A Three-Stage Iterative Procedure for Space-Time Modeling

Phillip E. Pfeifer and Stuart Jay Deutsch

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Reprinted from
TECHNOMETRICS
Volume 22, Number 1, February 1980

TECHNOMETRICS ©, VOL. 22, NO. 1, FEBRUARY 1980

A Three-Stage Iterative Procedure for Space-Time Modeling

Phillip E. Pfeifer and Stuart Jay Deutsch
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

A three-stage iterative procedure for building space-time models is presented. These models fall into the general class of STARIMA models and are characterized by autoregressive and moving average terms lagged in both time and space. This model class collapses into the ARIMA model class in the absence of spatial correlation. The theoretical properties of STARIMA models are presented and the model building procedure is described and illustrated by a substantive example.

KEY WORDS

Space-time modeling
STARIMA
STAR
STMA
Time series modeling
Three-stage model building procedure

1. INTRODUCTION

A flexible class of empirical models is the multiplicative autoregressive moving average model family. These models, together with the model building procedure commonly referred to as the Box-Jenkins method (see Box and Jenkins, 1970), have proven very useful in a wide spectrum of statistical analyses (Box and Tiao, 1975; Deutsch, 1978; Deutsch and Alt, 1977; Deutsch and Ogelsby, 1979; Deutsch and Wu, 1974; McMichael and Hunter, 1972). Since these models are univariate, however, they are applicable only to single series of data. Thus, although statistics are often available over a region in space, univariate models can only deal with the history at one particular point or region in space, or effectively work with data from several distinct regions which have been aggregated to form a larger but single region of space. An alternative to univariate time series modeling is multivariate time series modeling (Goldberger, 1964; Granger and Newbold, 1977; Hannon, 1970; Phadke and Wu, 1974). These models attempt to simultaneously describe and forecast a set of N observable time series. When these N series represent patterns of the N regions, the interrelationships between the different regions can be taken into account and thus a better systems description should result.

A further refinement of a general multivariate time series model can occur if the system to be modeled exhibits systematic dependence between the observations at each region and the observations at neighboring regions. This phenomenon is labeled "spatial correlation," and was referred to by Cliff and Ord (1973):

> If the presence of some quality in a county of a country makes its presence in neighboring counties more or less likely, we say that the phenomenon exhibits spatial autocorrelation.

Models that explicitly attempt to explain these dependencies across space are referred to as space-time models.

The purpose of this paper is to describe the extension of the three-stage iterative model building procedure developed by Box and Jenkins to accommodate the space-time, time series model class of STARMA models. In Section 2 of this paper we propose a comprehensive class of space-time models that are characterized by the autoregressive and moving average forms of univariate time series lagged in both time and space. These types of models have been referred to as space-time autoregressive moving average (STARMA) models (Cliff et al., 1975; Cliff and Ord, 1973; Marten and Oeppen, 1975). Section 3 develops the general procedures for the identification stage in which a tentative STARMA model is selected. The identification of a tentative model from

Received July 1978; revised February 1979

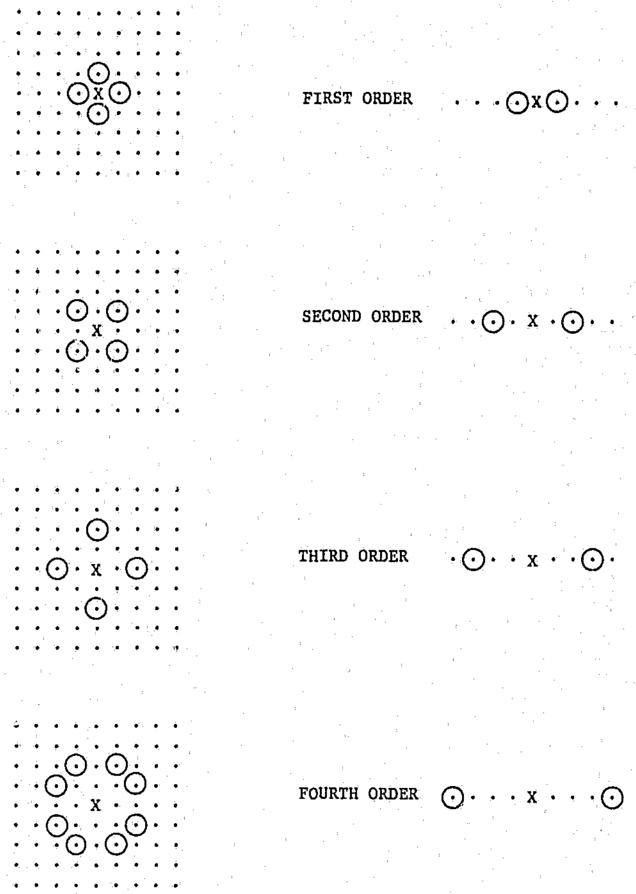

FIGURE 1. Spatial order in two- and one-dimensional systems.

data structure is the first stage of the three-stage iterative model building procedure. Section 4 deals with the second stage of the modeling procedure, estimation, or the fitting of the tentative model. The last stage, diagnostic checking, is the subject of Section 5. Here the adequacy of the fitted model is tested so as to give direction for updating the tentative model if it is inadequate. Section 6 presents a substantial application of the space-time model class and its associated modeling procedures to actual reported crime data for the city of Boston.

2. THE SPACE-TIME AUTOREGRESSIVE MOVING AVERAGE MODEL

The STARMA model class is characterized by linear dependence lagged in both space and time. Assume that observations $z_i(t)$ of the random variable $Z_i(t)$ are available at each of N fixed locations in space ($i = 1, 2, \cdots N$) over T time periods. The N locations in space will be referred to as sites and can represent a variety of situations. For instance, these sites could be the counties of a state, the districts of a city, or the cities of a nation. The autoregressive form of the space-time model would express the observation at time t and site i, $z_i(t)$ as a linear combination of past observations at zone i and neighboring zones. If the same relationship holds for every site in the system, the process is said to exhibit spatial stationarity and is thus amenable to these forms of space-time models.

To assist in the formulation of this space-time model, the following definition of the spatial lag operator is needed. Let $L^{(l)}$, the spatial lag operator of spatial order l, be such that

$$L^{(0)}z_i(t) = z_i(t)$$

$$L^{(l)}z_i(t) = \sum_{j=1}^{N} w_{ij}^{(l)} z_j(t)$$

where $w_{ij}^{(l)}$ are a set of weights with

$$\sum_{j=1}^{N} w_{ij}^{(l)} = 1$$

for all i and $w_{ij}^{(l)}$ nonzero only if sites i and j are l^{th} order neighbors. The matrix representation of the set of weights $w_{ij}^{(l)}$ is $\mathbf{W}^{(l)}$, an $N \times N$ square matrix with each row summing to one. If $\mathbf{z}(t)$ is an $(N \times 1)$ column vector of the observations $z_i(t)$, $i = 1, 2, \cdots N$, then

$$L^{(0)}\mathbf{z}(t) = \mathbf{W}^{(0)}\mathbf{z}(t) = \mathbf{I}_N\mathbf{z}(t)$$

and

$$L^{(l)}\mathbf{z}(t) = \mathbf{W}^{(l)}\mathbf{z}(t) \text{ for } l > 0.$$

The specification of the form of weights $w_{ij}^{(l)}$ for various positive l's is a matter left up to the model builder who may choose weights to reflect the configuration of, for example, the county system. The $w_{ij}^{(l)}$ may be chosen to reflect physical properties of the observed system such as the length of the common boundary between contiguous counties i and j, the distance between the centers of counties, natural barriers such as rivers or mountains and even the ease of accessibility of county i to county j. This last factor might include such things as the number of roads between i and j, the amount of public transportation available connecting the two, and even the flow rates upon these avenues.

These weights, however, must reflect a hierarchical ordering of spatial neighbors. First order neighbors are those "closest" to the site of interest. Second order neighbors should be "farther" away than first order neighbors, but "closer" than third order neighbors. For regularly spaced systems, a workable definition of spatial order is available (see Besag, 1974). Figure 1 shows the first four spatial order neighbors of a particular site for both a two-dimensional grid system and a one-dimension line of sites. This definition of spatial order represents an ordering in terms of euclidean distance of all sites surrounding the location of interest.

With this definition of spatial order in hand, we are now ready to present the STARMA model. Analogous to univariate time series, $z_i(t)$ will be expressed as a linear combination of past observations and errors. Here, however, instead of allowing dependence of $z_i(t)$ only with past observations and errors at site i, dependence is allowed with neighboring sites of various spatial order. In particular

$$z_i(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \phi_{kl} L^{(l)} z_i(t-k) - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} L^{(l)} \epsilon_i(t-k) + \epsilon_i(t) \quad (1)$$

where p is the autoregressive order, q is the moving average order, λ_k is the spatial order of the k^{th} autoregressive term, m_k is the spatial order of the k^{th} moving average term, ϕ_{kl} and θ_{kl} are parameters, and the $\epsilon_i(t)$ are random normal errors with

$$E[\epsilon_i(t)] = 0$$

$$E[\epsilon_i(t)\epsilon_j(t+s)] = \begin{cases} \sigma^2 & i=j, s=0 \\ 0 & \text{otherwise} \end{cases}$$

This model is referred to as a STARMA $(p_{\lambda_1,\lambda_2,\cdots\lambda_p}, q_{m_1,m_2,\cdots m_q})$ model.

