95232

MULTIVARIANT CONSIDERATIONS

IN

SPACE-TIME MODELING*

Prepared By

Stuart Jay Deutsch

and

C.C. Wang

*A final report for grant #81-IJ-CX-0020 from the National Institute of Justice to the Georgia Institute of Technology; Principal Investigator Dr. Stuart Jay Deutsch, Professor, School of Industrial and Systems Engineering, Atlanta, GA 30332

95232

MULTIVARIANT CONSIDERATIONS

IN

SPACE-TIME MODELING*

Prepared By

Stuart Jay Deutsch

and

C.C. Wang

*A final report for grant #81-IJ-CX-0020 from the National Institute of Justice to the Georgia Institute of Technology; Principal Investigator Dr. Stuart Jay Deutsch, Professor, School of Industrial and Systems Engineering, Atlanta, GA 30332

ACKNOWLEDGEMENT

This work was partially performed under Grant Number 81-IJ-CX-0020 from 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Autoregressive Moving Average Model Class (ARMA models) and the associated three stage iterative model building techniques of identification, estimation and diagnostic checking was proposed by Box and Jenkins [1970]. Since its inception these methods have had wide spread success in problems of description, forecasting and control for univariate systems. These univariate time series methods were adapted by Box and Tiao [1975] to allow their use in intervention analysis or the detection of a change in level of a time varying attribute or the detection of a change in level of a time varying attribute after the process is intentionally altered to induce change. The intervention models as the univariate time series models have similarly had wide applicability and success in numerous engineering, science and social science problem settings.

The Box-Jenkins univariate models have been generalized to accomodate systems that vary in space and time by Pfeifer and Deutsch [1980a]. These models are referred to as Space-Time Autogregressive Moving Average Models (STARMA models). The STARMA model class takes the form

$$\underset{\sim}{z}_t = \sum_{k=1}^{p}\sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} \underset{\sim}{z}_{t-k} - \sum_{k=1}^{q}\sum_{\ell=0}^{m_k} \theta_{k\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t$$

where

$\underset{\sim}{Z}_t$ is the LN×1 vector of observations at time t,

LN is the number of locations in the system,

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,

$\phi_{k\ell}$ is the autoregressive parameter at temporal lag k and spatial lag ℓ,

$\theta_{k\ell}$ is the moving average parameter at temporal lag k and spatial lag ℓ,

$W^{(\ell)}$ is the LN×LN matrix of weights for spatial order ℓ, and

$\underset{\sim}{\varepsilon}_t$ is the random normally distributed error vector at time t with

$$E(\underset{\sim}{\varepsilon}_t) = 0$$

$$E(\underset{\sim}{\varepsilon}_t \underset{\sim}{\varepsilon}_{t+s}') = \begin{cases} G & s=0 \\ 0 & s\neq 0 \end{cases}$$

$$E(\underset{\sim}{Z}_t \underset{\sim}{\varepsilon}_{t+s}') = 0 \quad \text{for } s > 0$$

By constraining the number of locations to one (LN=1) the STARMA model class collapse to the ARMA univariate model class.

A complete literature review leading to the STARMA models is contained in Pfeifer [1979]. In addition to formulating the space-time moving average class, Deutsch and Pfeifer developed the methodological considerations for a three stage iterative model building procedure for the space-time model, parallel to that proposed by Box and Jenkins' for univariate series (see Pfeifer and Deutsch [1979, 1980a, 1980b, 1980c, 1980d, 1980e, 1981a, 1981b]). The STARMA model class and associated model building was extended by Pfeifer and Deutsch [1981c] to describe seasonal phenomena and by Deutsch and Pfeifer [1981d] to incorporate contemporaneously correlated innovations. All other appropriate literature citations are contained in context in each chapter.

The purpose of this final report is to extend the currently available methodological procedures/capability of the current state-of-the-art in intervention analysis and system description of the STARMA time series methods. Each of the modeling extensions is stimulated by real world problems. Thus each of the modeling extensions in each chapter are substantively illustrated with case studies. These example applications are in behavioral science, criminology, air pollution and water resources settings.

In Chapter II, the multi-consequence intervention modeling procedure is developed based on the univariate ARMA process. The multi-sequence ARMA(p,d,q)MCI allows for the description of a change in mean level and covariance in a process due to an intervention initiated at $t \leq n_1$ and takes the form, pre-intervention:

$$\Phi_p(B)\nabla^d(Z_t-\mu) = \theta_q(B)a_t, \quad t=1,2,\ldots,n_1$$

post-intervention:

$$\Psi_p(B)\nabla^d(Z_t-\mu-\delta(t)) = \Gamma_q(B)a_t, \quad t=n_1+1,\ldots,n_1+n_2$$

where $\delta(t)$ is the realized intervention effect, and $\Phi_p(B)$, $\theta_q(B)$, $\Psi_p(B)$, $\Gamma_q(B)$ are functions of process parameters that contain the information of the process covariance structure. Since $\Psi_p(B)$ and $\Gamma_q(B)$ can be distinguished from $\Phi_p(B)$ and $\Theta_q(B)$, respectively, the above model has the capability of describing the process covariance change. $\delta(t)$ can be expressed as,

$$\delta(t) = k(t)\delta$$

where $k(t)$ are known numerical values that can be computed from the model parameter values and the intervention model specification, and δ is the intrinsic program utility. The realized intervention effect $\delta(t) = \delta$ if $k(t) = 1$, which is the non-environment influenced situation since the intrinsic program utility is realized fully. However, when $k(t) \neq 1$, then the intrinsic program utility is not realized fully but is masked by the ecclectic environmental process in place. Situations arise where ~~that~~ the intervention effect is not known to be ~~the~~ influenced by

the environment ~~or not~~, and it is necessary to identify the nature of environment influence to build the intervention model. A dynamic component identification procedure is developed in this chapter to identify the interactive relationship between the process environment and the intervention program utility. The biases caused by model mis-specification are derived for the point estimate as well as the interval estimate of δ. The statistics needed for testing the hypothesis of multi-consequence are also developed. A behavioral science example is given to illustrate the non-environmental modeling building situation and a criminology example is used to illustrate the environmental influence situation.

The univariate multi-consequence intervention model in Chapter II is followed by the space-time intervention modeling chapter. In Chapter III, the system contains more than one location and the intervention program is assumed to be initiated at any chosen locations. The single space-time intervention model, STARMA$(p_{\underset{\sim}{\lambda}},d,q_m)I_{\underset{\sim}{m}}$, is formulated as;

$$\Phi_{p,\underset{\sim}{\lambda}}(B)\nabla^d(\underset{\sim}{Z}_t-\mu) = [(1-I_m)\Phi_{p,\underset{\sim}{\lambda}}(B)\nabla^d + I_m\Theta_{q,m}(B)]\underset{\sim}{\xi}(t) + \theta_{q,m}(B)\underset{\sim}{\varepsilon}_t$$

where

$\underset{\sim}{\xi}(t) = \underset{\sim}{\delta}\xi_t$

$\underset{\sim}{\delta}$ is the intrinsic program utility vector, and

ξ_t is the indicator variable, which it takes the value 0 for pre-intervention periods and 1 for the post-intervention periods. In the above model, the I_m parameter is embedded to distinguish whether the intervention effect is the environmentally influenced or non-environmentally influenced.

An alternative representation is developed to decompose the process into two mutually exclusive components, the random component and the deterministic component. The diffusion behavior of the deterministic component is physically interpreted and two diffusion process types, the regenerating diffusion type and the relocation diffusion type, are characterized by the stationary and non-stationary $(p_{\underset{\sim}{\lambda}}, d, q_{\underset{\sim}{m}})$ I_m model. Simulations are performed to illustrate the diffusion phenomena characterized by the (STAR)I_m, (STMA)I_m, (STARMA)I_m processes. Situations arise where that the nature of the intervention program is unknown, so a procedure for the identification of the dynamic component is developed. A transformation formula that transforms the space-time intervention model into linear model form is developed, and the results of linear model are applied to obtain the point estimation and the interval estimation. A substantive air pollution example that contains two interventions is given to illustrate the model building procedures described in this chapter.

In previous STARMA modeling procedures, the system relationships were assumed to be spatially and temporally uniform. In Chapter IV, non-equal diffusion preference systems are considered. The component of STARMA model to be modified in order to describe diffusion prefe-

rence is examined in scaled weight matrices. Two approaches, the strip region approach and the angular region approach, are proposed to construct the non-equally preferential neighbor structure that reflects the non-equally preferential diffusion processes. Simulations are performed for one-direction preference system and two-direction preference system to illustrate the relationship between the non-equally preferential diffusion process and the corresponding weight matrices. Then the theoretical residual patterns, due to unexhausted non-equally preferential structures are analyzed. This analysis is followed by simulation examples to illustrate the analytical conclusion. The air pollution data used in Chapter III is revisited to illustrate the construction of the non-equally preferential neighbor structure and the construction of non-equally preferential models.

In Chapter V, the purely spatial model is introduced. The purely spatial model contains only contemporaneous terms, therefore, it doesn't have the capability to capture any spatial-temporal correlation structure. The existence conditions that corresponds to the stationary conditions and the invertible conditions of the space-time model are derived. For use in identification, the purely spatial autocorrelation functions and the purely spatial partial autocorrelation functions are defined and their statistical properties derived. Computational algorithms are developed to make the computer effort more efficient. Pattern recognition is also proposed to assist in identification. The expectation values of the sample purely spatial autocorrelation function are computed for the low order models and charts are developed for identification. The purely spatial process has no observation and noise initial value problem in estimation, however, the initial parameter values are still needed to start the required iterative estimation procedures. Charts to obtain initial parameter values for low order spatial models are developed. Two application examples are given to illustrate the purely spatial model building procedure. The first example is from hydrology and the other is from criminology. In these examples, the purely spatial modeling procedures are applied to enhance the descriptive capability of the system.

In Chapter VI, two topics are addressed: the coupling and reparameterizing of the aggregate purely spatial model and the space-time model and the coupling of purely spatial models. The coupled models are capable of capturing the contemporaneous purely spatial correlative structure, as well as, the space-time correlative structure. Three potential space-time models are coupled with the purely spatial model, and the resulting models are then reparameterized. Here two modeling sequences are possible: 1) building the purely spatial observation model first and then the space-time residual model, or 2) building the space-time observation model first and then the purely spatial residual model. The coupled and reparameterized models may be dependent on the modeling sequence or may be independent of the modeling sequence. The systems that give rise to each are distinguished and the appropriate modeling procedures are described. Two processes that were modeled in Chapter V as separate but simultaneous space-time and purely spatial models, are presented to illustrated the coupling and reparameterizing procedures.

The purely spatial structures contained in each observation period may be identical or may be different. If the purely spatial structures in each observation period are identical, then the system is said to be ergodic and the aggregate purely spatial model captures the same purely spatial structure with better precision. However, if the purely spatial structures in each observation period are different, then the purely spatial structures are mixed and the resulting aggregated purely spatial model describes an average correlative structure. The second topic in Chapter VI is the development of the modeling procedures for the ergodic process. Here statistics are developed to test the process' ergodic property. Since these homogeneity assumptions may not hold and may mask the ergodic property, therefore a modeling procedure that estimates and corrects the outliers to obtain homogeneity is developed to model the potential ergodic processes that contain outliers. An example is given to illustrate the modeling procedure of the ergodic process with outliers. The masking effect of the outliers are illustrated in details. Forecast functions are constructed for the ergodic model as well as the model which assumes but doesn't verify the homogeneity assumption to compare the forecasting and descriptive capabilities. Differences are explained in detail.

The STARMA model captures the spatial-temporal correlated structures, the multivariate ARMA model captures the inter-category correlated structures. A natural generalization of the STARMA model and the multivariate ARMA model is the MULSTARMA(MULtivariate STARMA) model that captures all the spatial-temporal, inter-category correlated structures. The MULSTARMA($\underset{\sim}{\xi}$,$\underset{\sim}{p}$,$\underset{\sim}{q}$,$\underset{\sim}{\lambda}$,$\underset{\sim}{m}$) model class takes the form;

$$\underset{\sim}{z}^h(t) = \sum_{g=1}^{\xi} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} W^{(\ell)} \underset{\sim}{z}^g(t-k)$$

$$- \sum_{g=1}^{\xi} \sum_{k=1}^{q^{hg}} \sum_{\ell=0}^{m_k^{hg}} \theta_{k\ell}^{hg} W^{(\ell)} \underset{\sim}{\varepsilon}^g(t-k) + \underset{\sim}{\varepsilon}^h(t).$$

$$h=1,2,\ldots,\xi.$$

A specification of the system parameters ξ, the category number, and the model parameters $\underset{\sim}{p}$, $\underset{\sim}{q}$, $\underset{\sim}{\lambda}$, $\underset{\sim}{m}$ serves to define one MULSTARMA model from the general familty of models. In this MULSTARMA model class, the spatial-temporal, inter-category correlated structures are captured by the estimatable parameters ϕ_k^{hg}, θ_k^{hg} and the weight matrix. To illustrate the model structure, a series of flow charts are plotted. These flow charts contain filters that correspond to identifiable terms in the MULSTAR model formulation. The stationary regions and the invertible regions are derived. Then the Λ-weight representation for the stationary process is derived. The Λ-weight representation express the process observation as summation of past errors, which allow derivation of statistical properties to be made easier. The multivariate space-time autocorrelation function and the multivariate space-time partial autocorrelation function are defined, and the statistical properties are derived to help identify the candidate model. Due to the high dimensionality of the model parameter space,

computational efficiency deserves attention. Therefore, efficient computation approaches are introduced and justified. Based on linear model theory, estimation procedures are developed. A criminology example is used to illustrate the multivariate space-time modeling procedure. In this example, the employed model is used in constructing forecast functions, that rely on the spatial, temporal and inter-category structure, and is augmented for use in intervention modeling.

The final report concludes in Chapter VIII with a discussion of the conclusions of this work.

CHAPTER II

MULTICONSEQUENCE INTERVENTION MODEL

The analysis of correlated time series data for changes in level using autoregressive integrated moving average (ARIMA) processes was first introduced by Box and Tiao (1965) using the ARIMA(0,1,1) form. Their work was extended by Glass, Willson and Gottman (1975) to include other types of ARIMA processes while focusing the statistical methods in a quasi-experimental design/interrupted time series framework. Here, an intervention is thought of as affecting the observation, Z_t, between n_1 and n_1+1. The process is described as follows:

Preintervention:

$$\phi_p(B)\nabla^d(Z_t - \mu) = \theta_q(B)a_t \quad t=1,2,\ldots,n_1 \tag{2-1}$$

Postintervention:

$$\phi_p(B)\nabla^d(Z_t - (\mu + \delta)) = \theta_q(B)a_t \quad t=n_1+1,\ldots n_1+n_2 \tag{2-2}$$

Where:

$\phi_p(B) = 1-\phi_1B-\phi_2B^2-\ldots-\phi_pB^p$

$\theta q(B) = 1-\theta_1B-\theta_2B^2-\ldots-\theta_qB^q$

$\nabla = (1-B)$

B is a backward shift operator such that

$$(BZ_t=Z_{t-1}, B\varepsilon_t=\varepsilon_{t-1}) \quad ,$$

μ is the process mean,

δ is the magnitude of change induced by a modification of the process, that is associate with a type of intervention activity,

the ϕ's and θ's are autoregressive and moving average parameters, respectively and the

a_t's are innovations distributed normally and independent with mean zero and variance σ_a^2 .

These models assume (1) only a single consequence in that after intervention only the mean level not the covariance can change and (2) the change in the process after time T is instantaneous (e.g., having been fully realized in the n_1+1^{st} observation).

This chapter focuses on the multiconsequence intervention model that allows for both a change in the mean level of the process as well as the possibility for a modification in the covariance structure. The model structures contain a mean shift function to allow for instantaneous or transient modification, depending upon the nature of the interrupted time series experiment and/or the environmental process in place. In Section 2.1 the multiconsequence model structures are described along with its mean shift function that allows the estimation of an intervention program's intrinsic utility which is not identical to the program's realized effect. The next section contains the maximum likelihood estimation procedures for the multiconsequence intervention model of order (p,d,q) for the mean shift function known or unknown. Section 2.3

develops the necessary covariance matrix specifications for low order multiconsequence models that are used in estimation and tests of significance of the mean shift function and other model parameters. These hypothesis tests are contained in Section 2.4. In Section 2.5, the effects on the ability of ~~statistically determining~~ to determine the significance of an intervention program intrinsic value due to the misspecification of the mean shift function form and/or the use of single consequence intervention model form when there is change in the covariance structure are addressed. Section 2.6 discusses the bias considerations from misspecification intervention model form both with respect to the autoregressive and moving average parameters and the mean shift function. The modeling procedures are algorithmically described in Section 2.7 for situations that arise in the analysis of interrupted correlated time series designs. The ability to statistically detect a given magnitude of a mean level change in correlated interrupted time series design is dependent upon the number of pre and post observations, and the magnitude of the change in the covariance structure after intervention. Section 2.8 analyses these effects and develops guidelines for designing interrupted time series experiments from power considerations. Tables for pre and post intervention sample sizes are given for designing experiments in Section 2.9. Lastly, in Section 2.10, two substantive examples are given to illustrate the modeling procedures described. One example corresponds to the analysis of a direct stimulus-response interrupted time series experiment in which the environmental process does not influence the intervention changes. The second example illustrates the modeling procedures in experiments

in which the existing environmental process affects the realized modifications.

2.1 Multiconsequence Model Structure

The effect of an intervention is not typically expected to be directly transmitted to the observations. Depending upon the nature of the intervention activities, the transmittal of the intervention effects, δ, can be direct or influenced by the ecclectic environmental phenomena in place. When an intervention program is typified by direct stimulus to subjects whose response is monitored, the mean level of the intervention effect of the stimuli would be transmitted directly to the attribute being monitored. Similarly when the ecclectic environmental process is random or weakly correlated, the transmittal of the intervention effect would also be direct. These situations are termed the instantaneous case. On the other hand, when the intervention activities are not direct stimulus to subjects whose response is monitored but rather the stimulus is intended to alter the overall ecclectic environmental phenomena and in turn modify behavior of a segment of the population, the mean level of the intervention effect would be filtered by the environmental process causing a delayed steady state realization in the attribute monitored. The situation is labeled the transient case.

In either situation, once the intervention activities are initiated, in addition to modification of the realized mean level of the attribute monitored, a second simultaneous consequence, the change in the postintervention covariance structure can occur. This may be thought of as occurring due to behavior modification of the subjects or treatment group in the instantaneous case or due to the formation of a new ec- cletic environmental process immediately after intervention for the mean level transient case.

A comprehensive model that describes both instantaneous and transient mean level changes and postintervention modification of the covariance structure is the multiconsequence intervention model form: preintervention;

$$\Phi_p(B)\nabla^d(Z_t - \mu) = \Theta_q(B)a_t \quad t=1,2,\ldots,n_1 \qquad (2\text{-}3)$$

postintervention;

$$\Psi_p(B)\nabla^d(Z_t - \mu - \delta(t)) = \Gamma_q(B)a_t \quad t=n_1+1,\ldots,n_1+n_2 \qquad (2\text{-}4)$$

where

$\Psi_p(B) = 1-\psi_1B-\psi_2B^2-\ldots-\psi_pB^p$

$\Gamma_q(B) = 1-\gamma_1B-\gamma_2B^2-\ldots-\gamma_qB^q$

the ψ's and γ's are postintervention autoregressive and moving average parameters respectively and

$\delta(t)$ is the mean shift function.

Alternatively the postintervention model can be expressed in terms of an intervention transfer function operator, T(B), and the intervention input variable, ξ_t where $\xi_t=0$ for $t<n_1$ and 1 for $t\geq n_1$;

$$\Psi p(B)\nabla^d(Z_t - \mu) = T(B)\xi_t\delta + \Gamma_q(B)a_t \quad t=n_1+1,\ldots,n_2+n_2 \qquad (2\text{-}5)$$

where δ is the intrinsic utility of the process intervention.

From the expectations of Equations (2-4) and (2-5), the mean shift function can be expressed as;

$$\delta(t) = T(B)\xi_t\delta/\Psi_p(B)\nabla^d \quad t=n_1+1,n_1+2,\ldots,n_1+n_2 \qquad (2\text{-}6)$$

Figure 2-1(a) and (b) are schematic representations of the coupling of the pre and post intervention structures for the multiconsequence intervention model for systems in which the environment influences or does not influence the realized program utility, respectively. As seen from Figure 2-1(a), even when allowing for environmental influences, mathematically when $T(B) = \Psi_p(B)$, the postintervention environmental process does not affect the change in mean associated with the intervention in that $\delta(t) = \xi_t\delta$ as in Figure 2-1(b). The more similar T(B) is to $\Psi_p(B)$ the smaller the influence of the environmental process. Also mathematically, when $T(B) = \Gamma_q(B)$, the environmental process directly influences the magnitude of the realized intervention $\delta(t)$. It should be noted that this influence or variation in the observed mean shift from the program utility, δ, is dependent upon the full transfer function associated with the postintervention environment, namely $T(B)/\nabla^d\Psi_p(B)$. The choice of the mean shift function, $\delta(t)$ in modeling interrupted time series data is critical in estimating the intrinsic program utility δ. As seen from Equation (2-6) when $\Psi_p(B)\nabla^d = T(B)$, the

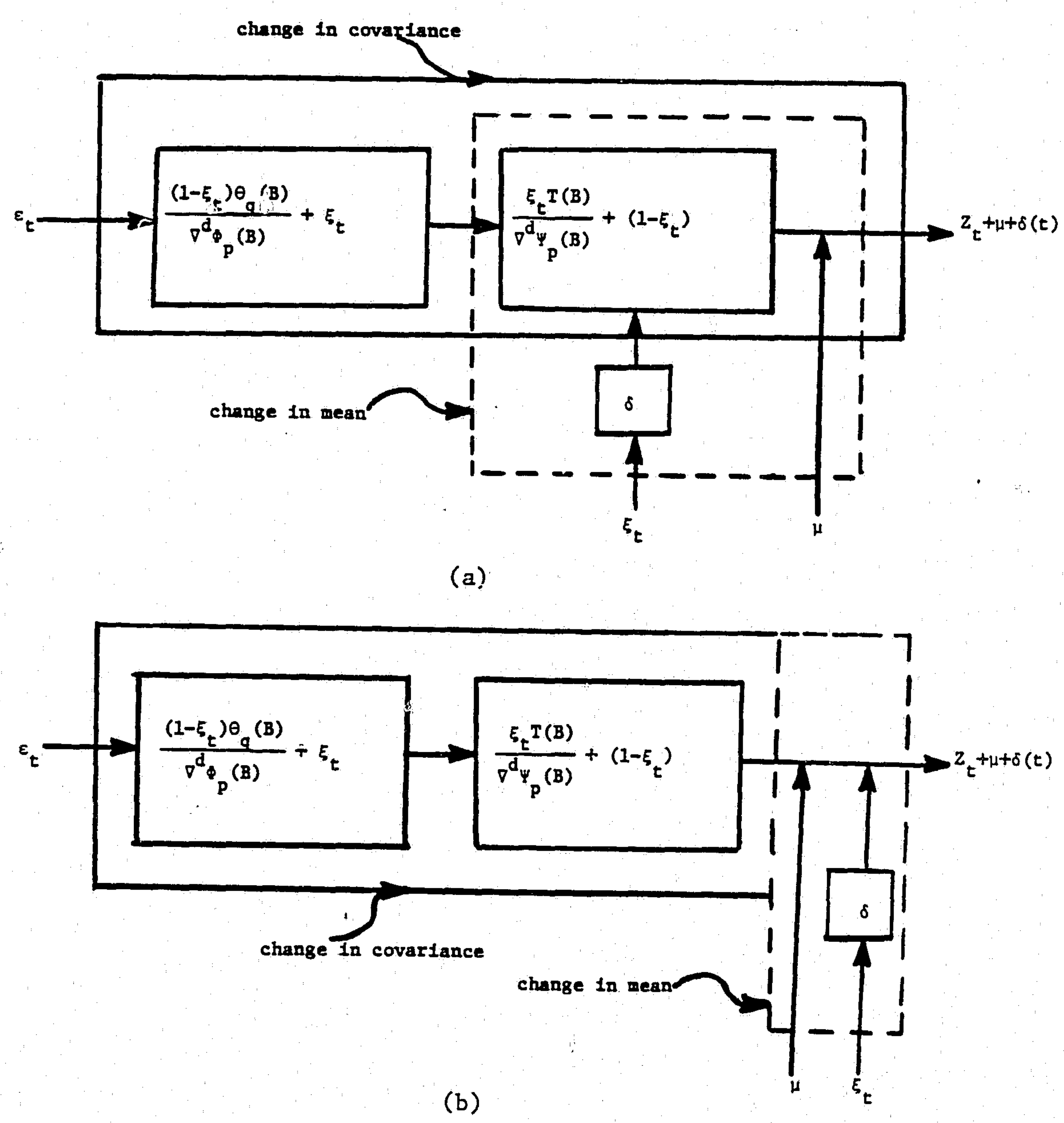


Figure 2-1. Transfer Function Representation of Multiconsequences Intervention Model

(a) Realized change in mean level affected by environmental process

(b) Realized change in mean level not affected by environmental process

mean shift function contributes to the postintervention process level in magnitude equal to the program utility. This is the instantaneous model associated with direct stimulus-subject response experiments. Otherwise, the intrinsic program utility is filtered -- the transient case. Throughout this paper the mean shift function is employed in each of the different considerations presented with regard to the modeling of interrupted time series data with multiconsequence intervention models.

2.2 Maximum Likelihood Estimation Considerations

The procedures for developing maximum likelihood estimators (M.L.E.) described in this section are for various combinations of apriori information about the parameters for different modeling situations. We will first assume that the mean shift function, $\delta(t)$ is known and differentiable and the model parameters $(\underset{\sim}{\phi},\underset{\sim}{\theta},\underset{\sim}{\psi},\underset{\sim}{\gamma})$, are known for the ARIMA(p,d,q)MCI model class. Next the situation in which the mean shift function is unknown is addressed and a sequential procedure for identifying the form of the mean shift function is described. We then relax the assumption that the model parameters are all known.

In the following we will denote

$$\underset{\sim}{Z} = [Z_1, Z_2, \ldots Z_{n_1}, Z_{n_1+1}, \ldots Z_{n_1+n_2}]^t, \quad N=n_1+n_2$$

$$\underset{\sim}{\mu} = E(\underset{\sim}{Z}) = [\mu(1), \mu(2), \ldots \mu(n_1+n_2)]^t \quad \text{and}$$

$$\underset{\sim}{\beta}^t = [\underset{\sim}{\phi}^t, \underset{\sim}{\theta}^t, \underset{\sim}{\psi}^t, \underset{\sim}{\gamma}^t]$$

denotes transpose. Can be confusing since t is used as time parameter.

where

$$\mu(t) = \begin{cases} \mu & t \le n_1 \\ \mu + \delta(t) & n_1 + 1 \le t \le n_1 + n_2 \end{cases}$$

with n_1 and n_2 being the number of pre and post intervention observations, respectively.

2.2.1 Case I: Mean Shift Function Form Known (Model Parameters Known)

1. d=0, ARIMA(p,0,q)MCI model

Suppose that the mean shift function is known to be ~~of~~ some arbitary function of m parameters, $\delta(t) = f(t-n_1, \underset{\sim}{\delta})$ with $\underset{\sim}{\delta}^t = [\delta_1, \delta_2, \ldots \delta_m]^t$ being unknown mean-shift measurement parameters. The mean shift function $\delta(t)$ is first order differentiable when $(t-n_1) \ge 1$. The joint distribution of $\underset{\sim}{Z}$ is given by,

$$f_{\underset{\sim}{z}}(\underset{\sim}{Z}, \underset{\sim}{\mu}_z, \underset{\sim}{\beta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-N/2} |M_N^{(p,0,q)}|^{1/2} \quad (2\text{-}7)$$

$$\operatorname{Exp}\{-(\underset{\sim}{Z} - \underset{\sim}{\mu}_z)^t M_N^{(p,0,q)} (\underset{\sim}{Z} - \underset{\sim}{\mu}_z)/2\sigma_a^2\}$$

where $M_N^{(p,0,q)}$ is the inverse of the covariance matrix of $\underset{\sim}{Z}$. When $\underset{\sim}{\beta}$ is known, $M_N^{(p,0,q)}$ is determined. The maximization of the log likelihood function is equivalent to maximizing the quadratic function,

$$Q(\underset{\sim}{\mu}_z, \underset{\sim}{\beta}, \sigma_a^2) = -(\underset{\sim}{Z} - \underset{\sim}{\mu}_z)^t M_N^{(p,0,q)} (\underset{\sim}{Z} - \underset{\sim}{\mu}_z)/\sigma_a^2 \quad . \quad (2\text{-}8)$$

Since μ, $\delta_1, \ldots \delta_m$ are independent mathematical variables and the mean shift function is known and differentiable, the partial derivative of Q with respect to $\mu, \delta_i, i=1,\ldots m$, set equal to zero,

$$\begin{cases} \left.\dfrac{\partial Q}{\partial \mu}\right|_{\mu=\hat{\mu}, \delta_i=\hat{\delta}_i, i=1,2,\ldots m} = 0 \\ \left.\dfrac{\partial Q}{\partial \delta_k}\right|_{\mu=\hat{\mu}, \delta_i=\hat{\delta}_i, i=1,2,\ldots m} = 0 \qquad k=1,2,\ldots m \end{cases} \quad (2\text{-}9)$$

yields the M.L.E. of $\mu, \delta_i, i=1,2,\ldots,m$.

The mean shift function of the form,

$$\delta(t) = \delta_1 \sum_{i=1}^{t-n_1} \delta_2^{i-1} \quad (2\text{-}10)$$

in which $|\delta_2| < 1$ is differentiable with respect to δ_1, the scale factor, and δ_2, the shape factor. When m=2, the M.L.E. for μ, δ_1 and δ_2 are obtained by solving the following normal equations;

$$\begin{cases} (\underset{\sim}{Z} - \hat{\mu}\underset{\sim}{1} - \hat{\underset{\sim}{D}})^t M_N^{(p,0,q)} \underset{\sim}{1} = 0 \\ \hat{\underset{\sim}{D}}_0^t M_N^{(p,0,q)} (\underset{\sim}{Z} - \hat{\mu}\underset{\sim}{1} - \hat{\underset{\sim}{D}}) = 0 \\ \hat{\underset{\sim}{D}}_1^t M_N^{(p,0,q)} (\underset{\sim}{Z} - \hat{\mu}\underset{\sim}{1} - \hat{\underset{\sim}{D}}) = 0 \end{cases} \quad (2\text{-}11)$$

where $\underset{\sim}{1}$ is $(n_1+n_2) \times 1$ vector with every elements 1.

$\hat{\underset{\sim}{D}} = \hat{\delta}_1(0,0,\ldots 1,1+\hat{\delta}_2,1+\hat{\delta}_2+(\hat{\delta}_2)^2,\ldots \sum_{i=1}^{n_2}(\hat{\delta}\ )^{i-1})^t$, with the first n_1 elements zero,

$\hat{\underset{\sim}{D}}_0 = (0,0,\ldots 1,1+\hat{\delta}_2,1+\hat{\delta}_2+(\hat{\delta}_2)^2,\ldots \sum_{i=1}^{n_2}(\hat{\delta}\ )^{i-1})^t$, with the first n_1 elements zero, and

$\hat{\underset{\sim}{D}}_1 = \hat{\delta}_1(0,0,\ldots 1,1+2\hat{\delta}_2,\ldots \sum_{i=1}^{n_2}(i-1)(\hat{\delta}\ )^{i-2})^t$, with the first (n_1+1) elements zero.

The solution of these simultaneous equations may not be unique and the function value of Q is computed in order to determine the global optimum solution that maximize the quadratic form Q. Similarly, we may have m=1, i.e.,

$$\delta(t) = f(t-n_1,\delta\ )$$

which results in the normal equations;

$$\begin{cases} \underset{\sim}{1}^t\ M_N^{(p,0,q)}\ \underset{\sim}{Z} - \hat{\mu}\underset{\sim}{1}^t\ M_N^{(p,0,q)}\ \underset{\sim}{1} - \underset{\sim}{1}^t\ M_N^{(p,0,q)}\ \hat{\underset{\sim}{D}} = 0 \\ \hat{\underset{\sim}{\mu}}_\delta^t\ M_N^{(p,0,q)}\ \underset{\sim}{Z} - \hat{\mu}\hat{\underset{\sim}{\mu}}^t\ M_N^{(p,0,q)}\ \underset{\sim}{1} - \hat{\underset{\sim}{\mu}}^t\ M_N^{(p,0,q)}\ \hat{\underset{\sim}{D}} = 0 \end{cases} \qquad (2\text{-}12)$$

where

$\hat{\underset{\sim}{D}} = (0,0,\ldots f(1,\hat{\delta}\ ),f(2,\hat{\delta}\ ),\ldots f(n_2,\hat{\delta}\ ))^t$, with the first n_1 elements zero and

$\hat{\underset{\sim}{\mu}}_\delta = \frac{\partial}{\partial\delta}\hat{\underset{\sim}{D}}$.

A useful mean shift function takes the form

$$\delta(t) = \begin{cases} 0 & t \le n_1 \\ \delta K(t) & n_1+1 \le t \end{cases} \qquad (2\text{-}13)$$

where K(t) is a known numerical value. In this situation, the normal equations result in a unique solution,

$$\hat{\mu} = \frac{[\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{K}][\underset{\sim}{1}^t M_N^{(p,0,q)}\underset{\sim}{Z}] - [\underset{\sim}{1}^t M_N^{(p,0,q)}\underset{\sim}{K}][\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{Z}]}{[\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{K}][\underset{\sim}{1}^t M_N^{(p,0,q)}\underset{\sim}{1}] - [\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{1}]^2}$$

$$\hat{\delta} = \frac{[\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{Z}][\underset{\sim}{1}^t M_N^{(p,0,q)}\underset{\sim}{1}] - [\underset{\sim}{Z}^t M_N^{(p,0,q)}\underset{\sim}{1}][\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{1}]}{[\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{K}][\underset{\sim}{1}^t M_N^{(p,0,q)}\underset{\sim}{1}] - [\underset{\sim}{K}^t M_N^{(p,0,q)}\underset{\sim}{1}]^2} \qquad (2\text{-}14)$$

where

$\underset{\sim}{K} = [0,0,\ldots K(n_1+1),K(n_1+2),\ldots K(n_1+n_2)]^t$, with the first n_1 elements zero.

In the analysis of intervention program two specific forms of $K(t)$ have wide applicability. With $d=0$ and $T(B) = \Psi_p(B)$, Equation (2-6) reduces to $\delta(t) = \xi_t\delta$. However, from Equation (2-13) we see $\delta(t) = \delta K(t)$ for $n \geq n_1$, i.e. ~~e.g.~~, depending upon the value of $K(t)$, the realized modification in the process, $\delta(t)$ is not necessarily identical to the intrinsic value or worth of intervention program's activities. When,

$$K(t) = \begin{cases} 0 & t \leq n_1 \\ 1 & \text{elsewhere} \end{cases} \tag{2-15}$$

or

$$K(t) = \begin{cases} 1 & n_1+1 \leq t \leq n_1+T_p \\ 0 & \text{elsewhere} \end{cases}, \tag{2-16}$$

the intrinsic value of the intervention program activities are directly realized and thus are unaffected by the environmental process. Equation (2-15) representing a step function for the intervention programs active life (e.g., $\xi_t=0$, $t \leq n_1$ and $\xi_t=1$ for $t > n_1$; $K(t) = S_t^{n_1}$) while Equation (2-16) represents a pulse function for the intervention program activities that initiates at time n_1+1 and terminates at n_1+T_p where T_p is the number of active periods for the program (e.g., $K(t) = P_t^{n_1}(T_p)$).

The remaining case results from situations in which the environmental process $T(B)$ affects the realized value of modification in level, when $T(B) = \Gamma_q(B)$,

$$K(t) = \begin{cases} 0 & t \leq n_1 \\ \Psi_p^{-1}(B)\Gamma_q(B)\,\xi_t & t \leq n_1+1 \end{cases}. \tag{2-17}$$

The specific form of $K(t)$ is obtained using the recursive equation $\Psi_p(B)K(t) = \Gamma_q(B)\xi_t$ with initial condition $K(t)=0$ for $t \leq n_1$. For example, when $p=q=1$, (i.e., an ARIMA(1,0,1)MCI model) with $\xi_t = P^{n_1}(T_p)$, we have,

$$K(t) - \Psi_1 K(t-1) = 1 - \gamma_1$$

with

$$K(t \leq n_1) = 0,\ K(n_1+1) = 1.$$

This results in

$$K(t) = \begin{cases} 0 & t \leq n_1 \\ \dfrac{1-\gamma_1}{1-\Psi_1} + \left(\dfrac{\gamma_1-\Psi_1}{1-\Psi_1}\right)\Psi_1^{\,t-n_1-1} & n_1+1 \leq t \leq n_1+T_p \\ \Psi_1^{\,t-T_p-n_1-1}(\Psi_1 K(n_1+T_p) - \gamma_1) & t \geq n_1+T_p+1 \end{cases}. \tag{2-18}$$

In Equation (2-18) by setting $T_p = n_2$, and deleting the last equation, we get the $K(t)$ expression for the step function situation, i.e., $\xi_t = S_t^{n_1}$, since the $\xi_t = P_t^{n_1}(T_p = n_2)$.

2. $d \geq 1$

For nonstationary processes, we initially assume that the mean shift function is known, but let $d \geq 1$. In this situation, we difference the original observation vector, $\underset{\sim}{Z}$, to obtain $\underset{\sim}{Z}^* = \nabla^d \underset{\sim}{Z}$, which is stationary. For start up consideration we use, $\underset{\sim}{Z}^* = [0,0,\ldots Z^*_{d+1},\ldots Z^*_{n_1+n_1}]^t$, with the first d elements zero. Since there are now (n_1-d) preintervention data, Equations (2-3) and (2-4) become,

$$\Phi_p(B)\ Z^*_t = \Theta(B)\ a_t \qquad t=d+1,d+2,\ldots n_2$$

$$\Psi_p(B)\ (Z^*_t - \delta^*(t)) = \Gamma_q(B)\ a_t \qquad t=n_1+1,\ldots n_1+n_2$$

where $\delta^*(t) = \nabla^d \delta(t)$.

This form is different from the stationary case where d=0 in that the mean level μ is eliminated and the interpretation of the mean shift function changes. We still, however, have $E(Z_t) = \delta(t)$, $E(Z^*_t) = \delta^*(t)$ while $\delta(t)$ can be obtained once $\delta^*(t)$ is obtained by applying the recessive relation $\delta^*(t) = \nabla^d \delta(t)$,

$$\delta^*(t) = \sum_{i=0}^{d} (-1)^2 \binom{d}{i} \delta(t-i) \quad . \tag{2-21}$$

For example, when d=1 and $\delta^*(t) = \delta, P^{n_1}_t(T_p=1)$ results in $\delta(t) = \delta\ S^{n_1}_t$ which is used in conjunction with $\underset{\sim}{Z}^*$ for modeling.

Since the parameter μ drops out of the model when $d \geq 1$, there is one less parameter to estimate and this makes the simultaneous Equations (2-9) and (2-11) simpler by reducing each by the variable, $\hat{\mu}$, and deleting one normal equation, i.e., $\partial Q/\partial\mu|_{\mu=\hat{\mu}} = 0$. Consider the situation ~~that~~ where $\delta^*(t)$ is one parameter function and can be expressed as $\delta^*(t) = \delta^* \cdot K^*(t)$ with $K^*(t)$ a known numerical value, (similar to that of Equation (2-13). The resulting estimator for $\hat{\delta}^*$ is

$$\hat{\delta}^* = \frac{\underset{\sim}{Z}^{*t} M_N^{*(p,0,q)} \underset{\sim}{K}^*}{\underset{\sim}{K}^{*t} M_N^{*(p,0,q)} \underset{\sim}{K}^*} \tag{2-22}$$

where $M_N^{*(p,0,q)}$ is the covariance matrix of $\underset{\sim}{Z}^*$, and $\underset{\sim}{K}^* = [K^*(d+1), K^*(d+2),\ldots K^*(n_1+n_2)]^t$ with $K^*(t) = 0$ for $t \leq n_1$.

2.2.2 Case II: Mean Shift Function and Influence of Environmental Process Unknown (Model Parameters Known)

In the previous section the form of the mean shift function, $\delta(t)$, was assumed and the time varying proportionality coefficient $K(t)$ was determined to couple a program's intrinsic value δ to the estimated mean shift function $\hat{\delta}(t)$. This proportionality coefficient was developed for the situations in which the program environment did not influence or influenced the realized effects of the activities. It is not unusual, however, to be unable to specify whether the affect of the program is influenced by the environment process. In addition, the form of the mean shift function suggested, which couples $\delta(t)$ to δ (e.g., Equation (2-13), may not be known. In these situations the sequential estimation of

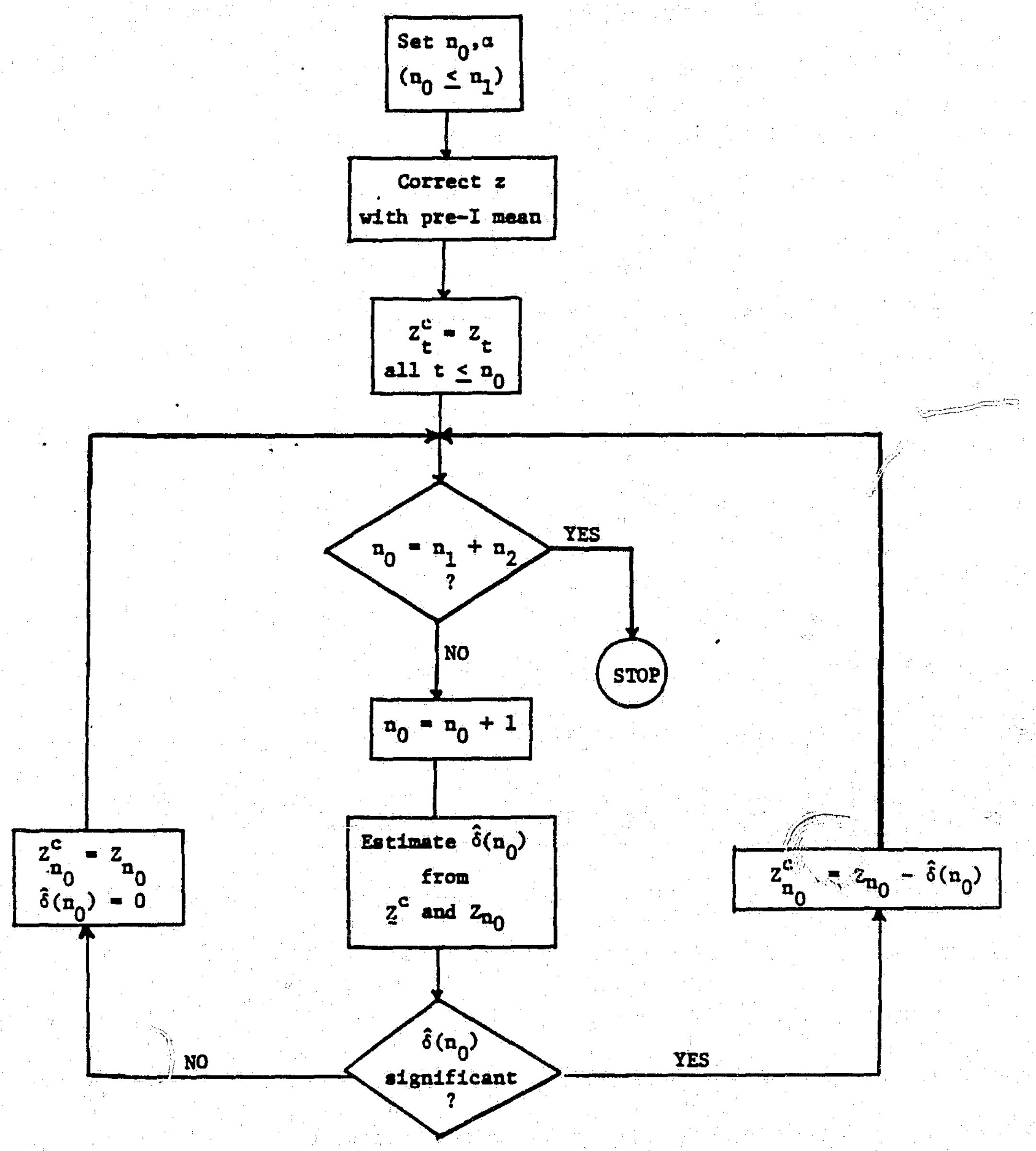


Figure 2.2. Identification Procedure for Dynamic Component

$\hat{\delta}(t)$, $t=n_1+1,\ldots,n_1+n_2$ for $n_2=1$ is used to identify both whether the program's affect is influenced by the environment and to determine an appropriate of the mean shift function.

In estimating the mean shift function, $\delta(t)$, the value of $K(t)$ must be specified which determine whether the environmental process influences or does not influence the mean shift function. However, from Equation (2-18) the value of $K(t)$ in the situation where the environmental process has an influence, the estimation of $\delta(t)$ is unaffected when $n_2=1$. Figure 2.2 delineates the sequential procedure for estimating the mean shift function.

3. d=0

Initially, the M.L.E. of μ using the n_1 preintervention observations are obtaining using,

$$\hat{\mu} = \frac{\underset{\sim}{1}^t M_{n_1}^{(p,0,q)} \underset{\sim}{z}}{\underset{\sim}{1}^t M_{n_1}^{(p,0,q)} \underset{\sim}{1}} \tag{2-23}$$

where $M_{n_1}^{(p,0,q)}$ is the covariance matrix of the first n_1 observations. If μ is not equal to zero the n_1+n_2 observations are corrected by subtracting $\hat{\mu}$. Then the M.L.E. of $\delta(t)$ is obtained using Equation (2-22), with $\underset{\sim}{K}^*$ specified as $\underset{\sim}{e}_t$, i.e., the (n_1+n_2) unit vector with $n_2=1$, the vector with 1 at the t^{th} position,

$$\hat{\delta}(t) = \frac{\underset{\sim}{z}_t^t \underset{\sim}{M}_t}{M_{tt}} \tag{2-24}$$

where

$\underset{\sim}{Z}_t$ is the vector $(Z_1, Z_2, \ldots Z_{t-1}, Z_t)$,

M_{tt} is the (t,t) element of $M_t^{(p,0,q)}$ and

$\underset{\sim}{M}_t$ is the $t^{\underline{th}}$ column of $M_t^{(p,0,q)}$.

If $\hat{\delta}(n_1+1)$ is significant, the $(n_1+1)^{\underline{st}}$ observation is corrected by subtracting $\hat{\delta}(n_1+1)$ and entered into the set of n_1 observation. The process is repeated with n_1+1 preintervention observations until all n_1+n_2 observations are exhausted.

4. $d \geq 1$

When the ARIMA(p,d,q)MCI process is nonstationary, e.g., $d \geq 1$, μ is set equal to zero with (n_1-d) preintervention data points. The remaining identification procedure is identical to the stationary situation. Note that the mean shift function estimates obtained are for $\underset{\sim}{Z}^*$ not $\underset{\sim}{Z}$, i.e. ~~e.g.~~, they are $\delta^*(t)$ not $\delta(t)$. In this case, we obtain $\hat{\delta}(t)$ from $\hat{\delta}^*(t)$ by applying Equation (2-21) recursively.

From the pattern exhibited by $\delta(t)$ vs. time for $t=n_1+1, n_1+2, \ldots, n_1+n_2-1$ and the form of $\underset{\sim}{\xi}_t$ which is known to be a pulse function of length T_p or a step function $(T_p \to \infty)$ the appropriate multiconsequence intervention representation is determined. If $\hat{\delta}(t)$ is constant for each of the T_p periods, the environmental process does not exert any influence on the realized mean level modification, ~~e.g.~~, $K(t)=1$) for $t=n_1+1, n_1+2, \ldots, n_1+T_p$ and zero elsewhere. Thus the structural model exhibited in Figure 2-1(b) is selected. In identifying this configuration the appropriated form of the mean shift function follows as $\delta(t) = K(t)\delta$ with $K(t)=1$. However, if the environmental process influences the realized

values of the mean modification, an appropriate form of K(t) must be established. The identification of the appropriate represent of K(t) may proceed in two ways: 1) pattern matching to known theoretical patterns of K(t) arising from specified forms of $\delta(t)$, or 2) use of regression to fit an empirical form of $\delta(t)$ to data of $\hat{\delta}(t)$ versus t and then computing K(t).

2.2.3 Case III: Estimating μ, $\delta(t)$ Given β_u Unknown (Model Parameters Not All Known)

So far we have assumed that all the model parameters are known. Often, however, some or all of these model parameters $\underset{\sim}{\beta}$ are unknown. We will denote $\underset{\sim}{\beta}_u$ as the unknown parameter of $\underset{\sim}{\beta}$ and $\underset{\sim}{\beta}_u = (\underset{\sim}{\phi}_u, \underset{\sim}{\theta}_u, \underset{\sim}{\psi}_u, \underset{\sim}{\gamma}_u)$ with $\underset{\sim}{\phi}_u, \underset{\sim}{\theta}_u, \underset{\sim}{\psi}_u, \underset{\sim}{\gamma}_u$ as the unknown parameters of $\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\psi}, \underset{\sim}{\gamma}$ respectively. From the earlier estimation results cited, we know that the M.L.E. of μ and the mean shift parameters, δ_i, are not independent of the model parameters (i.e., giving the model parameters, the M.L.E. are determined without searching through the $(\mu, \underset{\sim}{\delta})$ subspace). That is,

$$\underset{\underset{\sim}{\beta},\mu,\underset{\sim}{\delta},\sigma_a^2}{\text{Max}} \{L(\underset{\sim}{\beta},\mu,\underset{\sim}{\delta},\sigma_a^2)\}$$

$$= \underset{\mu,\underset{\sim}{\delta},\sigma_a^2}{\text{Max}} \{L(\beta,\mu,\underset{\sim}{\delta},\sigma_a^2)\} \qquad \text{when } \underset{\sim}{\beta} \text{ all known}$$

$$= \underset{\sigma_a^2}{\text{Max}} [\underset{\mu,\underset{\sim}{\delta}}{\text{Max}}\, L(\underset{\sim}{\beta},\mu,\underset{\sim}{\delta},\sigma_a^2)]$$

In the procedures for obtaining $\hat{\mu}$, $\hat{\underset{\sim}{\delta}}$ and $\hat{\sigma}_a$ when $\underset{\sim}{\beta}$ is known the least squares estimators (L.S.E.) are equivalent to M.L.E. due to the normal distribution assumption of the residuals. However, when not all the model parameters are known, the L.S.E. are not equivalent to the M.L.E. From Equation (3-7), we see that the unknown model parameters $\underset{\sim}{\beta}_u$ would appear not only in the quadratic form but also in the term $|M_N^{(p,0,q)}|^{1/2}$. Therefore, to maximize the likelihood function when not all model parameters are known is equivalent to maximize ing the whole function. But MAX $Q(\mu,\underset{\sim}{\delta},\sigma_a^2) = -1$, so it is equivalent to Maximize ing $(\hat{\sigma}_a^2)^{-N/2}\cdot|M_N^{(p,0,q)}|^{1/2}$. Consider the maximization of the likelihood function;

$$\max_{\underset{\sim}{\beta},\mu,\underset{\sim}{\delta},\sigma_a^2} \{L(\underset{\sim}{\beta},\mu,\underset{\sim}{\delta},\sigma_a^2)\}$$

$$= \max_{\underset{\sim}{\beta}_u} \{\max_{\sigma_a^2}[\max_{\mu,\underset{\sim}{\delta}} L\ (\underset{\sim}{\beta},\mu,\underset{\sim}{\delta},\sigma_a^2)]\}$$

$$= \max_{\underset{\sim}{\beta}_u} \ \{|M_N^{(p,0,q)}|^{1/2}\cdot \max(\sigma_a^2)^{-N/2}\}$$

$$= \max_{\underset{\sim}{\beta}_u} \ \{|M_N^{(p,0,q)}|^{1/2}(\hat{\sigma}_a^2)^{-N/2}\}\ .$$

This is equivalent to minimizing,

$$N\ln(\sigma_a^2) - \ln|M_N^{(p,0,q)}|$$

Thus, to obtain the M.L.E. for $(\underset{\sim}{\beta}_u,\mu,\underset{\sim}{\delta},\sigma_a^2)$, we search through the $\underset{\sim}{\beta}_u$ subspace, apply previous results and minimize the value of $N\ln(\sigma_a^2) - \ln|M_N^{(p,0,q)}|$ where $\hat{\sigma}_a^2$ is the L.S.E. (or M.L.E.) of σ_a^2 for given $\underset{\sim}{\beta}$.

2.3 Covariance Matrix for Low-Order Models

As just seen, the determination of the form of the covariance matrix is central to the development of the M.L.E. In developing this matrix we will focus on the four submatrices that emerge from partitioning realizations of the process into those associated with: the pre-intervention period and the postintervention period (e.g., $t \le n_1$ and $n_1 < t \le n_1+n_2$ respectively). These submatrices will be denoted c_{ij} as follows;

$$\Sigma^{(p,0,q)} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n_1 \\ n_1+1 \\ \vdots \\ n_1+n_2 \end{array} \begin{array}{c} 1,2,\ldots,n_1,n_1+1,\ldots,n_1+n_2 \\ \left[\begin{array}{c|c} c_{11} & c_{12} \\ \hline c_{21} & c_{22} \end{array}\right] \end{array}$$

In the following the c_{ij}'s of the covariance matrix for the ARIMA (1,0,1)MCI model will be constructed. It should be noted that its evaluation for $\phi = \psi = 0$, or $\theta = \gamma = 0$ results in the covariance for the ARIMA(0,0,1)MCI and ARIMA(1,0,0)MCI models respectively. For nonstationary processes, the corresponding covariance matrices are obtained using the following results with μ=0 and Z_t^* where $Z_t^* = \nabla^d Z_t$.

For the ARIMA(1,0,1)MCI model;

$(Z_t - \mu) = \phi_1(Z_{t-1} - \mu) + a_t - \theta_1 a_{t-1}, \qquad 1 \le t \le n_1$

$(Z_{n_1+1} - \mu - \delta(t)) = \psi_1(Z_{n_1} - \mu) + a_{n_1+1} - \gamma_1 a_{n_1} \qquad t=n_1+1$

$(Z_t - \mu - \delta(t)) = \psi_1(Z_{t-1} - \mu - \delta(t)) + a_t - \gamma_1 a_{t-1}, \qquad n_1+2 \le t \le n_1+n_2$

(i) Submatrix c_{11} corresponds to the preintervention time series model. The development of the covariance elements are found in Box and Jenkins (1970). The pertinent results are;

for k=0;

$$\mathrm{Var}(Z_t) = \frac{1+\theta_1^2-2\theta_1\phi_1}{1-\phi_1^2}\ \sigma_a^2 \tag{2-25}$$

for $k \ge 1$

$$\mathrm{Cov}(Z_t, Z_{t+k}) = \phi_1^{k-1} \frac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1-\phi_1^2}\ \sigma_a^2 \tag{2-26}$$

(ii) Submatrix c_{12} and c_{21}

$1 \le t \le n_1, \qquad n_1+1 \le t+k \le n_1+n_2 \qquad k \ge 1$

For $t=n_1$, $k=1$

$$\mathrm{Cov}(Z_{n_1}, Z_{n_1+1}) = \psi_1 \mathrm{Var}(Z_{n_1}) - \gamma_1\ \sigma_a^2$$
$$= [\psi_1 \frac{1+\theta_1^2-2\theta_1\phi_1}{1-\theta_1^2} - \gamma_1]\ \sigma_a^2$$

For $1 \le t \le n_1-1$, $t+k=n_1+1$

$$\mathrm{Cov}(Z_t, Z_{n_1+1}) = \psi_1 \mathrm{Cov}(Z_t, Z_{n_1})$$
$$= \psi_1\phi_1^{(n_1-t-1)} \frac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1-\phi_1^2}\ \sigma_a^2$$

Thus,

$$\mathrm{Cov}(Z_t, Z_{n_1+1}) = \begin{cases} \psi_1\phi_1^{(n_1-t-1)} \dfrac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1-\phi_1^2}\ \sigma_a^2 & 1 \le t \le n_1-1 \\ (\psi_1 \dfrac{1+\theta_1-2\theta_1\phi_1}{1-\phi_1^2} - \gamma_1)\ \sigma_a^2 & t=n_1 \end{cases} \tag{2-27}$$

Similarly, the covariance for $k \ge 2$ can be shown to be,

$$\mathrm{Cov}(Z_t,Z_{t+k}) = \begin{cases} \psi_1^{(t+k-n_1)}\phi_1^{n_1-t-1}\dfrac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1-\phi_1^2}\sigma_a^2 & 1 \le t \le n_1-1 \\ \psi_1^{k-1}\left(\psi_1\dfrac{1+\theta_1^2-2\theta_1\phi_1}{1-\phi_1^2}-\gamma_1\right)\sigma_a^2 & t=n_1 \end{cases} \tag{2-28}$$

(iii) Submatrix c_{22}

For $1 \le k \le n_2$, $n_1+1 \le t+k \le n_1+n_2$ and $t \ge n_1+1$,

$$\mathrm{Var}(Z_{n_1+k}) = \psi_1^2\,\mathrm{Var}(Z_{n_1+k-1}) + (1+\gamma_1^2-2\psi_1\gamma_1)\,\sigma_a^2$$

$$= \psi_1^{2k}\frac{1+\theta_1^2-\theta_1\phi_1}{1-\phi_1^2}\sigma_a^2 + \frac{1-\psi_1^{2k}}{1-\psi_1^2}(1+\gamma_1^2-2\psi_1\gamma_1)\,\sigma_a^2 \tag{2-29}$$

Since

$$E(Z_t,a_{t+k-1}) = \begin{cases} \sigma_a^2 & k=1 \\ 0 & k\ge 2 \end{cases}$$

then

$$\mathrm{Cov}(Z_t,Z_{t+k}) = \psi_1\cdot\mathrm{Cov}(Z_t,Z_{t+k-1}) - \gamma_1\,E(Z_t a_{t+k-1})$$ (2.29a)

$$= \psi_1^{k-1}(\psi_1\,\mathrm{Var}(Z_t)-\gamma_1)\,\sigma_a^2$$

$$\underset{\sim}{c}_{11} = \begin{array}{c|cccccc} & 1 & 2 & 3 & & n_1-1 & n_1 \\ \hline 1 & g_0 & g_1 & \phi_1 g_1 & \cdots & \phi_1^{n_1-3}g_1 & \phi_1^{n_1-2}g_1 \\ 2 & g_1 & g_0 & g_1 & \cdots & \phi_1^{n_1-4}g_1 & \phi_1^{n_1-3}g_1 \\ 3 & \phi_1 g_1 & g_1 & g_0 & \cdots & \phi_1^{n_1-5}g_1 & \phi_1^{n_1-4}g_1 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ n_1-1 & \phi_1^{n_1-3}g_1 & \phi_1^{n_1-4}g_1 & \phi_1^{n_1-5}g_1 & \cdots & g_0 & g_1 \\ n_1 & \phi_1^{n_1-2}g_1 & \phi_1^{n_1-3}g_1 & \phi_1^{n_1-4}g_1 & \cdots & g_1 & g_0 \end{array}$$

Where $g_0 = \dfrac{1+\theta_1^2-2\theta_1\phi_1}{1-\phi_1^2}$, $g_1 = \dfrac{(\phi_1-\theta_1)(1-\theta_1\phi_1)}{1-\phi_1^2}$

$$\underset{\sim}{c}_{12} = \begin{array}{c|cccc} & n_1+1 & n_1+2 & \cdots & n_1+n_2 \\ \hline 1 & \psi_1\phi_1^{n_1-2}g_1 & \psi_1^2\phi_1^{n_1-2}g_1 & \cdots & \psi_1^{n_2}\phi_1^{n_1-2}g_1 \\ 2 & \psi_1\phi_1^{n_1-3}g_1 & \psi_1^2\phi_1^{n_1-3}g_1 & \cdots & \psi_1^{n_2}\phi_1^{n_1-3}g_1 \\ 3 & \psi_1\phi_1^{n_1-4}g_1 & \psi_1^2\phi_1^{n_1-4}g_1 & \cdots & \psi_1^{n_2}\phi_1^{n_1-4}g_1 \\ \vdots & \vdots & \vdots & & \vdots \\ n_1-1 & \psi_1 g_1 & \psi_1^2 g_1 & & \psi_1^{n_2}g_1 \\ n_1 & \psi_1 g_0-\gamma_1 & \psi_1(\psi_1 g_0-\gamma_1) & \cdots & \psi_1^{n_2-1}(\psi_1 g_0-\gamma_1) \end{array}$$

Figure 2-3. The Covariance Matrix of the ARIMA(1,0,1)MCI Model

$$
\underset{\sim}{C}_{22} = \begin{array}{c} \\ n_1+1 \\ n_1+2 \\ n_1+3 \\ \vdots \\ n_1+n-1 \\ n_1+n \end{array}
\begin{array}{c}
\begin{array}{ccccc} n_1+1 & n_1+2 & \cdots & n_1+n_2-1 & n_1+n_2 \end{array} \\
\left[\begin{array}{ccccc}
g_{22}^{(1)} & h_{22}^{(1)} & \cdots & \psi_1^{n_2-3}h_{22}^{(1)} & \psi_1^{n_2-2}h_{22}^{(1)} \\
h_{22}^{(1)} & g_{22}^{(1)} & \cdots & \psi_1^{n_2-4}h_{22}^{(1)} & \psi_1^{n_2-4}h_{22}^{(2)} \\
\psi_1 h_{22}^{(1)} & h_{22}^{(2)} & \cdots & \psi_1^{n_2-5}h_{22}^{(3)} & \psi_1^{n_2-4}h_{22}^{(3)} \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
\psi_1^{n-3}h_{22}^{(1)} & \psi_1^{n-4}h_{22}^{(2)} & \cdots & g_{22}^{(n_2-1)} & h_{22}^{(n_2-1)} \\
\psi_1^{n-2}h_{22}^{(1)} & \psi_1^{n-3}h_{22}^{(2)} & \cdots & h_{22}^{(n_2-1)} & g_{22}^{(n2)}
\end{array}\right]
\end{array}
$$

Where $g_{22}^{(k)} = \psi_1^{2k}\, g_0 + \dfrac{1-\psi_1^{2k}}{1-\psi_1^2}(1+\gamma_1^2 - 2\psi_1\gamma_1)$, $k \geq 1$

$h_{22}^{(k)} = \psi_1 g_{22}^{(k)} - \gamma_1$

When evaluated with,

$\phi = \psi = 0,$ $\quad \Sigma^{(1,0,1)} = \Sigma^{(0,0,1)}$

$\theta = \gamma = 0,$ $\quad \Sigma^{(1,0,1)} = \Sigma^{(1,0,0)}$

$\left.\begin{array}{c} \phi = \psi \neq 0 \\ \text{or} \\ \theta = \gamma \neq 0 \end{array}\right\}$, multiconsequence covariance = single consequence covariance.

Figure 2-3. (Cont'd)

(2.29a)

Substitute Equation (2-29) into the above equation, we get

$$
\operatorname{Cov}(Z_t, Z_{t+k}) = \psi_1^{|k-1|}\left[\psi_1\left(\psi_1^{2\ell}\,\frac{1+\theta_1^2-2\theta_1\phi_1}{1-\phi_1^2} + \frac{1-\psi_1^{2\ell}}{1-\psi_1^2}(1+\gamma_1^2-2\psi_1\gamma_1)\right)-\gamma_1\right]\sigma_a^2 \quad (2\text{-}30)
$$

where $\ell = t-n_1$.

The results are summarized in matrix form in Figure 2-3 for ease in evaluation of the covariance for other low order ARIMA(p,0,q)MCI processes.

2.4 Hypothesis Testing and Confidence Intervals

In Section 2.2, the M.L.E. procedures for three different cases of unknown combinations of the model parameters $(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\psi}, \underset{\sim}{\gamma})$ or the mean shift function, $\delta(t)$, were detailed. Case I described the situation in which the form of the mean shift function was known along with the model parameters and δ and μ were to be estimated. When the influence of the environmental process on the realized postintervention mean level is unknown (and therefore the form of the mean shift function) but the preintervention model parameters, $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$, are known, the estimation procedures for the form of mean shift was described in Case II. Case III described the M.L.E. procedures for estimating δ, μ and $\underset{\sim}{\beta}_u$ when the form of the mean shift function is known and $\underset{\sim}{\beta}_u$ is a subset of the model parameters. In this section the corresponding hypothesis tests of significance for δ, μ, $(\underset{\sim}{\phi}-\underset{\sim}{\psi})$ and $(\underset{\sim}{\theta}-\underset{\sim}{\gamma})$ needed for modeling are presented. Also the corresponding confidence intervals for δ and μ and the pre and post intervention model parameters are developed.

Case I:

From the M.L.E. of δ for d=0, our immediate interest is to test whether δ is significant, i.e., we want to test;

$$H_0: \ \delta = 0$$
$$H_1: \ \delta \neq 0$$

Since $\hat{\delta} \sim N(\delta, \sigma^2_{\hat{\delta}})$ with,

$$\sigma^2_{\hat{\delta}} = \frac{\underset{\sim}{1}^t M_N^{(p,o,q)} \underset{\sim}{1}}{(\underset{\sim}{K}^t M_N^{(p,o,q)} \underset{\sim}{K})(\underset{\sim}{1}^t M_N^{(p,o,q)} \underset{\sim}{1}) - (\underset{\sim}{K}^t M_N^{(p,o,q)} \underset{\sim}{1})^2} \ \sigma^2_a \qquad (2\text{-}31)$$

the statistic W where,

$$W_\delta = \frac{|\hat{\delta}|}{\sqrt{\sigma^2_{\hat{\delta}}}},$$

is ~ $N(0,1^2)$. For a selected α level we reject H_0 when $W_\delta > t_{1-\alpha/2,N-2}$. The corresponding $(1-\alpha)$ 100% confidence interval for δ is, $[\delta \pm W_\delta t_{\alpha/2,N-2}]$. When $d \geq 1$, the same procedure applies with the exception that,

$$\sigma^2_{\hat{\delta}} = \frac{\sigma^2_a}{K^{*t} M_N^{*(p,o,q)} K^*} \qquad (2\text{-}32)$$

Case II:

In the identification of the form of the mean shift function the significance of the preintervention mean must be determined prior to estimation of δ_i when $n_2=1$ for $i=n_1+1,\ldots,n_1+n_2-1$. Since $\hat{\mu} \sim N(\mu,\sigma^2_{\hat{\mu}})$ with,

$$\sigma^2_{\hat{\mu}} = \frac{\sigma^2_a}{\underset{\sim}{1}^t M_N^{(p,0,q)} \underset{\sim}{1}}, \qquad (2\text{-}33)$$

to test the hypotheses;

$$H_0: \ \mu = 0$$
$$H_1: \ \mu \neq 0,$$

we compute the statistic W_μ where

$$W_\mu = \frac{|\hat{\mu}|(\underset{\sim}{1}^t M_{n_1}^{(p,0,q)} \underset{\sim}{1})}{\sigma^2_a}.$$

We reject H_0 when $W_\mu > t_{1-\alpha/2,n_1-1}$ or accept H_0 when $W_\mu < t_{1-\alpha/2,n_1-1}$. Similarly, in testing the significance of each $\delta(i)$ in order to determine correction of the i^{th} observed value before estimation of $\delta(i+1)$, we test the hypothesis,

$$H_0: \ \delta(i) = 0$$
$$H_1: \ \delta(i) \neq 0$$

by computing W where,

$$W = \frac{|\hat{\delta}(i)|M_{tt}}{\hat{\sigma}^2}.$$

The null hypothesis is rejected when $W > t_{1-\alpha/2,t-1-N_\delta}$ or accepted when $W < t_{1-\alpha/2,t-1-N_\delta}$ where N_δ is the number of previous rejections of the null hypotheses in this sequential procedure.

