Modeling Multiple Time Series With Applications

G. C. TIAO and G. E. P. BOX*

An approach to the modeling and analysis of multiple time series is proposed. Properties of a class of vector autoregressive moving average models are discussed. Modeling procedures consisting of tentative specification, estimation, and diagnostic checking are outlined and illustrated by three real examples.

KEY WORDS: Multiple time series; Vector autoregressive moving average models; Cross-correlations; Partial autoregression; Intervention analysis; Transfer function.

1. INTRODUCTION

Business, economic, engineering and environmental data are often collected in roughly equally spaced time intervals, for example, hour, week, month, or quarter. In many problems, such time series data may be available on several related variables of interest. Two of the reasons for analyzing and modeling such series jointly are

1. To understand the dynamic relationships among them. They may be contemporaneously related, one series may lead the others or there may be feedback relationships.
2. To improve accuracy of forecasts. When there is information on one series contained in the historical data of another, better forecasts can result when the series are modeled jointly.

Let

$$\{Z_{1t}\}, \ldots, \{Z_{kt}\}, \quad t = 0, \pm 1, \pm 2, \ldots \tag{1.1}$$

be k series taken in equally spaced time intervals. Writing

$$\mathbf{Z}_t = (Z_{1t}, \ldots, Z_{kt})', \tag{1.2}$$

we shall refer to the k series as a k-dimensional vector of multiple time series. Models that are of possible use in representing such multiple time series, considerations of their properties, and methods for relating them to actual data have been extensively discussed in the literature. See in particular Quenouille (1957), Whittle (1963), Hannan (1970), Zellner and Palm (1974), Brillinger (1975), Dunsmuir and Hannan (1976), Box and Haugh (1977), Granger and Newbold (1977), Parzen (1977), Wallis (1977), Chan and Wallis (1978), Deistler, Dunsmuir, and Hannan (1978), Hallin (1978), Jenkins (1979), Hsiao (1979), Akaike (1980), Hannan (1980), Hannan, Dunsmuir, and Deistler (1980), and Quinn (1980). There are, however, considerable divergences of view. The object of this article is to describe an approach to the modeling and analysis that we have developed over a considerable period of time and that we are finding effective. Our main emphasis will be on motivating, describing, and illustrating the various methods used in an iterative model building process. Much, if not all, of the underlying theory can be found in the references given and, therefore, will not be repeated. Section 2 presents a short review of the widely used univariate ($k = 1$) time series and transfer function models as developed in Box and Jenkins (1970). Section 3 discusses a class of vector autoregressive moving average models. Model building procedures are discussed in Section 4 and applied to two actual examples in Section 5. A comparison with some alternative approaches and some concluding remarks pertaining to the analysis of fitting results are given in Section 6.

2. UNIVARIATE TIME SERIES AND TRANSFER FUNCTION MODELS

When $k = 1$ we shall write $\mathbf{Z}_t = Z_t$ in (1.2). An important class of models for discrete univariate series originally proposed by Yule (1927) and Slutsky (1937) and developed by such authors as Bartlett, Kendall, Walker, Wold, and Yaglom are stochastic difference equations of the form

$$\phi_p(B) z_t = \theta_q(B) a_t, \tag{2.1}$$

where $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$. In (2.1) the a_t's are independently identically and normally distributed random shocks (or white noise) with zero mean and variance σ^2; B is the back-shift operator such that $BZ_t = Z_{t-1}$; and $z_t = Z_t - \eta$ is the deviation of the observation Z_t from some convenient location η.

Relationships between k series $\{z_{1t}\}, \ldots, \{z_{kt}\}$ can sometimes be represented by linear *transfer function* models of the form

$$z_{ht} = \sum_{i \in k(h)} [\omega_{s_{hi}}(B) B^{b_{hi}} / \delta_{r_{hi}}(B)]\, z_{it} + [\theta_{q_h}(B)/\varphi_{p_h}(B)]\, a_{ht}, \quad (h = 1, 2, \ldots, k) \tag{2.2}$$

where $z_{0t} \equiv 0$, $k(h)$ is the set $(1, \ldots, h - 1)$; $\omega_{s_{hi}}(B)$, $\delta_{r_{hi}}(B)$, $\varphi_{p_h}(B)$, and $\theta_{q_h}(B)$ are polynomials in B;

* G.C. Tiao is Professor of Statistics and Business, and G.E.P. Box is Vilas Research Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. The authors are grateful to W.R. Bell, I. Chang, M.R. Grupe, G.B. Hudak, and R.S. Tsay for computing assistance. This research was partially supported by the U.S. Bureau of the Census under JSA 80-10, the Army Research Office, Durham, NC under Grant No. DAAG29-78-G00166, and the Alcoa Foundation.

© Journal of the American Statistical Association
December 1981, Volume 76, Number 376
Theory and Methods Section

the b_{hi}'s are nonnegative integers; and $\{a_{1t}\}, \ldots, \{a_{kt}\}$ are k independent Gaussian white-noise processes with zero means and variances $\sigma_1^2, \ldots, \sigma_k^2$. In particular, intervention models of this form with one or more of the z_h's indicator variables have proved useful (Box and Tiao 1975; Abraham 1980).

Transfer function models of the form (2.2), however, assume that the series, when suitably arranged, possess a triangular relationship, implying for example that z_1 depends only on its own past; z_2 depends on its own past and on the present and past of z_1; z_3 on its own past and on the present and past of z_2 and z_1; and so on. On the other hand, if z_1 depends on the past of z_2, and also z_2 depends on the past of z_1, then we must have a model that allows for this *feedback*.

3. MULTIPLE STOCHASTIC DIFFERENCE EQUATION MODELS

3.1 The Vector ARMA Model

A useful class of models obtained by direct generalization of the Yule-Slutsky ARMA models that allow for feedback relationships among the k series is obtained from (2.2) by letting $k(h)$ be the set $(1, \ldots, k)$ excluding h. These models can be alternatively expressed as the vector autoregressive moving average ARMA models (Quenouille 1957),

$$\boldsymbol{\varphi}_p(B)\mathbf{z}_t = \boldsymbol{\theta}_q(B)\mathbf{a}_t, \tag{3.1}$$

where

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

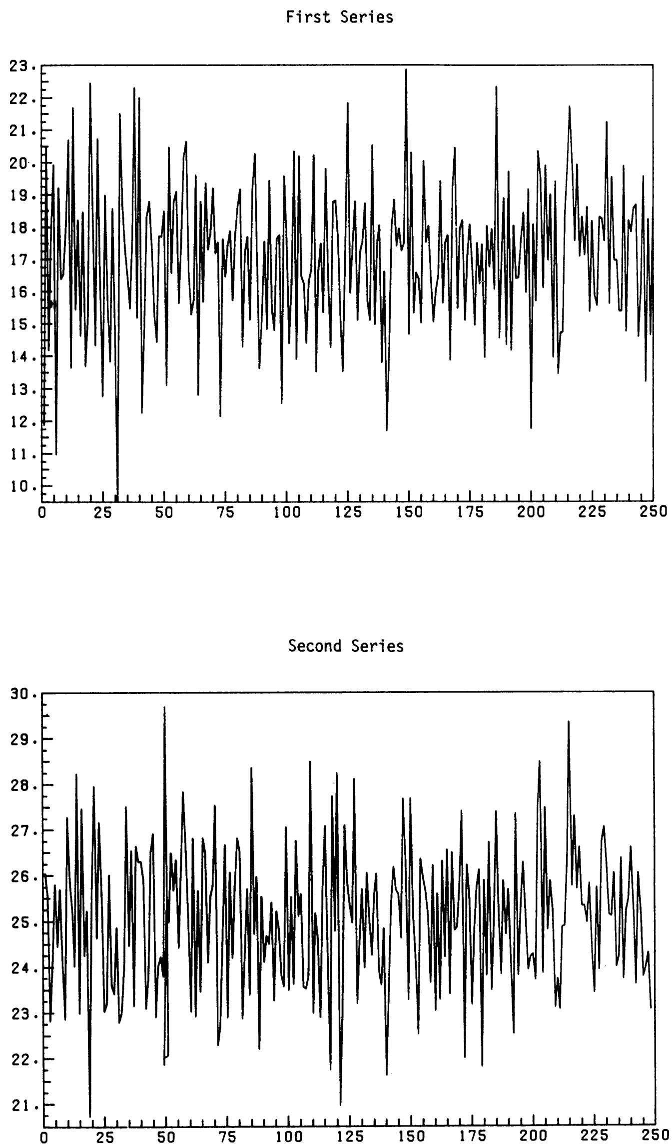
are matrix polynomials in B, the $\boldsymbol{\varphi}$'s and $\boldsymbol{\theta}$'s are $k \times k$ matrices, $\mathbf{z}_t = \mathbf{Z}_t - \boldsymbol{\eta}$ is the vector of deviations from some origin $\boldsymbol{\eta}$ that is the mean if the series is stationary, and $\{\mathbf{a}_t\}$ with $\mathbf{a}_t = (a_{1t}, \ldots, a_{kt})'$ is a sequence of random shock vectors identically independently and normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}$. We shall suppose that the zeros of the determinantal polynomials $|\boldsymbol{\varphi}_p(B)|$ and $|\boldsymbol{\theta}_q(B)|$ are on or outside the unit circle. The series $\mathbf{z}_t$ will be stationary when the zeros of $|\boldsymbol{\varphi}_p(B)|$ are all outside the unit circle, and will be invertible when those of $|\boldsymbol{\theta}_q(B)|$ are all outside the unit circle. Properties of such models have been discussed by, for example, Hannan (1970), Anderson (1971), and Granger and Newbold (1977).

