Likelihood Function of Stationary Multiple Autoregressive Moving Average Models

STEVEN C. HILLMER and GEORGE C. TIAO*

Procedures to estimate parameters in multivariate autoregressive moving average (ARMA) models are developed. Gaussian errors are assumed. Exact maximum likelihood estimation procedures are developed for pure moving average models. Approximate procedures are obtained to estimate stationary mixed ARMA models. Properties of the estimates and an example are given.

KEY WORDS: Time series; Multiple autoregressive moving average models; Maximum likelihood estimation; Multiplicative ARMA models; Noninvertible models.

1. INTRODUCTION

This article concerns some aspects of the estimation problem associated with stationary multiple autoregressive moving average (ARMA) models. Let $\{\mathbf{z}_t\}$ be a k-dimensional vector-valued time series. We suppose that $\mathbf{z}_t$ is measured from some vector mean level $\boldsymbol{\mu}$ and follows the model

$$\boldsymbol{\phi}_p(B)\mathbf{z}_t = \boldsymbol{\theta}_q(B)\mathbf{a}_t \ , \tag{1.1}$$

where

$$\boldsymbol{\phi}_p(B) = (\mathbf{I} - \boldsymbol{\phi}_1 B - \ldots - \boldsymbol{\phi}_p B^p) \ ,$$
$$\boldsymbol{\theta}_q(B) = (\mathbf{I} - \boldsymbol{\theta}_1 B - \ldots - \boldsymbol{\theta}_q B^q) \ ,$$

B is the back shift operator such that $B\mathbf{z}_t = \mathbf{z}_{t-1}$, the $\boldsymbol{\phi}_j$'s and the $\boldsymbol{\theta}_j$'s are $k \times k$ unknown parameter matrices, and the $\mathbf{a}_t$'s are $k \times 1$ vectors identically and independently distributed as $N(\mathbf{0}, \boldsymbol{\Sigma})$. For stationarity, we shall require that the zeros of the determinantal polynomial in B, $|\boldsymbol{\phi}_p(B)|$, all lie outside the unit circle. Further, we shall assume that the model is identified; conditions to guarantee this can be found in Hannan (1969).

The problem of parameter estimation for (1.1) has been considered by Wilson (1973), Hannan (1970), Nicholls (1976), Reinsel (1976), Phadke and Kedem (1976), and Osborn (1977). Except for the last two works cited, the estimation procedures proposed are all based on approximations to the model obtained under the assumption that the zeros of $|\boldsymbol{\theta}_q(B)|$ all lie outside the unit circle, that is, that the model is invertible.

In the analysis of time series, it is frequently necessary to difference the data to achieve stationarity. Overdifferencing, however, will lead to models that are not invertible. This situation often occurs in the analysis of seasonal data as illustrated by Abraham (1975). Noninvertible models arising from differencing can also occur during analysis of multiple time series. For example, suppose we have two series $\{z_{1t}\}$ and $\{z_{2t}\}$ such that $z_{1t} = z_{1(t-1)} + a_{1t} - \alpha a_{1(t-1)}$ and $z_{2t} = \beta z_{1t} + a_{2t}$. Thus, individually, both series are nonstationary, and we might be led to consider the first differences $W_{it} = (1 - B)z_{it}$, $i = 1, 2$. The bivariate model for the differenced series is $\mathbf{W}_t = \mathbf{a}_t^* - \boldsymbol{\theta}\mathbf{a}_{t-1}^*$, where $\mathbf{a}_t^* = (a_{1t}, \ \beta a_{1t} + a_{2t})'$ and $\boldsymbol{\theta} = \left(\begin{smallmatrix} \alpha & 0 \\ \beta(\alpha-1) & 1 \end{smallmatrix}\right)$ so that $|\boldsymbol{\theta}(B)| = (1 - \alpha B)(1 - B)$ and we have a noninvertible model.

In Section 2, we obtain the exact likelihood function for the multiple pure moving average model and for a model of a multiplicative form. Section 3 compares properties of the exact likelihood with an approximation based on the assumption of invertibility. Specifically, we show that, in situations in which (1.1) is noninvertible, estimates obtained by maximizing the approximate likelihood function are inferior to estimates obtained by maximizing the exact likelihood. In Section 4, we extend the results to the ARMA model (1.1) and develop an approximation that does not depend on the assumption of invertibility. Section 5 gives an illustrative example using two actual time series.

2. LIKELIHOOD FUNCTION FOR THE MOVING AVERAGE MODEL

2.1 General Considerations

We obtain in this section the exact likelihood function of the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_q)$ and $\boldsymbol{\Sigma}$ for the MA(q) model, that is, $\boldsymbol{\phi}_p(B) = \mathbf{I}$ in (1.1), given n observations $\mathbf{Z}' = (\mathbf{z}'_1, \ldots, \mathbf{z}'_n)$. Let the $(nk) \times (nk)$ matrix $\mathbf{D}_{\lambda,n}$, the $(qk) \times (qk)$ matrices $\mathbf{C}_{\lambda,q}$ and $\mathbf{F}_{\lambda,q}$, as well as the $(nk) \times (qk)$ matrix $\boldsymbol{\Delta}_{n,q}$ be defined as

$$\mathbf{D}_{\lambda,n} = \begin{bmatrix} \mathbf{I} & & & & \mathbf{0} \\ -\boldsymbol{\lambda}_1 & \ddots & & & \\ \vdots & & & & \\ -\boldsymbol{\lambda}_q & & \ddots & & \\ & \ddots & & & \\ \mathbf{0} & -\boldsymbol{\lambda}_q & \ldots & -\boldsymbol{\lambda}_1 & \mathbf{I} \end{bmatrix} ,$$

* Steven C. Hillmer is Assistant Professor, School of Business, University of Kansas, Lawrence, KS 66045. George C. Tiao is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. This research was partially supported by the U.S. Army research office grant DAAG29-76-G-0304. The authors wish to thank an associate editor and the referees for their helpful comments.

© Journal of the American Statistical Association
September 1979, Volume 74, Number 367
Theory and Methods Section

$$\mathbf{C}_{\lambda,q} = \begin{bmatrix} \lambda_q & \lambda_{q-1} & \cdots & \lambda_1 \\ & & \ddots & \vdots \\ & \ddots & & \lambda_{q-1} \\ 0 & & & \lambda_q \end{bmatrix},$$

$$\mathbf{F}_{\lambda,q} = \begin{bmatrix} \mathbf{I} & & & 0 \\ -\lambda_1 & \ddots & & \\ \vdots & \ddots & & \\ -\lambda_{q-1} & \cdots & -\lambda_1 & \mathbf{I} \end{bmatrix}, \quad \mathbf{\Delta}_{n,q} = \begin{bmatrix} \mathbf{I}_{kq} \\ \mathbf{0} \end{bmatrix}, \quad (2.1)$$

where $\lambda_1, \ldots, \lambda_q$ are $k \times k$ matrices. Note that

$$\mathbf{D}_{\lambda,n+q} = \begin{bmatrix} \mathbf{F}_{\lambda,q} & \mathbf{0} \\ -\mathbf{\Delta}_{n,q}\mathbf{C}_{\lambda,q} & \mathbf{D}_{\lambda,n} \end{bmatrix}$$

and

$$\mathbf{D}_{\lambda,n}^{-1} = \begin{bmatrix} \mathbf{I} & & & 0 \\ \mathbf{\Pi}_1(\lambda) & \ddots & & \\ \vdots & \ddots & & \\ \mathbf{\Pi}_{n-1}(\lambda) & \cdots & \mathbf{\Pi}_1(\lambda) & \mathbf{I} \end{bmatrix}, \quad (2.2)$$

where

$$\mathbf{\Pi}_j(\lambda) = \lambda_1\mathbf{\Pi}_{j-1}(\lambda) + \ldots + \lambda_q\mathbf{\Pi}_{j-q}(\lambda) , \quad j = 1, \ldots, n-1 , \quad (2.3)$$

with $\mathbf{\Pi}_0(\lambda) = \mathbf{I}_k$ and, for $q \geq 2$, $\mathbf{\Pi}_{-1}(\lambda) = \ldots = \mathbf{\Pi}_{-q+1}(\lambda) = \mathbf{0}$. Also, we shall let $\mathbf{\Sigma}_n = \mathbf{I}_n \otimes \mathbf{\Sigma}$ where $\otimes$ denotes the Kronecker product. Since the elements of the matrices $\mathbf{D}_{\lambda,n}$, $\mathbf{C}_{\lambda,q}$, and $\mathbf{F}_{\lambda,q}$ in (2.1) are all $k \times k$ matrices, it will be convenient to introduce the terminology *block row* and *block column*. For examples, $[\mathbf{I}, -\lambda'_1, \ldots, -\lambda'_q, \mathbf{0}]'$ will be called the first block column of $\mathbf{D}_{\lambda,n}$, and $[\lambda_q, \ldots, \lambda_1]$ the first block row of $\mathbf{C}_{\lambda,q}$.