Case III:

After estimation of the pre and post intervention model parameters the question whether there has been a significant change in the covariance structure should be tested before different postintervention parameter are retained in the model. To test the significance of the change in covariance, we test the hypothesis,

$$H_0: \underset{\sim}{\beta}_{u1} = \underset{\sim}{\beta}_{u2}$$

$$H_1: \underset{\sim}{\beta}_{u1} \neq \underset{\sim}{\beta}_{u2}$$

where $\underset{\sim}{\beta}_{u1} \varepsilon (\underset{\sim}{\phi},\underset{\sim}{\theta})$ and $\underset{\sim}{\beta}_{u2} \varepsilon (\underset{\sim}{\psi},\underset{\sim}{\gamma})$. The statistic,

$$W_{\underset{\sim}{\beta}} = [N\ell n(\hat{\sigma}_a^2) + \ell n(|M_N^{(p,0,q)}|)]_{\underset{\sim}{\beta}_{u1}=\underset{\sim}{\beta}_{u2}} - [N\ell n(\hat{\sigma}_a^2 + \ell n(|M_N^{(p,0,q)}|)]_{\underset{\sim}{\beta}_{u1}\neq\underset{\sim}{\beta}_{u2}}$$

is computed (see Wilks (1938)). The null hypothesis is rejected when $W_{\underset{\sim}{\beta}} > \chi^2_{\alpha,n-m}$ or accepted when $W_{\underset{\sim}{\beta}} \leq \chi^2_{\alpha,n-m}$ where n is the number of unknown model parameters and m is the dimension of $\underset{\sim}{\beta}_i$, i=1,2.

2.5 Anticipated Causes of Bias in Estimation of the Shift in Process Level

The misspecification of the mean shift function form and/or the use of a single consequence intervention model form when there is change in the covariance structure affects the ability to statistically determine the significance of an intervention program's intrinsic value.

From the estimator for δ, the program's intrinsic utility, Equation (2-14), the elements of $\underset{\sim}{K}$ which are determined from the form of K(t) are seen to weight the observation vector $\underset{\sim}{Z}$. For example, when we have $\underset{\sim}{\xi}_t$ as a step function coupled with a direct stimulus response intervention model each observation after $t > n_1$ is weighted equally since K(t) = 1 for $t > n_1$. However, when $\underset{\sim}{\xi}_t$ is a pulse function regardless of the specification of K(t) each postintervention observation is not equally weighted. Also when K(t) is of the form appropriate ~~of~~ for an indirect stimulus response experiment regardless of the form of $\underset{\sim}{\xi}_t$ all postintervention observations are not in general equally weighted. The consequences of having weights larger or smaller than appropriate is to underestimate or overestimate δ, respectively.

Figure 2.4 portrays various configurations of K(t) that results from ~~for a~~ step function for the (1,0,1)MCI model. Each exhibit in Figure 2.4(a), (b) and (c) represents $\underset{\sim}{\xi}_t$ as step functions with ψ_1 = 0.0, 0.6 and -0.6 for a range of γ values, respectively. The cases of $\psi_1 = \gamma_1$ correspond to the direct stimulus response form of K(t), e.g., K(t) = 1 $t \geq 2$, however, in comparisons between cases a, b and c, the observations are weighted differentially being dependent upon the postintervention autocorrelative structure.

Within Figure 2.4, $K(t) = 1-\gamma$ for $t \geq 2$, and thus for a fixed γ_1 postintervention observations Z_t, $t \geq n_1+2$ are weighted equally in estimation of δ. However, postintervention observations Z_t for $t \geq n_1+2$ are weighted $(1-\gamma)$ times the $Z_{n_1+1}{}^{st}$ value weight in estimation of δ. Thus realizations Z_t, $t \geq n_1+2$ exert greater influence than Z_{n_1+1} on the estimation of δ when $\gamma < 0$ and less influence when $\gamma > 0$. In the latter case as $\gamma \rightarrow 1$ only the $Z_{n+1}{}^{st}$ observation influences estimation of δ. Similar differentiable behavior for the relative contributions of Z_t, $t \geq n_1+1$ in estimation of δ are seen for the step functions in Figures 2.4(a) and (b). When ξ_t is a pulse function, the $K(t)$ weights behave the same as the step function values for $t \leq n_1+T_p$ and eventually decay to zero as t increases for $t > n_1+T_p+1$. Thus as in the step function case, a pulse function also behaves similarly in that the consequence of not assigning an appropriate mean shift function form manifests itself in improper specification of $K(t)$ and consequent overestimation of the magnitude of δ if $K(t)$ chosen is less than the appropriate $K(t)$ or underestimation of δ if the $K(t)$ chosen is larger than the appropriate $K(t)$.

Suppose the mean shift function form $K(t)$ is correctly specified/identified but a single consequence model is used inappropriately. Since $K(t)$ is correct the point estimates of δ are not inflated or deflated. However, the tests of its significance is still impaired due to the difference in magnitude of the $Var(\hat{\delta})$ that occur. For example, when the system should be treated as a (1,0,1)MCI process and is not, the resulting model parameters for the single consequence formulation are:

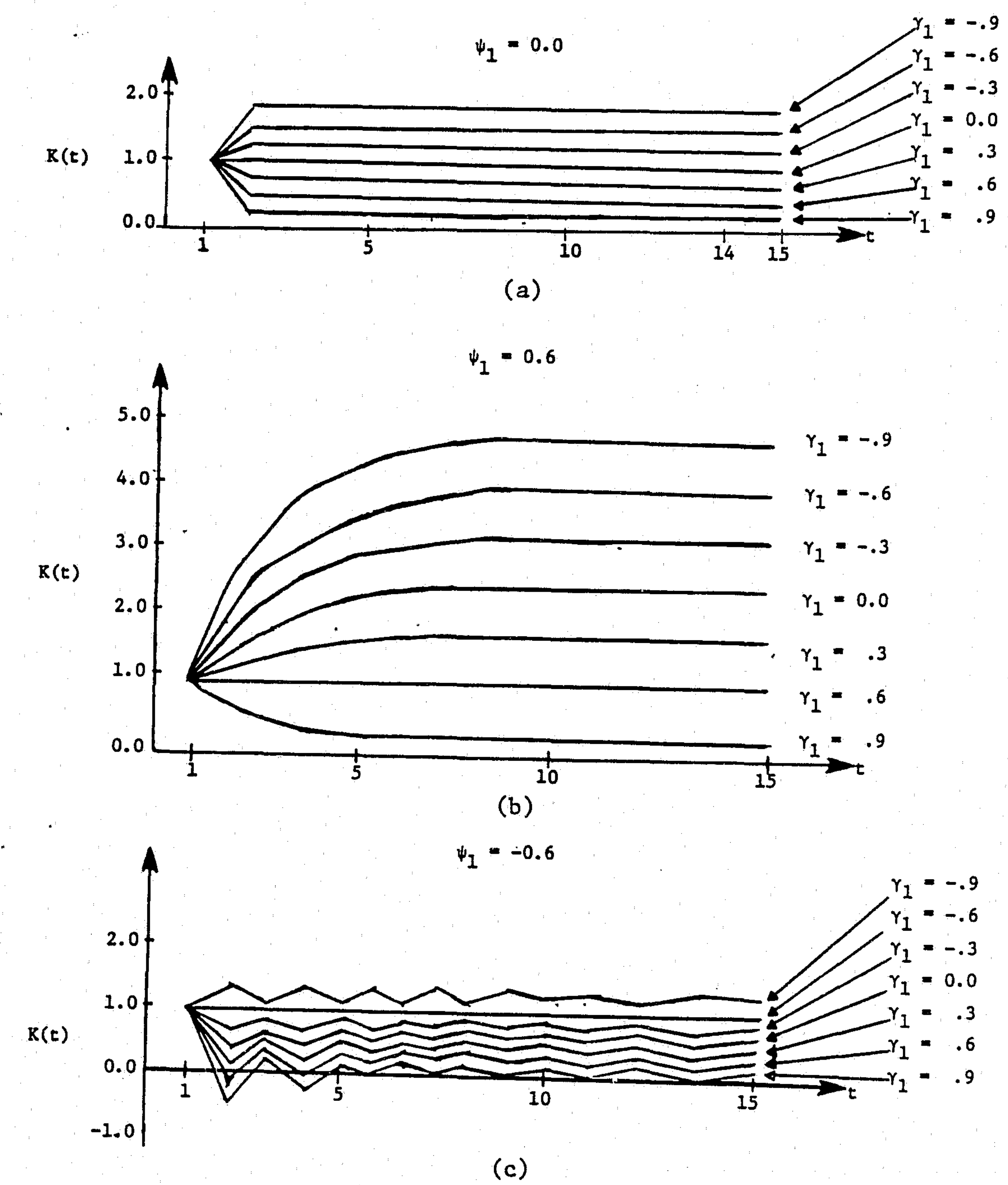


Figure 2.4. Variations in K(t) Weights

$$\hat{\phi}_S = \frac{n_1\phi_1 + n_2\psi_1}{n_1 + n_2}$$

and

$$\hat{\theta}_S = \frac{n_1\theta_1 + n_2\gamma_1}{n_1 + n_2}$$

where $\hat{\phi}_S$ and $\hat{\theta}_S$ are approximate solution to the estimation problem,

$$\min_{\phi_S,\theta_S} \left[\hat{\sigma}^2_A(\phi_S,\theta_S|\phi_1,\theta_1,\psi_1,\gamma_1,\sigma^2_a)\right]$$

with $\hat{\sigma}^2_A$ defined as $SS/(n_1+n_2)$ where

$$SS = \sum_{t=1}^{n_1+n_2} (\hat{A}_t)^2 \quad .$$

where σ^2_A, ϕ_1, θ_1, ψ_1, γ_1 is the true (1,0,1)MCI model parameters and $\hat{A}_t$'s are the residuals computed from the single consequence model, e.g., the model with $\phi_1 = \psi_1 = \phi_S$, $\theta_1 = \gamma_1 = \theta_S$.

Table 2.1 illustrates representative values of $M = \left(\sqrt{\mathrm{Var}(\hat{\delta})\mathrm{MCI}/\mathrm{Var}(\hat{\delta})\mathrm{SCI}}\right)^{-1}$ for preintervention parameter values of

Table 2.1. Values of $\left(\sqrt{\mathrm{Var}(\hat{\delta})\mathrm{MCI}/\mathrm{Var}(\hat{\delta})\mathrm{SCI}}\,\right)^{-1}$ for Determining Consequences of Over/Under Estimation of Significance of $\hat{\delta}$ when K(t) is Correctly Specified

$n_1 = n_2 = 10$ $\phi_1 = 0.60$

$\theta_1 = -0.60$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	1.25	1.01	.80	.67	.67	.84	1.14	1.49	1.80	1.93	2.02
-.80	1.15	.84	.72	.66	.66	.73	.85	1.00	1.15	1.24	1.34
-.60	1.28	.84	.72	.67	.67	.71	.79	.89	1.00	1.04	1.11
-.40	1.37	.87	.74	.69	.69	.71	.78	.87	.96	1.00	1.03
-.20	1.42	.89	.75	.70	.69	.72	.78	.86	.96	1.00	1.00
0.00	1.45	.91	.77	.71	.70	.72	.77	.86	.97	1.03	1.01
.20	1.46	.92	.77	.72	.70	.72	.77	.86	.99	1.08	1.05
.40	1.48	.94	.79	.72	.70	.71	.76	.85	1.00	1.15	1.14
.60	1.55	.98	.82	.74	.70	.70	.74	.84	1.00	1.23	1.30
.80	1.73	1.10	.90	.79	.72	.69	.72	.82	1.01	1.32	1.63
.95	2.04	1.30	1.04	.88	.76	.70	.70	.80	1.04	1.44	2.03

$n_1 = n_2 = 10$ $\phi_1 = 0.60$

$\theta_1 = 0.00$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	.90	.83	.78	.81	.95	1.21	1.56	1.96	2.32	2.58	2.94
-.80	.83	.75	.74	.77	.84	.95	1.10	1.27	1.43	1.56	1.81
-.60	.93	.74	.72	.74	.79	.86	.96	1.07	1.15	1.20	1.36
-.40	1.02	.76	.72	.72	.76	.82	.90	.99	1.06	1.06	1.16
-.20	1.08	.77	.71	.71	.74	.79	.86	.95	1.01	1.00	1.05
0.00	1.12	.77	.71	.70	.72	.76	.83	.92	1.00	.99	.98
.20	1.13	.77	.70	.69	.70	.74	.80	.89	98	1.00	.95
.40	1.13	.77	.69	.67	.68	.71	.77	.86	.97	1.04	.96
.60	1.14	.78	.69	.66	.66	.68	.73	.82	.95	1.09	1.04
.80	1.20	.82	.71	.66	.64	.65	.69	.77	.91	1.14	1.26
.95	1.35	.92	.78	.70	.66	.63	.65	.72	.87	1.16	1.59

$n_1 = n_2 = 10$ $\phi_1 = 0.60$

$\theta_1 = 0.60$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	.96	1.00	1.09	1.21	1.35	1.52	1.72	1.96	2.26	2.71	3.42
-.80	.97	.96	1.01	1.08	1.15	1.23	1.31	1.38	1.48	1.67	2.09
-.60	1.06	.94	.96	1.01	1.07	1.12	1.17	1.19	1.20	1.25	1.52
-.40	1.15	.93	.93	.96	1.01	1.06	1.10	1.11	1.09	1.06	1.24
-.20	1.22	.93	.90	.93	.97	1.01	1.05	1.07	1.03	.96	1.07
0.00	1.24	.91	.87	.89	.92	.97	1.02	1.04	1.01	.91	.95
.20	1.23	.88	.83	.84	.88	.92	.98	1.01	1.00	.88	.86
.40	1.17	.83	.78	.79	.82	.87	.93	.98	1.00	.89	.80
.60	1.09	.77	.72	.72	.75	.79	.86	.93	1.00	.93	.77
.80	1.00	.71	.65	.65	.66	.70	.76	.84	.95	1.02	.84
.95	.96	.68	.62	.60	.61	.62	.66	.73	.85	1.02	1.05

Table 2.1. (Cont'd)

$n_1 = n_2 = 10$ $\phi_1 = 0.00$

$\theta_1 = -0.60$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	.99	.95	.95	1.01	1.11	1.25	1.41	1.57	1.73	1.98	2.75
-.80	1.12	.96	.94	.96	.99	1.03	1.07	1.11	1.14	1.24	1.74
-.60	1.31	1.01	.96	.96	.98	1.00	1.01	1.00	.97	.98	1.31
-.40	1.43	1.05	.99	.98	.98	1.00	1.00	.98	.93	.88	1.10
-.20	1.50	1.08	1.00	.99	.99	1.00	1.01	.99	.93	.84	.97
0.00	1.54	1.10	1.02	1.00	1.00	1.02	1.03	1.02	.96	.83	.88
.20	1.55	1.11	1.02	1.00	1.01	1.02	1.05	1.05	1.01	.86	.83
.40	1.54	1.11	1.02	1.00	1.01	1.03	1.06	1.08	1.07	.93	.80
.60	1.54	1.11	1.02	1.00	1.00	1.02	1.06	1.10	1.14	1.06	.84
.80	1.59	1.15	1.04	1.01	1.00	1.01	1.04	1.10	1.18	1.25	1.04
.95	1.73	1.24	1.11	1.05	1.01	1.00	1.02	1.08	1.17	1.34	1.44

$n_1 = n_2 = 10$ $\phi_1 = 0.00$

$\theta_1 = 0.00$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	.98	1.01	1.07	1.15	1.25	1.36	1.49	1.65	1.88	2.30	3.45
-.80	1.01	.99	1.01	1.05	1.08	1.11	1.14	1.16	1.21	1.39	2.06
-.60	1.10	.98	.99	1.01	1.03	1.04	1.04	1.02	.99	1.03	1.46
-.40	1.19	1.00	.98	.99	1.00	1.01	1.00	.96	.90	.87	1.17
-.20	1.26	1.01	.98	.98	.99	1.00	.99	.95	.87	.79	.98
0.00	1.30	1.01	.97	.97	.99	1.00	.99	.95	.87	.74	.85
.20	1.31	1.01	.96	.96	.98	.99	1.00	.97	.89	.73	.76
.40	1.30	.99	.94	.95	.96	.99	1.01	1.00	.93	.76	.69
.60	1.28	.97	.92	.92	.94	.97	1.00	1.03	1.00	.84	.66
.80	1.26	.95	.90	.89	.91	.94	.98	1.03	1.07	1.01	.74
.95	1.29	.97	.91	.88	.88	.90	.93	.99	1.06	1.13	1.02

$n_1 = n_2 = 10$ $\phi_1 = 0.00$

$\theta_1 = 0.60$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	1.11	1.13	1.15	1.18	1.22	1.27	1.35	1.48	1.73	2.24	3.61
-.80	1.10	1.09	1.09	1.10	1.10	1.10	1.10	1.10	1.15	1.37	2.15
-.60	1.14	1.07	1.06	1.06	1.05	1.03	1.00	.96	.93	.99	1.51
-.40	1.20	1.07	1.05	1.04	1.02	1.00	.96	.90	.84	.82	1.18
-.20	1.27	1.08	1.04	1.02	1.01	.98	.94	.87	.79	.72	.98
0.00	1.32	1.08	1.04	1.02	1.00	.98	.93	.86	.76	.66	.83
.20	1.36	1.08	1.03	1.02	1.00	.98	.94	.86	.75	.62	.71
.40	1.36	1.07	1.02	1.01	1.00	.99	.96	.89	.77	.61	.61
.60	1.31	1.02	.98	.98	.99	1.00	.98	.94	.83	.63	.54
.80	1.20	.93	.90	.91	.93	.96	.99	1.00	.94	.74	.51
.95	1.08	.84	.81	.82	.84	.87	.91	.95	.97	.88	.60

Table 2.1.. (Cont'd)

$n_1 = n_2 = 10$ $\phi_1 = -0.60$

$\theta_1 = -0.60$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	.98	1.00	1.05	1.11	1.19	1.30	1.41	1.56	1.78	2.20	3.44
-.80	1.01	.99	1.00	1.02	1.04	1.06	1.07	1.09	1.13	1.30	2.04
-.60	1.07	1.00	1.00	1.00	1.00	1.00	.98	.94	.91	.95	1.44
-.40	1.12	1.02	1.00	1.00	.99	.98	.95	.90	.83	.80	1.14
-.20	1.15	1.04	1.02	1.01	1.00	.99	.95	.89	.80	.72	.95
0.00	1.18	1.05	1.03	1.02	1.02	1.00	.97	.91	.81	.68	.82
.20	1.19	1.07	1.04	1.04	1.04	1.03	1.01	.95	.84	.68	.72
.40	1.19	1.07	1.05	1.05	1.05	1.06	1.05	1.01	.91	.71	.65
.60	1.20	1.07	1.06	1.06	1.06	1.08	1.09	1.08	1.01	.81	.62
.80	1.22	1.09	1.07	1.07	1.07	1.09	1.11	1.13	1.14	1.01	.71
.95	1.29	1.15	1.11	1.09	1.08	1.09	1.10	1.14	1.18	1.19	1.02

$n_1 = n_2 = 10$ $\phi_1 = -0.60$

$\theta_1 = 0.00$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	1.05	1.07	1.12	1.17	1.24	1.32	1.42	1.58	1.83	2.35	3.80
-.80	1.03	1.03	1.05	1.09	1.09	1.11	1.11	1.13	1.18	1.39	2.21
-.60	1.03	1.01	1.02	1.03	1.03	1.03	1.01	.97	.93	.99	1.52
-.40	1.04	1.00	1.00	1.01	1.01	.99	.96	.90	.83	.81	1.19
-.20	1.05	1.00	1.00	1.00	.99	.98	.94	.88	.79	.72	.97
0.00	1.07	1.00	1.00	1.00	.99	.98	.94	.87	.77	.66	.82
.20	1.08	1.00	1.00	1.00	1.00	.98	.96	.89	.78	.63	.70
.40	1.07	1.00	.99	1.00	1.01	1.01	1.00	.94	.82	.63	.61
.60	1.06	.98	.98	1.00	1.02	1.04	1.04	1.01	.91	.69	.55
.80	1.04	.96	.96	.98	1.01	1.04	1.08	1.09	1.06	.86	.57
.95	1.04	.96	.95	.96	.98	1.01	1.05	1.09	1.12	1.06	.77

$n_1 = n_2 = 10$ $\phi_1 = -0.60$

$\theta_1 = 0.60$

γ_1 \ ψ_1	-.95	-.80	-.60	-.40	-.20	0.00	.20	.40	.60	.80	.95
-.95	1.08	1.08	1.09	1.11	1.14	1.19	1.26	1.39	1.64	2.18	3.73
-.80	1.05	1.04	1.04	1.04	1.04	1.04	1.03	1.04	1.09	1.32	2.20
-.60	1.03	1.02	1.01	1.00	.99	.97	.94	.90	.87	.94	1.52
-.40	1.03	1.00	.98	.97	.96	.93	.89	.84	.77	.77	1.18
-.20	1.03	.99	.97	.96	.94	.91	.86	.80	.72	.67	.96
0.00	1.04	.99	.97	.95	.93	.90	.85	.78	.69	.60	.80
.20	1.06	1.00	.97	.95	.93	.90	.85	.78	.67	.56	.68
.40	1.07	1.00	.98	.97	.95	.92	.87	.80	.68	.54	.58
.60	1.07	1.00	1.00	.99	.99	.97	.93	.85	.73	.55	.49
.80	1.02	.96	.97	.99	1.01	1.03	1.02	.98	.86	.63	.44
.95	.92	.87	.88	.91	.94	.98	1.01	1.02	.98	.80	.50

ϕ_1 = -0.6, 0.0 or 0.6 and θ_1 = -0.6, 0.0 or 0.6 tabulated for values of ψ_1 and γ_1, where MCI and SCI refer to multiconsequence and single consequence. In computing the tabled entities the values of $\hat{\phi}_S$ and $\hat{\theta}_S$ are used in determining Var($\hat{\delta}$)SCI. If the tabulated multiplier, M, is equal to 1.0000 then the standardization of $\hat{\delta}$ in hypothesis testing is unaffected by ignoring the multiconsequence phenomena. Although, only when the assumption of a single consequence structure is appropriate (e.g., $\phi_1 = \psi_1$, $\theta_1 = \gamma_1$) is the hypothesis test unaffected. There are, however, many combinations of parameters that reasonably mimic correct estimation of significance of δ. These combinations are typically those with relatively small changes in corresponding pre and post parameters from correct values. However, large variations from correct values can also mimic by chance due to the cancellation or offsetting of individual parameter contributions. As noted, more typically a severe overestimation of significance is obtained when M >> 1 and severe underestimation of significance occurs when M << 1.

2.6 Bias of $\hat{\mu}$ and $\hat{\delta}$

In Section 2.2, a flexible mean shift function,

$$\delta(t) = \begin{cases} 0 & t \le n_1 \\ \delta K(t) & n_1+1 \le t \end{cases}$$

was described where δ is the intrinsic utility of an intervention programs activities and K(t) is the dynamic coefficient whose specification is determined by whether the intervention effect is influenced or not influ-

enced by the environmental process, the indirect or direct stimulus-response model forms. The M.L.E. estimators of μ and δ were determined to be:

$$\hat{\mu} = \frac{[\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{K}][\underset{\sim}{1}^t M_N^{(p,0,q)} \underset{\sim}{Z}] - [\underset{\sim}{1}^t M_N^{(p,0,q)} \underset{\sim}{K}][\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{Z}]}{[\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{K}][\underset{\sim}{1}^t M_N^{(p,0,q)} \underset{\sim}{1}] - [\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{1}]^2}$$

and

$$\hat{\delta} = \frac{[\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{Z}][\underset{\sim}{1}^t M_N^{(p,0,q)} \underset{\sim}{1}] - [\underset{\sim}{Z}^t M_N^{(p,0,q)} \underset{\sim}{1}][\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{1}]}{[\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{K}][\underset{\sim}{1}^t M_N^{(p,0,q)} \underset{\sim}{1}] - [\underset{\sim}{K}^t M_N^{(p,0,q)} \underset{\sim}{1}]^2}$$

where $M_N^{(p,0,q)}$ is the inverse of the covariance matrix of $\underset{\sim}{Z}$ and $\underset{\sim}{K}$ = $[0,0,\ldots,K(n_1+1),K(n_1+2),\ldots,K(n_1+n_2)]^t$ with the first n_1 elements zero.

Since $E(\underset{\sim}{Z}) = \mu\underset{\sim}{1}+\delta\underset{\sim}{K}$, the estimators for the mean, μ, and program utility, δ, are unbiased when the $\underset{\sim}{K}$ vector is correctly assigned, since $E(\hat{\mu}) = \mu$ and $E(\hat{\delta}) = \delta$. However, when the $\underset{\sim}{K}$ vector is incorrectly assigned, the estimators are biased since, we have $\delta(t) = \delta K^0(t)$ rather than $\delta(t) = \delta K(t)$, eg.

$$E(\hat{\mu}) = \mu + \frac{[\underline{K}^{0^t} M_N^{(p,0,q)} \underline{K}^0][\underline{1}^t M_N^{(p,0,q)} \underline{K}] - [\underline{1}^t M_N^{(p,0,q)} \underline{K}^0][\underline{K}^0(t) M_N^{(p,0,q)} \underline{K}]}{[\underline{K}^{0^t} M_N^{(p,0,q)} \underline{K}^0][\underline{1}^t M_N^{(p,0,q)} \underline{1}] - [\underline{K}^{0^t} M_N^{(p,0,q)} \underline{1}]^2} \delta \qquad (2\text{-}34)$$

$$E(\hat{\delta}) = \frac{[\underline{K}^{0^t} M_N^{(p,0,q)} \underline{K}][\underline{1}^t M_N^{(p,0,q)} \underline{1}] - [\underline{K}^t M_N^{(p,0,q)} \underline{1}][\underline{K}^{0^t} M_N^{(p,0,q)} \underline{1}]}{[\underline{K}^{0^t} M_N^{(p,0,q)} \underline{K}^0][\underline{1}^t M_N^{(p,0,q)} \underline{1}] - [\underline{K}^{0^t} M_N^{(p,0,q)} \underline{1}]^2} \delta \qquad (2\text{-}35)$$

It should be noted that only when the correct specification of model structure, either direct stimulus response or indirect stimulus response is made, in which case does $\underline{K}^0 \rightarrow \underline{K}$, are the estimates of μ and δ unbiased. In this situation even when the model parameter vector of the noise process, composed of $\underline{\phi}$, $\underline{\theta}$, $\underline{\psi}$, $\underline{\gamma}$, is not correctly assigned, i.e., the matrix $M_N^{(p,0,q)}$ is not appropriate for the data, the estimates $\hat{\mu}$ and $\hat{\delta}$ are still unbiased. In addition, regardless whether the correct model specification is or is not made, the estimates of $\hat{\mu}$ and $\hat{\delta}$ are unbiased when $n_2 = 1$. (Effectively K(1) for the direct stimulus or indirect stimulus response model is identical.) This fact is exploited in the identification of the appropriate model structure which determines the specific form of K(t) in the mean shift function in the following section. Lastly, if the direct stimulus response structure is appropriate and identified correctly, the estimates of μ and δ are unbiased even when the noise process parameters are misspecified (eg. $K(t) \neq f(\underline{\phi},\underline{\theta},\underline{\psi},\underline{\gamma}) = 1$ all $t > n_1$).

The magnitude of bias due to misspecification of the K vector obtained from Equations (2-34) and (2-35) may be expressed as

$$E\begin{bmatrix}\hat{\mu}\\ \hat{\delta}\end{bmatrix} = \begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}\mu\\ \delta\end{bmatrix}, \qquad (2\text{-}36)$$

where A=1 and C=0 regardless of the form of K(t) or the correct specification of the parameters of the noise process. Both B and D are functions of the pre and post intervention noise process parameters. Thus given that the form of K(t) is not correct, the specification of a particular combination of noise process parameter values determines the magnitude of the bias in the estimates $\hat{\mu}$ and $\hat{\delta}$. Combinations of parameters giving rise to values of B close to zero and D close to 1 result in small biases.

The practical consequences of incorrectly specifying a multiconsequence or single consequence model or the misspecification of the noise parameter values for the correct specification for the indirect stimulus response form is illustrated in Example 1 in terms of the bias in $\hat{\mu}$ and $\hat{\delta}$.

Example 1

Table 2.2 gives computed values of B and D for the true parameters representing a multiconsequence process (e.g., $\phi_T = 0.6$, $\psi_T = 0.3$, $\theta_T = 0.3$ and $\gamma_T = 0.6$) and a single consequence process (e.g., $\phi_T = \psi_T = 0.6$ and $\theta_T = \gamma_T = 0.3$). Entries a, b and c in Table 2.2 are for the multiple consequence case where the preintervention parameters used are both set equal to the true parameters, equal to the true parameters plus 0.3 or minus 0.3, respectively. For comparison entries d and e are for the single consequence true process in which the preintervention parameters are set equal to the true parameters and equal to the true parameters plus 0.3, respectively. Entries a through e are for $n_1 = n_2 = 10$ while entry f corresponds to the parameter value of entry e only differing in that $n_1 = n_2 = 20$.

From Equation (2-36) the bias in $\hat{\mu}$ and $\hat{\delta}$ is seen to be proportional to δ; the bias in $\hat{\mu}$ being $B\delta$ and the bias in $\hat{\delta}$ being $(D-1)\delta$. From Table 2.2 the estimate of $\hat{\mu}$ is seen to be unbiased when $\gamma_F = 0.0$ regardless of the other misspecifications of model parameters. Typically, however, depending upon the combinations of the misspecified parameters, mild to severe bias are noted, being in some cases several times the magnitude of δ, with μ either overestimated (B and δ opposite signs), or underestimated (both B and δ of the same sign). Similar observations both with regard to the magnitude of the bias and direction of the bias (overestimation of underestimation) are noted for δ.

The ramification of ignoring the need of a multiconsequence model and using a single consequence model allowing only for a shift in the process label is illustrated in Table 2.2(a) - (c). Although a small nonconsequential bias in the mean is expected the intervention impact is overestimated by approximately 40%. Similarly severe biases both positive and negative are illustrated in Table 1(d) in situations in which a single consequence model is appropriate while the multiconsequence form is fit. The potential magnitudes and signs of the biases clearly indicate the need to consider the multiconsequence formulation in real world evaluations of intervention experiments.

As n_1 and n_2 increase, the bias in the estimates of μ and δ

asymptotically stabilizes to a fixed bias. The comparison of entry e and f which corresponds to an increase from $n_1 = n_2 = 10$ to $n_1 = n_2 =$ 20 illustrates a rapid convergence for most combinations of parameters β except severe departures of the postintervention autoregressive parameter from its true value. Thus increasing sample size does not offset the biases induced by parameter misspecification.

From Table 2.2, the relative robustness of the estimators μ and δ are observable. The estimator μ is more robust than the estimator δ for departures of $\underset{\sim}{\beta}$ from $\underset{\sim}{\beta}_T$. Departures toward the boundaries of the stationary or invertibility region are more severe. The estimator μ is more robust to departures from true value of the preintervention parameters ϕ_T and θ_T than the corresponding postintervention parameters. Departures in the autoregressive parameters exert a greater effect than equivalent magnitude departures in the moving average parameters.

Regardless whether the misspecification of noise parameters cause a bias in the estimates of μ and δ, it does affect the detection of the true significance since the $Var(\hat{\mu})$ and $Var(\hat{\delta})$ are not correct for improperly specified noise parameters. For the indirect stimulus response model, which may result in small biases in the estimate of $\hat{\delta}$ due to noise parameter misspecification resulting in $\underset{\sim}{K}^0 = \underset{\sim}{K}$, the statistical significance of $\hat{\delta}$ may be severely distorted.

<u>Example 2</u>

Table 2.3 gives entries of h where

$$h = [E(Var(\hat{\delta}/\hat{\sigma}_a)_{MISS}/E(Var(\hat{\delta}/\hat{\sigma}_a)_{COR}]^{1/2} \tag{2-37}$$

Table 2.2. Value of B and D for Misspecifications of $\tilde{\beta}$ Parameters (Entries are B above, D below)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.6,\ \theta_F = 0.3$

$\gamma_F \backslash \psi_F$	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.03 0.61	0.06 0.48	0.05 0.40	-0.03 0.35	-0.15 0.29	-0.31 0.20	-0.48 0.06
-0.6	0.00 0.74	0.02 0.61	0.02 0.50	0.01 0.40	-0.02 0.30	-0.06 0.20	-0.12 0.08
-0.3	0.00 0.91	0.01 0.76	0.01 0.62	0.01 0.49	0.00 0.36	-0.02 0.23	-0.04 0.09
0.0	0.00 1.15	0.00 0.98	0.00 0.81	0.00 0.63	0.00 0.46	0.00 0.29	0.00 0.11
0.3	0.06 1.41	0.03 1.28	0.00 1.10	-0.01 0.88	-0.01 0.64	0.00 0.40	0.02 0.16
0.6	0.37 0.95	0.28 1.13	0.18 1.22	0.08 1.19	0.00 1.00	-0.02 0.65	0.03 0.26
0.9	0.58 0.12	0.56 0.26	0.53 0.41	0.49 0.58	0.39 0.78	0.19 0.96	-0.02 0.65

(a)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.9,\ \theta_F = 0.6$

$\gamma_F \backslash \psi_F$	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.05 0.58	0.16 0.39	0.24 0.27	0.25 0.20	0.15 0.18	0.08 0.16	-0.38 0.06
-0.6	0.01 0.73	0.05 0.58	0.08 0.46	0.08 0.36	0.05 0.27	0.00 0.19	-0.08 0.08
-0.3	0.00 0.91	0.02 0.75	0.03 0.60	0.03 0.47	0.02 0.35	0.00 0.22	-0.02 0.09
0.0	0.00 1.15	0.00 0.98	0.00 0.81	0.00 0.63	0.00 0.46	0.00 0.29	0.00 0.14
0.3	0.12 1.28	0.05 1.22	0.01 1.09	-0.02 0.90	-0.03 0.67	-0.02 0.41	0.01 0.16
0.6	0.43 0.72	0.35 0.91	0.24 1.05	0.12 1.10	0.00 1.00	-0.06 0.69	0.00 0.27
0.9	0.58 0.23	0.56 0.35	0.54 0.47	0.50 0.61	0.42 0.76	0.24 0.91	-0.05 0.66

(b)

Table 2.2. (Cont'd)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.3,\ \theta_F = 0.0$

$\gamma_F \backslash \psi_F$	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.01 0.62	0.03 0.51	-0.01 0.44	-0.08 0.38	-0.19 0.30	-0.31 0.20	-0.44 0.05
-0.6	0.00 0.74	0.01 0.62	0.01 0.51	-0.01 0.41	-0.03 0.31	-0.07 0.20	-0.11 0.08
-0.3	0.00 0.91	0.00 0.76	0.01 0.63	0.00 0.50	-0.01 0.36	-0.02 0.23	-0.04 0.09
0.0	0.00 1.15	0.00 0.98	0.00 0.81	0.00 0.63	0.00 0.46	0.00 0.29	0.00 0.11
0.3	0.05 1.45	0.02 1.29	0.00 1.10	-0.01 0.88	-0.01 0.64	0.00 0.39	0.02 0.16
0.6	0.33 1.11	0.24 1.26	0.15 1.31	0.06 1.23	0.00 1.00	-0.01 0.64	0.04 0.25
0.9	0.58 0.11	0.56 0.25	0.53 0.42	0.47 0.61	0.37 0.83	0.16 1.00	-0.01 0.64

(c)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.6,\ \gamma_T = 0.3$

$\phi_F = 0.6,\ \theta_F = 0.3$

$\gamma_F \backslash \psi_F$	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	-0.03 1.60	0.13 1.32	0.02 1.09	-0.02 0.87	-0.10 0.66	-0.20 0.43	-0.32 0.17
-0.6	-0.05 1.91	-0.03 1.59	0.00 1.29	-0.00 1.02	-0.01 0.75	-0.04 0.48	-0.08 0.20
-0.3	-0.07 2.33	-0.04 1.94	-0.02 1.58	0.00 1.24	0.00 0.90	-0.01 0.58	-0.03 0.25
0.0	0.00 2.81	0.00 2.40	0.00 1.98	0.00 1.56	0.00 1.15	0.00 0.73	0.00 0.31
0.3	0.46 2.61	0.30 2.58	0.17 2.38	0.09 2.02	0.02 1.54	0.00 1.00	0.00 0.43
0.6	1.71 -1.23	1.42 0.00	1.07 1.06	0.68 1.77	0.30 1.95	0.04 1.52	0.02 0.70
0.9	1.77 -3.80	1.75 -2.96	1.71 -2.07	1.62 -1.10	1.42 0.02	0.91 1.29	0.04 1.54

(d)

Table 2.2. (Cont'd)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.6,\ \gamma_T = 0.3$

$\phi_F = 0.9,\ \theta_F = 0.6$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	-0.06 1.63	0.03 1.30	0.11 1.02	0.14 0.78	0.09 0.59	-0.05 0.40	-0.25 0.17
-0.6	-0.15 2.01	-0.07 1.64	-0.01 1.30	0.02 1.01	0.02 0.73	0.00 0.47	-0.06 0.20
-0.3	-0.16 2.46	-0.09 2.01	-0.04 1.61	-0.01 1.24	0.01 0.90	0.00 0.57	-0.02 0.25
0.0	0.00 2.81	0.00 2.40	0.00 1.98	0.00 1.56	0.00 1.15	0.00 0.73	0.00 0.31
0.3	0.82 1.68	0.57 2.00	0.35 2.06	0.18 1.88	0.06 1.51	0.00 1.00	0.01 0.44
0.6	2.00 -2.32	1.76 -1.10	1.45 0.06	1.03 1.03	0.55 1.58	0.12 1.46	0.00 0.71
0.9	1.77 -3.47	1.76 -2.68	1.73 -1.86	1.67 -1.00	1.53 -0.05	1.14 1.03	0.11 1.50

(e)

$n_1 = n_2 = 20$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.6,\ \theta_F = 0.3$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.02 0.59	0.04 0.48	0.03 0.39	-0.01 0.33	-0.10 0.26	-0.20 0.18	-0.33 0.06
-0.6	0.00 0.71	0.01 0.59	0.01 0.48	0.00 0.38	-0.01 0.28	-0.04 0.17	-0.07 0.06
-0.3	0.00 0.87	0.00 0.73	0.01 0.60	0.00 0.46	0.00 0.33	-0.01 0.20	-0.02 0.07
0.0	0.00 1.12	0.00 0.95	0.00 0.77	0.00 0.60	0.00 0.43	0.00 0.26	0.00 0.09
0.3	0.04 1.47	0.01 1.29	0.00 1.08	-0.01 0.85	-0.01 0.61	0.00 0.36	0.01 0.12
0.6	0.29 1.29	0.22 1.38	0.13 1.39	0.06 1.26	0.00 1.00	-0.01 0.62	0.02 0.20
0.9	0.56 0.28	0.55 0.36	0.53 0.44	0.51 0.55	0.44 0.70	0.28 0.90	-0.01 0.62

(f)

where MISS and COR denotes misspecified and correct, respectively, for the same combinations of the noise parameters used in Table 2.2. When $h > 1$, the misspecification of these parameters result in an overestimation of $\mathrm{Var}(\hat{\delta}/\hat{\sigma}_a)$ and consequently, and underestimation of the significance of $\hat{\delta}$. Similarly, when $h < 1$, we effectively underestimate $\mathrm{Var}(\hat{\delta}/\hat{\sigma}_a)$ and overestimate the significance of $\hat{\delta}$. As seen from the range exhibited for h in Table 2.3, severe overestimation or underestimation of $\mathrm{Var}(\hat{\delta}/\hat{\sigma}_a)$ can occur as values of h from 0.1 to 2.4 are seen.

The composite consequence of the bias in the point estimates and the over or under estimation of h is evaluable using the results of both Table 2.2 and Table 2.3. The scale factor to measure departure of the test statistics in testing whether H_0: $\delta = 0$ versus H_1: $\delta \neq 0$ is D/h. When this ratio is one, the test statistic adequately portrays the time significance or non significance. When this scale factor is close to one, this might be due to the compensation of severe departures in B and h or due to minor departures in B and h from one. In the former situation although adequate portrayal of the significance of the intervention effect δ is obtained severe error in the magnitude of the intervention program utility for subsequent policy inferences such as the value of a certain program type is present.

More typically values of the scale factor result in severe misestimation of the test statistics. For example, from entry a of Table 2.2 and Table 2.3 for $\psi_F = 0.6$ and $\gamma_F = -0.6$, the scale factor is .27 while when $\psi_F = 0.0$ and $\gamma_F = 0.6$ the scale factor is 1.25. When testing

Table 2.3. Computed Values of h for Misspecified Parameters $\underset{\sim}{\beta}$

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.6,\ \theta_F = .3$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.98	0.93	0.69	0.48	0.35	0.28	0.24
-0.6	0.95	1.05	1.02	0.92	0.81	0.75	0.70
-0.3	0.89	1.00	1.05	1.03	0.98	0.98	1.02
0.0	0.81	0.93	1.02	1.05	1.03	1.05	1.19
0.3	0.71	0.84	0.95	1.02	1.03	1.04	1.22
0.6	0.57	0.70	0.84	0.95	1.00	0.99	1.08
0.9	0.33	0.43	0.57	0.73	0.88	0.93	0.85

(a)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.9,\ \theta_F = 0.6$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	1.12	1.43	1.54	0.93	0.47	0.28	0.24
-0.6	1.44	1.90	2.17	1.87	1.28	0.86	0.69
-0.3	1.45	1.90	2.27	2.21	1.74	1.26	1.02
0.0	1.35	1.79	2.22	2.35	2.01	1.52	1.20
0.3	1.17	1.59	2.07	2.37	2.18	1.69	1.30
0.6	0.91	1.27	1.78	2.27	2.27	1.79	1.27
0.9	0.49	0.72	1.13	1.73	2.07	1.76	1.21

(b)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \psi_T = 0.6$

$\phi_F = 0.3,\ \theta_F = 0.0$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.86	0.78	0.59	0.44	0.34	0.29	0.24
-0.6	0.81	0.88	0.87	0.81	0.76	0.74	0.70
-0.3	0.75	0.84	0.88	0.89	0.89	0.94	1.02
0.0	0.68	0.77	0.84	0.88	0.91	0.99	1.19
0.3	0.59	0.69	0.77	0.83	0.87	0.94	1.20
0.6	0.49	0.58	0.68	0.76	0.81	0.85	1.05
0.9	0.28	0.36	0.46	0.58	0.70	0.76	0.78

(c)

Table 2.3. (Cont'd)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.6,\ \gamma_T = 0.3$

$\phi_F = 0.6,\ \theta_F = 0.3$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.86	0.92	0.78	0.58	0.44	0.36	0.32
-0.6	0.80	0.92	0.97	0.96	0.91	0.87	0.84
-0.3	0.70	0.82	0.92	0.98	1.01	1.07	1.15
0.0	0.58	0.70	0.82	0.92	0.99	1.08	1.28
0.3	0.45	0.56	0.68	0.80	0.90	1.00	1.24
0.6	0.39	0.38	0.49	0.62	0.75	0.87	1.03
0.9	0.12	0.16	0.22	0.30	0.43	0.60	0.74

(d)

$n_1 = n_2 = 10$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.6,\ \gamma_T = 0.3$

$\phi_F = 0.9,\ \theta_F = 0.6$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	0.98	1.42	1.73	1.13	0.59	0.36	0.31
-0.6	1.20	1.66	2.07	1.95	1.43	1.01	0.83
-0.3	1.14	1.56	1.99	2.10	1.79	1.37	1.15
0.0	0.98	1.35	1.79	2.06	1.92	1.56	1.30
0.3	0.75	1.06	1.48	1.86	1.91	1.63	1.31
0.6	0.47	0.69	1.04	1.48	1.71	1.57	1.22
0.9	0.17	0.26	0.43	0.72	1.02	1.15	1.06

(e)

$n_1 = n_2 = 20$ $\phi_T = 0.6,\ \theta_T = 0.3,\ \psi_T = 0.3,\ \gamma_T = 0.6$

$\phi_F = 0.6,\ \theta_F = 0.3$

γ_F \ ψ_F	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9
-0.9	1.02	0.94	0.72	0.54	0.43	0.36	0.31
-0.6	0.99	1.06	1.04	0.97	0.92	0.94	0.99
-0.3	0.92	1.02	1.06	1.06	1.06	1.16	1.44
0.0	0.85	0.95	1.02	1.06	1.08	1.20	1.66
0.3	0.76	0.87	0.96	1.02	1.05	1.13	1.64
0.6	0.62	0.74	0.85	0.95	1.00	1.03	1.36
0.9	0.34	0.44	0.56	0.71	0.86	0.93	0.92

(f)

at the $\alpha = 0.05$ level of significance for values of the true parameter δ between the critical value of the test and approximately four times the critical value would result in the incorrect inference of accepting H_0 when H_1 should be preferred. In the latter case, values of the true parameter less than the critical value of the test but greater than approximately 0.80 of the critical value results in the incorrect inference of rejecting H_0 and H_0 is true.

2.7 Modeling Procedures

For the modeling of multiconsequence intervention time series experiments the time occurrence of an intervention program is known which determines the preintervention and postintervention series size n_1 and n_2 respectively. Also from the experiment it is known whether the intervention activities are directed specifically to subjects whose response forms the time series monitored or is in fact directed to the general environment from which the time series responses are monitored. In the former case we start with the instantaneous mean shift multiconsequence model structure with known mean shift function $\delta(t) = \delta K(t)$ where $K(t)$ is zero for $t \leq n_1$ and one for $t > n_1$, while for the latter we start with the transient mean level change model structure to allow for the realized change in mean level to be influenced by the environmental process. As discussed in Section 2.2.2, when this distinction cannot be easily made, the analysis should be initiated with the more general indirect stimulus-response case or transient mean level shift structure with unknown mean shift function form to identify the appropriate model form.

The steps for modeling the indirect-stimulus response case are;

1) Estimate μ, $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}$ from the n_1 preintervention observations.
2) Identif~~ication of~~ the form of the mean shift function by sequential estimation of $\delta(t)$.
   Recall that when $n_2 = 1$, $K(t)$ is the same for both the transient or instantaneous mean level cases. Therefore, incorrect values of postintervention parameters do not bias the estimation of $\delta(t)$. Thus $\underset{\sim}{\psi}$ and $\underset{\sim}{\gamma}$ may be set equal to $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}$ (e.g., the single consequence model).
3) From the sequential estimates of the mean shift function with $n_2 = 1$ determine the form of $K(t)$.
4) Estimate $\underset{\sim}{\delta}$, $\underset{\sim}{\psi}$ and $\underset{\sim}{\gamma}$ from the n_2 postintervention observations.
5) Estimate $\underset{\sim}{\mu}$, $\underset{\sim}{\delta}$, $\underset{\sim}{\phi}$, $\underset{\sim}{\theta}$, $\underset{\sim}{\psi}$ and $\underset{\sim}{\gamma}$ simultaneously from the n_1+n_2 observations.

Figure 2.5 pictorially describes the modeling procedures. The estimation procedures called for in each step here are those delineated in Section 2.2. As seen from this figure, the primary difference in the modeling of direct stimulus-response time series experiments from the indirect-stimulus response experiments is that in the latter, the mean shift function form is not known and has to be identified. The estimation procedures for the mean shift is described in Section 2.2.2.

2.8 Considerations in Designing Interrupted Time Series Experiments

The design of interrupted time series experiments consists of specifying the appropriate number of pre and post intervention observations. By appropriate, we refer to the individual size of n_1 and n_2 and

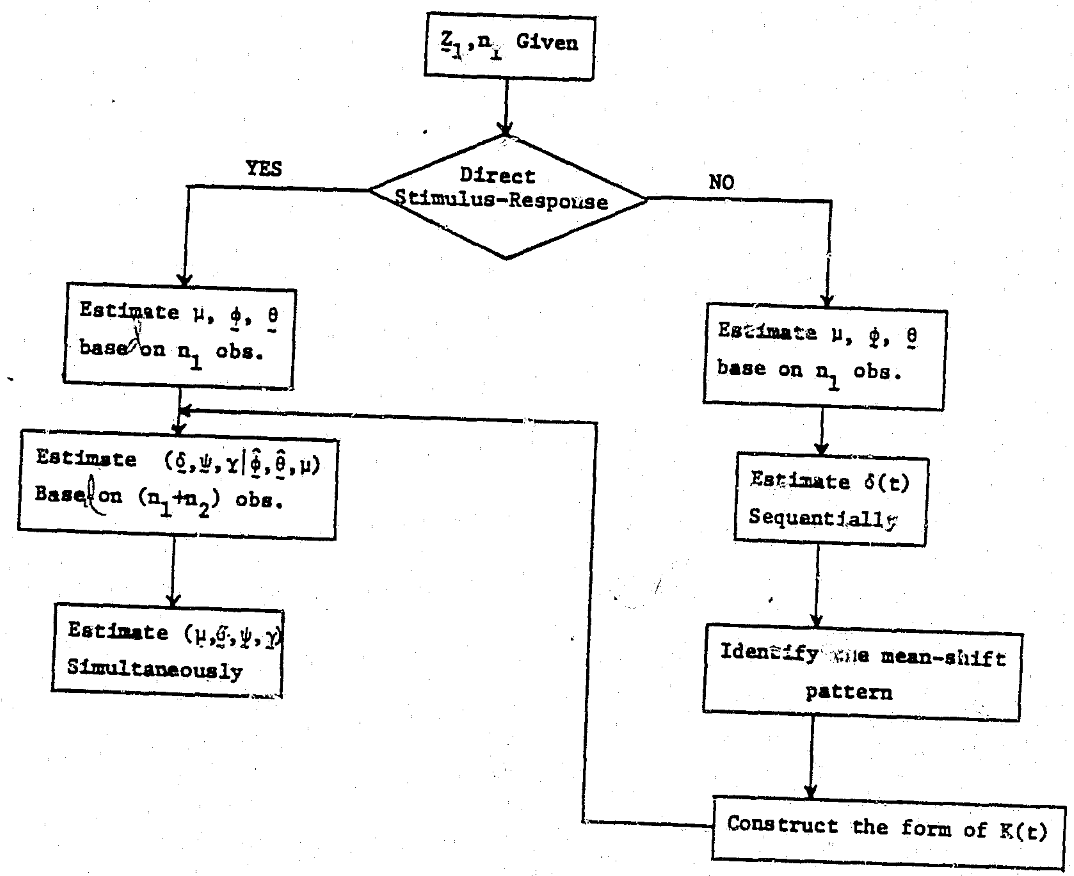


Figure 2.5. Multiconsequence Intervention Model Building Procedure

also their relative size to be able to;

change in structure

1) statistically detect a given magnitude covariance, as measured by the change in the corresponding pre and post intervention autoregressive and moving average parameters, with Type I error α_1 and,

2) statistically detect a standardized magnitude of an intervention program's utility δ/σ with Type I error α_2.

In this section we describe each of these considerations in turn and then focus the results derived for interrupted time series designs.

2.8.1 Sample Sizes for Detecting a Change in Covariance

Intuitively, the smaller difference between $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$, the more difficult it is to reject the hypotheses H_0: $\underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2$. Further the more observations we have, the better chance that we will be able to detect any real difference that exists. In developing specific sample sizes, we will assume that both the multiconsequence model or the single consequence model forms have been correctly identified and estimated. That is, we assume K(t) is assigned correctly and $\hat{\mu} = \mu$, $\hat{\delta} = \delta$ with probability 1. We will denote $Z_t^c = Z_t - \mu$, $t \leq n_1$ and $Z_t^c - \mu - \delta\cdot K(t)$, $t \geq n_1+1$, i.e., $Z^c \sim N(0, \sum_N^{(p,0,q)})$, where $N = n_1 + n_2$.

Thus, when the correct model is a multiconsequence intervention form,

$$\phi_p(B)Z_t^c = \Theta_q(B)a_t \qquad t \leq n_1$$

$$\psi_p(B)Z_t^c = \Gamma_q(B)a_t \qquad t \geq n_1+1 \qquad (2\text{-}38)$$

the incorrect single consequence model will be,

$$\hat{\phi}_p^0(B) Z_t^c = \hat{H}_q^0(B) A_t \qquad \text{all } t. \qquad (2\text{-}39)$$

For known parameter values of $\phi_p(B)$, $\psi_p(B)$, $\hat{H}_q(B)$, $\Gamma_q(B)$ of the correct model, the parameters β^0 of the incorrect single consequence form contained in $\hat{\phi}_p^0(B)$ and $\hat{H}_q^0(B)$ can be estimated so as to minimize $\sum_{t=1}^{n_1+n_2} A_t^2$. Since the multiconsequence form is correct, the resulting model of the residuals of the single consequence model is,

$$\hat{\phi}_p^{0^{-1}}(B)\hat{H}_q^0(B)\hat{A}_t = \phi_p^{-1}(B) H_q(B) a_t \qquad t \le n_1$$

$$\hat{\phi}_p^{0^{-1}}(B)\hat{H}_q^0(B)\hat{A}_t = \Psi_p^{-1}(B)\Gamma_q(B) a_t \qquad t \ge n_1+1 \qquad (2\text{-}40)$$

or

$$\hat{A}_t = \hat{\phi}_p^0(B)\phi_p^{-1}(B)\hat{H}_q^{0^{-1}}(B) H_q(B) a_t \qquad t \le n_1$$

$$\hat{A}_t = \hat{\phi}_p^0(B)\Psi_p^{-1}(B)\hat{H}_q^{0^{-1}}(B)\Gamma_q(B) a_t \qquad t \ge n_1+1 \qquad (2\text{-}41)$$

The test statistic W for the null hypothesis that a single consequence model is appropriate (e.g.: H_0: $\beta_1 = \beta_2$) is

$$(W_\beta | n_1, n_2, \beta) = \min_{\hat{\beta}^0} (n_1+n_2) \ln\left(\sum_{t=1}^{n_1+n_2} \hat{A}_t^2/\sigma_\varepsilon^2\right) + \ln(M_N^{(p,0,q)}(\hat{\beta}^0))$$

$$- \ln(M_N^{(p,0,q)}(\beta))$$

where β_1 and β_2 are the pre-invention and the post intervention model parameter vectors, respectively, $\beta^t = (\beta_1^t, \beta_2^t)$, and $\hat{\beta}^0$ is the corresponding single consequence model parameter vector, and $M_N^{(p,0,q)}(\hat{\beta})$ and $M_N^{(p,0,q)}(\beta)$ are the inverses of the covariance matrix of Z^c constructed with $\hat{\beta}^0$ and β, respectively. Taking the expectation of $(W_\beta | n_1, n_2, \beta)$ results in,

$$E(W_\beta | n_1, n_2, \beta) = \min_{\hat{\beta}^0} (n_1+n_2) \ln\left[\sum_{t=1}^{n_1+n_2} E(\hat{A}_t^2)\right] + \ln(|M_N^{(p,0,q)}(\hat{\beta}^0)|)$$

$$- \ln(|M_N^{(p,0,q)}(\beta)|) \qquad (2\text{-}42)$$

where $E(\hat{A}_t^2)$ is approximately computed from Equation (2-40) and Equation (2-41) by applying the covariance generating function and $\hat{\beta}^0$ is obtained by searching through the $(\beta | \beta_1 = \beta_2)$ subspace. If the $E(W_\beta | n_1, n_2, \beta) > \chi^2_{\alpha, q}$ then the sample size (n_1, n_2) is large enough to statistically detect the differences between β_1 and β_2 at the $(1-\alpha)$ level.

For example, for the ARIMA(1,0,1)MCI model with $\theta_1, \phi_1, \gamma_1, \psi_1$ given,

$$Z_t^c = \phi_1 Z_{t-1}^c - \theta_1 a_{t-1} + a_t \qquad t \leq n_1$$

$$Z_t^c = \psi_1 Z_{t-1}^c - \gamma_1 a_{t-1} + a_t \qquad t \geq n_1+1 \tag{2-43}$$

and the alternative single consequence model,

$$Z_t^c = \hat{\phi}_1^0 Z_{t-1}^c - \hat{\theta}_1^0 \hat{A}_{t-1} + \hat{A}_t \qquad \text{all } t$$

where $\hat{\phi}_1^0$ and $\hat{\theta}_1^0$ are the parameters the minimize $\Sigma E(\hat{A}_t^2|n_1,n_2,\underset{\sim}{\beta})$. The residuals of the single consequence set are;

$$\hat{A}_t = \left(\frac{1-\theta_1 B}{1-\hat{\theta}_1^0 B}\right)a_t + (\phi_1 - \hat{\phi}_1^0)B\left(\frac{1-\theta_1 B}{1-\phi_1 B}\right)a_t \qquad t \leq n_1$$

$$\hat{A}_t = \left(\frac{1-\gamma_1 B}{1-\hat{\theta}_1^0 B}\right)a_t + (\psi_1 - \hat{\phi}^0)B\left(\frac{1-\gamma_1 B}{1-\psi_1 B}\right)a_t \qquad t \geq n_1+1 . \tag{2-44}$$

To compute $\sum E(\hat{A}_t^2|n_1,n_2,\underset{\sim}{\beta})$ we apply the covariance generating function which results in,

$$E(\hat{A}_t^2) = [1 + \frac{(\hat{\theta}_1^0-\theta_1)^2}{1-\hat{\theta}_1^{0^2}} + \frac{2(\hat{\theta}_1^0-\theta_1)(\phi_1-\hat{\phi}_1^0)(1-\hat{\theta}_1^0\theta_1)}{1-\phi_1\hat{\theta}_1^0} + \frac{(\phi_1-\hat{\phi}_1^0)^2(1+\theta_1^2-2\theta_1\phi_1)}{1-\phi_1^2}]\sigma_a^2 \quad t \leq n_1$$

$$E(\hat{A}_t^2) = [1 + \frac{(\hat{\theta}_1^0-\gamma_1)^2}{1-\hat{\theta}_1^{0^2}} + \frac{2(\hat{\theta}_1^0-\gamma_1)(\psi_1-\hat{\phi}_1^0)(1-\hat{\theta}_1^0\gamma_1)}{1-\psi_1\hat{\theta}_1^0} + \frac{(\psi_1-\hat{\phi}_1^0)^2(1+\gamma_1^2-2\gamma_1\psi_1)}{1-\psi_1^2}]\sigma_a^2 \quad t \geq n_1+1 \tag{2-45}$$

Thus,

$$E(W_{\underset{\sim}{\beta}}|n_1,n_2,\underset{\sim}{\beta}) = \min_{\hat{\theta}_1^0,\hat{\phi}_1^0}: \quad (n_1+n_2)\ \ell n\ f_1 + \ell n\ f_2 - \ell n\ f_3 \tag{2-46}$$

where

$$f_1 = \frac{n_1}{n_1+n_2} E(\hat{A}_t^2|t \leq n_1) + \frac{n_2}{n_1+n_2} E(\hat{A}_t^2|t \geq n_1+1),$$

$$f_2 = |\Sigma_N^{(1,0,1)}(\hat{\phi}_1^0,\hat{\theta}_1^0)|^{-1} \text{ and}$$

$$f_3 = |\Sigma_N^{(1,0,1)}(\phi_1,\theta_1,\psi_1,\gamma_1)|^{-1}.$$

For a fixed preintervention sample size n_1, the general behavior of the sample size n_2 to detect the covariance change for the ARIMA (1,0,1)MCI model is indicated by the sign of the third term for each equation labeled (4-8), which is the only term that can be negative. When $\phi_1 > \psi_1$ and $\theta_1 > \gamma_1$ or $\phi_1 < \psi_1$ and $\theta_1 < \gamma_1$ each of the third terms are negative otherwise they are positive. In the former cases, n_2 increase when $|\phi_1-\psi_1|$ is small and decrease when $|\phi_1-\psi_1|$ is large. From Equation (2-46), sample size requirements for the (1,0,0)MCI or (0,0,1)MCI model can be obtained by setting the appropriate subset of the noise parameters to zero. For example, when $\phi_1 = \psi_1 = \hat{\phi}_1^0 = 0$ we obtain the $E(W_\beta|n_1,n_2,\theta_1,\gamma_1)$ for the (0,0,1) process.

<u>Example 3</u>

In the following numerical example, the magnitude of the post intervention sample size required to determine a given size covariance change is addressed. Tables 2.4(a) - (c) illustrate the computed sample size requirements for the (0,0,1) process for $n_1 = Kn_2$ with the tabled values being n_2 for values of K = 1.0, 0.5 and 2.0. Thus, equivalent time histories of the pre and post intervention segments are available when K = 1.0 while more preintervention history is available when K > 1.0 and more postintervention history than preintervention history. In each table, e.g., the required size of the n_2 sample is seen to be sensitive to the combination of parameter values (θ_1, γ_1) for situations in which $|\theta_1-\gamma_1|$ is relatively small (e.g., values around the diagonal where $\theta_1=\gamma_1$). As expected the larger the magnitude of $|\theta_1-\gamma_1|$ the smaller the requirements for the postintervention sample size. In fact, for larger values contained in the top right or bottom left the sample size becomes

Table 2-4. Sample Sizes of n_2 for detecting Covariance Change of STMA(1) Process Given $n_2 = k\,n_1$.

(a) K = 1.0

Significant Level = 0.15

	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	–	28	10	6	4	4	3	3	2	2	2
-0.80	72	–	42	13	7	5	4	3	3	3	2
-0.60	25	71	–	71	19	10	6	5	4	3	3
-0.40	14	24	92	–	92	24	12	7	5	4	4
-0.20	9	13	27	105	–	105	27	13	8	6	5
0.00	6	8	13	28	109	–	109	28	13	8	6
0.20	5	6	8	13	27	105	–	105	27	13	9
0.40	4	4	5	7	12	24	92	–	92	24	14
0.60	3	3	4	5	6	10	19	71	–	71	25
0.80	2	3	3	3	4	5	7	13	42	–	72
0.95	2	2	2	3	3	4	4	6	10	28	–

Significant Level = 0.05

	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	–	77	23	14	10	8	7	6	6	5	5
-0.80	242	–	139	39	20	13	10	8	7	6	5
-0.60	82	241	–	241	64	31	19	14	11	9	8
-0.40	45	82	314	–	314	82	39	24	16	13	11
-0.20	29	43	93	358	–	358	93	43	26	18	15
0.00	20	27	45	96	373	–	373	96	45	27	20
0.20	15	18	26	43	93	358	–	358	93	43	[illegible]
0.40	11	13	16	24	39	82	314	–	314	82	45
0.60	8	9	11	14	19	31	64	241	–	241	82
0.80	6	6	7	8	10	13	20	39	139	–	242
0.95	5	5	6	6	7	8	10	14	23	77	–

Table 2-4. (Cont'd)

(b) K = 0.5

Significant Level = 0.15

	−0.95	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	0.95
−0.95	–	28	10	6	4	4	3	3	3	2	2
−0.80	7Y	–	42	13	7	5	4	3	3	3	3
−0.60	25	71	–	71	20	10	7	5	4	3	3
−0.40	14	25	92	–	92	25	12	8	6	4	4
−0.20	9	13	28	105	–	105	28	13	8	6	5
0.00	7	9	14	29	109	–	109	29	14	9	7
0.20	5	6	8	13	28	105	–	105	28	13	9
0.40	4	4	6	8	12	25	92	–	92	25	14
0.60	3	3	4	5	7	10	20	71	–	71	25
0.80	3	3	3	3	4	5	7	13	42	–	72
0.95	2	2	3	3	3	4	4	6	10	28	–

Significant Level = 0.05

	−0.95	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	0.95
−0.95	–	78	23	14	11	9	7	7	6	5	5
−0.80	243	–	140	40	21	14	11	9	8	7	7
−0.60	83	242	–	242	65	32	20	15	12	10	9
−0.40	46	83	315	–	315	83	40	25	17	14	12
−0.20	30	45	94	359	–	359	94	45	27	19	16
0.00	21	28	46	97	375	–	374	97	46	28	21
0.20	16	19	27	45	94	359	–	359	94	45	30
0.40	12	14	17	25	40	83	315	–	315	83	46
0.60	9	10	12	15	20	32	65	242	–	242	83
0.80	7	7	8	9	11	14	21	40	140	–	243
0.95	5	6	6	7	7	9	11	14	23	78	–

Table 2-4. (Cont'd)

(c) K = 2.0

Significant Level = 0.15

	−0.95	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	0.95
−0.95	–	28	10	6	4	4	3	3	2	2	2
−0.80	72	–	42	13	7	5	4	3	3	2	2
−0.60	24	70	–	70	19	9	6	4	3	3	3
−0.40	13	24	92	–	92	24	11	7	5	4	3
−0.20	9	13	27	104	–	104	27	13	8	5	4
0.00	6	8	13	28	109	–	109	28	13	8	6
0.20	4	5	8	13	27	104	–	104	27	13	9
0.40	3	4	5	7	11	24	92	–	92	24	13
0.60	3	3	3	4	6	9	19	70	–	70	24
0.80	2	2	3	3	4	5	7	13	42	–	72
0.95	2	2	2	3	3	4	4	6	10	28	–

Significant Level = 0.05

	−0.95	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	0.95
−0.95	–	76	22	13	9	8	6	6	5	5	4
−0.80	241	–	137	38	19	13	9	8	6	6	5
−0.60	81	239	–	239	62	30	18	13	10	8	7
−0.40	44	81	313	–	313	81	37	22	15	12	10
−0.20	28	42	91	357	–	357	91	42	25	17	14
0.00	19	26	44	95	372	–	372	95	44	26	19
0.20	14	17	25	42	91	357	–	357	91	42	28
0.40	10	12	15	22	37	81	313	–	313	81	44
0.60	7	8	10	13	18	30	62	239	–	239	81
0.80	5	6	6	8	9	13	19	38	137	–	241
0.95	4	5	5	6	6	8	9	13	22	76	–

fixed without variation for different parameter combinations. Further, in comparison to the case exhibited in which $n_1 = n_2$, increasing the relative length of the preintervention history to the postintervention history results in smaller magnitudes of n_2, for comparable size changes in covariance for a given α level while decreasing the relative size of the preintervention history to the postintervention history dictates a larger postintervention history n_2 for equivalent magnitude changes in covariance for a given α level. This tradeoff is exploited, in Section 2.9, in the generation of optimal designs for interrupted time series experiments.

2.8.2 Sample Sizes for Detecting a Change in Mean Level

The test of significance of δ involves the evaluation of the hypothesis,

$$H_0:\ \delta=0$$

$$H_1:\ \delta\neq 0,$$

in which the test statistic W_δ is,

$$W_\delta = \frac{\hat{\delta}}{\sqrt{\mathrm{Var}(\hat{\delta})}} \quad \text{or} \quad \frac{\hat{\delta}/\sigma a}{\sqrt{\mathrm{Var}(\hat{\delta}/\sigma_a)}} \tag{2-47}$$

For a preselected type I level, α_2, we accept H_0 when $|W_\delta| < t_{\alpha_2/2,n_1+n_2}$. In Equation (2-47), we see the ability to detect a standardized magnitude of the program utility is dependent upon the behavior of $\mathrm{Var}(\hat{\delta}/\sigma_a)$ with variations in the pre and post intervention

CONTINUED

1 OF 8

sample sizes n_1 and n_2.