The same model in vector form is

$$\mathbf{z}(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^{(l)} \mathbf{z}(t-k) - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^{(l)} \boldsymbol{\varepsilon}(t-k) + \boldsymbol{\varepsilon}(t) \quad (2)$$

with $\boldsymbol{\varepsilon}(t)$ normal with mean zero and

$$E[\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(t+s)'] = \begin{cases} \sigma^2 \mathbf{I}_N & s=0 \\ 0 & \text{otherwise} \end{cases}$$

Two special subclasses of the STARMA model are of note. When $q = 0$, only autoregressive terms remain, and hence the model class carries the label space-time autoregressive or STAR model. The model

$$\mathbf{z}(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^{(l)} \mathbf{z}(t-k) + \boldsymbol{\varepsilon}(t) \quad (3)$$

is referred to as a STAR $(p_{\lambda_1,\lambda_2,\cdots\lambda_p})$ model.

Models that contain no autoregressive terms ($p = 0$) are referred to as STMA models. The model form

$$\mathbf{z}(t) = \boldsymbol{\varepsilon}(t) - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^{(l)} \boldsymbol{\varepsilon}(t-k) \quad (4)$$

is a STMA $(q_{m_1,m_2,\cdots m_q})$ model.

In order for the STARMA model to represent a *stationary* process, one in which the covariance structure of $\mathbf{z}(t)$ does not change with time, certain conditions must be met. These conditions are called the stationarity conditions and require that every x_u that solves

$$\det\left[x_u^p \mathbf{I} - \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^{(l)} x_u^{p-k}\right] = 0$$

lie inside the unit circle ($|x_u| < 1$). Effectively this requirement serves to determine a region of possible ϕ_{kl} values that will result in a stationary process.

If the same conditions are applied to the moving average terms in the STARMA model, namely if every x_u that solves

$$\det\left[x_u^q \mathbf{I} - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^{(l)} x_u^{q-k}\right] = 0$$

lies inside the unit circle ($|x_u|<1$), then the model is said to be *invertible*. The invertibility property implies that $\mathbf{z}(t)$ can be expressed as a weighted linear combination of past observations with weights that converge to zero. It is clear that all STAR models are invertible and all STMA models are stationary.

3. IDENTIFICATION OF STARMA MODELS

The most pressing questions encountered when attempting to utilize these forms of space-time models are: which of the model forms (STAR, STMA, STARMA) is most appropriate for the data at hand, and what are the temporal and spatial orders ($p,q,\boldsymbol{\lambda},\mathbf{m}$) of the model form? These questions are answered in the identification stage of the three-stage model building procedure.

Identification is the process by which one subclass of the general model class is chosen that exhibits theoretical properties most closely matching those estimated from the data. The techniques of identification involve summarizing and categorizing the data to yield information that best matches the observed process with a subclass of models. In univariate time series modeling, the primary tools in identification are the autocorrelation and partial autocorrelation functions. Choosing between the three general subclasses of models (AR, MA, ARMA) is a matter of determining whether the partial autocorrelation function cuts off, the autocorrelation function cuts off, or they both tail off.

To identify space-time models, it is usually a good idea to combine the N^2 possible cross-covariances between all possible pairs of sites in a logical manner consistant with the forms associated with the proposed model class. The result is labeled the space-time autocovariance function, a function expressing the covariance between points lagged *both* in space and time.

Using the definition of the spatial lag operator presented previously, an average covariance between the weighted l^{th} order neighbors of any site and the weighted k^{th} order neighbors of the same site at s

time lags in the future would be

$$\gamma_{lk}(s) = E\left\{\sum_{i=1}^{N} \frac{L^{(l)} z_i(t)\, L^{(k)} z_i(t+s)}{N}\right\}. \quad (5)$$

Here $\gamma_{lk}(s)$ is referred to as the space-time covariance between l^{th} and k^{th} order neighbors at time lag s. This formulation assumes that $E\{z_i(t)\} = 0$, and hence, will often require a centering of the system about the overall sample mean $\bar{z}$.

In vector notation the space-time covariance function can be expressed as

$$\gamma_{lk}(s) = E\left\{\frac{[\mathbf{W}^{(l)}\,\mathbf{z}(t)]'\,[\mathbf{W}^{(k)}\,\mathbf{z}(t+s)]}{N}\right\} \quad (6)$$

which can be seen to be equivalent to

$$\gamma_{lk}(s) = tr\left\{\frac{\mathbf{W}^{(k)\prime}\,\mathbf{W}^{(l)}\,\mathbf{\Gamma}(s)}{N}\right\} \quad (7)$$

where $\mathbf{\Gamma}(s) = E\,[\mathbf{z}(t)\,\mathbf{z}(t+s)']$ and tr[$\mathbf{A}$] is the trace of $\mathbf{A}$ defined on square matrices as the sum of the diagonal elements.

The usual estimator of $\mathbf{\Gamma}(s)$

$$\hat{\mathbf{\Gamma}}(s) = \sum_{t=1}^{T-s} \frac{\mathbf{z}(t)\,\mathbf{z}(t+s)'}{T-s} \quad (8)$$

may be substituted into (7) to obtain sample estimates, $\hat{\gamma}_{lk}(s)$, of the space-time autocovariance function. Alternately one could estimate directly via (5) as

$$\hat{\gamma}_{lk}(s) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T-s} L^{(l)}z_i(t)\,L^{(k)}z_i(t+s)}{N(T-s)}. \quad (9)$$

In particular note that

$$\gamma_{00}(s) = \frac{1}{N}\,\text{tr}\,[\mathbf{\Gamma}(s)]$$

which equals the average of the s^{th} lag autocovariance for all N sites. Also

$$\gamma_{10}(s) = \frac{1}{N}\,\text{tr}\,[\mathbf{W}^{(1)}\mathbf{\Gamma}(s)]$$

is the average over all sites of the covariance between each site and its weighted first order neighbors s time lags previous.

The space-time covariance has the important property that

$$\gamma_{lk}(s) = \gamma_{kl}(-s). \quad (10)$$

This follows from (7) and the fact that tr [$\mathbf{AB}$] = tr [$\mathbf{BA}$] and $\mathbf{\Gamma}(s) = \mathbf{\Gamma}(-s)'$. In particular from (7)

$$\gamma_{lk}(s) = \frac{1}{N}\,\text{tr}\,[\mathbf{W}^{(k)\prime}\mathbf{W}^{(l)}\mathbf{\Gamma}(s)]$$

which also equals

$$\frac{1}{N}\,\text{tr}\,[\mathbf{\Gamma}(s)'\mathbf{W}^{(l)\prime}\mathbf{W}^{(k)}]$$

since tr [$\mathbf{A}$] = tr [$\mathbf{A}'$]. Rearranging we have

$$\gamma_{lk}(s) = \frac{1}{N}\,\text{tr}\,[\mathbf{W}^{(l)\prime}\mathbf{W}^{(k)}\mathbf{\Gamma}(s)']$$

which equals

$$\frac{1}{N}\,\text{tr}\,[\mathbf{W}^{(l)\prime}\mathbf{W}^{(k)}\mathbf{\Gamma}(-s)] \text{ or } \gamma_{kl}(-s).$$

The definition of space-time autocorrelation is not as straightforward a matter as it is in the univariate domain since there are several possible scalings that might be used (see Martin and Oeppen, 1975). One is interested in a definition that leads to sample autocorrelations that have constant variance at all spatial lags. A suitable definition for the space-time autocorrelation between l^{th} and k^{th} order neighbors s times lags apart is

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{[\gamma_{ll}(0)\,\gamma_{kk}(0)]^{1/2}}. \quad (11)$$

This definition is preferred because the variance of its sample estimate has been shown to be relatively constant for all spatial lags (Pfeifer, 1979).

The sample estimates of the space-time autocorrelation coefficients follow quite naturally to be

$$\hat{\rho}_{lk}(s) = \frac{\hat{\gamma}_{lk}(s)}{[\hat{\gamma}_{ll}(0)\hat{\gamma}_{kk}(0)]^{1/2}}$$

$$= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T-s} L^{(l)}z_i(t)L^{(k)}z_i(t+s)}{\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(L^{(l)}z_i(t))^2\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(L^{(k)}z_i(t))^2\right]^{1/2}} \quad (12)$$

Once the sample autocorrelation function has been calculated via (12) the selection of a candidate model from the STARMA model family is still not an easy task. Although each of the particular models within the STARMA family has a unique space-time autocorrelation function, it is often hard to distinguish between some of the model forms. In particular, even if it is known that the process is autoregressive, but of unknown order, it is not easy to see from $\hat{\rho}_{lk}(s)$ what the order is. This particular problem can, however, be overcome by using the space-time partial correlation function.