Some Simple Examples. To illustrate the behavior of observations from these models, Figure 1 shows two series with 250 observations generated from the bivariate ($k = 2$) first order moving average [MA(1)] model, $\mathbf{z}_t = (\mathbf{I} - \boldsymbol{\theta}B)\mathbf{a}_t$, with

$$\boldsymbol{\theta} = \begin{bmatrix} .2 & .3 \\ -.6 & 1.1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}. \tag{3.2}$$

Figure 2 shows two series with 150 observations gener-

Figure 1. Data Generated From a Bivariate MA(1) Model With Parameter Values in (3.2)



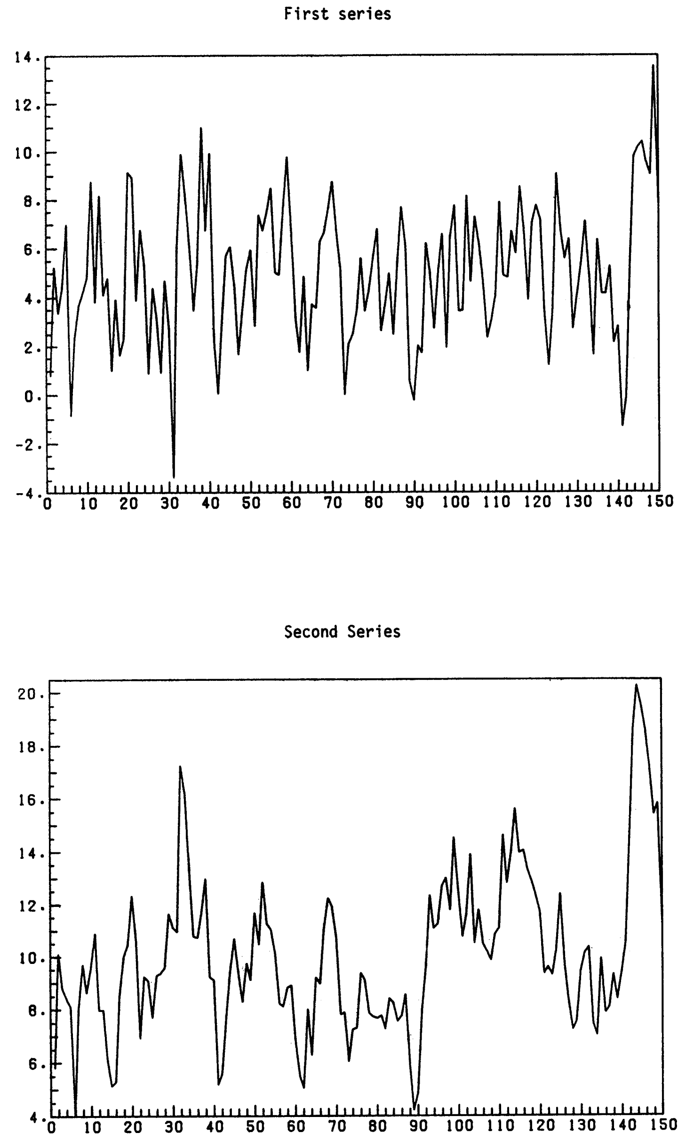
ated from the bivariate first order autoregressive [AR(1)] model, $(\mathbf{I} - \boldsymbol{\varphi}B)\mathbf{z}_t = \mathbf{a}_t$, with

$$\boldsymbol{\varphi} = \begin{bmatrix} .2 & .3 \\ -.6 & 1.1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}. \tag{3.3}$$

While in both cases the series are seen to be stationary, observations from the autoregressive model are seen to have more "momentum" than those from the moving average model.

In practice, time series often exhibit nonstationary behavior. When several such series are considered jointly, nonstationarity may be modeled by allowing the zeros of $|\boldsymbol{\varphi}(B)|$ in (3.1) to lie on the unit circle. A particular example is the model $(1 - B)\mathbf{z}_t = (\mathbf{I} - \boldsymbol{\theta}B)\mathbf{a}_t$, that is, after differencing each series we obtain a vector MA(1) model. This is a vector analog of the commonly used univariate nonstationary model $(1 - B)z_t = (1 - \theta B)a_t$. However,

Figure 2. Data Generated From a Bivariate AR(1) Model With Parameter Values in (3.3)



it should be noted here that for vector time series, linear combinations of the elements of $\mathbf{z}_t$ may often be stationary, and simultaneous differencing of all series can lead to unnecessary complications in model fitting. See, for example, the discussion in Box and Tiao (1977) and Hillmer and Tiao (1979).

Tranfer Function Model. For the vector model in (3.1), in general, all elements of $\mathbf{z}_t$ are related to all elements of $\mathbf{z}_{t-j}$ ($j = 1, 2, \ldots$) and there can be feedback relationships between all the series. However, if the $\mathbf{z}_t$'s can be arranged so that the coefficient matrices $\boldsymbol{\varphi}$'s and $\boldsymbol{\theta}$'s are all lower triangular, then (3.1) can be written as a transfer function model of the form (2.2). More generally, if the $\boldsymbol{\varphi}$'s and $\boldsymbol{\theta}$'s are all lower block triangular, then we obtain a generalization of the transfer function form of (2.2) in which both the input vector series and the output vector series are allowed to have feedback relationships. Furthermore, relationships between the vector transfer function model and the econometric linear simultaneous equation model have been discussed in Zellner and Palm (1974) and Wallis (1977).

3.2 Cross-Covariance and Cross-Correlation Matrices

For a stationary vector time series $\{\mathbf{Z}_t\}$ with mean vector $\boldsymbol{\eta}$, let $\boldsymbol{\Gamma}(l)$ be the lag l cross-covariance matrix

$$\begin{aligned}\boldsymbol{\Gamma}(l) &= E(\mathbf{z}_{t-l}\mathbf{z}'_t)\\ &= \{\gamma_{ij}(l)\}, \quad l = 0, \pm 1, \pm 2, \ldots \\ &\qquad\qquad i, j = 1, \ldots, k\end{aligned} \tag{3.4}$$

and let $\boldsymbol{\rho}(l) = \{\rho_{ij}(l)\}$ be the corresponding cross-correlation matrix.

When the vector ARMA model in (3.1) is stationary, it is well known that

$$\boldsymbol{\Gamma}(l) = \begin{cases} \displaystyle\sum_{j=l-r}^{l-1} \boldsymbol{\Gamma}(j)\boldsymbol{\varphi}'_{l-j} - \sum_{j=0}^{r-l} \boldsymbol{\psi}_j \boldsymbol{\Sigma} \boldsymbol{\theta}'_{j+l}, & l = 0, \ldots, r \\ \displaystyle\sum_{j=1}^{r} \boldsymbol{\Gamma}(l-j)\boldsymbol{\varphi}_j', & l > r, \end{cases} \tag{3.5}$$

where the $\boldsymbol{\psi}_j$'s are obtained from the relationship

$$\boldsymbol{\psi}(B) = \boldsymbol{\varphi}^{-1}(B)\boldsymbol{\theta}(B) = (\mathbf{I} + \boldsymbol{\psi}_1 B + \cdots),$$

$\boldsymbol{\theta}_0 = -\mathbf{I}$, $r = \max(p, q)$, and it is understood that (a) if $p < q$, $\boldsymbol{\varphi}_{p+1} = \cdots = \boldsymbol{\varphi}_r = \mathbf{0}$, and (b) if $q < p$, $\boldsymbol{\theta}_{q+1} = \cdots = \boldsymbol{\theta}_r = \mathbf{0}$.

In particular, when $p = 0$, that is, we have a vector MA(q) model, then

$$\boldsymbol{\Gamma}(l) = \begin{cases} \displaystyle\sum_{j=0}^{q-l} \boldsymbol{\theta}_j \boldsymbol{\Sigma} \boldsymbol{\theta}'_{j+l}, & l = 0, \ldots, q \\ \mathbf{0} \quad l > q. \end{cases} \tag{3.6}$$

Thus, all auto- and cross-correlations are zero when $l > q$. On the other hand, for a vector autoregressive model the auto- and cross-correlations in general will decay gradually to zero as $|l|$ increases.

3.3 A Determinantal Criterion for ARMA Models and the Partial Autoregression Matrices

From the moment equations in (3.5) for a stationary ARMA (p, q) model, we see that the autocovariance matrices $\boldsymbol{\Gamma}(l)$'s and the autoregressive coefficient matrices $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_p$ are related as follows:

$$\begin{bmatrix} \mathbf{A}(p, m) & \mathbf{b}(p, m) \\ \mathbf{g}'(p, m) & \boldsymbol{\Gamma}(m) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{p-1} \\ \boldsymbol{\varphi}'_p \end{bmatrix} = \begin{bmatrix} \mathbf{c}(p, m) \\ \boldsymbol{\Gamma}(p+m) \end{bmatrix}, \quad m = q, q+1, \ldots, \tag{3.7}$$

where

$$\mathbf{A}(p, m) = \begin{bmatrix} \boldsymbol{\Gamma}(m)\ \boldsymbol{\Gamma}(m-1) & \cdots & \boldsymbol{\Gamma}(m-p+2) \\ \boldsymbol{\Gamma}(m+1) & \ddots & \vdots \\ \vdots & \ddots & \boldsymbol{\Gamma}(m-1) \\ \boldsymbol{\Gamma}(m+p-2) \ldots \boldsymbol{\Gamma}(m+1) & & \boldsymbol{\Gamma}(m) \end{bmatrix},$$

$$\mathbf{b}(p, m) = \begin{bmatrix} \boldsymbol{\Gamma}(m-p+1 \\ \vdots \\ \boldsymbol{\Gamma}(m-1) \end{bmatrix},$$

$$\mathbf{c}(p, m) = \begin{bmatrix} \boldsymbol{\Gamma}(m+1) \\ \vdots \\ \boldsymbol{\Gamma}(m+p-1) \end{bmatrix},$$

$\mathbf{g}'(p, m) = [\boldsymbol{\Gamma}(m+p-1), \ldots, \boldsymbol{\Gamma}(m+1)]$, and $\boldsymbol{\Phi}'_{p-1} = [\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{p-1}]$. Consider now the $k \times k$ matrix

$$\mathbf{D}(l, m) = [d_{ij}(l, m)] \quad \begin{matrix} l = 1, 2, \ldots \\ m = 0, 1, \ldots, \end{matrix} \tag{3.8}$$

where $d_{ij}(l, m)$ is the determinant

$$d_{ij}(l, m) = \det \begin{bmatrix} \mathbf{A}(l, m) & \mathbf{c}_j(l, m) \\ \mathbf{g}'_i(l, m) & \gamma_{ij}(l+m) \end{bmatrix}, \quad i, j = 1, \ldots, k$$

$\mathbf{c}_j(l, m)$ is the jth column of $\mathbf{c}(l, m)$, $\mathbf{g}'_i(l, m)$ is the ith row of $\mathbf{g}'(l, m)$, and $\gamma_{ij}(l+m)$ is the (i, j)th element of $\boldsymbol{\Gamma}(l+m)$. It follows from (3.7) that for an ARMA (p, q) model

$$\mathbf{D}(l, m) = \mathbf{0} \quad \text{for} \quad l > p \quad \text{and} \quad m \geq q. \tag{3.9}$$

This provides a multivariate generalization of the results in Gray, Kelley, and McIntire (1978) for univariate ARMA models.