For stationary processes,

$$\lim_{n_2 \to \infty} E[\mathrm{Var}(\hat{\delta}/\sigma_a)] = 1/M,$$

where M is a constant that is determined by the model parameters as well as n_1. For example in the case of the (1,0,1)MCI model,

$$M = [(\frac{1-\phi_1}{1-\theta_1})^2(n_1 - 2\frac{\theta_1(1-\theta_1^{n_1})}{1-\theta_1} + \frac{\theta_1^2(1-\theta_1^{2n_1})}{1-\theta_1^2}] + (\frac{1-\psi_1}{1-\gamma_1})^2(\frac{-2\gamma_1-1}{1-\gamma_1^2})$$

$$+ 2(\frac{1-\phi_1}{1-\theta_1})(1-\theta_1^{n_1})[\frac{\gamma_1(1-\psi_1)}{(1-\gamma_1)(1-\gamma_1^2)} - \frac{\gamma_1}{(1-\gamma_1^2)}(1 + (\frac{1-\psi_1}{1-\gamma_1})\gamma_1)]$$

$$- 2\frac{\gamma_1(1-\psi)}{(1-\gamma_1^2)(1-\gamma_1)} + (\frac{1-\phi_1}{1-\theta_1})^2(1-\theta_1^{n_1})^2(\frac{\gamma_1}{1-\gamma_1^2}) - 1$$

$$+ 2\frac{(1-\psi_1)(\psi_1-\gamma_1)\gamma_1}{(1-\gamma_1)^3} + 2\psi_1 - 2(\frac{1-\psi_1}{1-\gamma_1})^2\frac{\gamma_1^3-1}{(1-\gamma_1)(1-\gamma_1^2)}$$

$$+ (\frac{\psi_1-\gamma_1}{1-\gamma_1})^2(\frac{\gamma_1^2}{1-\gamma_1^2}). \qquad (2\text{-}48)$$

Note that,

$$\lim_{n_2\to\infty} E[Var(\hat{\delta}/\sigma_a)] = \lim_{n_2\to\infty} E[Var(\hat{u}/\sigma_a)] > 0$$

when $n_1 < \infty$ and

$$\lim_{n_1\to\infty} [\lim_{n_2\to\infty} E[Var(\hat{\delta}/\sigma_a)]] = \lim_{n_1\to\infty} [\lim_{n_2\to\infty} [Var(\hat{u}/\sigma_a)]] = 0.$$

Since for a given value of n_1, $E(Var(\hat{\delta}/\sigma_a))$ doesn't go to zero as n_2 increases there is a lower bound of $|\delta|$, below which a given magnitude of δ cannot be statistically detected even as n_2 goes to infinity. That is when $|\delta/\sigma_a|$ falls in the region $[0, t_{\alpha_2/2,\infty}\sqrt{\lim_{n_2\to\infty} E(Var(\hat{\delta}/\sigma_a))}]$ we cannot detect the significance of the mean-shift. Alternatively from the width of the $(1-\alpha_2)100\%$ confidence interval for μ/σ_a which is $2t_{\alpha_2/2,\infty}\sqrt{\lim_{n_2\to\infty} E\ Var(\hat{\delta}/\sigma_a)}$, $|(\mu+\delta)/\sigma_a|$ falls in the $(1-\alpha_2)100\%$ confidence interval for μ, therefore the magnitude of mean shift cannot be detected at the α_2 significance level.

For nonstationary processes, $d \geq 1$, $\mu = 0$ and $Var(\hat{\mu}) = 0$. In this case where only δ is estimated $\lim_{n_2\to\infty} E[Var(\hat{\delta}/\alpha_a)] = 0$. Therefore any mean shift magnitude, $\delta \neq 0$, can be detected to be statistically significant by increasing the postintervention observations n_2.

Example 4

In the following numerical example, the effect of the preintervention correlative structure on the ability to detect a minimum threshold shift in the process mean for stationary processes and the magnitude of the postintervention sample size n_2 required to determine a given change in process level is illustrated. Table 2.5 displays numerical values of the $\lim_{n_2\to\infty} E[Var(\hat{\delta}/\sigma_a)]$ for $d=0$, $n_1 = 10$ and $-1.0 < \psi_1, \gamma_1 < 1.0$. For a basis of comparison of the effect of varying the preintervention parameters ϕ_1, θ_1 on the $\lim_{n_2\to\infty} E[Var(\hat{\delta}/\sigma_a)]$, entry a contains the case where $\phi_1 = \theta_1 = 0.0$. Entries b-e are for $\phi_1 = 0.4, -0.4$ and $\theta_1 = 0.4$, respectively for all other conditions in entry a. Within each entry a-e, the closer

Table 2.5. Computed Values of $\underset{n_2 \to \infty}{\text{Limit}}\ E[\mathrm{Var}(\delta/\sigma_a)]$ Given $n_1 = 10$

$\phi_1 = 0.0$, $\theta_1 = 0.0$ ($n_1 = 10$)

γ_1 \ ψ_1	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	0.022	0.023	0.025	0.026	0.026	0.027	0.026	0.025	0.024	0.022	0.021
-0.8	0.059	0.061	0.063	0.065	0.066	0.065	0.064	0.062	0.059	0.056	0.054
-0.6	0.081	0.083	0.085	0.086	0.086	0.086	0.084	0.082	0.079	0.076	0.072
-0.4	0.090	0.092	0.094	0.095	0.095	0.095	0.093	0.091	0.088	0.085	0.082
-0.2	0.093	0.095	0.097	0.098	0.099	0.099	0.098	0.096	0.094	0.090	0.088
0.0	0.092	0.094	0.097	0.098	0.100	0.100	0.100	0.098	0.097	0.094	0.092
0.2	0.087	0.089	0.093	0.095	0.097	0.099	0.099	0.099	0.098	0.096	0.094
0.4	0.077	0.081	0.085	0.089	0.092	0.095	0.097	0.098	0.098	0.097	0.096
0.6	0.063	0.066	0.071	0.076	0.081	0.086	0.090	0.093	0.096	0.097	0.098
0.8	0.039	0.043	0.047	0.053	0.059	0.065	0.072	0.080	0.087	0.093	0.098
0.95	0.012	0.013	0.016	0.018	0.022	0.027	0.033	0.041	0.054	0.072	0.092

(a)

$\phi_1 = 0.4$, $\theta_1 = 0.0$ ($n_1 = 10$)

γ_1 \ ψ_1	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	0.024	0.027	0.030	0.034	0.037	0.039	0.041	0.041	0.040	0.038	0.035
-0.8	0.090	0.097	0.107	0.116	0.124	0.128	0.129	0.127	0.121	0.113	0.106
-0.6	0.158	0.169	0.182	0.193	0.200	0.203	0.200	0.193	0.182	0.169	0.158
-0.4	0.202	0.215	0.229	0.240	0.246	0.247	0.242	0.232	0.218	0.202	0.189
-0.2	0.223	0.236	0.252	0.263	0.270	0.270	0.265	0.254	0.240	0.222	0.209
0.0	0.222	0.236	0.253	0.266	0.275	0.278	0.275	0.266	0.253	0.236	0.222
0.2	0.202	0.216	0.234	0.250	0.263	0.270	0.272	0.268	0.258	0.244	0.231
0.4	0.166	0.179	0.198	0.216	0.233	0.247	0.256	0.259	0.256	0.247	0.237
0.6	0.118	0.129	0.146	0.165	0.184	0.203	0.220	0.234	0.242	0.242	0.238
0.8	0.062	0.069	0.080	0.093	0.109	0.128	0.150	0.174	0.198	0.218	0.228
0.95	0.016	0.018	0.021	0.026	0.031	0.039	0.051	0.067	0.091	0.128	0.167

(b)

Table 2.5. (Cont'd)

$\phi_1 = -0.4$, $\theta_1 = 0.0$ ($n_1 = 10$)

γ_1 \ ψ_1	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	0.018	0.018	0.018	0.018	0.018	0.018	0.017	0.016	0.015	0.014	0.014
-0.8	0.038	0.038	0.038	0.038	0.038	0.037	0.037	0.036	0.034	0.033	0.032
-0.6	0.046	0.046	0.047	0.047	0.046	0.046	0.045	0.044	0.043	0.042	0.041
-0.4	0.049	0.049	0.050	0.050	0.050	0.049	0.049	0.048	0.047	0.046	0.045
-0.2	0.049	0.050	0.050	0.051	0.051	0.051	0.050	0.050	0.049	0.048	0.047
0.0	0.049	0.049	0.050	0.051	0.051	0.051	0.051	0.051	0.050	0.049	0.049
0.2	0.047	0.048	0.049	0.050	0.050	0.051	0.051	0.051	0.051	0.050	0.050
0.4	0.043	0.045	0.046	0.047	0.048	0.049	0.050	0.050	0.051	0.051	0.051
0.6	0.037	0.039	0.041	0.042	0.044	0.046	0.047	0.049	0.050	0.050	0.051
0.8	0.026	0.028	0.030	0.032	0.035	0.038	0.040	0.043	0.050	0.049	0.050
0.95	0.009	0.010	0.011	0.013	0.015	0.018	0.021	0.025	0.031	0.038	0.045

(c)

$\phi_1 = 0.0$, $\theta_1 = 0.4$ ($n_1 = 10$)

γ_1 \ ψ_1	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	0.015	0.016	0.016	0.015	0.015	0.015	0.014	0.013	0.012	0.012	0.011
-0.8	0.031	0.032	0.032	0.031	0.031	0.030	0.029	0.029	0.028	0.026	0.026
-0.6	0.037	0.038	0.038	0.037	0.037	0.037	0.036	0.035	0.034	0.034	0.033
-0.4	0.039	0.040	0.040	0.040	0.040	0.040	0.039	0.038	0.038	0.037	0.036
-0.2	0.040	0.030	0.040	0.040	0.040	0.040	0.040	0.040	0.039	0.039	0.038
0.0	0.039	0.040	0.040	0.040	0.041	0.041	0.041	0.040	0.040	0.040	0.039
0.2	0.038	0.038	0.040	0.040	0.040	0.040	0.041	0.041	0.041	0.040	0.040
0.4	0.035	0.036	0.037	0.038	0.039	0.039	0.040	0.040	0.040	0.041	0.041
0.6	0.031	0.032	0.033	0.034	0.036	0.037	0.038	0.039	0.040	0.040	0.041
0.8	0.022	0.023	0.025	0.026	0.028	0.030	0.032	0.034	0.036	0.038	0.040
9.5	0.008	0.009	0.010	0.011	0.013	0.015	0.017	0.020	0.024	0.029	0.034

(d)

Table 2.5. (Cont'd)

$\phi_1 = 0.0,\ \theta_1 = -0.4 \quad (n_1 = 10)$

γ_1 \ ψ_1	-0.95	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	0.95
-0.95	0.024	0.026	0.028	0.031	0.033	0.035	0.035	0.035	0.034	0.032	0.030
-0.8	0.078	0.083	0.089	0.095	0.099	0.101	0.100	0.098	0.093	0.088	0.083
-0.6	0.123	0.129	0.136	0.142	0.145	0.145	0.143	0.138	0.132	0.124	0.117
-0.4	0.148	0.154	0.161	0.166	0.168	0.168	0.165	0.159	0.152	0.144	0.137
-0.2	0.158	0.164	0.171	0.176	0.179	0.179	0.176	0.171	0.164	0.155	0.148
0.0	0.156	0.163	0.171	0.177	0.181	0.182	0.181	0.177	0.171	0.163	0.156
0.2	0.145	0.152	0.161	0.169	0.175	0.179	0.180	0.178	0.174	0.168	0.162
0.4	0.124	0.132	0.142	0.152	0.161	0.168	0.172	0.174	0.174	0.170	0.166
0.6	0.094	0.101	0.112	0.123	0.134	0.145	0.155	0.162	0.167	0.169	0.168
0.8	0.053	0.058	0.067	0.076	0.088	0.101	0.115	0.130	0.145	0.158	0.165
0.95	0.015	0.016	0.019	0.023	0.028	0.035	0.044	0.058	0.077	0.107	0.139

(e)

the parameters to zero the larger the $\lim_{n_2\to\infty} E[\mathrm{Var}(\hat{\delta}/\sigma_a)]$. Also as θ_1 decreases from zero or ϕ_1 increases from zero the $\lim_{n_2\to\infty} E[\mathrm{Var}(\hat{\delta}/\sigma_a)]$ increases. Conversely as θ_1 increases from zero and ϕ_1 decreases from zero the $\lim_{n_2\to\infty} E[\mathrm{Var}(\hat{\delta}/\sigma_a)]$ decreases. The smaller this value is, the greater the ability to detect a small magnitude of the mean shift parameter. Thus, if the time series is smooth (e.g., described by moderate to long length runs above and below the mean, $\phi_1 > 0$, $\theta_1 < 0$) small magnitude changes in the mean shift parameter are harder to detect than when the series is oscillatory (e.g., described by a high frequency component such as runs above or below the mean of short duration).

Table 2-6(a) illustrates the sample size of n_1 or n_2 need to detect a given standardized shift in process level, δ/σ_a, for the stationary (1,0,1)MCI models for $\alpha = 0.05$, $\theta_1 = 0.975$ and $\gamma_1 = 0.816$. For a given column entry, n_1, and row entry δ/σ_a, the tabled entry is the minimum value of n_2 such that $|W| > t_{\alpha_2/2, n_1+n_2-2}$. Some entries contain ∞ which denotes the inability to detect the corresponding δ/σ_a for the combination of (n_1,n_2). Table 2-6(b) illustrates the sample size requirement for the nonstationary model (1,1,1)MCI under the same parameter values and hypothesis test specification of Table 2-6(a). Since when d=1 the $\lim_{n_2\to\infty} E[\mathrm{Var}(\delta/\sigma_a)]$ converges to zero, all non zero magnitudes of δ/σ_a will be detectable for large enough n_2, as shown for the parameter set illustrated. In Table 2.6 for a fixed preintervention history n_1 the larger the magnitude of δ/σ_a the smaller the postintervention sample size n_2. Similarly for a fixed δ/σ_a as n_1 increases n_2 decreases

Table 2.6. Sample Sizes for Detecting δ/σ_a

$d = 0$

$\lvert\delta/\sigma_a\rvert$ \ n_1	5	10	20	30	40	50	75	100	200	1000
0.05	∞	∞	163	82	70	66	62	61	60	59
0.10	∞	106	26	23	22	21	21	21	20	20
0.20	∞	13	10	10	10	9	9	9	9	9
0.30	15	8	7	7	7	6	6	6	6	6
0.40	9	6	5	5	5	5	5	5	5	5
0.50	7	5	4	4	4	4	4	4	4	4
0.60	5	4	4	4	4	4	4	4	4	4
0.70	4	4	3	3	3	3	3	3	3	3
0.80	4	3	3	3	3	3	3	3	3	3
0.90	3	3	3	3	3	3	3	3	3	3
1.00	3	3	3	2	2	2	2	2	2	2
1.50	2	2	2	2	2	2	2	2	2	2
2.00	2	2	2	2	2	1	1	1	1	1
2.50	1	1	1	1	1	1	1	1	1	1

(a) (0,0,1)MCI with $\theta_1 = 0.98$, $\gamma_1 = 0.82$

$d \geq 1$

$\lvert\delta/\sigma_a\rvert$	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.50	2.00
n_2	59	20	9	6	5	4	4	3	3	3	2	2	1

(b) (0,1,1)MCI with $\theta_1 = 0.98$, $\gamma_1 = 0.82$

except when δ/σ_a is large and n_2 remains constant. For nonstationary processes the postintervention sample size is seen to be independent of the size of the preintervention sample size. Lastly, under the same conditions (comparing corresponding tabled values for d=0 or d=1) a nonstationary process is seen to be more sensitive in detecting δ/σ_a since fewer postintervention observations n_2 are required.

2.9 Optimal Design of Interrupted Time Series Experiments

In evaluating experimental programs in field settings, the cost of experiments is a paramount importance. Further, the ability to evaluate whether the program exhibited its intended impact is lost if the appropriate sample sizes in the pre and post period are not selected. Many field experiments are often evaluated to have had no impact or non-reproducible impact when geographical locations are changed. These outcomes are not necessarily program related but are design related in that sufficient data is not recorded. Similarly, the cost of field experiments can be reduced by specification of required sample sizes required to draw the desired level of inference, with regard to potential changes in covariance and the magnitude of shift in the process level, as related to an intervention programs goal.

The optimal design for the multiconsequence intervention model is the solution to the nonlinear programming problem:

$$\text{Min:}\quad c_1n_1 + c_2n_2$$

$$\text{s.t.:}\quad E(W_{\underline{\beta}}|n_1,n_2,\underline{\beta}) > \chi^2_{1-\alpha_1,s}$$

$$E(W_\delta) > t_{1-\alpha_2/2,n_1+n_2-(p+q)}$$

where c_1 and c_2 are the costs for obtaining pre and post intervention observations respectively and $(W_{\underline{\beta}}|n_1,n_2,\underline{\beta})$ and W_δ are the test statistics needed to detect a given magnitude change in covariance and process level, respectively. The constraints are nonlinear and convex.

<u>Example 5</u>

Suppose the interrupted time series model is known to be a (0,0,1)MCI form with $\theta_1 = 0.975$ and $\gamma_1 = 0.816$, for which the sample sizes n_1 and n_2 are desired to be able to determine this magnitude change in covariance and simultaneously detect shifts in mean level as small as $\delta/\sigma_a = 0.05$. Figure 2.6 graphically illustrated the optimal design subject to (a) the mean-shift constraint only, (b) the covariance-change constraints. When $c_1 = c_2$ and only the mean-shift constraint is considered, the minimal cost feasible design ($n_1 = 37$, $n_2 = 74$) is read from the tangential point of the $c_1 = c_2$ cost curve (a) and the mean-shift constraint. If $c_1 = 0.25c_2$ and only the covariance-change constraint is considered, the minimal cost feasible design ($n_1 = 128$, $n_2 = 34$) is read from the tangential point of the $c_1 = 0.25c$ cost curve (b) and the covariance-change constraint. In most situations both mean-shift and covariance-change constraints are considered. In such situations, the active constraint consists of two portions: (1) $n_1 \leq 89$

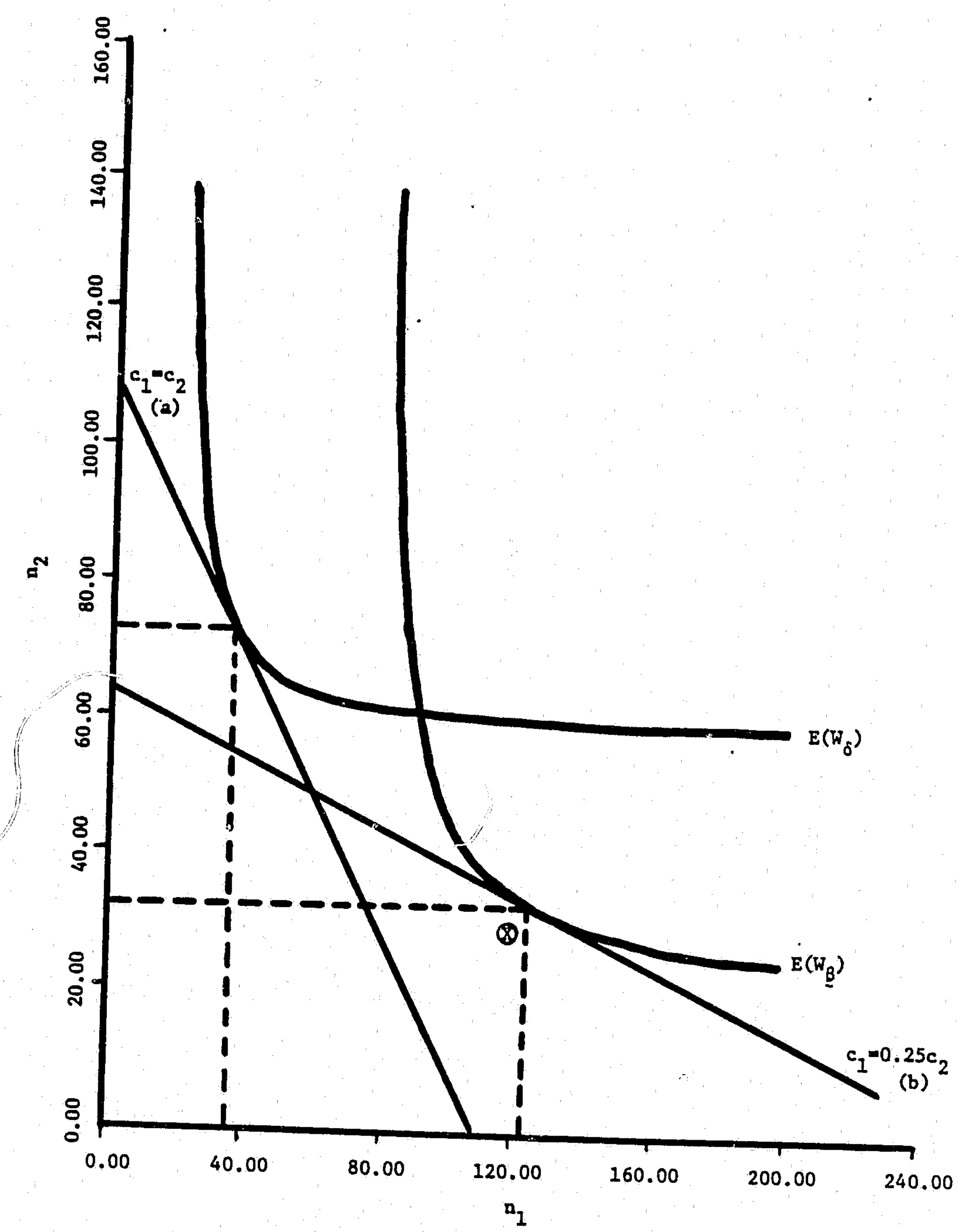


Figure 2.6. Optimal Design of Experiments, (0,0,1)MCI with $\theta_1=0.98$, $\gamma_1=0.82$, $\alpha_1=0.05$, $\alpha_2=0.05$

portion of the covariance-change constraint and (2) $n_1 \geq 90$ portion of the mean-shift constraint. Since the Min|Slope| = 3 for the upper-left portion and Max|Slope| = 1/45 for the lower-right portion, the minimal cost design will occur at (n_1 = 90, n_2 = 63) when $1/45 \leq c_1/c_2 \leq 3$ it will occur at the tangential point of the cost curve and the covariance-change constraint (the upper-left portion of the active constraint) when $c_1/c_2 > 3$, or the optimal design will occur at the tangential point of the cost curve and the mean-shift constraint (the lower-right of the active constraint) when $0 \leq c_1/c_2 < 1/45$.

2.10 Modeling Examples

Two data case studies are developed here to illustrate the different model structures, their associated model building and design considerations discussed in this chapter.

2.10.1 A Direct Stimulus Response Structure: Talk Out Data

The data reported by Hall et al. [1971] records the daily number of "talk outs" of twenty-seven pupils in the second grade of an all-black urban poverty area school for a total time period of forty days. The first twenty days were denoted as the preintervention history before the commencement of the intervention. Beginning on the twenty first day, the teacher initiated a program of systematic praise for not talking out. Figure 2.7(a) contains the time series data.

From the preintervention history $t \leq 20$ or $n_1 = 20$, applying the usual univariate time series modeling of identification, estimation and diagnostic checking of Box and Jenkins [1970], resulted in an ARIMA(0,0,1) model with $\hat{\theta}_1 = -0.195$,

$$Z_t' = (1+0.195B)a_t, \quad t \leq 20.$$

Confusing here

The direct stimulus response structure was postulated apriori since the intervention consisted of applying the stimulus (praise) directly to the subjects (students) whose response ("talk outs") was the attribute of interest to evaluate the worth of the program. Figure 2.7(b) contains the results of the identification procedure for the form of the mean shift function $\delta(t)$ using $n_1 = n_t$ and $n_2 = 1$ sequentially for $20 \leq t \leq 39$. Confirmation of the initially postulated direct stimulus response structure is seen here since $\delta(t)$ after an initial transient due to the "learning curve" associated with the intervention has little variation about the steady state gain level. Therefore the mean shift function $\delta(t) = K(t)\delta$ is specified by,

$$K(t) = \begin{cases} 7/15 & t=21 \\ 11/15 & t=22 \\ 1 & t \geq 23. \end{cases}$$

The (0,0,1)MCI model was fit with parameters $\hat{\mu} = 19.35$, $\hat{\delta} = -15.18$, $\hat{\theta}_1 = -0.175$ and $\hat{\gamma}_1 = -0.601$. Diagnostic checking of the residuals proved adequate. However, testing the hypothesis that $\theta_1 = \gamma_1$ resulted in favoring the alternate hypothesis for all values of $\alpha < 0.2869$. The test of the intrinsic utility of the intervention program, ~~e.g.~~ i.e. H_0: $\delta=0$, was not preferred for all values of $\alpha_2 > 0.001$. Thus the final model is the single consequence model,

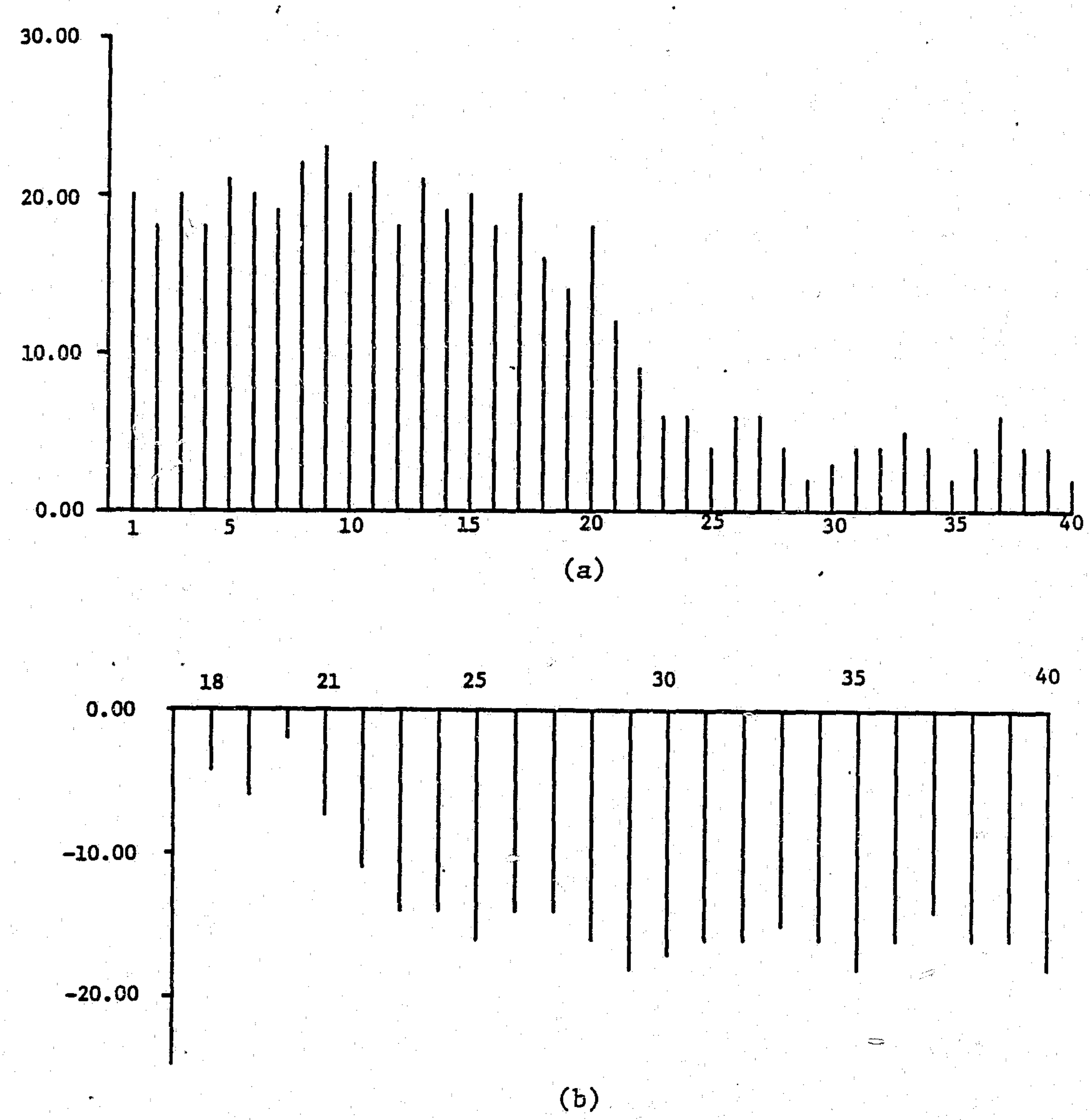


Figure 2.7. Talk Out

(a) Time Series Data

(b) Mean Shift Function

$$Z_t = 19.35 + 0.23a_{t-1} + a_t \qquad t \leq 20$$

$$Z_t = 19.35 - K(t)(15.13) + 0.23a_{t-1} + a_t \qquad t = 21,22$$

$$Z_t = 4.22 + 0.23a_{t-1} + a_t \qquad t \geq 23,$$

which results in the conclusion that the intervention program reduces "talk outs" by 15.13 per day in steady state.

For the sample sizes used in the "talkout" data, $n_1 = n_2 = 20$, there was not sufficient power to detect the estimated change in covariance. For the estimated covariance change with $n_1 = 20$, 23 and 58 postintervention observations for $\alpha = 0.15$ and 0.05 respectively are needed to assess significance. For the mean shift detection, however with $n_1 = n_2 = 20$ and $\alpha = 0.05$ magnitudes of $\delta/\sigma_a = 0.78$ can be detected. Since $\hat{\delta}/\hat{\sigma}_a = 9.10$ the sample size is sufficient. If the change in covariance is real and was not detected because sufficient data was not available, using the single consequence model form for the talk out data would represent a model misspecification in the parameters. For the estimated parameter values, the corresponding h value (Equation (2-37)) is 0.7789 which would represent an overestimation of the significance of δ/σ_a by 28.4% or alternatively, using the single consequence model with $\alpha = 0.05$ would effectively be $\alpha = 0.126$ if the multiconsequence form was needed. Since the estimated magnitude of δ was very large in this data set these facts do not alter the policy inference in the final model described.

2.10.2 An Indirect Stimulus Response Structure; Gun Control

The data reported by Deutsch and Alt [1977] records the number of reported assaults with a gun in Boston by month. Figure 2-8(a) illustrates this data for the period from January 1966 to October 1976. In April 1975, the State of Massachusetts formally put into operation a gun control law that mandates a one-year minimum sentence on conviction of carrying a firearm without a special license, thereby eliminating judicial descretion in sanctions. In order to assess policy implications of gun control legislation, the impact of the enactment of the legislative intervention in reducing gun related assaults were desired.

Although the gun control law was formally implemented in April 1975, there were several associated activities with this intervention strategy for several months prior to this date. Thus, the exact time point where the impact might first be measured was uncertain. Therefore, the preintervention data for the period from January 1966 to September 1974 was used to construct the preintervention model. Applying the usual Box and Jenkins model building procedure (Box and Jenkins [1970]) resulted in an ARIMA(0,1,1) × $(0,1,1)_{12}$ model,

t = 2

$$(1-B)(1-B^{12})z_t = (1-\theta_1 B)(1-\theta_{12}B^{12})a_t$$

with $\hat{\theta}_1 = 0.83$ and $\hat{\theta}_{12} = 0.78$.

The indirect-stimulus response structure was chosen since the intervention program was intended to modify the ~~eclectric~~ environmental process from which the behavior modification of a subset of the population is monitored. Figure 2.8(b) contains the results of the identification procedure for the form of the mean shift function $\delta(t)$ using

$n_1 = n_t$ and $n_2 = 1$ sequentially for $106 \leq t \leq 141$. From the mean shift function plot, the impact of the intervention program is seen to start in February (perhaps due to the planned massive publicity campaign in early 1975) and to continue through to post intervention period. The configuration of the mean shift function prior to reaching a steady state and the variation about this form confirms the environmental influence.

The mean shift function for the indirect stimulus response structure is $\delta(t) = f(K(t))$ where $K(t) = (1-\theta_1 B)(1-\theta_{12}B^{12})$. As seen from Figure 2.8(b), a plot of $\delta(t)$ vs. $K(t)$ would indicate a slope and intercept term are needed to describe this relationship. Thus,

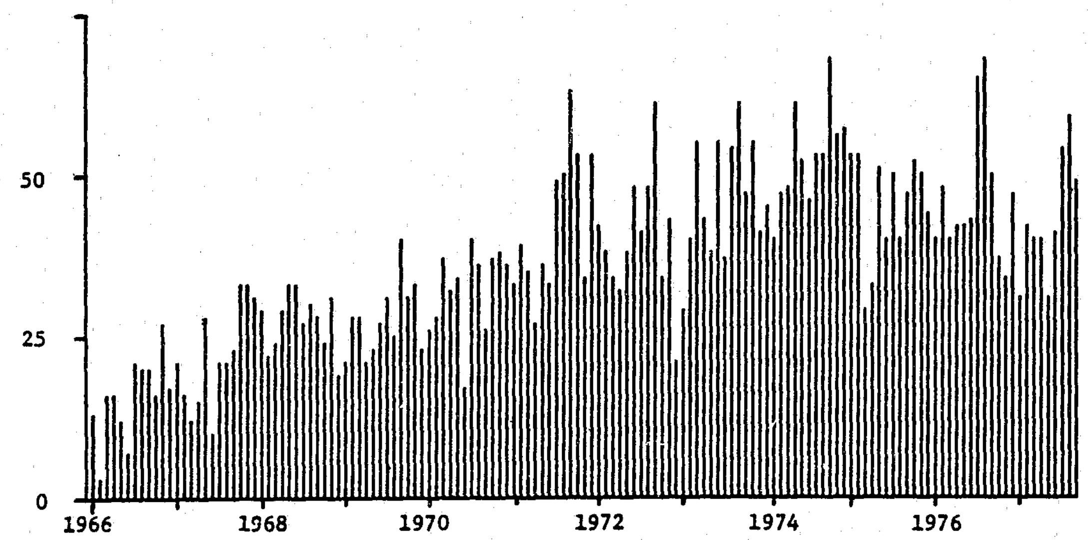
$$\delta(t) = \delta_0 + \delta_1 K(t) \qquad t \geq 110$$

and

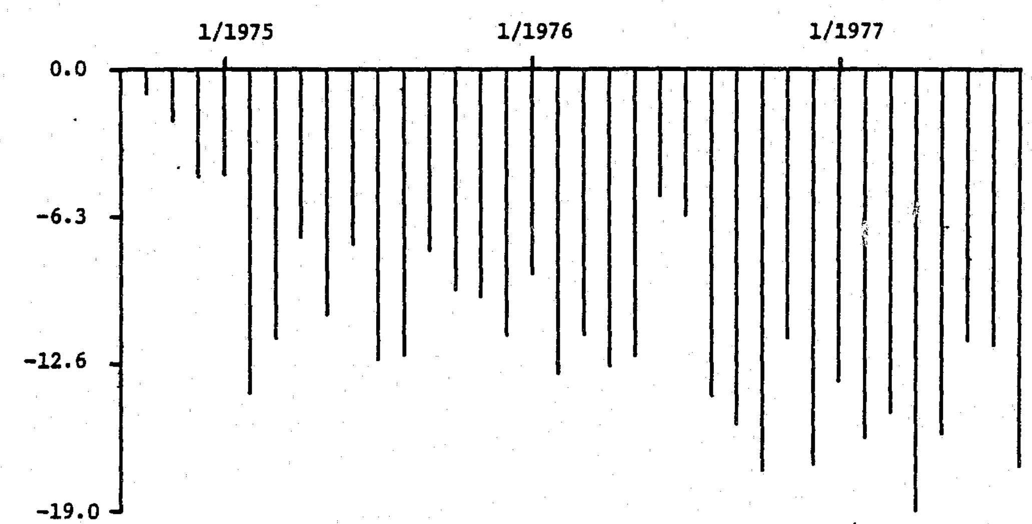
$$\delta(t) = [\delta_0 + \delta_1 K(t)]\xi_t$$

where

$$\xi_t = \begin{cases} 0 & t \leq 110 \\ 1 & t \geq 111. \end{cases}$$



(a)



(b)

Figure 2.8. Assault with a Gun

(a) Time Series Data

(b) Mean Shift Function

The intrinsic program utility δ occurs when $K(t) = 1$, ~~e.g.~~ i.e. since, $\delta(t) = K(t)\delta = \delta_0 + \delta_1 K(t)$, $\delta = \delta_0 + \delta_1$. For all other values of $K(t)$ the environmental influence will mask the intrinsic program value $((\delta(t) - \delta)$ is the masking effect of the environment).

The maximum likelihood estimates for the parameters of the $(0,1,1) \times (0,1,1)_{12}$MCI model are $\hat{\delta}_0 = -14.79$, $\hat{\delta}_1 = -3.89$, $\hat{\theta}_1 = 0.98$ and $\hat{\gamma}_1 = 0.82$ Diagnostic checking of the residuals proved adequate. In testing the hypothesis of no change in covariance, the null hypothesis H_0: $\theta_1 = \gamma_1$ is not preferred for all $\alpha > 0.06$. Similarly the hypothesis H_0: $\delta_i = 0$, $i = 0,1$ is not preferred for all $\alpha > 0.001$. Thus the final model is,

$$(1-B)(1-B^{12})Z_t = (1-0.98B)(1-0.78B^{12})a_t \qquad t \leq 110$$

$$(1-B)(1-B^{12})Z_t = -14.79 + (-3.89K(t)) + (1-.82B)(1-.78B^{12})a_t \qquad t > 110$$

Hard to follow

Thus the result of the legislative intervention of gun control was to reduce the observed number of reported gun assaults between 19 to 15 per month. Further this impact is seen to manifest itself in steadystate through the postintervention period. The steady state decrease is approximately 15 per month and the intrinsic utility of the program being approximately 19 per month.

The sample size used in this analysis was n_1=110 and n_2=31 which is denoted by the circled X in Figure 2.6. This point is seen to be just below the $\alpha = 0.05$ constraint for $E(W_\beta)$. (The hypothesis test for change in covariance was significant for values of $\alpha > 0.06$). The program goal was to reduce the number of assaults with a gun by 5%. At the time of program the number of reported assaults per month was 52, therefore $\delta/\sigma_a \simeq 0.29$. For $n_1 = 110$ only 4 postintervention observations are needed. Since 31 were used to statistically detect the change in covariance a feasible design was employed. However, if fewer postintervention data points were used and the existing change in covariance could not be statistically detected (resulting in the reduced single consequence model), the estimate of δ would have been biased. The bias in δ would be 0.30 (D=1.30, Equation 2-36)), thus there would have been an overestimation of 30%. Also, the ratio of the variance of the standardized model (Equation 2-37) would be 1.46, resulting in D/h = 0.89 or an underestimation of the level of significance of δ. If the change in covariance was not statistically detected because of too short a post-intervention period (n_2) the corresponding single consequence model would have resulted in a 30% overestimation of the monthly reductions in assaults with a gun. Similarly, for n_1 fixed at 110, 59 postintervention data points would be needed to statistically detect values of $\delta/\sigma_a = 0.05$. In this case $\delta/\sigma_a = 1.79$. However, if the program in-intervention resulted in values of $0.05 < \delta/\sigma_a < 0.10$, which should result in favorable policy implications, the sample size would not have been large enough and therefore incorrect policy inferences would be drawn.

CHAPTER III

THE SPACE-TIME INTERVENTION MODELING

In Chapter II multiconsequence intervention modeling was studied in detail. The multiconsequence intervention modeling is apropriate for observations at one location. In this chapter, space-time intervention modeling, which is appropriate for modeling an intervention process at multiple locations, will be studied. The space-time intervention modeling procedure shares the same modeling strategies as the multiconsequence intervention modeling, i.e., the dynamic components identification procedure and the three-stage iterative model building scheme. The space-time intervention model class is an adaptation of the model formulation of the space-time model class that has been developed by Deutsch and Pfeifer [1980a, 1980b, 1981].

In the next section the space-time intervention model class $(\mathrm{STARMA})I_m$ that allows for both environmental and non-environmental influence on the intervention is described for the single intervention processs as well as the multiple intervention process. The physical properties of the space-time intervention model are discussed in section 3.2. Here four elementary diffusion mechanisms; translation, domain-change, growth, and contraction are described and ~~corresponded~~ to the $(\mathrm{STARMA})I_m$ model class. Simulations of the diffusion processes described by low order $(\mathrm{STARIMA})I_m$ models are presented to illustrate the physical characteristics of diffusion speed, amplitude and influenced area in section 3.3. The intervention model building procedure, that includes pre-intervention model building, dynamic component identification, the coupling of these components and diagnostic checking is given in section 3.4. In section 3.5, the M.L. estimators of the mean-shift function are developed first for the pre-intervention noise model parameters known situation, and then is extended to the case where all parametrs are unknown. A case study of the air pollution quality as measured by ambient carbon monoxide levels in Los Angeles is contained in section 3.6.

3.1 The Space-Time Intervention Model

A linear stationary time series data-generating process can be expressed in transfer function form as:

$$\underset{\sim}{Z}_t = \underset{\sim}{\mu} + T_e(B)\ \underset{\sim}{\varepsilon}_t, \quad t=1,2,3,\ldots,n$$

where $\underset{\sim}{Z}_t$ is the observation or output vector, $\underset{\sim}{\mu}$ is the mean vector, $\underset{\sim}{\varepsilon}_t$ is the residual or input vector assumed to be $NID(0, \sigma^2 \underset{\sim}{I})$, all of which is of dimension (LN x 1) where LN is the number of locations and $T_e(B)$ is the transfer function for environment e which couples the system input and output (Figure 3-1(a)).

The intervention at time τ, $t < \tau < t+1$ is thought of as "switching on", $\xi_t = 0,1$ causing an additional potential contribution $\underset{\sim}{\delta}$ associated with the utility of the intervention program. When a system intervention occurs this potential shift can cause a modification in the data generation process in two ways. The program utility modifica-

tion to the output is either affected by the environmental process (Figure 3-1(b)) or unaffected by the environment (Figure 3-1(c)).

For a single region, univariate time series, Box and Tiao [1975] have proposed intervention models of the characteristics exhibited in Figure 3-1(c), in which the intervention component is independent of the environmental data-generating process. For the single location intervention model of Box and Tiao [1975] the observation vector is decomposed into two components:

$$Z_t = D_t + N_t$$

where D_t, the dynamic component, monitors the mean level changes of the pre-intervention environmental noise model N_t. When this eclectic noise model influences the measurable changes in process level, as in Figure 3-1(b), this estimated change~~s~~ in process level confounds the intrinsic utility of the intervention and the environmental noise process. The more general framework to allow separation of these components and thus unbiased estimation of an intervention's intrinsic impact is that described in Figure 3-1(d). The transfer function $T_\xi(B)$ will not be a function of the environmental noise process parameters when there is no environmental influence on the intervention. The general structure of Figure 3-1(d) is used throughout the balance of this chapter in regard to LN site systems.

We denote the variable for a single intervention, $\underset{\sim}{\xi}(t)$, as the product of the program utility $\underset{\sim}{\delta}$ and the switch variable ξ_t, which

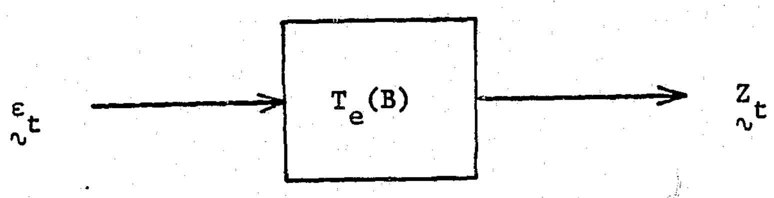


Figure 3-1(a). Non-intervention Data Generating Process.

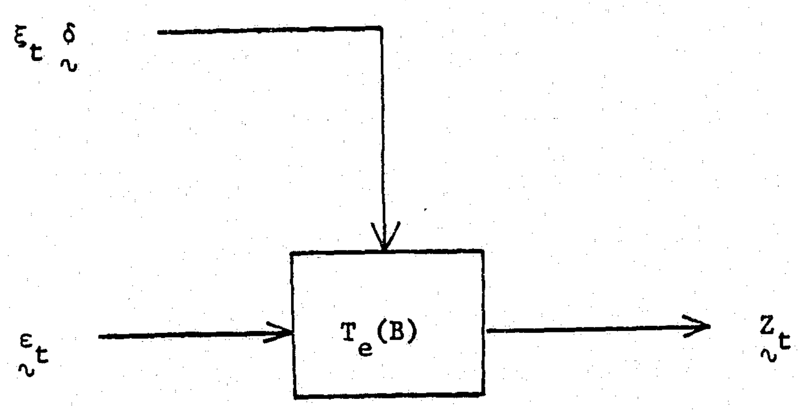


Figure 3-1(b). Intervention Data Generating Process (Effected by the Environment).

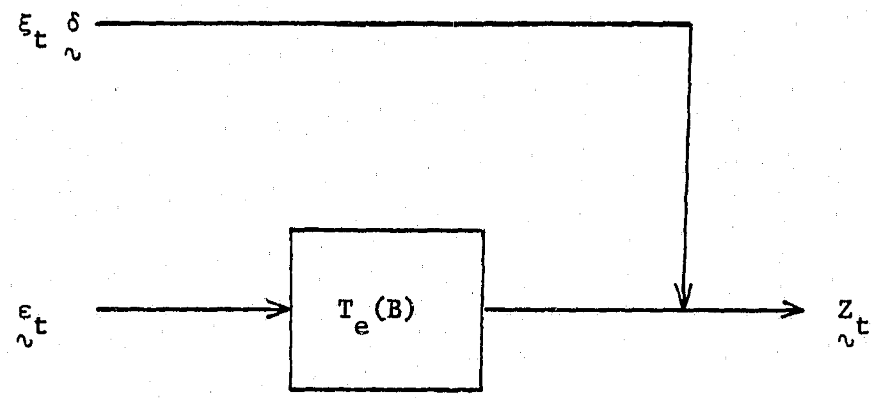


Figure 3-1(c). Intervention Data Generating Process (Uneffected by the Environment).

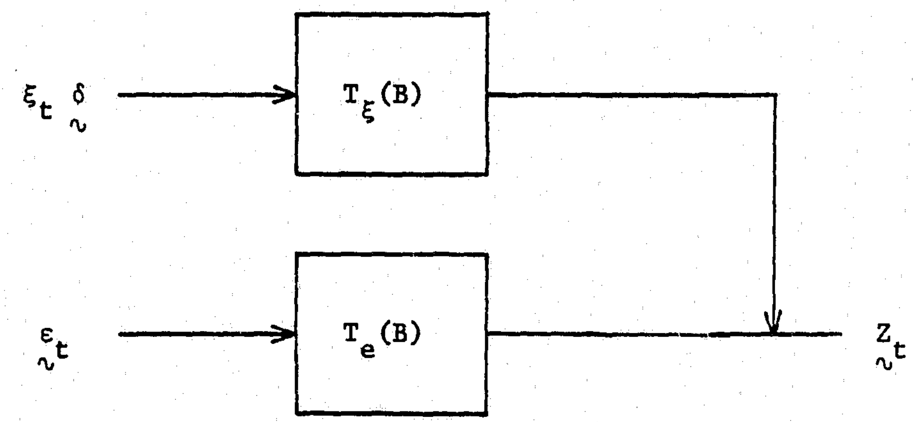


Figure 3-1(d). Intervention Data Generating Process with $T_\xi(B)$ as the Intervention Transfer Function.

either takes the form of a pulse or step function. The pulse function is defined as:

$$P_t^{n_1}(T_p) = \begin{cases} 1 & n_1+1 \le t \le n_1+T_p \\ 0 & \text{elsewhere} \end{cases}$$

Thus, the pulse function corresponds to the situation in which an intervention "switches on" for a period of length T_p, after the n_1+1^{st} observation. The step function is defined as an infinite period intervention occurring between $t = n_1$, and $t = n_1+1$;

$$S_t^{n_1} = \begin{cases} 1 & t \ge n_1+1 \\ 0 & t \le n_1 \end{cases}$$

The space-time single intervention model formulation for LN regions is;

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\nabla^d(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \left|(1-I_m)\underset{\sim}{\Phi}_{p,\lambda}(B)\nabla^d + I_m\underset{\sim}{\Theta}_{q,m}(B)\right|\underset{\sim}{\xi}(t) \qquad (3\text{-}1)$$

$$+ \underset{\sim}{\Theta}_{q,m}(B)\underset{\sim}{\varepsilon}_t$$

where

$$I_m = \begin{cases} 1 & \text{realized intervention effected by environmental process} \\ 0 & \text{realized intervention not effected by environmental process,} \end{cases}$$

$\underset{\sim}{Z}_t$ is an LN x 1 vector of observations at time t,

$$\underset{\sim}{\Phi}_{p,\lambda}(B) = I - \sum_{K=1}^{p}\sum_{\ell=0}^{\lambda_K}\phi_{K\ell}W^{(\ell)}B^K,$$

$$\nabla^d = (I-B)^d,$$

$$\underset{\sim}{\Theta}_{q,m}(B) = I - \sum_{K=1}^{q}\sum_{\ell=0}^{m_K}\Theta_{K\ell}W^{(\ell)}B^K,$$

B is an LN x LN backward shift operator such that

$$B^K\underset{\sim}{Z}_t = \underset{\sim}{Z}_{t-K},$$

$\phi_{K\ell}$ is the autoregressive parameter at temporal lag K and spatial lag ℓ,

$\theta_{K\ell}$ is the moving average parameter at temporal lag K and spatial lag ℓ,

$W^{(\ell)}$ is an LN x LN matrix of weights for spatial lag ℓ,

$\underset{\sim}{\xi}(t) = \xi_t\,\underset{\sim}{\delta}$, $\underset{\sim}{\delta}$ is the intrinsic utility of the intervention program,

$\underset{\sim}{\varepsilon}_t$ is the random normally distributed innovation vector at time t with

$$E|\underset{\sim}{\varepsilon}_t| = \underset{\sim}{0}$$

$$E|\underset{\sim}{\varepsilon}_t\,\underset{\sim}{\varepsilon}'_{t+K}| = \begin{cases} G & K=0 \\ 0 & K\ne 0 \end{cases}$$

$$E[\underset{\sim}{Z}_t \underset{\sim}{\varepsilon}'_{t+K}] = 0 \quad \text{for } K > 0$$

Parameter p is the autoregressive order of the model and $\underset{\sim}{\lambda}$ is a vector with components λ_K specifying the spatial order of the Kth autoregressive term. Likewise, q is the moving average order of the model, and $\underset{\sim}{m}$ is the vector of moving average spatial orders. Parameter d is specifying the number of differences needed to induce stationarity in the original series.

3.1.1 An Alternative Representation

The single intervention space-time model, equation (3-1), assuming d=0 without loss of generality, can be expressed as,

$$\underset{\sim}{Z}_t = \underset{\sim}{\mu} + \Lambda_D(B)\ \underset{\sim}{\xi}(t) + \Lambda_N(B)\ \underset{\sim}{\varepsilon}_t \tag{3-2}$$

where

$$\Lambda_D(B) = \Phi_{p,\underset{\sim}{\lambda}}(B)^{-1}[(1-I_m)\Phi_{p,\underset{\sim}{\lambda}}(B) + I_m\Theta_{q,\underset{\sim}{m}}(B)]$$

$$\Lambda_N(B) = \Phi_{p,\underset{\sim}{\lambda}}(B)^{-1}\ \Theta_{q,\underset{\sim}{m}}(B)$$

Thus $\underset{\sim}{Z}_t$ is the summation of two components: the deterministic component and the random component, $\underset{\sim}{D}(t)$ and $\underset{\sim}{N}(t)$ respectively, i.e.,

$$\underset{\sim}{Z}_t = \underset{\sim}{D}(t) + \underset{\sim}{N}(t) \tag{3-3}$$

$$\underset{\sim}{D}(t) = \underset{\sim}{\mu} + \Lambda_D(B)\ \underset{\sim}{\xi}(t) \tag{3-4}$$

$$\underset{\sim}{N}(t) = \Lambda_N(B)\ \underset{\sim}{\varepsilon}_t \tag{3-5}$$

Taking expectation of equation 3-3 yields,

$$E[\underset{\sim}{Z}_t] = \underset{\sim}{D}(t) \tag{3-6}$$

Since

$$E[\underset{\sim}{N}(t)] = \Lambda_N(B)\ E[\underset{\sim}{\varepsilon}_t] = \underset{\sim}{0}$$

Therefore the expectation value of the process observation $\underset{\sim}{Z}_t$ is the deterministic component, $\underset{\sim}{D}(t)$. It should be noted that the intervention variable $\underset{\sim}{\xi}(t)$ only appears in the deterministic component as does the effect of the intervention.

Equation 3-4 can be expressed in recursive form,

$$\Phi_{p,\underset{\sim}{\lambda}}(B)(\underset{\sim}{D}(t)-\underset{\sim}{\mu}) = [(1-I_m)\Phi_{p,\underset{\sim}{\lambda}}(B) + I_m\Theta_{q,\underset{\sim}{m}}(B)]\ \underset{\sim}{\xi}(t). \tag{3-7}$$

The realized mean shift function at time t, $\underset{\sim}{\delta}(t)$, is

$\underset{\sim}{\delta}(t) = \underset{\sim}{D}(t) - \underset{\sim}{\mu}$, thus,

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\ \underset{\sim}{\delta}(t) = \left[(1-I_m)\underset{\sim}{\Phi}_{p,\lambda}(B) + I_m\underset{\sim}{\Theta}_{q,m}(B)\right]\ \underset{\sim}{\xi}(t). \qquad (3\text{-}8)$$

In equation 3-8 if we set $I_m = 1$, the realized intervention is effected by the environment, we have the deterministic component of the intervention model,

$$\underset{\sim}{\Phi}_{p,\lambda}(B)(\underset{\sim}{D}(t)-\underset{\sim}{\mu}) = \underset{\sim}{\Theta}_{q,m}(B)\ \underset{\sim}{\xi}(t),$$

with the realized mean shift function, $\underset{\sim}{\delta}(t)$ of the form

$$\underset{\sim}{\delta}(t) = \underset{\sim}{D}(t) - \underset{\sim}{\mu} = \underset{\sim}{\Phi}_{p,\lambda}(B)^{-1}\underset{\sim}{\Theta}_{q,m}(B)\ \underset{\sim}{\xi}(t)$$

or

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\underset{\sim}{\delta}(t) = \underset{\sim}{\Theta}_{q,m}(B)\ \underset{\sim}{\xi}(t) \qquad (3\text{-}9)$$

Since $\underset{\sim}{\delta}(t)$ for the mixed process is a function of the intervention variable $\underset{\sim}{\xi}(t)$ as well as the environmental noise model parameters, the realized shift-mean doesn't reach its steady state level instantaneously but rather converges to the steady state level. For the $STARMA(\underset{\sim}{q}_m)I_m$ process ($p=0$ and $\underset{\sim}{\Phi}_{p,\lambda}(B) = I$) equation 3-9 reduces to

$$\underset{\sim}{\delta}(t) = \underset{\sim}{\xi}(t) - \sum_{K=1}^{q}\sum_{\ell=0}^{m_K}\Theta_{K\ell}W^{(\ell)}\ \underset{\sim}{\xi}(t-K). \qquad (3\text{-}10)$$

Thus, the steady state level for $\underset{\sim}{\delta}(t)$, i.e. $(I-\sum_{K=1}^{q}\sum_{\ell=0}^{m_K}\Theta_{K\ell}W^{(\ell)}B^K)\underset{\sim}{\xi}(t)$, is reached in q periods. However, for the $STAR(\underset{\sim}{P}_\lambda)I_m$ model in which $q=0$ and $\underset{\sim}{\Theta}_{q,m}=I$ equation 3-9 reduces to

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\ \underset{\sim}{\delta}(t) = \underset{\sim}{\xi}(t) \qquad (3\text{-}11)$$

Thus, as with the mixed model, the $STAR(\underset{\sim}{P}_\lambda)I_m$ model's mean shift will only as $t \to \infty$ approach the steady state level.

For the case where there is no environmental influence, $I_m=0$, in equation 3-8, we have the deterministic component of the intervention model given as,

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\left[\underset{\sim}{\delta}(t)-\underset{\sim}{\xi}(t)\right] = 0$$

or

$$\underset{\sim}{\delta}(t) = \underset{\sim}{\xi}(t). \qquad (3\text{-}12)$$

Thus when there is no environmental influence, the total effect of the intervention is fully realized instantaneously with transition to the steady state level at t+1 for an intervention having occurred in the interval t, t+1.

3.1.2 Multiple Interventions

In the previous section only a single intervention was introduced. Often several interventions can occur. In this section we will generalize the (STARMA)I_m model to allow for multiple interventions and discuss the interpretation of the resulting mean-shift function.

The STARMA($\underset{\sim}{P}_\lambda$,0,$\underset{\sim}{q}_m$)$I_m$ model is generalized for η different interventions by

$$\underset{\sim}{\Phi}_{p,\lambda}(B)(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \sum_{r=1}^{\eta} \left|(1-I_{m,r})\, \underset{\sim}{\Phi}_{p,\lambda}(B) + I_{m,r}\, \underset{\sim}{\Theta}_{q,m}(B)\right| \underset{\sim}{\xi}^{(r)}(t) \qquad (3\text{-}13)$$

$$+ \underset{\sim}{\Theta}_{q,m}(B)\, \underset{\sim}{\varepsilon}_t.$$

where

$$I_{m,r} = \begin{cases} 1 & \text{realized effect of the r-th intervention effect by the environmental process.} \\ 0 & \text{realized effect of the r-th intervention not effected by the environmental process.} \end{cases}$$

$\underset{\sim}{\xi}^{(r)}(t)$ is the indicator variable for the r-th intervention and is the product of the switch variable $\xi_t^{(r)}$ and the program utility of the r-th intervention, $\underset{\sim}{\delta}^{(r)}$. This model will be referred to as the STARMA($\underset{\sim}{P}_\lambda$,0,$\underset{\sim}{q}_m$)$I_{m,\eta}$ model.

The random component of the STARMA($\underset{\sim}{P}_\lambda$,0,$\underset{\sim}{q}_m$)$I_{m,\eta}$ model is, exactly the same as that of the corresponding single intervention STARMA($\underset{\sim}{P}_\lambda$,0,$\underset{\sim}{q}_m$)$I_m$ model. The deterministic component of the STARMA($\underset{\sim}{P}_\lambda$,0,$\underset{\sim}{Q}_m$)$I_{m,\eta}$ model is,

$$\underset{\sim}{D}(t) = \underset{\sim}{\mu} + \sum_{r=1}^{\eta} \underset{\sim}{\Phi}_{p,\lambda}(B)^{-1} \left|(1-I_{m,r})\, \underset{\sim}{\Phi}_{p,\lambda}(B) + I_{m,r}\, \underset{\sim}{\Theta}_{q,m}(B)\right| \underset{\sim}{\xi}^{(r)}(t),$$

and the realized mean shift function $\underset{\sim}{\delta}(t)$ is

$$\underset{\sim}{\delta}(t) = \underset{\sim}{D}(t) - \underset{\sim}{\mu} \qquad (3\text{-}14)$$

$$= \sum_{r=1}^{\eta} \underset{\sim}{\Phi}_{p,\lambda}(B)^{-1}\left|(1-I_{m,r})\, \underset{\sim}{\Phi}_{p,\lambda}(B) + I_{m,r}\, \underset{\sim}{\Theta}_{q,m}(B)\right| \underset{\sim}{\xi}^{(r)}(t).$$

Equation 3-14 reveals that the realized mean-shift function can be decomposed into η components, each corresponding to the η intervention

$$\underset{\sim}{\delta}(t) = \sum_{r=1}^{n} \underset{\sim}{\delta}^{(r)}(t) \qquad (3\text{-}15)$$

and

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\underset{\sim}{\delta}^{(r)}(t) = [(1-I_{m,r})]\underset{\sim}{\Phi}_{p,\lambda}(B) + I_{m,r}\Theta_{q,m}(B)]\xi^{(r)}(t). \qquad (3\text{-}16)$$

As in the single intervention model, when $I_{m,r} = 1$, the environmental influence situation, the realized mean shift function of the r-th intervention, $\delta^{(r)}(t)$, is a function of $\underset{\sim}{\phi}$, $\underset{\sim}{\Theta}$ as well as $\underset{\sim}{\delta}^{(r)}$. When $I_{m,r} = 0$, the environment is not involved and $\underset{\sim}{\delta}^{(r)}(t)$ is not a function of the environmental noise parameters. From equations 3-14 and (3-15), we see that the interventions have a superposition property. This means that they will add up to give the realized mean shift function $\underset{\sim}{\delta}(t)$. When the situation arises that sequential interventions interact with each other, e.g., r_1th intervention occurs before r_2th intervention and have interactions (e.g. a synergy) equation (3-15) and (3-16) are still the appropriate model, however, $\underset{\sim}{\delta}^{(r_2)}(t)$, $\underset{\sim}{\xi}^{(r_2)}(t)$ and $\underset{\sim}{\delta}^{(r_2)}$ will be interpreted differently. In this case, the mean shift function of r_2th intervention is confounded with the (r_1,r_2) interaction, i.e. the $\underset{\sim}{\delta}^{(r_2)}(t)$ is the summation of the mean shift function of r_2th intervention and the (r_1,r_2) interaction effect. Thus the proper

interpretation of $\underset{\sim}{\delta}^{(r_2)}(t)$ is the mean shift function of the r_2th intervention given that the r_1th intervention has been introduced. Also $\underset{\sim}{\xi}^{(r_2)}(t)$ as well as $\underset{\sim}{\delta}^{(r_2)}$ should be interpreted in this way.

3.2 Physical Representations of the Model

When an intervention program is introduced in one or more locations, the effect may also be realized at other locations via a diffusion process. In this section, we will discuss the capabilities of the STARIMA$(\underset{\sim}{p}_\lambda,d,\underset{\sim}{q}_m)I_m$ models in describing diffusion phenomena.

3.2.1 Diffusion Mechanism

Spatial-temporal diffusion is the spread of a phenomena within a given space through time, so as to alter the distributional pattern of the phenomena over time. Four distinct types of the diffusion mechanisms are useful in classifying the spatial-temporal diffusion process.

1) Translation: occurs when members of the population and their relative positions do not change, while the positions of the members are translated from time t to t + Δt.

2) Domain-change: occurs when members of the population do not change but their relative positions as well as the nonzero population location number change from t to t + Δt.

3) Growth: occurs when new members are created and into the

population between time t and t + Δt, increasing the population in size.

4) Contraction: occurs when some members of the population die out between time t and t + Δt, decreasing the population size.

These four diffusion mechanisms are illustrated in figure 3-2. A more detailed discussion of these mechanisms is contained in Brown [1968].

There are two distinct types of spatial-temporal diffusion process. Each are described by combinations of the four diffusion mechanisms. The types of spatial-temporal diffusion processes are;

I. Regenerating type diffusion: occurs when new members are generated, added into the population and/or some old members die out, and the population size changes while individual locations may/may not change. (A combination of type 1, type 2, type 3 and 4 mechanisms with growth rate ≠ contraction rate.) We will say that it is a regenerating (+) type process when the net gain in the size of total population is positive, is a regenerating (-) type process when the net gain is negative.

II. Relocation type diffusion: occurs when members of the population change their relative positions and/or locations while the population size stays constant (a combination of the type 1, type 2 mechanism and/or type 3 and type 4 mechanisms with growth rate = contraction rate).

3.2.2 Characterization of the Diffusion Processes Types by the STARIMA(P_λ,d,q_m)I_m Models.

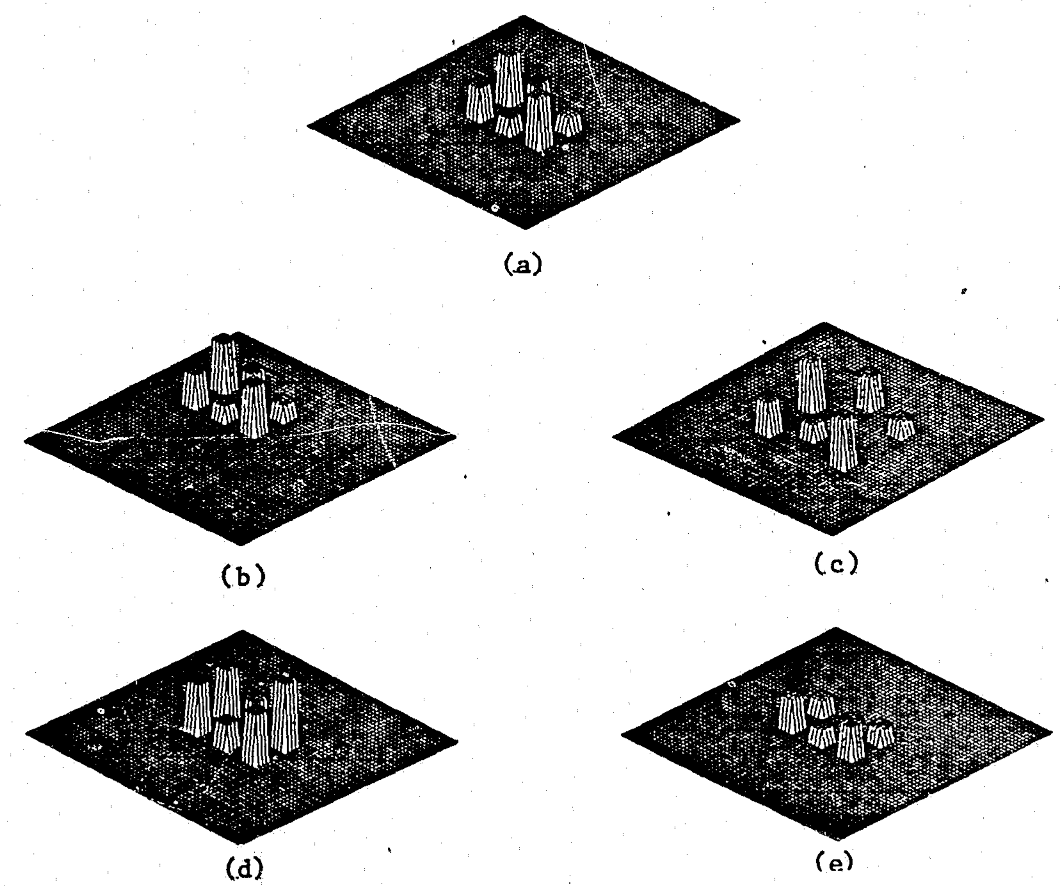


Figure 3-2. Diffusion Mechanisms
(a) Population Distribution at Time t
(b) Population Distribution at Time t+Δt through Translation
(c) Population Distribution at Time t+Δt through Domain-Change
(d) Population Distribution at Time t+Δt through Growth
(e) Population Distribution at Time t+Δt through Contraction

In the previous section two types of general diffusion processes were described. Here we focus on the STARIMA($\underset{\sim}{P_\lambda}$,d,$\underset{\sim}{q_m}$)$I_m$ models to assess whether and which subgroups of models in this class describe these two types of diffusion and if so which model subgroups correspond to what type of diffusion process. To assess this issue, we use the equation of mean shift function of the STARIMA($\underset{\sim}{P_\lambda}$,d,$\underset{\sim}{q_m}$)$I_m$ model,

$$\nabla^d \underset{\sim}{\Phi_{p,\lambda}}(B)\ \underset{\sim}{\delta}(t) = \lfloor (1-I_m)\nabla^d \underset{\sim}{\Phi_{p,\lambda}}(B) + I_m \underset{\sim}{\Theta_{q,m}}(B) \rfloor\ \underset{\sim}{\xi}(t).$$

This equation can be expressed alternatively as,

$$\underset{\sim}{\delta}(t) = \Lambda(B)\ \underset{\sim}{\xi}(t) \qquad (3\text{-}17)$$

with

$$\Lambda(B) = \lfloor \nabla^d \underset{\sim}{\Phi_{p,\lambda}}(B) \rfloor^{-1} \lfloor (1-I_m)\nabla^d \underset{\sim}{\Phi_{p,\lambda}}(B) + I_m \underset{\sim}{\Theta_{q,m}}(B) \rfloor \qquad (3\text{-}18)$$

$$\Lambda(B) = \sum_{k=0}^{\infty} \Lambda^{(K)} B^K \qquad (3\text{-}19)$$

Where $\Lambda^{(K)}$ is obtained recursively from equation (3-2), and $\Lambda^{(0)} = I$.