As was the case with the space-time autocorrelation function, alternate definitions exist for the space-time partial autocorrelation function (see Martin and Oeppen, 1975). In light of the proposed model class, however, the appropriate definition follows quite directly from the form of the STAR model. Pre-multiplying both sides of the general STAR($k_{\lambda,\lambda,\ldots}$) model

$$\mathbf{z}(t) = \sum_{j=1}^{k}\sum_{l=0}^{\lambda} \phi_{jl}\mathbf{W}^{(l)}\mathbf{z}(t-j) + \boldsymbol{\varepsilon}(t)$$

$$
\begin{array}{c}
s=1 \\ \\ \\ \\ s=2 \\ \\ \\ \\ \\ s=k
\end{array}
\left[\begin{array}{c}
\gamma_{00}(1) \\ \gamma_{10}(1) \\ \vdots \\ \gamma_{\lambda 0}(1) \\ \hline
\gamma_{00}(2) \\ \gamma_{10}(2) \\ \vdots \\ \gamma_{\lambda 0}(2) \\ \hline
\vdots \\ \hline
\gamma_{00}(k) \\ \gamma_{10}(k) \\ \vdots \\ \gamma_{\ell 0}(k)
\end{array}\right]
=
\left[\begin{array}{c|c|c|c}
\begin{matrix} \gamma_{00}(0) & \gamma_{01}(0) & \cdots & \gamma_{0\lambda}(0) \\ \gamma_{10}(0) & \gamma_{11}(0) & \cdots & \gamma_{1\lambda}(0) \\ & \vdots & & \\ \gamma_{\lambda 0}(0) & \gamma_{\lambda 1}(0) & \cdots & \gamma_{\lambda\lambda}(0) \end{matrix} &
\begin{matrix} \gamma_{00}(-1) & \gamma_{01}(-1) & \cdots & \gamma_{0\lambda}(-1) \\ \gamma_{10}(-1) & \gamma_{11}(-1) & \cdots & \gamma_{1\lambda}(-1) \\ & \vdots & & \\ \gamma_{\lambda 0}(-1) & \gamma_{\lambda 1}(-1) & \cdots & \gamma_{\lambda\lambda}(-1) \end{matrix} & \cdots & (1-k) \\ \hline
(1) & (0) & & (2-k) \\ \hline
\vdots & \vdots & & \vdots \\ \hline
(k-1) & (k-2) & \cdots & (0)
\end{array}\right]
\left[\begin{array}{c}
\phi_{10} \\ \phi_{11} \\ \vdots \\ \phi_{1\lambda} \\ \hline
\phi_{20} \\ \phi_{21} \\ \vdots \\ \phi_{2\lambda} \\ \hline
\vdots \\ \hline
\phi_{k0} \\ \phi_{k1} \\ \vdots \\ \phi_{k\ell}
\end{array}\right]
\qquad (14)
$$

$$\underbrace{\qquad}_{j=1} \qquad \underbrace{\qquad}_{j=2} \qquad \underbrace{\quad}_{j=k}$$

DISPLAY 1. The space-time analog of the Yule-Walker equations.

by $[\mathbf{W}^{(h)}\mathbf{z}(t-s)]'$ gives

$$\mathbf{z}(t-s)'\mathbf{W}^{(h)'}\mathbf{z}(t) = \sum_{j=1}^{k} \sum_{l=0}^{\lambda} \phi_{jl}\mathbf{z}(t-s)'\mathbf{W}^{(h)'}\mathbf{W}^{(l)}\mathbf{z}(t-j) + \mathbf{z}(t-s)'\mathbf{W}^{(h)'}\boldsymbol{\varepsilon}(t).$$

Taking expected values and dividing both sides by N yields

$$\gamma_{h0}(s) = \sum_{j=1}^{k} \sum_{l=0}^{\lambda} \phi_{jl}\gamma_{hl}(s-j) \qquad (13)$$

since $E[\mathbf{z}(t-s)'\boldsymbol{\varepsilon}(t)] = 0$ for $s > 0$.

This system of equations for $s = 1, 2, \cdots k$ and $h = 0, 1, \cdots \lambda$ may be written as shown in Display 1 (equation (14)). This system is the space-time analog of the Yule-Walker equations for univariate time series.

The last coefficient, ϕ'_{kl}, obtained from solving the system of equations as $l = 0, 1 \cdots \lambda$ for $k = 1, 2, \cdots$ is called the space-time partial correlation function of spatial order λ. For example, the space-time correlation function of spatial order 2 has the series of terms $\phi'_{10}, \phi'_{11}, \phi'_{12}, \phi'_{20}, \phi'_{21}, \phi'_{22}, \phi'_{30}, \cdots$ that represent the last coefficient obtained by successively fitting the systems of equations for $\lambda = 2$. For spatial $\lambda = 3$ the function has elements $\phi'_{10}, \phi'_{11}, \phi'_{12}, \phi'_{13}, \phi'_{20}, \phi'_{21}, \phi'_{22}, \phi'_{23}, \phi'_{30}, \cdots$ etc.

The choice of λ, the spatial order of the space-time partial correlation function, is left to the investigator. It is important that λ be at least as large as the maximum spatial order of any hypothesized model. Too large a λ, however, results in an unduly large amount of computational effort. In making the choice on the value of λ, the size of the system should be taken into account. Larger systems might warrant a fairly high λ, say 3 or 4, but for moderate systems $\lambda = 2$ will likely suffice.

The space-time partial correlation function could be estimated by successively fitting STAR $(k_{\lambda,\lambda,\cdots,l})$ models for $l = 0, 1, \cdots \lambda$ for each k, $k = 1, 2, \cdots$ and picking out the estimates $\hat{\phi}_{kl}$ of the last coefficient from each of these models. Fitting STAR models, however, involves enough computational effort to warrant an alternative, approximate estimation procedure. If the values of the parameters are not too close to their stationarity boundaries, approximate Yule-Walker type estimates can be employed. They are calculated from equation set (13) by replacing the theoretical $\gamma_{hl}(s)$ with their estimates $\hat{\gamma}_{hl}(s)$ and solving this system successively for $l = 0, 1, \cdots \lambda$ for $k = 1, 2, \cdots$.

Direct solution of (14) with estimates of γ replacing the theoretical values still involves a bit of computational effort. Unfortunately a strictly recursive method similar to that due to Durbin (1960) for univariate time series partial calculation is not possible. Some improvement over the successive solution of system (14), however, can be made.

In a manner completely analogous to that of univariate time series, STARMA processes are each characterized by a distinct space-time partial and autocorrelation function. Whereas univariate autoregressive models exhibit autocorrelation functions that decay exponentially with time and partial correlation functions that cut off after p lags, the STAR process exhibits a space-time correlation function that tails off with both space and time and partial autocorrelations that cut off after p lags in time and λ_p lags in space. Similarly, univariate moving average models have just the opposite, autocorrelations that cut off after q lags and partials that decay over time. The STMA $(q_{m_1 \cdots m_q})$ model similarly is characterized by an autocorrelation function that cuts off after q temporal lags and m_q spatial lags and partials that

TABLE 1—*Space-time autocorrelation functions for the STAR model.*

Space-Time Autocorrelation, $\rho_{\ell 0}(s)$

Spatial lag (ℓ) / Time lag (s)	5x5 system 0	1	2	3	7x7 system 0	1	2	3
1	0.607	0.467	0.216	0.155	0.598	0.447	0.181	0.145
2	0.424	0.413	0.228	0.159	0.410	0.394	0.189	0.149
3	0.319	0.353	0.222	0.155	0.303	0.333	0.184	0.149
4	0.250	0.298	0.206	0.144	0.234	0.281	0.170	0.135
5	0.202	0.254	0.186	0.133	0.187	0.237	0.153	0.124
6	0.166	0.217	0.166	0.121	0.152	0.200	0.136	0.112
7	0.139	0.186	0.147	0.110	0.125	0.171	0.120	0.101
8	0.117	0.160	0.130	0.099	0.105	0.145	0.105	0.091
9	0.100	0.138	0.114	0.090	0.089	0.125	0.097	0.081
10	0.086	0.119	0.100	0.080	0.075	0.108	0.080	0.072

Space-Time Partial Autocorrelations, $\phi'_{s\ell}$

Spatial lag (ℓ) / Times lag (s)	5x5 system 0	1	2	3	7x7 system 0	1	2	3
1	0.608	0.400	0	0	0.598	0.400	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0

tail off spatially and temporally. Mixed models exhibit partials and autocorrelations that both tail off. In the univariate case, they tail off only in time, whereas space-time mixed ARMA processes have space-time autocorrelation functions that decay with both time *and* space.