In the special case $m = q = 0$, (3.7) is a multivariate generalization of the Yule-Walker equations for autoregressive models in univariate time series. Analogous to the partial autocorrelation function for the univariate case, we may define a *partial autoregression matrix function* $\underset{\sim}{\mathcal{P}}(l)$ having the property that if the model is AR(p), then

$$\underset{\sim}{\mathcal{P}}(l) = \begin{cases} \boldsymbol{\varphi}_l, & l = p \\ \mathbf{0}, & l > p \end{cases}. \tag{3.10}$$

From (3.7), we define $\underset{\sim}{\mathcal{P}}(l)$ as

$$\underset{\sim}{\mathcal{P}}'(l) = \begin{cases} \boldsymbol{\Gamma}^{-1}(0)\boldsymbol{\Gamma}(1), \ l = 1 \\ [\boldsymbol{\Gamma}(0) - \mathbf{b}'(l, 0)\mathbf{A}^{-1}(l, 0)\mathbf{b}(l, 0)]^{-1} \\ [\boldsymbol{\Gamma}(l) - \mathbf{b}'(l, 0)\mathbf{A}^{-1}(l, 0)\mathbf{c}(l, 0)], \ l > 1. \end{cases} \tag{3.11}$$

4. MODEL BUILDING STRATEGY FOR MULTIPLE TIME SERIES

The models in (3.1) contain a dauntingly large number $\{k^2(p + q) + \frac{1}{2}k(k + 1)\}$ of parameters, complicating methods for model building. It is natural that attempts have been made to simplify the general form in the model building process, for example by Granger and Newbold (1977) and Wallis (1977). While we sympathize with this aspiration, we feel that so far at least these attempts have not been successful. In some comparisons made later in Section 6, we argue that they do not result in genuine simplification, nor do they provide feasible methods when k is greater than 2 or 3. We see no alternative but to provide for direct initial fitting of models of the form (3.1). It must, however, be added

1. that often models of rather low order (p and q small) provide adequate approximation,
2. that occasionally knowledge of the system might allow simplification a priori, although even here prudent checking of the adequacy of the simplifcation would be necessary (see Zellner and Palm 1974),
3. that considerable simplification is almost invariably possible after an initial model has been fitted,
4. that 2 and 3 imply that provision should be made to allow models to be fitted in which certain parameters are fixed or constrained in some other way,
5. that other methods of seeking simplifications, for example principal component analysis or canonical analysis (see Box and Tiao 1977), will often prove effective.

In brief, we feel that although the full form (3.1) needs to be fitted initially, subsequent iterations will usually lead to simplification.

In what follows we sketch an iterative approach consisting of (a) tentative specification (identification), (b) estimation, and (c) diagnostic checking for the vector ARMA models in (3.1). A computer package to carry out this analysis has been completed (Tiao et al. 1979) consisting of three main programs: (a) Preliminary Analysis, (b) Stepwise Autoregression, and (c) Estimation and Forecasting.

4.1 Tentative Specification

The aim here is to employ statistics (a) that can be readily calculated from the data and (b) that facilitate the choice of subclass of models worthy of further examination.

Sample Cross-Correlations. The sample cross-correlations $\hat{\rho}_{ij}(l)$,

$$\hat{\rho}_{ij}(l) = \sum(Z_{it} - \bar{Z}_i)(Z_{j(t+l)} - \bar{Z}_j) / \{\sum(Z_{it} - \bar{Z}_i)^2 \sum(Z_{jt} - \bar{Z}_j)^2\}^{1/2}$$

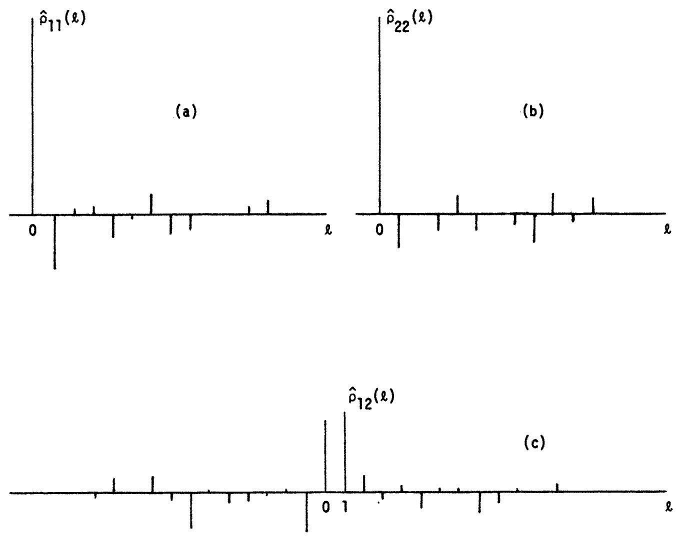
where $\bar{Z}_i$ is the sample mean of the ith component series of $\mathbf{Z}_t$, are particularly useful in spotting low order vector moving average models, since from (3.6) $\rho_{ij}(l) = 0$ for $l > q$.

For the data shown in Figure 1, which were generated from a bivariate MA(1) model, Figures 3(a)–(c) show, respectively, the sample autocorrelations $\hat{\rho}_{11}(l)$ and $\hat{\rho}_{22}(l)$, and the sample cross-correlations $\hat{\rho}_{12}(l)$. The large values occurring at $|l| = 1$ would lead to tentative specification of the model as an MA(1). However, graphs of this kind become increasingly cumbersome as the number of series is increased. Furthermore, identification is not easy from a listing of sample cross-correlation matrices $\hat{\boldsymbol{\rho}}(l)$ like that in Table 1(a), particularly when k is greater than 4 or 5.

In this circumstance, we have found the following simple device of great practical value. Instead of the numerical values, a plus sign is used to indicate a value greater than $2n^{-1/2}$, a minus sign a value less than $-2n^{-1/2}$, and a dot to indicate a value inbetween $-2n^{-1/2}$ and $2n^{-1/2}$. The motivation is that if the series were white noise, for large n the $\hat{\rho}_{ij}(l)$'s would be normally distributed with mean 0 and variance n^{-1}. The symbols can be arranged either as in Table 1(b) or as in Table 1(c). We realize that the variances of the $\hat{\rho}_{ij}(l)$'s can be considerably greater than $n^{-1/2}$ when the series are highly autocorrelated, so that these indicator symbols, if taken literally, can lead to overparameterization. However, we do not interpret these indicator symbols in the sense of a formal significance test, but as a rather crude "signal-to-noise ratio" guide. Taken together they can give useful and assimilable indicators of the general correlation pattern.

Table 2 shows sample cross-correlation matrices in terms of these indicator symbols for the series in Figure 2 generated from an AR(1) model. The persistence of large correlations suggests the possibility of autoregressive behavior. In general, the pattern of indicator symbols for the cross-correlation matrices makes it very easy to identify a low order moving average model.

Figure 3. Sample Auto- and Cross-Correlations for the Data in Figure 1



Sample Partial Autoregression and Related Summary Statistics. For an AR(p) process, the partial autoregression matrices $\mathcal{P}(l)$ in (3.11) are zero for $l > p$. They are therefore particularly useful for identifying an autoregressive model. Estimates of $\mathcal{P}(l)$ and their standard er-

Table 1. Cross-Correlations Matrices $\hat{\boldsymbol{\rho}}$ (l) for the Data in Figure 1

(a) Sample cross-correlation matrices $\hat{\boldsymbol{\rho}}(l)$ for the data in Figure 1

Lag 1-6

$$\begin{bmatrix} -.28 & .37 \\ -.21 & -.19 \end{bmatrix} \begin{bmatrix} .03 & .08 \\ .02 & .01 \end{bmatrix} \begin{bmatrix} .04 & -.03 \\ -.01 & -.08 \end{bmatrix} \begin{bmatrix} -.11 & .04 \\ -.03 & .09 \end{bmatrix} \begin{bmatrix} -.02 & -.09 \\ -.02 & -.08 \end{bmatrix} \begin{bmatrix} .10 & .01 \\ .01 & -.00 \end{bmatrix}$$

Lag 7-12

$$\begin{bmatrix} -.11 & .01 \\ -.17 & -.06 \end{bmatrix} \begin{bmatrix} -.09 & -.12 \\ -.03 & -.16 \end{bmatrix} \begin{bmatrix} .01 & -.06 \\ .08 & .10 \end{bmatrix} \begin{bmatrix} -.00 & .02 \\ .01 & -.04 \end{bmatrix} \begin{bmatrix} .03 & .00 \\ .08 & .08 \end{bmatrix} \begin{bmatrix} .06 & .04 \\ -.01 & .01 \end{bmatrix}$$

(b) $\hat{\boldsymbol{\rho}}(l)$ in term of indicator symbols

Lag 1-6

$$\begin{bmatrix} - & + \\ - & - \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

Lag 7-12

$$\begin{bmatrix} \cdot & \cdot \\ - & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & - \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

(c) Pattern of correlations for each element in the matrix over all lags

	z_1	z_2
z_1	−	+ . . . .
z_2	− . . . . . −	− . . . . . . − . . . .

Table 2. Sample Cross-Correlation Matrices $\hat{\boldsymbol{\rho}}(l)$ for the Data in Figure 2 in Terms of Indicator Symbols

Lag 1-6

$$\begin{bmatrix} + & \cdot \\ + & + \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ + & + \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ + & + \end{bmatrix}\begin{bmatrix} \cdot & - \\ + & + \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ + & + \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ + & \cdot \end{bmatrix}$$

Lag 7-12

$$\begin{bmatrix} - & - \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} - & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

rors can be obtained by fitting autoregressive models of successively high order $l = 1, 2, \ldots$ by standard multivariate least squares.

It is well known (see, e.g., Anderson 1971) that for a stationary AR(p) model asymptotically the estimates $\varphi'_1, \ldots, \varphi'_p$ are jointly normally distributed. A useful summary of the pattern of the partials is obtained by listing indicator symbols, assigning a plus (minus) sign when a coefficient in $\hat{\boldsymbol{\Phi}}(l)$ is greater (less) than 2 (-2) times its estimated standard errors, and a dot for values in between.