When we let $I_m = 0$ in equation (3-18), which corresponds to the non-environmental influence case, we have, $\Lambda^{(0)} = I$ and $\Lambda^{(K)} = 0$ for

all K > 0, i.e.,

$$\underset{\sim}{\delta}(t) = \underset{\sim}{\xi}(t).$$

The more interesting and broader characterization of the different diffusion type occurs in the environmental influence structure with $I_m = 1$. The balance of this section will address this situation.

For the $I_m = 1$ case we will assume for descriptive purposes that an intervention program is introduced at location i between time $t = n_1$ and $t = n_1+1$ for one period only, i.e.,

$$\underset{\sim}{\xi}(t) = P_t^{n_1}(T_p=1)\ \underset{\sim}{\delta} \qquad (3\text{-}20)$$

where $\underset{\sim}{\delta} = (0,0,\ldots,\delta^0,0,\ldots)$ with δ^0 at location i.

Then we have,

$$\begin{cases} \underset{\sim}{\delta}(t) = \underset{\sim}{0} & t < n_1 \quad \text{(pre-intervention)} \\ \underset{\sim}{\delta}(t) = \Lambda^{(K)}\underset{\sim}{\delta}, & t = n_1+K+1, \quad k > 0 \quad \text{(post-intervention)} \end{cases}$$

From the previous discussion, we know that the major difference between the regenerating processes and the relocation processes is the conservation of population. For the intervention described in equation (3-

20), the conservation of population can be stated as

$$
\begin{cases}
\sum_{j=1}^{LN} \delta_j(t) = 0 \qquad t < n_1 & \text{(pre-intervention)} \\[2ex]
\sum_{j=1}^{LN} \delta_j(t) = \delta^0 \qquad t > n_1 + 1 & \text{(post-intervention)}
\end{cases}
$$

where $\delta_j(t)$ is the jth element of $\underset{\sim}{\delta}(t)$.

The first equation can always be satisfied, since $\underset{\sim}{\delta}(t) = \underset{\sim}{0}$ for $t < n_1$, and the second one is equivalent to;

$$\sum_{j=1}^{LN} \Lambda_{ji}^{(K)} = 1 \quad \text{all } K > 20,\ i = 1,2,\ldots,LN. \tag{3-21}$$

where $\Lambda_{ji}^{(K)}$ is the (j,i) element of $\Lambda^{(K)}$ matrix.

3.2.2.1 Processes and Pulse Inputs

For the stationary STARMA($\underset{\sim}{P_\lambda}$,0,$\underset{\sim}{q_m}$)$I_m$ model equation (3-21) can't be satisfied for all k > 0 no matter if it is a STAR, STMA or STARMA model. Therefore the stationary STARMA($\underset{\sim}{P_\lambda}$,0,$\underset{\sim}{q_m}$)$I_m$ models are always kind of the regenerating (-) or regenerating (+) type diffusion processes. Only a non-stationary STARIMA($\underset{\sim}{P_\lambda}$,0,$\underset{\sim}{q_m}$)$I_m$ model can satisfy equation (3-21). For example, the STARIMA(0,1,$\underset{\sim}{q_m}$)I_m model will satisfy

equation (3-21) with the model parameters constrainted as follows

$$\sum_{j=1}^{LN} \Big[\sum_{\ell=0}^{m_K} (\Theta_{k\ell} W^{(\ell)}) \Big]_{j,i} = 0 \quad \text{for} \quad K = 1,2,\ldots,q. \tag{3-22}$$

where $\Big[\sum_{\ell=0}^{m_K} (\Theta_{K\ell} W^{(\ell)}) \Big]_{j,i} = 0$ is the (j,i) element of $\Big[\sum_{\ell=0}^{m_K} (\Theta_{K\ell} W^{(\ell)}) \Big]$ matrix.

A special case of the STARIMA(0,1,$\underset{\sim}{q_m}$)I_m model class is the STARIMA(0,1,0)I_m model i.e., p=0, d=1, q=0. Since the necessary and sufficient conditions of the relocation type diffusion requires the conservation of population as well as the capability to move to its neighbors, to have the moving capability, it is required that $\Lambda^{(K)} \neq I$, $\Lambda^{(K)} \neq 0$ for at least one K. Even though the STARIMA(0,1,$\underset{\sim}{q_m}$)I_m has the moving capability, the STARIMA(0,1,0)I_m model does not. For example, for a ring system with 5 locations the first and second order weight structure is

$$
W^{(1)} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}
\qquad
W^{(2)} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}
$$

We have a STARIMA$(0,1,q_m)I_m$ model with $m_1=2$, $m_2=2$, ..., $m_q=2$. When $\Theta_{K0} + 2\Theta_{K1} + 2\Theta_{K2} = 0$ for all $k=1,2,\ldots,q$, we can see that the requirements imposed by equation (3-22) are satisfied, since

$$\sum_{j=1}^{LN} \Big| \sum_{\ell=0}^{M_k} (\Theta_{K\ell} W^{(\ell)}) \Big|_{ji} = \Theta_{K0} + 2\Theta_{K1} + 2\Theta_{K2}.$$

However, if one of the Θ_{K1}, Θ_{K2}, $K=1,2,\ldots,q$ is not equal to zero, i.e. $\Theta_{k1} \neq 0$, then at least $\Lambda^{(K)} \neq I$ and $\Lambda^{(K)} \neq 0$, so the moving capability requirement is satisfied and this is a relocation diffusion process.

3.2.2.2 Homogeneously Nonstationary Processes

The homogeneously nonstationary STAR(P_{λ}) process can be of relocation diffusion type if and only if

$$\sum_{i=1}^{LN} \Big| \sum_{K=1}^{t} \sum_{\ell=0}^{\lambda K} \sum_{j=1}^{LN} \phi_{K\ell}\, W^{(\ell)}_{i,j} e_{r,j} \Big| = 1 \qquad (3\text{-}23)$$

for $t = 1,2,\ldots,P$ and $r = 1,2,\ldots,LN$.

where $e_{r,j}$ is the j-th element of vector e_r, which is the unit vector with 1 at location r. Equation (3-23) guarantees that any input to any location in this system will stay in the system forever. This condition is very restrictive in that it requires conservation of population for all time t.

A less restrictive constraint to ensure relocation diffusion is to not require absolute conservation of population for all time t but rather to consider asymptotic population conservation as $t \to \infty$. That is a set of conditions that will guarantee that any input will change the steady state of the system and the system will stay at the new steady state forever or until there is another input. Such a set of conditions allows a transient period for the system to adjust itself to a new steady state and the difference between the new state and the original state is the conserved part of the input and stays in the system forever. This kind of restrictions will be referred to as the relaxed conditions for the relocation processes.

The stationarity conditions discussed in Hannan [1970] applied to the STAR process are as follows: if every Z_u that solves,

$$\mathrm{Det}\Big| Z_u^P I - \sum_{k=1}^{P} \Big(\sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \Big) Z_u^{P-K} \Big| = 0$$

lies inside the unit circle, then the STAR($\underset{\sim}{P_\lambda}$) process will be stationary. This implies that any linear combinations of these inputs will damp out when $t \to \infty$. However, if all the Z_u's lie on the unit circle, then no linear combinations of these inputs will damp out. Thus, all these linear combinations will be conservative. So the relaxed conditions for relocation in the STAR($\underset{\sim}{P_\lambda}$) process is; if every Z_u that solves,

$$\mathrm{Det}\left| Z_u^P I - \sum_{K=1}^{P} \left(\sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \right) Z_u^{P-K} \right| = 0 \qquad (3\text{-}24)$$

lies on the unit circle, then these homogeneously nonstationary STAR($\underset{\sim}{P_\lambda}$) process will be of relocation diffusion type.

To compare these constraints for a relocation type diffusion process, equation (3-23) can be restated as follows; if every Z_u that solves,

$$\mathrm{Det}\left| Z_u^t I - \sum_{K=1}^{t} \left(\sum_{\ell=0}^{\lambda K} \phi_{K\ell} W^{(\ell)} \right) Z_u^{t-K} \right| = 0, \quad t = 1,2,\ldots,P \qquad (3\text{-}25)$$

has the same solution $Z_u = 1$, then the homogeneously nonstationary STAR($\underset{\sim}{P_\lambda}$) process is of the relocation diffusion type. Note that equation (3-24) is a relaxation of equation (3-25) in the sense that the

allowed solution set of equation (3-25) is a subset of the allowed solution set of equation (3-24) and the equation set (3-24) is a subset of the equation set (3-25). Thus when equation (3-25) is satisfied, equation (3-24) is also satisfied. The reverse is however not true.

3.2.2.3 Steady State Gain For the Step Function Input

In previus sections the intervention program input variable $\underset{\sim}{\xi}(t)$ has been assumed to be the pulse function, in this section the step input situation will be discussed. It is assumed that

$$\underset{\sim}{\xi}(t) = S_t^{n_1} \underset{\sim}{\delta}$$

where $\underset{\sim}{\delta} = (0,0,\ldots,\delta^0,0,\ldots)$ with δ^0 at location i.

The step input function can be viewed as superposition of a sequence of pulse functions with the output of this system being the superposition of the outputs of the sequence of pulse function inputs. This can be seen in equation (3-7) by letting

$$\underset{\sim}{\xi}(t) = \underset{\sim}{\xi}_1(t) + \underset{\sim}{\xi}_2(t)$$

$$\underset{\sim}{\delta}_1(t) = \Lambda(B)\, \underset{\sim}{\xi}_1(t), \quad \underset{\sim}{\delta}_2(t) = \Lambda(B)\, \underset{\sim}{\xi}_2(t)$$

Therefore; $\underset{\sim}{\delta}(t) = \underset{\sim}{\delta}_1(t) + \underset{\sim}{\delta}_2(t) = \Lambda(B)\left(\underset{\sim}{\xi}_1(t) + \underset{\sim}{\xi}_2(t)\right) = \Lambda(B)\underset{\sim}{\xi}(t)$.

So at any time point, the necessary and sufficient conditions for the system population to be conserved is that the system input rate equals the dying-out rate. In the following we will consider the strictly conservative system in which the system reaches the steady state gain instantaneously, maintaining its population, and the asymptotically conservative system, in which the system reaches the steady state gain by passing through transient states. Thus the latter differs from the first in that the strictly conservative system has constant population for all $t > n_1$.

The have the instantaneous and strictly conservative system, the system has the followng property:

$$\begin{cases} \underset{\sim}{\delta}(t) = 0 & \text{(pre-intervention)} \\ \underset{\sim}{\delta}(t) = \underset{\sim}{\delta}^S & \text{(post-intervention)} \end{cases}$$

where $\underset{\sim}{\delta}^S$ denotes the steady state gain of the system and is a constant vector.

Since $\underset{\sim}{\delta}(t) = \Lambda(B)\, \underset{\sim}{\xi}(t) = \sum_{K=0}^{\infty} \Lambda^{(K)} S_{t-K}^{n_1} \underset{\sim}{\delta}$,

the followng conditions should be satisfied;

$$\underset{\sim}{\delta}(t) = \underset{\sim}{\delta}(t+u) \qquad u = 1,2,\ldots,\infty,\ t > n_1$$

$$\text{i.e.} \sum_{K=0}^{\infty} \Lambda^{(K)} S_{t-K}^{n_1} \underset{\sim}{\delta} = \sum_{K=0}^{\infty} \Lambda^{(K)} S_{t+u-K}^{n_1} \underset{\sim}{\delta}.$$

This gives the following equations;

$$\sum_{K=0}^{t-n_1} \Lambda^{(K)} = \sum_{K=0}^{t+u-n_1} \Lambda^{(K)} \qquad u = 1,2,\ldots,\infty;\ t > n_1 \tag{3-26}$$

To satisfy these equations for $u = 1,2,\ldots,\infty$, the only solution is $\Lambda^{(0)} = I$, $\Lambda^{(K)} = 0$ all $K > 1$. As we have pointed out in section 3.2.2, this is the non-environment involved intervention process.

The asymptotically conservative system reaches the steady state after the transient periods. This means that the condition in equation 3-26 can be relaxed as follows;

$$\sum_{K=0}^{t-n_1} \Lambda^{(K)} = \sum_{K=0}^{t+u-n_1} \Lambda^{(K)} \qquad u = 1,2,\ldots,\infty;\ t > n_1 + T_0 \tag{3-27}$$

where T_0 is the transient periods.

The number of transient periods T_0 for the $\underset{\sim}{\text{STMA}}(q_m)I_m$ process is q, since

$$\sum_{K=0}^{t-n_1} \Lambda^{(K)} = \sum_{K=0}^{\mathrm{Min}\{(t+u-n),q\}} \Lambda^{(K)}.$$

so

$$\sum_{K=0}^{t-n_1} \Lambda^{(K)} = \sum_{K=0}^{t+u-n_1} \Lambda^{(K)} = \sum_{K=0}^{q} \Lambda^{(K)} \qquad \text{when } t > n_1 + T_0$$

The number of transient periods T_0 of the STAR(P_λ)$\underset{\sim}{I_m}$ and the STARMA($\underset{\sim}{P_\lambda}$,$\underset{\sim}{q_m}$)$I_m$ processes are much longer than those of the (STMA)I_m process. In fact, they will reach the steady state only when $T_0 \to \infty$. Since $\Lambda^{(K)} S_t^{n_1} \underset{\sim}{\delta}$ represents the contribution of the intervention input, in order to satisfy the asymptotically conservative conditions when $T_0 \to \infty$ it is necessary and sufficient to have

$$\lim_{K\to\infty} \Lambda^{(K)}\ S_t^{n_1}\ \underset{\sim}{\delta} = \underset{\sim}{0} \quad \text{or equivalently}$$

$$\lim_{K\to\infty} \Lambda^{(K)} = 0. \tag{3-28}$$

Solutions to satisfy equation (3-28) are discussed in Deutsch and Pfeifer [1979] and are those systems who's parameters lie inside the stationary boundary, i.e. every Z_u that solves equation (3-25) should lie inside the stationary boundary.

The diffusion type of the instantaneous steady state gain process can be interpreted as the space-time regenerating diffusion pro-

cess or the space-time relocation diffusion process. This depends on how we define the system and the system input.

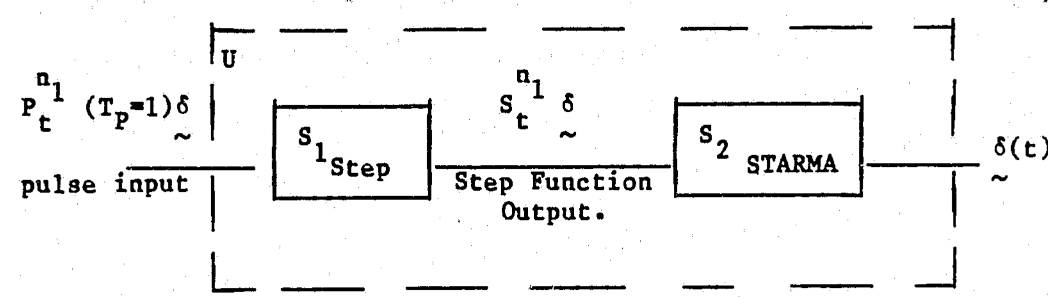


Figure 3-3. Permanent Impact System Diffusion Type Interpretation.

In figure 3-3, the step transformation block transforms the pulse input into a step function and the step function output of S_1 is input into the STARMA diffusion process. System U contains S_1 and S_2 as its subsystem, and the pulse input $P_t^{n_1}(T_p=1)\ \underset{\sim}{\delta}$ is input into U to give $\underset{\sim}{\delta}(t)$ as the output. From the viewpoint of the S_2 system, the diffusion type for the instantaneously steady state gain is of the regenerating type, since the total input $S_t^{n_1}\ \underset{\sim}{\delta}$ isn't conserved. While from the viewpoint of the system U, which has the pulse input, the diffusion type for the steady state gain is of the relocation type, since the pulse input $P_t^{n_1}(T_p=1)\ \underset{\sim}{\delta}$ has been conserved.

Similarly, the diffusion type of this asymptotically steady state gain process can be interpreted as the space time regenerating process or the relaxed space-time relocation diffusion process. From the viewpoint of the whole system U, which includes the permanent impact transformation block, the asymptotically steady state process can be interpreted as the relaxed space-time relocation diffusion process.

3.2.3 Properties of the Diffusion Process

Having described the types of diffusion processes, we now turn our attention to the properties of a given type of diffusion process. Three characteristics of a diffusion process are addressed; the sphere of influence of the process, the speed of the process and the amplitude of the process. In describing these characteristics it is helpful to formally state the following;

1. If location j is one of the ℓth order neighbors of location i, then location j is connected to location i by the ℓth-order neighbor-chain.
2. The set of locations in the space considered that are connected by the ℓth-order neighbor-chain and contains location i is called the ℓth-order connected regions of location i.
3. The connected regions of location i for $\mathrm{STARMA}(\underset{\sim}{P}_{\lambda},0,\underset{\sim}{q}_{m})I_m$ models is the union of all 1st-order connected regions, 2nd-order connected regions, ..., up to $\max\{\lambda_{max},m_{max}\}$th-order connected regins of location i, where $\lambda_{max} = \mathrm{Max}\{\lambda_1,\lambda_2,\ldots,\lambda_P\}$ and $m_{max} = \mathrm{Max}\{M_1,M_2,\ldots,M_q\}$.
4. The influenced regions of location i for $\mathrm{STARMA}(\underset{\sim}{P}_{\lambda},0,\underset{\sim}{q}_{m})I_m$ models is all regions that can, in the long run, be influenced by the

intervention program introduced at location i.

5. If an intervention program is introduced at location i, all the 1st order neighbors, 2nd order neighbors, ..., and ℓ*th order neighbors, but not (ℓ*+1)th order neighbors of location i are influenced at the very next moment, we will say that the diffusion speed of this $\mathrm{STARMA}(\underset{\sim}{P}_{\lambda},0,\underset{\sim}{q}_{m})I_m$ process v equals ℓ^*.

Since the characteristics of the diffusion process of the $\mathrm{STAR}(\underset{\sim}{P}_{\lambda})I_m$, $\mathrm{STMA}(\underset{\sim}{q}_{m})I_m$ and $\mathrm{STARMA}(\underset{\sim}{P}_{\lambda},0,\underset{\sim}{q}_{m})I_m$ process in the environment involved situation are quite different, the following section discusses each in turn. We will keep the same assumption on $\underset{\sim}{\xi}(t)$ as that in equation (3-20), i.e.

$$\underset{\sim}{\xi}(t) = P_t^{n_1}\,(T_p=1)\,\underset{\sim}{\delta}$$

where $\underset{\sim}{\delta} = [0,0,0,\ldots,\delta^0,0,\ldots]$

with δ^0 at location i.

3.2.3.1 AR Type

The diffusion process of $\mathrm{STAR}(\underset{\sim}{P}_{\lambda})I_m$ model is described in the recursive equation that is obtained by setting $\underset{\sim}{\Theta}_{q,m}(B) = I$ in eqution (3-9),

$$\underset{\sim}{\delta}(t) = \sum_{K=1}^{P} \sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \underset{\sim}{\delta}(t-K) + \underset{\sim}{\xi}(t),$$

This can be rewritten as,

$$\begin{cases} \underset{\sim}{\delta}(t) = 0 & t < n_1 \\ \underset{\sim}{\delta}(t) = \underset{\sim}{\delta} & t = n_1 + 1 \\ \underset{\sim}{\delta}(t) = \sum_{K=1}^{P} \sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \underset{\sim}{\delta}(t-K) & t > n_1 + 2. \end{cases} \tag{3-29}$$

From equation (3-29), we can see that for the STAR process that the influenced regions of any location i will be equal to the connected regions of that location, since the influence will be transmitted from location i to its neighbors and then retransmitted to all its connected regions before the effect of this intervention completely dies out as $t \to \infty$. In physical/engineering systems this is intuitively appealing in that a given location in the system will receive a stronger and quicker influence from those closer-connected locations. The term "closer-connected locations" does not necessary mean close in the sense of Euclidean distance. This implies that,

$$\lambda_{K+1} - \lambda_K < \lambda_K - \lambda_{K-1}$$

and thus the diffusion speed of the STAR process is λ_1.

The amplitude of diffusion for the stationary STAR$(\underset{\sim}{P_\lambda})I_m$ process can be computed recursively using equation (3-26). The STAR process is a relocation type diffusion when the parameters lie on the stationary boundary. The farther away the autoregressive parameter vector falls from this boundary the more rapid the loss of population. Thus, given ϕ_{K0}, K = 1, 2, ...P are fixed, the stronger the conservation of population tendency is, the larger the diffusion amplitude will be.

3.2.3.2 MA Type

By setting $\Phi_{p,\underset{\sim}{\lambda}}(B) = I$ in equation (3-9), we have the recursive equation for the realized mean-shift function $\underset{\sim}{\delta}(t)$ of the STMA$(\underset{\sim}{q_m})I_m$ model,

$$\underset{\sim}{\delta}(t) = \underset{\sim}{\xi}(t) - \sum_{K=1}^{q} \sum_{\ell=0}^{m_K} \theta_{K\ell} W^{(\ell)} \underset{\sim}{\xi}(t).$$

with the intervention variable $\underset{\sim}{\xi}(t) = P_t^{n_1}(T_p=1)\,\underset{\sim}{\delta}$, and $\underset{\sim}{\delta}=[0,0,0,\ldots,\delta^0,0,\ldots,0]$, this equation can be rewritten as,

$$\begin{cases} \underset{\sim}{\delta}(t) = 0 & t < n_1 \text{ or } t > n_1 + q + 2 \\ \underset{\sim}{\delta}(t) = \underset{\sim}{\delta} & t = n_1 + 1 \end{cases} \tag{3-30}$$

$$\underset{\sim}{\delta}(t) = \sum_{\ell=0}^{m_K} \theta_{K\ell} W^{(\ell)} \underset{\sim}{\delta}\, P_{t-K}^{n_1} \qquad t = n_1 + K + 1, \quad K < q$$

From equation (3-30), it is seen that the influenced regions of locations i will include only the 1st order neighbors, 2nd order neighbors, ..., up to the m_{max}-th order neighbors, where $m_{max} = \mathrm{Max}\{m_1, m_2, \ldots, m_q\}$. Also the diffusion speed between $t = n_1+1$ and $t = n_1 + 2$ is m_1, but the diffusion process will completely die out at $t = n_1 + q + 2$. Note that the STMA type diffusion is quite different from the STAR type in its ability to transmit the influence. In the STAR type diffusion processes, any connected location that has received influence has the ability to retransmit the influence to its neighbors. thus, whereas the STAR process has the received influence at time t behave as an intervention transmitted to (t+1) for all connected regions for all t, in the STMA type diffusion, the influenced region has no such ability to retransmit the influence to its neighbors. The only influence transmitted comes directly from the location that the intervention program is introduced at time t. Thus the influenced region of the $\mathrm{STMA}(\underset{\sim}{q_m})I_m$ process may not cover the whole connected region of location i in which the intervention was implemented.

From equation (3-30) the amplitude of the STMA process can be recursively computed. This amplitude is seen to depend on the magnitude of the moving average model parameters as well as that of the input amplitude $\underset{\sim}{\delta}$.

3.2.3.3 ARMA Type

For the $\mathrm{STARMA}(\underset{\sim}{P_\lambda}, 0, \underset{\sim}{q_m})I_m$ model, the recursive formula for $\underset{\sim}{\delta}(t)$ is,

$$\underset{\sim}{\delta}(t) = \sum_{K=1}^{P} \sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \underset{\sim}{\delta}(t-K) - \sum_{K=1}^{q} \sum_{\ell=0}^{m_K} \theta_{K\ell} W^{(\ell)} \underset{\sim}{\xi}(t). \qquad (3\text{-}31)$$

Thus the STARMA type diffusion is a combination of STAR and STMA type. For the $\underset{\sim}{\xi}(t)$ as described in equation (3-20), we can rewrite equation (3-31) as follows,

$$\underset{\sim}{\delta}(t) = \underset{\sim}{0} \qquad t < n_1 \qquad (3\text{-}32)$$

$$\underset{\sim}{\delta}(t) = \underset{\sim}{\delta} \qquad t = n_1 + 1$$

$$\underset{\sim}{\delta}(t) = \sum_{K=1}^{P} \sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \underset{\sim}{\delta}(t-K) - \sum_{\ell=0}^{m_u} \theta_{u\ell} W^{(\ell)} \qquad t = n_1 + u + 1, \quad u < q$$

$$\underset{\sim}{\delta}(t) = \sum_{K=1}^{P} \sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \underset{\sim}{\delta}(t-K) \qquad t > n_1 + q + 2$$

The diffusion speed of this process may depend on the elapsed time since the intervention was introduced. At $t = n_1+1$, the diffusion speed is $\mathrm{Max}\{\lambda_1, m_1\}$, while at $t = n_1+q+2$, the diffusion speed will be λ_1. Thus the moving average influence will cause a change in the speed of diffusion if m_1 is larger than λ_1 for q periods until the effect of the moving average term dissipates. The influenced regions of this mixed process is the same as the STAR process, i.e. all connected regions of location i where intervention is introduced.

The amplitude of the diffusion process for the mixed

STARMA(P_{λ},0,q_m)I_m process can be recursively computed using equation (3-32). Here we see that it is composed of two parts, due to the AR the MA components. From equation (3-31), we see that the input to the MA component comes only from $\xi(t)$ which, in part, forms subsequent $\delta(t)$ where as the input for the AR component comes from the lagged values of the mean shift function, $\delta(t-K)$. Thus the MA component shapes/-influences the AR component but the MA component is independent of the AR component. Thus, the total amplitude of the mixed process depends on two factors;

1) the individual and relative amplitudes of the AR and MA components and
2) the nature of the interaction of the MA component that can be constructive, reinforcing, or destructive.

3.3 Simulation of the Diffusion Process of the Low Order STARMA(P_{λ},0,q_m)I_m Models

In the following, we will illustrate the diffusion processes for low order STARMA(P_{λ},0,q_m)I_m model with Max{P,q} < 1 and Max{λ_1,m_1} < 2. Since there is no diffusion phenomena in the non-environmental influenced case, I_m = 0, we will only illustrate the environment involved case with I_m = 1.

All simulated illustrations are from (11 x 11) square regions. Thus there are 121 locations identified as (i,j)th location, i = 1,2,...,11, j = 1,2,...,11. The neighbor structure used is; The 1st order neighbors of the (i,j)-location are locations (i+1,j), (i-1,j), (i,j-1) and (i,j+1). The 2nd order neighbors of the (i,j) location are

(i-1,j-1), (i-1,j+1), (i+1,j-1) and (i+1,j+1). Thus all the (11 x 11) locations form the connected region. The intervention is introduced at the (6,6) location between time t = T_0-1 and t = T_0. The amplitude of this intervention δ^0 is set to 10 in all simulations. The intervention input is shown in Figure 3-4.

Figure 3-5 illustrates the AR type diffusion, figure 3-6 illustrates the MA type diffusion process and figure 3-7 the ARMA type diffusion process for model parameters selected to ensure stationarity. Thus these figures all illustrate regeneration type diffusion processes.

In figure 3-5(a-c), it is seen that the effect of the intervention spreads over all locations. Figure 3-5(a) and figure 3-5(b) are both STAR(1_1)I_m models, thus the speed of diffusion is the same (i.e. v = 1). However due to the different model parameter values, the amplitudes realized are different. The amplitudes in the figure 4-5(a), in which the process parameters are closer to the stationary bundary , are 1.00, 0.90, 0.56, 0.36 in the (5,6) location, which is one of the 1st order neighbors of the (6,6) location, at time T = 1,2,3,4, respectively, where T denotes the elapsed time since the intervention was introduced. The corresponding amplitudes in the figure 3-5(b), in which the process parameters are farther from the stationary boundary, are 1.00, 0.40, 0.21, 0.10, respectively. As expected, the diffusion amplitudes are bigger for the processes in which the process parameters are closer to the stationary boundary.

From the STAR(1_2)I_m model illustrated in figure 3-5(c), we can see that the velocity of spreading is twice those in figure 3-5(a) and (b) since λ_1 = 2. This is clearly illustrated by the number of regions influenced at a given time after the intervention is introduced between $t = T_0-1$ and $t = T_0$. From these plots, we see that at $t = T_0+2$, i.e., $T = 2$, figure 3-5(c) has 8 influenced locations while figure 3-5(a) and (b) have only 4 influenced locations.

For the MA type processes in figure 3-6, we can see that the effect of the pulse intervention disappear abruptly after $t = T_0+2$. It is also quite different from the STAR($1_{\lambda 1}$)I_m process in that the influenced regions are limited. In figure 3-6(a) and (b) the influenced region contains the 1st order neighbors of the (6,6) location and in figure 3-7(c), it contains the 1st order and 2nd order neighbors since the latter has a larger spatial influence. The speed of spreading for the STMA(1_2)I_m model is 2 at $t = T_0+1$, as seen in figure 3-6(c), while the speed of spreading for the STMA(1_1)I_m models is 1. After $t = T_0+2$, the effect disappears at all locations and the velocity of spreading for STMA(1_{m1})I_m model becomes 0 for $t > T_0+m_1$. Comparing figure 3-6(a) and (b), we see that different $\underset{\sim}{\theta}$-parameter values give different amplitudes. In figure 3-6(a), where the model parameters are, $\theta_{10}=-.4$, $\theta_{11}=-.4$, the amplitudes at the (5,6) and (6,6) locations at T=1 are 1.00 and 4.00 respectively, while the corresponding amplitudes in the figure 4-6(b), where the model parameters are $\theta_{10}=-.2$, $\theta_{11}=-.4$, are 1.00 and 2.00.

In figure 3-7(a-b), the infuenced regions contain all these (11 x 11) locations. In figure 3-7(a), the speed of spreading is constant all the time, i.e. v=2. But in figure 3-7(b), the speed is 2 at T = 1 and 1 after T = 2.

Comparing figure 3-7(a),(b) with those of the STAR and STMA models, we can see that due to the interference of AR and MA process, the amplitude of the STARMA models is not equal to the superposition of amplitudes of the corresponding STAR and STMA models. For example the amplitudes in figure 3-7(b), in which the parameter values are set to $\phi_{10}= 0.4$, $\phi_{11} = 0.4$, $\theta_{10} = -0.2$, $\theta_{11} = -0.4$, $\theta_{12} = -0.2$, are 2.00, 1.50, 1.06, 0.75 at the (5,6) location at T = 1,2,3,4 respectively. The corresponding amplitudes in figure 3-6(a), which is a STAR model in which the parameter values are set to $\phi_{10}=0.4$, $\phi_{11}=0.4$, are 1.00, 0.80, 0.56, 0.36 and the corresponding amplitudes in figure 3-6(c), which is a STMA model in which the parameter values are set to $\theta_{10}=-0.2$, $\theta_{11}=-0.4$, $\theta_{12}=-0.2$, are 1.00, 0.00, 0.00, 0.00. It is obvious that from T=2 on, the amplitudes in figure 3-7(b) are not equal to the superposition of those corresponding amplitudes in figure 3-5(a) and figure 3-4(c). this phenomena is due to the interference of STMA process on the STAR process, which has been discused in previous section.

A STAR(1_1)I_m process that satisfies equation 3-23, i.e. the strictly conservative conditions, is simulated and plotted in figure 3-8. In this simulated relocaiton process, the process parameters are set to $\phi_{10}=0.4$, $\phi_{11}=0.6$, and $\sum_{i=1}^{121} W_{ij}^{(1)} = 1$ for j = 1,2,...,121, so that the strictly conservative conditions, equation (3-23) are satisfied. Comparing the figures of (3-8) with their corresponding figures of 3-3(a), we see that the diffusion speed of the relocation process, which is a homogeneously nonstationary STAR(1_1)I_m process, is the same as that of the stationary STAR(1_1)I_m process in figure 3-5(a), i.e. 1 order neighbor per observation period. But the diffusion amplitudes of

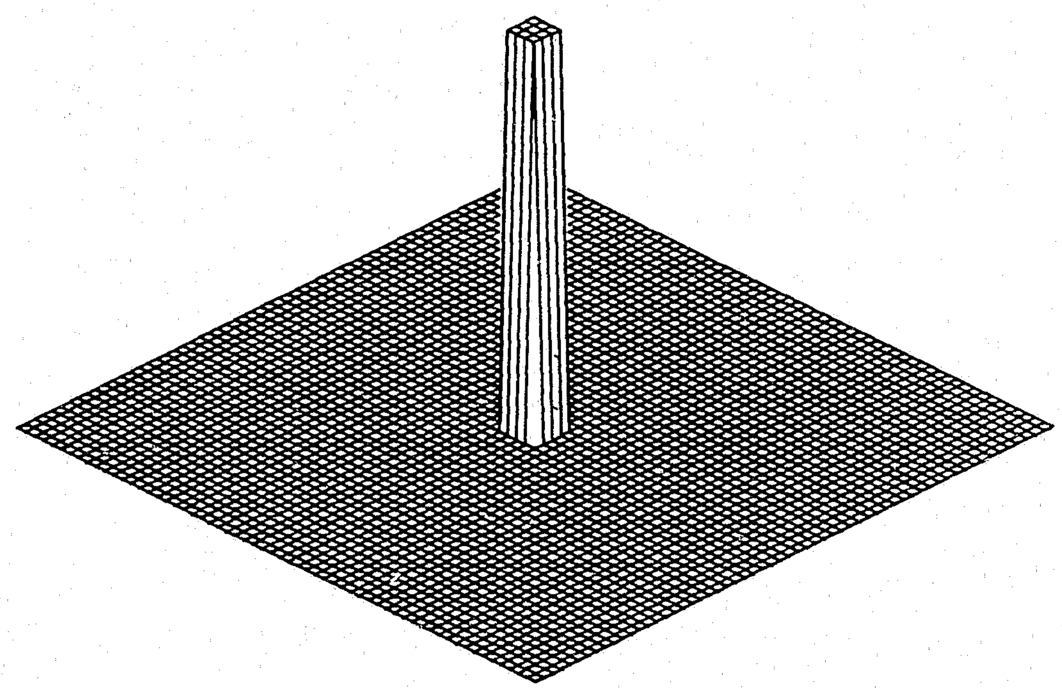


Figure 3-4. The Intervention Input $\delta=10$ at Location (6,6) of the 11X11 Regular Grid System.

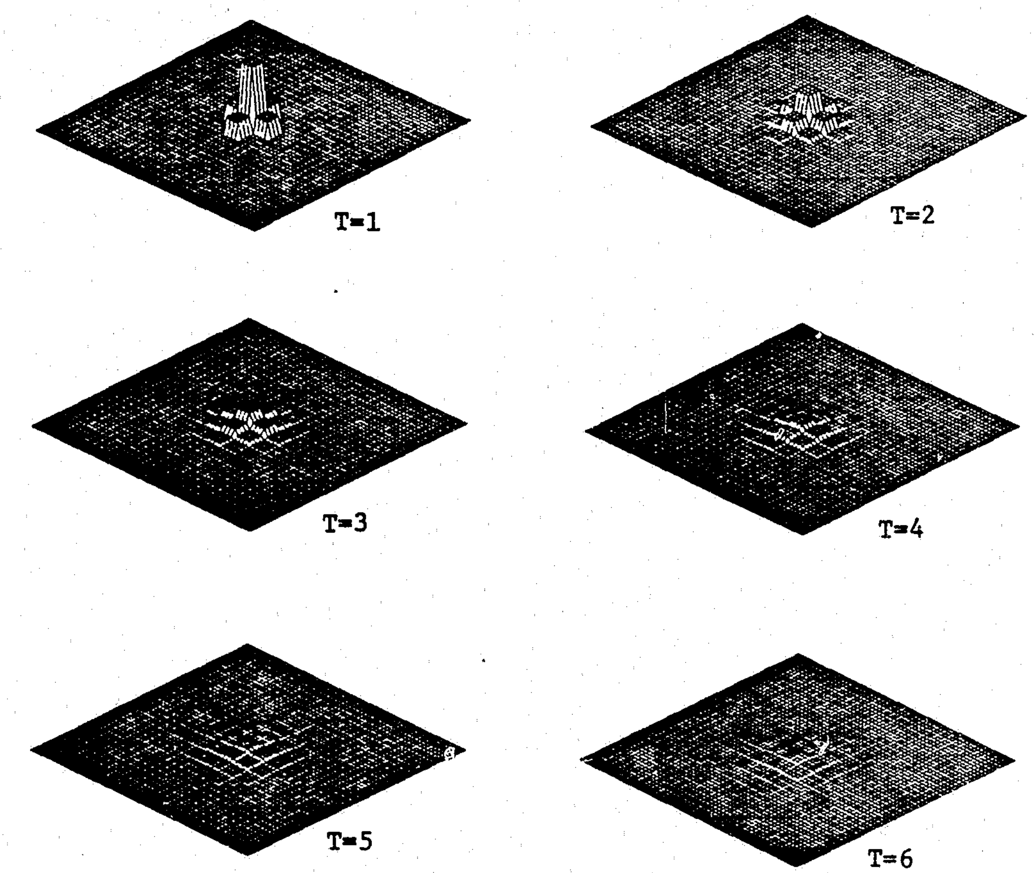


Figure 3-5(a). Diffusion Process of AR Type, STAR(1_1) Model with $\phi_{10}=0.4$, $\phi_{11}=0.4$

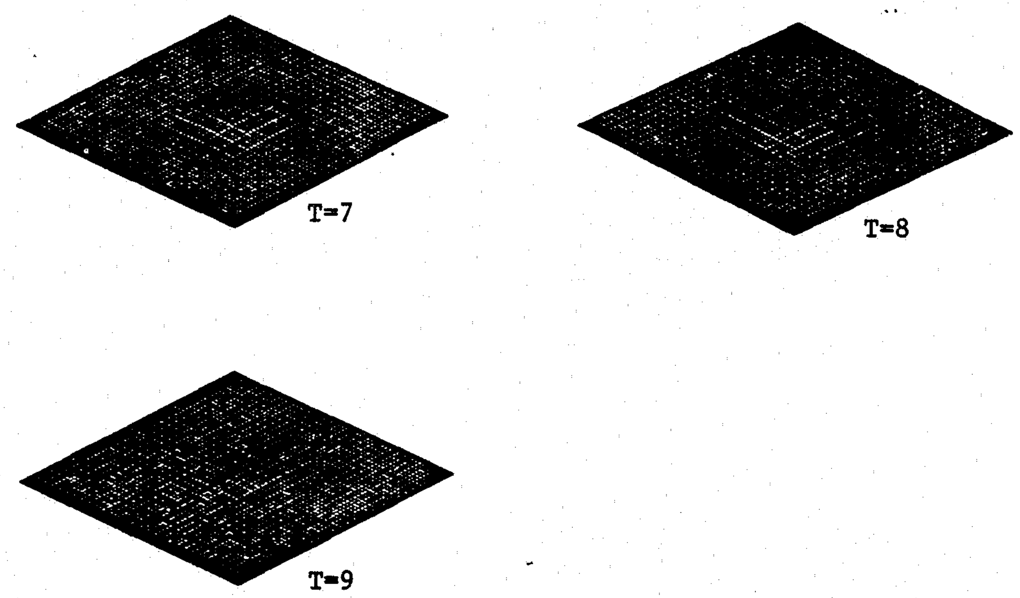


Figure 3-5(a). (Cont'd)

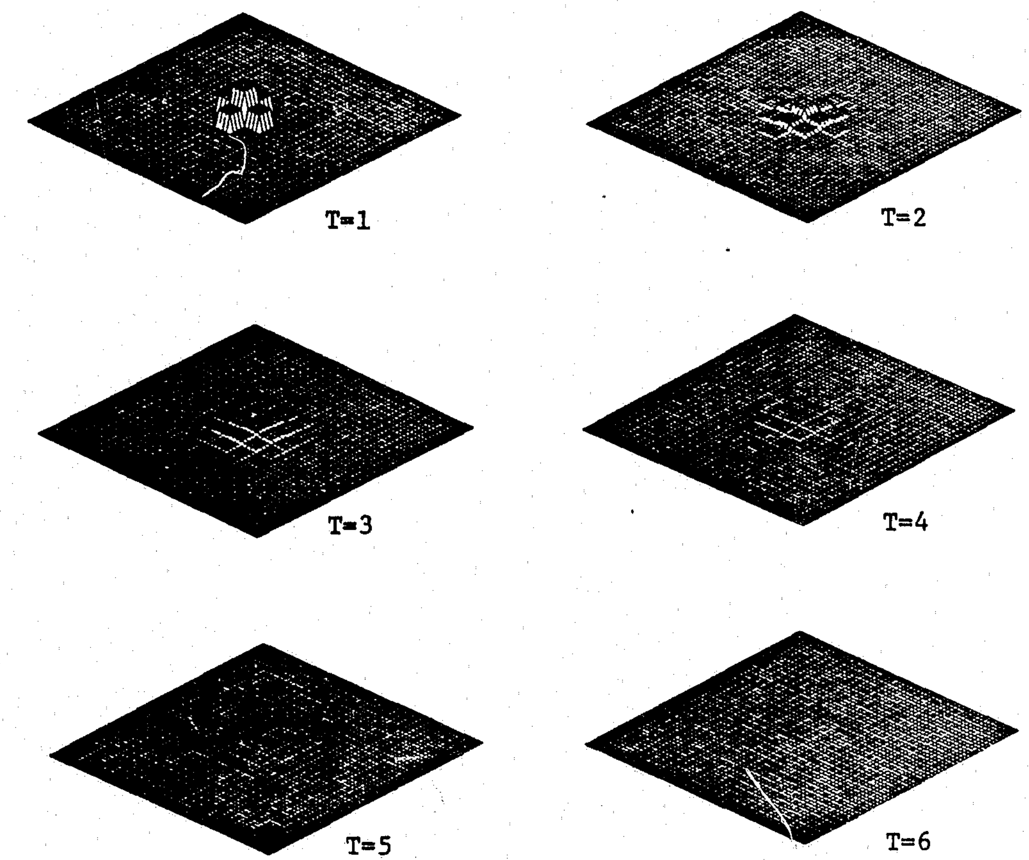


Figure 3-5(b). Diffusion Process of AR Type, STAR(1_1) Model with ϕ_{10}=0.2, ϕ_{11}=0.4

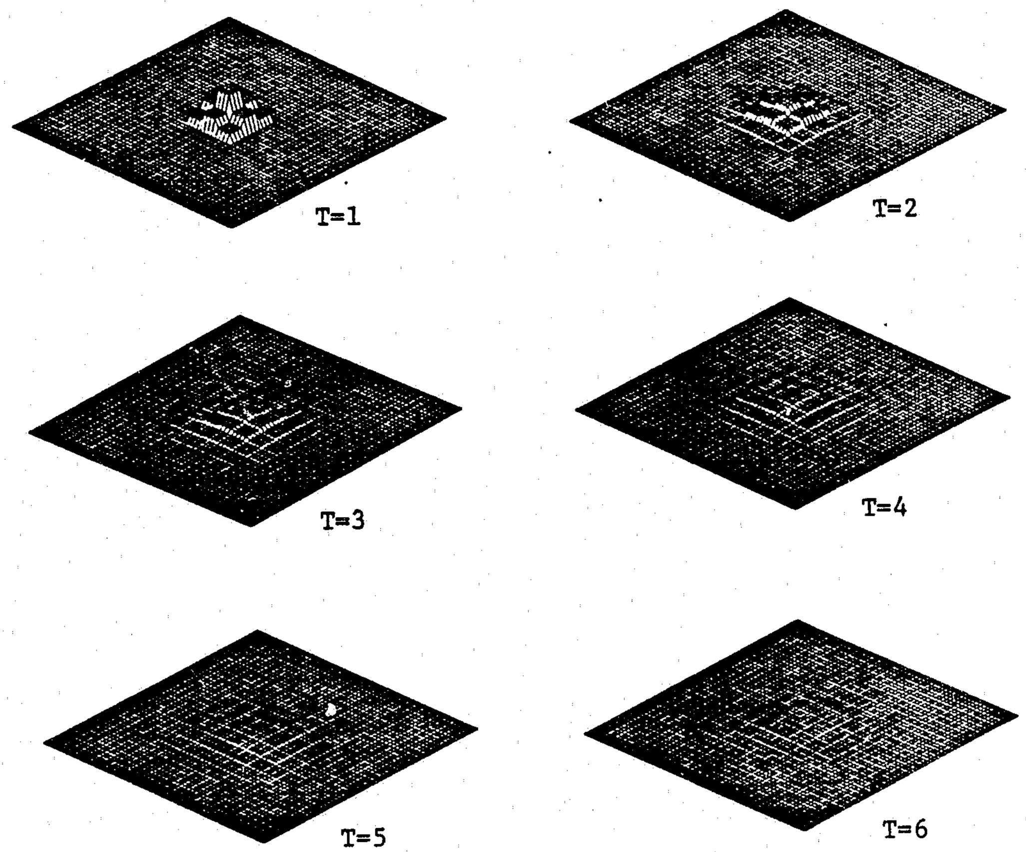


Figure 3-5(c). Diffusion Process of AR Type, STAR(1_1) Model with $\phi_{10}=0.2$, $\phi_{11}=0.4$, $\phi_{12}=0.2$

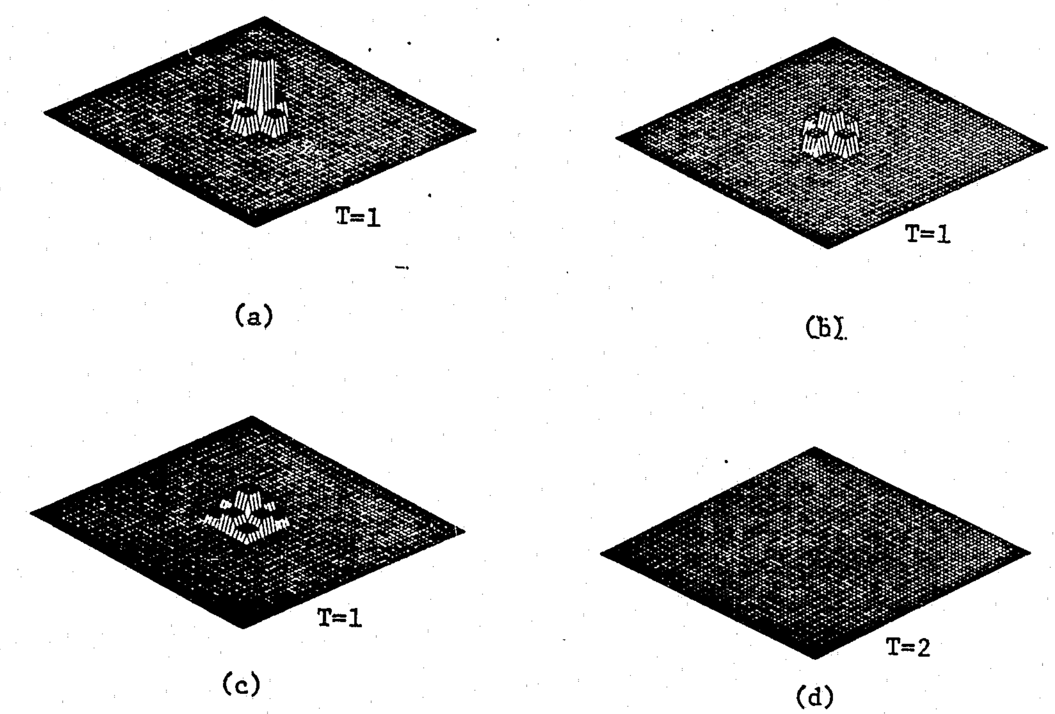


Figure 3-6. Diffusion Process of MA Type
(a) STMA(1_1) Model with $\theta_{10}=-0.4$, $\theta_{11}=-0.4$
(b) STMA(1_1) Model with $\theta_{10}=-0.2$, $\theta_{11}=-0.4$
(c) STMA(1_2) Model with $\theta_{10}=-0.2$, $\theta_{11}=-0.4$, $_{12}=-0.2$
(d) T=2 for Models in (a,b,c)

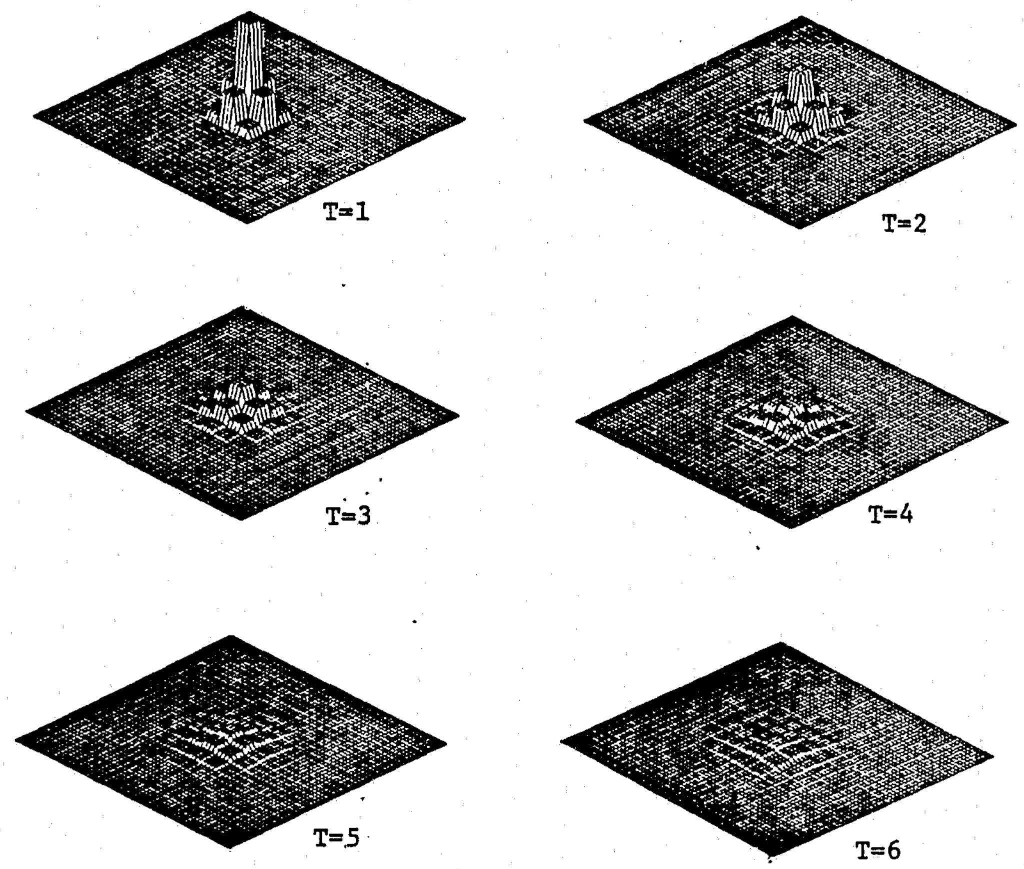


Figure 3-7(a). Diffusion Process of ARMA Type, STARIMA(1_2,0,1_1) Model with ϕ_{10}=0.2, ϕ_{11}=0.4, ϕ_{12}=0.2, θ_{10}=-0.4, θ_{11}=-0.4

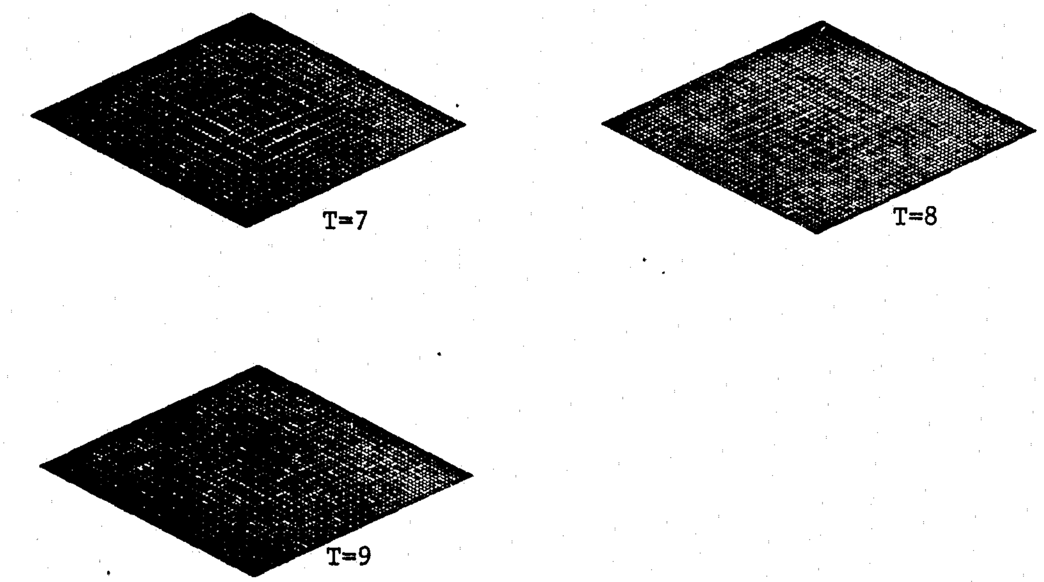


Figure 3-7(a). (Cont'd)

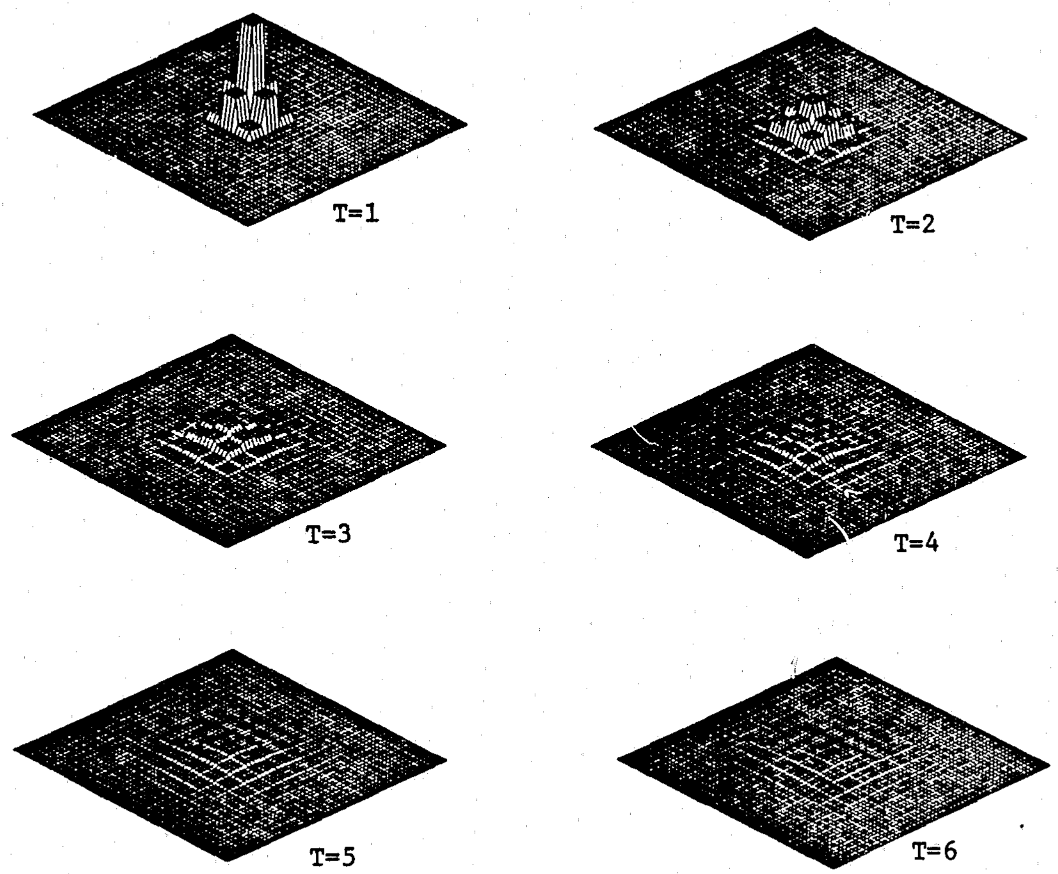


Figure 3-7(b). Diffusion Process of ARMA Type, STARIMA$(1_1,0,1_2)$ Process with $\phi_{10}=0.4$, $\phi_{11}=0.4$, $\theta_{10}=-0.2$, $\theta_{11}=-0.4$, $\theta_{12}=-0.2$

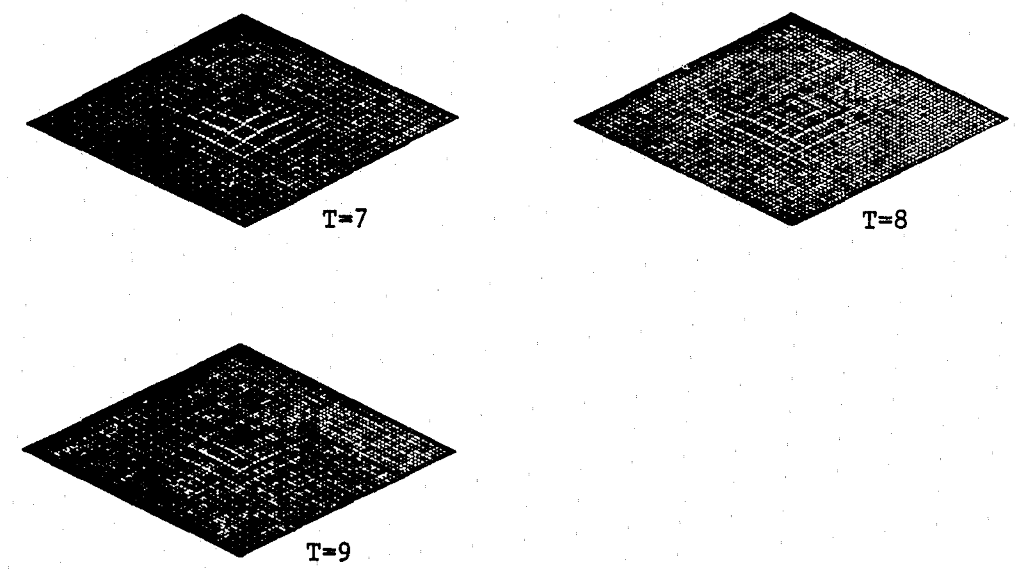


Figure 3-7(b). (Cont'd)

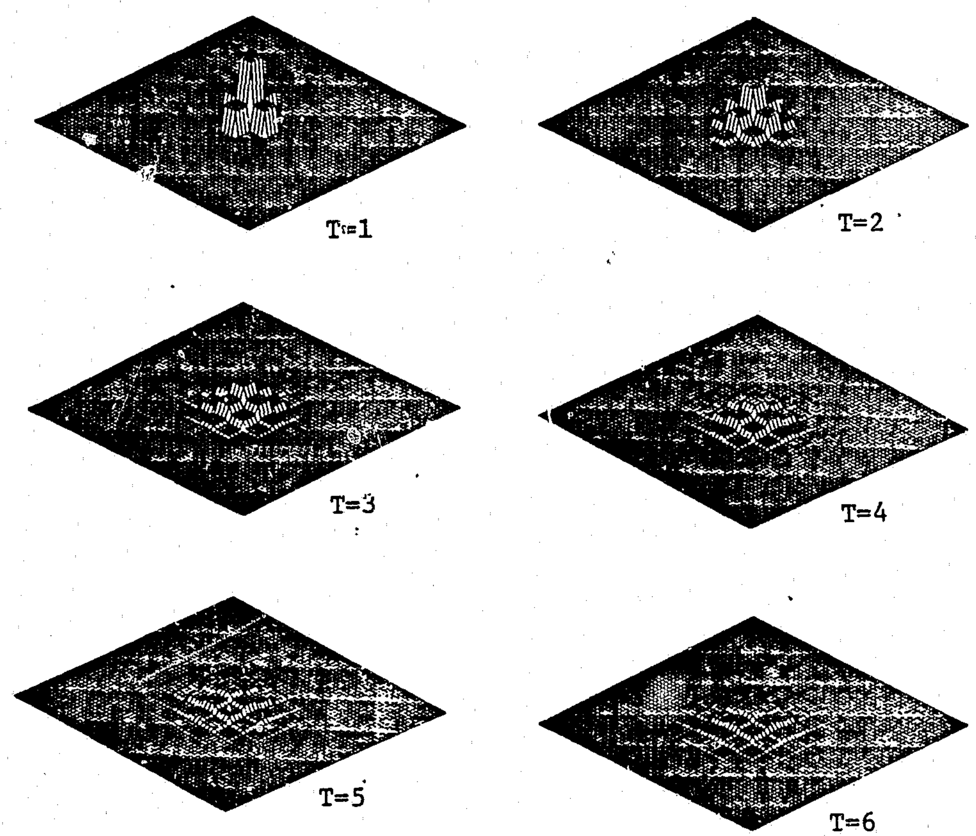


Figure 3-8. Diffusion Process of Relocation Type
$\phi_{10}=0.4$, $\phi_{11}=0.6$

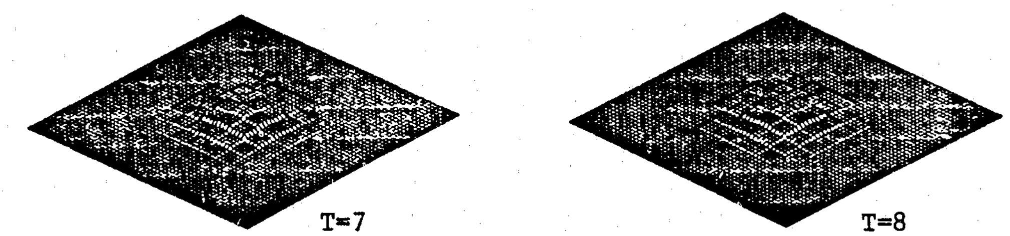


Figure 3-8. (Cont'd)

the relocation process are always higher than those of the stationary STAR diffusion process, which is a regenerating (-) diffusion process. From T=3 on it becomes even more obvious that the populations at every influenced locations of the relocation process are much higher than those in the correspoding regenerating (-) process. Since the total population at T=0, which is the intervention input, almost damps out in figure 3-5(a) at T=3, while the total population of the relocation process in figure 3-8 is conserved for all time, the longer the time elapsed, the greater the differences in the diffusion amplitude will be.

3.4 Modeling Space-Time Intervention Processes

In previous sections, we investigate the properties of the space-time intervention models. In this section, we describe how to build the space-time intervention model for a process. We still assume that there are n_1 pre-intervention observations and n_2 post-intervention observations.

We will build the space-time intervention model following the three steps listed below:

1. Build the model for the pre-intervention space-time process.
2. Build the dynamic model for the effect of the intervention. An important component of this step is to identify whether the intervention process is influenced by the environment or not and to identify the form of the impact from the data structure.
3. Estimation of the parameters of the total model and diagnostic checking of its adequacy.

To build the STARIMA model, first the space-time sample autocorrelation functions and the partial autocorrelation functions are computed and are used in accordance with the cut-off tail-off properties to select a candidate model. Then the model parameters are estimated and diagnostic checking is applied to check the adequacy of the proposed model. If the diagnostic checking doesn't show any inadequacy, the model is ready to be employed, otherwise the candidate model will be updated according to the remaining structure, then the parameters of the updated model are estimated. This iterative procedure is employed until an adequate model is found. These procedures and inferential statistics for building the pre-intervention model are completely described in Deutsch and Pfeifer (1980).

In building the dynamic component we employ the preintervention model to sequentially estimate $\underset{\sim}{\delta}(t)$ for each observations $t > n_1$ and correct these original observations by subtracting $\underset{\sim}{\delta}(t)$ when it is significant. We will denote $Z_{i,t}$ as the observation of location i at time t and $Z^c_{i,t} = Z_{i,t} - \delta_i(t)$ when $\delta_i(t)$ is significant, $Z^c_{i,t} = Z_{i,t}$ when $\delta_i(t)$ is nonsignificant, where $\delta_i(t)$ is the estimated mean shift at time t, location i. Thus, in this dynamic scheme, we start with using the n_1 observations to estimate $\underset{\sim}{\phi}$, $\underset{\sim}{\Theta}$, $\underset{\sim}{\mu}$ and correct all the observations (pre- and post-intervention) by subtracting $\underset{\sim}{\mu}$. Setting $n_2=1$ we use the $t<n_1+1$ to estimate $\underset{\sim}{\delta}(n_1+1)$ and correct Z_{i,n_1+1} if $\delta_i(n_1+1)$ is significant, and set $n_1+1 \to n_1$, repeating this procedure until all postintervention observations are exhausted. A detailed flow chart of this procedure is given in figure 3-9.

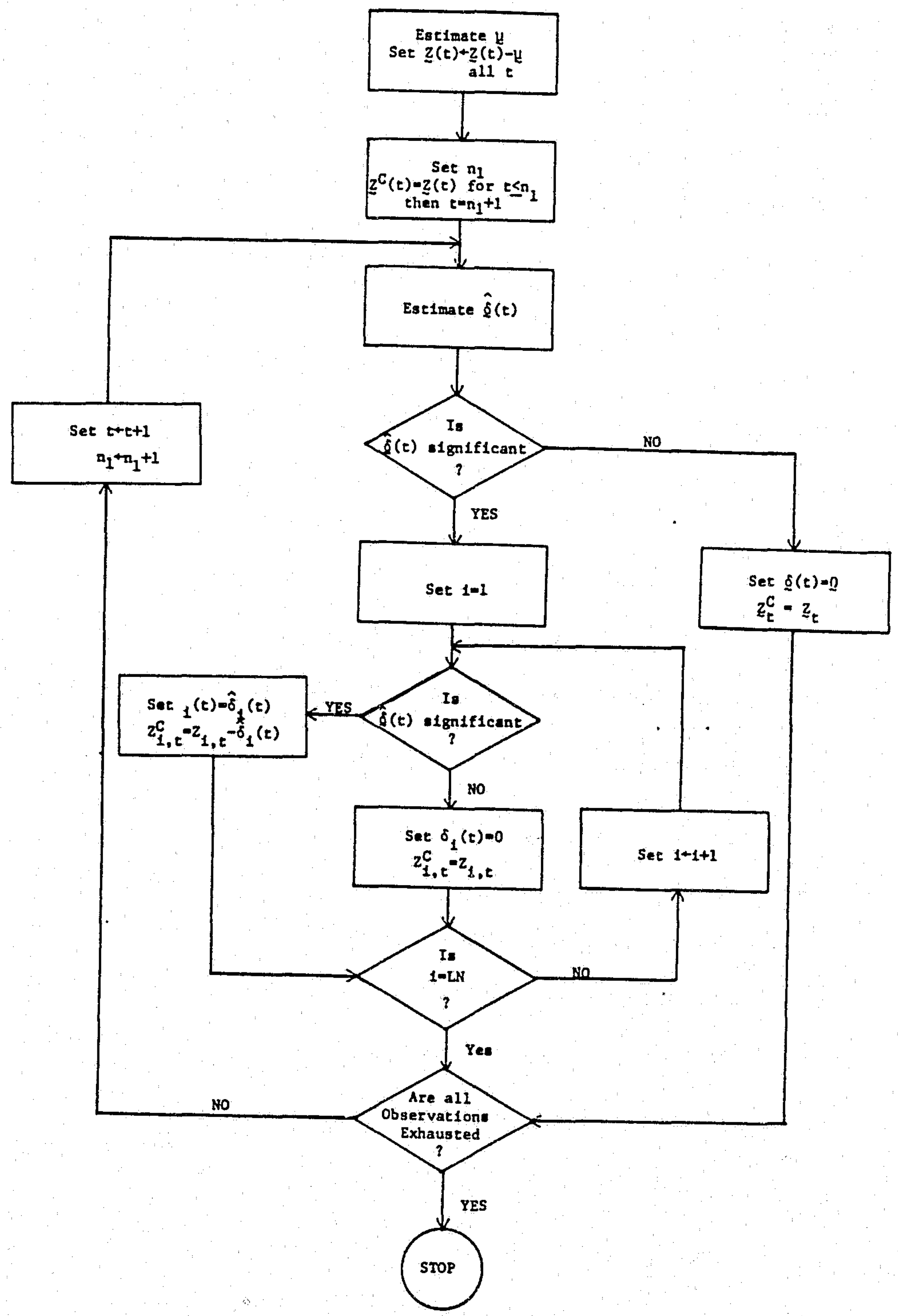


Figure 3-9. Procedure of Dynamic Model Building.