To illustrate the relationship between the theoretical space-time partial and autocorrelation functions and the three subclasses of the STARMA model family, the theoretical space-time autocorrelation and partial autocorrelation functions for three representative models of a STAR, STMA and mixed STARMA are calculated. First order models, both temporally and spatially, are chosen: the STAR (1_1) with $\phi_{10} = 0.5$ and $\phi_{11} = 0.4$,

$$\mathbf{z}(t) = 0.5\,\mathbf{z}(t-1) + 0.4\,\mathbf{W}^{(1)}\mathbf{z}(t-1) + \boldsymbol{\varepsilon}(t);$$

the STMA (1_1) with $\theta_{10} = -0.5$ and $\theta_{11} = -0.4$,

$$\mathbf{z}(t) = \boldsymbol{\varepsilon}(t) + 0.5\boldsymbol{\varepsilon}(t-1) + 0.4\mathbf{W}^{(1)}\boldsymbol{\varepsilon}(t-1);$$

and the STARMA $(1_1, 1_1)$ with $\phi_{10} = 0.5$, $\phi_{11} = 0.4$, $\theta_{10} = -0.5$, $\theta_{11} = -0.4$,

$$\mathbf{z}(t) = 0.5\mathbf{z}(t-1) + 0.4\mathbf{W}^{(1)}\mathbf{z}(t-1) + \boldsymbol{\varepsilon}(t) + 0.5\boldsymbol{\varepsilon}(t-1) + 0.4\mathbf{W}^{(1)}\boldsymbol{\varepsilon}(t-1).$$

Tables 1, 2, and 3 present the calculated space-time autocorrelations and space-time partials for these STAR, STMA and STARMA models respectively. Only $\rho_{l0}(s)$ is presented [not $\rho_{lk}(s)$ for positive l and k] since identification can usually proceed strictly on the basis of $\rho_{l0}(s)$ for $l = 0, 1, \cdots \lambda$ and $s = 1, 2, \cdots S$. For compactness we chose $l = 3$ and $S = 10$.

Two sizes of systems are presented. The three tables provide the correlative information for both a 5×5 and 7×7 square regular system of sites. The weighting scheme chosen in both instances incorporates the spatial order convention depicted in Figure 1, with the added property that each l^{th} order neighbor of any site is weighted equally. Equal scaled weights are chosen not only because they are useful in their own right, especially toward modeling regularly spaced homogeneous spatial systems, but also because they can serve as a pattern for more general weighting schemes. Thus, the four corner points, because they possess just 2 first order neighbors, will have $w_{ij}^{(1)} = 1/2$ for the two j values. The remaining nodes on the boundary each have 3 first order neighbors and thus, $w_{ij}^{(1)} = 1/3$ for the 3 neighboring sites. All other interior sites possess a full compliment of first order neighbors and have 4 nonzero $w_{ij}^{(1)}$ with value 1/4.

Tables 1, 2, and 3 serve to demonstrate the basic concepts of space-time identification. In each instance, the space-time autocorrelation and partial autocorrelation functions give an indication of both the type and order of the generating model. Table 1 shows space-time partials that cut off after one spatial and one temporal lag, characterizing the STAR (1_1) model. Table 2, on the other hand, shows a space-time autocorrelation function cutting off after initial first order terms both spatially and temporally, thus representing a STMA (1_1) model. Lastly, Table 3 exhibits autocorrelations and partials that tail off, a characteristic of a mixed model.

In practice, of course, the identification of a candidate model will never be as easy as it appears here since the model builder will not be dealing with the exact space-time autocorrelation functions of the underlying process, but rather a sample calculated from the observed data history. Thus the identification

TABLE 2—*Space-time autocorrelation functions for the STMA model.*

Space-Time Autocorrelations, $\rho_{\ell 0}(s)$

Spatial lag (ℓ) / Time lag (s)	5x5 system 0	1	2	3	7x7 system 0	1	2	3
1	0.384	0.173	0	0	0.385	0.167	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0

Space-Time Partial Autocorrelation, $\phi'_{s\ell}$

Spatial lag (ℓ) / Time lag	5x5 system 0	1	2	3	7x7 system 0	1	2	3
1	0.384	0.190	-0.041	-0.024	0.385	0.189	-0.043	-0.024
2	-0.190	-0.179	0.007	0.007	-0.190	-0.179	0.009	0.006
3	0.108	0.140	0.017	0.006	0.108	0.141	0.017	0.007
4	0.068	-0.107	-0.027	-0.011	0.067	-0.108	-0.028	-0.012
5	0.045	0.082	0.028	0.012	0.045	0.083	0.029	0.013
6	-0.032	-0.064	-0.126	-0.011	0.031	-0.064	-0.027	-0.012
7	0.024	0.050	0.022	0.009	0.023	0.050	0.023	0.010
8	-0.018	-0.039	-0.019	-0.008	-0.017	-0.039	-0.019	-0.008
9	0.014	0.031	0.015	0.006	0.013	0.030	0.016	0.007
10	0.011	-0.024	-0.012	-0.005	0.010	-0.024	-0.012	-0.005

process is complicated by the sample fluctuations exhibited by the estimates of $\rho_{l0}(s)$ and ϕ'_{sl}. An approximate measure of the variance of the sample space-time autocorrelations of a pure white noise process, (see equation (19) below), however, will help to determine the significance of observed correlations, and a more detailed knowledge of the relationships between the theoretical patterns of the space-time autocorrelation functions and the values of the parameters in the STARMA models (as presented in Pfeifer, 1979) will also aid the model builder in the identification stage. In any case, the identification stage represents a tentative evaluation of the form of the underlying model. This initial guess as to the form of the model is evaluated in the second and third stages of the modeling procedure, and if inadequacies or misinterpretations are discovered, the model builder returns to the identification stage to reevaluate his initial decision.

Note also from Tables 1, 2, and 3 that system size does indeed have a small effect on the theoretical space-time autocorrelation functions of the models. This is due to the relative influence of the boundary elements. The 5 × 5 system has 16 boundary sites, and 64% of the system is on the boundary. There are 24 boundary sites in a 7 × 7 system, representing 49% of the total. The effect of the boundary then decreases with system size and one would expect the statistical properties of the model to be fairly invariant to system size for large enough systems. In practice, the identification procedure will require a working knowledge of the relationship between size and shape of the system and the theoretical properties of the models.

TABLE 3—*Space-time autocorrelation functions for the STARMA model.*

Space Time Autocorrelations, $\rho_{\ell 0}(s)$

	5x5 system				7x7 system			
Spatial lag (ℓ) / Time lag (s)	0	1	2	3	0	1	2	3
1	0.815	0.664	0.372	0.270	0.808	0.643	0.357	0.256
2	0.594	0.577	0.369	0.265	0.578	0.556	0.352	0.251
3	0.458	0.492	0.347	0.250	0.439	0.470	0.330	0.237
4	0.367	0.418	0.317	0.232	0.346	0.397	0.300	0.219
5	0.300	0.357	0.284	0.212	0.280	0.336	0.268	0.199
6	0.250	0.305	0.252	0.192	0.231	0.286	0.236	0.179
7	0.210	0.263	0.223	0.173	0.192	0.244	0.208	0.161
8	0.178	0.227	0.196	0.156	0.162	0.209	0.182	0.144
9	0.153	0.196	0.173	0.140	0.137	0.180	0.159	0.129
10	0.131	0.171	0.152	0.125	0.117	0.156	0.139	0.114

Space-Time Partial Autocorrelations, $\phi'_{s\ell}$

	5x5 system				7x7 system			
Spatial lag (ℓ) / Time lag (s)	0	1	2	3	0	1	2	3
1	0.815	0.306	-0.052	-0.026	0.808	0.304	-0.054	-0.026
2	-0.309	-0.252	0.026	0.020	-0.310	-0.252	0.029	0.019
3	0.162	0.183	0.014	0.004	0.162	0.185	0.013	0.005
4	-0.096	-0.138	-0.028	-0.009	-0.095	-0.139	-0.028	-0.011
5	0.061	0.102	0.032	0.011	0.061	0.103	0.033	0.013
6	-0.042	-0.078	-0.029	-0.009	-0.041	-0.079	-0.030	-0.010
7	0.030	0.059	0.025	0.008	0.029	0.059	0.026	0.009
8	-0.020	-0.046	-0.020	-0.006	-0.021	-0.046	-0.021	-0.007
9	0.016	0.035	0.016	0.005	0.016	0.035	0.017	0.005
10	-0.012	-0.027	-0.013	-0.004	-0.012	-0.027	-0.013	-0.004

4. ESTIMATION OF THE STARMA MODEL

After a candidate model from the STARMA model family has been chosen during the identification phase of the modeling procedure it is necessary to estimate the parameters. The best estimates of the $\mathbf{\Phi}$ and $\boldsymbol{\theta}$ from many points of view are the maximum likelihood estimates.