To help tentatively determine the order of an autoregressive model, we may also employ the likelihood ratio statistics corresponding to testing the null hypotheses $\boldsymbol{\varphi}_l = \mathbf{0}$ against the alternative $\boldsymbol{\varphi}_l \neq \mathbf{0}$ when an AR(l) model is fitted. Let

$$\mathbf{S}(l) = \sum_{t=l+1}^{n} (\mathbf{z}_t - \hat{\boldsymbol{\varphi}}_1\mathbf{z}_{t-1} - \ldots - \hat{\boldsymbol{\varphi}}_l\mathbf{z}_{t-l}) \times (\mathbf{z}_t - \hat{\boldsymbol{\varphi}}_1\mathbf{z}_{t-1} - \ldots - \hat{\boldsymbol{\varphi}}_l\mathbf{z}_{t-l})' \tag{4.1}$$

Table 3. Indicator Symbols for Partial Autoregression and Related Statistics for Data in Figure 2

Lag l	*Indicator symbols*	*$M(l)^a \overset{a}{\sim} \chi^2_4$*	*Diagonal elements of $\hat{\boldsymbol{\Sigma}}$*
1	$\cdot \quad +$	356.96	5.30
	$- \quad +$		1.08
2	$\cdot \quad \cdot$	7.04	5.16
	$\cdot \quad +$		1.03
3	$\cdot \quad \cdot$	2.63	5.07
	$\cdot \quad \cdot$		1.03
4	$\cdot \quad \cdot$	4.38	5.01
	$\cdot \quad \cdot$		1.02
5	$\cdot \quad \cdot$	2.42	4.95
	$\cdot \quad \cdot$		1.01

[a] $\overset{a}{\sim}$ means approximately distributed as.

be the matrix of residual sum of squares and cross products after fitting an AR(l). The likelihood ratio statistic is the ratio of the determinants

$$U = |\mathbf{S}(l)| / |\mathbf{S}(l-1)|. \tag{4.2}$$

Using Bartlett's (1938) approximation, the statistic

$$M(l) = -(N - \tfrac{1}{2} - l \cdot k)\log_e U \tag{4.3}$$

is, on the null hypothesis, asymptotically distributed as χ^2 with k^2 degrees of freedom, where $N = n - p - 1$ is the effective number of observations, assuming that a constant term is included in the model.

Finally, a measure of the extent to which the fit is improved as the order is increased is provided by the diagonal elements of the residual covariance matrices $\hat{\boldsymbol{\Sigma}}$ corresponding to the successive AR models.

For illustration, the matrices of summary symbols, the $M(l)$ statistics, and the diagonal elements of the residual covariance matrices for the series in Figure 2 are shown in Table 3 for $l = 1, \ldots, 5$. They indicate that an AR(1) or at most an AR(2) would be adequate for the data.

For the series shown in Figure 1, the pattern of the partials and related statistics are given in Table 4. Notice here that if we had confined attention to autoregressive models as is advocated in Parzen (1977), we would have needed p to be as high as 7. This is not surprising since with the MA(1) model of (3.2) written in the autoregressive form $\mathbf{z}_t = \boldsymbol{\pi}_1\mathbf{z}_{t-1} + \boldsymbol{\pi}_2\mathbf{z}_{t-2} + \ldots + \mathbf{a}_t$, we find

$$\boldsymbol{\pi}_1 = \begin{bmatrix} -.2 & -.3 \\ .6 & -1.1 \end{bmatrix}, \quad \boldsymbol{\pi}_2 = \begin{bmatrix} .14 & -.39 \\ .78 & -1.03 \end{bmatrix}, \ldots, \quad \boldsymbol{\pi}_6 = \begin{bmatrix} .23 & -.25 \\ .49 & -.51 \end{bmatrix}, \tag{4.4}$$

$$|\boldsymbol{\pi}_1| = .4, \quad |\boldsymbol{\pi}_2| = .16, \quad \ldots, \quad |\boldsymbol{\pi}_6| = .0041.$$

Thus, although the determinants $|\boldsymbol{\pi}_j|$ decrease rapidly towards zero as j increases, the elements of $\boldsymbol{\pi}_j$ converge to zero very slowly so that many autoregressive terms would be needed to provide an adequate approximation.

In general, the pattern of the partial autoregression matrices, the $M(l)$ statistic, and the diagonal elements of the residual covariance matrix are useful to distinguish between moving average and low order autoregressive models and to select tentatively the appropriate order for the latter.

Table 4. Pattern of Partial Autoregression and Related Statistics for Data in Figure 1

Lag	Pattern of $\hat{\mathcal{P}}(l)$		$M(l) \xrightarrow{\sim} \chi_4^2$	$\hat{\Sigma}$
1	−	−	123.2	4.78
	+	−		1.88
2	·	·	75.9	4.75
	+	−		1.43
3	+	·	35.2	4.63
	+	−		1.23
4	·	·	27.5	4.63
	+	·		1.08
5	·	·	16.6	4.61
	+	·		1.04
6	·	·	13.5	4.53
	+	·		.98
7	·	−	16.5	4.38
	+	·		.94
8	·	·	8.1	4.31
	·	−		.91

Sample Residual Cross-Correlation Matrices After AR Fit. After each AR(l) fit, $l = 1, \ldots, p$, cross-correlation matrices of the residuals $\hat{\mathbf{a}}_t$'s may be readily obtained. Table 5 shows indicator symbols for residual correlations after fitting AR(1) and AR(2) to the AR data plotted in Figure 2. Again a plus sign is used to indicate values greater than $2n^{-1/2}$, a minus sign for values less than $-2n^{-1/2}$, and a dot for in-between values. They verify that there is no need to go beyond an AR(2) model.

It is perhaps worth emphasizing here again that these indicator symbols are proposed as a rough preliminary device to help arrive at an initial model. They should not be treated as "exact significance testing." In a recent paper by Li and McLeod (1980), expressions have been obtained for the asymptotic distributions of the residual autocorrelations. As in the univariate case, the low order autocorrelations have variance considerably less than $n^{-1/2}$.

For mixed vector autoregressive moving average models in general, however, both the population cross-correlation matrices $\boldsymbol{\rho}(l)$ and the partial autoregression matrices $\mathcal{P}(l)$ decay only gradually toward $\mathbf{0}$. In some situations, the order of mixed models may be tentatively identified by inspection of patterns in residual cross-correlations after the AR fit, but in others study of residual correlations could be misleading. For illustration, consider the case of a stationary ARMA(1, 1) model

$$(\mathbf{I} - \boldsymbol{\varphi}B)\mathbf{z}_t = (\mathbf{I} - \boldsymbol{\theta}B)\mathbf{a}_t. \tag{4.5}$$

If an AR(1) model is fitted to $\{\mathbf{z}_t\}$, then the estimate $\hat{\boldsymbol{\varphi}}$ will be biased. In fact, asymptotically $\hat{\boldsymbol{\varphi}}$ converges in probability to

$$\hat{\boldsymbol{\varphi}} \to \boldsymbol{\varphi}_0 = \boldsymbol{\Gamma}'(1)\boldsymbol{\Gamma}(0)^{-1}. \tag{4.6}$$

Thus the residuals $\hat{\mathbf{a}}_t = \mathbf{z}_t - \boldsymbol{\varphi}_0\mathbf{z}_{t-1}$ approximately follow the model

$$\hat{\mathbf{a}}_t = (\mathbf{i} - \boldsymbol{\varphi}_0 B)(\mathbf{I} - \boldsymbol{\varphi}B)^{-1}(\mathbf{I} - \boldsymbol{\theta}B)\mathbf{a}_t. \tag{4.7}$$

Table 5. Indicator Symbols for Residual Cross Correlations for the AR (1) Data of Figure 2

AR (1) Lag 1-6

$$\begin{bmatrix} \cdot & \cdot \\ \cdot & - \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} + & \cdot \\ \cdot & \cdot \end{bmatrix}$$

Lag 7-12

$$\begin{bmatrix} - & \cdot \\ - & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

AR (2) Lag 1-6

$$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} + & \cdot \\ \cdot & \cdot \end{bmatrix}$$

Lag 7-12

$$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

For $k = 1$, $\{\hat{a}_t\}$ follows an ARMA(1, 2) model so that the autocorrelations of $\hat{a}_t$ are

$$\rho_{\hat{a}}(j) = \varphi\rho_{\hat{a}}(j - 1), j > 2 \quad (4.8)$$

and $\rho_{\hat{a}}(1)$ and $\rho_{\hat{a}}(2)$ are functions of φ and θ. Table 6 gives values of $\rho_{\hat{a}}(1)$ and $\rho_{\hat{a}}(2)$ for various combinations of values of φ and θ. For each combination, the first value is $\rho_{\hat{a}}(1)$ and the second $\rho_{\hat{a}}(2)$.

We see that if the true value of φ is large in magnitude, residual autocorrelations would lead to the choice of an MA(1) model for $\hat{a}_t$ and therefore the correct identification. For intermediate values of φ, a moving average of order 2 or higher might be selected, resulting in overparametrization.

In Gray, Kelley, and McIntire (1978) and Beguin, Gouricroux, and Monfort (1980), methods have been proposed to determine the order of univariate ARMA model. These methods are essentially equivalent to estimating, for $k = 1$, the determinant $D(l, m)$ in (3.8) using sample estimates of the autocovariances and selecting the orders of autoregressive and moving average polynomials on the basis of the property in (3.9). We are currently studying sampling properties of estimates of appropriate functions of $\mathbf{D}(l, m)$ in the vector case.

4.2 Estimation

Once the order of the model in (3.1) has been tentatively selected, efficient estimates of the associated parameter matrices $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_p)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_q)$, and $\boldsymbol{\Sigma}$ are determined by maximizing the likelihood function. Approximate standard errors and correlation matrix of the estimates of elements of the $\boldsymbol{\varphi}_j$'s and $\boldsymbol{\theta}_j$'s can also be obtained.

Conditional Likelihood. For the ARMA (p, q) model, we can write

$$\mathbf{a}_t = \mathbf{z}_t - \boldsymbol{\varphi}_1\mathbf{z}_{t-1} - \cdots - \boldsymbol{\varphi}_p\mathbf{z}_{t-p} + \boldsymbol{\theta}_1\mathbf{a}_{t-1} + \cdots + \boldsymbol{\theta}_q\mathbf{a}_{t-q}. \quad (4.9)$$

As in the univariate case discussed in Box and Jenkins (1970), the likelihood function can be approximated by a "conditional" likelihood function as follows. The series is regarded as consisting of the $n - p$ vector observations $\mathbf{z}_{p+1}, \ldots, \mathbf{z}_n$. The likelihood function is then determined from $\mathbf{a}_{p+1}, \ldots, \mathbf{a}_n$, using the preliminary values $\mathbf{z}_1, \ldots, \mathbf{z}_p$ and conditional on zero values for $\mathbf{a}_p, \ldots, \mathbf{a}_{p-q-1}$. Thus, as shown in Wilson (1973),

$$l_c(\boldsymbol{\varphi}, \boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{z}) \propto |\boldsymbol{\Sigma}|^{-(n-p)/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\varphi}, \boldsymbol{\theta})\}, \quad (4.10)$$

where $\mathbf{S}(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \sum_{t=p+1}^{n} \mathbf{a}_t\mathbf{a}'_t$. Properties of the maximum likelihood estimates obtained from (4.10) have been discussed in Nicholls (1976, 1977) and Anderson (1980).