As shown on the flow chart, we use the n_1 observations, i.e., $Z_{i,t}$, $t < n_1$, $i = 1,2,\ldots,LN$, to determine the model class, i.e., STARMA, STAR or STMA, and to estimate $\underset{\sim}{\phi}$, $\underset{\sim}{\Theta}$. For each estimate of $\underset{\sim}{\delta}(t)$, we perform the hypothesis test $H_0 : \underset{\sim}{\delta}(t) = \underset{\sim}{0}$ vs. $H_1 : \underset{\sim}{\delta}(t) \neq \underset{\sim}{0}$. If H_0 is rejected, we perform all the LN hypothesis-testing $H_{0i} : \delta_i(t) = 0$ vs. $H_{1i} : \delta_i(t) \neq 0$, $i = 1,2,\ldots,LN$, for each individual location and set $\delta_i(t) = \delta_i(t)$, $Z^c_{i,t} = Z_{i,t} - \delta_i(t)$, if H_{0i} is rejected. Set $\delta_i(t) = 0$ thus $Z^c_{i,t} = Z_{i,t}$ if H_{oi} is not rejected. While if H_0 is not rejected, then $\underset{\sim}{\delta}(t) = \underset{\sim}{0}$ and $\underset{\sim}{Z}^c_t = \underset{\sim}{Z}_t$.

From this sequential procedure, the plot of $\underset{\sim}{\delta}(t)$ vs. t allows for the determination of whether there is environmental influence in the intervention process. If the plot reveals no transient behavior and is deterministic, no environmental influence to the mean shift function is present. Alternatively, when these characteristics are present the parameter estimators for the intervention process are those correspoding to $I_m = 1$ structure for model fitting using all n_2 postintervention data. Recall from Chapter II that if $n_2 > 1$ estimates of $\underset{\sim}{\delta}(t)$ and therefore, the intrinsic utility of a program $\underset{\sim}{\delta}$ will be biased if the incorrect structure $I_m = 0$ or $I_m = 1$ is employed.

3.5 Estimation for Space-Time Intervention Models

Once the pre-intervention space-time model is built in step one of the modeling procedures, the model parameters, i.e. $(\underset{\sim}{\phi},\underset{\sim}{\Theta})$ which have already been estimated may be treated as known in determing estimates of the other parameters. However, sometimes we do not have complete confidence in the correctness of these estimates perhaps due to the smallness of the length of the preintervention history, n_1. In this case it is desirable to refine the whole intervention model by estimating all the model parameters simultaneously, eg. treating all parameters as unknown. In this section, we first will develop the formula for estimating $(\underset{\sim}{\mu},\underset{\sim}{\delta})$ given that $(\underset{\sim}{\phi},\underset{\sim}{\Theta})$ known. Then we move to the situation in which $(\underset{\sim}{\phi},\underset{\sim}{\Theta})$ are treated as unknown. Here the L.S. estimates of $(\underset{\sim}{\mu},\underset{\sim}{\delta},\underset{\sim}{\phi},\underset{\sim}{\Theta})$ are gotten by searching through the $(\underset{\sim}{\phi},\underset{\sim}{\Theta})$ subspace, and the approximate hypotheses testing statistics, the approximate confidence intervals are developed based on linearization. Lastly, the estimation for the multiconsequence space-time intervention process is treated. Here in addition to the intervention potentially causing a change in the mean level of the process at any of the LN locations a simultaneous change in the covariance can occur causing the preintervention parameters $\underset{\sim}{\phi},\underset{\sim}{\Theta}$ to change to $\underset{\sim}{\psi},\underset{\sim}{\gamma}$ after the intervention. For this situation, the conditional estimation of $(\underset{\sim}{\psi},\underset{\sim}{\gamma},\underset{\sim}{\mu},\underset{\sim}{\delta}|\underset{\sim}{\phi},\underset{\sim}{\Theta})$ are discussed. In the rest of this section, we will assume that ξ_t, i.e. the intervention variable, is well identified and is known.

3.5.1 Transformation to Linear Model Form

In this section we assume that the preintervention model parameters, $\underset{\sim}{\phi}$ and $\underset{\sim}{\Theta}$, are known. A recursive formula is developed to transform the original STARIMA$(P_{\lambda_1,\ldots,\lambda_p},0,q_{m_1,\ldots,m_q})I_m$ model into linear model form, i.e.

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\rho} + \underset{\sim}{\varepsilon} \qquad (3\text{-}33)$$

where

$\underset{\sim}{Y}$ is $[LN \times (n_1+n_2)]$ by 1 vector with each element as a function of the observations and model parameters.

X is $[LN \times (n_1+n_2)]$ by 2 LN matrix with each element as a function of model parameters only.

$$\underset{\sim}{\rho}^t = [\underset{\sim}{\mu}^t,\underset{\sim}{\delta}^t] = [\mu_1,\mu_2,\ldots,\mu_{LN},\delta_1,\delta_2,\ldots,\delta_{LN}]$$

μ_i, $i = 1,2,\ldots,LN$ is the preintervention mean value at location i.

δ_i, $i = 1,2,\ldots,LN$ is the intrinsic mean shift at location i and

$$\underset{\sim}{\varepsilon} \sim N(0,I\sigma_a^2)$$

In the linear model form,

$$
X = \left[\begin{array}{cccc|ccc}
X_{1,1}(1) & X_{1,2}(1) & \cdots & X_{1,LN}(1) & X_{1,LN+1}(1) & \cdots & X_{1,2LN}(1) \\
X_{2,1}(1) & X_{2,2}(1) & \cdots & X_{2,LN}(1) & X_{2,LN+1}(1) & \cdots & X_{2,2LN}(1) \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
X_{LN,1}(1) & X_{LN,2}(1) & \cdots & X_{LN,LN}(1) & X_{LN,LN+1}(1) & \cdots & X_{LN,2LN}(1) \\
X_{1,1}(2) & X_{1,2}(2) & \cdots & X_{1,LN}(2) & X_{1,LN+1}(2) & & X_{1,2LN}(2) \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
X_{LN,1}(n_1) & X_{LN,2}(n_1) & \cdots & X_{LN,LN}(n_1) & X_{LN,LN+1}(n_1) & \cdots & X_{LN,2LN}(n_1) \\
\hline
X_{1,1}(n_1+1) & X_{1,2}(n_1+1) & \cdots & X_{1,LN}(n_1+1) & X_{1,LN+L}(n_1+1) & \cdots & X_{1,2LN}(n_1+1) \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
X_{LN,1}(n_1+n_2) & X_{LN,2}(n_1+n_2) & \cdots & X_{LN,LN}(n_1+n_2) & X_{LN,LN+1}(n_1+n_2) & \cdots & X_{LN,2LN}(n_1+n_2)
\end{array}\right]
$$

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1(1) \\ Y_2(1) \\ \vdots \\ Y_{LN}(1) \\ Y_1(2) \\ \vdots \\ Y_{LN}(n_1+n_2) \end{bmatrix} \qquad \underset{\sim}{\rho} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{LN} \\ \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{LN} \end{bmatrix} \qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1(1) \\ \varepsilon_2(1) \\ \vdots \\ \varepsilon_{LN}(1) \\ \varepsilon_1(2) \\ \vdots \\ \varepsilon_{LN}(n_1+n_2) \end{bmatrix}$$

The closed form transformation formula for the STARMA$(1_{\lambda 1},0,1_{m1})I_m$ model which will be derived is also applicable to either the STAR$(1_\lambda,0,0)I_m$ model or the STMA$(0,0,1_{m1})I_m$ model by deleting certain terms by setting the appropriate model parameters to zero. Once we have the transformed linear form, we will apply the results of linear model theory, e.g. F. A. Graybill [1976], to get the L.S. estimator for $\underset{\sim}{\mu}$, $\underset{\sim}{\delta}$, and construct the hypotheses testing statistics and confidence intervals.

The STARMA$(P_{\lambda 1,\ldots,\lambda p},0,q_{m1,\ldots,mq})I_m$ model form is,

$$\underset{\sim}{Z}_t = S_t^{-\infty}\underset{\sim}{\mu} + \xi_t\underset{\sim}{\delta} + \sum_{K=1}^{P}\sum_{\ell=0}^{\lambda_K}\phi_{K\ell}W^{(\ell)}B^K(\underset{\sim}{Z}_t - S_t^{-\infty}\underset{\sim}{\mu} - (1-I_m)\xi_t\underset{\sim}{\delta}) \tag{3-34}$$

$$- \sum_{K=1}^{q}\sum_{\ell=0}^{m_K}\Theta_{K\ell}W^{(\ell)}B^K(\underset{\sim}{\varepsilon}_t + I_m\,\xi_t\,\underset{\sim}{\delta}) + \underset{\sim}{\varepsilon}_t$$

where $S_t^{-\infty}$ is step function and ξ_t is the intervention variable.
Equation 3-34 can also be expressed as,

$$\underset{\sim}{\Phi}_{p,\lambda}(B)\underset{\sim}{Z}_t = \underset{\sim}{\Phi}_{p,\lambda}(B)\, S_t^{-\infty}\, \underset{\sim}{\mu} \tag{3-35}$$

$$+ \left((1-I_m)\underset{\sim}{\Phi}_{p,\lambda}(B)+I_m\, \underset{\sim}{\Theta}_{q,m}(B)\right) \xi_t\, \underset{\sim}{\delta} + \underset{\sim}{\Theta}_{q,m}(B)\, \underset{\sim}{\varepsilon}_t.$$

By multiplying every term by $\underset{\sim}{\Theta}_{q,m}(B)^{-1}$ we get the lienar model form of,

$$\frac{\underset{\sim}{\Phi}_{p,\lambda}(B)}{\underset{\sim}{\Theta}_{q,m}(B)}\underset{\sim}{Z}_t = \frac{\underset{\sim}{\Phi}_{p,\lambda}(B)}{\underset{\sim}{\Theta}_{q,m}(B)} S_t^{-\infty}\underset{\sim}{\mu} + \frac{(1-I_m)\underset{\sim}{\Phi}_{p,\lambda}(B)+I_m\underset{\sim}{\Theta}_{q,m}(B)}{\underset{\sim}{\Theta}_{q,m}(B)}\xi_t\,\underset{\sim}{\delta} + \underset{\sim}{\varepsilon}_t. \tag{3-36}$$

By defining

$$\underset{\sim}{Y}(t) = \frac{\underset{\sim}{\Phi}_{p,\lambda}(B)}{\underset{\sim}{\Theta}_{q,m}(B)}\underset{\sim}{Z}_t. \tag{3-37}$$

$$X_1(t) = \frac{\underset{\sim}{\Phi}_{p,\lambda}(B)}{\underset{\sim}{\Theta}_{q,m}(B)} S_t^{-\infty}. \tag{3-38}$$

$$X_2(t) = \frac{(1-I_m)\underset{\sim}{\Phi}_{p,\lambda}(B) + I_m\underset{\sim}{\Theta}_{q,m}(B)}{\underset{\sim}{\Theta}_{q,m}(B)}\xi_t. \tag{3-39}$$

where $\underset{\sim}{Y}(t)$ is a vector and $X_1(t)$, $X_2(t)$ are LN by LN matrix, the model is reparameterized to the standard linear model form;

$$\underset{\sim}{Y}(t) = X_1(t)\,\underset{\sim}{\mu}_2 + X_2(t)\,\underset{\sim}{\delta} + \underset{\sim}{\varepsilon}_t. \tag{3-40}$$

where $\underset{\sim}{Y}(t)$, $X_1(t)$, $X_2(t)$ of given STARMA$(\underset{\sim}{P}_\lambda,0,\underset{\sim}{q}_m)I_m$ model can be derived by using equations 3-37, 3-38, 3-39 recursively.

In the following, we will use ξ_t as a pulse function, $P_t^{n_1}(T_p)$, without the loss of generality since, if we let $T_p = \infty$, the pulse function becomes a step function. Also implicitly, we assume that $n_1 > \mathrm{Max}(p,q)$. To obtain the initial values for $\underset{\sim}{Y}(t)$ and $X_1(t)$, we will replace the unrealized values of $\underset{\sim}{Z}_t$ by their expected values, e.g. $\underset{\sim}{Z}_t|_{t<0} = \underset{\sim}{\mu}$. Thus, $X_2(t) = 0$ for $t < n_1$.

Based on thes initial conditions, we will develop a general transformation formula by using equations 3-37, 3-38, 3-39.
Define

$$W_{\phi K} = \sum_{\ell=0}^{\lambda_K} \phi_{K\ell} W^{(\ell)} \qquad K = 1,2,\ldots,P$$

$$W_{\theta K} = \sum_{\ell=0}^{m_K} \theta_{K\ell} W^{(\ell)} \qquad K = 1,2,\ldots,q$$

$$W_{\psi K} = \sum_{\ell=0}^{\lambda_K} \psi_{K\ell} W^{(\ell)} \qquad K = 1,2,\ldots,P$$

$$W_{\gamma K} = \sum_{\ell=0}^{m_K} \gamma_{K\ell} W^{(\ell)} \qquad K = 1,2,\ldots,q$$

From equation 3-37,

$$\underset{\sim}{\theta}_{q,m}(B)\ \underset{\sim}{Y}(t) = \underset{\sim}{\phi}_{p;\lambda}(B)\ \underset{\sim}{Z}_t$$

or

$$\underset{\sim}{Y}(t) = \sum_{K=1}^{q} W_{\theta K} B^K \underset{\sim}{Y}(t) + \underset{\sim}{Z}_t - \sum_{K=1}^{P} W_{\phi K} B^k \underset{\sim}{Z}_t. \tag{3-41}$$

Imposing the initial conditions, we get;

$$\underset{\sim}{Y}(t) = \begin{cases} (I - \sum_{K=1}^{q} W_{\theta K})^{-1} (I - \sum_{K=1}^{P} W_{\phi K})\ \underset{\sim}{\mu} & t < 0 \\ \sum_{K=1}^{q} W_{\theta K}\ \underset{\sim}{Y}(t-K) + \underset{\sim}{Z}_t - \sum_{K=1}^{P} W_{\phi K}\ \underset{\sim}{Z}_{t-K} & t > 0 \end{cases} \tag{3-42}$$

From equation 3-38

$$\underset{\sim}{\theta}_{q,m}(B)\ X_1(t) = \underset{\sim}{\phi}_{p,\lambda}(B)\ S_t^{-\infty}$$

or

$$X_1(t) = \sum_{K=1}^{q} W_{\theta K} B^K X_1(t) + I - \sum_{K=1}^{P} W_{\phi K}. \tag{3-43}$$

Again imposing the initial condition, we have,

$$X_1(t) = \begin{cases} (I - \sum_{K=1}^{q} W_{\theta K})^{-1} (I - \sum_{K=1}^{P} W_{\phi K}) & t < 0. \\ \sum_{K=1}^{q} W_{\theta K}\ X_1(t-K) + I - \sum_{K=1}^{P} W_{\phi K}. & t > 0. \end{cases} \tag{3-44}$$

From equation 3-33

$$\underset{\sim}{\theta}_{q,m}(B)\ X_2(t) = \left((1-I_m)\ \underset{\sim}{\phi}_{p,\lambda}(B) + I_m\ \underset{\sim}{\theta}_{q,m}(B)\right)\ P_t^{n_1}(T_p).$$

or

$$X_2(t) = \sum_{K=1}^{q} W_{\theta K} B^K X_2(t)$$

$$+ \left(I - (i-I_m) \sum_{K=1}^{P} W_{\phi K} B^K - I_m \sum_{K=1}^{q} W_{\theta K} B^K\right) P_t^{n_1}(T_p). \tag{3-45}$$

Using the stated conditions, we have

$$X_2(t) = \begin{cases} 0 & t < n_1 \\ \sum_{K=1}^{q} W_{\Theta K}\, X_2(t-K) + P_t^{n_1}(T_p) - (1-I_m) \sum_{K=1}^{P} W_{\phi K}\, P_t^{n_1}(Tp) & \\ \quad - I_m \sum_{K=1}^{q} W_{\Theta K}\, P_{t-K}^{n_1}(Tp) & t > n_1 \end{cases} \qquad (3\text{-}46)$$

The estimator $\underset{\sim}{\rho}$ from linear model theory is unbiased if the X_1 and X_2 matrix are constructed correctly. However if we choose the wrong model structure, i.e., the wrong I_m parameter value, the estimates will not only be biased, but simply incorrect.

This transformation to lienar model form can be made more compact and computationally simpler to give the transformation formula of STARMA $(1_{\lambda 1},0,1_{m1})I_m$ model by defining the iterative functions, $L_{i,Z}(t)$ and $U_{ij}^{(K)}$.

$$L_i(t) = \begin{cases} 0 & \text{for } t < 1 \\ \sum_{j=1}^{LN} \left| \sum_{\ell=0}^{\lambda_1} \phi_{1\ell} W_{ij}^{(\ell)} Z_{j,t} - \sum_{\ell=0}^{m_1} \Theta_{1\ell} W_{ij}^{(\ell)} \left(Z_{j,t} - L_{j,Z}(t-1)\right) \right| & \text{for } 2 < t < n_1 + n_2 \end{cases} \qquad (3\text{-}47)$$

$$i = 1,2,\ldots,LN$$

$$U_{ij}^{(K)} = \begin{cases} \sum_{\ell=0}^{\lambda_1^0} \phi_{1\ell} W_{ij}^{(\ell)} - \sum_{\ell=0}^{m_1} \Theta_{i\ell} W_{ij}^{(\ell)} & K=0 \\ 0 & K \leq -1 \\ \sum_{\ell=0}^{m_1} \sum_{g=1}^{LN} \Theta_{1\ell} W_{ig}^{(\ell)} U_{gj}^{(K-1)} & K > 1 \end{cases} \qquad (3\text{-}48)$$

Using these definitions, the transformation formula becomes

For $i,j = 1,2,\ldots,LN$.

$$Y_i(t) = X_{i,t} - L_{i,8}(t) \qquad 1 < t < n_1+n_2 \qquad (3\text{-}49)$$

$$X_{ii}(t) = 1 - \sum_{k=0}^{t-2} u_{ii}(k) \qquad 1 < t < n_1+n_2 \qquad (3\text{-}50)$$

$$X_{ij}(t) = \sum_{k=0}^{t-2} U_{ij}(k) \qquad 1 < t < n_1+n_2,\ i \neq j \qquad (3\text{-}51)$$

$$X_{i,i+LN}(t) = \begin{cases} 0 & t < n_1 \\ 1 - I_m \sum_{k=0}^{t-n_1-2} U_{i,i}(k) & n_1+1 < t n_1 + T_p \\ -I_m \sum_{k=t-n_1-T_p-1}^{t-n_2-2} U_{i,i}(k) & t < n_1 T_p + 1 \end{cases} \qquad (3\text{-}52)$$

$$X_{i,j+LN}(t) = \begin{cases} 0 & t < n_1+1 \\ -I_m \sum_{K=0}^{t-n_1-2} U_{ij}^{(K)} & n_1+2 < t < n_1+T_p \\ -I_m \sum_{K=t-n_1-n_p-1}^{t-n_1-2} U_{ij}^{(K)} & n_1+T_p+1 < t \end{cases} \qquad (3\text{-}53)$$

To obtain the transformation formula for the situation that ξ_t is presumed to be a step function. We just set $T_p = n_2$ or equivalently, $T_p = \infty$. Not that T_p appears only in equations (3-52), (3-53), so only these equations will be changed. For the step function situation therefore,

$$X_{i,i+LN}(t) = \begin{cases} 0 & t < n_1 \\ 1 - I_m \sum_{K=0}^{t-n_1-2} U_{ij}^{(K)} & n_1+1 < t < n_1 n_2 \end{cases} \qquad (3\text{-}54)$$

and

$$\bar{X}_{i,j+LN}(t) = \begin{cases} 0 & t < n_1+1 \\ -I_m \sum_{K=0}^{t-n_1-2} U_{ij}^{(K)} & t > n_1+2 \end{cases} \qquad (3\text{-}55)$$

Note that equations (3-54), (3-55) are obtained by eliminating the last

part, i.e. $t > n_1+T_p+1$, of equations (3-52), (3-53). To obtain the transformation formula for the STAR(1_{λ_1})I_m intervention model and the STMA(1_{m_1})I_m intervention model, we only set $\tilde{\Theta} = \tilde{0}$ and $\tilde{\phi} = \tilde{0}$, respectively in the derived pulse or step formulation.

3.5.2 Least Squares Estimates and Hypothesis Tests for $\tilde{\phi}$, $\tilde{\Theta}$ Known

In order to apply the result of linear model theory to get the M.L. estimators, it is important to be sure that the X-matrix of the transformed form of all these models are of full rank (of rank 2LN). This can be seen by noting that $X_{i,i}(1) = 1$, $X_{i,j}(1) = 0$, $i \neq j$, $X_{i,i+LN}(t<n_1) = 0$, $X_{i,i+LN}(n_1+1) = 1$, $X_{i,j+LN}(t<n_1+1) = 0$, $i \neq j$. The X-matrix is of the form

$$X = \left[\begin{array}{cc} I_{LN \times LN} & \\ X_1(2) & \\ X_1(3) & 0 \\ \vdots & \\ X_1(n_1) & \\ \hline X_1(n_1+1) & I_{LN\times LN} \\ X_1(n_1+2) & X_2(n_1+2) \\ \vdots & \vdots \\ X_1(n_1 n_2) & X_2(n_1+n_2) \end{array}\right]$$

where I_{LNxLN} is LN by LN identity matrix, 0 is $(LNxN_1)$ by LN null matrix and $X_1(t>2)$, $X_2(t>n_1+2)$ are LN by LN matrices as stated in equation (3-40). Therefore the transformed X-matrix always consists of 2 LN independent columns and is of 2 LN rank.

Assume that all the ϕ's and Θ's are known, then all the $X_{i,j}$'s in the transformed model would be known. Applying the results from the linear model, we can immediately get the least square estimation for $\underset{\sim}{\rho}$ as

$$\underset{\sim}{\rho} = X^{\dagger} \underset{\sim}{Y} \tag{3-56}$$

where $X^{\dagger}$ is the generalized inverse of matrix X. That is, when X, the $[LNx(n_1+n_2)] \times (2LN)$ matrix, is of rank 2 LN then $X^{\dagger} = (X^1X)^{-1}X^1$ and the L.S. estimator becomes $\underset{\sim}{\rho} = (X'X)^{-1}X'\underset{\sim}{Y}$. Also, the sampling distribution of the quantity,

$$\frac{\rho_i - \rho_i}{\left[\frac{(\underset{\sim}{Y}-\underset{\sim}{Y})'(\underset{\sim}{Y}-\underset{\sim}{Y})C^{ii}}{M-2LN}\right]^{1/2}} \qquad \begin{array}{l} M = (n_1+n_2)(LN) \\ c^{ii} = \text{the ith diagonal} \\ \quad\ \text{element of } (X'X)^{-1} \end{array} \tag{3-57}$$

is student t-distribution. Therefore, the one-at-a-time $100(1-\alpha)\%$ confidence interval is,

$$\rho_i \pm t_{\alpha/2,(M-2LN)} \left[\frac{(\underset{\sim}{Y}-\underset{\sim}{Y})'(\underset{\sim}{Y}-\underset{\sim}{Y})C^{ii}}{M-2LN}\right]^{1/2} . \tag{3-58}$$

Also the simultaneous confidence interval is contained in the region

$$\frac{(\underset{\sim}{\rho}-\underset{\sim}{\rho})'(X'X)^{-1}(\underset{\sim}{\rho}-\underset{\sim}{\rho})}{(2LN)\sigma^2} < F\alpha, 2LN, (M-2LN) \tag{3-59}$$

where $\qquad \sigma^2 = SS_E/(M-2LN)$.

To test the hypothesis $H_0 : H \underset{\sim}{\rho} = 0$, where H is a q x 2LN matrix of rank q, with q < 2LN, the testing statistics is:

$$W = \frac{\underset{\sim}{\rho}\ H'[H(X'X)^{-1}H']\ H\underset{\sim}{\rho}}{q\sigma^2} \tag{3-60}$$

H_0 is rejected if $W > F\alpha, q, [(LN)(n_1+n_2-2)]$. For example, if we would like to test the hypothesis that the shift is insignificant in all regions, we test,

$$H_0 : \underset{\sim}{\delta} = \underset{\sim}{0}.$$

with H matrix of the form $[0 \,\vdots\, I_{LN}]$.

3.5.3 L.S. Estimators and Hypothesis Tests for $\underset{\sim}{\phi}$, $\underset{\sim}{\Theta}$ Unknown

The problem described previously of estimating $\underset{\sim}{\mu}$, $\underset{\sim}{\delta}$ when $\underset{\sim}{\phi}$, $\underset{\sim}{\Theta}$ are known was a linear estimation problem. When $\underset{\sim}{\phi}$, $\underset{\sim}{\Theta}$ are also unknown and we would like to estimate $(\underset{\sim}{\phi},\underset{\sim}{\Theta},\underset{\sim}{\mu},\underset{\sim}{\delta})$ simultaneously, the problem becomes one of nonlinear estimation. Since there are $(m_1+\lambda_1+2LN + 2)$

paramters seaching over the sum of squares function is tedious. For a more efficient solution the model form is linearized and an interative linear least squares appraoch is used. To linearize the transformation form $\underset{\sim}{Y} = x\,\underset{\sim}{\rho} + \varepsilon$ an initial guess value is selected $({}^{o}\underset{\sim}{\phi}, {}^{o}\underset{\sim}{\Theta}, {}^{o}\underset{\sim}{\mu}, {}^{o}\underset{\sim}{\delta})$. In the following we will restrict ourself to the $p\leq 1$, $q\leq 1$ models for simplicity. The discussion can be naturally extended to other models ($p\geq 2$ or $q\geq 2$). Recall that elements of X-matrix are functions of $\underset{\sim}{\Theta}$ and $\underset{\sim}{\phi}$ only, so to get the linearized approximation at $\underset{\sim}{\beta} = {}^{o}\underset{\sim}{\beta}$, $\underset{\sim}{\rho} = {}^{o}\underset{\sim}{\rho}$, we will have the linearized models of the following form:

$$\underset{\sim}{Y} = \sum_{\ell=0}^{\lambda_1} \left(\frac{\partial X}{\partial \phi_{1\ell}}\bigg|_{\underset{\sim}{\beta} = {}^{o}\underset{\sim}{\beta}}\right)(\phi_{1\ell} - {}^{o}\phi_{1\ell})\,\underset{\sim}{\rho} \qquad (3\text{-}61)$$

$$+ \sum_{\ell=0}^{m_1} \left(\frac{\partial X}{\partial \Theta_{1\ell}}\bigg|_{\underset{\sim}{\beta} = {}^{o}\underset{\sim}{\beta}}\right)(\Theta_{1\ell} - {}^{o}\Theta_{1\ell})\,{}^{o}\underset{\sim}{\rho} + X_0\,\underset{\sim}{\rho} + \underset{\sim}{\varepsilon}.$$

where X_0 is the X-matrix evaluated at $({}^{o}\underset{\sim}{\phi}, {}^{o}\underset{\sim}{\Theta}, {}^{o}\underset{\sim}{\mu}, {}^{o}\underset{\sim}{\delta})$, ${}^{o}\beta^{t} = ({}^{o}\underset{\sim}{\phi}^{t}, {}^{o}\underset{\sim}{\Theta}^{t})$ and ${}^{o}\underset{\sim}{\rho} = ({}^{o}\underset{\sim}{\mu}, {}^{o}\underset{\sim}{\delta})$, which is the M.L. estimates of $(\underset{\sim}{\mu}, \underset{\sim}{\delta})$ given that $(\underset{\sim}{\phi}^{t}, \underset{\sim}{\Theta}^{t}) = ({}^{o}\underset{\sim}{\phi}^{t}, {}^{o}\underset{\sim}{\Theta}^{t})$.

Let $${}^{o}X^{\Delta} = \left(\frac{\partial X}{\partial \phi_{10}}\bigg|_{\underset{\sim}{\beta}={}^{o}\underset{\sim}{\beta}}\right){}^{o}\underset{\sim}{\rho},\ \left(\frac{\partial X}{\partial \phi_{11}}\bigg|_{\underset{\sim}{\beta}={}^{o}\underset{\sim}{\beta}}\right){}^{o}\underset{\sim}{\rho},$$

$$\ldots,\ \left(\frac{\partial X}{\partial \phi_{1\lambda_1}}\bigg|_{\underset{\sim}{\beta}={}^{o}\underset{\sim}{\beta}}\right){}^{o}\underset{\sim}{\rho},\ \left(\frac{\partial X}{\partial \Theta_{10}}\bigg|_{\underset{\sim}{\beta}={}^{o}\underset{\sim}{\beta}}\right){}^{o}\underset{\sim}{\rho},\ \ldots,\ \left(\frac{\partial X}{\partial \Theta_{1m_1}}\bigg|_{\underset{\sim}{\beta}={}^{o}\underset{\sim}{\beta}}\right){}^{o}\underset{\sim}{\rho})$$

$${}^{o}\underset{\sim}{\beta}^{\Delta} = \underset{\sim}{\beta} - {}^{o}\underset{\sim}{\beta} = \begin{matrix} \underset{\sim}{\phi} - {}^{o}\underset{\sim}{\phi} \\ \underset{\sim}{\Theta} - {}^{o}\underset{\sim}{\Theta} \end{matrix}$$

where ${}^{o}X^{\Delta}$ matrix is an $[LN\cdot(n_1+n_2)]$ by $(\lambda_1 + m_1 + 2)$ matrix, $\underset{\sim}{\beta}^{\Delta}$ is an $(\lambda_1 + m_1 + 2)$ vector, and we have the linearized model at $(\underset{\sim}{\beta}, \underset{\sim}{\rho}) = ({}^{o}\underset{\sim}{\beta}, {}^{o}\underset{\sim}{\rho})$ as

$$\underset{\sim}{Y} = {}^{o}X^{\Delta}\,\underset{\sim}{\beta}^{\Delta} + X_0\,\underset{\sim}{\rho} + \underset{\sim}{\varepsilon} \qquad (3\text{-}62)$$

Since the linearized approximate models are linear in $(\underset{\sim}{\phi}, \underset{\sim}{\Theta}, \underset{\sim}{\mu}, \underset{\sim}{\delta})$, so we could apply the results of lienar model theory to get the L.S. estimates of ${}^{o}\underset{\sim}{\beta}^{\Delta}$ and ${}^{\Theta}\underset{\sim}{\rho}$, i.e., ${}^{o}\underset{\sim}{\hat\beta}^{\Delta}$ and ${}^{1}\underset{\sim}{\rho}$. Once the L.S. estimates ${}^{o}\underset{\sim}{\beta}^{\Delta}$, ${}^{1}\underset{\sim}{\rho}$ are obtained, the linearized model is then established at $(\underset{\sim}{\beta}, \underset{\sim}{\rho}) = ({}^{1}\underset{\sim}{\beta}, {}^{1}\underset{\sim}{\rho})$, where $({}^{1}\underset{\sim}{B}, {}^{1}\underset{\sim}{\rho}) = ({}^{o}\underset{\sim}{\beta} + {}^{o}\underset{\sim}{\beta}^{\Delta}, {}^{1}\underset{\sim}{\rho})$, and the linear model theory is then applied again to get the L.S. estimates ${}^{1}\underset{\sim}{\beta}^{\Delta}$ and ${}^{\Theta}\underset{\sim}{\rho}$. This procedure is repeated until both of the following two stopping rules are satisfied.

1. ${}^{i}\underset{\sim}{\beta}^{\Delta} < \underset{\sim}{\varepsilon}_{\beta}$, where $\underset{\sim}{\varepsilon}_{\beta}$ is a vector of arbitrarily small positive number, and ${}^{i}\underset{\sim}{\beta}^{\Delta}$ is the L.S. estimate of ${}^{i}\underset{\sim}{\beta}^{\Delta}$ at i-th iteration.
2. $SS_i - SS_{i-1} < \varepsilon$, where ε is an arbitrarily small positive number, and SS_i, SS_{i-1} are the sum of squares at

i-th iteration and $(i-1)^{th}$ iteration

respectively.

Once the iterative procedure stops, the L.S. estimates of $\underset{\sim}{\beta}$ and $\underset{\sim}{\rho}$ are obtained. To construct the confidence intervals and to perform the hypothesis tests, the linearized model is constructed at $(\underset{\sim}{\beta},\underset{\sim}{\rho}) = (\underset{\sim}{\beta},\underset{\sim}{\rho})$ and then the results of the linear model, which are discussed in 3.5.2, are applied directly to obtain the confidence intervals as well as the test statistics.

To illustrate the linearization in more detail, let us linearize the ξ_t pulse function situation for STARMA$(1_{\lambda 1},0,1_{m1})I_m$ intervention model. The results are given below of the linear expression form in equation 3-62,

$$\underset{\sim}{Y} = {}^{o}X^{\Delta}\ {}^{o}\underset{\sim}{\beta}^{\Delta} + X_0\underset{\sim}{\rho} + \underset{\sim}{\varepsilon}$$

where

$${}^{o}\underset{\sim}{\beta}^{\Delta} = |(\underset{\sim}{\phi}-{}^{o}\phi)^t,\ (\underset{\sim}{\Theta}-{}^{o}\underset{\sim}{\Theta})^t|^t, \text{ and}$$

$${}^{o}X^{\Delta}_{i,\ell}(t) = -\sum_{j=1}^{LN}\sum_{K=0}^{t-2} D^{(K)}_{ij\ell}\ {}^{o}\mu_j - I_m \sum_{j=1}^{LN}\sum_{K=t-n_1-T_p-1}^{t-n_1-2} D^{(K)}_{ij\ell}\ {}^{o}\delta_j$$

$$1 < \ell < \lambda_1 + 1, \quad i = 1,2,\ldots,LN$$

$${}^{o}X^{\Delta}_{i,\ell+\lambda_1}(t) = -\sum_{j=1}^{LN}\sum_{K=0}^{t-2} E^{(K)}_{ij\ell}\ {}^{o}\mu_j - I_m \sum_{j=1}^{LN}\sum_{K=t-n_1-T_p-1}^{t-n_1-2} E^{(K)}_{ij\ell}\ {}^{o}\delta_j$$

CONTINUED

2 OF 8

with $D_{ij\ell}^{(K)}$ and $E_{ij\ell}^{(K)}$ defined recursively as follows.

$$D_{ij\ell}^{(0)} = W_{ij}^{(\ell)}$$

$$D_{ij\ell}^{(K)} = \sum_{r=1}^{m_1} \sum_{g=1}^{LN} {}^{o}\Theta_{1\gamma}\, W_{ig}^{(r)}\, D_{gj\ell}^{(K-1)}$$

$$E_{ij\ell}^{(0)} = -W_{ij}^{(\ell)}$$

$$E_{ij\ell}^{(K)} = \sum_{g=1}^{LN} W_{ig}^{(\ell)}\, U_{gj}^{(K-1)} + \sum_{\gamma=0}^{m_1} \sum_{g=1}^{LN} {}^{o}\Theta_{1\gamma}\, E_{gj\ell}^{(K-1)}$$

Note that all the $D_{ij\ell}$'s and $E_{ij\ell}$'s are functions of $\underset{\sim}{\Theta}$, not $\underset{\sim}{\phi}$, and are evaluated at $\underset{\sim}{\Theta} = {}^{o}\underset{\sim}{\Theta}$. This is intuitively true because the moving average parameters are nonlinear while the autoregressive parameters are linear in their own nature.

3.5.4 L.S. Estimators for the Multi-Consequence Space-Time Model

When the intervention is believed to change the process covariance structure, we perform conditional L.S. estimation, i.e. estimate $(\underset{\sim}{\psi},\underset{\sim}{\gamma},\underset{\sim}{\delta}|\underset{\sim}{\phi},\underset{\sim}{\Theta},\underset{\sim}{\mu})$, where $\underset{\sim}{\phi},\underset{\sim}{\Theta},\underset{\sim}{\mu}$ are the L.S. estimates of pre-intervention model parameters and the pre-intervention process mean respectively. To get the L.S. estimates of $(\underset{\sim}{\psi},\underset{\sim}{\gamma},\underset{\sim}{\delta}|\underset{\sim}{\phi},\underset{\sim}{\Theta},\underset{\sim}{\mu})$, we need to search through the $(\underset{\sim}{\psi},\underset{\sim}{\gamma})$ space. The transformation needed to express this model in lienar model form can be obtained by modifying equation

(3-42), (3-46). To get the $\underset{\sim}{Y}(t)$ vector ahd $X_2(t)$ matrix,

$$\underset{\sim}{Y}(t) = \sum_{k=t-n_1}^{q} W_{\Theta K}\, \underset{\sim}{Y}(t-K) + \underset{\sim}{Z}_t - \underset{\sim}{\mu} - \sum_{K=t-n_1}^{P} W_{\phi K}\, \underset{\sim}{Z}_{t-K}$$

$$+ \sum_{K=1}^{\min\{q,t-n_1-1\}} W_{\gamma K}\, \underset{\sim}{Y}(t-K) + \sum_{K=1}^{\min\{P,t-n_1-1\}} W_{\psi K}\, \underset{\sim}{Z}_{t-K}$$

$$X_2(t) = \sum_{K=t-n_1}^{q} W_{\Theta K}\, X_2(t-K) + P_t^{n_1}(T_p) - (1-I_m) \sum_{K=t-n_1}^{P} W_{\phi K} P_t^{n_1}(T_p)$$

$$- I_m \sum_{K=t-n_1+1}^{q} W_{\Theta K}\, P_{t-K}^{n_1}(T_p)t \sum_{K=1}^{\min\{q,t-n_1-1\}} W_{\gamma K}\, X(t-K)$$

$$-(1-I_m) \sum_{K=1}^{\min\{P,t-n_1-1\}} W_{\psi K}\, P_t^{n_1}(T_p) - I_m \sum_{K=1}^{\min\{q,t-n_1-1\}} W_{\gamma K}\, P_{t-K}^{n_1}(T_p),$$

which results in the liear model form,

$$\underset{\sim}{Y}(t) = X_2(t)\, \underset{\sim}{\delta} + \underset{\sim}{\varepsilon}_t$$

Note that $\underset{\sim}{Y}(t)$, $X_2(t)$, $t<n_1$ are known when the pre-intervention model has been built.

Once the L.S. estimates $(\underset{\sim}{\psi},\underset{\sim}{\gamma},\underset{\sim}{\delta}|\underset{\sim}{\phi},\underset{\sim}{\Theta},\underset{\sim}{\mu})$ are obtained, the lienarized model, which has been discussed in the previous section, is constructed and the results of the linear model are applied to construct the confidence interval as well as the hypothesis testing stat.

3.6 An Example: Los Angeles Carbon Monoxide (CO) Data

In this section the Los Angeles CO Data is reanalyzed to illustrate the space-time intervention modeling procedures. This data was previously analyzed by Box, Tiao and Hamming [1975] using single site intervention models. The data are the monthly averages of hourly measurements of carbon monoxide, that ~~was~~ were recorded from March 65 to December 71 at six geographically distributed locations in the Los Angeles Basin, i.e., Azusa, Burbank, Lennox, Long Beach, Downtown LA and LA County. Two events (or interventions) occurred, which were expected to reduce the measurement level of carbon monoxide at these locations. The first intervention was the air quality legislation that required an engine design change. This law, enacted in January 1966, required the engine to be designed more efficiently so as to produce less air pollutants. The second intervention introduced in April 1968 was the change in the method of calibration of the measuring instruments. We will denote I_1 as the first intervention, i.e. the engine design change, and I_2 as the second intervention, i.e. the change of calibration method.

In the following sections, the geological environment of these six locations in the Los Angeles Basin will be described and followed by the construction of weight matrix. Then these two interventions are analyzed to determine the appropriate dynamic component model. In the next section, the space-time intervention modeling procedures are followed step by step to build the pre-I_1 space-time noise model, to identify the model form of the dynamic components, to build the post-I_1, pre-I_2 space-time intervention model and then the whole space-time multiple interventions model for the whole process. The implications revealed in the built model are discussed to draw the conclusions of the physical interpretation. Also comparisons are made between the space-time multiple intervention model and those univariate time series intervention models in Box, Tiao and Hamming [1975].

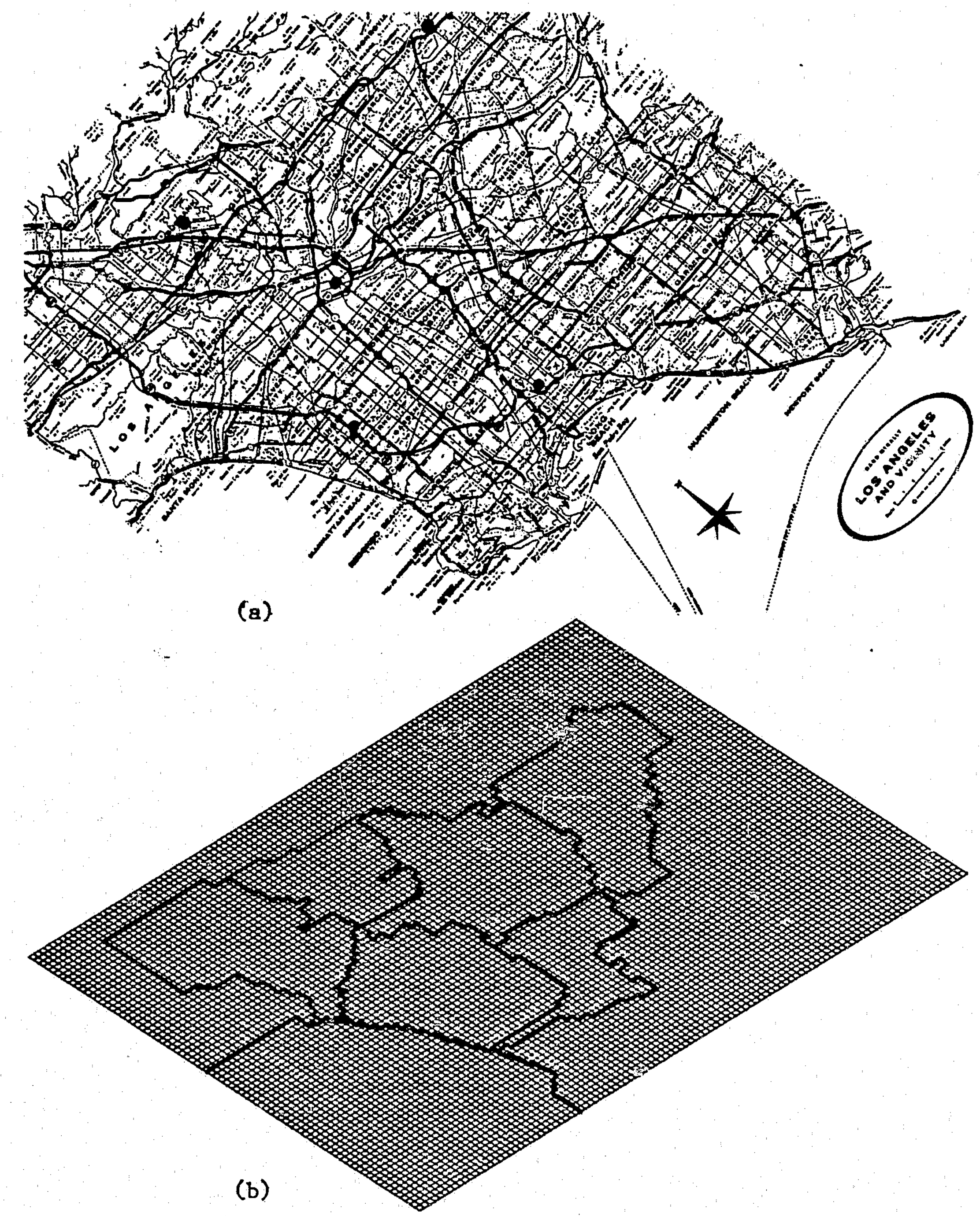
3.6.1 The System, Structure and Data

The relative position of these six locations are illustrated in the actual map and computer drawn facsimile in figures 3-10(a) and (b) respectively. The distances in miles as measured from this map are given in table 3.1. In each case these measurements represent not centroidal distances but distances between measurement sites.

Table 3.1 Distances between Gauge Sites

	Azusa	Burbank	Lennox	Long Beach	Downtown LA	LA County
Azusa		22.1	29.5	26.8	19.5	35.4
Burbank	22.1		19.0	26.8	11.3	13.3
Lennox	29.5	19.0		12.8	10.7	21.2
Long Beach	26.8	26.8	12.8		15.6	33.3
Downtown LA	19.5	11.3	10.7	15.6		20.2
LA County	35.4	13.3	21.2	33.3	20.2	

Any two locations are assigned the same order neighbor to each other according to the distances listed below:



(a)

(b)

Figure 3-10. The Map and the Relative Positions of Lennox, Long Beach, LA County, Burbank, Downtown LA and Azusa.

Distance Range	Order Assigned
0 - 16 miles	1
17 - 25 miles	2
26 - 34 miles	3

The resulting neighbor structure for this assignment is:

Location	Index	1st Order Neighbor	2nd Order Neighbor	3rd Order Neighbor
Azusa	1	-	2,5	3,4,6
Burbank	2	5,6	1,3	4
Lennox	3	4,5	2,6	1
Long Beach	4	3,5	-	1,2,6
Downtown LA	5	2,3,4	1,6	-
LA County	6	2	3,5	1,4

The scaled weight matrix is constructed according to the inverse distance and is listed below.

$$W^{(1)} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.44 & 0.56 \\ 0 & 0 & 0 & 0.54 & 0.46 & 0 \\ 0 & 0 & 0.45 & 0 & 0.55 & 0 \\ 0 & 0.29 & 0.29 & 0.42 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

$$W^{(2)} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0.53 & 0 & 0 & 0.47 & 0 \\ 0.54 & 0 & 0.46 & 0 & 0 & 0 \\ 0 & 0.48 & 0 & 0 & 0 & 0.52 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.51 & 0 & 0.49 & 0 \end{array}\right] \end{array}$$

$$W^{(3)} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & 0.32 & 0.29 & 0 & 0.39 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0.32 & 0.30 & 0 & 0 & 0 & 0.38 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.52 & 0 & 0 & 0.48 & 0 & 0 \end{array}\right] \end{array}$$

The data of each location are plotted in the figures 3-11(a) to 3-11(f). Table 3-2 contains the sample space-time autocorrelation function and the standardized sample space-time autocorrelation functions for the 82 time points and 6 locations. In the following sections the modeling of this substantial, statistically significant spatially and temporally correlated information will be conducted.

3.6.2 Initial Considerations of the Forms of the Interventions

The first intervention, enacted in January 1966, required an improvement in engine design to reduce the air pollutants contained in

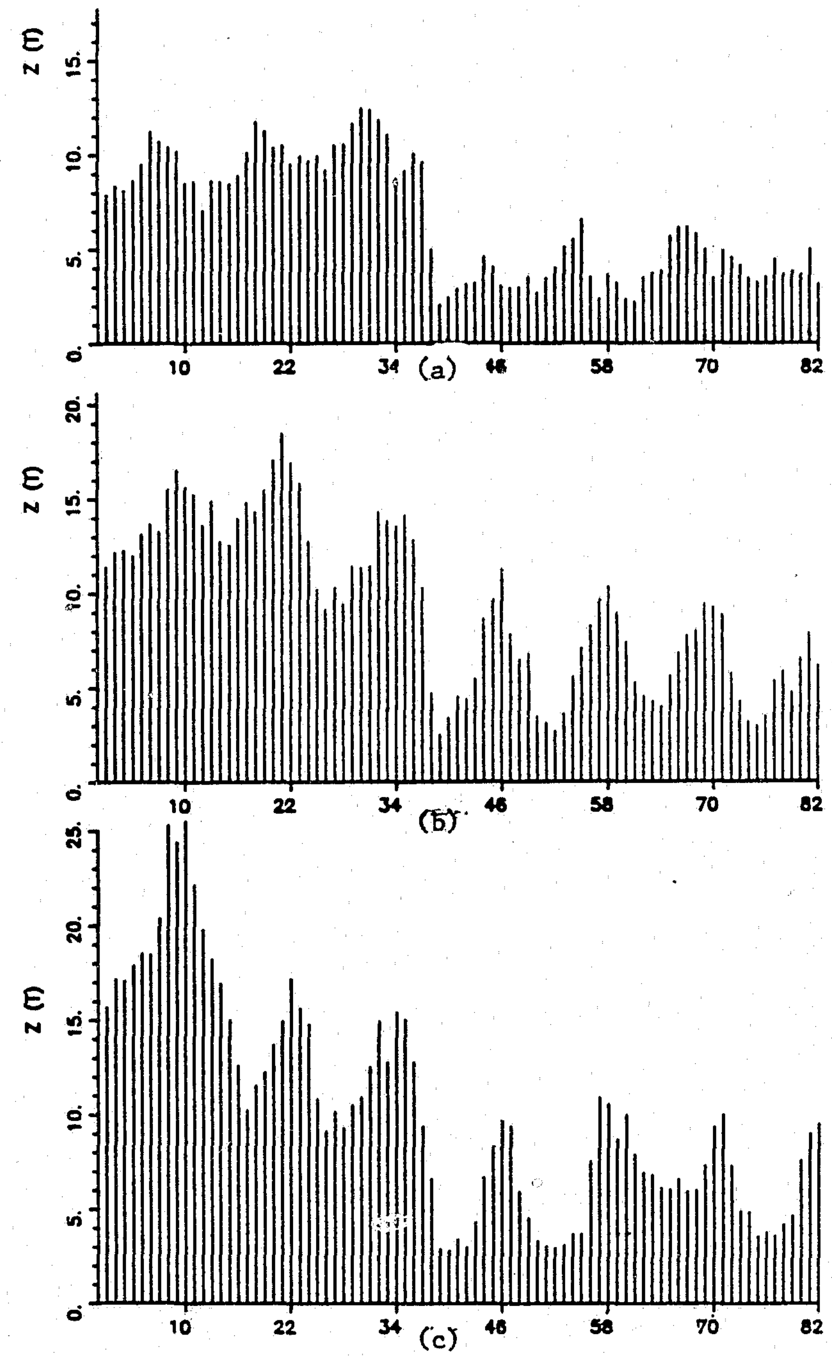


Figure 3-11. CO Measurements
(a) at Azusa
(b) at Burbank
(c) at Lennox

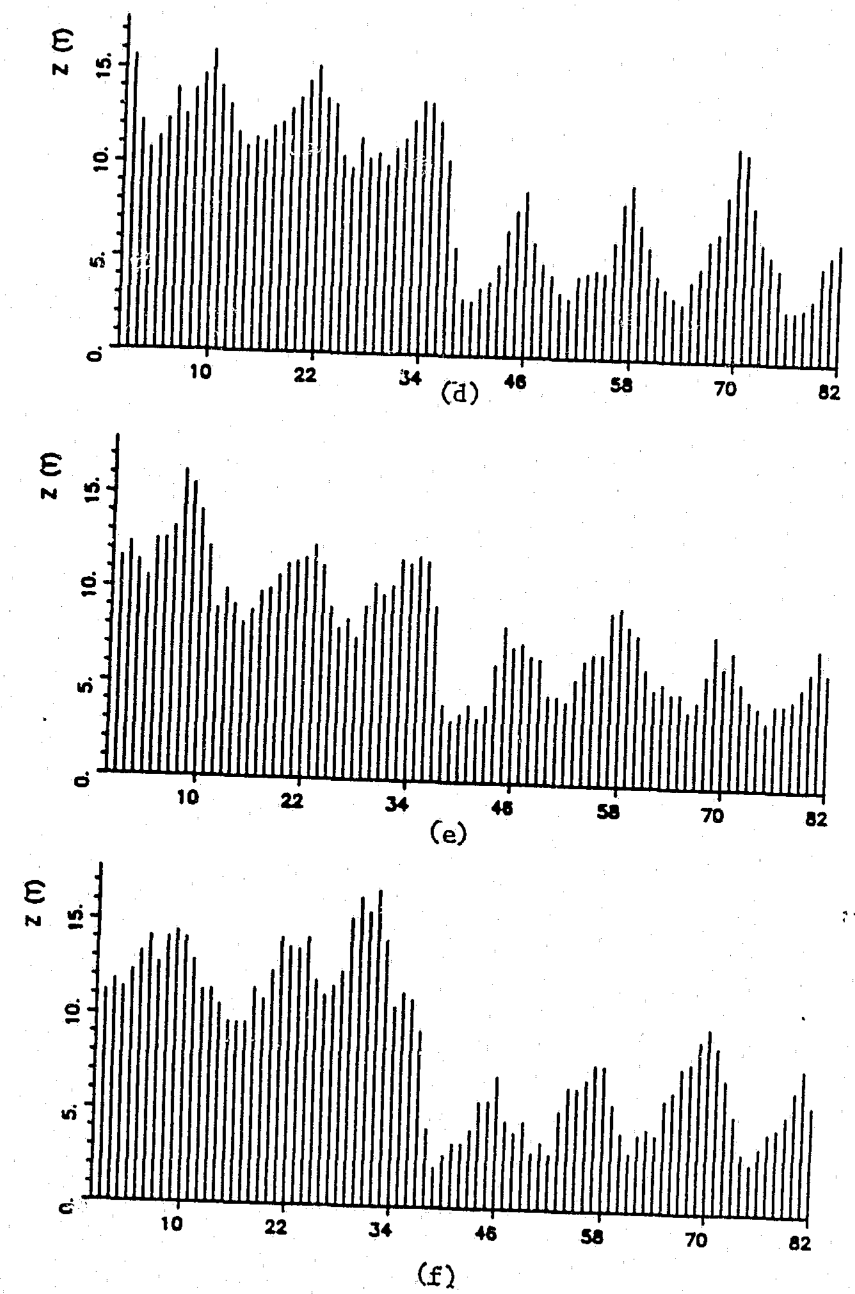


Figure 3-11. (Cont'd)
(d) at Long Beach
(e) at Downtown LA
(f) at LA County

Table 3-2 The Sample Space-Time Autocorrelation Functions and the Standardized Smaple S-T Autocorrelation Functions for the Observations from March 1965 to December 1971.

	Space-Time Autocorrelations			Standardized S-T Autocorrelations		
Space Lag / Time Lag	0	1	2	0	1	2
1	0.94	0.79	0.75	20.52	17.25	16.39
2	0.89	0.71	0.70	18.36	15.47	15.14
3	0.74	0.62	0.63	15.94	13.35	13.60
4	0.63	0.52	0.56	13.52	11.20	11.98
5	0.53	0.44	0.49	11.43	9.49	10.59
6	0.47	0.39	0.45	9.92	8.35	9.49
7	0.44	0.39	0.43	9.29	8.24	9.02
8	0.45	0.42	0.43	9.40	8.76	9.08
9	0.48	0.46	0.45	9.97	9.69	9.38
10	0.52	0.51	0.47	10.68	10.62	9.71
11	0.55	0.56	0.49	11.29	11.43	10.00
12	0.56	0.57	0.49	11.47	11.71	10.06
13	0.53	0.54	0.47	10.69	11.02	9.54
14	0.46	0.47	0.42	9.17	9.41	8.52
15	0.35	0.36	0.35	6.96	7.11	7.00
16	0.25	0.25	0.28	5.02	4.99	5.58
17	0.17	0.16	0.22	3.38	3.22	4.35
18	0.12	0.10	0.18	2.35	2.07	3.52
19	0.10	0.09	0.16	2.03	1.80	3.23
20	0.12	0.11	0.17	2.32	2.21	3.37
21	0.16	0.16	0.20	3.19	3.16	3.88
22	0.22	0.22	0.23	4.18	4.14	4.40
23	0.25	0.25	0.24	4.74	4.76	4.54
24	0.25	0.25	0.23	4.73	4.76	4.28
25	0.22	0.22	0.20	4.09	4.15	3.68
26	0.15	0.16	0.15	2.78	2.93	2.70
27	0.04	0.05	0.07	0.80	1.03	1.32
28	-0.07	-0.06	-0.01	-1.26	-1.10	-0.31
29	-0.18	-0.17	-0.11	-3.25	-3.08	-2.03
30	-0.26	-0.25	-0.19	-4.64	-4.46	-3.31

the exhausted gas. Since the old cars will be replaced by the newly designed car with improved engine design gradually, the effect of this impact would be expected to increase perhaps linearly until the steady state is reached, i.e. all the cars on the roads are designed under the impact of this air quality legislation. Thus from a modeling standpoint, the indicator variable $\xi_t^{(1)}$ is not 0,1 since there is not instantaneous total resource implementation. Rather $\xi_t^{(1)}$, the indicator variable for the first intervention should be of the form $\gamma_t^{(1)}\xi_t$, where $\gamma_t^{(1)}$ reflects the resource implementation. Also this legislative intervention didn't involve any action that will reduce the quantity of air pollutants directly. Instead it put constraints on the air pollutant generators (the engine) that produce the source input, i.e. the noise input of the STARMA process, of the whole system. Recall, that this noise input is the only source of input of the STARMA process, so the effect of this engine design change legislation will enter the environment process that the noise follows. That is, this intervention takes the form of the environmental influence situation, i.e. $I_{m,1} = 1$.

Assuming a constant change over rate and an eight year useful life of an automobile, we have the intervention STARMA $(\underset{\sim}{P_\lambda},0,\underset{\sim}{q_m})\ I_m$ process takes the following general form:

$$\Phi_2(B)(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \Theta_2(B)(\underset{\sim}{\varepsilon}+\underset{\sim}{\xi}^{(1)}(t)) \tag{3-63}$$

where $\Phi_2(B)$, $\Theta_2(B)$ are process parameters and

$\utilde{\xi}^{(1)}(t) = \xi_t^{(1)}\utilde{\delta}^{(1)}$

$\utilde{\delta}^{(1)} = \lfloor \delta_1^{(1)}, \delta_2^{(1)}, \ldots, \delta_6^{(1)} \rfloor^t$ with

$\delta_i^{(1)}$ as the intrinsic program utility at the location i, and

$$\xi_t^{(1)} = \begin{cases} 0 & t<10 \\ 0.065 + 0.125 \left\lfloor \frac{t-10}{12} \right\rfloor & t>11 \end{cases}$$

$\lfloor a \rfloor$ ≡ the largest integer contained in $(-\infty, a)$, ~~a is real number.~~

The second intervention, the change of instrument calibration method, will not have any impact on the existent air pollutants, but only potentially changes in the measurement readings. Suppose that for two instruments, one is calibrated by the old method and the other by the new method, were available at the same time and they were used to measure the pollution level at the same location simultaneously. Assume that $\utilde{Z}_t^0$ and $\utilde{Z}_t^N$ will have a one-to-one corresponding relationship of the form,

$$\utilde{Z}_t^0 = H(\utilde{Z}_t^N) \quad \text{or} \quad \utilde{Z}_t^N = H^{-1}(\utilde{Z}_t^0) \qquad (3\text{-}64)$$

where H is an arbitrary function with H^{-1} exists as its inverse function. H function may be linear may be nonlinear. Here we will assume in general that H function is nonlinear and well behaved, i.e. $H(\utilde{\mu}+\utilde{\varepsilon})$ can be approximated by $H(\utilde{\mu}) + H'(\utilde{\mu})\utilde{\varepsilon}$ when $\utilde{\varepsilon}$ is a

small quantity.

Since $\utilde{Z}^0 = \utilde{\mu}^0 + \utilde{A}_t{}^0$ and $\utilde{Z}_t^N = \utilde{\mu}^N + \utilde{A}_t^N$, where $\utilde{\mu}^0$, $\utilde{\mu}^N$ are mean values of $\utilde{Z}_t^0$ and $\utilde{Z}_t^N$ measurements respectively, $\utilde{A}_t^0$ and $\utilde{A}_t^N$ are the random portions.

$$\utilde{Z}_t^0 = H(\utilde{Z}_t^N) = H(\utilde{\mu}^N) + H'(\utilde{\mu}^N)\,\utilde{A}_t^N, \text{ and}$$

$$\utilde{Z}_t^0 = \lfloor H(\utilde{\mu}^N) - H'(\utilde{\mu}^N)\,\utilde{\mu}^N \rfloor + H'(\utilde{\mu}^N)\,\utilde{Z}_t^N, \qquad (3\text{-}65)$$

or

$$\utilde{Z}_t^0 - \utilde{\mu}^0 = \lfloor H(\utilde{\mu}^N) - H'(\utilde{\mu}^N)(\utilde{\mu}^N - \utilde{\mu}^0) \rfloor + H'(\utilde{\mu}^N)(\utilde{Z}_t^N - \utilde{\mu}^0) \qquad (3\text{-}66)$$

Since the I_2 intervention doesn't have any impact on the existent air pollution level, in that it only changes the level readings, the post I_1 process will not be changed by I_2. What will change is the units that was used to describe this process. Thus, I_2 is an non-environment involved intervention with $I_{m,2} = 0$.

The $\utilde{Z}_t^0$ in the equation 3-66, which is of the same descriptive unit as that of the $\utilde{Z}_t$ in the equation 3-63, is not available after I_2 but can be obtained through the transformation equation 3-66. The process that is described by equation 3-63 in terms of $\utilde{Z}_t^0$ for the post I_1 pre-I_2 periods then can be expressed in terms of $\utilde{Z}_t^N$, that is available for the post-I_2 periods, by the following equation.

$$\Phi_2(B)\left(\underset{\sim}{\delta_3}+H'(\underset{\sim}{\mu}^N)(\underset{\sim}{Z}_t^N-\underset{\sim}{\mu}^0)\right) = \Theta_2(B)\left(\underset{\sim}{\varepsilon}_t+\underset{\sim}{\xi}^{(1)}(t)\right) \qquad (3\text{-}67)$$

where

$\underset{\sim}{\delta}_3 = H(\underset{\sim}{\mu}^N) - H'(\underset{\sim}{\mu}^N)(\underset{\sim}{\mu}^N-\underset{\sim}{\mu}^0)$ is function of $\underset{\sim}{\mu}^0$,

because $\underset{\sim}{\mu}^N$ is function of $\underset{\sim}{\mu}^0$ too. Since this unit transformation will be independent among locations, $H'(\underset{\sim}{\mu}^N)$ should be a diagonal matrix, i.e. $H'(\underset{\sim}{\mu}^N) = D$, of which the diagonal elements d_{ii} are functions of μ_i, the mean level of location i. To simplify the notation we will denote $\underset{\sim}{Z}_t^N$ as $\underset{\sim}{Z}_t$ for the post-I_2 periods because it is available for analysis in those post I_2 periods. The model, equation 3-67, is rewritten as

$$\Phi_2(B)\left(\underset{\sim}{\delta}_3+D(\underset{\sim}{Z}_t-\underset{\sim}{\mu})\right) = \Theta_2(B)\left(\underset{\sim}{\varepsilon}_t+\underset{\sim}{\xi}^{(1)}(t)\right) \qquad (3\text{-}68)$$

To summarize the previous discussion, we have the following general multiple intervention STARMA model form;

$$\begin{cases} \text{Pre-}I_1: & \Phi_1(B)(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \Theta_1(B)\underset{\sim}{\varepsilon}_t \\[2ex] \text{Post-}I_1 \\ \text{Pre-}I_2: & \Phi_2(B)(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \Theta_2(B)\left(\underset{\sim}{\varepsilon}_t+\underset{\sim}{\xi}^{(1)}(t)\right) \\[2ex] \text{Post-}I_2: & \Phi_2(B)\left(\underset{\sim}{\delta}_3+D(\underset{\sim}{Z}_t-\underset{\sim}{\mu})\right) = \Theta_2(B)\left(\underset{\sim}{\varepsilon}_t+\underset{\sim}{\xi}^{(1)}(t)\right) \end{cases} \qquad (3\text{-}69)$$

where $\underset{\sim}{\xi}^{(1)}(t)$ is described in equation (3-63), and

$\underset{\sim}{\delta}_3$ is the measure level shift,

$\Phi_1(b)$, $\Theta_1(b)$ are the pre-I_1 model parameters, and

$\Phi_2(b)$, $\Theta_2(b)$ are the post-I_1 model parameters.

From equation (3-69), we have

$$\mathrm{Var}(\underset{\sim}{Z}_t^c) = \mathrm{Var}(\Phi_2(B)^{-1}\,\Theta_2(b)\underset{\sim}{\varepsilon}_t),\ t\ \varepsilon\ \text{pre-}I_2,\ \text{post-}I_1 \text{ periods},$$

and

$$\mathrm{Var}(D\underset{\sim}{Z}_t^c) = \mathrm{Var}\left(\Phi_2(B^{-1})\ \Theta_2(b)\ \underset{\sim}{\varepsilon}_t\right),\ t\ \varepsilon\ \text{post-}I_2 \text{ periods}.$$

So

$$D^2 = \mathrm{Var}(\underset{\sim}{Z}_t^c)\ t\ \varepsilon\ \text{post-}I_2 \ /\ \mathrm{Var}(\underset{\sim}{Z}_t^c)\ t\ \varepsilon\ \text{pre-}I_2,\ \text{post-}I_1, \qquad (3\text{-}70)$$

where $\underset{\sim}{Z}_t^c$ is the mean corrected observations in the dynamic components identification procedure. Since the D-matrix is a diagonal matrix with its ith diagonal element d_{ii} interpreted as the scaled factor between readings from the old instrument and the new instrument measured at location i, the element d_{ii} can be estimated by

$$d_{ii} = \left|\mathrm{Var}(Z_{i,t})_{t\ \varepsilon\ \text{post-}I_2} \ /\ \mathrm{Var}(\underset{\sim}{Z}_{i,t})_{t\ \varepsilon\ \text{pre-}I_2,\ \text{post-}I_1}\right|^{1/2}. \qquad (3\text{-}71)$$

3.6.3 Modeling the LA CO Data

In the following the results of each step of the model building procedures are presented. This includes;

(a) Modeling the pre-I_1 space-time process.

(b) Modeling the post-I_1 dynamic components.

(c) Modeling the post-I_1 space-time intervention process.

(d) Overall diagnostic checking and model update.