Because our basic model formulation has errors that are pure white noise, the distribution of

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}(1) \\ \boldsymbol{\varepsilon}(2) \\ \vdots \\ \boldsymbol{\varepsilon}(T) \end{bmatrix}$$

is multivariate normal with mean $\mathbf{0}$ and variance-covariance matrix equal to $\sigma^2\mathbf{I}_{NT}$. Specifically we have.

$$f(\boldsymbol{\varepsilon}|\mathbf{\Phi},\mathbf{\Theta}\sigma^2) = (2\,\Pi)^{-TN/2}\,|\sigma^2\mathbf{I}_{NT}|^{-1/2}\exp\left(-\frac{1}{2\sigma^2}\,\boldsymbol{\varepsilon}'\mathbf{I}\boldsymbol{\varepsilon}\right)$$

$$= (2\,\Pi)^{-TN/2}\,(\sigma^2)^{-TN/2}\exp\left(\frac{-S(\mathbf{\Phi},\mathbf{\Theta})}{2\sigma^2}\right)$$

where

$$S(\mathbf{\Phi},\mathbf{\Theta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_i(t)^2 .$$

Since the $\boldsymbol{\varepsilon}(t)$ are unobservable random errors and the $\mathbf{z}(t)$ are the quantities actually observed, it is necessary to recursively calculate the $\boldsymbol{\varepsilon}(t)$ from the observed $\mathbf{z}(t)$. The appropriate equations are

$$\boldsymbol{\varepsilon}(t) = \mathbf{z}(t) - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}\mathbf{W}^{(l)}\,\mathbf{z}(t-k)$$

$$+ \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}\mathbf{W}^{(l)}\boldsymbol{\varepsilon}(t-k) \text{ for } t = 1, 2, \cdots T. \tag{15}$$

Immediately we see that the first few $\boldsymbol{\varepsilon}$'s and consequently the entire $\boldsymbol{\varepsilon}$ vector, if moving average terms are present in the model, are functions of observations and errors at times before time 1, the initial epoch observed. Hence without a priori knowledge of these initial starting values, one cannot easily calculate exact m.l.e. of $\mathbf{\Phi}$ and $\mathbf{\Theta}$. This difficulty is best sidestepped by substituting zero, the unconditional mean for all values of $\mathbf{z}(t)$ and $\boldsymbol{\varepsilon}(t)$ with $t < 1$.

The *conditional* likelihood function of $\mathbf{\Phi}$, $\mathbf{\Theta}$ and σ^2 is then

$$L(\mathbf{\Phi},\mathbf{\Theta},\sigma^2 1\mathbf{z}) = (2\Pi)^{-TN/2}\,(\sigma^2)^{-TN/2}\exp\left(-\frac{S_*(\mathbf{\Phi},\mathbf{\Theta})}{2\sigma^2}\right)$$

where $S^*(\mathbf{\Phi},\mathbf{\Theta})$ is the conditional sum of squares function

$$S_*(\mathbf{\Phi},\mathbf{\Theta}) = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

and the $\hat{\boldsymbol{\varepsilon}}$ vector is calculated via (15) with $\mathbf{z}(t)$ and $\boldsymbol{\varepsilon}(t)$ set equal to zero for $t < 1$. The conditional

$$\begin{bmatrix} \mathbf{z}(1) \\ \mathbf{z}(2) \\ \mathbf{z}(3) \\ \vdots \\ \mathbf{z}(T) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{z}(1) & W^{(1)}\mathbf{z}(1) & \mathbf{0} \\ \mathbf{z}(2) & W^{(1)}\mathbf{z}(2) & \mathbf{z}(1) \\ \vdots & \vdots & \vdots \\ \mathbf{z}(T-1) & W^{(1)}\mathbf{z}(T-1) & \mathbf{z}(T-2) \end{bmatrix} \begin{bmatrix} \phi_{10} \\ \phi_{11} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}(1) \\ \boldsymbol{\varepsilon}(2) \\ \boldsymbol{\varepsilon}(3) \\ \vdots \\ \boldsymbol{\varepsilon}(T) \end{bmatrix}$$

DISPLAY 2. The STAR (2_{10}) in general linear model form.

m.l.e.'s of $\sigma^2, \mathbf{\Phi}, \mathbf{\Theta}$ are thus

$$\hat{\sigma}^2 = \frac{S_*(\hat{\mathbf{\Phi}}, \hat{\mathbf{\Theta}})}{TN}$$

and $\hat{\mathbf{\Phi}}$, $\hat{\mathbf{\Theta}}$ that minimize $S_*(\hat{\mathbf{\Phi}}, \hat{\mathbf{\Theta}})$. The closeness of the conditional m.l.e to the exact m.l.e. is a direct function of T. For small T, conditional maximum likelihood estimation is most inappropriate and the more complicated exact m.l.e. procedure must be employed. For moderate or larger values of T however, the conditional procedure closely approximates the exact m.l.e. and is usually adopted at a great savings in computational effort.

Because these conditional maximum likelihood estimates are also least squares estimates, i.e., those parameter values that minimize the residual sum of squares, estimation of the STAR model is based on standard linear regression theory. As an example, consider the STAR(2_{10}) model

$$\mathbf{z}(t) = \phi_{10}\mathbf{z}(t-1) + \phi_{11}\mathbf{W}^{(1)}\mathbf{z}(t-1) + \phi_{20}\mathbf{z}(t-2) + \varepsilon(t)\,.$$

In general linear model form $\mathbf{Y} = \mathbf{XB} + \varepsilon$, this model for $t = 1, 2, \cdots T$ can be written as shown in Display 2. Here, zero vectors have been substituted for the unobserved z vectors, for those times before the system was under observation. The least squares normal equations $(\mathbf{X'X})\mathbf{\Phi} = \mathbf{X'Z}$ in this instance are as shown in Display 3. The parameter vector $[\hat{\phi}_{10}, \hat{\phi}_{11}, \hat{\phi}_{20}]'$ that solves this system of equations is the conditional least squares estimate of the parameters of the STAR(2_{10}) model.

Approximate confidence regions for the parameter values of the STAR model can be constructed using

$$\frac{(\mathbf{\Phi} - \hat{\mathbf{\Phi}})' \mathbf{X'X} (\mathbf{\Phi} - \hat{\mathbf{\Phi}})}{\dfrac{K\, S_*(\hat{\mathbf{\Phi}})}{TN-\mathrm{K}}} \sim F(K, TN-K), \qquad (16)$$

a result from linear regression theory. We should point out that because of the time series nature of the STAR model, the linear regression assumptions about the independent or regressor variable do not hold. Specifically the $\mathbf{X}$ matrix is stochastic rather than fixed in repeated samples. Additionally the regressor variables are not independent of the residual errors and thus the classic results concerning the distribution of the least squares estimates of the parameters do not follow directly. It has been shown, however (Mann and Wald, 1943), that the properties associated with the estimators in the classical linear regression model are possessed in the limit for models such as the STAR. In any case the classical regression results will be used here with the knowledge that they are only approximate. In expression (16), $\mathbf{X'X}$ is the appropriate moment matrix for the model at hand, $\hat{\mathbf{\Phi}}$ is the least squares parameter vector estimate, $S_*(\hat{\mathbf{\Phi}})$ is the residual sum of squares and K is the dimension of $\mathbf{\Phi}$. Also

$$\hat{\sigma}^2 = \frac{S_*(\hat{\mathbf{\Phi}})}{TN-K} \quad \text{and} \quad \frac{(TN-K)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(TN-K)$$

allow confidence intervals for σ^2 to be constructed.

Since only STAR models are linear in form, it is necessary to estimate the parameters of the STMA and STARMA models using any of a variety of nonlinear optimization techniques. Gradient methods (where the gradient is numerically approximated) have found use, as has linearization, an iterative technique that at each stage "linearizes" the nonlinear model using Taylor's expansion and solves approximate normal equations for the next guess at the optimum parameters. The estimation algorithm used in this paper is due to Marquardt (1963) and combines the desirable properties of both these into a single procedure referred to as Marquardt's compromise.

The nonlinear nature of STMA and STARMA models presents special problems in the determination of confidence regions for the parameters. The sum of squares surface $S(\mathbf{\Phi},\mathbf{\Theta})$ and correspondingly the likelihood function is not symmetric, as is the case for linear models, and thus, no closed expression is available for an exact confidence region. To compensate, an approximate likelihood function is used from which an approximate confidence region is calculable.