It has been shown in Hillmer and Tiao (1979) that this approximation can be seriously inadequate if n is not sufficiently large and one or more zeros of $|\boldsymbol{\theta}_q(B)|$ lie on or close to the unit circle. Specifically, this would lead to estimates of the moving average parameters with large bias.

Exact Likelihood Function. For univariate ARMA models, the exact likelihood function has been considered by Tiao and Ali (1971), Newbold (1974), Dent (1977), Ansley (1979), and others. For vector models, this function has been studied by Osborn (1977) for the pure moving average case and by Phadke and Kedem (1978), Nicholls and Hall (1979), and Hillmer and Tiao (1979). It takes the form

$$l(\boldsymbol{\varphi}, \boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{z}) \propto l_c(\boldsymbol{\varphi}, \boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{z}) l_1(\boldsymbol{\varphi}, \boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{z}), \quad (4.11)$$

where l_1 depends (a) only on $\mathbf{z}_1, \ldots, \mathbf{z}_p$ if $q = 0$ and (b) on all the data vectors $\mathbf{z}_1, \ldots, \mathbf{z}_n$ if $q \neq 0$. Estimation algorithms have been developed and incorporated in our computer package for the vector MA(q) model. For the general ARMA(p, q) model, it has been shown that a close approximation to the exact likelihood can be obtained by considering the transformation

$$\mathbf{w}_t = (\mathbf{I} - \boldsymbol{\varphi}_1 B - \cdots - \boldsymbol{\varphi}_p B^p)\mathbf{z}_t$$

so that (4.12)

$$\mathbf{w}_t = \boldsymbol{\theta}_q(B)\mathbf{a}_t$$

and then applying the results for MA(q) to $\mathbf{w}_t$, $t = p + 1, \ldots, n$.

Because estimation of moving average parameters using the exact likelihood is rather slow, we presently employ the conditional method in the preliminary stages of iterative model building and switch to the exact method towards the end.

Table 6. Asymptotic Values of $\rho_{\hat{a}}$ (1) and $\rho_{\hat{a}}$ (2)

φ \ θ	−.95	−.50	.50	.95
−.95	—	.265 .085	−.381 −.03	−.481 −.036
−.50	.049 −.222	—	−.223 −.201	−.321 −.267
.50	.321 −.267	.223 −.201	—	−.049 −.222
.95	.481 −.036	.381 −.03	−.265 .085	—

4.3 Diagnostic Checking

To guard against model misspecification and to search for directions of improvement, a detailed diagnostic analysis of the residual series $\{\hat{\mathbf{a}}_t\}$, where

$$\hat{\mathbf{a}}_t = \mathbf{z}_t - \hat{\boldsymbol{\varphi}}_1\mathbf{z}_{t-1} - \cdots - \hat{\boldsymbol{\varphi}}_p\mathbf{z}_{t-p} + \hat{\boldsymbol{\theta}}_1\hat{\mathbf{a}}_{t-1} + \cdots + \hat{\boldsymbol{\theta}}_q\hat{\mathbf{a}}_{t-q}, \quad (4.13)$$

is performed. Useful diagnostic checks include (a) plots of standardized residual series against time and/or other variables and (b) cross-correlation matrices of the resid-

Table 7. Pattern of Sample Cross-Correlations for the SCC Data

	Z_1 Stocks	Z_2 Cars	Z_3 Commodities
Z_1 Stocks	+ + + + + + + + ...	− −	− −
Z_2 Cars	. . . . + + +	+ + + + + + + + + + + + + + + ...	+ + + + + + + + + + + + + + ...
Z_3 Commodities	− − − − − − − . . . . + + + + +	+ + + + + + + + + + + + + + + + + + . .	+ + + + + + + + + + + + + + + + + . . .

uals $\hat{\mathbf{a}}_t$. As before, the structures of the correlations are summarized by indicator symbols. Overall χ^2 tests based on the sample cross correlations of the residuals have been proposed in recent papers by Hosking (1980) and Li and McLeod (1980). However, as is noted in Box and Jenkins (1970), such overall tests are not substitutes for more detailed study of the correlation structure.

5. ANALYSES OF TWO EXAMPLES

We now apply the model building approach introduced in the preceding section to the following sets of data:

1. The Financial Time Ordinary Share Index, U.K. Car Production and the Financial Time Commodity Price Index: Quarterly Data 3/1952–4/1967, obtained from Coen, Gomme, and Kendall (1969). This will be referred to as the SCC data.
2. The Gas Furnace Data given in Box and Jenkins (1970).

5.1 The SCC Data

The three series are

Z_{1t}: Financial Time Ordinary Share Index
Z_{2t}: U.K. Car Production
Z_{3t}: Financial Time Commodity Price Index

The authors of the original study were interested in the possibility of predicting Z_{1t} from lagged values of Z_{2t} and Z_{3t} using a standard regression analysis in which Z_{1t} was treated as a dependent variable and $Z_{2(t-6)}$ and $Z_{3(t-7)}$ as regressors or independent variables. For a critical evaluation of this approach, see Box and Newbold (1970). Here we consider what structure is revealed by the present multiple time series analysis, in which the three series are jointly modeled.

Tentative Specification. We see in Table 7 that the original series show high and persistent auto- and cross-correlations. Examination of the partials and related statistics in Table 8 shows that for $l > 2$ most of the elements of $\mathcal{P}(l)$ are small compared with their estimated standard errors and the $M(l)$ statistic fails to show significant improvement. Table 9 shows that the pattern of the cross-correlations of the residuals after AR(2) is consonant with estimated white noise. However, note that there is one large residual correlation at lag 1 after the AR(1) fit, suggesting also the possibility of an ARMA(1, 1) model.

Estimation. Both an AR(2) and an ARMA(1, 1) model were fitted using the exact likelihood method* but results are given only for the ARMA(1, 1) model, which produced a marginally better representation. For this model,

$$(\mathbf{I} - \boldsymbol{\varphi} B)\mathbf{Z}_t = \boldsymbol{\theta}_0 + (\mathbf{I} - \boldsymbol{\theta} B)\mathbf{a}_t, \qquad (5.1)$$

where $\boldsymbol{\theta}_0$ is a vector of constants, Table 10 shows the initial unrestricted fit and also the fits for two simpler models obtained by setting to zero those coefficients whose estimates were small compared to their standard errors.

Diagnostic Checking. Table 11 suggests that the restricted ARMA(1, 1) model provides an adequate representation of the data.

Implication of the Model. The final model implies that the system is approximated by

$$(1 - .98B)Z_{1t} = a_{1t} \qquad (5.2a)$$

$$(1 - .93B)Z_{2t} = .2 + a_{2t} \qquad (5.2b)$$

$$(1 - .83B)Z_{3t} = 2.8 + .40a_{1(t-1)} + (1 + .41B)a_{3t}. \qquad (5.2c)$$

Upon substituting (5.2a) into (5.2c), we get

$$(1 - .83B)Z_{3t} = 2.8 + .40(1 - .98B)Z_{1(t-1)} + (1 + .41B)a_{3t}. \qquad (5.2d)$$

Thus all three series behave approximately as random walks with slightly correlated innovations. From the point of view of forecasting, (5.2d) is of some interest since it implies that ordinary share $Z_{1(t-1)}$ is a *leading indicator* at lag 1 for the commodity index Z_{3t}. Its effect is small, however, as can be seen for example by the improvement achieved over the corresponding best fitting univariate model, which was

$$(1 - .78B)Z_{3t} = 3.63 + (1 + .53B)a_t, \sigma^2 = .151 \qquad (5.3)$$

The residual variance of .151 from the univariate model is not much larger than the value .134 for a_{3t} obtained

* For this example, estimates from the conditional likelihood for the ARMA(1, 1) case are very close to the exact results.

Table 8. Partial Autoregression and Related Statistics: SCC Data

Lag	Indicator Symbols for Partials	$M(l)$ Statistic $\stackrel{a}{\sim} \chi_9^2$	Diagonal Elements of $\hat{\Sigma} \times 10$
	+ · ·		.44
1	· + ·	301.3	.89
	· · +		1.62
	− · ·		.40
2	· · ·	18.6	.84
	− + −		1.23
	· · ·		.37
3	· · ·	9.6	.81
	· · ·		1.21
	· · ·		.36
4	· · ·	3.6	.79
	· · ·		1.19
	· + ·		.32
5	· + ·	11.9	.70
	· · ·		1.11

from the final vector model. Although the multiple time series analysis fails to reveal anything very surprising for this example, it shows what is there and does not mislead.

5.2 The Gas Furnace Data

The two series consist of (a) input gas rate and (b) output as CO_2 concentration at 9-second intervals from a gas furnace. We shall let Z_{1t} = gas rate + .057 and Z_{2t} = CO_2 − 5.35. This set of data was employed in Box and Jenkins (1970) to illustrate a procedure of identification, fitting, and checking of a transfer function model of the form (2.3) for $k = 2$ relating two time series one of which is *known* to be input for the other. Using this approach, the following models were found for the input Z_{1t} and the output Z_{2t};

$$(1 - 1.97B + 1.37B^2 - .34B^3)Z_{1t} = a_{1t},$$
$$\hat{\sigma}^2_{a_1} = .0353 \quad (5.4a)$$

$$Z_{2t} = \frac{\omega(B)}{\delta(B)} B^b Z_{1t} + \varphi(B)^{-1} a_{2t}, \hat{\sigma}^2_{a_2} = .0561, \quad (5.4b)$$

Table 9. Pattern of Cross-Correlation Matrices of Residuals: SCC Data

Lag: 1	2	3	4	5	6	7	8
(a) AR(1) model							
· · +	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·
· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·
· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · +	· · ·
(b) AR(2) model							
· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · ·
· · ·	· · ·	· · ·	− · ·	· · ·	· · ·	· · ·	· · ·
· · ·	· · ·	· · ·	· · ·	· · ·	· · ·	· · +	· · ·

Table 10. Estimation Results for the Model (5.1): SCC Data (exact likelihood)