2.2 The MA(*q*) Model

To derive the likelihood function of $(\theta, \mathbf{\Sigma})$, we see from (1.1) that the model for the observations $\mathbf{Z}$ is

$$\mathbf{Z} = \mathbf{D}_{\theta,n}\mathbf{a} - \mathbf{\Delta}_{n,q}\mathbf{C}_{\theta,q}\mathbf{a}_* , \quad (2.4)$$

where $\mathbf{a}' = (\mathbf{a}'_1, \ldots, \mathbf{a}'_n)$ and $\mathbf{a}'_* = (\mathbf{a}'_{-q+1}, \ldots, \mathbf{a}'_0)$. Letting $\mathbf{b} = \mathbf{P}\mathbf{a}_*$ where $\mathbf{P}$ is an arbitrarily chosen $(qk) \times (qk)$ nonsingular matrix such that the absolute value of $|\mathbf{P}| = 1$, we can write

$$\begin{bmatrix} \mathbf{a}_* \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^{-1} & \mathbf{0} \\ \mathbf{D}_{\theta,n}^{-1}\mathbf{\Delta}_{n,q}\mathbf{C}_{\theta,q}\mathbf{P}^{-1} & \mathbf{D}_{\theta,n}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{Z} \end{bmatrix} = \mathbf{Y} - \mathbf{X}\mathbf{b} , \quad (2.5)$$

where $\mathbf{Y}' = \mathbf{Z}'[\mathbf{0}, \mathbf{D}_{\theta,n}^{-1\prime}]$ and

$$\mathbf{X}' = -\mathbf{P}^{-1\prime}[\mathbf{I}, (\mathbf{\Delta}_{n,q}\mathbf{C}_{\theta,q})'\mathbf{D}_{\theta,n}^{-1\prime}] .$$

Applying standard least squares theory to integrate out $\mathbf{b}$, we find

$$\ell(\theta, \mathbf{\Sigma}|\mathbf{Z}) \propto |\mathbf{\Sigma}|^{-\frac{1}{2}(n+q)}|\mathbf{X}'\mathbf{\Sigma}_{n+q}^{-1}\mathbf{X}|^{-\frac{1}{2}} \cdot\exp(-\tfrac{1}{2}[\mathbf{Y}'\mathbf{\Sigma}_{n+q}^{-1}\mathbf{Y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{\Sigma}_{n+q}^{-1}\mathbf{X}\hat{\mathbf{b}}]) , \quad (2.6a)$$

where $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{\Sigma}_{n+q}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}_{n+q}^{-1}\mathbf{Y}$. This can be alternatively expressed as

$$\ell(\theta, \mathbf{\Sigma}|\mathbf{Z}) \propto |\mathbf{\Sigma}|^{-\frac{1}{2}(n+q)}|\mathbf{A}_{n,q}|^{-\frac{1}{2}} \cdot\exp(-\tfrac{1}{2}[\mathbf{Z}'\mathbf{H}_n\mathbf{Z} - \hat{\mathbf{a}}'_*\mathbf{A}_{n,q}\hat{\mathbf{a}}_*]) , \quad (2.6b)$$

where $\mathbf{H}_n = \mathbf{D}_{\theta,n}^{-1\prime}\mathbf{\Sigma}_n^{-1}\mathbf{D}_{\theta,n}^{-1}$,

$$\mathbf{A}_{n,q} = \mathbf{\Sigma}_q^{-1} + (\mathbf{\Delta}_{n,q}\mathbf{C}_{\theta,q})'\mathbf{H}_n(\mathbf{\Delta}_{n,q}\mathbf{C}_{\theta,q}) ,$$

and $\hat{\mathbf{a}}_* = -\mathbf{A}_{n,q}^{-1}(\mathbf{\Delta}_{n,q}\mathbf{C}_{\theta,q})'\mathbf{H}_n\mathbf{Z}$.

An equivalent form of (2.6b) can be found in Osborn (1977), and the special case $q = 1$ was discussed in Box, Hillmer, and Tiao (1976). For the MA(q) model of a single time series ($k = 1$), this result reduces to those in Tiao and Ali (1971), Newbold (1974), Ljung and Box (1976), and Dent (1977). In addition, if the term $\hat{\mathbf{a}}'_*\mathbf{A}_{n,q}\hat{\mathbf{a}}_*$ in the exponent and a factor $|\mathbf{\Sigma}|^{-\frac{1}{2}q}|\mathbf{A}_{n,q}|^{-\frac{1}{2}}$ were ignored, then (2.6a) reduces to the likelihood function (conditional on $\mathbf{a}_* = \mathbf{0}$) discussed in Wilson (1973),

$$\ell(\theta, \mathbf{\Sigma}|\mathbf{Z}, \mathbf{a}_* = \mathbf{0}) \propto |\mathbf{\Sigma}|^{-\frac{1}{2}n}\exp(-\tfrac{1}{2}\mathbf{Z}'\mathbf{H}_n\mathbf{Z}) . \quad (2.7)$$

A comparison of the properties of (2.6b) and (2.7) will be given in Section 3.

In evaluating the likelihood (2.6a) for given $(\theta, \mathbf{\Sigma})$ note that

1. If we write $\mathbf{Y}' = (\mathbf{y}'_{-q+1}, \ldots, \mathbf{y}'_n)$, then $\mathbf{y}_{-q+1} = \ldots = \mathbf{y}_0 = \mathbf{0}$ and $\mathbf{y}_j = \mathbf{z}_j + \theta_1\mathbf{y}_{j-1} + \ldots + \theta_q\mathbf{y}_{j-q}$, $j = 1, \ldots, n$.
2. If $\mathbf{P} = \mathbf{F}_{\theta,q}$, then the matrix $\mathbf{X}$ consists of the first q block columns of $\mathbf{D}_{\theta,n+q}^{-1}$ so that the elements of $\mathbf{X}$ can be determined from θ through the difference relationship (2.3).

Thus, the elements of $\mathbf{X}$ and $\mathbf{Y}$ can all be calculated recursively, and the only matrix inversion needed is in computing the $k \times k$ matrix $\mathbf{\Sigma}^{-1}$ and the $(qk) \times (qk)$ matrix $(\mathbf{X}'\mathbf{\Sigma}_{n+q}^{-1}\mathbf{X})^{-1}$. As q increases, computation of the latter becomes increasingly laborious even for a high-speed computer. In situations in which the model assumes special structures, however, suitable choices of the matrix $\mathbf{P}$ in (2.5) can help alleviate the computational burden. We now turn to discuss one such situation.

2.3 Multiplicative Model

In representing multiple time series exhibiting a strong seasonal behavior, a useful model is

$$\mathbf{z}_t = \eta_{q_1}(B)\omega_{q_2}(B^s)\mathbf{a}_t , \quad (2.8)$$

where $\eta_{q_1}(B) = \mathbf{I} - \eta_1 B - \ldots - \eta_{q_1}B^{q_1}$, $\omega_{q_2}(B^s) = \mathbf{I} - \omega_1 B^s - \ldots - \omega_{q_2}B^{q_2 s}$, and $(\eta_1, \ldots, \eta_{q_1}, \omega_1, \ldots, \omega_{q_2})$ are all $k \times k$ matrices. This model is a generalization of the univariate multiplicative moving average model proposed in Box and Jenkins (1970). While (2.8) can be viewed as an MA($q_1 + q_2 s$) model, considerable savings in the number of parameters can be achieved by employing the multiplicative form.

