **48**, 343-357.

Walker, A. M. (1962), "Large sample estimation of parameters for autoregressive processes with moving-average residuals," Biometrika, **49**, pp. 117-131.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 15	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): T. W. Anderson	8. CONTRACT OR GRANT NUMBER(s): N00014-67-A-0112-0030
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Stanford University, Stanford, California 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-034
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics & Probability Program Code 436, ARlington, Virginia 22217	12. REPORT DATE: December 1973
	13. NUMBER OF PAGES: 48
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

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

Maximum likelihood estimation, autoregressive process with moving average residuals, covariance matrices with linear structure, estimation of linear transformations, multivariate normal distribution.

20. Abstract (see other side)

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)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

The autoregressive process with moving average residuals is a stationary process $\{y_t\}$ satisfying $\sum_{s=0}^{p} \beta_s y_{t-s} = \sum_{j=0}^{q} \alpha_j v_{t-j}$, where the sequence $\{v_t\}$ consists of independently identically distributed (unobservable) random variables. The distribution of $y_1, \cdots, y_T$ can be approximated by the distribution of the T-component vector $\underset{\sim}{y}$ satisfying $\sum_{s=0}^{p} \beta_s \underset{\sim}{K}_s \underset{\sim}{y} = \sum_{j=0}^{q} \alpha_j \underset{\sim}{J}_j \underset{\sim}{v}$, where $\underset{\sim}{v}$ has covariance matrix $\sigma^2 \underset{\sim}{I}, \underset{\sim}{K}_s = \underset{\sim}{J}_s = \underset{\sim}{L}^s$, and $\underset{\sim}{L}$ is the $T \times T$ matrix with 1's immediately below the main diagonal and 0's elsewhere. Maximum likelihood estimates are obtained when $\underset{\sim}{v}$ has a normal distribution. The method of scoring is used to find estimates defined by linear equations which are consistent, asymptotically normal, and asymptotically efficient (as $T \to \infty$). Several special cases are treated. It is shown how to calculate the estimates.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)