4.6.3.1 Modeling the pre-I_1 space-time process

The observations from March, 1965 to December, 1965 comprise the pre-I_1 periods. The sample space-time autocorrelation functions in the table 3-3(a) and the sample space-time partial autocorrelation functions in the table 3-3(b) suggest the candidate model STAR($2\underset{\sim}{\lambda}$) model with $\underset{\sim}{\lambda} = (1,1)$, i.e.

$$\underset{\sim}{Z}_t - \underset{\sim}{\mu} = \sum_{k=1}^{2} \sum_{\ell=0}^{1} \Phi_{K\ell} W^{(\ell)} (\underset{\sim}{Z}_{t-K} - \underset{\sim}{\mu}) + \underset{\sim}{\varepsilon}_t \qquad 1 < t < 0 \qquad (3\text{-}72)$$

The M.L. point estimaters and their associated 95% confidence intervals are;

		95% CI
Φ_{10}	= 0.668	(0.359, 0.976)
Φ_{11}	= 0.336	(-0.044, 0.715)
Φ_{20}	= -0.130	(-0.459, 0.201)
Φ_{21}	= -0.056	(-0.482, 0.370)

Table 3-3(a).
The Space-Time Autocorrelation Functions and The Standardized S-T Autocorrelation Functions of the Pre-I_1 Observations.

	Space-Time Autocorrelations/Standardized S-T Autocorrelations					
Space Lag	0	1	2	0	1	2
Time Lag						
1	0.68	0.57	0.50	4.438	3.716	3.298
2	0.38	0.38	0.35	2.321	2.329	2.148
3	0.01	0.06	0.12	0.106	0.329	0.704
4	-0.26	-0.24	-0.11	-1.274	-1.196	-0.578
5	-0.52	-0.59	-0.43	-2.222	-2.541	-1.834

Table 3-3(b).
The Space-Time Partial Autocorrelation Functions and The Standardized S-T Partials of The Pre-I_1 Observations.

	Space-Time Autocorrelations/Standardized S-T Autocorrelations					
Space Lag	0	1	2	0	1	2
Time Lag						
1	0.686	0.252	0.089	5.033	1.849	0.655
2	-0.214	-0.144	-0.025	-1.481	-0.999	-0.175
3	-0.381	-0.319	-0.001	-2.471	-2.065	-0.008
4	-0.149	-0.126	-0.100	-0.896	-0.757	-0.603
5	-0.263	-0.570	-0.370	-1.449	-3.124	-2.026

$$\sigma^2 = 1.4377$$

The extra sum of squares associated with Φ_{21} was insignificant. Since Φ_{21} is insignificant, it was deleted which results in the STAR($2_{1,0}$) model. The M.L. point estimates and their associated 95% confidence intervals are:

		95% CI
Φ_{10}	= 0.671	(0.366, 0.976)
Φ_{11}	= 0.319	(-0.007, 0.624)
Φ_{20}	= -0.146	(-0.449, 0.158)
σ^2	= 1.438	

The sample space-time autocorrelation functions and the sample partial autocorrelation functions of the residuals of this model are listed in tables 3-4(a) and 3-4(b) respectively. No additional structure is seen here and thus these residuals approximate to be uncorrelated. Thus the STAR($2_{1,0}$) model is adequate for the pre-I_1 process.

3.6.3.2 Modeling the Post-I_1 Dynamic Components

Following the dynamic component modeling procedures as shown in the figure 3-9, the mean shift function $\delta_i(t)$ i=1,2,...,6 for 11<t<82 (e.g. for the post-I_1 period) are estimated. These estimated values, $\delta_i(t)$, are plotted in figure 3-12(a)-(f) for locations 1 through 6 respectively. The $\delta_i(t)$ in the pre-I_2 period characterizes the effect of I_1 intervention, while the $\delta_1(t)$ in the post-I_2 periods should be

Table 3-4(a) The Sample Space-Time Autocorrelation Functions and The Standardized Sample S-T Autocorrelation Functions of the Pre-I_1 Residuals.

	Space-Time Autocorrelations			Standardized S-T Autocorrelations		
Space Lag / Time Lag	0	1	2	0	1	2
1	-0.019	-0.001	-0.004	-0.125	-0.005	-0.024
2	0.113	0.091	0.102	0.679	0.543	0.612
3	0.100	0.140	0.081	0.546	0.766	0.441
4	-0.107	-0.067	-0.060	-0.525	-0.326	-0.294
5	-0.105	-0.187	-0.144	-0.445	-0.792	-0.611
6	-0.022	-0.041	-0.060	-0.077	-0.144	-0.207
7	-0.278	-0.146	-0.261	-0.682	-0.357	-0.639

(b). The Sample Space-Time Partial Autocorrelation Functions and The Standardized Sample S-T Partials of The Pre-I_1 Residuals.

	Space-Time Autocorrelations			Standardized S-T Autocorrelations		
Space Lag / Time Lag	0	1	2	0	1	2
1	-0.019	0.011	0.010	-0.141	0.078	0.072
2	0.112	0.069	0.049	0.781	0.481	0.342
3	0.106	0.156	-0.019	0.680	1.009	-0.122
4	-0.121	-0.050	-0.006	-0.728	-0.300	-0.034
5	-0.150	-0.286	-0.090	-0.823	-1.568	-0.491
6	-0.031	0.057	0.181	-0.149	0.278	0.885
7	-0.212	0.059	-0.173	-0.899	0.251	-0.732

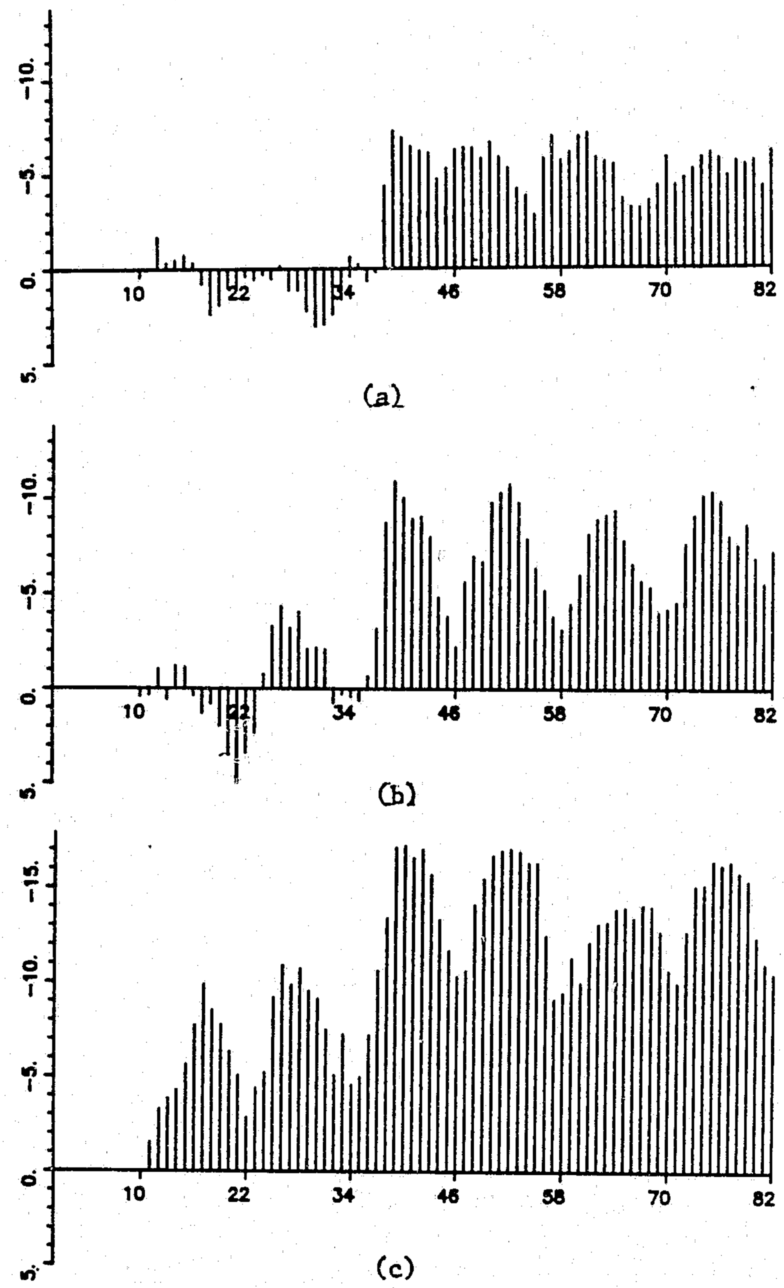


Figure 3-12. The Dynamic Component $\hat{\delta}_1(t)$
(a) at Azusa
(b) at Burbank
(c) at Lennox

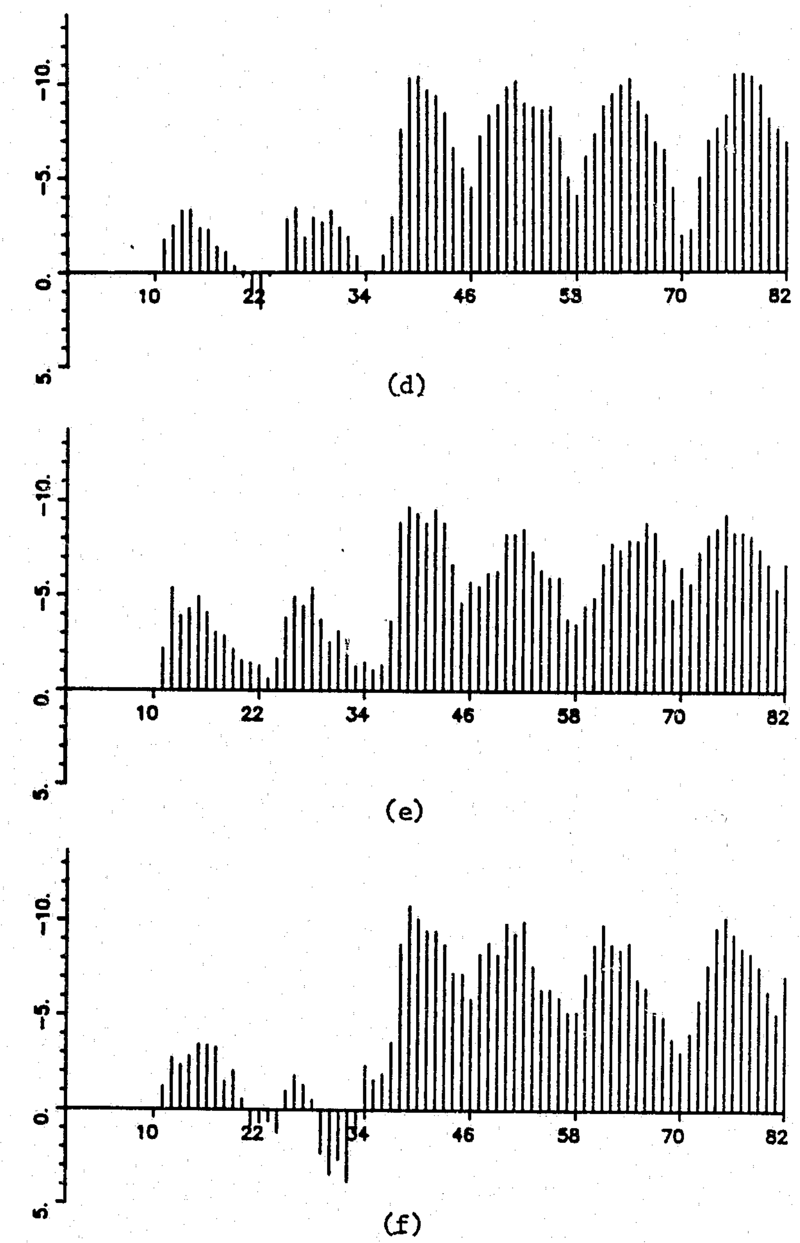


Figure 3-12. (Cont'd)
(d) at Long Beach
(e) at Downtown LA
(f) at LA County

interpreted as the effect of I_2 given that I_1 has been initiated in this system.

From figure 3.12 the I_2 intervention is seen to shift the mean shift function instantaneously for each location suggesting its effect to be additive not interactive with I_1. Thus, although in general, the effect of multiple interventions that overlaps should be viewed as conditional depending on the physical nature of the intervention they may be additive. Thus the I_2 intervention which is solely a measurement change does not appear to interact with the engine design change as expected apriori. In addition, from the data plots in figure 3-11(a)-(f), we see that at April, 1978 the air pollution levels of these six locations were about the same. Thus even under a nonlinear transformation assumption, i.e. the transformation between readings of different calibration methods are nonlienar nature and state dependent, the I_2 intervention had the same level change effect and the same scale factor at these six locations would be appropriate. That is, the mean shift vector $\underset{\sim}{\delta_3} = \delta_3 \underset{\sim}{1}$ and the scale factor matrix between measurements, $D = dI$.

From the initiation point of the engine design modification the pattern of $\delta_i(t)$ which is not a constant shift but rather exhibits seasonal fluctuations about a change in mean level indicates that the I_1 intervention is environmentally influenced. Figures 3-13(a),(b) exhibit the K(t) values for the six locations and figures 3-14(a)-(f) contains the $\delta(t)$ versus K(t) plots for each location. From the latter plots the appropriate form of the dynamics of the intervention can be identified. From these plots we see an apparant change in slop between the post-I_1, pre-I_2 segment and the post-I_2 segment. Regression fits

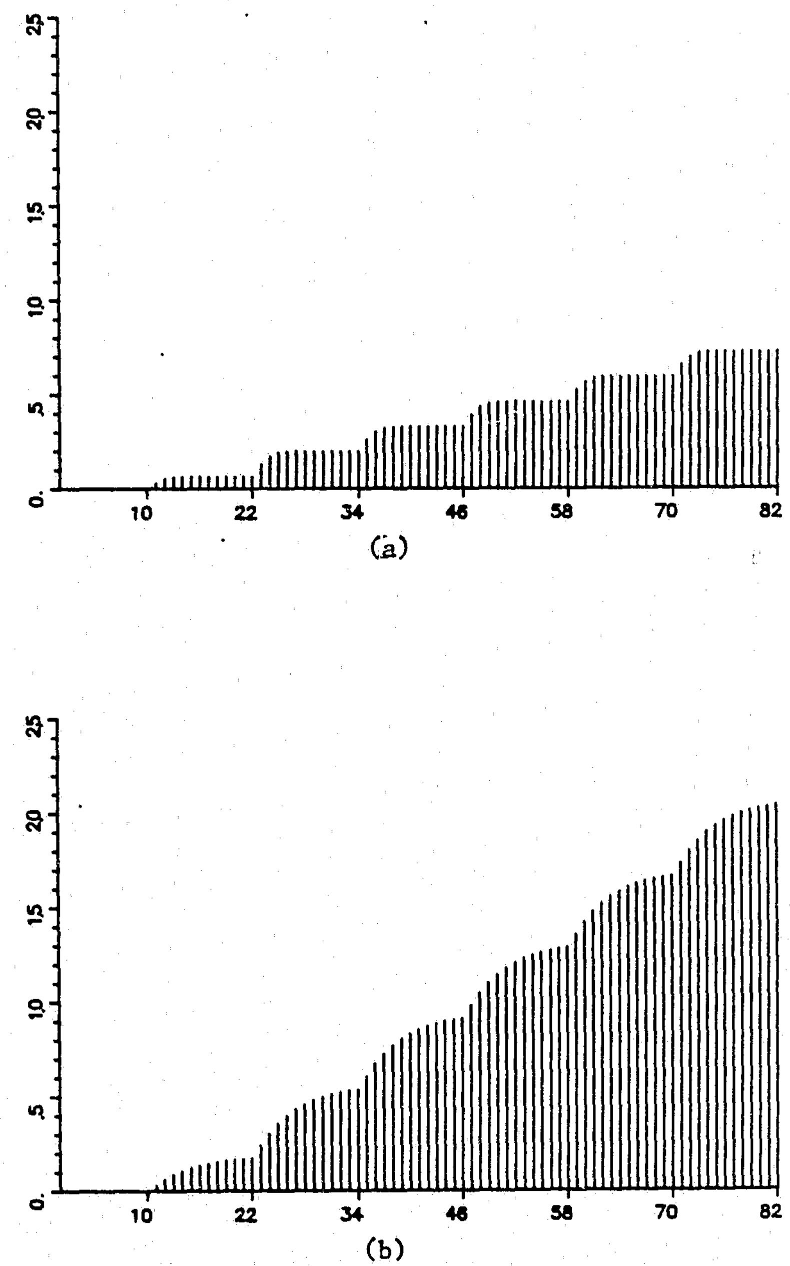


Figure 3-13. K(t) Values for Six Measurement Sites.
(a) Azusa
(b) Other Sites

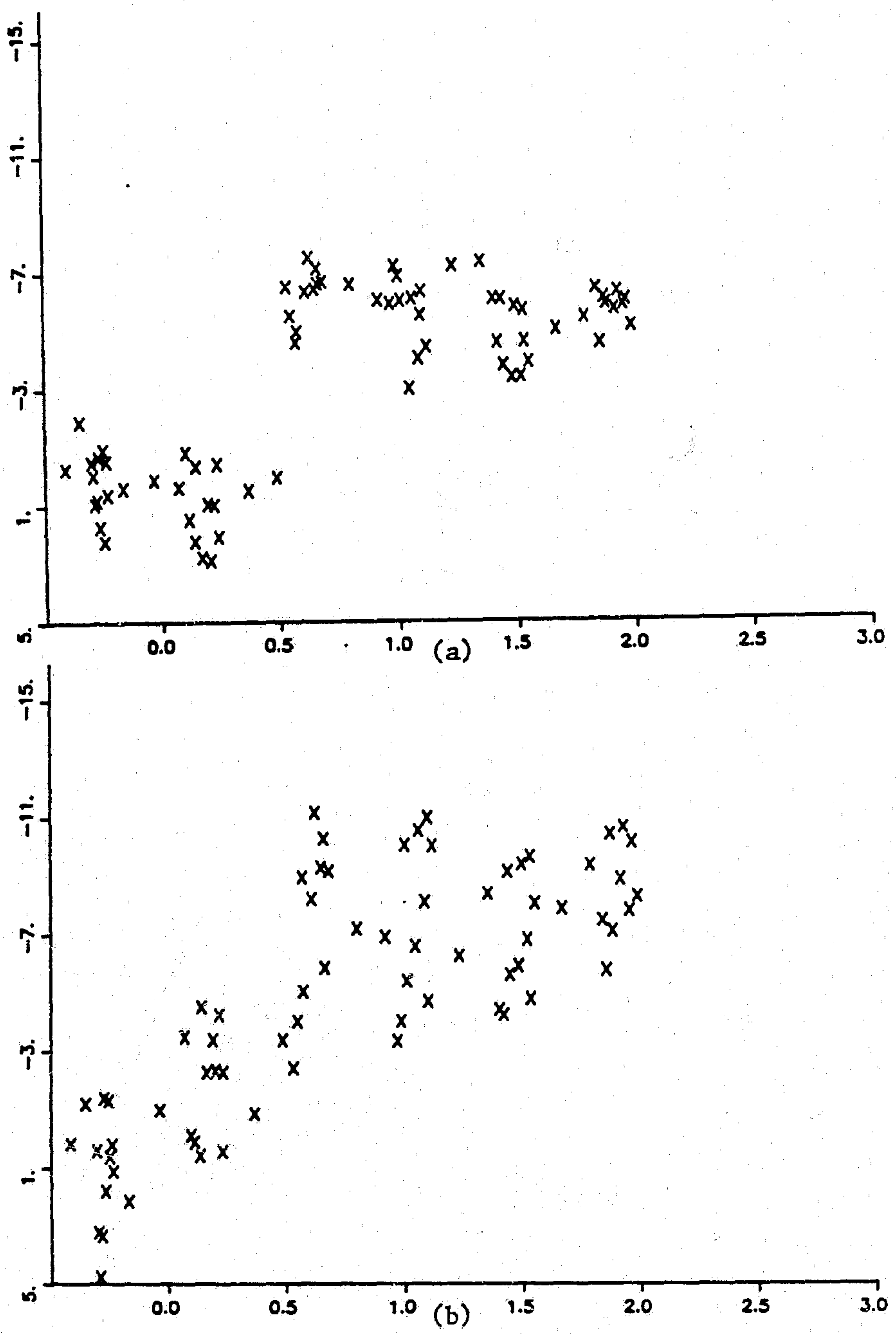


Figure 3-14. Post-I_1 $\underset{\sim}{\hat{\delta}}(t)$ vs. K(t) Plot.

(a) Azusa
(b) Burbank

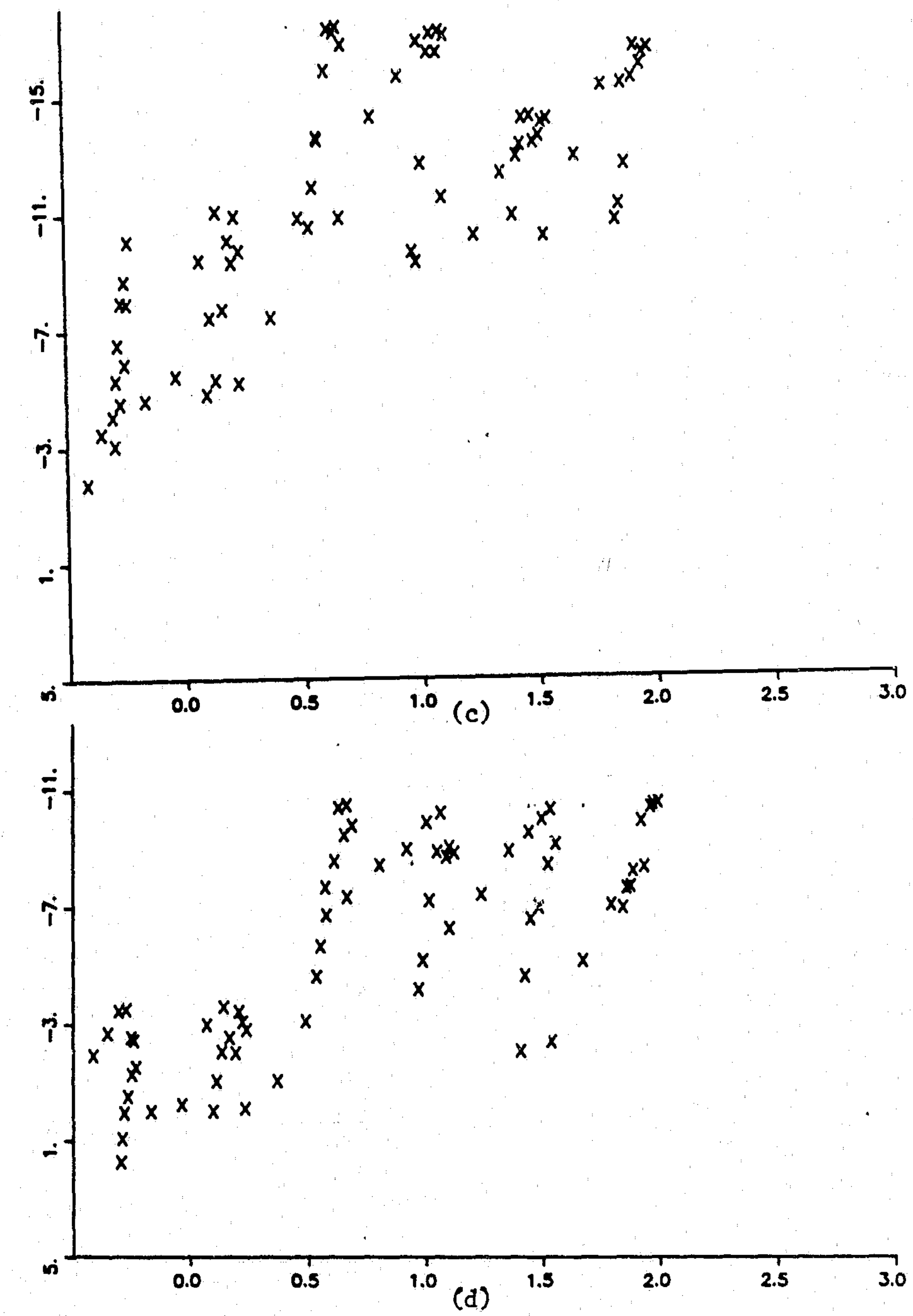


Figure 3-14. (Cont'd)
(c) Lennox
(d) Long Beach

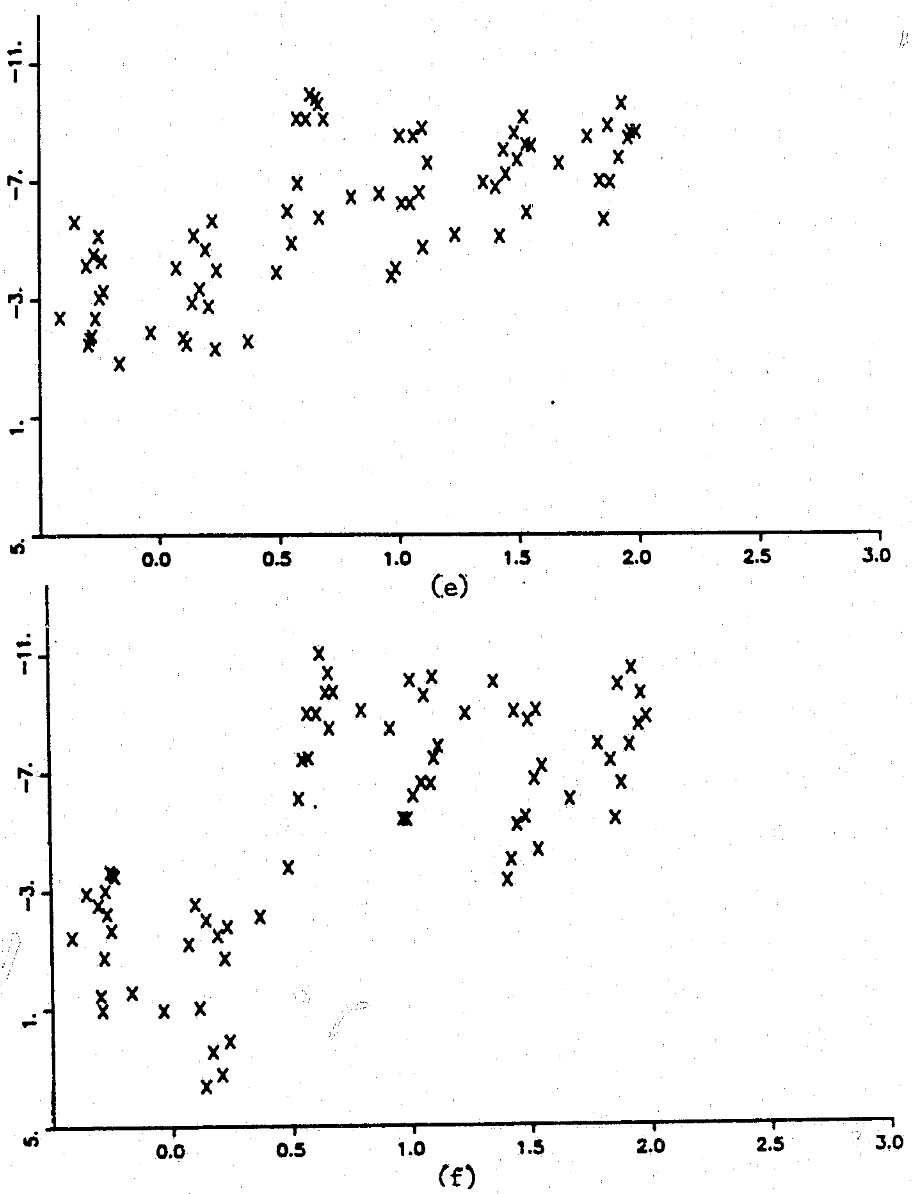


Figure 3-14. (Cont'd)
(e) Downtown LA
(f) LA County

to each of these segments for all six locations were developed for first and second order polynomials of the form;

$$\delta(t) = \beta_0 + \beta_1\left(K(t)-\bar{K}_t\right) + \beta_2\left(K(t)-\bar{K}_t\right)^2.$$

Table 3-5 contains the estimated parameters. For the post-I_1, pre-I_2 segment only in locations 3, 4 and 5 (Lennox, Long Beach and Downtown LA) is there a significant shift in carbon monoxide tentatively identified. Further only in Lennox is there any indication that there is a 2nd order transient effect. For the post I_2 period, all sites are seen to have an instantaneous measurement effect (e.g. the slope β_1 being statistically significant). Basing on the above considerations, the following tentative model for post-I_1 periods is identified.

$$\begin{cases} \text{Post-}I_1,\ \text{pre-}I_2: \\ (I-\psi_{10}\, IB-\psi_{11}\, W^{(1)}B-\psi_{20}\, IB^2)(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \underset{\sim}{A}_t + \underset{\sim}{\xi}^{(1)}(t), \\ \\ \text{Post-}I_2: \\ (I-\psi_{10}\, IB-\psi_{11}\, W^{(1)}B-\psi_{20}\, IB^2)(\underset{\sim}{\delta}_3\, \underset{\sim}{1}+d(\underset{\sim}{Z}_t-\underset{\sim}{\mu})) = \underset{\sim}{A}_t + \xi^{(2)}(t) \end{cases} \qquad (3\text{-}73)$$

where $\underset{\sim}{\xi}^{(1)}(t) = \underset{\sim}{\xi}_t^{(1)}\, \underset{\sim}{\delta}^{(1)}$ as defined in equation 3-63,

$\underset{\sim}{\xi}^{(2)}(t) = \underset{\sim}{\xi}_t^{(1)}\, \underset{\sim}{\delta}^{(2)}$, $\underset{\sim}{\delta}^{(2)} = [\delta_1^{(2)}, \delta_2^{(2)}, \ldots, \delta_6^{(2)}]^t$,

δ_3 and d are the mean change and the scale faactor between readings of different calibration methods.

Table 3-5 Mean Shift Function Forms.

(Post I_1, Pre I_2)				
location	β_0 (Var)	β_1 (Var)	β_2 (Var)	σ^2
1	0.748(0.152)	2.435(2.249)	-	1.337
2	-0.174(0.169)	-2.425(1.168)	-	4.399
3	-8.268(0.434)	-3.639(1.169)	11.200(13.306)	4.014
4	-1.437(0.082)	-0.079(0.568)	-	2.138
5	-2.916(0.768)	0.826(0.530)	-	1.947
6	-0.637(0.136)	2.483(0.935)	-	3.522

(Post I_2)				
location	β_0 (Var)	β_1 (Var)	β_2 (Var)	σ^2
1	-5.335(0.041)	0.181(0.506)	-	1.902
2	-7.279(0.119)	-0.292(0.145)	-	5.495
3	-13.810(0.132)	0.264(0.195)	-	6.076
4	-7.683(1.09)	-0.094(0.161)	-	4.992
5	-7.115(0.059)	-0.147(0.087)	-	2.708
6	-7.513(0.081)	0.449(0.119)	-	3.711

3.6.3.3 Modeling the Post-I_1 Space-Time Intervention Model

Keeping the pre-I_1 model, equation (3-72) unchanged, the M.L. estimation has been performed to give the conditional M.L. estimates of $\underset{\sim}{\psi}$, $\underset{\sim}{\delta}^{(1)}$, $\underset{\sim}{\delta}^{(2)}$, δ_3. This resulted in non-significance of $\psi_{11} = 0.004$. The associated extra sum of square $SS_E = 0.432$ being less than the corresponding critical F for reasonable α levels. Therefore the ψ_{11} term was dropped from the model, and the M.L. estimation for the following reduced model was obtained.

$$
\begin{cases}
\text{Pre-}I_1,\ \text{Post-}I_2: & (I-\psi_{10}\ IB-\psi_{20}\ IB^2)(\underset{\sim}{Z}_t-\underset{\sim}{\mu}) = \underset{\sim}{A}_t + \underset{\sim}{\xi}^{(1)}(t) \\
\text{Post-}I_2: & (I-\psi_{10}\ IB-\psi_{20}\ IB^2)(\delta_3\ \underset{\sim}{1} + d(\underset{\sim}{Z}_t-\underset{\sim}{\mu})) = \underset{\sim}{A}_t + \xi^{(2)}(t)
\end{cases}
\tag{3-74}
$$

The conditional M.L. estimation results are;

Model Parameter	C.M.L. Estimate	95% C.I.
ψ_{10}	0.900	(0.821, 0.979)
ψ_{20}	-0.247	(-0.324, -0.170)
$\delta_1^{(1)}$	1.647	(-1.119, 4.414)
$\delta_2^{(1)}$	-2.574	(-5.350, 0.193)
$\delta_3^{(1)}$	-14.895	(-17.660,-12.130)
$\delta_4^{(1)}$	-3.674	(-6.440, -0.907)

$\delta_5^{(1)}$	-5.677	(-8.443, -2.910)
$\delta_6^{(1)}$	-1.382	(-4.148, 1.385)
$\delta_1^{(2)}$	-2.105	(-1.148, 0.727)
$\delta_2^{(2)}$	-1.503	(-2.441, -0.566)
$\delta_3^{(2)}$	-5.410	(-6.347, -4.472)
$\delta_4^{(2)}$	-1.687	(-2.628, -0.750)
$\delta_5^{(2)}$	-1.300	(-2.237, -0.362)
$\delta_6^{(2)}$	-1.534	(-2.471, -0.596)
δ_3	-5.123	(-6.128, -4.117)

and

$\sigma_A^2 = 1.517$, $d = 1.048$

3.6.4 Refining the Noise Model and Diagnostic Checking

From the estimated residuals $\underset{\sim}{A}_t$ of the post-I_1 model (equation 3-76), the sample space-time autocorrelation functions and the sample space-time partial autocorrelation functions were estimated. Table 3-6(a) and 3-6(b) contain these autocorrelation functions and their standardized forms. The autocorrelations are seen to repeat in blocks of size 11 indicating the need of seeeasonal components in the noise model. Furthr, within each block, the autocorrelation and partial autocorrelation functions appar to tail off for spatial log 0 and 1. Thus a tentative noise model for the post-I_1 period is the seasonal STARMA$(1_1,0,1_1) \times (1_1,0,0)_{11}$ model. It's form is,

$$(I - \phi_{10}^N IB - \phi_{11}^N W(1)B)(I - \phi_{11,0}^N IB^{11} - \phi_{11,1}^N W^{(1)}B^{11}) \underset{\sim}{A}_t \quad (3\text{-}75)$$

$$= (I - \theta_{10}^N IB - \theta_{11}^N W^{(1)}B) \underset{\sim}{\varepsilon}_t .$$

The M.L. point estimates and their associated 95% confidence intervals are;

	95% C.I.
$\phi_{10}^N = 0.698$	(0.311, 1.085)
$\phi_{11}^N = -0.145$	(-0.614, 0.325)
$\phi_{11,0}^N = 0.060$	(-0.057, 0.178)
$\phi_{11,1}^N = 0.389$	(0.244, 0.535)
$\theta_{10}^N = 0.575$	(0.137, 1.014)
$\theta_{11}^N = -0.211$	(-0.746, 0.324)
$\sigma_\varepsilon^2 = 1.236.$	

The residuals of this fitted model $\underset{\sim}{\varepsilon}_t$, were computed and their sample space-time autocorrelation functions were estimated. Table 3-7 contains these autocorrelations and their standardized forms. From this table no additional identifiable structure is seen. An overall test of the adequacy of this model in that there is no additional structure (the residuals are uncorrelated) can be made using χ^2 and F tests. The standard portmanteau χ^2 test used in univriate modeling is not appropriate since the zeroth, first and second spatial lags are not

Table 3-6

(a) The Sample Space-Time Autocorrelation Functions and The Standardized S-T Autocorrelation Functions of the Rersiduals of the Post-I_1 Model.

Space Lag / Time Lag	Space-Time Autocorrelations 0	1	2	Standardized S-T Autocorrelations 0	1	2
1	0.16	0.17	0.05	3.38	3.49	1.05
2	0.20	0.12	0.10	4.13	2.54	2.15
3	0.15	0.07	0.09	3.11	1.59	1.84
4	-0.08	-0.19	-0.10	-1.61	-3.83	-2.14
5	-0.17	-0.16	-0.05	-3.51	-3.15	-1.10
6	-0.15	-0.23	-0.02	-3.03	-4.58	-0.53
7	-0.27	-0.25	-0.19	-5.26	-4.92	-3.78
8	-0.06	-0.03	0.03	-1.21	-0.64	0.75
9	0.03	0.09	0.09	0.67	1.88	1.74
10	0.04	0.03	-0.04	0.82	0.74	-0.84
11	0.26	0.39	0.22	5.04	7.38	4.25
12	0.19	0.22	0.05	3.59	4.17	0.98
13	0.15	0.26	0.07	2.90	4.84	1.46
14	0.21	0.21	0.16	3.89	4.00	2.95
15	-0.18	-0.14	-0.09	-3.33	-2.69	-1.77
16	-0.13	-0.13	-0.08	-2.45	-2.35	-1.45
17	-0.23	-0.22	-0.09	-4.27	-3.93	-1.68
18	-0.36	-0.37	-0.22	-6.43	-6.63	-3.94
19	-0.23	-0.23	-0.15	-4.12	-4.09	-2.79
20	-0.18	-0.21	-0.11	-3.14	-3.67	-1.95
21	-0.16	-0.10	-0.13	-2.90	-1.74	-2.26
22	0.26	0.22	0.18	4.42	3.87	3.14
23	0.15	0.17	0.02	2.53	2.89	0.47
24	0.23	0.21	0.08	3.96	3.55	1.38
25	0.32	0.31	0.21	5.26	5.11	3.50
26	0.06	0.02	-0.05	1.00	0.36	-0.97
27	-0.00	0.03	0.06	-0.13	0.52	1.02
28	-0.08	-0.11	-0.03	-1.34	-1.84	-0.48
29	-0.25	-0.29	-0.24	-3.95	-4.57	-3.83
30	-0.23	-0.21	-0.12	-3.71	-3.36	-1.99

Table 3-6

(b) The Sample Space-Time Partial Autocorrelation Functions and The Standardized S-T Partials of the Rersiduals of the Post-I_1 Model.

Space Lag / Time Lag	Space-Time Autocorrelations 0	1	2	Standardized S-T Autocorrelations 0	1	2
1	0.16	0.13	-0.08	3.43	2.87	-1.68
2	0.17	-0.01	0.04	3.51	-0.33	0.97
3	0.10	-0.06	0.05	2.21	-1.39	1.21
4	-0.15	-0.31	-0.05	-3.19	-6.41	-1.17
5	-0.19	-0.02	0.03	-3.86	-0.59	0.63
6	-0.08	-0.10	0.13	-1.67	-2.16	2.77
7	-0.13	-0.04	-0.11	-2.75	-0.79	-2.19
8	0.04	0.14	0.08	0.82	2.90	1.72
9	0.14	0.20	0.10	2.78	3.95	2.03
10	0.02	-0.07	-0.09	0.48	-1.45	-1.91
11	0.14	0.25	0.03	2.81	4.87	0.75
12	0.02	0.02	-0.09	0.49	0.43	-1.86
13	0.03	0.19	0.00	0.62	3.75	0.06
14	0.04	0.02	0.11	0.88	0.52	2.07
15	-0.23	-0.22	0.04	-4.36	-4.13	0.85
16	-0.09	-0.00	-0.00	-1.65	-0.08	-0.14
17	-0.16	-0.12	-0.08	-2.93	-2.20	-1.49
18	-0.11	-0.04	-0.09	-1.98	-0.84	-1.65
19	-0.02	-0.04	-0.15	-0.37	-0.88	-2.77
20	-0.14	-0.23	-0.09	-2.48	-4.12	-1.74
21	-0.07	0.02	0.01	-1.33	0.45	0.29
22	0.17	0.05	0.12	2.99	0.94	2.11
23	0.04	-0.01	0.07	0.80	-0.33	1.23
24	0.16	0.08	0.08	2.72	1.44	1.37
25	0.11	0.02	0.04	1.90	0.48	0.80
26	0.01	0.05	-0.06	0.20	0.99	-1.13
27	-0.01	0.23	0.16	-0.22	3.84	2.75
28	-0.03	0.10	-0.01	-0.64	1.69	-0.29
29	0.07	0.15	-0.19	1.23	2.47	-3.12
30	-0.11	0.02	-0.19	-1.77	0.36	-3.04

independent. However since the current model doesn't have any seecond order spatial terms the sample autocorrelations for the second order spatial terms can be checked for adequacy, (being uncorrelated or informationless using the χ^2 statistic). If the computed χ^2 statistic is insignificant, the magnitude of the statistic could have come about by chance alone and thus can be used to check the adequacy of the proceeding spatial lags usig an F test.

From Table 3-7 we have

$$\sum_{K=1}^{30} \rho_{K0}^2 = 77.71$$

$$\sum_{K=1}^{30} \rho_{K1}^2 = 58.27$$

$$\sum_{K=1}^{30} \rho_{K2}^2 = 42.15$$

For $\alpha = 0.05$, the theoretical $\chi^2_{30,.05} = 43.77$ and

$$\chi^2 = 42.15 < \chi^2_{30,.05}.$$

Thus the second spatial lag is uncorrelated and the magnitude 42.15 is associated with chance error. For the first spatial lag,

$$F_1 = 58.27/42.15 = 1.38.$$

Table 3-7
The Space-Time Autocorrelation Functions and the Standardized S-T Autocorrelation Functions of the Rersiduals, $\underset{\sim}{\varepsilon}_t$, of the Post-$I_1$ Noise Model.

	Space-Time Autocorrelations			Standardized S-T Autocorrelations		
Space Lag / Time Lag	0	1	2	0	1	2
1	-0.03	0.00	-0.02	-0.78	0.19	-0.49
2	0.03	-0.01	0.00	0.78	-0.26	0.12
3	0.08	0.04	0.03	1.70	0.81	0.67
4	0.01	-0.04	-0.04	0.28	-0.88	-0.85
5	-0.09	-0.03	-0.01	-1.95	-0.70	-0.21
6	-0.04	-0.07	0.04	-0.87	-1.49	0.82
7	-0.13	-0.08	-0.11	-2.59	-1.64	-2.19
8	-0.02	0.01	0.06	-0.54	0.19	1.29
9	0.06	0.12	0.09	1.30	2.38	1.88
10	0.00	-0.03	-0.08	0.12	-0.62	-1.54
11	-0.03	-0.04	-0.02	-0.61	-0.90	0.41
12	0.09	0.07	0.00	1.73	1.47	0.14
13	0.04	0.09	0.00	0.85	1.83	0.00
14	0.10	0.06	0.06	1.94	1.17	1.21
15	-0.13	-0.11	-0.03	-2.46	-2.00	-0.64
16	-0.04	-0.02	-0.05	-0.88	-0.44	-0.92
17	-0.08	-0.05	-0.02	-1.51	-1.05	-0.50
18	-0.14	-0.12	-0.03	-2.58	-2.19	-0.70
19	-0.09	-0.06	-0.08	-1.58	-1.18	-1.44
20	-0.11	-0.15	-0.09	-2.00	-2.63	-1.59
21	-0.06	-0.00	-0.03	-1.15	-0.16	-0.67
22	0.11	0.07	0.09	1.99	1.20	1.61
23	0.04	0.02	-0.02	0.70	0.36	-0.42
24	0.04	-0.00	-0.02	0.69	-0.02	-0.40
25	0.09	0.07	0.05	1.61	1.19	0.96
26	0.03	0.00	-0.04	0.54	0.05	-0.67
27	-0.00	0.05	0.07	-0.03	0.81	1.20
28	0.00	-0.00	0.00	0.04	-0.03	0.14
29	-0.06	-0.06	-0.09	-1.01	-1.04	-1.48
30	-0.05	-0.01	-0.01	-0.90	-0.20	-0.28

$$F_1 = 1.38 < F_{30,30,.05} = 1.84$$

Also since the first spatial lag is uncorrelated,

$$F_0 = 77.71/(\frac{42.15 + 58.27}{2}) = 1.44 < F_{30,60,.05} = 1.65$$

Thus the space-time residuals are uncorrelated in both space and time. Therefore the model is accepted as statistically adequate.

Therefore we have the overall intervention model as follows;

(3-76)

Pre-I_1 process:

$$(I - (\Phi_{10}\, I + \Phi_{11} W^{(1)})\, B - \Phi_{20}\, IB^2)(\underset{\sim}{Z}_t - \underset{\sim}{\mu}) = \underset{\sim}{\varepsilon}_t , \quad t < 10$$

Post-I_1, Pre-I_2 process:

$$(I - \psi_{10}\, IB - \psi_{20}\, IB^2)(\underset{\sim}{Z}_t - \underset{\sim}{\mu}) = \underset{\sim}{A}_t + \xi_t^{(1)} \underset{\sim}{\delta}^{(1)} , \quad 11 < t < 37$$

Post-I_2 process:

$$(I - \psi_{10}\, IB - \psi_{20}\, IB^2)(\delta_3 \underset{\sim}{1} + d(\underset{\sim}{Z}_t - \underset{\sim}{\mu})) = \underset{\sim}{A}_t + \xi_t^{(1)} \underset{\sim}{\delta}^{(2)}, \quad 38 < t < 82$$

The noise process of $\underset{\sim}{A}_t$:

$$(I - (\Phi^N_{10}\, I + \Phi^N_{11}\, W^{(1)})B)(I - (\Phi^N_{11,0}\, I + \Phi^N_{11,1}\, W^{(1)})B^{11})\, \underset{\sim}{A}_t$$

$$= (I - (\Theta^N_{10}\, I + \Theta^N_{11}\, W^{(1)})B)\, \underset{\sim}{\varepsilon}_t$$

Where the model parameter values Φ_{10}, Φ_{11}, Φ_{20} have been listed in equation 3-72, ψ_{10}, ψ_{20}, $\underset{\sim}{\delta}^{(1)}$, $\underset{\sim}{\delta}^{(2)}$, δ_3 and d have been listed in equation 3-74, Φ^N_{10}, Φ^N_{11}, $\Phi^N_{11,0}$, Θ^N_{10}, Θ^N_{11} have been listed in equation 3-75, and ξ_t has been listed in equation 3-63.

Note that the autoregressive operator,

$$(I-(0.698I-0.145W^{(1)})B)(I - (0.060I+0.389W^{(1)})B^{11})$$

$$= I - (0.698I-0.1445W^{(1)})B - (0.060I+0.389W^{(1)})B^{11}$$

$$- (-0.04I-0.263W^{(1)}+0.056W^{(2)})B^{12}$$

reveals that the noise process contains both seasonal terms of an 11 months lag and a 12 months lag. Also, the space-time autocorrelation seasonal pattern of 11 months lag of $\underset{\sim}{A}_t$ in the table 3-6(a) is similar to the seasonal pattern of 12 months lag. This indicates that the noise process doesn't repeat the seasonal mechanism exactly every 12 months, instead the seasonal mechanism repeats itself somewhere between 11 months and 12 months. That is, for this data a non-integer seasonal lag between S=11 and S=12 would be appropriate. If this were done a model with even fewer parameters would be obtained, since the correlative structure associated with S=11 and S=12 in the current model form are similar.

3.6.5 Checking Alternative Forms: The No Feedback Structure

The current model which is statistically adequate, suggests another modeling alternative. This alternative is:

$$(3\text{-}77)$$

Pre-I_1 process:

$$(I - (\phi_{10}I + \phi_{11}W^{(1)})B - \Phi_{20}IB^2)\ (\underset{\sim}{Z}_t - \underset{\sim}{\mu}) = \underset{\sim}{\varepsilon}_t$$

Post - I_1, pre-I_2 process:

$$(I-\psi_{10}IB - \psi_{20}IB\ )(\underset{\sim}{Z}_t-\underset{\sim}{\mu})$$

$$= \frac{(I - (\Theta^{+}_{10}I+\Theta^{+}_{11}W^{(1)})B)}{(I - (\Phi^{+}_{1o}I+\Phi^{+}_{11}W^{(1)})B)(I - (\Phi^{+}_{11,0}I+\Phi^{+}_{11,1}W^{(1)})B^{11})}\ (\underset{\sim}{\varepsilon}_t+\xi_t\delta^{(1)})$$

Post-I_1, Pre-I_2 process:

$$(I - \psi_{10}IB - \psi_{20}IB^2)(\delta_3\underset{\sim}{1}+d(\underset{\sim}{Z}_t-\underset{\sim}{\mu}))$$

$$= \frac{I - (\Theta^{+}_{10}I+\Theta^{+}_{11}W^{(1)})B}{(I - (\Phi^{+}_{10}I+\Phi^{+}_{11}W^{(1)})B)(I - (\Phi^{+}_{11,0}I+\Phi^{N}_{11,1}W^{(1)})B^{11})}\ (\underset{\sim}{\varepsilon}_t+\xi_{\underset{\sim}{t}}\delta^{(2)})$$

This model differs from the model, equation 3-76, in one respect. Here the intervention effect follows the exactly environmental process that the noise input follows, whereas the interventional effect of the model, equation 3-76, follows a process that is similar to the environment process. These two modeling alternatives' structures are illustrated in the Figure 3-15(a) and (b).

Figure 3-15(a) illustrate the situation that there is no feedback from the environment process, thus the intervention effect directly enters the evironmental process and is influenced by it solely. In the figure 3-15(b), the situation where there is feedback of some degree and the intervention effect that enters the environmet process will follow a modified environmental process $T_\xi(B)$. The

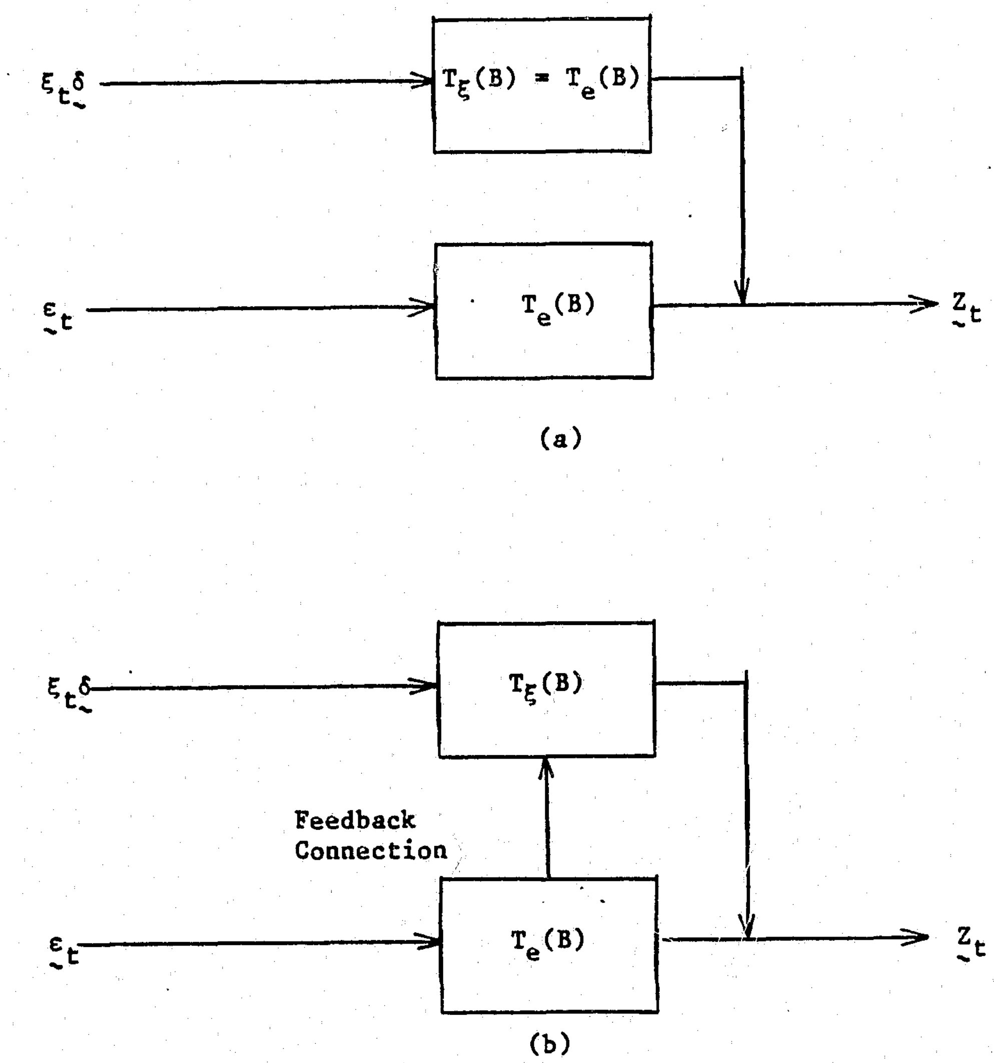


Figure 3-15 (a). Intervention Process Follows Exactly the Environment Process, i.e., No Environment Feedback.

(b). Intervention Process Gets Feedback from the Environment Process and Follows the Environment Process Partially.

existence of the feedback loop can be interpreted as the intervention effect behaves non-linearly, e.g. the magnitude of the realized effect is dependent upon the level the system is operating. Thus when there is no feedback the intervention exerts the same influence regardless of the operating level of the system.

The post-I_1 model of the non-feedback model can be rewritten equivaletly as;

(3-78)

Post-I_1, Pre-I_2 process:

$$(I - \sum_{\ell=0}^{1} \phi^{+}_{11,\ell} W^{(\ell)} B^{11})(I - \sum_{K=1}^{2} \sum_{\ell=0}^{1} \phi^{+}_{K\ell} W^{(\ell)} B^{K})(\underset{\sim}{Z}_t - \underset{\sim}{\mu})$$
$$= (I - (\Theta^{+}_{10} + \Theta^{+}_{11} W^{(1)})B)(\underset{\sim}{\varepsilon}_t + \xi_t \underset{\sim}{\delta}^{(1)})$$

Post-I_2 process:

$$(I - \sum_{\ell=0}^{1} \phi^{+}_{11,\ell} W^{(\ell)} B^{11})(I - \sum_{K=1}^{2} \sum_{\ell=0}^{1} \phi^{+}_{K\ell} W^{(\ell)} B^{K})(\delta_3 \underset{\sim}{1} + d(\underset{\sim}{Z}_t - \underset{\sim}{\mu}))$$
$$= (I - (\Theta^{+}_{10} I + \Theta^{+}_{11} W^{(1)})B)(\underset{\sim}{\varepsilon}_t + \xi_t \underset{\sim}{\delta}^{(2)})$$

The conditional M.L. estimates of the model parameters and the associated 95% confidence intervals are;

			95% C.I.
ϕ^{+}_{10}	=	1.514	(1.249, 1.779)
ϕ^{+}_{11}	=	0.053	(-0.268, 0.375)

ϕ^{+}_{20}	=	-0.636	(-0.904, -0.368)
ϕ^{+}_{21}	=	0.093	(-0.434, 0.248)
$\phi^{+}_{11,0}$	=	0.092	(0.037, 0.146)
$\phi^{+}_{11,1}$	=	0.243	(0.177, 0.308)
Θ^{+}_{10}	=	0.923	(0.646, 1.201)
Θ^{+}_{11}	=	-0.400	(-0.739, -0.069)
$\delta^{(1)}_{1}$	=	5.349	(7.869, 2.830)
$\delta^{(1)}_{2}$	=	-1.985	(0.535, -4.504)
$\delta^{(1)}_{3}$	=	-11.040	(-8.521, -13.56)
$\delta^{(1)}_{4}$	=	-3.370	(-0.850, -5.889)
$\delta^{(1)}_{5}$	=	-4.182	(-1.663, -6.701)
$\delta^{(1)}_{6}$	=	-0.739	(1.781, -3.258)
$\delta^{(2)}_{1}$	=	-0.863	(1.136, -2.861)
$\delta^{(2)}_{2}$	=	-1.183	(-0.469, -1.898)
$\delta^{(2)}_{3}$	=	-3.577	(-2.856, -4.297)
$\delta^{(2)}_{4}$	=	-1.287	(-0.506, -2.007)
$\delta^{(2)}_{5}$	=	-0.954	(-0.228, -1.680)
$\delta^{(2)}_{6}$	=	-1.075	(-0.344, -1.805)
δ_3	=	-4.906	(-4.057, -5.755)
b	=	1.202	
σ^2_{ε}	=	1.259	

The residuals of the non-feedback model were computed and their sample space-time autocorrelation function estimated. Table 3.8 contains these estimates and their standardized form. Since the zeroth spatial lag, 1st spatial lag and the 2nd spatial lag autocorrelation function, i.e. ρ_{K0}, ρ_{K1}, ρ_{K2}, K=1,2,...,30, are not independent, so it is not appropriate to perform the χ^2 test on all these 90 autocorrelations. Instead, the χ^2 test is performed on the 2nd lag autocorrelation functions first, and then followed by the F test to test the significance of differences of the population of $\rho_{K,0}$, ρ_{K1} and ρ_{k2}.

From table 3-8, we have

$$\sum_{K=1}^{30} \rho_{K0}^2 = 92.68,$$

$$\sum_{K=1}^{30} \rho_{K1}^2 = 83.84,$$

$$\sum_{K=1}^{30} \rho_{K2}^2 = 78.12.$$

Even with $\alpha = 0.01$,

$$\chi^2 = 78.12 > \chi^2_{30,.01} = 50.89,$$

this test can't be past to conclude that the ρ_{K2}, K=1,2,...,30 are uncorrelated.

Based on these diagnostic checks the non-feedback model alternative is seen to be statistically inadequate.

Table 3-8
The Sample Space-Time Autocorrelation Functions and the Standardized Sample S-T Autocorrelation Functions of the Residuals of the Post-I_1 Model without Any Feedback from the Environment Process.

Space Lag / Time Lag	Space-Time Autocorrelations 0	1	2	Standardized S-T Autocorrelations 0	1	2
1	-0.01	0.01	-0.01	-0.38	0.32	-0.38
2	-0.05	-0.03	0.00	-1.02	-0.83	0.18
3	0.08	0.05	0.05	1.62	1.17	1.10
4	0.06	0.00	-0.04	1.23	0.11	-0.91
5	-0.00	0.07	0.03	-0.17	1.46	0.78
6	0.02	-0.01	0.06	0.39	-0.19	1.25
7	-0.09	-0.03	-0.10	-1.84	-0.72	-2.10
8	0.02	0.07	0.06	0.47	1.45	1.30
9	0.11	0.15	0.10	2.12	3.03	2.07
10	-0.02	-0.10	-0.11	-0.47	-1.89	-2.10
11	-0.06	-0.03	0.03	-1.13	-0.71	0.61
12	0.07	0.05	-0.00	1.45	1.09	-0.08
13	0.04	0.10	-0.00	0.82	1.97	-0.13
14	0.10	0.06	0.09	1.85	1.14	1.81
15	-0.24	-0.22	-0.04	-4.41	-4.13	-0.86
16	-0.08	-0.06	-0.09	-1.47	-1.23	-1.63
17	-0.07	-0.00	-0.02	-1.37	-0.16	-0.37
18	-0.13	-0.07	-0.03	-2.37	-1.41	-0.58
19	-0.08	-0.04	-0.09	-1.47	-0.77	-1.67
20	-0.11	-0.14	-0.10	-1.96	-2.44	-1.85
21	-0.09	-0.00	-0.03	-1.51	-0.01	-0.53
22	-0.14	-0.11	-0.15	2.47	1.93	2.68
23	0.01	-0.00	-0.04	0.25	-0.02	-0.71
24	0.00	-0.07	-0.04	0.10	-1.26	-0.82
25	0.10	0.05	0.09	1.73	0.84	1.48
26	0.02	-0.01	-0.03	0.36	-0.17	-0.51
27	-0.00	0.09	0.13	-0.03	1.44	2.23
28	0.01	0.01	0.01	0.31	0.19	0.26
29	-0.06	-0.07	-0.16	-1.03	-1.17	-2.56
30	-0.02	-0.07	0.00	-0.32	1.17	0.10

3.6.6. Model Interpretation

In the model, equation (3-76), the intrinsic utility of I_1, which is represented by the terms that contain $\xi_t^{(1)}$, followed the noise component $\underset{\sim}{A}_t$, and it is interpereted as the situation that the realized effect of I_1 was influenced by the environment. While the intrinsic utility of I_2, which includes the mean shift $\delta_3 \underset{\sim}{1}$ and the scale factor d, didn't follow the noise process and the effect was realized instantaneously, this is interpreted as the situation that the realized effect of I_2 wasn't influenced by the environment process. In the model, equation (3-76), the intrinsic utility of I_1 didn't follow exactly the process that the white noise followed, i.e. the interventional input followed a modified environmental process, this modified environment process has been found its interpretation in the existence of environmental feedback.

A non-lienear investment-return system has the property that the farer the state is from the system equilibrium state, the bigger the gain will be, and the state of the non-linear system converges to the equilibrium state at a reducing rate. When the system is far away from the equilibrium, the convergence rate is high. The convergence rate, at the very beginning, is dramatically, non-linearly reduced, and this period is usually referred to as the transient period. After the transient period, the system state converges to the equilibrium state at a steady state convergence rate, and it is referred to as the steady state. The mean values of each location at pre-I_1 and pre-I_2 are listed below.

		Pre-I_1	Pre-I_2
1.	Azusa	9.3	10.7
2.	Burbank	13.5	11.2
3.	Lennox	20.0	9.6
4.	Long Beach	13.3	10.4
5.	Downtown LA	13.0	9.0
6.	LA County	12.9	12.2

The tendency, that the mean levels moved to about the same level at all locations, is seen in the above mean level list. In the model, equaiton (3-74), it is seen that between I_1 and I_2 only four locations, i.e. Lennox, Long Beach, Downtown LA and Burbank, receive significant impact. Here we see that Lennox has the highest pre-I_1, mean, Long Beach and Downtown LA, that are the first order neighbors of Lennox, and Burbank have high pre-I_1, mean, too. The absolute values of $\delta_i^{(1)}$, $i = 2,3,4,4$ are larger than the corresponding absolute values of $\delta_i^{(2)}$, $i = 2,3,4,5$. This means that the system state converged to the equilibrium state, which may be somewhere between 9.0 and 12.2, at reducing convergence rate, since the scaled factor $d = 1.048$ can't have such a slope-reduction power. These observations confirm that this is a nonlinear investment-return system. Since the pre-I_1 mean is highest at Lennox, so the gain of the engine design change is the biggest. The pre-I_2 period, when $\delta_3(t)$ is appropriately fitted into the second order form is interpreted as the transient period, and the steady state was reached when I_2 was initiated.

The effect of I_1 and I_2 on these six locations are plotted in figure 3-15(a)-(b) to compare the relative effects at the same time. The height of these three dimension plots reflect the magnitude of the effect at that location, at that plotted time.

The figures 3-16(a) and 3-16(b) show the effect of I_1 at t=22 and t=34, i.e. 1 year after I_1 and 2 years after I_1 respectively. The figures 3-15(c) and 3-15(d) show the effect at t=37 and t=38, i.e. immediately before and after the introduction of I_2 respectively. Here we see a significant reading shift at all these six locations by the same amount. The figures 3-16(e)-(h) show the effect of I_1 and I_2 after the initiation of I_2. The figures 3-16(e), (f), (g), (h) show the effect at t=46, 58, 70, 82, i.e. 3 years, 4 years, 5 years and 6 years after the introducing of I_1, respectively. From these plots, all the time we see that Lennox has the shortest effect, Downtown LA the second and Long Beach the third. Note that Downtown LA and Long Beach are the 1st order neighbors of Lennox.

From the model 3-76, we read that there was no diffusion phenomena in the intervention effect, and the magnitude of the diffusion mechanism of the noise process among 1-st order neighbors was reduced, since $|\phi_{11}| = 0.319$ while $|\phi_{11}^N| = 0.145$, it is interpreted that the diffusion mechanism among the 1-st order neighbors has been reduced by more than one half in strength due to the I_1 intervention. This phenomena is consistent with the fact that the I_1 intervention reduced the difference of the carbon monoxide levels among these six locations, i.e. I_1 forced the CO levels of these six locations to

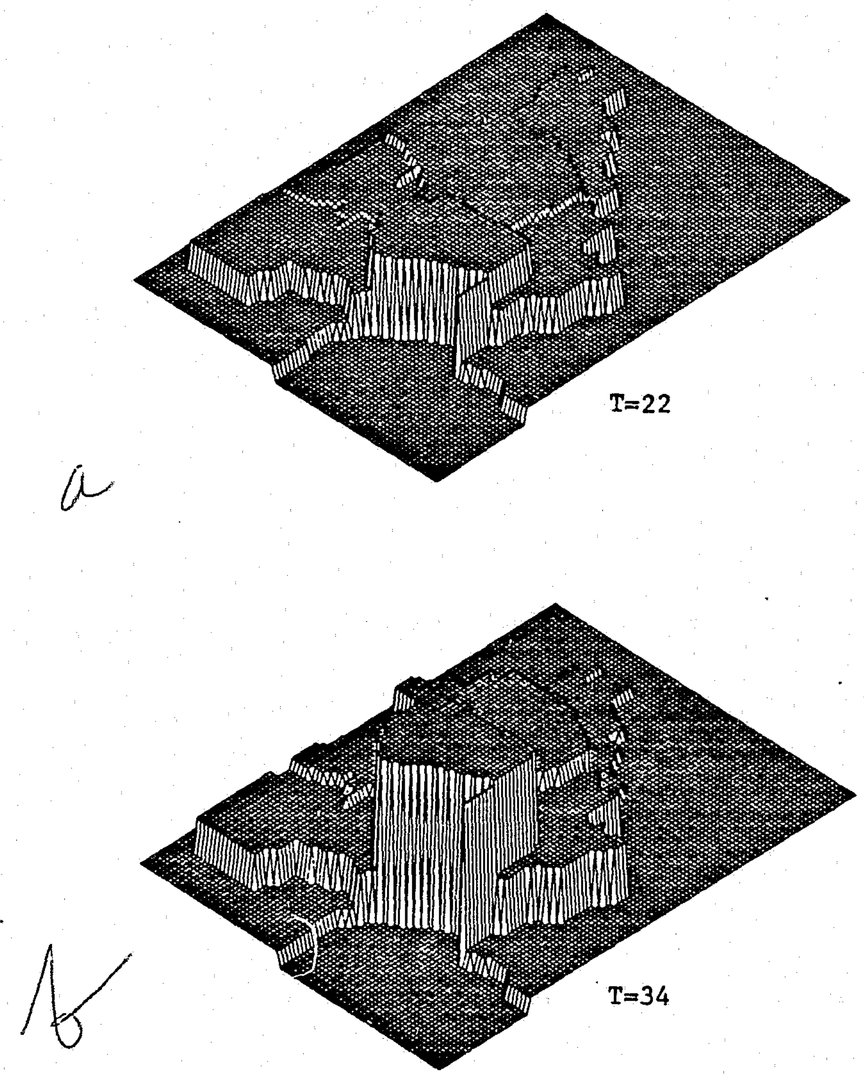


Figure 3-16. The Mean Shift $\hat{\delta}(t)$.