Let $\mathbf{D}_{\eta,n}$, $\mathbf{D}_{\omega,n}$, and $\mathbf{D}_{\theta,n}$ correspond, respectively, to models with moving average polynomials $\eta_{q_1}(B)$, $\omega_{q_2}(B^s)$, and $\theta_q(B) = \eta_{q_1}(B)\omega_{q_2}(B^s)$. For the likelihood function of the parameters $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_{q_1}, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_{q_2})$ and $\boldsymbol{\Sigma}$ in (2.8), we shall choose $\mathbf{P}$ in the following manner. Let us construct an $(n+q)k \times (n+q)k$ matrix $\mathbf{J}$ and partition the matrix $\mathbf{D}_{\omega,n+q}$ as follows:

$$\mathbf{J} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\eta,n+q_1} \end{bmatrix} = \left[\begin{array}{cc|c} \mathbf{Q} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\eta,q_1} & \mathbf{0} \\ \hline \mathbf{0} & -\boldsymbol{\Delta}_{n,q_1}\mathbf{C}_{\eta,q_1} & \mathbf{D}_{\eta,n} \end{array}\right] ,$$

$$\mathbf{D}_{\omega,n+q} = \left[\begin{array}{c|c} \mathbf{L}_1 & \mathbf{0} \\ \hline \mathbf{L}_2 & \mathbf{D}_{\omega,n} \end{array}\right] , \quad (2.9)$$

where $\mathbf{Q} = \dot{\mathbf{Q}} \otimes \mathbf{I}_k$ and $\dot{\mathbf{Q}}' = [\dot{\mathbf{Q}}'_1, \ldots, \dot{\mathbf{Q}}'_s]$ is a $(q_2 s) \times (q_2 s)$ orthogonal matrix such that $\dot{\mathbf{Q}}'_j$ consists of the jth, $(s+j)$th, $\ldots$, $[(q_2-1)s+j]$th columns of $\mathbf{I}_{q_2 s}$, $j = 1, \ldots, s$. It is readily seen that $\mathbf{D}_{\theta,n} = \mathbf{D}_{\eta,n}\mathbf{D}_{\omega,n}$ and $-\boldsymbol{\Delta}_{n,q}\mathbf{C}_{\theta,q} = [\mathbf{0}, -\boldsymbol{\Delta}_{n,q_1}\mathbf{C}_{\eta,q_1}]\mathbf{L}_1 + \mathbf{D}_{\eta,n}\mathbf{L}_2$. Thus, if we let

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\eta,q_1} \end{bmatrix} \mathbf{L}_1 ,$$

then (2.5) can be written as

$$\begin{bmatrix} \mathbf{a}_* \\ \mathbf{a} \end{bmatrix} = \mathbf{D}_{\omega,n+q}^{-1}\mathbf{J}^{-1} \begin{bmatrix} \mathbf{b} \\ \mathbf{Z} \end{bmatrix} . \quad (2.10)$$

In particular, the $\mathbf{X}$ matrix in (2.5) is

$$-\mathbf{X} = [\mathbf{M}_1, \ldots, \mathbf{M}_s, \mathbf{U}] , \quad \mathbf{U} = \begin{bmatrix} \mathbf{0} \\ \mathbf{R} \end{bmatrix} , \quad (2.11)$$

where $\mathbf{M}_j$ is the $(n+q)k \times q_2 k$ matrix consisting of the jth, $(s+j)$th, $\ldots$, $[(q_2-1)s+j]$th block columns of $\mathbf{D}_{\omega,n+q}^{-1}$, $(j = 1, \ldots, s)$, and $\mathbf{R}$ consists of the first, $\ldots$, q_1th block columns of $\mathbf{D}_{\theta,n+q_1}^{-1}$. It is easy to verify that

$$\mathbf{X}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{X} = [\mathbf{d}_{ij}] , \quad i, j = 1, \ldots, s+1 , \quad (2.12)$$

where (a) for $i = 1, \ldots, s$, $\mathbf{d}_{ii} = \mathbf{M}'_i\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{M}_i$, $\mathbf{d}_{i,s+1} = \mathbf{M}'_i\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{U}$, and $\mathbf{d}_{ij} = \mathbf{0}$, $j = 2, \ldots, s$ and $j > i$; (b) $\mathbf{d}_{s+1,s+1} = \mathbf{U}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{U}$. Thus, standard partitional inverse formula can be applied to obtain $(\mathbf{X}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{X})^{-1}$, involving computing the inverse of the $(q_2k) \times (q_2k)$ matrices $\mathbf{d}_{11}, \ldots, \mathbf{d}_{ss}$ and the $(q_1k) \times (q_1k)$ matrix $\mathbf{d}_{s+1,s+1}$ instead of a $(qk) \times (qk)$ matrix.

3. COMPARISON OF THE EXACT AND THE CONDITIONAL LIKELIHOOD FOR MA(q) MODEL

This section compares some properties of estimates of $\boldsymbol{\theta}$ obtained by maximizing the entire likelihood (2.6b) with those obtained by maximizing the conditional likelihood (2.7). We shall be concerned only with the situation when the model is noninvertible, that is, $|\boldsymbol{\theta}_q(B)|$ has one or more zeros on the unit circle.

Dunsmuir and Hannan (1976) have shown that the maximum likelihood estimates with respect to (2.6b) are consistent both for invertible and noninvertible models. Reinsel (1976) has indicated that the consistency property also holds for estimates corresponding to (2.7). In what follows we shall concentrate mainly on finite sample situations.

3.1 Expected Value of the Log-Likelihood

To appreciate the effect of the condition $\mathbf{a}_* = \mathbf{0}$ on the likelihood function, it is useful to consider first the expected value of the negative of the log-likelihood function. We shall illustrate the situation for the case of the univariate $(k = 1)$ MA(1) model,

$$z_t = a_t - \theta_0 a_{t-1} , \quad (3.1)$$

where θ_0 is the true value. We discuss the case $\theta_0 = 1$ and, for simplicity, shall assume that σ^2 is known to be unity. The case $\theta_0 = -1$ behaves in a similar manner.

It is straightforward to verify from (2.7) that

$$-(1/n)E_{\mathbf{Z}|\theta_0=1}[\ln \ell(\theta | \mathbf{Z}, a_* = 0)] = C + R_0(\theta) , \quad (3.2)$$

where

$$R_0(\theta) = \frac{1}{1+\theta}\left[1 + \frac{\theta}{n}\frac{(1-\theta^{2n})}{(1-\theta^2)}\right] . \quad (3.3)$$

Also, using (2.6b), we obtain $-(1/n)E_{\mathbf{Z}|\theta_0=1}[\ln \ell(\theta|\mathbf{Z})] = C' + R_1(\theta)$, where

$$R_1(\theta) = R_0(\theta) + \frac{1}{2n}\ln\frac{1-\theta^{2(n+1)}}{1-\theta^2} - \frac{\theta^2(1-\theta^{2n})(1-\theta^{2n+1}) + n\theta^{2n+1}(1+\theta)(1-\theta)^2}{n(1-\theta^2)(1+\theta)(1-\theta^{2(n+1)})} . \quad (3.4)$$

Note from (3.3) that (a) $R_0(-1) = n$ and $R_0(0) = R_0(1) = 1$; (b) when $n > 1$, $R'_0(\theta) < 0$ for $-1 < \theta \leq 0$; and (c) for sufficiently large n, $R_0(\theta)$ is approximately equal to $(1+\theta)^{-1}$ if $-1 < \theta < 1$. It follows that the value of θ minimizing $R_0(\theta)$ must lie in the open interval (0, 1); however, for large enough n, this value can be made as close to 1 as desired. Now $R_1(\theta)$ corresponding to the exact likelihood differs from $R_0(\theta)$ by the two terms shown in (3.4).

The following tabulation gives values of θ that minimize $R_0(\theta)$ for various n:

n	10	50	100	200	500	1,000
$R_0(\theta)$	.72	.87	.90	.93	.96	.97

The entries imply that very large sample sizes would be required to produce good estimates when maximizing the conditional likelihood (2.7). On the other hand, it can be verified that $\theta = 1$ will minimize $R_1(\theta)$ for any n so there is no bias relative to the exact likelihood. The case $n = 10$, seemingly too small a sample size to be of interest, is in fact of practical significance. Specifically, in the analysis of monthly seasonal data, a model of the form $y_t = a_t - \theta_0 a_{t-12}$, where y_t is some differences of the original data, has been frequently employed. Writing $t = 12T + j$, $Y_{jT} = y_{12T+j}$, and $\alpha_{jT} = a_{12T+j}$, $j = 1, \ldots, 12$, we see that this situation is equivalent to having 12 independent series, one for each month of the year, such that each follows the MA(1) model (3.1). If, as is often

1. Simulation Results for MA(1) and Seasonal MA(1) Models*

		Parameters				*Eigenvalues*		*100 × Mean Squared Error*			
True Values		θ_{11} .72	θ_{12} .56	θ_{21} .21	θ_{22} .58	λ_1 1.0	λ_2 .3	θ_{11}	θ_{12}	θ_{21}	θ_{22}
n	*Method*	*Averages of Estimates*				*Averages of Estimates*					
50	E[a]	.71	.55	.21	.56	.98	.29	.849	1.848	.801	2.685
	C	.69	.51	.17	.50	.90	.29	1.720	2.699	1.762	3.744
100	E	.72	.56	.21	.57	1.00	.29	.290	.848	.324	.953
	C	.72	.53	.19	.53	.96	.29	.835	1.137	.510	1.538
200	E	.72	.57	.21	.57	1.00	.29	.113	.382	.108	.406
	C	.70	.54	.19	.54	.95	.29	.185	.493	.201	.665
400	E	.72	.56	.21	.57	1.00	.29	.054	.191	.057	.207
	C	.70	.54	.19	.55	.95	.30	.011	.238	.121	.354
1,000	E	.72	.56	.21	.58	1.00	.30	.015	.057	.016	.063
	C	.71	.55	.20	.56	.98	.29	.034	.093	.041	.122
108[b]	E	.70	.55	.17	.56	.94	.32	.611	1.183	.828	1.564
	C	.60	.42	.05	.40	.68	.32	2.066	2.997	2.893	4.389

[a] E: Exact likelihood; C: Conditional likelihood.
[b] Seasonal MA*(1) model.

the case, we have a data span of about 10 years, then the results corresponding to $n = 10$ will be applicable.