	$\hat{\theta}_0$	$\hat{\varphi}$	$\hat{\theta}$	$\hat{\Sigma}$
(1) Full Model	$\begin{bmatrix} 1.11 \\ (.64) \\ 1.74 \\ (.82) \\ 4.08 \\ (1.47) \end{bmatrix}$	$\begin{bmatrix} .81 & .15 & -.06 \\ (.08) & (.07) & (.04) \\ -.07 & .98 & -.09 \\ (.10) & (.10) & (.05) \\ -.32 & .30 & .76 \\ (.18) & (.17) & (.08) \end{bmatrix}$	$\begin{bmatrix} -.29 & .23 & .06 \\ (.15) & (.11) & (.07) \\ -.45 & .20 & -.15 \\ (.22) & (.17) & (.11) \\ -.79 & .57 & -.44 \\ (.28) & (.21) & (.13) \end{bmatrix}$	$\begin{bmatrix} .037 & & \\ .022 & .078 & \\ .013 & .022 & .129 \end{bmatrix}$
(2) Restricted Model (intermediate)	$\begin{bmatrix} .13 \\ (.09) \\ .59 \\ (.05) \\ 2.48 \\ (1.10) \end{bmatrix}$	$\begin{bmatrix} .90 & .08 & \cdot \\ (.06) & (.06) & \\ \cdot & .92 & -.02 \\ & (.04) & (.04) \\ \cdot & \cdot & .85 \\ & & (.07) \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -.40 & \cdot & -.41 \\ (.23) & & (.12) \end{bmatrix}$	$\begin{bmatrix} .042 & & \\ .022 & .079 & \\ .017 & .021 & .131 \end{bmatrix}$
(3) Restricted Model (final)	$\begin{bmatrix} .12 \\ (.08) \\ .24 \\ (.10) \\ 2.76 \\ (1.07) \end{bmatrix}$	$\begin{bmatrix} .98 & & \cdot \\ (.03) & & \\ \cdot & .93 & \cdot \\ & (.04) & \\ \cdot & \cdot & .83 \\ & & (.06) \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -.40 & \cdot & -.41 \\ (.23) & & (.12) \end{bmatrix}$	$\begin{bmatrix} .045 & & \\ .024 & .085 & \\ .019 & .023 & .134 \end{bmatrix}$

Table 11. Pattern of Residual Cross-Correlations After Final Restricted ARMA(1,1) Model Fit: SCC Data

	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$
$\hat{a}_1$	. . . . −.	. . . .	. . . .
$\hat{a}_2$	. . . .	. . . .	. . . .
$\hat{a}_3$	. . . .	. . . . −.	. . . .

where $\omega(B) = -(.53 + .37B + .51B^2)$, $\delta(B) = 1 - .57B$, $\varphi(B) = 1 - 1.53B + .63B^2$, and the $\{a_{1t}\}$ and $\{a_{2t}\}$ series are assumed independent.

Particularly when we are dealing with econometric rather than engineering models, feedback relationships may not be known a priori; it is of interest, therefore, to analyze the data using the present approach where no distinction is made between an input and output variable and the fact that no feedback could occur in the system is not used.

Tentative Specification. In Table 12, we see that the auto- and cross-correlations of the original data in part (a) are persistently large in magnitude, ruling out low order moving average models; the $M(l)$ statistic (χ^2_4) in part (b) suggests that an AR(6) model might be appropriate; and the residual cross correlation pattern after an AR(6) fit in part (c) seems to verify the appropriateness of this model.

Estimation Results. Estimation results corresponding to an unrestricted AR(6) model

$$(\mathbf{I} - \varphi_1 B - \cdots - \varphi_6 B^6)\mathbf{Z}_t = \mathbf{a}_t \tag{5.5}$$

are as follows:

$$\hat{\varphi}_1 = \begin{bmatrix} 1.93 & -.05 \\ (.06) & (.05) \\ .06 & 1.55 \\ (.08) & (.06) \end{bmatrix} \quad \hat{\varphi}_2 = \begin{bmatrix} -1.20 & .10 \\ (.13) & (.08) \\ -.14 & -.59 \\ (.16) & (.11) \end{bmatrix}$$

$$\hat{\varphi}_3 = \begin{bmatrix} .17 & -.08 \\ (.15) & (.09) \\ -.44 & -.17 \\ (.19) & (.11) \end{bmatrix} \quad \hat{\varphi}_4 = \begin{bmatrix} -.16 & .03 \\ (.15) & (.09) \\ .15 & .13 \\ (.19) & (.11) \end{bmatrix}$$

$$\hat{\varphi}_5 = \begin{bmatrix} .38 & -.04 \\ (.14) & (.08) \\ -.12 & .06 \\ (.18) & (.10) \end{bmatrix} \quad \hat{\varphi}_6 = \begin{bmatrix} -.22 & .03 \\ (.08) & (.03) \\ .25 & -.04 \\ (.11) & (.04) \end{bmatrix}$$

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} .0345 & \\ -.0023 & .0566 \end{bmatrix}, \quad \hat{\rho}(a_1, a_2) = .045 \tag{5.6}$$

If we let

$$\hat{\varphi}_l = \{\hat{\varphi}_{ij}^{(l)}\},$$

Table 12. Tentative Identification for the Gas Furnace Data

(a) Pattern of cross-correlations of the original data

	Z_{1t}	Z_{2t}
Z_{1t}	+ + + + + + + + + + + +	− − − − − − − − − − − −
Z_{2t}	− − − − − − − − − − . .	+ + + + + + + + + + + +

(b) M statistic for partial autoregression

Lag l	1	2	3	4	5	6	7	8	9	10	11
$M(l)$	1650	665	31.7	22.5	5.6	12.9	1.8	8.0	3.5	0	2.0

(c) Pattern of cross-correlations of the residuals after AR(6) fit

	$\hat{a}_{1t}$	$\hat{a}_{2t}$
$\hat{a}_{1t}$	. . . . −	. . . .
$\hat{a}_{2t}$	. . . .	. . . .

we see that $\hat{\varphi}_{12}^{(l)}$ are small compared with their standard errors over all lags, confirming (as in this case is known from the physical nature of the apparatus generating the data) that there is a unidirectional relationship between Z_{1t} and Z_{2t} involving no feedback. Also, $\hat{\varphi}_{21}^{(l)}$ is small for $l = 1$, 2, and the residuals $\hat{a}_{1t}$ and $\hat{a}_{2t}$ are essentially uncorrelated, implying a delay of 3 periods. It should be noted also that the variances for a_{1t} and a_{2t} are very close to those for a_{1t} and a_{2t} in (5.4), and their correlation is negligible.

To facilitate comparison with (5.4), we set $\varphi_{11}^{(l)} = 0$ for $l > 3$, $\varphi_{12}^{(l)} = 0$ for all l, $\varphi_{21}^{(l)} = 0$ for $l = 1$, 2, and $\varphi_{22}^{(l)} = 0$ for $l = 5, 6$. Estimation results for this restricted AR(6) model are then

$$\hat{\varphi}_1 = \begin{bmatrix} 1.98 & . \\ (.06) & \\ . & 1.53 \\ & (.06) \end{bmatrix} \quad \hat{\varphi}_2 = \begin{bmatrix} -1.38 & . \\ (.10) & \\ . & -.58 \\ & (.11) \end{bmatrix} \quad \hat{\varphi}_3 = \begin{bmatrix} .35 & . \\ (.06) & \\ -.53 & -.14 \\ (.07) & (.10) \end{bmatrix}$$

$$\hat{\varphi}_4 = \begin{bmatrix} . & . \\ .11 & .12 \\ (.16) & (.04) \end{bmatrix} \quad \hat{\varphi}_5 = \begin{bmatrix} . & . \\ -.04 & . \\ (.17) & \end{bmatrix} \quad \hat{\varphi}_6 = \begin{bmatrix} . & . \\ .21 & . \\ (.11) & \end{bmatrix}$$

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} .0359 & -.0029 \\ & .0561 \end{bmatrix}, \quad \hat{\rho}(a_1, a_2) \doteq 0 \tag{5.7}$$

Examination of the pattern of the cross-correlations of the residuals suggests that the model is adequate.

Implication of the Bivariate Model. The final AR(6) model (5.7) can be written

$$\begin{bmatrix} \varphi_{11}(B) & \\ \varphi_{21}(B) & \varphi_{22}(B) \end{bmatrix} \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \tag{5.8}$$

where $\varphi_{11}(B) = 1 - 1.98B + 1.38B^2 - .35B^3$, $\varphi_{21}(B) = (.53 - .11B - .21B)B^3$, and $\varphi_{22}(B) = (1 - 1.53B + .58B^2 + .14B^3 - .12B^4)$. Assuming a_{1t} and a_{2t} are uncorrelated, the input model $\varphi_{11}(B)Z_{1t} = a_{1t}$ with $\mathrm{Var}(a_{1t}) = .0359$ is essentially the same as (5.4a). Now the model relating the output Z_{2t} to the input Z_{1t} is

$$Z_{2t} = -\frac{\varphi_{21}(B)}{\varphi_{22}(B)} Z_{1t} + \frac{1}{\varphi_{22}(B)} a_{2t} \tag{5.9}$$

with $\mathrm{Var}(a_{2t}) = .0561$. The noise model $\varphi_{22}{}^{-1}(B)a_{2t}$ is not very different from the corresponding one $\varphi^{-1}(B)a_{2t}$ in (5.4b), but the dynamic model $-\varphi_{21}(B)\varphi_{22}{}^{-1}(B)Z_{1t}$ at first sight appears markedly different from the first term on the right side of (5.4b). The reason is that in the form (5.9) the denominators of the dynamic model and of the noise model are constrained to be identical. This restriction is not present in the transfer function model (5.4b). The less restrictive form can however be written in the form of (5.9) if we set $\varphi_{22}(B) = \varphi(B)$ and $-\varphi_{21}(B) = \omega(B)B^b\{\varphi(B)\delta^{-1}(B)\}$. For this example, the factor $\varphi(B)\delta^{-1}(B) \doteq 1 - .96B$, and it is then seen that the models are in fact very similar. This may be confirmed by comparing the impulse response weights in Table 13, where $\omega(B)B^b\delta^{-1}(B) = \sum_{j=0}^{\infty} v_j B^j$ and $-\varphi_{21}(B)\varphi_{22}{}^{-1}(B) = \sum_{j=0}^{\infty} v^*{}_j B^j$.

Further Analysis of Stepwise AR Results. It is instructive to examine for this data the changes in the fitted autoregressive models as the order is increased. Using indicator symbols (and omitting the dots) Table 14 shows the situation for $p = 1, \ldots, 6$. The residual covariance matrix for each order is also given. The following observations may be made.

1. If only AR(1) or AR(2) were considered, one might be led to believe mistakenly that there was a feedback relationship between these two series.
2. The unidirectional dynamic relationship becomes clear when the order of the model, p, is increased to three. Since the input series Z_{1t} essentially follows a univariate AR(3) model, this suggests that the present procedure will correctly identify the one-sided causal dynamic relationship once the input model is appropriately selected.
3. The delay $b = 3$ emerges when the order p is increased to 4. Since only very marginal improvement in the fit occurs for $p > 4$, this is saying that the delay is correctly identified only when the model is specified essentially correctly.