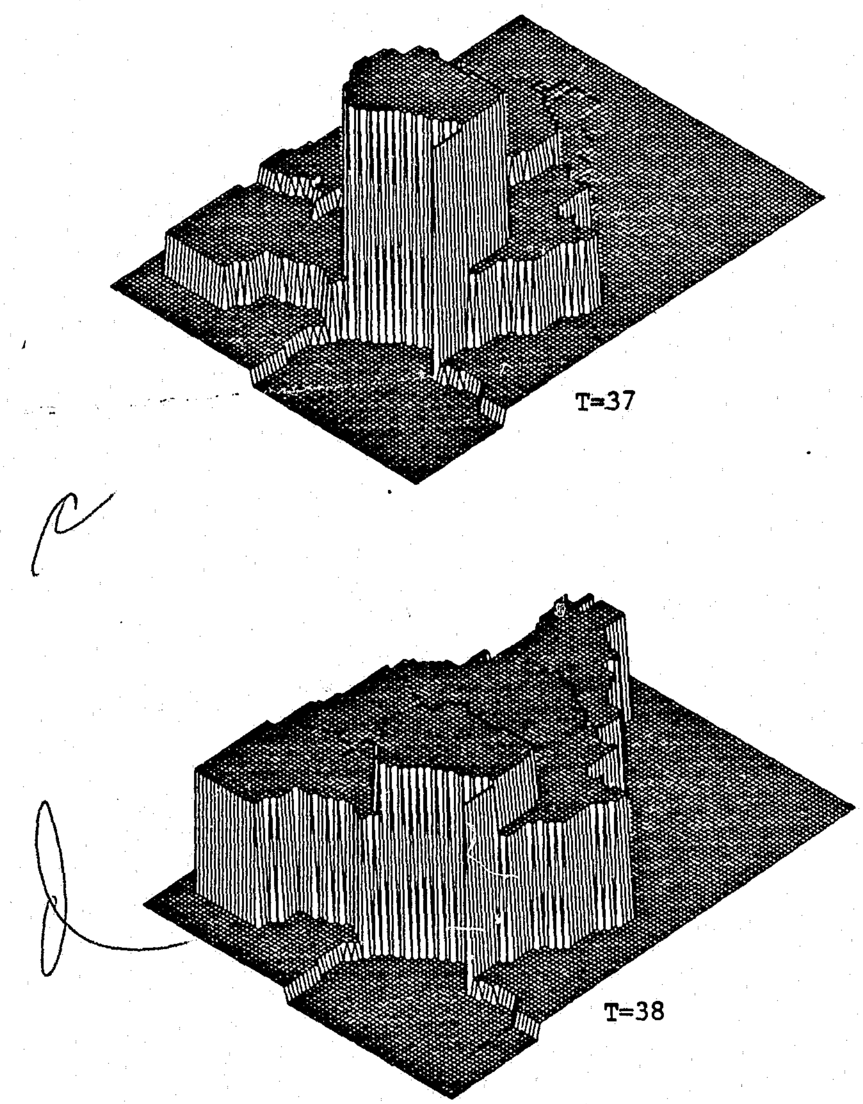
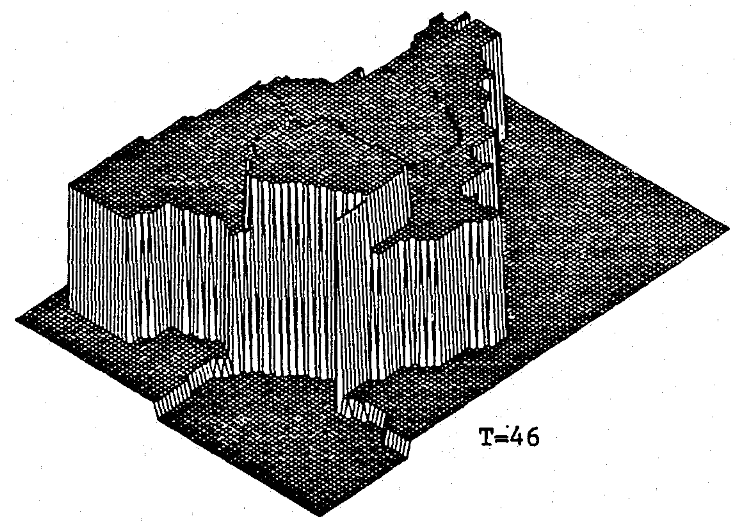


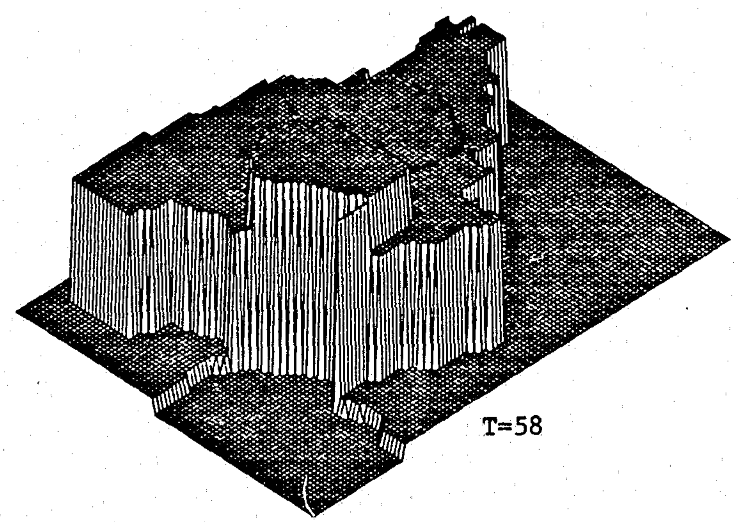
Figure 3-16. (Cont'd)

e



T=46

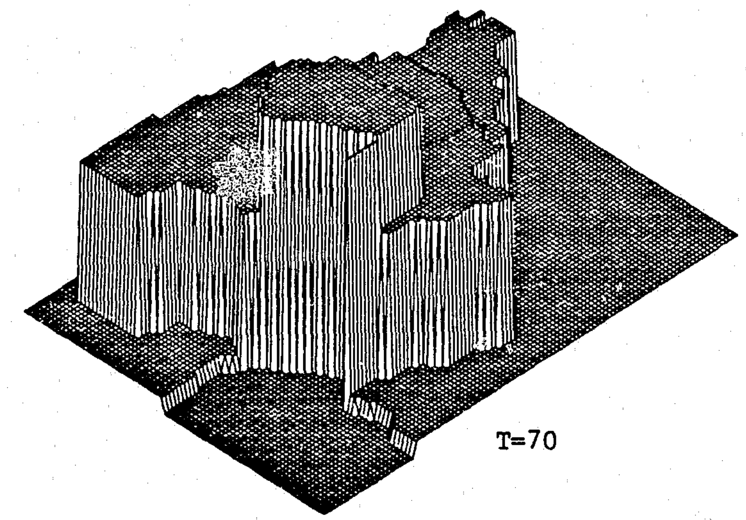
f



T=58

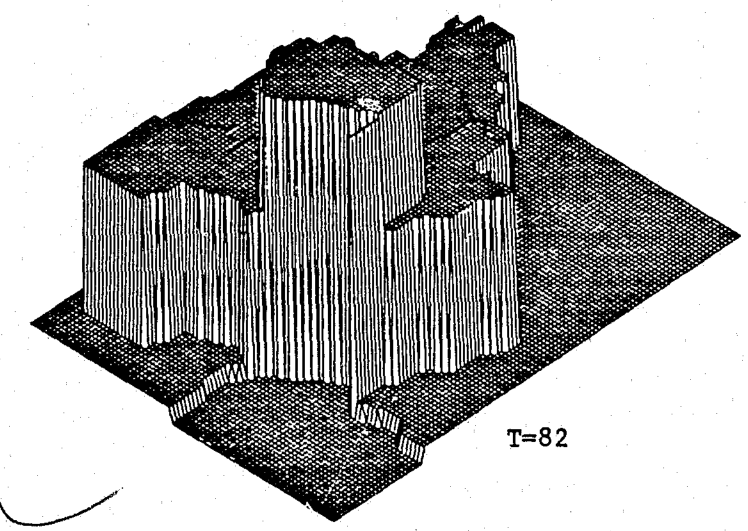
Figure 3-16. (Cont'd)

g



T=70

h



T=82

Figure 3-16. (Cont'd)

approach the same equilibrium level under the effectiveness of the I_2 legislation, so the driving force of the diffusion mechanism was weakened, this resulted in the reduced diffusion rate of the noise process and no diffusion phenomena of the intervention effect, i.e. $\underset{\sim}{\xi}_t^{(1)}$ and $\underset{\sim}{\xi}_t^{(2)}$.

All the models, i.e. the pre-I_1 model, post-I_1 model and noise model in the model equation (3-76) are either (STAR) model or (STARMA) model. Recall that it has besen discussed in section 3.2.3.1. that the inputs, i.e. $\underset{\sim}{\varepsilon}_t$, of the (STAR) model or the (STARMA) model of non-zero spatial order vector will diffuse through space and the influenced regions will be the whole connected regions. The diffused process of the (STAR) model and the (STARMA) model are of the AR type and the diffused particles will not die out immediately in the very next period like those of the MA type diffusion process. This long lasting property of the diffused particles of the AR diffusion type is quite matched to the fact that the carbon monoxide is essentially inert and will last long in the air.

3.6.7. Comparison to A Univariate Intervention Analysis Approach

Box, Tiao and Hamming [1975] have analyzed the Los Angeles CO data at seven locations. these locations are: Downtown LA, Lennox, Long Beach, Burbank, Azusa, Pasadena and West LA. They built the univariate intivention model for each individual location. The model for each location was;

$$\underset{\sim}{Z}_t = w\xi_{1t} + \beta_1 \frac{\xi_{2t}}{1-B^{12}} + B_2 \frac{\xi_{et}}{1-B^{12}} + \frac{(1-\theta_1 B)(1-\theta_2 B^{12})}{(1-\theta B)(1-B^{12})} \underset{\sim}{\varepsilon}_t \quad (3\text{-}79)$$

where $\xi_{1t} = \begin{cases} 0 & t < \text{April, 1968} \\ 1 & t > \text{April, 1968} \end{cases}$

$\xi_{2t} = \begin{cases} 0 & t < \text{January, 1966} \\ 1 & t > \text{January, 1966} \end{cases}$

$\xi_{3t} = 1 - \xi_{2t}$

The estimates of the model parameters are listed in table 3-9.

Comparing the model form, we see that they are different in three major respects:

1. The univariate intervention model, equation (3-79), doesn't imply any geological information, i.e. this model doesn't have the capability to model the pollutant's diffusion through the neighbors (space). The space-time intervention model, equation (3-76), has the capability to explain the space correlated structures.
2. The space-time intervention model considers the transient period effect. $\underset{\sim}{\delta}^{(1)}$ in the pre-I_2 period, and the steady state effect, $\underset{\sim}{\delta}^{(2)}$ in the post-I_2 period. The univariate

intervention models consider only the steady state efffect, this is not true as was revealed in the dynamic component identification step.

3. The univariate models have different mean shift effects of $I_2$for each location. This means that different locations was operating at significantly different pollution levels at pre-I_2 has been assumed. As is seen from the data itself, it is seen that at pre-I_2 the pollution level at all the locations are about the same. The space-time intervention incorporates the non-linear transformation conditions and has only one mean-shift mean effects and a transformation scaled factor.

Comparing the modeling procedures, we see that the space-time interventional modeling procedures contain the dynamic components identification procedure, which is a necessary procedure to determine the intervention effect formulation, i.e. environment involved or non-environment involved. A mistaken model formulation will result in a misleading model, from which the incorrect conclusions will be drawn. The univariate intervention modeling procedures do not contain the dynamic components identification procedure and do not have the capability to model the environment involved intervention process.

Comparing the analysis results of the univariate intervention models, equation (3-79), with the space-time intervention model, equation (3-76), we have the following:

1. Both models aggre that I_2, the change of calibration method, has significant negative effect on the measurements at all locations. The univariate models have different mean shift effects for each location, and the space-time intervention model has only an overall mean shift and a transformation scale factor for all the locations.

2. The univariate intervention models concluded that I_1 has the significant impact at Azusa, Downtown LA, Lennox, Long Beach and West LA. The space-time intervention model concluded that I_1 has the significant impact at Downtown LA, Lennox and Long Beach. Both models agree that the impacts of I_1 at Downtown LA, Lennox and Long Beach were significant and decreasing the pollution levels. But they don't agree in the impact of I_1 at Azusa. The space-time intervention model concluded that the impact of I_1 at Azusa was non-significant, where the univariate intervention model concluded that the impact of I_1 at Azusa was significant and positive, i.e. the pollution level was raised. The results of the space-time intervention model is then justifiable to be closer to what was happening, since thé the impact of I_1 is expected to reduced the pollution levels at all locations and the conclusion of the univariate intervention model disagree this expectation at Azusa.

Comparing the model parsimony, we see that the space-time intervention, equation (3-76), contains 25 model parameters, while the univariate intervention models, equation (3-79), needs 36 model parameters for 6 locations, so the space-time intervention model is more parsimonious.

Table 3-9. Estimates of Parameters in the Univariate Intervention Model.

Location	Measurement effect April 1968 $\hat{\omega}$	Trend before 1966 $\hat{\beta}_1$	Trend since 1966 $\hat{\beta}_2$	Trend Difference $\hat{\beta}_2-\hat{\beta}_1$	Noise Model Parameters $\hat{\phi}$	Noise Model Parameters $\hat{\theta}_1$	Noise Model Parameters $\hat{\theta}_2$	$\hat{\sigma}_\varepsilon$
Azusa	-6.50	-0.05	0.29*	0.34	0.84	0.42	0.97	0.89
	(0.70)	(0.13)	(0.15)	(0.23)	(0.07)	(0.12)	(0.04)	
Pasadena	-5.44	-0.29	-0.04	0.25	0.60	0.05	0.65	1.37
	(0.93)	(0.20)	(0.23)	(0.33)	(0.12)	(0.15	(0.07)	
Burbank	-5.72	0.40	-0.33	-0.73	0.78	0.22	0.79	1.37
	(1.10)	(0.25)	(0.26)	(0.40)	(0.08)	(0.12)	(0.05)	
LA County	-5.17	0.11	-0.43	-0.54	0.79	0.28	0.82	1.10
	(0.87)	(0.19)	(0.20)	(0.31)	(0.07)	(0.11)	(0.05)	
Downtown LA	-4.32	0.09	-0.28*	-0.37	0.71	0.20	0.83	0.97
	(0.73)	(0.12)	(0.16)	(0.23)	(0.10)	(0.13)	(0.05)	
Lennox	-5.08	0.51	-0.36*	-0.87	0.79	0.19	0.59	1.05
	(0.85)	(0.55)	(0.25)	(0.61)	(0.08)	(0.13)	(0.08)	
Long Beach	-5.29	0.40	-0.45*	-0.85	0.77	0.27	0.81	1.06
	(0.84)	(0.17)	(0.19)	(0.29)	(0.09)	(0.13)	(0.06)	

CHAPTER IV

NON-EQUAL DIFFUSION PREFERENCE MODELS

In previous chapters, we have considered STARIMA models in which, every location has an equally weighted influence from those locations that share it as a common neighbor of the same spatial order. Situations arise that this equal diffusion preference mechanism is not appropriate and a non-equal diffusion preference phenomenon is needed. For example, point pollution in the air and/or in the sea diffuse from one region to its neighbors with near neighbor regions exert stronger and quicker influence than the distant neighbor regions. Without the wind and/or marine currents, the pollution diffusion mechanism will exhibit equal preference for all directions. However, when there is wind and/or current, the pollution of one location will be effected most strongly by the downward regions and most weakly by the leeward regions. Thus the diffusion mechanism will not be isotropic. The nature extention of these STARIMA models is thus the extension to accomodate the modeling of non-equal diffusion preferences.

The interpretation of the diffusion in the weight matrix is discussed for the unscaled weight matrix as well as the scaled weight matrix in the first section. A discussion of the need for constructing the non-equally preferential weight matrices and the methods for their construction are then described. In Section 4.2, one-direction preferential space-time processes and two-direction preferential space-time processes are simulated to illustrate the diffusion processes of the non-equally preferential process. Model inadequacies due to the ignored non-equally preferential structure are studied in Section 4.3. In Section 4.4, two methods of model updating to account for detected nonisotropic behavior are proposed. The first method is based on the decomposition of the equally preferential weight matrix into the non-equally preferential weight matrices. The second method adds the potential non-equally preferential terms into the equally preferential models. The maximum likelihood estimation procedure that is based on the results of the linear model theory is briefly discussed in Section 4.5 for the STAR, STMA and STARIMA models. In Section 4.6, the Ambient Carbon Monoxide observations at Los Angeles during the pre-I_1, that have been modeled in Chapter III without isotropic characterization is used to illustrate the non-equal diffusion modeling methods developed in this chapter.

4.1 The Interpretation of the Diffusion Preference Weight Matrix

Here we are going to examine the physical meaning of the weights in the weight matrix. Let $W_{ij}^{(\ell)}$ be the (i,j) element of the ℓ^{th} order weight matrix $W^{(\ell)}$ and $W_{ij}^{(\ell)} \neq 0$ only when location i is an ℓ^{th} order neighbor of location j. $\phi_{k\ell}W_{ij}^{(\ell)}$ indicates the strength of influence that the observation at location j has on the k time period laged observation of its ℓ^{th} order neighbor of location i, i.e., $Z_j(t-k)$ on $Z_i(t)$. Similarly $\theta_{k\ell}W_{ij}^{(\ell)}$ indicates the influence from

$\varepsilon_j(t-k)$ to $Z_i(t)$. Since all elements of $W^{(\ell)}$ are premultiplied by $\phi_{k\ell}$ or $\theta_{k\ell}$ to express the strength of influence, then amplitude of $W_{ij}^{(\ell)}$ is a measure of how strong the state of location i depends on the state of location j as an ℓ^{th} order neighbor. The larger $W_{ij}^{(\ell)}$ is, the stronger the dependence. Let $\underset{\sim}{R}_i^{(\ell)}$ denote the i^{th} row of weight matrix $W^{(\ell)}$ and $\underset{\sim}{C}_j^{(\ell)}$ the j^{th} column. Then $\underset{\sim}{R}_i^{(\ell)}$ contains the information of the influence on location i from those locations that share location i as their common ℓ^{th} order neighbor and $\underset{\sim}{C}_j^{(\ell)}$ contains the information of the influence that location j has on all its ℓ^{th} order neighbor.

4.1.1 The Boundary Effect on the Scaled Weight Matrices

We have defined the unscaled weight matrix $W^{(\ell)}$ as

$$W_{ij}^{(\ell)} = \begin{cases} 1 & \text{if location i is an } \ell^{th} \text{ order neighbor of location j} \\ 0 & \text{otherwise .} \end{cases} \qquad (4\text{-}1)$$

In the unscaled weight matrix of ℓ^{th} spatial order, all the nonzero elements are equal to 1. This means that every location i is equally influenced by all the locations that share the location i as their common ℓ^{th} order neighbor, and every location j has equal influence on all its ℓ^{th} order neighbors.

In previous chapters, we have used the scaled weight matrix $W^{(\ell)}$ which is defined as

$$W_{ij}^{(\ell)} = \begin{cases} 1/n_i^{(\ell)} & \text{if location j has location i as its } \ell^{th} \text{ order neighbor} \\ 0 & \text{otherwise} \end{cases} \tag{4-2}$$

where $n_i^{(\ell)}$ is the number of ℓ^{th} order neighbors of location i. From this definition, every non-zero element in each row are equal. This means that the scaled weight matrix still keeps the property that all locations that share the location i as their common ℓ^{th} order neighbor have equal influence on the location i. However, not all the non-zero elements of the same column are equal, since the $n_i^{(\ell)}$ of those locations for boundary sites is smaller than that of those locations in the central area. This implies that every location will have larger (or equal) influence on those locations on (or near) the boundary. When this boundary is extended to infinity, all the $n_i^{(\ell)}$ will be equal and the scaled weight matrix will have exactly the same physical interpretation as that of the unscaled weight matrix.

Alternatively, if we define the scaled weight matrix as

$$W_{ij}^{(\ell)} = \begin{cases} 1/n_j^{(\ell)} & \text{if location j has location i as its } \ell^{th} \text{ order neighbor,} \\ 0 & \text{otherwise} \end{cases} \tag{4-3}$$

where $n_j^{(\ell)}$ is the number of locations that have the location i as their common ℓ^{th} order neighbors, then the scaled weight matrix is scaled in columns. In this case we have the property that every location i has

equal influence to each of its ℓ^{th} order neighbors, but, for any location which is a common ℓ^{th} order neighbor to some boundary locations and some non-boundary locations, the influence from those boundary locations will be stronger than that from the non-boundary locations. As an example, consider the 3 × 3 regular grid system;

1	4	7
2	5	8
3	6	9

The unscaled weight matrix is listed in Figure 4-1.

$$W^{(1)} = \begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1&2&3&4&5&6&7&8&9 \end{array} \\ \left[\begin{array}{ccccccccc} 0&1&0&1&0&0&0&0&0\\ 1&0&1&0&1&0&0&0&0\\ 0&1&0&0&0&1&0&0&0\\ 1&0&0&0&1&0&1&0&0\\ 0&1&0&1&0&1&0&1&0\\ 0&0&1&0&1&0&0&0&1\\ 0&0&0&1&0&0&0&1&0\\ 0&0&0&0&1&0&1&0&1\\ 0&0&0&0&0&1&0&1&0 \end{array}\right] \end{array}$$

Figure 4-1. The Unscaled Weight Matrix $W^{(1)}$ of the 3 × 3 Regular Grid System are listed in Tables 4-6 (e) and (f) for simulation 3 and simulation 4, respectively. These sample space-time autocorrelations shown in Tables 4-6 (e) and (f) clearly indicate the unexhausted structures.

In previous discussion for STAR models, it has been shown that if the ${}^N\Phi_K\Phi_K^{-1}$ matrix in Equation 4-24 is close to identity matrix I, then the residuals will behave closely to the white noise. However, if the ${}^N\Phi_K\Phi_K^{-1}$ matrix are far away from the identity matrix I, then the residuals will show model inadequacy and repeatedly modeling the residuals from the previous mistaken equal preference space-time model can't exhaust the process structure even when enough observation for appropriate power are available. In these simulations, we have ${}^N\Phi_1\Phi_1^{-1}$ of the first two simulations listed in Figure 4-19(a) and ${}^N\Phi_1\Phi_1^{-1}$ of the last two simulations listed in Figure 4-19(b). The matrix in Figure 4-19(a) is very closed to identity matrix, because in the first two simulations, the strength of the equally preferential components, that is represented by ϕ_{10}, is relatively stronger than the strength of the non-equally preferential components, that is represented by ${}^N\phi_{11}$. The matrix in Figure 4-19(b) is far away from the identity matrix, because in the last two simulations, the strength of the non-equally preferential component is relatively stronger than the strength of the equally preferential components. Comparing the matrix shown in Figure 4-19(a) with that in Figure 4-19(b) and looking back to the statement ${}^N\Phi_K\Phi_K^{-1}$ matrix is related to the sample space-time autocorrelations of the estimated residuals, we see that the diagnostic checking passes the mistaken models of simulation 1 and simulation 2, while the inadequacies for the mistaken models of

Table 4-6(a). The Sample Space-Time Autocorrelations of the Residuals from the Mistaken Model for Simulation 1.

The Sample S-T Autocorreltions/
The Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	0	1
1	0.03	0.04	0.97	1.24
2	-0.00	-0.02	-0.13	-0.74
3	-0.01	-0.03	-0.28	-0.82
4	-0.02	0.04	-0.69	1.29
5	-0.00	-0.07	-0.23	-1.94
6	-0.01	0.01	-0.41	0.51
7	-0.02	-0.03	-0.62	-1.04
8	-0.02	-0.02	-0.73	-0.70
9	-0.10	0.02	-2.78	0.56
10	0.04	-0.00	1.11	-0.08
11	0.00	0.00	0.14	0.11
12	-0.01	0.04	-0.30	1.10
13	-0.02	0.00	-0.63	0.22
14	-0.02	0.01	-0.73	0.46
15	-0.02	-0.03	-0.56	-0.84

Table 4-6(b). The Sample Space-Time Autocorrelations of the Residuals from the Mistaken Model for Simulation 2.

The Sample S-T Autocorrelations/
The Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	0	1
1	0.02	0.00	1.21	0.43
2	-0.01	-0.00	-0.82	-0.34
3	-0.03	-0.00	-1.70	-0.26
4	-0.03	0.00	-1.39	0.43
5	0.00	-0.01	0.19	-0.58
6	-0.00	0.01	-0.34	0.72
7	0.02	-0.02	0.92	-1.22
8	-0.01	-0.04	-0.80	-2.02
9	-0.03	0.00	-1.55	0.37
10	0.02	-0.00	0.95	-0.17
11	0.00	0.02	0.20	0.89
12	-0.00	0.00	-0.25	0.24
13	-0.04	-0.01	-1.85	-0.61
14	-0.02	-0.00	-1.00	-0.22
15	-0.02	0.00	-0.96	0.25

Table 4-6(c). The Sample Space-Time Autocorrelations of the Residuals from the Mistaken Model for Simulation 3.

The Sample S-T Autocorrelations/
The Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	0	1
1	0.12	0.14	3.47	4.08
2	-0.22	-0.01	-6.22	-0.38
3	-0.15	-0.20	-4.26	-5.66
4	-0.11	-0.05	-3.04	-1.45
5	0.02	-0.04	0.73	-1.33
6	0.11	0.02	3.13	0.79
7	0.05	-0.04	1.44	-1.28
8	-0.11	-0.03	-3.08	-1.00
9	-0.16	0.02	-4.32	0.54
10	0.07	0.00	2.02	0.24
11	0.09	0.04	2.37	1.13
12	0.00	0.06	0.12	1.66
13	-0.00	-0.00	-0.05	-0.04
14	-0.01	-0.03	-0.49	-0.92
15	-0.08	-0.03	-2.01	-0.87

Table 4-6(d). The Sample Space-Time Autocorrelations of the Residuals from the Mistaken Model for Simulation 4.

The Sample S-T Autocorrelations/
The Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	0	1
1	0.11	0.13	4.95	5.81
2	-0.21	0.00	-9.53	0.36
3	-0.14	-0.14	-6.47	-6.60
4	-0.06	-0.07	-2.84	-3.45
5	0.02	-0.01	1.06	-0.78
6	0.04	0.01	1.99	0.86
7	0.03	-0.05	1.56	-2.20
8	-0.07	-0.07	-3.11	-3.17
9	-0.09	0.00	-3.97	0.05
10	0.05	0.02	2.18	1.20
11	0.08	0.06	3.80	2.97
12	0.03	0.04	1.36	1.92
13	-0.03	-0.02	-1.36	-0.92
14	-0.04	-0.02	-1.85	-1.23
15	-0.04	0.00	-1.75	0.34

Table 4-6(e). The Sample Space-Time Autocorrelations of the Residuals from the STMA (1_1) Model of the Residuals from the Mistaken Model for Simulation 3.

The Sample S-T Autocorrelations/
The Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	0	1
1	-0.01	0.01	-0.38	0.36
2	-0.18	0.03	-5.18	0.83
3	-0.11	-0.19	-3.08	-5.40
4	-0.06	0.00	-1.82	0.01
5	0.01	-0.04	0.45	-1.08
6	0.11	0.05	3.10	1.44
7	0.04	-0.05	1.32	-1.40
8	-0.09	-0.02	-2.51	-0.62
9	-0.16	0.01	-4.28	0.36
10	0.08	0.00	2.32	0.03
11	0.07	0.02	1.88	0.76
12	-0.01	0.05	-0.34	1.53
13	-0.00	-0.00	-0.17	-0.07
14	-0.00	-0.01	-0.00	-0.33
15	-0.07	-0.03	-1.96	-0.78

Table 4-6(f). The Sample Space-Time Autocorrelations of the Residuals from the STMA (1_1) Model of the Residuals from the Mistaken Model for Simulation 4.

The Sample S-T Autocorrelations/
The Standardized S-T Autocorrelations

Space Lag Time Lag	0	1	0	1
1	0.13	0.24	5.86	10.76
2	-0.19	0.01	-8.56	0.64
3	-0.14	-0.15	-6.61	-6.96
4	-0.08	-0.08	-3.78	-3.86
5	0.01	-0.03	0.58	-1.71
6	0.03	0.00	1.40	0.30
7	0.02	-0.06	1.11	-2.65
8	-0.07	-0.07	-3.29	-3.27
9	-0.09	0.00	-4.01	0.11
10	0.05	0.03	2.47	1.49
11	0.09	0.07	4.05	3.32
12	0.03	0.04	1.62	2.11
13	-0.02	-0.01	-1.18	-0.63
14	-0.04	-0.03	-1.80	-1.41
15	-0.03	0.01	-1.55	0.50

1.00	0.27	-0.07	0.02	0.00	0.00	0.00	-0.02	0.07	-0.27
-0.27	1.00	0.27	-0.07	0.02	0.00	0.00	0.00	-0.02	0.07
0.07	-0.27	1.00	0.27	-0.07	0.02	-0.00	0.00	0.00	-0.02
-0.02	0.07	-0.27	1.00	0.27	-0.07	0.02	-0.00	0.00	0.00
0.00	-0.02	0.07	-0.27	1.00	0.27	-0.07	0.02	-0.00	0.00
0.00	0.00	-0.02	0.07	-0.27	1.00	0.27	-0.07	0.02	-0.00
0.00	0.00	0.00	-0.02	0.07	-0.27	1.00	0.27	-0.07	0.02
0.02	-0.00	0.00	0.00	-0.02	0.07	-0.27	1.00	0.27	-0.07
-0.07	0.02	-0.00	0.00	0.00	-0.02	0.07	-0.27	1.00	0.27
0.27	-0.07	0.02	-0.00	0.00	0.00	-0.02	0.07	-0.27	1.00

Figure 4-19(a). The ${}^{N}\phi_1\phi_1^{-1}$ Matrix for the Simulations with

$\phi_{10} = 0.6$, ${}^{N}\phi_{11} = 0.3$.

0.05	0.70	1.55	-1.18	0.01	0.50	-0.29	-0.05	0.15	-0.15
-0.45	-0.10	1.07	1.43	-1.28	0.12	0.48	-0.32	-0.02	0.35
0.47	-0.16	-0.38	0.86	1.67	-1.32	0.04	0.54	-0.33	-0.11
0.42	0.44	-0.67	0.19	0.75	1.49	-1.17	0.03	0.48	-0.68
-1.66	-0.21	1.77	-0.93	-0.23	1.10	1.47	-1.31	0.12	1.15
1.23	-0.75	-0.81	0.78	-0.10	-0.28	0.77	1.68	-1.29	0.04
2.22	1.48	-3.06	1.12	0.67	-0.86	0.20	0.82	1.44	-2.75
-5.84	-0.23	5.97	-3.31	-0.66	1.80	-0.74	-0.35	1.09	3.56
2.60	-3.14	-0.81	1.81	-0.69	-0.38	0.51	-0.13	-0.15	1.66
1.45	1.14	-2.11	0.71	0.49	-0.59	0.12	0.18	-0.16	0.03

Figure 4-19(b). The ${}^{N}\phi_1\phi_1^{-1}$ Matrix for the Simulations with

$\phi_{10} = 0.2$, ${}^{N}\phi_{11} = 0.7$.

simulation 3 and simulation 4 are detected.

4.4 Updating the Non-Equally Preferential Diffusion STARMA Process

In this section, two topics of modeling the non-equally preferential STARMA processes are discussed. First, the topic of testing the significance of the isotropic property of the space-time process is studied. This is followed by the topic of testing the significance of the non-equally preferential model parameters. In this section, we will apply the M.L estimation procedure to get the test statistics. Also, it is assumed that the equal preference STARMA model has been built already.

4.4.1 Testing the Significance of the Isotropic Property

In this section, the discussion on testing the isotropic property is started using a simple system, the STARMA $(1_1,0,1_1)$ process of a circular line system, and then it is generalized to the arbitrary system.

Assume that we have a circular line system with the preferential direction toward the right as in Figure 4-18, and assume that the system follows the STARMA $(1_1,0,1_1)$ process. For simplicity it is assumed that the preference is absolute, and the 1st order weight matrix ${}^{N}W^{(1)}$ is;

$$
{}^{N}W^{(1)} = \begin{array}{c} 1 \\ 2 \\ \vdots \\ \\ LN \end{array} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}
$$

Then the true model for this sytem is

$$
\underset{\sim}{Z}_t = \sum_{\ell=0}^{1} {}^{N}\phi_{1\ell} W^{(\ell)} \underset{\sim}{Z}_{t-1} - \sum_{\ell=0}^{1} {}^{N}\theta_{1\ell} {}^{N}W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t . \qquad (4\text{-}28)
$$

If this process is mistaken as an equal preference process, then the mistaken weight matrix $W^{(1)}$ will be assigned as follows;

$$W^{(1)} = \begin{array}{c} 1 \\ 2 \\ \\ \\ LN \end{array} \begin{bmatrix} 0. & 0.5 & 0. & & 0. & 0. & 0.5 \\ 0.5 & 0. & 0.5 & \cdots & 0. & 0. & 0. \\ \vdots & & & & & & \\ 0. & 0. & \cdots & & 0.5 & 0. & 0.5 \\ 0.5 & 0. & \cdots & & 0. & 0.5 & 0. \end{bmatrix}$$

and the model that is built in terms of $W^{(1)}$ is,

$$\underset{\sim}{Z}_t = \sum_{\ell=0}^{1} \phi_{1\ell} W^{(\ell)} \underset{\sim}{Z}_{t-1} - \sum_{\ell=0}^{1} \theta_{1\ell} W^{(\ell)} \underset{\sim}{A}_{t-1} + \underset{\sim}{A}_t . \qquad (4\text{-}29)$$

Here ϕ_{11}, θ_{11} and ${}^{N}\phi_{11}$, ${}^{N}\theta_{11}$ will have the relations,

$$0 \le 0.5\phi_{11} \le {}^{N}\phi_{11}, \quad 0 \le 0.5\theta_{11} \le {}^{N}\theta_{11},$$

and the residuals of model 4-29 will have the tendency to have dependence on the residuals of the right-hand-side neighbors and on the residuals of the left-hand-side neighbors. Since model 4-29 doesn't exhaust the dependence on the right-hand-side neighbors, while it overdraws the dependence on the left-hand-side neighbor, these two dependencies will be of opposite sign.

To test the significance of the non-equally preferential tendency in both directions, we reconstruct the model, Equation (4-29), in terms of ${}^{L}W(1)$ and ${}^{R}W(1)$ that are given below to obtain model,

Equation (4-30).

$${}^{L}W(1) = \begin{array}{c} 1 \\ 2 \\ 3 \\ \\ \\ LN \end{array} \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0.5 \\ 0.5 & 0 & & 0 & 0 & 0 \\ 0 & 0.5 & & 0 & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & & 0.5 & 0 & 0 \\ 0 & 0 & & 0 & 0.5 & 0 \end{bmatrix}$$

$${}^{R}W(1) = \begin{array}{c} 1 \\ 2 \\ 3 \\ \\ LN \end{array} \begin{bmatrix} 0 & 0.5 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0.5 & \cdots & & & \\ \vdots & & & & & & \\ 0 & 0 & 0 & & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & & 0 & 0 & 0 \end{bmatrix}$$

Note that the only non-zero elements in ${}^{L}W(1)$ (or ${}^{R}W(1)$) are the elements that correspond to the 1st order left-hand-side neighbors (or right-hand-side neighbors). Note also that ${}^{L}W(1) + {}^{R}W(1) = W^{(1)}$, and ${}^{R}W(1) = {}^{N}W(1)$.

$$\begin{aligned} \underset{\sim}{Z}_t = {} & \phi_{10} Z_{t-1} + {}^{R}\phi_{11}\, {}^{R}W(1) \underset{\sim}{Z}_{t-1} + {}^{L}\phi_{11}\, {}^{L}W(1) \underset{\sim}{Z}_{t-1} \\ & - \theta_{10} \underset{\sim}{R}_{t-1} - {}^{R}\theta_{11}\, {}^{R}W(1) \underset{\sim}{R}_{t-1} - {}^{L}\theta_{11}\, {}^{L}W(1) \underset{\sim}{R}_{t-1} + \underset{\sim}{R}_t \end{aligned} \qquad (4\text{-}30)$$

where R_t is the residual.

M.L estimation will give the estimates of parameters in Equation (4-30) as

$$^{R}\phi_{11} = 2^{N}\phi_{11}, \quad {}^{L}\phi_{11} = 0$$

$$^{R}\theta_{11} = 2^{N}\theta_{11}, \quad {}^{L}\theta_{11} = 0$$

which clearly indicates that this one-direction circular line system is of non-equally directional preference.

In general, when we have built an equally preferential space-time model,

$$\underset{\sim}{Z}_t = \sum_{K=1}^{p}\sum_{\ell=0}^{\lambda_K} \phi_{K\ell}W^{(\ell)}\underset{\sim}{Z}_{t-K} - \sum_{K=1}^{q}\sum_{\ell=0}^{m_K} \theta_{K\ell}W^{(\ell)}\underset{\sim}{\varepsilon}_{t-K} + \underset{\sim}{\varepsilon}_t, \tag{4-31}$$

and it is suspected that the true process may be highly preferential along some given direction, then we construct the model based on the principle of overfitting,

$$\underset{\sim}{Z}_t = \sum_{K=1}^{p} \phi_{K0}\underset{\sim}{Z}_{t-K} - \sum_{K=1}^{q} \theta_{K0}\underset{\sim}{R}_{t-K} + \underset{\sim}{R}_t$$
$$+ \sum_{K=1}^{p}\sum_{\ell=1}^{\lambda_K} {}^{L}\phi_{K\ell}{}^{L}W^{(\ell)}\underset{\sim}{Z}_{t-K} + \sum_{K=1}^{p}\sum_{\ell=1}^{\lambda_K} {}^{R}\phi_{K\ell}{}^{R}W^{(\ell)}\underset{\sim}{Z}_{t-K}$$

$$- \sum_{K=1}^{q}\sum_{\ell=1}^{m_K} {}^{L}\theta_{K\ell}{}^{L}W^{(\ell)}\underset{\sim}{R}_{t-K} + \sum_{K=1}^{p}\sum_{\ell=1}^{m_K} {}^{R}\theta_{K\ell}{}^{R}W^{(\ell)}\underset{\sim}{R}_{t-K} \tag{4-32}$$

where ${}^{L}W^{(\ell)}$ and ${}^{R}W^{(\ell)}$, $\ell=1,2,\ldots$ are constructed according to the approaches that were described in Section 4.1.3. Then the M.L. estimation is performed to give ${}^{L}\hat{\phi}$, ${}^{L}\hat{\theta}$, ${}^{R}\hat{\phi}$, ${}^{R}\hat{\theta}$, and is followed by hypothesis testing that tests the significance of the non-equally preferential structure;

$$H_0: \ {}^{L}\underset{\sim}{\beta} = {}^{R}\underset{\sim}{\beta}$$

vs.

$$H_1: \ {}^{L}\underset{\sim}{\beta} \neq {}^{R}\underset{\sim}{\beta}$$

where ${}^{L}\underset{\sim}{\beta} = ({}^{L}\underset{\sim}{\phi}^t, {}^{L}\underset{\sim}{\theta}^t)^t$ and ${}^{R}\underset{\sim}{\beta} = ({}^{R}\underset{\sim}{\phi}^t, {}^{R}\underset{\sim}{\theta}^t)^t$.

If the H_0 hypothesis is not rejected, we will keep the equally preferential space-time model, Equation (4-31). If the H_0 hypothesis is rejected, then the non-equally preferential space-time model, Equation (4-32), will be considered to be appropriate.

To test the hypothesis H_0: ${}^{L}\underset{\sim}{\beta} = {}^{R}\underset{\sim}{\beta}$, we compute the statistics suggested in Wilks [1938],

$$W_{\underset{\sim}{\beta}} = LN \cdot N[\ell n(\hat{\sigma}_\varepsilon^2) - \ell n(\hat{\sigma}_R^2)] \tag{4-33}$$

The null hypothesis is rejected when $W_{\underset{\sim}{\beta}} > \chi^2_{\alpha,n}$ or accepted when $W_{\underset{\sim}{\beta}} \leq \chi^2_{\alpha,n}$, where n is the total number of parameters in ${}^{L}\underset{\sim}{\beta}$ or ${}^{R}\underset{\sim}{\beta}$. If the hypothesis test shows that the non-equally preferential diffusion process is significant, then based on this conclusion, we may update the weight matrix by setting,

$$\phi_W(\ell) = {}^{L}\hat{\phi}_{K\ell}\,{}^{L}W(\ell) + {}^{R}\phi_{K\ell}\,{}^{R}W(\ell),$$

$$\theta_W(\ell) = {}^{L}\hat{\theta}_{K\ell}\,{}^{L}W(\ell) + {}^{R}\hat{\theta}_{K\ell}\,{}^{R}W(\ell) \qquad (4\text{-}34)$$

and scale them to have $\sum_{j=1}^{LN} \theta_{W_{ij}}(\ell) = 1$ and $\sum_{j=1}^{LN} \phi_{W_{ij}}(\ell) = 1$, i=1,2,...,LN. Then the model in Equation (4-32) can be updated in terms of $\phi_W(\ell)$ and $\theta_W(\ell)$ as follows,

$$\begin{aligned}\underset{\sim}{Z}_t = {} & \sum_{K=1}^{p} \phi_{K0}\underset{\sim}{Z}_{t-K} - \sum_{K=1}^{q} \theta_{K0}\underset{\sim}{R}_t + \underset{\sim}{R}_t \\ & + \sum_{K=1}^{p}\sum_{\ell=1}^{\lambda_K} {}^{N}\phi_{K\ell}\,\phi_W(\ell)\underset{\sim}{Z}_{t-K} - \sum_{K=1}^{q}\sum_{\ell=1}^{m_K} {}^{N}\theta_{K\ell}\,\theta_W(\ell)\underset{\sim}{R}_{t-K}. \qquad (4\text{-}35)\end{aligned}$$

4.4.2 Testing the Significance of the Non-Equal Preference Structures

In this section the proper procedure of adding in the non-equal diffusion preference components to update the equal diffusion preference STARIMA models is discussed. Assume that we have the candiate equally preferential STARIMA model built as follows

$$\underset{\sim}{Z}_t = \sum_{k=1}^{p}\sum_{\ell=0}^{\lambda_k} \phi_{k\ell}W^{(\ell)}\underset{\sim}{Z}_{t-k} - \sum_{k=1}^{q}\sum_{\ell=0}^{m_k} \theta_{k\ell}W^{(\ell)}\underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t \qquad (4\text{-}36)$$

where

$$E(\underset{\sim}{\varepsilon}_t\underset{\sim}{\varepsilon}'_{t+k}) = \begin{cases} G, & k=0 \\ 0, & \text{otherwise} \end{cases}$$

Also, assume that we have the weight matrices, ${}^{N}W(1), {}^{N}W(2), \ldots, {}^{N}W^{(n_0)}$, of which the corresponding non-equal diffusion preference mechanism is probably significant to the data generating process. We need a procedure that allows us to select the significant ones to add in the model and discard the non-significant ones. Two such procedures are discussed here. One is by fitting the whole model with all components that probably are significant, another is by the overfitting technique, Draper and Smith [1966].

First, we will discuss the procedure of fitting the whole model with all possible components. Without loss of generality, the whole model can be assumed to be of the form;

$$\begin{aligned}\underset{\sim}{Z}_t = {} & \sum_{k=1}^{p}\sum_{\ell=0}^{\lambda_k} \phi_{k\ell}W^{(\ell)}\underset{\sim}{Z}_{t-k} + \sum_{k=1}^{p}\sum_{\ell=0}^{\xi_k} {}^{N}\phi_{k\ell}\,{}^{N}W^{(\ell)}\underset{\sim}{Z}_{t-k} - \sum_{k=1}^{p}\sum_{\ell=0}^{m_k} \\ & \theta_{k\ell}W^{(\ell)}\varepsilon_{t-k} + \sum_{k=1}^{p}\sum_{\ell=0}^{\zeta_k} {}^{N}\theta_{k\ell}\,{}^{N}W^{(\ell)}\underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t \qquad (4\text{-}37)\end{aligned}$$

where

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t+k}') = \begin{cases} G, & k=0 \\ 0, & \text{otherwise} \end{cases}$$

An estimation procedure has been applied to get the estimates $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\theta}}$ and $S(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})$ for model, Equation (4-36). Although we add in the non-equal diffusion preference terms, model, Equation (4-37), can still be estimated by the same procedure to get the results $\hat{\boldsymbol{\phi}}$, ${}^N\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\theta}}$, ${}^N\hat{\boldsymbol{\theta}}$ and $S(\hat{\boldsymbol{\phi}},{}^N\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}},{}^N\hat{\boldsymbol{\theta}})$. Also note that the model, Equation (4-37), is a more generalized model in Equation (4-32) presented previously.

According to Whittle [1953], we constructed the estimated variance-covariance matrix of the estimated parameters.

$$V(\hat{\boldsymbol{\beta}}) \simeq \left[\frac{\partial^2 S(\boldsymbol{\beta})}{\partial\beta_i \partial\beta_j}\right]^{-1}_{\hat{\boldsymbol{\beta}}}$$

where

$$\boldsymbol{\beta} = (\boldsymbol{\phi}', {}^N\boldsymbol{\phi}', \boldsymbol{\theta}', {}^N\boldsymbol{\theta}')'$$

and we tested the hypothesis H_0: $H\boldsymbol{\beta} = \mathbf{0}$ by computing

$$W_f = (H\hat{\boldsymbol{\beta}})' (V(\hat{\boldsymbol{\beta}}))^{-1}(H\hat{\boldsymbol{\beta}}).$$

H_0 will be rejected if $W_f < kF(k, LN\cdot T-k_0, 1-\alpha)$ where k is the rank of H matrix and k_0 is the number of parameters included in model, Equation (4-36), i.e.,

$$k_0 = \sum_{k=1}^{p} (\lambda_k + \xi_k + 1) + \sum_{k=1}^{q} (m_k + \zeta_k + 1).$$

H is the null matrix with the diagonal elements that correspond to the tested parameters replaced by 1. So, k is also the number of parameters that will be tested simultaneously.

H_0 is accepted, then all parameters that are subjected to this test will be non-significant and the corresponding terms will be dropped from the model. When H_0 is rejected, some of the parameters that are subjected to this test are significant. (This does not necessarily imply that all of the parameters that are subjected to this test are significant). By this procedure, we can perform a series of tests using the same variance-covariance matrix to find out the significant components.

To apply the overfitting technique, the procedure described in Draper and Smith is employed. We can add any number of new components in and apply the extra sum of squares principle to test their significance. We keep only those components that pass the test.

After the significant non-equal diffusion preference, components are selected and added into the model, the updated model will be subjected to diagnostic checking again.

4.5 Non-Equal Preference Space-Time Model Estimation

In previous sections, the detection of non-equal preference structure, the test for the significance of isotropic property as well as the procedures for updating the non-equally preferential diffusion model from the already built equally preferential diffusion model have been studied in detail. In this section the M.L. estimation procedures that are appropriate for the non-equal preference model building are discussed. The application of the linear model to obtain the M.L. estimates when the covariance of the noise G is known is discussed in Section 4.5.1. Here the M.L. estimation procedures for STAR, STMA and STARMA are discussed separately. These discussions are followed by the M.L. estimation procedure when G is unknown in Section 4.5.2.

4.5.1 The Application of the Linear Model for the M.L. Estimation When the Covariance of the Noise is Known

In the model building procedure, we need to estimate the candidate model parameters. Here we will briefly review the estimation technique for estimating the STARIMA model parameters by applying the linear model transformation and searching through the parameter space. For computational convenience and for desirable properties of the estimates, we will limit ourselves here to the conditional M.L. estimation procedure only.

4.5.1.1 STAR Model Parameter Estimation. Consider the following model

$$\underset{\sim}{Z}_t = \sum_{k=1}^{p} \Phi_k \underset{\sim}{Z}_{t-k} + \underset{\sim}{\varepsilon}_t \tag{4-38}$$

where

$$\Phi_k = \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)}, \quad k=1,2,\ldots,p.$$

$\underset{\sim}{\varepsilon}_t$ are normally distributed with

$$E[\underset{\sim}{\varepsilon}_t] = \underset{\sim}{0}$$

and

$$E[\underset{\sim}{\varepsilon}_t \underset{\sim}{\varepsilon}'_{t+k}] = \begin{cases} G, & k=0 \\ 0, & \text{otherwise.} \end{cases}$$

Since the joint density of the (LN·T) × 1 random error vector $\underset{\sim}{\varepsilon}_t$ is

$$f(\underset{\sim}{\varepsilon}_t|\Phi,G) = (2\pi)^{-LN*T/2}|G|^{-T/2} \exp\{-1/2 \sum_{t=1}^{T} \underset{\sim}{\varepsilon}'_t G^{-1} \underset{\sim}{\varepsilon}_t\} \tag{4-39}$$

where Φ represents the p Φ_k matrices, i.e., $\Phi_1,\Phi_2,\ldots,\Phi_p$, LN is the location number and T is the total observation period. The transformation of the $\underset{\sim}{Z}$'s is

$$\underset{\sim}{\varepsilon}_t = \underset{\sim}{Z}_t - \sum_{k=1}^{p} \Phi_k \underset{\sim}{Z}_{t-k}$$

which has a unit Jacobin. Thus the density function of $\underset{\sim}{Z}_t$ conditional on Φ, G and $\underset{\sim}{Z}_0,\ldots,\underset{\sim}{Z}_{1-p}$ is

$$f(\underset{\sim}{Z}|\Phi,G,\underset{\sim}{Z}_0,\underset{\sim}{Z}_{-1},\ldots,\underset{\sim}{Z}_{1-p})$$

$$= (2\pi)^{-LN\cdot T/2}|G|^{-T/2} \exp\{-1/2 \sum_{t=1}^{T} [\underset{\sim}{Z}_t - \sum_{k=1}^{p} \Phi_k \underset{\sim}{Z}_{t-1}]' G^{-1}$$

$$[\underset{\sim}{Z}_t - \sum_{k=1}^{p} \Phi_k \underset{\sim}{Z}_{t-k}]\}$$

Remember Φ is a linear function of $\underset{\sim}{\phi}$ for STAR models. Letting

$$SS(\Phi,G) = \frac{1}{2} \sum_{t=1}^{T} [\underset{\sim}{Z}_t - \sum_{k=1}^{p} B(k)\underset{\sim}{Z}_{t-k}]' G^{-1} [\underset{\sim}{Z}_t - \sum_{k=1}^{p} B(k)\underset{\sim}{Z}_{t-k}]\},$$

We obtain the log conditional likelihood function as

$$\ell n\, L(\underset{\sim}{\phi},G|\underset{\sim}{Z},\underset{\sim}{Z}_0,\ldots,\underset{\sim}{Z}_{1-p}) = -T/2\ \ell n|G| - SS(\Phi,G). \qquad (4\text{-}40)$$

Usually, we set $\underset{\sim}{Z}_t$, $t \leq 0$ to its unconditional mean, $\underset{\sim}{Z}_t = \underset{\sim}{0}$, $t \leq 0$. From Equation (4-40) we see that to maximize the conditional likelihood function for given G is equivalent to minimize the SS(Φ) terms. We use SS(Φ) instead of SS(Φ,G) because G is given. We can get the

conditional maximum likelihood estimators for STAR models by applying the results of linear models,

$$\hat{\underset{\sim}{\phi}} = (X'X)^{-1}X\underset{\sim}{Y} \qquad (4\text{-}41)$$

where

$\underset{\sim}{Y} = [\underset{\sim}{Y}'(1), \underset{\sim}{Y}'(2),\ldots,Y'(t)]'$
with $\underset{\sim}{Y}(t) = C^{-1}\underset{\sim}{Z}_t$, $\quad t=1,2,\ldots,T.$

$X = [X_1,X_2,\ldots,X_p]$
X_k is the matrix defined as;

$X_k = [\underset{\sim}{X}_{k1},\underset{\sim}{X}_{k2},\ldots,\underset{\sim}{X}_{k\lambda_k}]$
$\underset{\sim}{X}_{k\ell} = C^{-1}[(W^{(\ell)}\underset{\sim}{Z}_{1-k})', (W^{(\ell)}\underset{\sim}{Z}_{2-k})',\ldots, (W^{(\ell)}\underset{\sim}{Z}(T-k))']$
$k=1,2,\ldots,p;\ \ell=0,1,2,\ldots,\lambda_k.$

and $CC' = G$, since G is positive definite matrix.

<u>4.5.1.2 STMA Model Parameter Estimation</u>. The STMA models are non-linear in the model parameters, and it is impossible to get a closed form expression for the estimators of model parameters. Consider the STMA model,

$$\underset{\sim}{Z}(t) = -\sum_{k=1}^{q} \Theta_k \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t \qquad (4\text{-}42)$$

where

$$\underset{\sim}{\varepsilon} = (\underset{\sim}{\varepsilon}_1', \underset{\sim}{\varepsilon}_2', \ldots, \underset{\sim}{\varepsilon}_T').$$

The transformation from $\underset{\sim}{\varepsilon}$'s to $\underset{\sim}{Z}$'s is

$$\underset{\sim}{\varepsilon}_t = \underset{\sim}{Z}_t + \sum_{k=1}^{q} Ⓗ_k \underset{\sim}{Z}_{t-k}, \qquad t=1,2,\ldots,T \tag{4-43}$$

which has unitary Jacobin. Thus, the conditional log likelihood function is

$$\ell n\, L(\underset{\sim}{\theta}, G | \underset{\sim}{Z}, \underset{\sim}{\varepsilon}_0, \underset{\sim}{\varepsilon}_{-1}, \ldots, \underset{\sim}{\varepsilon}_{1-q}) = (-LN \cdot T/2) \cdot \ell n\, 2\pi$$

$$-(LN \cdot T/2) \cdot \ell n |G| - SS(\underset{\sim}{\theta}, G). \tag{4-44}$$

Usually, we set $\underset{\sim}{\varepsilon}(t)$, $t \leq 0$ to their unconditional expections, $\underset{\sim}{\varepsilon}(t) = 0$, $t \leq 0$. From Equation (4-43), we see that for given G, to maximize the likelihood function is equivalent to minimize $SS(\underset{\sim}{\theta})$. We search through $\underset{\sim}{\theta}$ space to get $\hat{\underset{\sim}{\theta}}$ that minimizes $SS(\underset{\sim}{\theta})$. $SS(\underset{\sim}{\theta})$ can be derived easily by using the recursive Equation (4-44) with the $\underset{\sim}{\varepsilon}_t$, $t \leq 0$ set to zero.

4.5.1.3 Mixed STARIMA Parameter Estimation. The mixed STARIMA(p,d,q) models are linear in autoregressive parameters but non-linear in moving average parameters. So for a given G, $\underset{\sim}{\theta}$, we can have the autoregressive parameters estimated by applying results from linear models. Consider the STARIMA(p,0,q) model.

$$\underset{\sim}{Z}_t = \sum_{k=1}^{p} \Phi_k \underset{\sim}{Z}_{t-k} - \sum_{k=1}^{q} Ⓗ_k \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t$$

$$t=1,2,\ldots,T. \tag{4-45}$$

When $\underset{\sim}{\theta}$ is given, Equation (4-45) can be

$$\underset{\sim}{Y} = X\underset{\sim}{\phi} + \underset{\sim}{\varepsilon}^* \tag{4-46}$$

where

$$\underset{\sim}{\phi} = (\phi_{10}, \phi_{11}, \ldots, \phi_{1\lambda_1}, \ldots, \phi_{p0}, \phi_{p1}, \ldots, \phi_{p\lambda_p})$$

$$\underset{\sim}{Y} = A^{-1}\underset{\sim}{Z}$$

with $\underset{\sim}{Z} = (\underset{\sim}{Z}_1', \underset{\sim}{Z}_2', \ldots, \underset{\sim}{Z}_t')$

$$A = A_q C_T$$

where A_q is the (LN•T) by (LN•T) null matrix with the T diagonal (LN × LN) blocks replaced by $I_{LN \times LN}$ and the k^{th}, $1 \leq k \leq q$, subdiagonal (LN × LN) blocks replaced by $Ⓗ_k$,

C_T is the (LN•T) by (LN•T) null matrix with the T diagonal (LN × LN) blocks replaced by c, $cc' = G$, and

$X = C_T^{-1} B_T$ with $B_T = [B_1, B_2, \ldots, B_p]$, B_k is the (LN•T) by λ_k matrix defined as,

$$B_k = [\underset{\sim}{B}_{k1}, \underset{\sim}{B}_{k2}, \ldots, \underset{\sim}{B}_{k\lambda_k}]$$

$$\underset{\sim}{B}_{k\ell} = [(W^{(\ell)}\underset{\sim}{Z}_{1-k})', (W^{(\;)}\underset{\sim}{Z}_{2-k})', \ldots, (W^{(\ell)}\underset{\sim}{Z}_{T-k})']'$$

and

$$\underset{\sim}{\varepsilon}^* = C_T^{-1}\underset{\sim}{\varepsilon}$$

with $\underset{\sim}{\varepsilon} = (\underset{\sim}{\varepsilon}'(1), \underset{\sim}{\varepsilon}'(2), \ldots, \underset{\sim}{\varepsilon}'(T))$.

Since $\underset{\sim}{\varepsilon}^* \sim NID(0,I)$, the conditional maximum likelihood estimator for $\underset{\sim}{\phi}$ with $\underset{\sim}{\theta}$ are,

$$\hat{\underset{\sim}{\phi}}(\underset{\sim}{\theta}) = (X'X)^{-1}X'Y.$$

Once $\hat{\underset{\sim}{\phi}}(\underset{\sim}{\theta})$ is computed, we can recursively compute the $\underset{\sim}{\varepsilon}(t)$ and get $SS(\underset{\sim}{\theta})$. By searching through the $\underset{\sim}{\theta}$ space, we obtain the conditional maximum likelihood estimates for $(\underset{\sim}{\phi}', \underset{\sim}{\theta}')'$.

4.5.2 The M.L. Estimation Procedure When the Covariance of the Noise Is Unknown

The estimation procedure described above is for the situation that G is known. It is not an unusual case that G is unknown. In such situations, we will apply the two-stage estimation procedure. At the first stage, we assume that $G = I\sigma^2$ to get the estimates. Following the estimation, we check the $G = I\sigma^2$ assumption. If $G = I\sigma^2$ assumption is acceptable, then we have the appropriate estimates. If $G = I\sigma^2$ assumption is not acceptable, then we will use $(\hat{\underset{\sim}{\varepsilon}}\hat{\underset{\sim}{\varepsilon}}')$ as the true

covariance and apply the estimation procedure described above to get the second stage estimates and accept this second stage estimates as the appropriate conditional maximum likelihood estimates.

4.6 The Example of the Non-Equally Preferential Diffusion Process Modeling

The pre-I_1 observations of the Los Angeles ambient Carbon Monoxide data are used to serve as an example to illustrate the non-equally preferential space-time modeling procedure. In the paper by Tiao, Box and Hamming [1975] the effect of the wind speeds on the air quality law was found to be of little effect. Their explanation was that wind speed at 8 A.M. varies little between seasons. From the space-time intervention analysis in Chapter III, we found that the effects of the intervention Program I_1 are more similar in locatins located geographically in directions that are approximately perpendicular to the coast line. In the rest of this section, the directions that are perpendicular to the coast line will be denoted by the phrase "the | directions", and the directions that are parallel to the coast line will be denoted by the phrase "the ‖ directions".

4.6.1 Building the Model of Preference by the Strip Region Approach

The geographical locations of 1. Azusa, 2. Burbank, 3. Lennox, 4. Long Beach, 5. Downtown LA, 6. LA County are marked in the map of the Figure 3-10(a). The non-equally preferential weight matrices ${}^{|}W^{(1)}$, ${}^{\|}W^{(1)}$ has been obtained in the example of Section 4.1.3.1 by applying the strip region approach. ${}^{|}W^{(1)}$, ${}^{\|}W^{(1)}$ were listed in Table 4-2 (a) and (b), respectively.

Using this neighbor structures we built the following two STAR models:

1. The Equally Preferential Model

$$\underset{\sim}{Z}_t = \hat{\phi}_{10}\underset{\sim}{Z}_{t-1} + \hat{\phi}_{20}\underset{\sim}{Z}_{t-2} + \hat{\phi}_{11}({}^{\|}W^{(1)} + {}^{|}W^{(1)})\underset{\sim}{Z}_{t-1} + \underset{\sim}{\varepsilon}_t \qquad (4\text{-}47)$$

where

$\hat{\phi}_{10} = 0.6923$ (0.4229, 0.9617)

$\hat{\phi}_{20} = -0.1404$ (-0.3908, 0.1101)

$\hat{\phi}_{11} = 0.1371$ (-0.0891, 0.3634)

$\hat{\sigma}^2_\varepsilon = 1.9937$

The sample space-time autocorrelation functions and the standardized sample S-T autocorrelation functions are listed in the Table 4-7. The sum of squares of the ℓ^{th} order autocorrelations SS(ℓ) are

$$SS(2) = 0.034753$$

$$SS(1) = 0.06260$$

$$SS(0) = 0.062731$$

and

$$\chi^2_{(2)} = (6)(10)S(2) = 2.0852$$

$$\chi^2_{(1)} = (6)(10)SS(1) = 3.756,\ F_1 = \chi^2_{(1)}/\chi^2_{(2)} = 1.80127$$

$$\chi^2_{(0)} = (6)(10)SS(0) = 3.7639,\ F_0 = \chi^2_{(0)}/(\frac{\chi^2_{(1)}+\chi^2_{(2)}}{2}) = 1.2887.$$

Since $\chi^2_{(2)} < \chi^2_{0.1,4} = 7.78$, $F_1 < F_{0.1,4,4} = 4.11$ and $F_0 < F_{0.1,4,8} = 2.81$ diagnostic checking doesn't reveal any inadequacy of the equally preferential model.

2. The Non-Equally Preferential Model

$$\underset{\sim}{Z}_t = {}^{N}\hat{\phi}_{10}\underset{\sim}{Z}_{t-1} + {}^{N}\hat{\phi}_{20}\underset{\sim}{Z}_{t-2} + {}^{\|}\hat{\phi}_{11}{}^{\|}W^{(1)}\underset{\sim}{Z}_{t-1} + {}^{|}\hat{\phi}_{11}{}^{|}W^{(1)}\underset{\sim}{Z}_{t-1} + \underset{\sim}{R}_t \qquad (4\text{-}48)$$

where

		95% C.I.
${}^{N}\hat{\phi}_{10}$	= 0.7464	(0.4770, 1.0160)
${}^{N}\hat{\phi}_{20}$	= -0.1621	(-0.4125, 0.0884)
${}^{|}\hat{\phi}_{11}$	= -0.0043	(-0.2214, 0.2128)
${}^{\|}\hat{\phi}_{11}$	= 0.2555	(0.0292, 0.4817)
σ^2_R	= 1.9608804	

For this above model, the model parameter ${}^{|}\phi_{11}$ is nonsignificant at a significant level equal 0.1, so the following non-equally proferential model without the perpendicular term, ${}^{|}\phi_{11}{}^{|}W^{(1)}\underset{\sim}{Z}_{t-1}$, is built.

This resulted in,

$$\underset{\sim}{Z}_t = {}^N\hat{\phi}_{10}\underset{\sim}{Z}_{t-1} + {}^N\hat{\phi}_{20}\underset{\sim}{Z}_{t-2} + {}^{||}\hat{\phi}_{11}{}^{||}W^{(1)}\underset{\sim}{Z}_{t-1} + \underset{\sim}{R}_t, \qquad (4\text{-}49)$$

where

		95% C.I.
${}^N\hat{\phi}_{10}$	$= 0.7450$	(0.4756, 1.0140)
${}^N\hat{\phi}_{20}$	$= -0.1625$	(-0.4310, 0.0879)
${}^{\|\|}\hat{\phi}_{11}$	$= 0.2558$	(0.0295, 0.4820)
$\hat{\sigma}^2_R$	$= 1.9608895$	

The extra sum of squares due to the ${}^{|}\hat{\phi}_{11}{}^{|}W^{(1)}\underset{\sim}{Z}_{t-1}$ terms,

$$SS({}^{|}\phi_{11}|{}^N\phi_{10}, {}^N\phi_{20}, {}^{||}\phi_{11}) = 0.000546$$

$$F_1 = 0.000546/1.9608804 = 0.000273 < F_{0.1,1,50}$$

Therefore the non-equally preferential model, Equation (4-49), without the perpendicular term is appropriate.

The sample space-time autocorrelation functions and the standardized sample S-T autocorrelation functions for the residuals of the model in Equation (4-49) are listed in the Table 4-8. The sum of squares of the ℓ^{th} order autocorrelations SS(ℓ) are:

CONTINUED

3 OF 8

$$SS(2) = 0.03035$$

$$SS(1) = 0.053986$$

$$SS(0) = 0.05547$$

and

$$\chi^2_{(2)} = (6)(10)SS(3) = 1.8210$$

$$\chi^2_{(1)} = (6)(10)SS(2) = 3.239,\ F_1 = \chi^2_{(1)}/\chi^2_{(0)} = 1.7787$$

$$\chi^2_{(0)} = (6)(10)SS(1) = 3.328,\ F_0 = \chi^2_{(0)}/(\frac{\chi^2_{(1)}+\chi^2_{(2)}}{2}) = 1.2798$$

Since $\chi^2_{(2)} < \chi^2_{0.1,4} = 7.78$, $F_1 < F_{0.1,4,4} = 4.11$ and $F_0 < F_{0.1,4,8} = 2.81$, so the diagnostic checking passes the adequacy of the non-equally preferential model.

Comparing the model in Equation (4-49), which is of preference in the $\|$ directions, with the equally preferential model 4-47, we see that the model 4-49 and the model 4-47 have the same number of model parameters, and the model of preference in the $\|$ directions is closer to the true process and the structures among neighbors in the $\|$ directions are underestimated in the equally preferential model 4-47, while the structures among neighbors in the $|$ directions are overdrawn. Also, the extra sum of squares test confirms that the diffusion process in the $|$ directions is non-significant and the diffusion phenomenia appears only in the $\|$ directions that are parallel to the coast line.

Table 4-7. The Sample Space-Time Autocorrelation Functions of the Residuals of the Equally Preferential Model for the Pre-I_1 Observations of the LA CO Data

Space-Time Autocorrelations/Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	2	0	1	2
1	-0.04	0.02	-0.03	-0.29	0.16	-0.19
2	0.10	0.12	0.11	0.61	0.75	0.67
3	0.11	0.17	0.09	0.65	0.96	0.53
4	-0.18	-0.12	-0.10	-0.92	-0.60	-0.53

Table 4-8. The Sample Space-Time Autocorrelation Functions of the Residuals of the Non-Equally Preferential Model for the Pre-I_1 Observations of the LA CO Data

Space-Time Autocorrelations/Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	2	0	1	2
1	-0.04	0.00	-0.03	-0.29	0.06	-0.22
2	0.11	0.09	0.09	0.68	0.54	0.59
3	0.11	0.17	0.09	0.63	0.95	0.52
4	-0.16	-0.12	-0.10	-0.80	-0.60	-0.49

It should be noted that the equal preference model, Equation (4-47), was not found to have any statistical model inadequacies. This is due to the following two reasons:

1. ${}^{\|}\Phi_1 = {}^{N}\hat{\phi}_{10} I + {}^{\|}\hat{\phi}_{11} {}^{\|}W^{(1)}$ with ${}^{N}\hat{\phi}_{10} = 0.745$, ${}^{\|}\hat{\phi}_{11} = 0.256$

$$\Phi_1 = \hat{\phi}_{10} I + \hat{\phi}_{11} W^{(1)} \text{ with } \hat{\phi}_{10} = 0.692,\ \hat{\phi}_{11} = 0.137$$

The ${}^{\|}\Phi_1$, Φ_1 as well as ${}^{\|}\Phi_1\Phi_1^{-1}$ matrices were computed and listed in Figures 4-20 (a), (b) and (c), respectively. Here we see that the product ${}^{\|}\Phi_1\Phi_1^{-1}$ is pretty close to the identity matrix I. According to the discussions in Section 4.3, we see that the inadequacies due to the ignoring non-equal preference structure are hard to be detected from the sample space-time autocorrelation functions of the residuals when the product matrix is the identity matrix.