3.2 Simulations

In order to compare small sample properties of the parameter estimates obtained by maximizing the exact and the conditional likelihood, we have performed a number of simulation experiments. One hundred realizations were generated from a bivariate ($k = 2$) MA(1) model $\mathbf{z}_t = \mathbf{a}_t - \boldsymbol{\theta}\mathbf{a}_{t-1}$ for various sample sizes n. In addition, 100 realizations were obtained from the bivariate seasonal MA*(1) model $\mathbf{z}_t = \mathbf{a}_t - \boldsymbol{\theta}\mathbf{a}_{t-12}$ for $n = 108$. The true $\boldsymbol{\theta}$ matrix had one of its eigenvalues equal to one. The elements of the $\mathbf{a}_t$'s were independent random normal deviates generated on a UNIVAC 1110 computer using the Box and Muller (1958) algorithm. For each realization, a nonlinear regression routine was employed to obtain parameter estimates by maximizing the exact and the conditional likelihood functions. The results are summarized in Table 1, where $(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$ are the elements and λ_1 and λ_2 are the eigenvalues of $\boldsymbol{\theta}$.

We make the following observations from Table 1:

1. The exact likelihood yields estimates that are in close agreement with the true $\boldsymbol{\theta}$ for all n considered.
2. For $n = 50$ and for the seasonal case, the conditional likelihood method performs very poorly. Specifically, the larger eigenvalues of the estimated $\boldsymbol{\theta}$ matrix are significantly smaller than unity.
3. The bias for the conditional likelihood estimates decreases as n increases.
4. For the larger eigenvalue, the estimates in these two methods behave very much like the corresponding values in the tabulation in Section 3.1.
5. The mean squared errors for the exact likelihood estimates are considerably smaller than those for the other estimates. For the MA(1) model, this situation mainly results from an increased variance for the conditional estimates; however, for the seasonal MA*(1) model, the larger bias for the conditional estimates accounts for most of the increase in the mean squared error.

Although the scope of the simulation experiments is limited, the results in the table strongly suggest that estimates based on the exact likelihood should perform substantially better than the conditional ones when the model is not invertible.

4. LIKELIHOOD FUNCTION FOR THE ARMA(*p, q*) MODEL

4.1 The Case $p = q$

For the ARMA (p, q) model, we first discuss the special case $p = q$. If we employ the notations in (2.1), the probabilistic structure of $\mathbf{Z}$ for the ARMA(q, q) model can be written

$$\mathbf{D}_{\phi,n}\mathbf{Z} = \mathbf{D}_{\theta,n}\mathbf{a} + \boldsymbol{\Delta}_{n,q}\mathbf{v} \;, \tag{4.1}$$

where $\mathbf{v} = \mathbf{C}_{\phi,q}\mathbf{Z}_* - \mathbf{C}_{\theta,q}\mathbf{a}_*$, $\mathbf{Z}'_* = (\mathbf{z}'_{-q+1}, \ldots, \mathbf{z}'_0)$, $\mathbf{v}$ and $\mathbf{a}$ being independent. For the likelihood function, consider first the distribution of $\mathbf{W} = \mathbf{D}_{\phi,n}\mathbf{Z}$. Let $\mathbf{Z}' = [\mathbf{Z}'_1, \mathbf{Z}'_2]$, $\mathbf{W}' = [\mathbf{W}'_1, \mathbf{W}'_2]$, and $\mathbf{a}' = [\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2]$, where $\mathbf{Z}_1$, $\mathbf{W}_1$, and $\boldsymbol{\alpha}_1$ consist of the first q $k \times 1$ vectors of $\mathbf{Z}$, $\mathbf{W}$, and $\mathbf{a}$, respectively. From (4.1),

$$\begin{bmatrix}\mathbf{W}_1\\ \mathbf{W}_2\end{bmatrix} = \begin{bmatrix}\mathbf{F}_{\theta,q} & \mathbf{0}\\ -\boldsymbol{\Delta}_{n-q,q}\mathbf{C}_{\theta,q} & \mathbf{D}_{\theta,n-q}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\end{bmatrix} + \begin{bmatrix}\mathbf{v}\\ \mathbf{0}\end{bmatrix}, \tag{4.2}$$

where $\mathbf{W}_1 = \mathbf{F}_{\phi,q}\mathbf{Z}_1$ and $\mathbf{W}_2 = \mathbf{D}_{\phi,n-q}\mathbf{Z}_2 - \boldsymbol{\Delta}_{n-q,q}\mathbf{C}_{\phi,q}\mathbf{Z}_1$.

Now the distribution of $\mathbf{W}$ can be written $p(\mathbf{W}) = p(\mathbf{W}_2)p(\mathbf{W}_1|\mathbf{W}_2)$. First, we see from (2.4) and (4.2) that $p(\mathbf{W}_2)$ corresponds to the density of $n - q$ observations from an MA(q) process. If we apply the results in (2.6b), the associated likelihood function of $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_q)$, $\boldsymbol{\theta}$, and $\boldsymbol{\Sigma}$ is

$$\ell_1(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma} | \mathbf{Z}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}n} |\mathbf{A}_{n-q,q}|^{-\frac{1}{2}} \cdot \exp\left(-\tfrac{1}{2}[\mathbf{W}'_2 \mathbf{H}_{n-q} \mathbf{W}_2 - \tilde{\boldsymbol{\alpha}}'_1 \mathbf{A}_{n-q,q} \tilde{\boldsymbol{\alpha}}_1]\right) , \quad (4.3)$$

where $\tilde{\boldsymbol{\alpha}}_1 = -\mathbf{A}_{n-q,q}^{-1}(\boldsymbol{\Delta}_{n-q,q}\mathbf{C}_{\theta,q})'\mathbf{H}_{n-q}\mathbf{W}_2$. Second, if we use standard multivariate normal theory, the likelihood associated with the conditional distribution $p(\mathbf{W}_1|\mathbf{W}_2)$ is found to be

$$\ell_2(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma} | \mathbf{Z}) \propto |\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q}|^{-\frac{1}{2}} \cdot \exp\left(-\tfrac{1}{2}\mathbf{W}'_{1\cdot}(\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q})^{-1}\mathbf{W}_{1\cdot}\right) , \quad (4.4)$$

where $\mathbf{W}_{1\cdot} = \mathbf{W}_1 - \mathbf{F}_{\theta,q}\tilde{\boldsymbol{\alpha}}_1$, and $\boldsymbol{\Gamma}$ is the covariance matrix of $\mathbf{v}$. Thus, the complete likelihood function is

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma} | \mathbf{Z}) \propto \ell_1(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma} | \mathbf{Z})\ell_2(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma} | \mathbf{Z}) . \quad (4.5)$$

For a pure MA(q) process ($\boldsymbol{\phi} = \mathbf{0}$) $\boldsymbol{\Gamma} = \mathbf{C}_{\theta,q}\boldsymbol{\Sigma}_q\mathbf{C}'_{\theta,q}$, and (4.5) reduces to the result in (2.6b). On the other hand, for an AR(q) ($\boldsymbol{\theta} = \mathbf{0}$) or the general ARMA($q$, q) model, explicit expressions for $\boldsymbol{\Gamma}$ are rather complicated. We shall show later in Section 4.3 that as n increases, $\ln \ell_1$ tends to dominate $\ln \ell_2$ irrespective of whether $|\boldsymbol{\theta}_q(B)|$ in (1.1) has zeros on the unit circle. Therefore, for large n, $\ell(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z})$ can be approximated by $\ell_1(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z})$.