Implications on General Time Series Model Building. The relative merit of the present procedure and more direct modeling of the system will depend on how much is known or how much we are prepared to assume. In some applications, particularly in engineering and most examples of intervention analysis, an adequate initial specification may be possible from knowledge of the nature of the problem. This may allow a flow diagram showing the feedback structure to be drawn and likely orders to be guessed for the various dynamic components. The resulting models can then be directly *fitted* in the manner described and illustrated in Box and MacGregor (1974, 1976) and Box and Tiao (1975). For a single input with feedback known to be absent, a prewhitening method is given in Box and Jenkins (1970) for *identifying* an unknown dynamic system, but extension of this identification method to multiple inputs is rather complex.

Particularly for economic and business examples, however, the feedback structure and orders of the multiple system are often unknown. The present multiple time series procedure has the great advantage that it allows *identification* of the feedback and dynamic structure. Furthermore,

1. A one-sided causal relationship, if it exists, will emerge in the identification process, and the stochastic structures of the input as well as the transfer function relationship between input and output will be modeled simultaneously.
2. Stochastic multiple input and multiple output situations are readily handled.
3. A useful method is provided for seeking leading indicators in economic and business applications. In this context it should be noted that a unidirectional dynamic relationship may not exist between two time series even when one variable is known to be the input for the other. One reason for this phenomenon is the effect of temporal aggregation. As shown in Tiao and Wei (1976), pseudo-feedback relationships could occur because of this temporal aggregation effect, and it would be a mistake to impose a transfer function model in such a situation.
4. However, when a simple transfer function structure of the form (2.2) *is* appropriate, the present multiple time series approach could rarely reproduce it directly—see, for example, (5.4b) and (5.9)—and some analysis of the fitted form might be necessary to reveal a more parsimonious and more easily understood structure.

6. COMPARISON WITH SOME OTHER APPROACHES AND CONCLUDING REMARKS

We have discussed various tools used in an iterative approach to modeling multiple time series and illustrated

Table 13. Impulse Response Weights for the Gas Furnace Data

	j												
	0	1	2	3	4	5	6	7	8	9	10	11	12
v_j	·	·	·	−.53	−.67	−.89	−.51	−.29	−.17	−.09	−.05	−.03	−.02
v_j^*	·	·	·	−.53	−.70	−.77	−.48	−.26	−.09	−.01	.01	.00	−.01

Table 14. Successive AR Fitting Results for the Gas Furnace Data

Order of AR	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	$\boldsymbol{\Sigma}$
1	$\begin{bmatrix} + & + \\ - & + \end{bmatrix}$						$\begin{bmatrix} .102 & .090 \\ & .346 \end{bmatrix}$
2	$\begin{bmatrix} + & - \\ + & + \end{bmatrix}$	$\begin{bmatrix} - & + \\ - & - \end{bmatrix}$					$\begin{bmatrix} .037 & -.004 \\ & .069 \end{bmatrix}$
3	$\begin{bmatrix} + & \\ & + \end{bmatrix}$	$\begin{bmatrix} - & \\ - & - \end{bmatrix}$	$\begin{bmatrix} + & \\ & + \end{bmatrix}$				$\begin{bmatrix} .036 & -.002 \\ & .063 \end{bmatrix}$
4	$\begin{bmatrix} + & \\ & + \end{bmatrix}$	$\begin{bmatrix} - & \\ & - \end{bmatrix}$	$\begin{bmatrix} & \\ - & \end{bmatrix}$	$\begin{bmatrix} & \\ + & + \end{bmatrix}$			$\begin{bmatrix} .036 & -.003 \\ & .059 \end{bmatrix}$
5	$\begin{bmatrix} + & \\ & + \end{bmatrix}$	$\begin{bmatrix} - & \\ & - \end{bmatrix}$	$\begin{bmatrix} & \\ - & \end{bmatrix}$	$\begin{bmatrix} & \\ & \end{bmatrix}$	$\begin{bmatrix} & \\ & \end{bmatrix}$		$\begin{bmatrix} .035 & -.003 \\ & .058 \end{bmatrix}$
6	$\begin{bmatrix} + & \\ & + \end{bmatrix}$	$\begin{bmatrix} - & \\ & - \end{bmatrix}$	$\begin{bmatrix} & \\ - & \end{bmatrix}$	$\begin{bmatrix} & \\ & \end{bmatrix}$	$\begin{bmatrix} + & \\ & \end{bmatrix}$	$\begin{bmatrix} - & \\ + & \end{bmatrix}$	$\begin{bmatrix} .035 & -.002 \\ & .057 \end{bmatrix}$

how they work in practice. Much further work is needed, especially in the identification of mixed autoregressive moving average models and in developing faster estimation algorithms and better tools for diagnostic checking. In spite of the imperfections of the present tools and the preliminary nature of the approach, we have felt it appropriate to present them here in order to (a) illustrate the potential usefulness of vector autoregressive moving average models in characterizing dynamic structures in the data and (b) stimulate further development of modeling procedures. Several alternative approaches to modeling multiple time series have been proposed in the literature. It may be of interest to discuss briefly those proposed by Granger and Newbold (1977), Wallis (1977), and Chan and Wallis (1978).

In the Granger and Newbold approach, one begins by fitting univariate ARMA models to each series,

$$\varphi_{p_j}(B)Z_{jt} = \theta_{q_j}(B)C_{jt}, \quad j = 1, \ldots, k \tag{6.1}$$

and then attempts to identify the dynamic structure of the k white noise residual series $\{C_{jt}\}$ by examination of their cross-correlations. A model of the form (2.2) with $k(h)$ being the set $(1, \ldots, k)$ excluding h is then fitted to the k residual series. This model and the prewhitening transformations (6.1) then determine the model for the original vector series. As the authors themselves pointed out, the procedure is complex and difficult to apply for $k > 2$. One major difficulty arises from the fact that the parameters in the model for the residuals are subject to various complicated nonlinear constraints. Also, it can be readily shown that even if the vector series $\{\mathbf{Z}_t\}$ follows a low order ARMA model (3.1), the corresponding model for the residual vector $\{\mathbf{C}_t\}$ where $\mathbf{C}'_t = (C_{1t}, \ldots, C_{kt})$ can be complex and difficult to identify in practice.

The Wallis and Chan approach uses the form (6.1) for each individual series and the fact that the model (3.1) can be written as

$$|\boldsymbol{\varphi}(B)|\,\mathbf{Z}_t = \mathbf{H}(B)\mathbf{a}_t, \tag{6.2}$$

where $\mathbf{H}(B) = \mathbf{A}(B)\boldsymbol{\theta}_q(B)$, $\mathbf{A}(B)$ is the adjoint matrix and $|\boldsymbol{\varphi}_p(B)|$ the determinant of $\boldsymbol{\varphi}_p(B)$. As in the G and N approach, an individual model is first constructed for each series. From the degrees of the moving average polynomials $\theta_{q_j}(B)$ of these individual models, the degree of $\mathbf{H}(B)$ is determined. Next, models of the form

$$\mathbf{D}_l(B)\mathbf{Z}_t = \mathbf{H}(B)\mathbf{a}_t, \tag{6.3}$$

where $\mathbf{D}_l(B)$ is a diagonal matrix polynomial in B of degree l, are fitted successively for $l = r, r - 1, \ldots,$ where r is some specified maximum order, to determine an appropriate value for l. A likelihood ratio test is then performed to check whether the diagonal elements of $\mathbf{D}_l(B)$ are identical, that is, of the form (6.2). Finally, from the fitted $\mathbf{H}(B)$ and $|\varphi(B)|$ or $\mathbf{D}_l(B)$, one guesses at the values of p and q in (3.1) and then proceeds to estimate the parameters in $\boldsymbol{\varphi}_p(B)$ and $\boldsymbol{\theta}_q(B)$. The efficacy of this approach is open to question on several grounds.

1. The degree of the polynomial $\mathbf{H}(B)$ in (6.2) can be higher than the maximum degree of $\theta_{q_j}(B)$ for the individual series. For example, suppose $k = 2$,

$$\mathbf{H}(B) = \begin{pmatrix} 1 & -h_1B \\ -h_2B & 1 \end{pmatrix}$$

and the two elements of $\mathbf{a}_t$ are independent. Then $q_1 = q_2 = 0$, but it would be a mistake to infer that $\mathbf{H}(B)$ is of degree zero.

2. For vector AR or ARMA models, the representation (6.2) is certainly nonparsimonious. Apart from the covariance matrix $\boldsymbol{\Sigma}$, for k series the maximum number of parameters in the original form (3.1) is $k^2(p + q)$, while the maximum number of parameters in the form (6.2) is $kp + [(k - 1)p + q]k^2$, representing an increase

of $pk(k - 1)^2$ parameters. The increase could be even greater if the diagonal form (6.3) is employed. Thus, assuming the degree of $\mathbf{H}(B)$ is correctly specified, even for k as low as 3 or 4, a very large number of additional parameters will have to be estimated merely to identify correctly a low order vector AR model, say $p = 1$ or 2.

3. Since the correspondence between the degrees of the determinantal polynomial $|\boldsymbol{\varphi}(B)|$ and $\mathbf{H}(B)$ and the values of (p, q) is not necessarily one to one, it is not clear how one determines p and q in (3.1) from the form (6.2).

4. The approach is made even more computationally burdensome because the authors propose to employ the exact likelihood method for moving average parameters throughout the processes of model building. Our experience, however, suggests that because this method converges relatively slowly it is better to use it only in the final stage of the estimation process.

The chief distinction between our approach and the two alternatives just discussed is that we believe it better to tackle the dynamic relationships of the k series in their entirety, employing tools such as the estimates of cross-correlation matrices and partial autoregression matrices to shed light directly on the structure. Simplifications of one kind or another will then often follow. At least for the tentative specification of the vector autoregressive or the vector moving average model, our procedures seem far simpler to use in practice and do not require the multitude of steps these alternative approaches need to arrive at even a simple model.

To illustrate these points, we briefly consider the mink-muskrat example which Chan and Wallis used to illustrate their methods. They treat two series $Y_{1t}{}^*$ and $Y_{2t}{}^*$ obtained after "detrending" the muskrat and mink series by first and second degree polynomials respectively. Proceeding through the various steps outlined above, they eventually arrive at an AR(1) model. However, it will be seen that this same model is suggested immediately by the simple procedures we propose. Table 15(a) shows the partial autoregression results for $l = 1, 2$ and Table 15(b) the residual cross correlations after the AR(1) fit. A very similar analysis of this set of data is given in Ansley and Newbold (1979). For various reasons, we do not wish to sanctify this AR(1) model. These include the question of whether any linear structural model is adequate for these series (see Tong and Lim 1980). Also, the validity of the detrending procedures and the suspicious behavior of a high autocorrelation at lag 10 occurring in the residuals seem suspect. Our only point is to show that the circuitous route adopted by Chan and Wallis to arrive at this model is unnecessary.