In general, the exact sum of squares surface can be approximated by expanding about the least squares estimates as follows,

$$S(\mathbf{\Phi},\mathbf{\Theta}) = S(\boldsymbol{\delta}) \approx S(\hat{\boldsymbol{\delta}}) + (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}})'\, \mathbf{Q}(\boldsymbol{\delta} - \hat{\boldsymbol{\delta}})$$

where

$$\boldsymbol{\delta}' = (\mathbf{\Phi}', \mathbf{\Theta}')$$

$$\mathbf{Q} = \tfrac{1}{2}\left[\frac{\partial\, S(\boldsymbol{\delta})}{\partial\, \delta_i\, \partial\, \delta_j}\right]_{\hat{\boldsymbol{\delta}}}$$

for $i = 1, 2, \cdots K$, $j = 1, 2, \cdots K$; K is the dimension of δ, or the number of parameters. Since

$$S(\boldsymbol{\delta}) = \sum_{t=1}^{T} \varepsilon(t)\, \varepsilon(t)'$$

we have

$$\frac{\partial S(\boldsymbol{\delta})}{\partial \delta_i} = \sum_{t=1}^{TN} 2\,\boldsymbol{\varepsilon}(t)' \left.\frac{\partial \boldsymbol{\varepsilon}(t)}{\partial \delta_i}\right|_{\hat{\boldsymbol{\delta}}} = 0$$

$$\frac{1}{2}\left.\frac{\partial^2 S(\boldsymbol{\delta})}{\partial \delta_i \partial \delta_j}\right|_{\hat{\boldsymbol{\delta}}} = \sum_{t=1}^{T} \boldsymbol{\varepsilon}(t)' \left.\frac{\partial^2 \boldsymbol{\varepsilon}(t)}{\partial \delta_i \partial \delta_j}\right|_{\hat{\boldsymbol{\delta}}} + \sum_{t=1}^{T} \left.\frac{\partial \boldsymbol{\varepsilon}(t)'}{\partial \delta_i}\frac{\partial \boldsymbol{\varepsilon}(t)}{\partial \delta_j}\right|_{\hat{\boldsymbol{\delta}}}$$

Now it can be shown that

$$\left.\frac{\delta^2 \boldsymbol{\varepsilon}(t)}{\partial \delta_i \partial \delta_j}\right|_{\hat{\boldsymbol{\delta}}}$$

will in general be a function of $\boldsymbol{\varepsilon}(t)$ occurring before time t and since we expect that if the model fits, $E[\boldsymbol{\varepsilon}(t)\,\boldsymbol{\varepsilon}(t-k)'] = 0$ for $k \geq 1$, the term is neglected. The matrix $\mathbf{Q}$ then can be written

$$\mathbf{Q} = \mathbf{X}'\mathbf{X}$$

where

$$\mathbf{X} = \begin{bmatrix} \left.\frac{\partial \boldsymbol{\varepsilon}(1)}{\partial \delta_1}\right|_{\hat{\boldsymbol{\delta}}} & \left.\frac{\partial \boldsymbol{\varepsilon}(1)}{\partial \delta_2}\right|_{\hat{\boldsymbol{\delta}}} & \cdots & \left.\frac{\partial \boldsymbol{\varepsilon}(1)}{\partial \delta_k}\right|_{\hat{\boldsymbol{\delta}}} \\ \left.\frac{\partial \boldsymbol{\varepsilon}(2)}{\partial \delta_1}\right|_{\hat{\boldsymbol{\delta}}} & \left.\frac{\partial \boldsymbol{\varepsilon}(2)}{\partial \delta_2}\right|_{\hat{\boldsymbol{\delta}}} & \cdots, & \\ \vdots & & & \\ \left.\frac{\partial \boldsymbol{\varepsilon}(T)}{\partial \delta_1}\right|_{\hat{\boldsymbol{\delta}}} & \left.\frac{\partial \boldsymbol{\varepsilon}(T)}{\partial \delta_2}\right|_{\hat{\boldsymbol{\delta}}} & \cdots & \left.\frac{\partial \boldsymbol{\varepsilon}(T)}{\partial \delta_k}\right|_{\hat{\boldsymbol{\delta}}} \end{bmatrix}$$

Thus the sum of squares function is approximated via

$$S(\boldsymbol{\delta}) = S(\hat{\boldsymbol{\delta}}) + (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}})'\,\mathbf{Q}(\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}) \qquad (17)$$

and an approximate $100(1-\alpha)\%$ confidence region for $[\boldsymbol{\Phi}, \boldsymbol{\theta}]' = \boldsymbol{\delta}$ is obtained via the relationship

$$S(\boldsymbol{\delta}) = S(\hat{\boldsymbol{\delta}}) + \frac{K}{TN-K} S(\hat{\boldsymbol{\delta}})\, F_{K,\,TN-K,\,\alpha} \qquad (18)$$

and the quadratic representation of $S(\boldsymbol{\delta})$, equation (17). As a preliminary to the construction of the confidence region, matrix $\mathbf{Q}$ must be numerically estimated. It should be noted that the exact sum of squares function, $S(\boldsymbol{\delta})$, will be replaced by the conditional sum of squares $S_*(\boldsymbol{\delta})$ in the calculation of these confidence regions when conditional maximum likelihood is employed.

As it turns out, matrix Q is also the calculated moment matrix used in the linearization estimation method. Thus, an added benefit of linearization is the ease with which approximate confidence regions can be calculated.

Confidence intervals on σ^2 are calculated via

$$(\sigma^2|\mathbf{z}(1),\mathbf{z}(2), \cdots \mathbf{z}(T)) \sim S_*(\boldsymbol{\delta})\, \chi_{TN-K}^{-2}$$

as in the linear model case.

5. DIAGNOSTIC CHECKING OF THE STARMA MODEL

After a candidate model has been selected and its parameters estimated, the model must be subjected

$$\begin{bmatrix} \sum_{t=1}^{T-1} \mathbf{z}(t)'\,\mathbf{z}(t) & \sum_{t=1}^{T-1} \mathbf{z}(t)'\,W^{(1)}\mathbf{z}(t) & \sum_{t=2}^{T-1} \mathbf{z}(t)'\mathbf{z}(t-1) \\ & \sum_{t=1}^{T-1} \mathbf{z}(t)'\,W^{(1)\prime}W^{(1)}\mathbf{z}(t) & \sum_{t=2}^{T-1} \mathbf{z}(t)'\,W^{(1)\prime}\mathbf{z}(t-1) \\ \text{symmetric} & & \sum_{t=1}^{T-2} \mathbf{z}(t)'\mathbf{z}(t) \end{bmatrix} \begin{bmatrix} \phi_{10} \\ \phi_{11} \\ \phi_{20} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T-1} \mathbf{z}(t)'\,\mathbf{z}(t+1) \\ \sum_{t=1}^{T-1} \mathbf{z}(t)'\,W^{(1)\prime}\,\mathbf{z}(t+1) \\ \sum_{t=1}^{T-2} \mathbf{z}(t)'\,\mathbf{z}(t+2) \end{bmatrix}$$

DISPLAY 3. The conditional least squares normal equations for the STAR (2_{10}) model.

to "diagnostic checks" to determine if the model does adequately represent the data. The model can "fail" in two important ways. Firstly, the model may insufficiently represent the observed correlation of the process. This inadequacy will surface in the form of significant correlation among the residuals of the fitted model. Secondly, the model may be unduly complex. In this instance, estimated parameters will prove to be statistically insignificant.

The first phase of the diagnostic checking stage is the examination of the residuals from the fitted model. If the fitted model adequately represents the data, these residuals should be white noise, i.e., should be distributed normally with mean zero and variance-covariance matrix equal to $\sigma^2\mathbf{I}_N$ and all autocovariances at nonzero lags equal to 0.

Various tests are available for testing the residuals for white noise. Probably the most useful test (especially in the context of the three-stage modeling procedure for space-time models) is that of calculating the sample space-time autocorrelations and partials of the residuals and comparing them to their theoretically derived variance. It has been shown (Pfeifer, 1979) that if the underlying process is pure white noise,

$$\operatorname{var}(\hat{\rho}_{l0}(s)) \approx \frac{1}{N(T-s)}. \qquad (19)$$

If the residuals are approximately white noise, the sample space-time autocorrelation functions should all be effectively zero. If the residuals are not random they may follow a pattern that can be represented by a STARMA model. Identifying this model and coupling it with the candidate that generated the residuals will usually lead to a better updated model.