4.2 The Pure Autoregressive Model

When $\boldsymbol{\theta} = \mathbf{0}$, $\mathbf{C}_{\theta,q} = \mathbf{0}$ so that $\mathbf{A}_{n-q,q} = \boldsymbol{\Sigma}_q^{-1}$ and $\tilde{\boldsymbol{\alpha}}_1 = \mathbf{0}$. Also, $\mathbf{H}_{n-q} = \boldsymbol{\Sigma}_{n-q}^{-1}$ and $\mathbf{F}_{\theta,q} = \mathbf{I}$. Thus

$$\ell_1(\boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(n-q)} \exp\left(-\tfrac{1}{2}\mathbf{W}'_2\boldsymbol{\Sigma}_{n-q}^{-1}\mathbf{W}_2\right) \quad (4.6)$$

and

$$\ell_2(\boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) \propto |\boldsymbol{\Gamma} + \boldsymbol{\Sigma}_q|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\mathbf{W}'_1(\boldsymbol{\Gamma} + \boldsymbol{\Sigma}_q)^{-1}\mathbf{W}_1\right) .$$

Specifically, ℓ_2 corresponds to the marginal distribution of $\mathbf{Z}_1$ and ℓ_1, the conditional distribution of $\mathbf{Z}_2$ given $\mathbf{Z}_1$. Consequently, ignoring the contribution of ℓ_2 to the likelihood function amounts to working solely with the conditional distribution $p(\mathbf{Z}_2|\mathbf{Z}_1)$. In this case, the parameters $\boldsymbol{\phi}$ can be estimated by least squares and, as is well known (Anderson 1971), these estimates have desirable asymptotic properties.

4.3 An Approximation to the Likelihood Function

We show in this section that the likelihood function in (4.5) can be closely approximated by $\ell_1(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z})$ in (4.3) when n is large. We first establish some useful preliminary results.

1. Let $\mathbf{B}_1$ and $\mathbf{B}_2$ be two positive definite symmetric matrices of the same dimension. We shall adopt the notation $\mathbf{B}_1 \geq \mathbf{B}_2$ to mean that $\mathbf{B}_1 - \mathbf{B}_2$ is positive semidefinite. Thus, $\mathbf{B}_1 \geq \mathbf{B}_2$ implies that $\mathbf{B}_1^{-1} \leq \mathbf{B}_2^{-1}$. Consider now the matrix $(\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q})^{-1}$ in (4.4). Since $\mathbf{A}_{n-q,q} \geq \boldsymbol{\Sigma}_q^{-1}$ and $\boldsymbol{\Gamma}$ is positive definite, repeated applications of the results mentioned yield

$$(\mathbf{F}_{\theta,q}\boldsymbol{\Sigma}_q\mathbf{F}'_{\theta,q} + \boldsymbol{\Gamma})^{-1} \leq (\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q})^{-1} \leq \boldsymbol{\Gamma}^{-1} . \quad (4.7)$$

This implies that the eigenvalues of

$$(\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q})^{-1}$$

are bounded for all $\boldsymbol{\theta}$.

2. To simplify the derivation of the asymptotic results, we shall assume in this section that n is a multiple of q. Letting $m = n/q$, we can write

$$\mathbf{D}_{n-q,\theta} = \begin{bmatrix} \mathbf{F}_{\theta,q} & & & \mathbf{0} \\ & \ddots & & \\ -\mathbf{C}_{\theta,q} & & & \\ & \ddots & & \\ \mathbf{0} & & -\mathbf{C}_{\theta,q} & \mathbf{F}_{\theta,q} \end{bmatrix} ,$$

$$\mathbf{D}_{n-q,\theta}^{-1} = \begin{bmatrix} \mathbf{I} & & & \mathbf{0} \\ \mathbf{G} & \ddots & & \\ \vdots & \ddots & & \\ \mathbf{G}^{m-2} & \ldots & \mathbf{G} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{F}_{\theta,q}^{-1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{F}_{\theta,q}^{-1} \end{bmatrix} , \quad (4.8)$$

where $\mathbf{G} = \mathbf{F}_{\theta,q}^{-1}\mathbf{C}_{\theta,q}$. Thus,

$$\mathbf{A}_{n-q,q} = \boldsymbol{\Sigma}_q^{-1} + \sum_{j=1}^{m-1} \mathbf{G}'^j\boldsymbol{\Sigma}_q^{-1}\mathbf{G}^j = \mathbf{R}_{(m-1)} , \quad \text{say} . \quad (4.9)$$

Also,

$$(\boldsymbol{\Delta}_{n-q,q}\mathbf{C}_{\theta,q})'\mathbf{H}_{n-q} = [\mathbf{G}'\mathbf{R}_{(m-2)}\mathbf{F}_{\theta,q}^{-1}, \mathbf{G}'^2\mathbf{R}_{(m-3)}\mathbf{F}_{\theta,q}^{-1}, \ldots, \mathbf{G}'^{m-1}\mathbf{R}_{(0)}\mathbf{F}_{\theta,q}^{-1}] ,$$

where $\mathbf{R}_{(0)} = \boldsymbol{\Sigma}_q^{-1}$. Thus, writing $\mathbf{W}'_2 = (\mathbf{w}'_2, \ldots, \mathbf{w}'_m)$, where each element is a $1 \times (qk)$ vector, we can express the quantity $\tilde{\boldsymbol{\alpha}}_1$ in (4.3) as

$$\tilde{\boldsymbol{\alpha}}_1 = -\mathbf{R}_{(m-1)}^{-1} \sum_{j=1}^{m-1} \mathbf{G}'^j\mathbf{R}_{(m-1-j)}\mathbf{F}_{\theta,q}^{-1}\mathbf{w}_{j+1} . \quad (4.10)$$

3. It is straightforward to verify that $\mathbf{G} = \mathbf{L}^q$ where

$$\mathbf{L} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \ddots & & \\ & \ddots & & \mathbf{0} & \mathbf{I} \\ \boldsymbol{\theta}_q & & \ldots & \boldsymbol{\theta}_2 & \boldsymbol{\theta}_1 \end{bmatrix} \quad (4.11)$$

and that the characteristic polynomial of $\mathbf{L}$ is $|\lambda^q\mathbf{I} - \lambda^{q-1}\boldsymbol{\theta}_1 - \ldots - \boldsymbol{\theta}_q| = 0$. Hence, the eigenvalues of $\mathbf{L}$ are reciprocals of the zeros of $|\boldsymbol{\theta}_q(B)|$.

4. Note that there exists a nonsingular matrix $\mathbf{H}$ such that $\mathbf{L} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^{-1}$ where $\boldsymbol{\Lambda}$ is the Jordan canonical form of $\mathbf{L}$,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \delta_1 & & \\ & \ddots & \ddots & \\ & & & \delta_{qk-1} \\ & & & \lambda_{qk} \end{bmatrix} , \quad (4.12)$$

and δ_i can be either 1 or 0. Thus, $\mathbf{G}^m = \mathbf{H}\boldsymbol{\Lambda}^n\mathbf{H}^{-1}$ where $\boldsymbol{\Lambda}^n$ is a $(qk) \times (qk)$ upper triangular matrix such that (a) the ith diagonal element is $\lambda_i{}^n$, and (b) for $n > qk$ the modulus of (upper) off-diagonal elements are bounded by $\binom{n}{qk-1}|\lambda_*|^{n-qk+1}$ where $|\lambda_*| = \max_i|\lambda_i|$.

Turning now to the likelihood ℓ_2 in (4.4), consider

$$-(2/n)\ln \ell_2(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) = -(1/n)\ln|\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q}|^{-1} + (1/n)\mathbf{W}'_{1\cdot}(\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q})^{-1}\mathbf{W}_{1\cdot} \quad (4.13)$$

From (4.7), the first term on the right side will approach zero as $n \to \infty$ for all $\boldsymbol{\theta}$.