Table 15. Identification of Muskrat-Mink Data

(a) Partial Autoregression and Related Statistics

Lag	*Partials*		$M(l) \overset{\sim}{\to} \chi_4^2$	*Diagonal elements of* $\hat{\boldsymbol{\Sigma}}$
1	+	−	111.7	.062
	+	+		.059
2	−	·	4.8	.0571
	·	·		.0572

(b) Cross-correlations of Residuals After AR(1) Fit

	a_1	a_2
a_1	. . . .	. − . . . .
a_2	. . . .	. . . . + . .

Before concluding this paper, it is worth noting that in modeling as well as analysis of vector time series one often finds it useful to perform various eigenvalue and eigenvector analyses. Specifically, writing (3.1) in the form

$$\mathbf{z}_t = \hat{\mathbf{z}}_{t-1}(1) + \mathbf{a}_t, \tag{6.4}$$

where $\hat{\mathbf{z}}_{t-1}(1)$ is the one step ahead forecast of $\mathbf{z}_t$ made at time $t - 1$, and denoting, for stationary series,

$$\boldsymbol{\Gamma}_{\mathbf{z}}(0) = E(\mathbf{z}_t \mathbf{z}'_t)$$

and

$$\boldsymbol{\Gamma}_{\hat{\mathbf{z}}}(0) = E(\hat{\mathbf{z}}_{t-1}(1)\hat{\mathbf{z}}_{t-1}(1)'),$$

it will often be informative to compute eigenvalues and eigenvectors of estimates of the following matrices:

$$(a)\ \boldsymbol{\Gamma}_{\mathbf{z}}(0), \quad (b)\ \hat{\boldsymbol{\Sigma}},$$

$$(c)\ \boldsymbol{\Gamma}_{z}(0)^{-1}\boldsymbol{\Gamma}_{\hat{\mathbf{z}}}(0), \quad (d)\ \boldsymbol{\varphi}_l, \text{ and } \boldsymbol{\theta}_l.$$

Such analyses are described in Quenouille (1957), Box and Tiao (1977), and Tiao et al. (1979). Also, the eigenvalues and eigenvectors of the spectral density matrix of the model should also be considered (see Brillinger 1975). These techniques are useful in (a) detecting exact concurrent or lagged linear relations between series, and (b) facilitating understanding and interpretation of the fitted model. In our opinion, this is one of the most important and challenging topics for further research.

[*Received January 1981. Revised June 1981.*]

REFERENCES

ABRAHAM, B. (1980), "Intervention Analysis and Multiple Time Series," *Biometrika*, 67, 73–78.

ABRAHAM, B., and BOX, G.E.P. (1978), "Deterministic and Forecast-Adaptive Time Dependent Models," *Applied Statistics*, 27, 120–130.

AKAIKE, H. (1980), "On the Identification of State Space Models and Their Use in Control," in *Directions in Time Series*, eds. D.R. Brillinger and G.C. Tiao, Institute of Mathematical Statistics, 175–187.

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

——— (1980), "Maximum Likelihood Estimation for Vector Autoregressive Moving Average Models," in *Directions in Time Series*, eds. D.R. Brillinger and G.C. Tiao, Institute of Mathematical Statistics, 49–59.

ANSLEY, C. (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process," *Biometrika*, 66, 59–65.

ANSLEY, C.F., and NEWBOLD, P. (1979), "Multivariate Partial Autocorrelations," *Proceedings of Business and Economic Statistics Section*, American Statistical Association, 349–353.

BARTLETT, M.S. (1938), "Further Aspects of the Theory of Multiple Regression," *Proceedings of the Cambridge Philosophical Society*, 34, 33–40.

BEGUIN, J.M., GOURIEROUX, C., and MONFORT, A. (1980), "Identification of a Mixed Autoregressive-Moving Average Process: The Corner Method," in *Time Series*, ed. O.D. Anderson, Amsterdam: North-Holland, 423–436.
BOX, G.E.P, and HAUGH, L. (1977), "Identification of Dynamic Regression Models Connecting Two Time Series," *Journal of the American Statistical Association*, 72, 121–130.
BOX, G.E.P., and JENKINS, G.M. (1970), *Time Series Analysis—Forecasting and Control*, San Francisco: Holden-Day.
BOX, G.E.P., and MACGREGOR, J.F. (1974), "The Analysis of Closed Loop Dynamic Stochastic Systems," *Technometrics*, 16, 391–398.
——— (1976), "Parameter Estimation with Closed-Loop Operating data," *Technometrics*, 18, 371–380.
BOX, G.E.P., and NEWBOLD, P. (1970), "Some Comments on a Paper by Coen, Gomme and Kendall," *Journal of the Royal Statistical Society*, Ser. A, 134, 229–240.
BOX, G.E.P., and TIAO, G.C. (1975), "Intervention Analysis with Applications to Environmental and Economic Problems," *Journal of the American Statistical Association*, 70, 70–79.
——— (1977), "A Canonical Analysis of Multiple Time Series," *Biometrika*, 64, 355–365.
BRILLINGER, D.R. (1975), *Time Series Data Analysis and Theory*, New York: Holt, Rinehart, and Winston.
CHAN, W.Y.T., and WALLIS, K.F. (1978), "Multiple Time Series Modelling: Another Look at the Mink-Muskrat Interaction," *Applied Statistics*, 27, 168–175.
COEN, P.G., GOMME, E.D., and KENDALL, M.G. (1969), "Lagged Relationships in Economic Forecasting," *Journal of the Royal Statistical Society*, Ser. A, 132, 133–163.
DEISTLER, M., DUNSMUIR, W., and HANNAN, E.J. (1978), "Vector Linear Time Series Models: Corrections and Extensions," *Advances in Applied Probability*, 10, 360–372.
DENT, W. (1977), "Computation of the Exact Likelihood Function of an ARIMA Process," *Journal of Statistical Computation and Simulation*, 5, 193–206.
DUNSMUIR, W., and HANNAN, E.J. (1976), "Vector Linear Time Series Models," *Advances in Applied Probability*, 8, 339–364.
GRANGER, C.W.J., and NEWBOLD, P. (1977), *Forecasting Economic Time Series*, New York: Academic Press.
GRAY, H.L., KELLEY, G.D., and McINTIRE, D.D. (1978), "A New Approach to ARMA Modelling," *Communications in Statistics*, B7, 1–77.
HALLIN, M. (1978), "Mixed Autoregressive-Moving Average Multivariate Processes with Time-Dependent Coefficients," *Journal of Multivariate Analysis*, 8, 567–572.
HANNAN, E.J. (1970), *Multiple Time Series*, New York: John Wiley.
——— (1980), "The Estimation of the Order of an ARMA Process," *Annals of Statistics*, 8, 1071–1081.
HANNAN, E.J., DUNSMUIR, W.T.M., and DEISTLER, M. (1980), "Estimation of Vector ARMAX Models," *Journal of Multivariate Analysis*, 10, 275–295.
HILLMER, S.C., and TIAO, G.C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74, 652–660.
HOSKING, J.R.M. (1980), "The Multivariate Portmanteau Statistic," *Journal of the American Statistical Association*, 75, 602–607.
HSIAO, C. (1979), "Autoregressive Modeling of Canadian Money and Income Data," *Journal of the American Statistical Association*, 74, 553–560.
JENKINS, G.J. (1979), *Practical Experiences with Modelling and Forecasting Time Series*, Channel Islands: GJP Ltd.
LI, W.K., and McLEOD, A.I. (1980), "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models," TR-80-03, University of Western Ontario.
NEWBOLD, P. (1974), "The Exact Likelihood Function for a Mixed Autoregressive Moving Average Process," *Biometrika*, 61, 423–427.
NICHOLLS, D.F. (1976), "The Efficient Estimation of Vector Linear Time Series Models," *Biometrika*, 63, 381–390.
——— (1977), "A Comparison of Estimation Methods for Vector Linear Time Series Models," *Biometrika*, 64, 85–90.
NICHOLLS, D.F., and HALL, A.D. (1979), "The Exact Likelihood of Multivariate Autoregressive-Moving Average Models," *Biometrika*, 66, 259–264.
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.
PARZEN, E. (1977), "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes," in *Multivariate Analysis-IV*, ed. P. Krishnaiah, Amsterdam: North-Holland, 283–295.
PHADKE, M.S., and KEDEM, G. (1978), "Computation of the Exact Likelihood Function of Multivariate Moving Average Models," *Biometrika*, 65, 511–519.
QUENOUILLE, M.H. (1957), *The Analysis of Multiple Time Series*, London: Griffin.
QUINN, B.G. (1980), "Order Determination for a Multivariate Autoregression," *Journal of the Royal Statistical Society*, Ser. B, 42, 182–185.
SLUTSKY, E. (1937), "The Summation of Random Causes as the Source of Cyclic Processes," *Econometrika*, 5, 105–146.
TIAO, G.C., and ALI, M.M. (1971), "Analysis of Correlated Random Effects: Linear Model with Two Random Components," *Biometrika*, 58, 37–51.
TIAO, G.C., BOX, G.E.P., GRUPE, M.R., HUDAK, G.B., BELL, W.R., and CHANG, I. (1979), "The Wisconsin Multiple Time Series (WMTS-1) Program: A Preliminary Guide," Department of Statistics, University of Wisconsin, Madison.
TIAO, G.C., and WEI, W.S. (1976), "Effect of Temporal Aggregation on the Dynamic Relationship of Two Time Series Variables," *Biometrika*, 63, 513–523.
TONG, H., and LIM, K.S. (1980), "Threshold Autoregression, Limit Cycles and Cyclical Data," *Journal of the Royal Statistical Society*, Ser. B, 42, 245–292.
WALLIS, K.F. (1977), "Multiple Time Series Analysis and the Final Form of Econometric Models," *Econometrika*, 45, 1481–1497.
WHITTLE, P. (1963), "On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix," *Biometrika*, 50, 129–134.
WILSON, G.T. (1973), "The Estimation of Parameters in Multivariate Time Series Models," *Journal of the Royal Statistical Society*, Ser. B, 35, 76–85.
YULE, G.U. (1927), "On a Method of Investigating Periodicities in Disturbed Series, With Special Reference to Wolfer's Sunspot Numbers," *Philosophical Transactions of the Royal Society of London*, Ser. A, 226, 267–298.
ZELLNER, A., and PALM, F. (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics*, 2, 17–54.