As an example, consider that the original candidate model was a STAR(1_1) with estimated parameters $\hat{\phi}_{10}$ and $\hat{\phi}_{11}$. Utilizing the backshift operator B

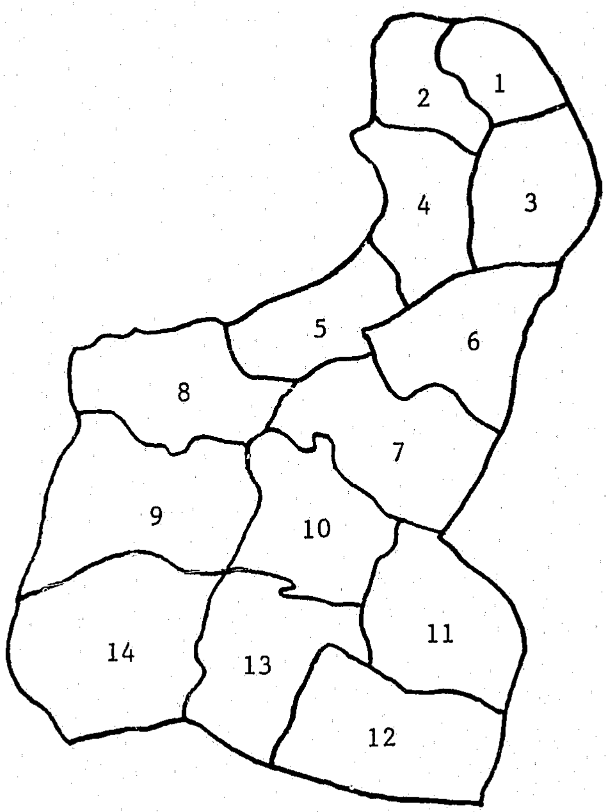

FIGURE 2. The 14 districts of Northeast Boston.

defined as

$$B^s \mathbf{z}(t) = \mathbf{z}(t - s)$$

one can express $a(t)$, the residuals from this model, as

$$\mathbf{a}(t) = (1 - \phi_{10}B - \phi_{11}\mathbf{W}^{(1)}B)\,\mathbf{z}(t).$$

Now suppose that our diagnostic checking of the $\mathbf{a}(t)$ showed a correlative structure suggesting that the $\mathbf{a}(t)$ themselves followed a STAR(1_0) process with parameter value approximately equal to ϕ_{10}^*. Here ϕ_{10}^* is not a least squares estimate, but rather a value suggested by the sample correlations of the residuals.

Hence, if

$$(1 - \phi_{10}^* B)\,\mathbf{a}(t) = \boldsymbol{\varepsilon}(t) \quad \text{where } \boldsymbol{\varepsilon}(t)$$

represents a true white noise process, then

$$\boldsymbol{\varepsilon}(t) = (1 - \phi_{10}^* B)(1 - \phi_{10}B - \phi_{11}\mathbf{W}^{(1)}B)\,\mathbf{z}(t)$$

is the updated model. Thus, the new candidate model would be a STAR(2_{11}):

$$\boldsymbol{\varepsilon}(t) = (1 - (\phi_{10}^* + \phi_{10})B - \phi_{11}\mathbf{W}^{(1)}B + \phi_{10}^*\phi_{10}B^2 + \phi_{10}^*\phi_{11}\mathbf{W}^{(1)}B^2)\,\mathbf{z}(t)\,.$$

To complete this example, it should be noted that if $\phi_{10}^*\ \phi_{11}$ was close to zero, the model builder might decide to nominate the STAR(2_{10}) as his next candidate. With the specification, or reidentification, of the next candidate model, the procedure once again returns to the estimation stage.

The second phase of the diagnostic checking stage involves checking the statistical significance of the estimated parameters. This is done via the confidence regions for the parameters presented in the last section. A calculated region that contains the null vector does not reject the hypothesis that $\mathbf{\Phi} = \mathbf{\Theta} = \mathbf{0}$ at the level of significance of the confidence region. A more useful test is one that tests the hypothesis that a particular ϕ_{kl} or θ_{kl} is zero with the remaining parameters in the model unrestricted. If we let $\hat{\delta}$ represent the least squares estimate of the full parameter vector (containing the parameter to be tested) and $\hat{\delta}^*$ be the least squares estimate of the parameter vector with δ_K (without loss of generality we assume the parameter to be tested is the last entry in the δ vector) constrained to be zero. The appropriate test for the hypothesis that $\delta_K = 0$ is based on the statistic

$$\frac{(TN - K)\,[S_*(\hat{\delta}^*) - S_*(\hat{\delta})]}{S_*(\hat{\delta})} \tag{20}$$

which is approximately distributed as an $F_{1,TN-K}$ under the null hypothesis.

Any estimated parameter that proves to be statistically insignificant should be removed from the model and the resulting simpler model would now be considered as the candidate for acceptance. The model building procedure then moves once again to the estimation stage. Removal of unsubstantiated parameters is necessary as part of the model builder's search for what has become known as a *parsimonious* model. Parsimonious models are those that are "efficient" in their use of parameters. They possess maximum simplicity and the smallest number of parameters consonant with representational accuracy.

The three-stage modeling procedure continues through identification, estimation and diagnostic checking until at some point the model at hand "passes" the diagnostic checking stage. To do this, the model must evidence parameters that are all significant and residuals that can effectively be consid-

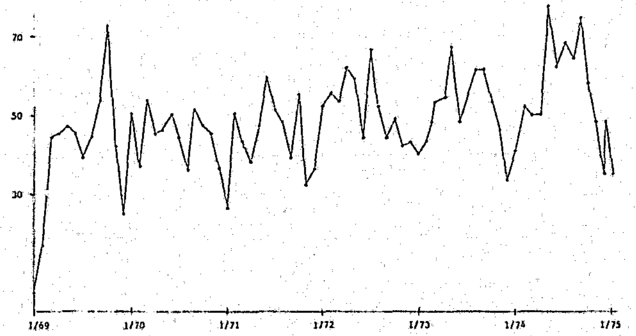

FIGURE 3. Total assault arrests in Northeast Boston.

ered to be white noise. At the conclusion of the modeling procedure the STARMA model is ready to employ.

6. SPACE-TIME MODELING BOSTON ASSAULT ARRESTS

To illustrate the space-time modeling procedure, raw arrest data from the 822 reporting areas of the city of Boston were combined to give monthly figures for the 14 areas of Northeast Boston depicted in Figure 2. Data were available for the years 1969 to 1974, and the crime type chosen to model was total assaults. Thus, the system of interest consists of $N = 14$ sites and was observed for $T = 72$ time periods. The total over all 14 sites of assault arrests is pictured in Figure 3.

Table 4 gives the delineation of the spatial neighbors of each of the 14 districts. The choices reflected in Table 4 were made to best depict the spatial ordering of Figure 1.

Weights for this system were also specified to agree with the equally weighted formulation presented in Section 2. As an example of this particular weighting scheme, $\mathbf{W}^{(1)}$ is presented in Figure 4. Note that each row sums to one, and that nonzero entries occur only for those pairs of points that are first order neighbors.

Initial identification of the assault data suggested that the system was nonstationary and required a first difference. This is evidenced by an autocorrelation function that decayed from the value of one at a very slow rate.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		$\frac{1}{2}$	$\frac{1}{2}$											
2	$\frac{1}{2}$			$\frac{1}{2}$										
3	$\frac{1}{3}$			$\frac{1}{3}$		$\frac{1}{3}$								
4		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$								
5				$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$						
6			$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$							
7					$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$				
8					$\frac{1}{3}$		$\frac{1}{3}$		$\frac{1}{3}$					
9								$\frac{1}{3}$		$\frac{1}{3}$				$\frac{1}{3}$
10							$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$	
11										$\frac{1}{2}$		$\frac{1}{2}$		
12											$\frac{1}{2}$		$\frac{1}{2}$	
13										$\frac{1}{3}$		$\frac{1}{3}$		$\frac{1}{3}$
14									$\frac{1}{2}$				$\frac{1}{2}$	

FIGURE 4. $\mathbf{W}^{(1)}$ for Northeast Boston.

TABLE 4—*Neighbors of each site for each spatial order.*

ORDER	1	2	3
Site 1	2,3	4	6
2	1,4	3	5,6
3	1,4,6	2	5,7
4	2,3,5,6	1	7,8
5	4,6,7,8		3,9,10,2
6	3,4,5,7		1,2,8,10
7	5,6,8,10	11	3,4,9,13
8	5,7,9	10	4,6,14
9	8,10,14	13	4,6,14
10	7,9,11,13	8,12,14	5,6
11	10,12	7,13	9
12	11,13	10	14
13	10,12,14	9,11	7
14	9,13	10	8,12

The sample space-time autocorrelation and partial autocorrelation function of the differenced series is presented in Table 5. Because the space-time partials tail off and the space-time autocorrelation function seems to cut off both spatially and temporally after one lag, the differenced series was tentatively identified as a STMA(1_0) process. Incorporating the difference into the model notation, this can be referred to as a STARIMA(0,1,1_0), where STARIMA stands for space-time autoregressive integrated moving average process, and the extra 1 is for the first difference.

Estimating this model gave $\hat{\theta}_{10}$ to be 0.803, and the residual sum of squares was 6504.679. The estimated variance of the parameter estimate was 0.000368, and thus an approximate 95% confidence interval for θ_{10} is (0.764, 0.841). Table 6 exhibits the sample space-time correlation functions for the residuals from this model.