Suppose now that $|\boldsymbol{\theta}_q(B)|$ has all its zeros lying outside the unit circle. We see from (4.12) that as n increases $\mathbf{G}^m$ approaches the null matrix. This in turn implies that the matrix $\mathbf{R}_{(m-1)}$ in (4.9) will tend to a finite, limiting positive definite matrix $\mathbf{R}$. Thus, as long as the elements of $\mathbf{Z}$ are finite, the elements of the vector $\tilde{\boldsymbol{\alpha}}_1$ in (4.10) will be bounded that, together with (4.7), implies that the second term on the right side of (4.13) will tend to zero. Now, in (4.3)

$$-(2/n)\ln \ell_1(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) = \ln|\boldsymbol{\Sigma}| + (1/n)\mathbf{W}'_2\mathbf{H}_{n-q}\mathbf{W}_2 + (1/n)(\ln|\mathbf{A}_{n-q,q}| - \tilde{\boldsymbol{\alpha}}'_1\mathbf{A}_{n-q,q}\tilde{\boldsymbol{\alpha}}_1) \quad (4.14)$$

As n increases, the last term approaches zero. Since $\mathbf{H}_{n-q} = \mathbf{D}_{n-q,\theta}^{-1'}\boldsymbol{\Sigma}_{n-q}^{-1}\mathbf{D}_{n-q,\theta}^{-1}$, we see from (4.8) that

$$\frac{1}{n}\mathbf{W}'_2\mathbf{H}_{n-q}\mathbf{W}_2 = \frac{1}{mq}\sum_{j=2}^{m}\mathbf{u}'_j\boldsymbol{\Sigma}_q^{-1}\mathbf{u}_j , \quad (4.15)$$

where $\mathbf{u}_j = \sum_{k=2}^{j}\mathbf{G}^{k-2}\mathbf{F}_{\theta,q}^{-1}\mathbf{w}_{j-(k-2)}$, whose elements are bounded for all j. It follows that $\mathbf{W}'_2\mathbf{H}_{n-q}\mathbf{W}_2$ increases with n. Therefore, we can conclude that the influence of ℓ_1 dominates that of ℓ_2, and, in fact, the likelihood function can be approximated by

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) \doteq |\boldsymbol{\Sigma}|^{-(n/2)}\exp(-\tfrac{1}{2}\mathbf{W}'_2\mathbf{H}_{n-q}\mathbf{W}_2) . \quad (4.16)$$

On the other hand, suppose $|\boldsymbol{\theta}_q(B)|$ has r, $(r \le kq)$, of its zeros lying on the unit circle. In this case, from (4.12) the elements of $\mathbf{G}^m$ can be of order as large as n^{r-1}, and hence those of $\mathbf{A}_{n-q,q}$ will be of order at least n. To show that ℓ_1 still dominates ℓ_2, note first that the exponent in (4.3) can be written alternatively as

$$\mathbf{W}'_2\mathbf{H}_{n-q}\mathbf{W}_2 - \tilde{\boldsymbol{\alpha}}_1\mathbf{A}_{n-q,q}\tilde{\boldsymbol{\alpha}}_1 = \mathbf{W}'\mathbf{H}_n\mathbf{W} - \mathbf{W}_{1\cdot}\mathbf{F}'_{\theta,q}{}^{-1}\mathbf{A}_{n-q,q}\mathbf{F}_{\theta,q}{}^{-1}\mathbf{W}_{1\cdot} \quad (4.17)$$

Since, from (4.7), $(1/n)(\boldsymbol{\Gamma} + \mathbf{F}_{\theta,q}\mathbf{A}_{n-q,q}^{-1}\mathbf{F}'_{\theta,q})^{-1}$ approaches the null matrix as n increases, we see that $(1/n)\mathbf{W}'_{1\cdot}\mathbf{F}'^{-1}_{\theta,q}\mathbf{A}_{n-q,q}\mathbf{F}_{\theta,q}^{-1}\mathbf{W}_{1\cdot}$ will dominate the second term on the right side of (4.13), and hence the desired result follows.

We may conclude from this analysis that for an ARMA(q, q) model, irrespective of whether $|\boldsymbol{\theta}_q(B)|$ has zeros on the unit circle, for large n, the contribution of $\ell_2(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z})$ can be ignored and the likelihood function can be approximated by $\ell_1(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z})$.

4.4 The General ARMA(*p*, *q*) Model

The approximation (4.3) for the case $p = q$ can be readily extended to the general ARMA(p, q) model. First, suppose $p > q$. Substituting p for q in (4.2) and writing $\mathbf{W}_1$ as $\mathbf{W}_{1p}$ and $\mathbf{W}_2$ as $\mathbf{W}_{2p}$ we have that

$$\begin{bmatrix}\mathbf{W}_{1p}\\ \mathbf{W}_{2p}\end{bmatrix} = \begin{bmatrix}\mathbf{F}_{\theta,p} & \mathbf{0}\\ -\boldsymbol{\Delta}_{n-p,p}\mathbf{C}_{\theta,p} & \mathbf{D}_{\theta,n-p}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\end{bmatrix} + \begin{bmatrix}\mathbf{v}\\ \mathbf{0}\end{bmatrix}, \quad (4.18)$$

where $\mathbf{W}_{1p} = \mathbf{F}_{\phi,p}\mathbf{Z}_1$, $\mathbf{W}_{2p} = \mathbf{D}_{\phi,n-p}\mathbf{Z}_2 - \boldsymbol{\Delta}_{n-p,p}\mathbf{C}_{\phi,p}\mathbf{Z}_1$, and it is understood that $\boldsymbol{\theta}_{q+1} = \ldots = \boldsymbol{\theta}_p = \mathbf{0}$ in $\mathbf{F}_{\theta,p}$, $\mathbf{C}_{\theta,p}$, and $\mathbf{D}_{\theta,n-p}$. It follows from the development in Sections 4.2 and 4.3 that the likelihood function can be approximated by (4.3) with appropriate change in notations. Further, since $-\boldsymbol{\Delta}_{n-p,p}\mathbf{C}_{\theta,p} = [\mathbf{0}, -\boldsymbol{\Delta}_{n-p,q}\mathbf{C}_{\theta,q}]$, we see from (4.18) that

$$\mathbf{W}_{2p} = \mathbf{D}_{\theta,n-p}\boldsymbol{\alpha}_2 - \boldsymbol{\Delta}_{n-p,q}\mathbf{C}_{\theta,q}\boldsymbol{\alpha}_{1*} , \quad (4.19)$$

where $\boldsymbol{\alpha}'_{1*} = (\mathbf{a}'_{p-q+1}, \ldots, \mathbf{a}'_p)$. Thus, the likelihood associated with $p(\mathbf{W}_2)$ in (4.3) can be alternatively written as

$$\ell_{1*}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(n-p+q)}|\mathbf{A}_{n-p,q}|^{-\frac{1}{2}} \cdot\exp(-\tfrac{1}{2}[\mathbf{W}'_{2p}\mathbf{H}_{n-p}\mathbf{W}_{2p} - \tilde{\boldsymbol{\alpha}}'_{1*}\mathbf{A}_{n-p,q}\tilde{\boldsymbol{\alpha}}_{1*}]) , \quad (4.20)$$

where $\tilde{\boldsymbol{\alpha}}_{1*} = -\mathbf{A}_{n-p,q}^{-1}(\boldsymbol{\Delta}_{n-p,q}\mathbf{C}_{\theta,q})'\mathbf{H}_{n-p}\mathbf{W}_{2p}$. The advantage of this form is that it involves inverting the $(qk) \times (qk)$ matrix $\mathbf{A}_{n-p,q}$ instead of the $(pk) \times (pk)$ matrix $\mathbf{A}_{n-p,p}$ implied by (4.3).

We now show that the approximation (4.20) also applies to the case $p < q$. First, suppose we split the $\mathbf{W}$ vector into $\mathbf{W}_1$ and $\mathbf{W}_2$ as was given in (4.2), where it is understood that in making the transformation $\mathbf{W} = \mathbf{D}_{\phi,n}\mathbf{Z}$, $\boldsymbol{\phi}_{p+1}, \ldots, \boldsymbol{\phi}_q$ are set to be equal to zero in $\mathbf{D}_{\phi,n}$. We can further decompose $\mathbf{W}_1$ into $\mathbf{W}_{1p}$ and $\mathbf{W}_{1*}$. Now the results in the preceding two sections show that the likelihood associated with $p(\mathbf{W}_2)$ dominates that with $p(\mathbf{W}_1|\mathbf{W}_2)$. It follows that the likelihood associated with $p(\mathbf{W}_2, \mathbf{W}_{1*})$ will dominate that with $p(\mathbf{W}_{1p}|\mathbf{W}_2, \mathbf{W}_{1*})$. But $\mathbf{W}'_{2p} = [\mathbf{W}'_{1*}, \mathbf{W}_2]$, and, since the structure of (4.19) also holds for $p < q$, the likelihood function can be closely approximated by ℓ_{1*} in (4.20).

In conclusion, we have shown that an approximation to the likelihood function for the general ARMA(p, q) model is obtained by (a) making the transformation $\mathbf{W} = \mathbf{D}_{\phi,n}\mathbf{Z}$, (b) ignoring the elements $\mathbf{W}_{1p}$, and (c) treating $\mathbf{W}_{2p}$ as $n - p$ vector observations from an MA(q) process. Note that, in obtaining the approximation (4.20), no assumption is made about the invertibility of the moving average polynomial $\boldsymbol{\theta}_q(B)$.

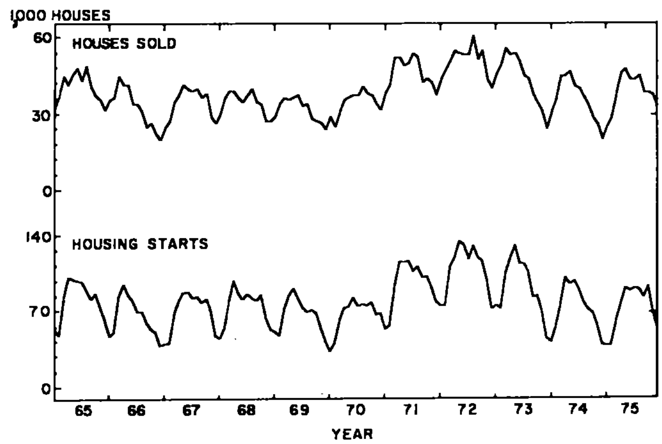
A computer program has been developed at the University of Wisconsin to calculate parameter estimates corresponding to (2.6a) for moving average models and to (4.20) for mixed models. Details are given in the Appendix.

If in (4.20), the factor $|\boldsymbol{\Sigma}|^{-\frac{1}{2}q}|\mathbf{A}_{n-p,q}|^{-\frac{1}{2}}$ and the term $\tilde{\boldsymbol{\alpha}}'_{1*}\mathbf{A}_{n-p,q}\tilde{\boldsymbol{\alpha}}_{1*}$ in the exponent were ignored, we would obtain the approximation

$$\ell_0(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\Sigma}|\mathbf{Z}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(n-p)} \cdot\exp(-\tfrac{1}{2}\mathbf{W}'_{2p}\mathbf{H}_{n-p}\mathbf{W}_{2p}) \quad (4.21)$$

proposed, for example, in Wilson (1973), which rests on

Housing Example



the assumption that all zeros of $|\theta_q(B)|$ are lying outside the unit circle. A comparison of these two approximations, (4.20) and (4.21), will be given in Section 5 in terms of an actual example.

5. AN EXAMPLE

As an illustrative example, we consider two monthly U.S. housing series for the period January 1965 through May 1975: (a) single-family housing starts z_{1t} and (b) houses sold z_{2t}. The data were obtained from *The Survey of Current Business* and are shown in the figure. Both series exhibit a strong seasonal behavior, and the dependence between the series is apparent.

5.1 Models for Individual Series

Following the modeling procedure in Box and Jenkins (1970), we find that individually each series can be well represented by a univariate model of the multiplicative form

$$(1 - B)(1 - B^{12})z_t = (1 - \theta B)(1 - \eta B^{12})a_t \ , \quad (5.1)$$

so that $y_{it} = (1 - B)(1 - B^{12})z_{it}$, $i = 1, 2$, would follow a stationary moving average model. The following tabulation compares the parameter estimates obtained by employing the exact (E) likelihood (2.6a) with those from the conditional (C) likelihood (2.7), where the numbers in parentheses are the associated estimated standard errors.

		θ	η	σ_a^2
Housing Starts	E	.28 (.09)	.91 (.06)	41.61
	C	.30 (.09)	.75 (.07)	50.49
Houses Sold	E	.16 (.10)	1.00 (.06)	11.93
	C	.24 (.10)	.72 (.08)	16.46

We observe that for both series the estimates of the seasonal moving average parameter are appreciably smaller for C than for E and that, judging from the estimates of σ_a^2, E yields a significantly better fit than does C. It appears that η is close to unity, especially for the houses-sold series, implying a possible deterministic seasonal structure, that is,

$$(1 - B)z_t = S_t + (1 - \theta B)a_t \ , \quad (5.2)$$

where $(1 - B^{12})S_t = 0$. Such a structure, however, would not be detected if the conditional likelihood method were employed.

5.2 Bivariate Model

In the individual analysis, we initially found that a first difference and a seasonal difference were required to achieve stationarity for both series. This situation has led us to consider modeling jointly the transformed series

$$\mathbf{y}_t = (1 - B)(1 - B^{12})\mathbf{z}_t \quad (5.3)$$

where $\mathbf{y}'_t = (y_{1t}, y_{2t})$ and $\mathbf{z}'_t = (z_{1t}, z_{2t})$.

Inspection of the sample cross-correlation matrices of $\mathbf{y}_t$ suggests the model

$$(\mathbf{I} - \boldsymbol{\phi}B)\mathbf{y}_t = (\mathbf{I} - \boldsymbol{\theta}B)(I - \boldsymbol{\eta}B^{12})\mathbf{a}_t \ , \quad (5.4)$$

where $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, and $\boldsymbol{\eta}$ are all 2×2 matrices and $\mathbf{a}'_t = (a_{1t}, a_{2t})$. Table 2 gives the estimates of $(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\eta})$ along with their eigenvalues and the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{a}_t$, based on the approximation ℓ_1^* in (4.20) and the approximation ℓ_0 in (4.21). For both these models examination of the residuals failed to indicate any apparent inadequacies. We see that estimates of $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ from these two methods are strikingly different. Specifically, for ℓ_{1*}, the estimated $\boldsymbol{\eta}$ has eigenvalues nearly on the unit circle, but this is not true for ℓ_0. Also, as in the analysis of individual models, the estimated residual variances are considerably smaller for ℓ_{1*} than those for ℓ_0, implying that ℓ_{1*} yields a better fit.

5.3 Implication of the Joint Model

Using the estimation results based on ℓ_{1*}, we shall study the implications of the fitted model. For this purpose, we shall set the elements of $(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\eta})$, whose estimates are small compared with their estimated standard errors, equal to zero. To simplify the fitted model further, we set the (1, 1)th elements of $\boldsymbol{\theta}$ and the (1, 1)th and the (2, 2)th element of $\boldsymbol{\eta}$ equal to one. It follows that approximately

$$\begin{aligned}(1 - .64B)y_{1t} &= 2.54y_{2(t-1)} \\ &\quad + (1 - B)(1 - B^{12})a_{1t} \ , \\ y_{2t} &= (1 - B^{12})a_{2t} \ . \end{aligned} \quad (5.5)$$

From (5.3), we can cancel the common factors $(1 - B)$ and $(1 - B^{12})$ so that, in terms of the original series $\mathbf{z}_t$, the fitted model can be written as

$$\begin{aligned}(1 - .64B)z_{1t} &= 2.54z_{2(t-1)} + \mu + S_{1t} + a_{1t} \ , \\ (1 - B)z_{2t} &= S_{2t} + a_{2t} \ , \end{aligned} \quad (5.6)$$

where μ is a constant and S_{1t} and S_{2t} are deterministic seasonal factors such that $(1 - B^{12})S_{it} = 0$, $i = 1, 2$. The model (5.6) implies that the series z_{2t} of houses sold is nonstationary; contains a deterministic seasonal component; and, except for an estimated correlation of .33 between a_{1t} and a_{2t}, is otherwise not affected by the series of housing starts. On the other hand, current housing starts z_{1t} depend not only on the previous month's housing

2. Parameter Estimate for a Bivariate Model Based on ℓ_{1*} and ℓ_0

Estimation Method	Estimates			
	ϕ	θ	η	Σ
ℓ_{1*} in (4.20)	$\begin{bmatrix} .64 & 2.54 \\ (.22) & (1.33) \\ -.04 & -.31 \\ (.14) & (.59) \end{bmatrix}$	$\begin{bmatrix} 1.19 & 1.57 \\ (.22) & (1.29) \\ -.18 & -.12 \\ (.15) & (.60) \end{bmatrix}$	$\begin{bmatrix} .96 & -.06 \\ (.07) & (.13) \\ .05 & 1.03 \\ (.05) & (.07) \end{bmatrix}$	$\begin{bmatrix} 29.58 & 5.92 \\ 5.92 & 11.09 \end{bmatrix}$
	$\lambda_1 = .52$, $\lambda_2 = -.19$	$\lambda_1 = .92$, $\lambda_2 = .15$	$\lambda_1 = .995 + .042i$, $\lambda_2 = .995 - .042i$	
ℓ_6 in (4.21)	$\begin{bmatrix} .65 & 1.67 \\ (.19) & (.40) \\ -.09 & -.38 \\ (.25) & (.74) \end{bmatrix}$	$\begin{bmatrix} 1.16 & .76 \\ (.16) & (.43) \\ .27 & .09 \\ (.25) & (.72) \end{bmatrix}$	$\begin{bmatrix} .80 & -.01 \\ (.07) & (.11) \\ .08 & .69 \\ (.05) & (.08) \end{bmatrix}$	$\begin{bmatrix} 36.81 & 7.09 \\ 7.09 & 15.58 \end{bmatrix}$
	$\lambda_1 = .47$, $\lambda_2 = -.20$	$\lambda_1 = .96$, $\lambda_2 = .10$	$\lambda_1 = .79$, $\lambda_2 = .70$	

starts $z_{1(t-1)}$ but also on the previous month's houses sold $z_{2(t-1)}$. It is clear that, apart from the seasonal effect, the z_{1t} series is nonstationary because of its dependence on z_{2t}, and there is really only one nonstationary component between these two series. Also, by comparing the fitting results between the individual model and the joint model, given in tabulation in Section 5.1 (under E) and Table 2 (under ℓ_{1*}), we see that the fits for the z_{2t} series are comparable, but there is a 30 percent reduction in the estimated residual variance of z_{1t} for the joint model, simply because of the additional information supplied by $z_{2(t-1)}$.