Examination of Table 6 reveals relatively large values for $\hat{\rho}_{10}(1)$ (compared to an approximate variance of (14×70^{-1})) and ϕ'_{11} suggesting that the residuals evidence some spatial correlation. For this reason a new model, the STARIMA (0, 1, 1_1), was

TABLE 5—*Space-time correlation functions of the differenced series.*

Space-Time Autocorrelations, $\hat{\rho}_{\ell 0}(s)$

Spatial lag (ℓ) / Time lag (s)	0	1	2	3
1	-0.484	0.007	0.041	0.013
2	0.023	-0.038	0.012	-0.027
3	-0.017	0.004	-0.045	0.049
4	0.026	0.013	0.019	-0.038
5	-0.056	0.039	0.005	0.017
6	0.043	-0.074	0.045	-0.021
7	-0.003	0.015	-0.091	-0.018
8	-0.032	0.015	0.053	0.037

Space-Time Partial Autocorrelations, $\phi'_{s\ell}$

Spatial lag (ℓ) / Time lag (s)	0	1	2	3
1	-0.484	0.068	0.007	0.020
2	-0.281	0.015	0.034	-0.011
3	-0.196	0.068	-0.026	0.070
4	-0.111	0.015	-0.025	0.028
5	-0.140	0.004	-0.021	0.030
6	-0.092	0.025	0.071	-0.007
7	-0.048	0.138	-0.036	-0.055
8	-0.073	0.008	-0.011	0.008

TABLE 6—*Sample space-time autocorrelation functions of the residuals from the model* $\mathbf{z}(t) - \mathbf{z}(t-1) = -0.803\ \varepsilon(t-1) + \varepsilon(t)$.

Space-Time Autocorrelations, $\hat{\rho}_{\ell 0}(s)$

Spatial lag (ℓ) / Time lag (s)	0	1	2	3
1	0.024	0.084	0.051	0.058
2	0.031	0.016	0.023	0.027
3	-0.007	0.027	-0.026	0.046
4	-0.002	0.038	0.021	0.016
5	-0.056	0.024	0.029	-0.014
6	0.000	-0.061	0.026	-0.041
7	-0.031	-0.009	-0.062	-0.026
8	-0.046	0.011	0.030	0.021

Space-Time Partial Autocorrelations, $\hat{\phi}'_{s\ell}$

Spatial lag (ℓ) / Time lag (s)	0	1	2	3
1	0.024	0.133	0.042	0.055
2	0.019	-0.007	0.016	0.010
3	-0.013	0.028	-0.042	0.045
4	-0.009	0.058	0.015	-0.044
5	-0.060	0.047	0.031	-0.035
6	-0.003	-0.094	0.039	-0.050
7	-0.022	0.003	-0.064	-0.021
8	-0.043	0.064	0.039	0.042

entertained. The extra "spatial" parameter θ_{11} in this model will hopefully describe the spatial structure of the residuals.

Estimation of the STARIMA (0, 1, 1_1) yielded $\hat{\theta}_{10} = 0.812$, $\hat{\theta}_{11} = -0.092$ and a residual sum of squares of 6457.266. The sample correlative properties of the residuals from this model are given in Table 7. We note that the first spatial order autocorrelation and partial autocorrelation have both decreased, and in general there seems to be a lack of structure in these residuals. The lone exception occurs at time lag 6 and spatial lag 1 pointing to a possibility that further investigation with respect to seasonal forms of the STARMA might prove useful. Testing the significance of the θ_{11} parameter is done via equation (20). Here $N = 14$, $T = 71$, $K = 2$, $S(\hat{\delta}^*) = 6504.679$ and $S(\hat{\delta}) = 6457.266$. Combining these via (20) gives an approximate $F_{1,992}$ value of 7.3. The theoretical critical F value at $\alpha = 0.01$ with 1 and 992 degrees of freedom is approximately 6.7. Thus, θ_{11} is significant with 99% confidence.

TABLE 7—*Sample space-time autocorrelation functions of the residuals from the model* $\mathbf{z}(t) - \mathbf{z}(t-1) = -0.812\ \varepsilon(t-1) + 0.092\ \mathbf{W}(1)\ \varepsilon(t-1) + \varepsilon(t)$.

Space-Time Autocorrelations, $\hat{\rho}_{\ell 0}(s)$

Spatial lag (ℓ) / Time lag (s)	0	1	2	3
1	0.023	0.045	0.036	0.046
2	0.023	0.062	0.014	0.008
3	-0.010	-0.014	-0.043	0.041
4	-0.004	0.029	0.015	-0.046
5	-0.053	0.020	0.034	-0.036
6	0.007	-0.120	0.042	-0.050
7	-0.023	-0.023	-0.062	-0.023
8	-0.044	0.036	0.041	0.040

Space-Time Partial Autocorrelations, $\hat{\phi}'_{s\ell}$

Spatial lag (ℓ) / Time lag (s)	0	1	2	3
1	0.023	0,030	0.036	0.042
2	0.028	-0.026	0.010	0.011
3	-0.008	-0.009	-0.036	0.030
4	-0.001	0.016	0.014	-0.029
5	-0.055	0.003	0.027	-0.028
6	-0.003	-0.074	0.026	-0.051
7	-0.032	-0.022	-0.058	-0.034
8	-0.046	0.004	0.034	0.015

Because the sample space-time autocorrelations and partials show a lack of structure (with the noted exception) characteristic of a white noise sequence, and since all parameters have proven statistically significant, the STARIMA (0, 1, 1_1) model

$$\mathbf{z}(t) - \mathbf{z}(t-1) = -0.812\varepsilon(t-1) + 0.092\mathbf{W}^{(1)}\varepsilon(t-1) + \varepsilon(t)$$

passes the diagnostic checking portion of the three-stage modeling procedure. This model is now ready to employ either as a forecasting function for the number of arrests for assault in the 14 districts of Northeast Boston or as part of a more sophisticated control system.

7. CONCLUSION

A three-stage iterative procedure is presented for building space-time autoregressive moving average (STARMA) models. These models are characterized by autoregressive and moving average terms lagged both in space and time and are useful toward modeling systems that exhibit spatial autocorrelation. The three model building stages, identification, estimation and diagnostic checking, are presented and illustrated with a substantive example.

8. ACKNOWLEDGEMENTS

This work was performed under Grants No. 76NI-99-091 and 78NI-AX-004 of the National Institute of Law Enforcement and Criminal Justice. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the United States Department of Justice.

The authors would like to thank the referees for their valuable comments.

REFERENCES

BESAG, J. S. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc., B, 36,* 197–242.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

BOX, G. E. P. and TIAO, G. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Assoc., 70,* 70–79.

CLIFF, A. D., HAGGETT, P., ORD, J. K., BASSETT, K. A. and DAVIES, R. B. (1975). *Elements of Spatial Structure: A Quantitative Approach.* New York: Cambridge University Press.

CLIFF, A. D. and ORD, J. K. (1973). *Spatial Autocorrelation.* London: Pioneer.

CLIFF, A. D. and ORD, J. K. (1975). Space-time modeling with an application to regional forecasting. *Trans. Inst. British Geographers, 66,* 119–128.

DEUTSCH, S. J. (1978). Stochastic models of crime rates. *Intl. J. Comparative and Applied Criminal Justice, 2,* 127–151.

DEUTSCH, S. J. and ALT, F. B. (1977). Evaluation of gun control law on gun-related crimes in Boston. *Evaluation Quarterly, 1,* 543–567.

DEUTSCH, S. J. and OGELSBY, G. B. (1979). Analysis of the toxic effects of physical and chemical properties of hypolimnetic waters by time series. *Environment International, 2, 3,* 133–138.

DEUTSCH, S. J. and WU, S. M. (1974). Analysis of wear during grinding by empirical-stochastic models. *Wear, 29,* 247–257.

DURBIN, J. (1960). The fitting of time series models. *Intl. Statist. Rev., 28,* 233.

GOLDBERGER, A. S. (1964). *Econometric Theory.* New York: John Wiley & Sons.

GRANGER, C. W. J. and NEWBOLD, P. (1977). *Forecasting Economic Time Series.* New York: Academic Press.

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

MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrika, 11, 173,* 270.

MARQUARDT, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Indust. and Applied Math., 11,* 431–441.

MARTIN, R. L. and OEPPEN, J. E. (1975). The identification of regional forecasting models using space-time correlation functions. *Trans. Inst. British Geographers, 66,* 95–118.

MCMICHAEL, F. C. and HUNTER, J. S. (1972). Stochastic modeling of temperature and flow in rivers. *Water Resources Research, 8,* 87–98.

PFEIFER, P. E. (1979). *Spatial-Dynamic Modeling.* Unpublished Ph.D. dissertation, Georgia Institute of Technology, Atlanta, Georgia.

PHADKE, M. S. and WU, S. M. (1974). Identification of multi-input–multi-output transfer function and noise model of a blast furnace from closed-loop data. *IEEE Trans., AC-19,* 944–951.

END