These data have illustrated a number of important points:

1. Improvement in residual variance can be achieved by considering series jointly.
2. Actual examples occur in which zeros of $|\boldsymbol{\theta}_q(B)|$ lie either on or near the unit circle.
3. Structure of this kind can be well estimated by using ℓ_{1*} in (4.20) but not when ℓ_0 in (4.21) is used.

APPENDIX—COMPUTATION OF THE MAXIMUM LIKELIHOOD ESTIMATES

We here sketch the key elements of a computer program to calculate the maximum likelihood estimates of parameters corresponding to (2.6a).

First, we observe that the exponent of (2.6a) can be expressed as

$$\mathbf{Y}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{Y} - \hat{\mathbf{b}}'\mathbf{X}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{X}\hat{\mathbf{b}} = \sum_{t=1-q}^{n} \hat{\mathbf{a}}'_t\boldsymbol{\Sigma}^{-1}\hat{\mathbf{a}}_t , \quad \text{(A.1)}$$

where $\hat{\mathbf{a}}_t = \mathbf{z}_t + \boldsymbol{\theta}_1\hat{\mathbf{a}}_{t-1} + \ldots + \boldsymbol{\theta}_q\hat{\mathbf{a}}_{t-q}$, $t = 1, \ldots, n$, and $(\hat{\mathbf{a}}'_{-q+1}, \ldots, \hat{\mathbf{a}}'_0) = \hat{\mathbf{a}}'_*$. Second, letting $\boldsymbol{\Sigma}^{-1} = \mathbf{H}'\mathbf{H}$ and $\mathbf{u}_t = \mathbf{H}\hat{\mathbf{a}}_t = (u_{1t}, \ldots, u_{kt})'$, we can write $\sum_{t=1-q}^{n} \hat{\mathbf{a}}'_t\boldsymbol{\Sigma}^{-1}\hat{\mathbf{a}}_t = \sum_{t=1-q}^{n} \sum_{j=1}^{k} u_{jt}^2$. Third, it can be readily shown that $|\boldsymbol{\Sigma}|^q|\mathbf{X}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{X}| \geq 1$. This, the likelihood function can be written as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\Sigma} | \mathbf{Z}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}n} \exp\left[-\tfrac{1}{2}S(\boldsymbol{\theta}, \boldsymbol{\Sigma})\right] , \quad \text{(A.2)}$$

where $S(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = u_0^2 + \sum_{t=1-q}^{n} \sum_{j=1}^{k} u_{jt}^2$ and

$$u_0^2 = \ln|\boldsymbol{\Sigma}|^q|\mathbf{X}'\boldsymbol{\Sigma}_{n+q}^{-1}\mathbf{X}| .$$

The program computes estimates of $(\boldsymbol{\Sigma}, \boldsymbol{\theta})$ iteratively as follows: Starting with an initial estimate $\boldsymbol{\theta}^{(0)}$ of $\boldsymbol{\theta}$, an estimate of $\boldsymbol{\Sigma}$ is formed $\boldsymbol{\Sigma}^{(0)} = (1/n) \sum_{t=1-q}^{n} \hat{\mathbf{a}}_t\hat{\mathbf{a}}'_t$. A standard nonlinear least squares routine is then employed to find the value $\boldsymbol{\theta}^{(1)}$ that minimizes $S(\boldsymbol{\theta}, \boldsymbol{\Sigma}^{(0)})$. A new estimate $\boldsymbol{\Sigma}^{(1)}$ based on $\boldsymbol{\theta}^{(1)}$ is found, and the process is repeated until convergence is achieved, that is, when the parameter changes are sufficiently small. Extension of the procedure to (4.20) is straightforward.

Strictly speaking, because of the term u_0^2 in $S(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, the parameter estimates so obtained are not the exact maximum likelihood estimates, unless $\boldsymbol{\Sigma}$ is known. We chose this procedure mainly because of the ready availability of, and our familiarity with, the nonlinear least squares routine. Alternative procedures that directly maximize the function (2.6a) are being investigated. The program also computes parameter estimates corresponding to the conditional likelihood (2.7), and this is done simply by setting u_0 and the elements of $\hat{\mathbf{a}}_0, \ldots, \hat{\mathbf{a}}_{-q+1}$ equal to zero in the preceding procedure. The complexity of the exact likelihood arises chiefly from the need to compute $\hat{\mathbf{a}}_*$ and u_0^2. Based on our limited experience, the central processing unit (CPU) time for the exact likelihood estimates is about 2.5 times that for the conditional estimates. The CPU time itself depends mainly on the total number of observations and the order of the moving average polynomial $\boldsymbol{\theta}_q(B)$. For two series, with $q \leq 2$ and $n \leq 160$, the longest CPU time on a UNIVAC 1110 computer we have encountered is 57 seconds for the exact and 36 seconds for the conditional likelihood. For three series, $q = 2$ and $n = 150$, the CPU time can be as high as 260 seconds for the exact method. In most cases considered, convergence is obtained within 10 iterations.

[Received December 1977. Revised October 1978.]

REFERENCES

Abraham, B. (1975), "Linear Models, Time Series and Outliers," unpublished PhD thesis, Dept. of Statistics, University of Wisconsin—Madison.

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

Box, G.E.P., Hillmer, S.C., and Tiao, G.C. (1976), "Analysis and Modelling of Seasonal Time Series," paper presented at NBER-CENSUS Conference on Seasonal Time Series, Washington, D.C.

———, and Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

———, and Muller, M.E. (1958), "A Note on the Generation of Random Normal Deviates," *Annals of Mathematical Statistics*, 29, 610–611.

Dent, Warren (1977), "Computation of the Exact Likelihood Function of an ARIMA Process," *Journal of Statistics, Computation and Simulation*, 5, 193–206.

Dunsmuir, W., and Hannan, E.J. (1976), "Vector Linear Time Series Models," *Journal of Applied Probability*, 10, 130–145.

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

——— (1970), *Multiple Time Series*, New York: John Wiley & Sons.

Ljung, G.M., and Box, G.E.P. (1976), "Studies in the Modelling of Discrete Time Series—Maximum Likelihood Estimation in the Autoregressive Moving Average Model," Technical Report No. 476, Dept. of Statistics, University of Wisconsin—Madison.

Newbold, P. (1974), "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process," *Biometrika*, 61, 423–426.

Nicholls, D.F. (1976), "The Efficient Estimation of Vector Linear Time Series Models," *Biometrika*, 63, 381–390.

Osborn, D.R. (1977), "Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes," *Journal of the Royal Statistical Society*, Ser. B, 39, 114–118.

Phadke, M.S., and Kedem, G. (1976), "Maximum Likelihood Estimation of Multivariate Autoregressive-Moving Average Models," paper presented at the annual meeting of the American Statistical Association, August 23–26, 1976.

Reinsel, G. (1976), "A Note on Approximations for the Inverse of the Covariance Matrix of a Stationary Vector Process," Technical Report No. 470, Dept. of Statistics, University of Wisconsin—Madison.

Tiao, G.C., and Ali, M.M. (1971), "Analysis of Correlated Random Effects: Linear Model With Two Random Components," *Biometrika*, 58, 37–51.

Wilson, G.T. (1973), "The Estimation of Parameters in Multivariate Time Series Models," *Journal of the Royal Statistical Society*, Ser. B, 35, 76–85.