2. In Equation (4-48), we found that ${}^{\|}\hat{\phi}_{11}$ can't be significantly distinguished from ${}^{|}\hat{\phi}_{11}$, although ${}^{\|}\hat{\phi}_{11}$ can be significantly distinguished from 0 while ${}^{|}\hat{\phi}_{11}$ can't be distinguished from 0 significantly. This is because the available observations couldn't offer enough power to test the difference of ${}^{|}\phi_{11}$ and ${}^{\|}\phi_{11}$.

4.6.2 Building the Non-Equally Preferential Model by the Weight Matrix Decomposition

In the last section it has been found that the space-time model of preference in the $\|$ directions is more appropriate to describe the pre-I_1 observations of the ambient carbon monoxide in Los Angeles. The neighbor structure in the $\|$ directions and the neighbor structure in the $|$ directions were obtained by applying the strip regions approach, the union of these two neighbor structures is not equal to the neighbor structure described in the Chapter III. So the models that have been constructed in the last section are not comparable to the space-time model that has been built in the Chapter III.

In Section 4.1.3.2 the weight matrices that were employed in the Chapter III were decomposed according to the neighbor structures obtained by applying the angular region approach. In this section, the non-equally preferential model, that is comparable with the equally preferential model, is constructed. It should be noted that ${}^{|}W^{(\ell)} + {}^{||}W^{(\ell)} = W^{(\ell)}$ for ℓ=1,2,3, where $W^{(\ell)}$'s are constructed in Section 3.6.1.

These non-equally preferential weight matrices are then employed to construct the non-equally preferential space-time model. The following results were obtained

$$\underset{\sim}{Z}_t = {}^{N}\hat{\phi}_{10}\underset{\sim}{Z}_{t-1} + {}^{N}\hat{\phi}_{20}\underset{\sim}{Z}_{t-2} + {}^{|}\hat{\phi}_{11}\,{}^{|}W^{(1)}\underset{\sim}{Z}_{t-1} + {}^{||}\hat{\phi}_{11}\,{}^{||}W^{(1)}\underset{\sim}{Z}_{t-1} + \underset{\sim}{R}_t \qquad (4\text{-}50)$$

where

		95% C.I.
${}^{N}\hat{\phi}_{10}$	= 0.6344	(0.3497, 0.9177)
${}^{N}\hat{\phi}_{20}$	= -0.1488	(-0.4001, 0.0998)
${}^{\mid}\hat{\phi}_{11}$	= 0.5126	(-0.1425, 1.1650)
${}^{\mid\mid}\hat{\phi}_{11}$	= 0.2967	(0.0622, 0.5336)
$\hat{\sigma}_R^2$	= 1.92486	

The model parameter ${}^{|}\phi_{11}$ is significant at the α = 0.1186 level. Therefore model, Equation (4-50), of non-equal preference appears more appropriate in describing the process. Before accepting the model, the residuals are subjected to diagnostic checking to examine the adequacy of this model.

The sample space-time autocorrelation functions of the estimated residuals are listed in Table 4-9. The portmanteau lack of fit test for the space-time system is performed. The sum of squares of the ℓ^{th} order space-time autocorrelations SS(ℓ) are computed to be 1.

$$SS(2) = 0.030084$$
$$SS(1) = 0.061734$$
$$SS(0) = 0.062276,$$

and

$$\chi^2_{(2)} = (6)(10)SS(2) = 1.805,$$
$$\chi^2_{(1)} = (6)(10)SS(1) = 3.704,\quad F_1 = \chi^2_{(1)}/\chi^2_{(2)} = 2.052$$
$$\chi^2_{(0)} = (6)(10)SS(0) = 3.737,\quad F_0 = \chi^2_{(0)}/\left(\frac{\chi^2_{(1)}+\chi^2_{(2)}}{2}\right) = 1.357$$

Since

$$\chi^2_{(2)} < \chi^2_{0.1,4} = 7.78,$$

$$F_1 < F_{0.1,4,4} = 4.11,$$

$$F_0 < F_{0.1,4,8} = 2.81,$$

the residuals do not contain any additional structure. Thus the non-equal preference model which contains the $|$ directions from the $||$ directions terms is concluded to be adequate.

Both equal preference model in Section 3.6 and the non-equal preference model, Equation (4-50), do not show any model inadequacies in the sample space-time autocorrelation functions of the model residuals, this is due to the following two reasons:

1. The 95% CI of ${}^{|}\hat{\phi}_{11}$ and the 95% CI of ${}^{||}\hat{\phi}_{11}$ are not disjoint sets, and ${}^{|}\phi_{11}$ can't be said to be significantly distinguished from ${}^{||}\hat{\phi}_{11}$.

2. Let ${}^{N}\Phi_1 = {}^{N}\hat{\phi}_{10} I + {}^{|}\hat{\phi}_{11}\,{}^{|}W^{(1)} + {}^{||}\phi_{11}\,{}^{||}W^{(1)}$ with

$${}^{N}\hat{\phi}_{10} = 0.634, \quad {}^{|}\hat{\phi}_{11} = 0.513, \quad {}^{||}\hat{\phi}_{11} = 0.297$$

$$\Phi_1 = \hat{\phi}_{10} I + \hat{\phi}_{11} W^{(1)} \text{ with } \hat{\phi}_{10} = 0.692,\ \hat{\phi}_{11} = 0.137.$$

The ${}^{N}\Phi_1$, Φ_1 as well as ${}^{N}\Phi_1\Phi_1^{-1}$ matrices were computed and listed in Figures 4-21 (a), (b) and (c), respectively. It is seen that ${}^{N}\Phi_1\Phi_1^{-1}$ is pretty close to the identity matrix I, according to the discussions in Section 4.3, since ${}^{N}\Phi_1\Phi_1^{-1} \doteq I$, so the inadequacies due to the ignored non-equal preference structure are not able to be detected from the sample space-time autocorrelation functions.

$$
{}^{||}\Phi_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{bmatrix}
0.745 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.745 & 0 & 0 & 0.128 & 0.128 \\
0 & 0 & 0.745 & 0.256 & 0 & 0 \\
0 & 0 & 0.128 & 0.745 & 0.128 & 0 \\
0 & 0.128 & 0 & 0.128 & 0.745 & 0 \\
0 & 0.256 & 0 & 0 & 0 & 0.745
\end{bmatrix}
$$

(a) The ${}^{||}\Phi_1$ Matrix

$$
\Phi_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{bmatrix}
0.692 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.692 & 0 & 0 & 0.068 & 0.206 \\
0 & 0 & 0.692 & 0.137 & 0.137 & 0 \\
0 & 0 & 0.068 & 0.692 & 0.068 & 0 \\
0 & 0.068 & 0.137 & 0.068 & 0.692 & 0 \\
0 & 0.274 & 0 & 0 & 0 & 0.692
\end{bmatrix}
$$

(b) The Φ_1 Matrix

Figure 4-20. The ${}^{||}\Phi_{11}$, Φ_1 and ${}^{||}\Phi_{11}\Phi_1^{-1}$ Matrices of the Models Based on the Weight Matrices That Were Obtained by Applying the Strip Region Approach.

$$\|\Phi_1\Phi_1^{-1} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{bmatrix} 1.08 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.13 & -0.02 & -0.01 & 0.08 & -0.15 \\ 0.00 & 0.03 & 1.11 & 0.17 & -0.02 & -0.01 \\ 0.00 & -0.01 & -0.01 & 1.06 & 0.07 & 0.00 \\ 0.00 & 0.01 & -0.23 & 0.12 & 1.10 & -0.03 \\ 0.00 & -0.07 & -0.00 & -0.00 & 0.01 & 1.10 \end{bmatrix}$$

(c) The $\|\Phi_1\Phi_1^{-1}$ Matrix

Figure 4-20. (Continued)

Table 4-9. The Sample Space-Time Autocorrelation Functions and the Standardized Sample S-T Autocorrelation Functions of the Residuals of the Non-Equally Preferential Model

Space-Time Autocorrelations/Standardized S-T Autocorrelations

Space Lag / Time Lag	0	1	2	0	1	2
1	-0.03	0.01	-0.04	-0.02	0.11	-0.03
2	0.11	0.09	0.08	0.68	0.57	0.53
3	0.11	0.18	0.10	0.64	0.98	0.57
4	-0.18	-0.13	-0.09	-0.90	-0.68	-0.46

$$N_{\Phi_1} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{bmatrix} 0.63 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.63 & 0 & 0.63 & 0.13 & 0.17 \\ 0 & 0 & 0.63 & 0.16 & 0.24 & 0 \\ 0 & 0 & 0.13 & 0.63 & 0.16 & 0 \\ 0 & 0.09 & 0.15 & 0.13 & 0.63 & 0 \\ 0 & 0.30 & 0 & 0 & 0 & 0.63 \end{bmatrix}$$

(a) The ${}^{N}\Phi_1$ Matrix

$$\Phi_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{bmatrix} 0.67 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.67 & 0 & 0 & 0.14 & 0.18 \\ 0 & 0 & 0.67 & 0.17 & 0.15 & 0 \\ 0 & 0 & 0.14 & 0.67 & 0.18 & 0 \\ 0 & 0.09 & 0.09 & 0.13 & 0.67 & 0 \\ 0 & 0.32 & 0 & 0 & 0 & 0.67 \end{bmatrix}$$

(b) The Φ_1 Matrix

$$N_{\Phi_1}\Phi_1^{-1} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{bmatrix} 0.94 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.95 & 0 & 0 & 0 & -0.01 \\ 0 & -0.03 & 0.93 & 0.03 & 0.16 & 0.01 \\ 0 & 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0.10 & 0.02 & 0.93 & 0 \\ 0 & -0.01 & 0 & 0 & 0 & 0.95 \end{bmatrix}$$

(c) The ${}^{N}\Phi_1\Phi_1^{-1}$ Matrix

Figure 4-21. The ${}^{N}\Phi_1$, Φ_1 and ${}^{N}\Phi_1\Phi_1^{-1}$ Matrices of the Models Based on the Weight Matrices That Were Obtained by Applying the Angular Region Approach.

4.6.3 Discussion of Modeling Results

In the non-equally preferential model, $|{}^{|}\hat{\phi}_{11}|$ is greater than $|{}^{||}\hat{\phi}_{11}|$, which revels the fact that the diffusion process is stronger in the $|$ directions than in the $||$ directions. The driving force of the diffusion process in the $||$ directions, between locations that are parallel to the coast line, is the unassisted diffusion mechanism, while the driving force of the diffusion process in the $|$ directions, between locations that are perpendicular to the coast line, should be the combination of the unassisted diffusion mechanism and the effect of the sea breeze that increases the normal speed of diffusion as well as the magnitude of the mass transfer. The extra driving force from the sea wind makes the diffusion process in the $|$ directions stronger. Recall that in Section 3.6, we have analyzed the intervention effect of the I_1 intervention and found that the effect of the I_1 intervention is stronger in the $|$ direction. This implies the same non-equally preferential diffusion involvements as those that are implied in the non-equally preferential noise model. Although the sea wind assists the diffusion mechanism in the $|$ directions, it is not strong enough to change the diffusion speed, the diffusion speed is one order neighbor/month (or 16 miles/month) for both $|$ directions and $||$ directions.

In the model, the average value of ${}^{|}\hat{\phi}_{11}$ and ${}^{||}\hat{\phi}_{11}$ is 0.405, and the estimated $\hat{\phi}_{11}$ of the equally preferential model that has been built in Section 3.6 is 0.309. The average value of the ${}^{|}\hat{\phi}_{11}$ and ${}^{||}\hat{\phi}_{11}$ in the model 4-48 is 0.126 and the estimated $\hat{\phi}_{11}$ of the equally preferential model, Equation (4-42) in the $|$ directions and the $||$ directions is 0.137. It is seen that the $\hat{\phi}_{11}$ is estimated to be approximately the average of $^{|}\hat{\phi}_{11}$ and $^{||}\hat{\phi}_{11}$ in the models that only the | directions and the || directions are considered to give the preliminary test of the significance of different preference.

In Section 3.6, the interventional model has been built for the Los Angeles CO data without considering the non-equally preferential structure. Since the non-equal preference structure is significant for the pre-I_1 process, the consequences of ignoring the non-equal preference structure arises. In particular, will it make the estimates of the intrinsic utilities biased? In the following, this question will be answered for the non-environmental influence situation, i.e., I_m=0, first and then for the environmental influence situation, i.e., I_m=1.

The non-environment involved non-equal preferential intervention model, Equation (4-51),

$$(I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} N_{\phi_{k\ell}} N_W(\ell))(Z_t - \underset{\sim}{\delta}\xi_t) = (I - \sum_{k=1}^{p} \sum_{\ell=0}^{m_k} N_{\theta_{k\ell}} N_W(\ell))\underset{\sim}{\varepsilon}_t, \quad (4\text{-}51)$$

where

$$\xi_t = \begin{cases} 0 & \text{pre-intervention periods} \\ 1 & \text{post-intervention periods,} \end{cases}$$

$\underset{\sim}{\delta}$ is the intrinsic utility vector,

can be put into the linear model form as

$$\underset{\sim}{Z}_t = \xi_t\underset{\sim}{\delta} + (I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} N_{\phi_{k\ell}} W(\ell))^{-1}(I - \sum_{k=1}^{p} \sum_{\ell=0}^{m_k} N_{\phi_{k\ell}} W(\ell))\underset{\sim}{\varepsilon}_t.$$

Since the coefficient matrix of $\underset{\sim}{\delta}$ will not contain any model parameters $\underset{\sim}{N_\phi}, \underset{\sim}{N_\theta}$, so the estimates of $\underset{\sim}{\delta}$ will be unbiased even if significant non-equally preferential structures are ignored. This is not the same situation for the environment involved intervention process.

The environment involved, non-equal preference intervention model, Equation (4-52),

$$(I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} N_{\phi_{k\ell}} W(\ell))(\underset{\sim}{Z}_t) = (I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} N_{\theta_{k\ell}} N_W(\ell))(\xi_t\underset{\sim}{\delta} + \underset{\sim}{\varepsilon}_t), \quad (4\text{-}52)$$

can be put into the linear model form as

$$\underset{\sim}{Z}_t = (I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} N_{\phi_{k\ell}} N_W(\ell))^{-1}(I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} N_{\theta_{k\ell}} N_W(\ell))(\xi_t\underset{\sim}{\delta}+\underset{\sim}{\varepsilon}_t)$$

If all the ${}^{N}\phi_{k\ell}{}^{N}W^{(\ell)}$, ${}^{N}\theta_{k\ell}{}^{N}W^{(\ell)}$ terms are correctly assigned, the estimates of $\underset{\sim}{\delta}$ is unbiased, otherwise it is biased. Further, the closer ${}^{N}\phi_{k\ell}{}^{N}W^{(\ell)}$, ${}^{N}\theta_{k\ell}{}^{N}W^{(\ell)}$ are to their corresponding correct matrices, the less the bias in the $\underset{\sim}{\delta}$ estimates. An efficient way to check the closeness of matrices of full rank, say A and B, is to check how far AB^{-1} or BA^{-1} is from the identity matrix I. If $AB^{-1} \approx I$ or $BA^{-1} \approx I$, then for practical purposes, we may say that $A \approx B$.

In Figure 4-21(c), it is seen that ${}^{N}\Phi_1\Phi_1^{-1}$ is pretty close to identity matrix I, where ${}^{N}\Phi_1$ is the estimate of the correct non-equal preference matrix and Φ_1 is the mistaken equal preference matrix. Since the 1st intervention I_1 is environment involved and the 2nd intervention I_2 is non-environment involved, we may conclude that the estimates of the intervention effect of I_1 is biased, while the estimates of the effect of I_2 is unbiased. Although the estimates of the effect of I_1 is biased, since ${}^{N}\Phi_1\Phi_1^{-1}$, the biasness is not severe.

CHAPTER V

THE PURELY SPATIAL ARMA MODELS

A process that is related to the space-time process is the purely spatial process. For this process the influence spreads at a pretty high speed and the equilibrium state is reached before the next observation so that no influence is transmitted to the next period. We will refer to the following model as the general purely spatial model.

$$\underset{\sim}{Z}_t = B(0)\underset{\sim}{Z}(t) - A(0)\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T. \tag{5-1}$$

where

$\underset{\sim}{Z}(t)$ is $LN \times 1$ column observation vector,

$\underset{\sim}{\varepsilon}(t)$ is multivariate normal random vector with mean $\underset{\sim}{0}$, and

$$E[\underset{\sim}{\varepsilon}_t \varepsilon'_{t+k}] = \begin{cases} G & k=0 \\ 0 & \text{otherwise} \end{cases}$$

$B(0)$, $A(0)$ are $LN \times LN$ square parameter matrices. LN is the location number of the system.

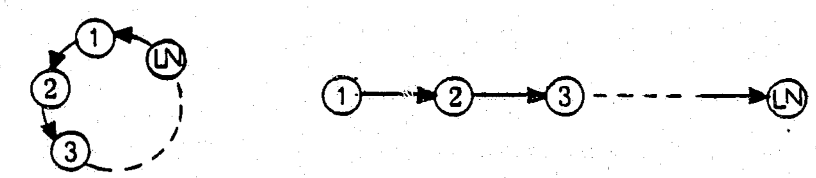
The existence conditions of the purely spatial model, that correspond to the stationary conditions of the space-time model, are studied in Section 5.1. The parameter constraints that form the existence regions are discussed and the necessary existence conditions are developed for the purely spatial AR, MA and ARMA models. In Section 5.2 the model identification problems are discussed. This includes defining the purely spatial autocorrelation function set, determining their characteristic properties and developing the expected sample purely spatial autocorrelation functions of low order models for the pattern recognition. In Section 5.2 charts/monograms are developed to yield initial estimates of the parameters for low order models. The M.L. estimation procedures and joint confidence intervals for purely spatial models assuming $G = \sigma^2 I$ are obtained in Section 5.4. Diagnostic checking considerations include the test of the noise sphericity assumption, the white noise assumption and significance tests of the model parameters, are detailed in Section 5.5. In Section 5.6, the M.L. estimation procedures for a more general noise covariance assumption of $G = \sigma^2 I$ is developed. An hydrology example, involving the Mohawk River Heights and a crimenology example, Assault in North-east Boston, are given in Section 5.7 to illustrate the purely spatial model building procedures.

5.1 Existence Conditions

Comparing the general purely spatial model with the following general space-time model,

$$\underset{\sim}{Z}_t = \sum_{k=1}^{p} B(k)\underset{\sim}{Z}_{t-k} - \sum_{k=1}^{q} A(k)\underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T, \qquad (5\text{-}2)$$

we see that the general purely spatial model contains no temporal back-shift operators while the general space-time model does. Both models, Equations (5-1) and (5-2), represent the structure that describes how the observations of one location are influenced by the observations and the white noise of other locations. The purely spatial model represents the structure that spells out only "instant influences" whereas the space-time model represents only the "delayed influences". To further compare and contrast these models, it is instructive to consider the backward spatial regressive structure of the One-Direction Circular Purely Spatial AR(1) System and the backward temporal regressive structure of the univariate time series AR(1) model. This One-Direction Circular Purely Spatial AR(1) System is a specific case of the purely spatial process, while the univariate time series AR(1) model is a space-time AR(1_0) process with only one location.



(a) (b)

Figure 5-1. (a) The One-Direction Ciruclar Purely Spatial AR(1) System
(b) The Univariate Time Series AR(1) Model

In the One-Direction Circular Purely Spatial AR(1) System, Figure 5-1, the location (i+1) is the first order neighbor of the location i for i=1,2,...,LN-1, and location 1 is the first order neighbor of location LN. This model can be expressed as,

$$\begin{cases} Z_{i,t} = \phi_{01}Z_{i-1,t} + \varepsilon_{i,t} & t=1,2,\ldots,T. \\ & i=2,3,\ldots,LN. \\ Z_{1,t} = \phi_{01}Z_{LN,t} + \varepsilon_{1,t} & t=1,2,\ldots,T. \end{cases} \tag{5-3}$$

Setting $t = t_0$, we have

$$\begin{cases} Z_{i,t_0} = \phi_{01}B_L Z_{i,t_0} + \varepsilon_{i,t_0} & i=2,3,\ldots,LN \quad (5\text{-}4) \\ Z_{1,t_0} = \phi_{01}B_L^{(LN-1)} Z_{LN,t_0} + \varepsilon_{1,t_0} & \quad (5\text{-}5) \end{cases}$$

where B_L is the spatial order backward operator and is defined as

$$B_L Z_{i,t_0} = Z_{i-1,t_0}, \quad 2 \le i \le LN$$

for the One-Direction Circular Purely Spatial System. Comparing Equations (5-4) and (5-5) with the following time series univariate AR(1) model,

$$Z_t = \phi_1 B Z_t + \varepsilon_t, \quad t=1,2,\ldots,LN \tag{5-6}$$

We see that the Equation (5-4) has the same backward regressive structure as that of the Equation (5-6), but Equation (5-5) cannot be equivalent to Equation (5-6) because Equation (5-6) contains B-operator, while Equation (5-5) contains a $B_L^{(LN-1)}$-operator which causes the influence that transmits through the system to be input back to the system again. So if the parameter ϕ_1 makes model, Equation (5-6), a non-stationary process, then due to the "instant feedback effect" of model, Equation (5-3), setting $\phi_{01} = \phi_1$ will make model, Equation (5-3), explosive immediately and non-existent. From the above discussion we see that the existence conditions of the purely spatial models correspond to the stationary, reversible conditions of the Space-Time Models. The major difference in physical meaning lies in the interpretation that one is "nonexistence immediately" while the other is "non-stationary or non-reversible in the long run".

A very important consideration concerning the purely spatial model is the existence condition, under which the vector process $\underset{\sim}{Z}(t)$ exists. By self-substitution n_0 times, the model, Equation (5-3), can be rewritten as,

$$\underset{\sim}{Z}_t = B^{n_0}(0)\underset{\sim}{Z}_t - \sum_{i=0}^{n_0-1} B^i(0)A(0)\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t \tag{5-7}$$

or

$$\underset{\sim}{\varepsilon}_t = A^{n_0}(0)\underset{\sim}{\varepsilon}_t - \sum_{i=1}^{n_0-1} A^i(0)B(0)\underset{\sim}{Z}_t + \underset{\sim}{Z}_t \tag{5-8}$$

From the Equations (5-7) and (5-8), it is clear that we should impose the following restrictions on the parametric matrix to avoid explosion.

The $\lim\limits_{n_0\to\infty} B^{n_0}(0)$, $\lim\limits_{n_0\to\infty} \sum_{i=0}^{n_0-1} B^i(0)A(0)$, $\lim\limits_{n_0\to\infty} A^{n_0}(0)$, $\lim\limits_{n_0\to\infty} \sum_{i=0}^{n_0-1} A^i(0)B(0)$ all must converge.

Any violation of these restrictions means that no such vector process $\underset{\sim}{Z}(t)$ can exist.

Since $\lim\limits_{n_0\to\infty} \sum_{i=0}^{n_0-1} B^i(0)$ and $\lim\limits_{n_0\to\infty} B^{n_0}(0)$ must converge implies that $\lim\limits_{n_0\to\infty} B^{n_0}(0) = 0$. Therefore the conditions for the purely model to exist are

$$\lim_{n_0\to\infty} B^{n_0}(0) = 0 \tag{5-9}$$

and

$$\lim_{n_0\to\infty} A^{n_0}(0) = 0$$

A matrix M is said to be power convergent when the powers $M, M^2, M^3, \ldots, M^n \ldots$ are a convergent sequence. It is well known in linear algebra that a matrix M is power convergent if and only if for each eigenvalue λ of M, $|\lambda| < 1$. Thus, Equation (5-9) is to satisfy these conditions, B(0) and A(0) must be power convergent. This is equivalent to constrain that each eigenvalue λ of B(0) and A(0) to be $|\lambda| < 1$.

5.1.1 The Existence Condition of the Purely Spatial AR(λ_0) Model

By setting

$$B(0) = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \text{ and } A(0) = 0$$

for the general purely spatial model class we obtain the purely spatial AR(λ_0) model,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T, \tag{5-10}$$

where $W^{(\ell)}$ is the weight matrix of the ℓ^{th} spatial order. Since all the eigenvalues λ of A(0) = 0 are 0, so the existence condition requires that $|\lambda| < 1$ for all the eigenvalues of $\sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}$. It is known that the eigenvalue λ of $K \times K$ matrix M satisfies $|\lambda| \le \gamma_i$ for $i=1,2,\ldots,K$, where $\gamma_i = \sum_{j=1}^{K} m_{ij}$ and m_{ij} is the (i,j) element of M. Under the assumption that all the weight matrix are normalized, the necessary condition of existence becomes,

$$\sum_{\ell=1}^{\lambda_0} |\phi_{0\ell}| < 1. \tag{5-11}$$

For example, the necessary existence conditions for AR(1) and AR(2) are $|\phi_{01}| < 1$ and $|\phi_{01}| + |\phi_{02}| < 1$, respectively. The existence region of AR(2) is the interior region of the diamond in Figure 5-2.

5.1.2 The Existence Condition of the Purely Spatial MA(m_0) Model

By setting

$$B(0) = 0 \quad \text{and} \quad A(0) = \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)}$$

for the general purely spatial model, we obtain the purely spatial MA(m_0) model,

$$\underset{\sim}{z}_t = - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T. \tag{5-12}$$

Following the same approach that has been made in the AR(λ_0) models, we obtain the necessary existence condition for the MA(m_0) models as

$$\sum_{\ell=1}^{m_0} |\theta_{0\ell}| < 1 . \tag{5-13}$$

For example, the necessary existence conditions for MA(1) and MA(2) are $|\theta_{01}| < 1$ and $|\theta_{01}| + |\theta_{02}| < 1$, respectively. The necessary existence region of MA(2) is the interior region of the diamond in Figure 5-3.

5.1.3 The Existence Condition of the Purely Spatial ARMA(λ_0,m_0) Model

By setting

$$B(0) = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}, \quad A(0) = \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)},$$

for the general purely spatial model we obtain the purely spatial ARMA(λ_0,m_0) model,

$$\underset{\sim}{z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{z}_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t. \tag{5-14}$$

The necessary existence condition for ARMA(λ_0,m_0) model is then,

$$\sum_{\ell=0}^{m_0} |\theta_{0\ell}| < 1 \quad \text{and} \quad \sum_{\ell=1}^{\lambda_0} |\phi_{0\ell}| < 1, \tag{5-15}$$

For example, the necessary existence region for ARMA(1,1) is $|\theta_{01}| < 1$ and $|\phi_{01}| < 1$, which is the interior region of the square in the Figure 5-4. The necessary existence region for ARMA(1,2) is bounded by $|\phi_{01}| < 1$ and $|\theta_{0\ell}| + |\theta_{02}| < 1$, which is the interior region of the

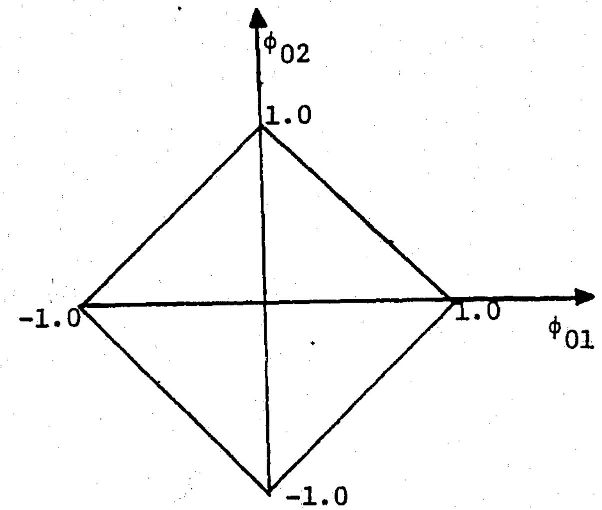


Figure 5-2. The AR(2) Existence Parameter Space

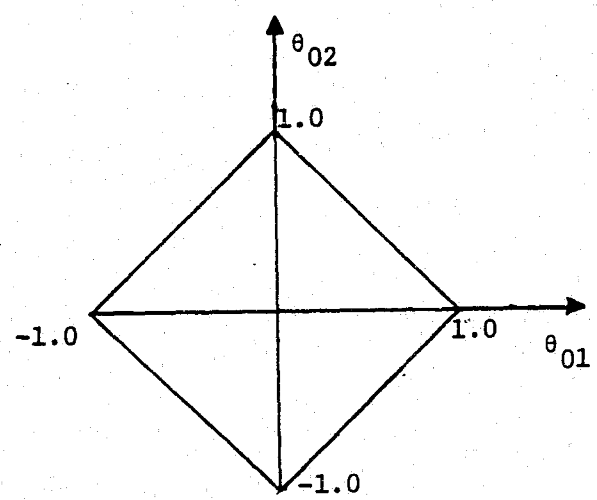


Figure 5-3. The MA(2) Existence Parameter Space

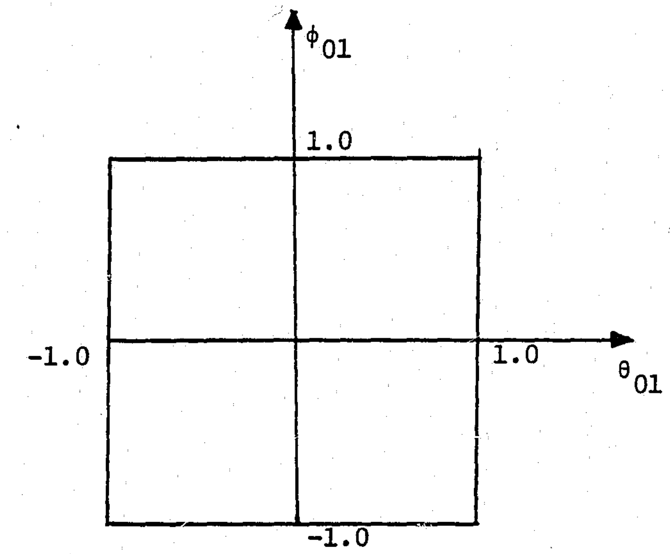


Figure 5-4. The ARMA(1,1) Existence Parameter Space

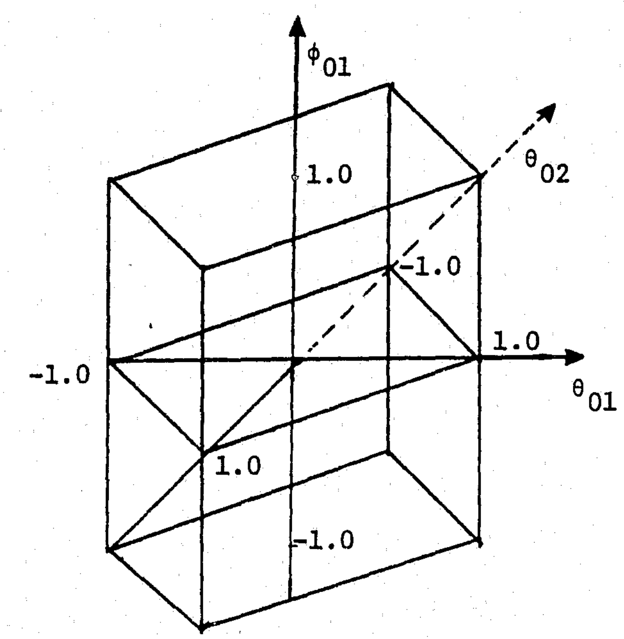


Figure 5-5. The ARMA(1,2) Existence Parameter Space

rectangular bar in the Figure 5-5.

In summary, low order model forms and the necessary existence conditions corresponding to that model form are shown in Table 5-1.

Table 5-1. The Necessary Existence Conditions of the Purely Spatial Models

	Model	Necessary Existence Area
$AR(\lambda_0)$	$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \underset{\sim}{\varepsilon}_t$, t=1,2,...,T.	$\sum_{\ell=1}^{\lambda_0} \lvert\phi_{0\ell}\rvert < 1$
$MA(m_0)$	$\underset{\sim}{Z}_t = -\sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t$, t=1,2,...,T.	$\sum_{\ell=1}^{m_0} \lvert\theta_{0\ell}\rvert < 1$
$ARMA(\lambda_0,m_0)$	$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t$, t=1,2,...,T.	$\sum_{\ell=1}^{\lambda_0} \lvert\phi_{0\ell}\rvert < 1$, and $\sum_{\ell=1}^{m_0} \lvert\theta_{0\ell}\rvert < 1$

It should be noted that the necessary existence region does not mean the best existence region, i.e., the largest region that the vector process can exist. Those model parameters that fall in the necessary existence region make the purely spatial process exist, but every parameter that does not fall in the necessary existence region will not necessarily make the purely spatial process explosive. According to the paper by Deutsch and Pfeifer [1980], the necessary existence condition for $AR(\lambda_0)$, $MA(m_0)$, $ARMA(\lambda_0,m_0)$ models of the regular grid system is also

the sufficient existence condition for λ_0, $m_0 \le 2$, i.e., the necessary existence region given in Table 5-1 is the best existence region for the regular grid system with λ_0, $m_0 \le 2$.

5.2 The Identification of the Purely Spatial Models

The identification of the candidate spatial model is addressed. The sample autocovariance $\hat{\gamma}_{ks}$ and sample autocorrelation function $\hat{\rho}_{ks}$ are defined along with their corresponding partial spatial autocorrelation function. Characteristic properties of these functions are addressed for different types of purely spatial process for the purpose of pattern recognition.

5.2.1 The Autocovariance and the Autocorrelation Function

The autocovariance function γ_{ks}, $k \ge 0$, $s \ge 0$ is defined as,

$$\gamma_{ks} = E(\hat{\gamma}_{ks}) = E\left(\frac{[W^{(k)}\underset{\sim}{Z}_t]'[W^{(s)}\underset{\sim}{Z}_t]}{LN}\right) = \frac{W'^{(k)}W^{(s)} \odot E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')}{LN} \tag{5-16}$$

The autocorrelation function ρ_{ks} is defined in terms of the autocovariance functions as,

$$\rho_{ks} = \frac{\gamma_{ks}}{[\gamma_{kk}\gamma_{ss}]^{1/2}} \tag{5-17}$$

In order to estimate ρ_{ks} and γ_{ks}, the $E(Z_tZ_t')$ has to be first computed. Note that in the Equation (5-16) only the term, $E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')$ contains the information about ϕ's and θ's. For the purely spatial $ARMA(\lambda_0,m_0)$

model, we have

$$E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t') = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})$$

$$(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})'(I - \sum_{\ell=1}^{m_0} \phi_{0\ell}W^{(\ell)})'^{-1}\sigma_\varepsilon^2. \qquad (5\text{-}18)$$

By setting $\theta_{0\ell} = 0$ for all ℓ, we have the $E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')$ of the purely spatial $AR(\lambda_0)$ model as,

$$E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t') = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})'^{-1}\sigma_\varepsilon^2. \qquad (5\text{-}19)$$

By setting $\phi_{0\ell} = 0$ for all ℓ, we have the $E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')$ of purely spatial $MA(m_0)$ model as,

$$E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t') = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})'\sigma_\varepsilon^2. \qquad (5\text{-}20)$$

For given $W^{(\ell)}$'s, it can be proven that the autocorrelation functions, defined in Equation (5-17), will cut off for the purely spatial $MA(m_0)$ model by substituting Equation (5-20) into Equation (5-17) and rearranging. The cut-off point depends on the weight matrices $W^{(\ell)}$'s. The $E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')$ of purely spatial $AR(\lambda_0)$ and $AR(\lambda_0,m_0)$ models contain the

inversed term $(I - \sum_{\ell=1}^{m_0} \phi_{0\ell}W^{(\ell)})^{-1}$, which can be expanded into infinite power series in terms of $(\sum_{\ell=1}^{m_0} \phi_{0\ell}W^{(\ell)})$. So substituting Equation (5-18) and Equation (5-19) into Equation (5-17), we result in the autocorrelation functions of purely spatial $AR(\lambda_0)$ and $ARMA(\lambda_0,m_0)$ models will tail off.

5.2.2 The Identification of the Purely Spatial $MA(m_0)$ Processes

Since the cut-off property of the autocorrelation functions is useful in identifying the MA processes, we will discuss the cut-off property of ρ_{ks} for the purely spatial $MA(m_0)$ model in more detail. The following definitions of sets that contain the information of the neighbor structure are given to help in studying the nature of this cut-off property. For each location i of the system, we associate two sets, which indicate the neighbor structure, with it, i.e., $S_i(k)$, $H_i(k)$.

$S_i(k)$: Set which contains all the k^{th} order neighbors of location i.

$H_i(k)$: Set which contains all the locations that location i is their k^{th} order neighbor.

For example, in the 5×5 regular grid system plotted in Figure 5-6(b), we have $S_{13}(1) = \{8,12,14,18\}$, $S_{13}(1) = \{8,12,14,18\}$.

For the purely spatial $MA(m_0)$ model, $\rho_{ks} = 0$ iff

$$S_{zz}(i,j) = \phi \quad \text{or} \quad S_{WW}(i,j,k,s) = \phi \quad \text{for all locations } i,j. \qquad (5\text{-}21)$$

where

$$S_{ZZ}(i,j) = \bigcup_{k_1=0}^{m_0} \bigcup_{k_2=0}^{m_0} [(S_i(k_1)) \bigcap (H_j(k_2))]$$

$$S_{WW}(i,j,k,s) = [(H_i(k)) \bigcap (S_j(s))], \text{ and } \phi$$

denotes the empty set.

Since the cut-off behavior is influenced by the neighbor structure, we need to be more specific about the type of neighbor structure before we can go further. We will use a line system and two dimension regular grid system in this section to serve as examples. These two systems as well as their neighbor structures are given in Figure 5-6. For practical purpose, these two systems can be treated as the representative for the line system and the 2-dimension system respectively given that there are more than or equal to 25 locations in the system since the boundary effect is negligible. See Deutsch and Pfeifer [1980e].

For the line system, let $e_Z(i,j)$ be the (i,j) element of the $E(\underset{\sim}{Z}_t\ \underset{\sim}{Z}_t')$, then

$$e_Z(i,j) = 0 \quad \text{iff} \quad |i-j| > 2m_0 \tag{5-22}$$

$$\text{i.e., } S_{ZZ}(i,j) = \{j\,|\,|i-j| > 2m_0\} = \phi$$

Let $d_{sk}(i,j)$ is the (i,j) element of $W^{\prime(s)}W^{(k)}$, then

$$d_{sk}(i,j) \neq 0 \quad \text{iff} \quad |i-j| = |k-s|. \tag{5-23}$$

$$\text{i.e., } S_{WW}(i,j,s,k) = \{j\,|\,|i-j| = |k-s|\} \neq \phi$$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

(a)

1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

(b)

Figure 5-6(a). The 1×25 Line System
(b). The 5×5 Regular Grid System

Combing Equations (5-22) and (5-23), we have the cut-off property of ρ_{sk} for $MA(m_0)$ as follows:

$$\rho_{ks} = 0 \ \text{ if } \ k > s + 2m_0 \ \text{ or } \ 0 \leq k < s - 2m_0 \tag{5-24}$$

Location / Order	1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10	11
2	1,3	4	5	6	7	8	9	10	11	12
3	2,4	1,5	6	7	8	9	10	11	12	13
4	3,5	2,6	1,7	8	9	10	11	12	13	14
5	4,6	3,7	2,8	1,9	10	11	12	13	14	15
6	5,7	4,8	3,9	2,10	1,11	12	13	14	15	16
7	6,8	5,9	4,10	3,11	2,12	1,13	14	15	16	17
8	7,9	6,10	5,11	4,12	3,13	2,14	1,15	16	17	18
9	8,10	7,11	6,12	5,13	4,14	3,15	2,16	1,17	18	19
10	9,11	8,12	7,13	6,14	5,15	4,16	3,17	2,18	1,19	20
11	10,12	9,13	8,14	7,15	6,16	5,17	4,18	3,19	2,20	1,21
12	11,13	10,14	9,15	8,16	7,17	6,18	5,19	4,20	3,21	2,22
13	12,14	11,15	10,16	9,17	8,18	7,19	6,20	5,21	4,22	3,23
14	13,15	12,16	11,17	10,18	9,19	8,20	7,21	6,22	5,23	4,24
15	14,16	13,17	12,18	11,19	10,20	9,21	8,22	7,23	6,24	5,25
16	15,17	14,18	13,19	12,20	11,21	10,22	9,23	8,24	7,25	6
17	16,18	15,19	14,20	13,21	12,22	11,23	10,24	9,25	8	7
18	17,19	16,20	15,21	14,22	13,23	12,24	11,25	10	9	8
19	18,20	17,21	16,22	15,23	14,24	13,25	12	11	10	9
20	19,21	18,22	17,23	16,24	15,25	14	13	12	11	10
21	20,22	19,23	18,24	17,25	16	15	14	13	12	11
22	21,23	20,24	19,25	18	17	16	15	14	13	12
23	22,24	21,25	20	19	18	15	16	15	14	13
24	23,25	22	21	20	19	18	17	16	15	14
25	24	23	22	21	20	19	18	17	16	15

Figure 5-6(c). The Neighbor Structure of the 1×25 Line System

Location	Spatial Order 1	2	3	4	5	6	7	8	9	10
1	2, 6	7	3,11	8,12	13	4,16	9,17	14,18	19	5,21
2	1, 3 7	6, 8	4,12	9,11 13	14	5,17	10,16 18	15,19	20	22
3	2, 4 8	7, 9	1, 5 13	6,10 12,14	11,15	18	17,19	16,20	23	22,24
4	3, 5 9	8,10	2,14	7,13 15	12	1,19	6,18 20	11,17	16	24
5	4,10	9	3,15	8,14	13	2,20	7,19	12,18	17	1,25
6	1,7 11	2,12	8,16	3,13 17	18	9,21	4,14 22	19,23	24	10
7	2, 6 8,12	1, 3 11,13	9,17	4,14 16,18	19	10,22	5,15 21,23	20,24	25	—
8	3, 7 9,13	2, 4 12,14	6,10 18	1, 5 11,15 17,19	16,20	23	22,24	21,25	—	—
9	4, 8 10,14	3, 5 13,15	7,19	2,12 18,20	17	6,24	1,11 23,25	16,22	21	—
10	5, 9 15	4,14	8,20	3,13 19	18	7,25	2,12 24	17,23	22	6
11	6,12 16	7,17	1,13 21	2, 8 18,22	3,23	14	9,19	4,24	15	10,20
12	7,11 13,17	6, 8 16,18	2,14 22	1, 3 9,19 21,23	4,24	15	10,20	5,25	—	—

Figure 5-6(d). The Neighbor Structure of the 5X5 Regular Grid System.

13	8,12 14,18	7, 9 17,19	3,11 15,23	3, 4 6,10 16,20 22,24	1, 5 21,25					
14	9,13 15,19	8,10 18,20	4,12 24	3, 5 7,17 23,25	2,22	11	6,16	1,21	—	—
15	10,14 20	9,19	5,13 25	4, 8 18,24	3,23	12	7,17	2,22	11	6,16
16	11,17 21	12,22	6,18	7,13 23	8	1,19	2,14 24	3, 9	4	20
17	12,16 18,22	11,13 21,23	7,19	6, 8 14,24	9	2,20	1, 3 15,25	4,10	5	—
18	13,17 19,23	12,14 22 24	8,16 20	7, 9 11 15 21,25	6,10	3	2, 4	1, 5	—	—
19	14,18 20,24	13,15 23,25	9,17	8,10 12,22	7	4,16	3, 5 11,21	2, 6	1	—
20	15,19 25	14,24	10,18	9,13 23	8	5,17	4,12 22	3, 7	2	16
21	16,22	17	11,23	12,18	13	6,24	7,19	8,14	9	1,25
22	17,21 23	16,18	12,24	11,13 19	14	7,25	6, 8 20	9,15	10	2
23	18,22 24	17,19	13,21 25	12,14 16,20	11,15	8	7, 9	6,10	3	2, 4
24	19,23 25	18,20	14,22	13,15 17	2	9,21	8,10 16	7,11	6	4
25	20,24	19	15,23	14,18	13	10,22	9,17	8,12	7	5,21

Figure 5-6(d). (Cont'd)

For the two dimension grid system, because there is no ordered location alignment and neighborhood structures are more complicated, the resulting cut-off orders of ρ_{ks} cannot be put in a closed form. However, for the 5×5 regular grid system, the following results for the MA(1), MA(2), MA(3) and MA(4) processes were obtained.

MA(1)

$$\rho_{0k} \begin{cases} \neq 0 & \text{if } k \leq 3 \\ = 0 & \text{otherwise} \end{cases}$$

$$\rho_{1k} \begin{cases} \neq 0 & \text{if } k \neq 5,\ k \leq 6 \\ = 0 & \text{otherwise} \end{cases}$$

$$\rho_{2k} \begin{cases} \neq 0 & \text{if } k \leq 7,\ k \neq 6 \\ = 0 & \text{otherwise} \end{cases}$$

$$\rho_{3k} \begin{cases} \neq 0 & \text{if } k \leq 10,\ k \neq 5,8,9 \\ = 0 & \text{otherwise} \end{cases}$$

MA(2)

$$\rho_{0k} \begin{cases} \neq 0 & \text{if } k \leq 5 \\ = 0 & \text{otherwise} \end{cases}$$

MA(3)

$$\rho_{0k} \begin{cases} \neq 0 & \text{if } k \leq 10 \\ = 0 & \text{otherwise} \end{cases}$$

MA(4)

$$\rho_{0k} \begin{cases} \neq 0 & \text{if } k \leq 12 \\ = 0 & \text{otherwise} \end{cases}$$

Note that for MA(1) and MA(2) processes, the ρ_{0k} of the line system cut off at k=1 and k=2, respectively. While the ρ_{0k} of the 2-dimension system doesn't cut off at k=1 and k=2, but cuts off at k=3 and k=5, respectively.

5.2.3 Partial Autocorrelation Function Sets

The partial autocorrelation function sets are helpful in identifying the purely spatial AR models. To analyze their characteristic properties we employ the following AR(λ_0) model,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots$$

Since

$$E[(W^{(k)}\underset{\sim}{\varepsilon}_t)'\underset{\sim}{Z}_t] = E[(W^{(k)}(I - \sum_{\ell=0}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\underset{\sim}{Z}_t)'\underset{\sim}{Z}_t],$$

and

$$E[(W^{(k)}\underset{\sim}{\varepsilon}_t)'\underset{\sim}{\varepsilon}_t] = 0, \quad k \neq 0,$$

We have the following simultaneous equations,

$$E[\underset{\sim}{Z}_t' W'^{(k)} \underset{\sim}{Z}_t] = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} \,[E(\underset{\sim}{Z}_t'(W'^{(\ell)}W'^{(k)} + W'^{(k)}W^{(\ell)})\underset{\sim}{Z}_t)$$
$$- \sum_{s=1}^{\lambda_0} \sum_{\ell=1}^{\lambda_0} \phi_{0s}\phi_{0\ell}\, E[\underset{\sim}{Z}_t' W'^{(s)} W'^{(k)} W^{(\ell)} \underset{\sim}{Z}_t], \qquad (5\text{-}25)$$
$$k=1,2,\ldots,\lambda_0 .$$

Defining,

$$C_{k\ell} = E(\underset{\sim}{Z}_t' W'^{(k)} W^{(\ell)} \underset{\sim}{Z}_t) = W'^{(k)} W^{(\ell)} \ominus E(\underset{\sim}{Z}_t \underset{\sim}{Z}_t')$$

$$D_{sk\ell} = E(\underset{\sim}{Z}_t' W'^{(s)} W'^{(k)} W^{(\ell)} \underset{\sim}{Z}_t) = W'^{(s)} W'^{(k)} W^{(\ell)} \ominus E(\underset{\sim}{Z}_t' \underset{\sim}{Z}_t)$$

Equation (5-25) can be rewritten as,

$$C_{k0} = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} \,(C_{k\ell} + D_{0k\ell}) - \sum_{s=1}^{\lambda_0} \sum_{\ell=1}^{\lambda_0} \phi_{0s}\phi_{0\ell} D_{sk\ell} \qquad (5\text{-}26)$$
$$k=1,2,3,\ldots,\lambda_0 ,$$

Equation (5-26) which contains λ_0 equations is a simultaneous quadratic equation.

We define the k_0^{th} partial autocorrelation function set $S(k_0)$ as the set of all the solutions of this equation with $\lambda_0 = k_0$. If $k_1 > k_2$, then the set $S(k_1)$ shares a solution with the set $S(k_2)$ if the set $S(k_1)$ has a solution with $\phi_{01}, \phi_{02}, \ldots, \phi_{0k2}$ equal to a solution contained in the solution set of $S(k_2)$ and has $\phi_{0(k2+1)}, \ldots, \phi_{0k1} = 0$ in the solution. Theoretically, if the true model is purely spatial $AR(\lambda_0)$, then we will have the situation that $S(1), S(2), \ldots, S(\lambda_0)$ share no solution, while $S(\lambda_0), S(\lambda_{0+1}), \ldots$ share a solution, i.e., $(\phi_{01}, \phi_{02}, \ldots, \phi_{0\lambda_0}, 0, 0, \ldots)$. In such situations, the partial autocorrelation function sets cuts-off at $k = \lambda_0$. In order to estimate the spatial partial autocorrelation substitute for $E(\underset{\sim}{Z}_t \underset{\sim}{Z}_t')$ with the $(\sum_{t=1}^{T} \underset{\sim}{Z}_t \underset{\sim}{Z}_t'/T)$, the sample autocovariance matrix, and then need to solve the simultaneous quadratic equations to obtain the sample partial autocorrelation function sets, $\hat{S}(1), \hat{S}(2), \ldots, \hat{S}(\lambda_0), \ldots$. The following example for a purely spatial AR(1) process in a 5×5 regular grid system is used to illustrate the cut-off property of the partial autocorrelation function sets.

Suppose we have an AR(1) model with $\phi_{01} = 0.9$, for a 5×5 regular grid system whose neighbor structure is given in Figure 5-6(d), we want to compute the partial autocorrelation set. At first, we set k=1, and obtain the following quadratic equation to solve for the ϕ_{01};

$$13.0374277151\phi_{01}^2 - 27.2580923337\phi_{01} + 13.971957651620 = 0$$

The solutions are $\phi_{01} = 0.9$ or $\phi_{01} = 1.19076$ or $s(1) = \{(0.9, 1.19076)\}$. Then we set k=2, and have the following two simultaneous quadratic equations to solve for ϕ_{01} and ϕ_{02};

$$k = 2 \quad \begin{cases} 13.0374277151\ \phi_{01}^2 + 23.2482853728\ \phi_{01}\phi_{02} + 10.9886205076\ \phi_{02}^2 \\ \qquad - 27.2580823337\ \phi_{01} - 23.3810108136\ \phi_{02} \\ \qquad\qquad + 13.9719576516 = 0 \\ 11.7515601773\ \phi_{01}^2 + 22.301871282\ \phi_{01}\phi_{02} + 10.2628662299\ \phi_{02}^2 \\ \qquad - 23.5812615590\ \phi_{01} - 23.4004319218\ \phi_{02} \\ \qquad\qquad + 11.7043716595 = 0 \end{cases}$$

The above simultaneous quadratic equations can be solved graphically. In Figure 5-7, the whole ellipse is plotted according to the first quadratic equation and the partial ellipse is plotted according to the second quadratic equation. Then the intersection of these two ellipses are the solutions of the above simultaneous quadratic equations. The solution set $S(2) = \{(0.9, 0), (1.04, 0.14)\}$ is read from Figure 5-7. Note that S(1), S(2) share the solution $\phi_{01} = 0.9$, $\phi_{02} = 0$. We see that the partial autocorrelation function sets cut-off after the 1st space lag for this purely spatial AR(1) model.

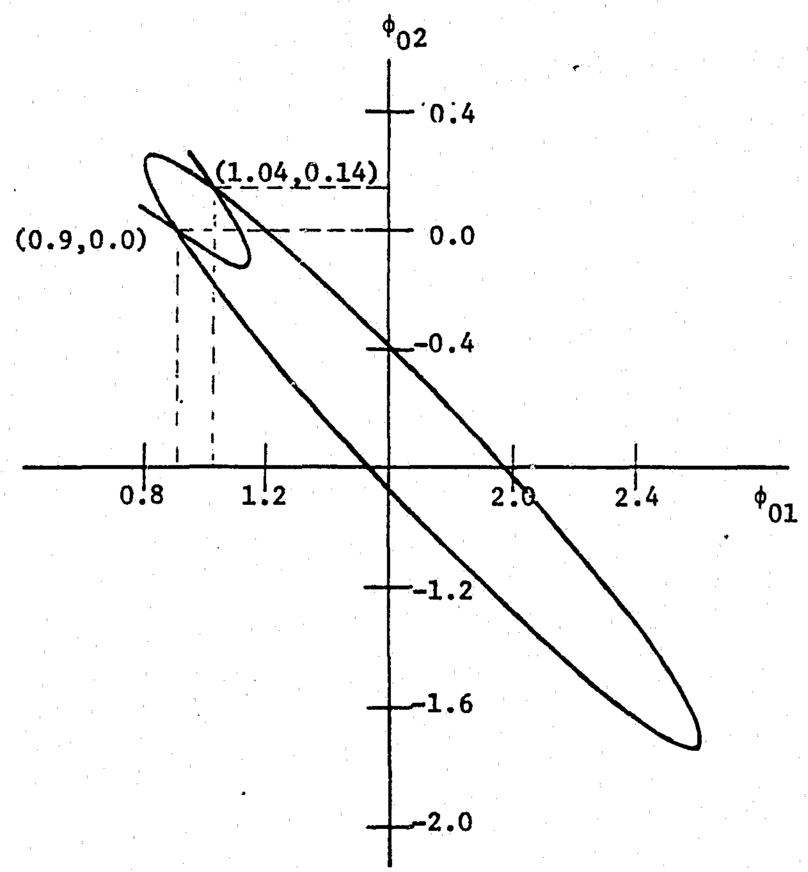


Figure 5-7. The Graphic Solutions of the Simultaneous Quadratic Equations in the Example of Identifying the Purely Spatial AR(1) Model.

Applying the power expansion on $(I - \sum_{\ell=0}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}$ of the purely spatial ARMA(λ_0,m_0) model, we can transform the purely spatial ARMA(λ_0,m_0) model into its equivalent infinite order purely spatial MA model with the amplitudes of the coefficients decay exponentially, so the purely spatial autocorrelation functions of the purely spatial ARMA(λ_0,m_0) model tail off. Applying the power expansion on $(I - \sum_{\ell=0}^{\lambda_0} \theta_{0\ell}W^{(\ell)})^{-1}$ of the purely spatial ARMA(λ_0,m_0) model, we can transform this model into its equivalent infinite order purely spatial AR model. So the purely spatial partial autocorrelation function sets tail off. For the purely spatial ARMA(λ_0,m_0) model, neither the autocorrelation function nore the partial autocorrelation function sets cut-off. The cut-off, tail-off properties of the autocorrelation function and partial autocorrelation function sets is summarized in Table 5-2.

Table 5-2. Characteristics of the Autocorrelation Functions for Purely Spatial ARMA Models

Model Class	Autocorrelation Function ρ_k	Partial Autocorrelation Functions Sets S(k)
AR(λ_0)	Tail-off	Cut-off at $k = \lambda_0$
MA(m_0)	Cut-off at $k = k_0$ K_0 is determined by m_0 and the neighbor structure	Tail-off
ARMA(λ_0,m_0)	Tail-off	Tail-off

From Table 5-2, we see that when the sample autocorrelation functions $\hat{\rho}_k$ tail-off while the sample autocorrelation function sets $\hat{S}(k)$ cut-off at $k = \lambda_0$, then the underlying process is an AR(λ_0) process. When $\hat{\rho}_k$ cut-off and $\hat{S}(k)$ tail-off, the candidate model is a MA model. Also when $\hat{\rho}_k$ and $\hat{S}(k)$ both tail-off the underlying process is an ARMA model.

5.2.4 The Pattern Recognition

Since all the informations are contained in the autocorrelation functions, equations for the partial autocorrelation is computationally difficult, the approach for identification will emphasize a pattern comparison approach based upon the sample to theoretical autocorrelation to identify the candidate model. To do this pattern comparison, it is necessary to develop the expectation values of the sample autocorrelation function for all the ARMA(λ_0,m_0) models.

The expectation values of the sample autocorrelation function can be derived from the general purely spatial model,

$$\underset{\sim}{Z}_t = B(0)\underset{\sim}{Z}_t - A(0)\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \qquad t=1,2,\ldots,T \tag{5-27}$$

where

$$\underset{\sim}{\varepsilon}_t \sim \mathrm{NID}(0,\sigma^2 I).$$

Defining

$$M_{ks} = [(I - B(0))^{-1}(I - A(0))]' W'^{(k)} W^{(s)} [(I - B(0))^{-1}(I - A(0))]$$

We have the following formula

$$E(\hat{\rho}_{ks}) \doteq \frac{\gamma_{ks}}{(\gamma_{kk}\gamma_{ss})^{1/2}} \left\{1 - \frac{\mathrm{cov}(\hat{\gamma}_{ks},\hat{\gamma}_{kk})}{2\gamma_{ks}\,\gamma_{kk}} - \frac{\mathrm{cov}(\hat{\gamma}_{ks},\hat{\gamma}_{ss})}{2\gamma_{ks}\,\gamma_{ss}} + \frac{\mathrm{cov}(\hat{\gamma}_{kk},\hat{\gamma}_{ss})}{4\gamma_{kk}\gamma_{ss}} + \frac{3\mathrm{var}(\hat{\gamma}_{kk})}{8\gamma_{kk}^2} + \frac{3\mathrm{var}(\hat{\gamma}_{ss})}{8\gamma_{ss}^2}\right\} \tag{5-28}$$

where

$$\gamma_{ks} = E(\hat{\gamma}_{ks}) = \left(\frac{1}{LN}\right)T_r(M_{ks})\sigma^2$$

$$\mathrm{var}(\hat{\gamma}_{ks}) = \left(\frac{\sigma^4}{TLN^2}\right)\left(M_{ks} \odot M_{ks} + M_{ks} \odot M'_{ks}\right)$$

$$\mathrm{cov}(\hat{\gamma}_{ks},\hat{\gamma}_{\ell m}) = \left(\frac{\sigma^4}{TLN^2}\right)\left(M_{ks} \odot M_{m} + M_{ks} \odot M'_{\ell m}\right)$$

Note that although $E(\hat{\rho}_{ks}) \neq \rho_{ks}$, where ρ_{ks} is defined in the Equation (5-17), the cut-off property of the MA process is still held for $E(\hat{\rho}_{ks})$, since the cut-off property of $E(\hat{\rho}_{ks})$ depends on γ_{ks} only. Also, according to the customary procedure, we have

$$\mathrm{var}(\hat{\rho}_{ks}) \doteq \frac{\mathrm{var}(\hat{\gamma}_{ks})}{\gamma_{kk}\gamma_{ss}} - \frac{\gamma_{ks}\gamma_{kk}\ \mathrm{cov}(\hat{\gamma}_{ks},\hat{\gamma}_{ss}) + \gamma_{ks}\gamma_{ss}\ \mathrm{cov}(\hat{\gamma}_{ks},\gamma_{kk})}{\gamma_{kk}^2\gamma_{ss}^2}$$

$$+ \frac{\gamma_{ks}^2[\gamma_{kk}^2\mathrm{var}(\hat{\gamma}_{ss}) + \gamma_{ss}^2\mathrm{var}(\hat{\gamma}_{kk}) + 2\gamma_{ss}\gamma_{kk}\ \mathrm{cov}(\hat{\gamma}_{ss},\hat{\gamma}_{kk})}{4\gamma_{kk}^3\ \gamma_{ss}^3} \quad (5\text{-}29)$$

Since $\hat{\rho}_{0S}$ are enough in identifying the potential model, so we define $\hat{\rho}_S = \hat{\rho}_{0S}$, and the pattern of $E(\hat{\rho}_S)$ is used to help in the model identification. The sign of $E(\hat{\rho}_S)$ depends on $\gamma_{k0} \equiv W'^{(k)} \odot E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')$, the $E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')$ depends on the neighbor structure as well as the true model parameters, i.e., $\phi_{01},\ldots,\phi_{0\lambda_0}$, and the magnitude of $E(\hat{\rho}_k)$ depends on γ_{k0} and γ_{kk}, so we have come up with the general conclusion that the sign and magnitude of $E(\hat{\rho}_k)$ depend heavily on the neighbor structure as well as the model parameters, i.e., $\lambda_0,\phi_{01},\ldots,\phi_{0\lambda_0}$.

In the purely white noise process, i.e., $A(0) = 0$, $B(0) = 0$, we have the simpler formula for $E(\hat{\rho}_S)$ that corresponds to the Equations (5-28) and (5-29),

$$E(\hat{\rho}_s) = \begin{cases} 1 & s = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{var}(\hat{\rho}_s) = \begin{cases} 0 & s = 0 \\ \dfrac{W^{(s)} \odot (W^{(s)} + W'^{(s)})}{(T \cdot LN)T_r(W'^{(s)}W^{(s)})} & s > 0 \end{cases} \quad (5\text{-}30)$$

The expectation values of the sample autocorrelation functions $E(\hat{\rho}_k)$, k=1,2,...,10 of the purely spatial ARMA(λ_0,m_0) models with $\lambda_0 + m_0 \leq 2$ for the 1×25 line system and the 5×5 regular grid system are plotted. These plots are useful in identifying the low order purely spatial ARMA models.

In Figures 5-8 (a) - (d), the expected sample autocorrelation functions $E(\hat{\rho}_k)$ for the 1×25 line system of the AR(1), MA(1), ARMA(1,1), AR(2) and MA(2) models are plotted, and the expected sample autocorrelation functions for the 5×5 regular grid system are plotted in Figures 5-9 (a) - (d). Comparing the plots of 1×25 line system with those of the 5×5 regular grid system, we see that most of the patterns are similar except those of the AR(1), ARMA(1,1) models with negative ϕ_{01} values and the AR(2) models with (ϕ_{01},ϕ_{02}) negative. The similarity comes from the same nature of the model specification and the difference comes from the different natures of the neighbor structures between the 1-dimension line system and the 2-dimension regular grid system.

From the plots in Figures 5-8(a) and 5-8(d), we see that MA(m_0) models do have cut-off correlations as we expect for the line system. From the plots in Figures 5-9(a) and 5-9(d), we see that the autocorrelations of the MA(m_0) models for the 2-dimension regular grid system cut-off and follow exactly what have been derived in Section 5.2.2.

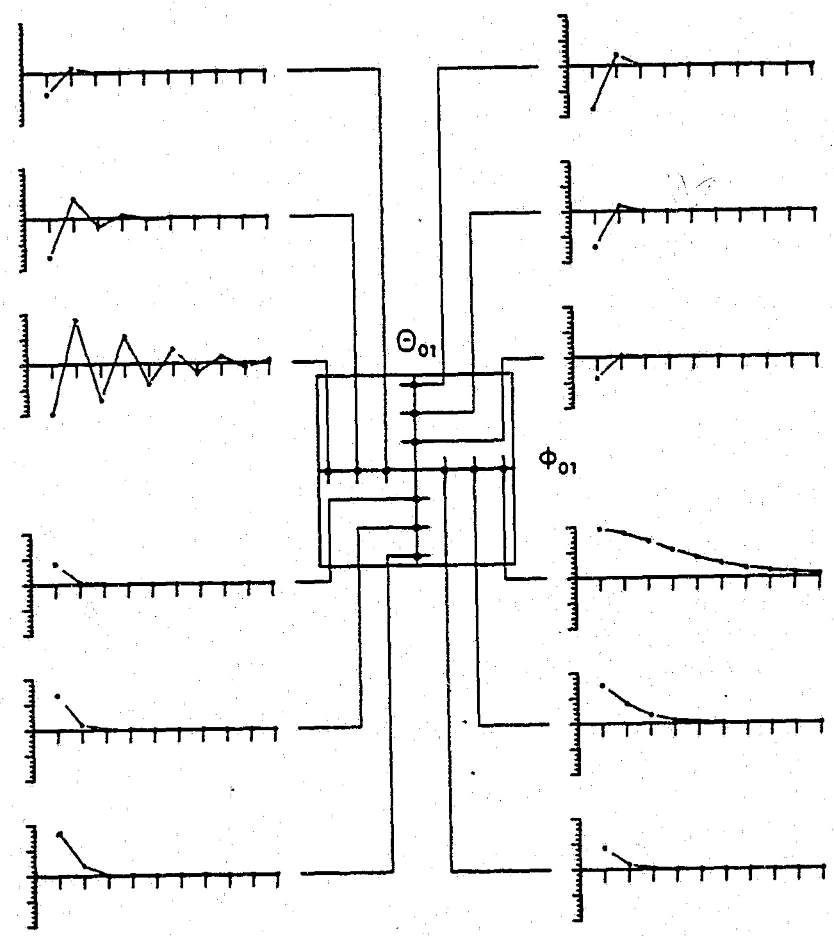


Figure 5-8(a). The Expected Sample Autocorrelation Functions $E(\hat{\rho}_k)$, $k=1.2,\ldots,10$ for the Purely Spatial AR(1) and MA(1) Models of the 1X25 Line System.

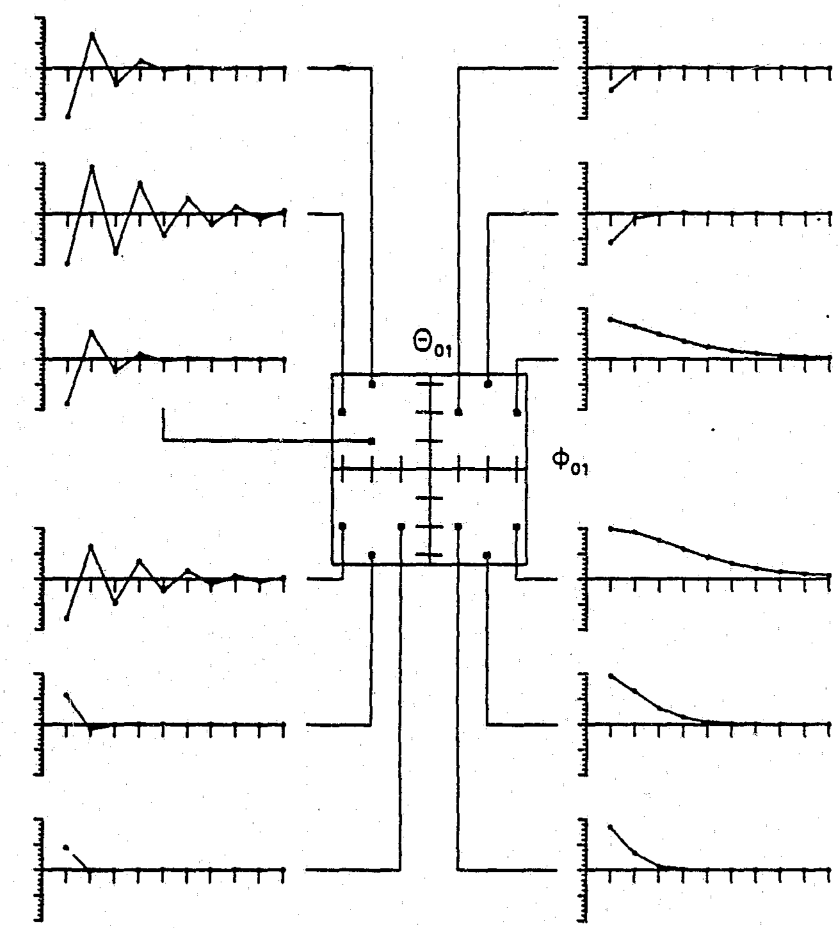


Figure 5-8(b) The Expected Sample Autocorrelation Functions $E(\hat{\rho}_k)$, $k=1,2,\ldots,10$ for the Purely Spatial ARMA(1,1) Models of the 1X25 Line System.

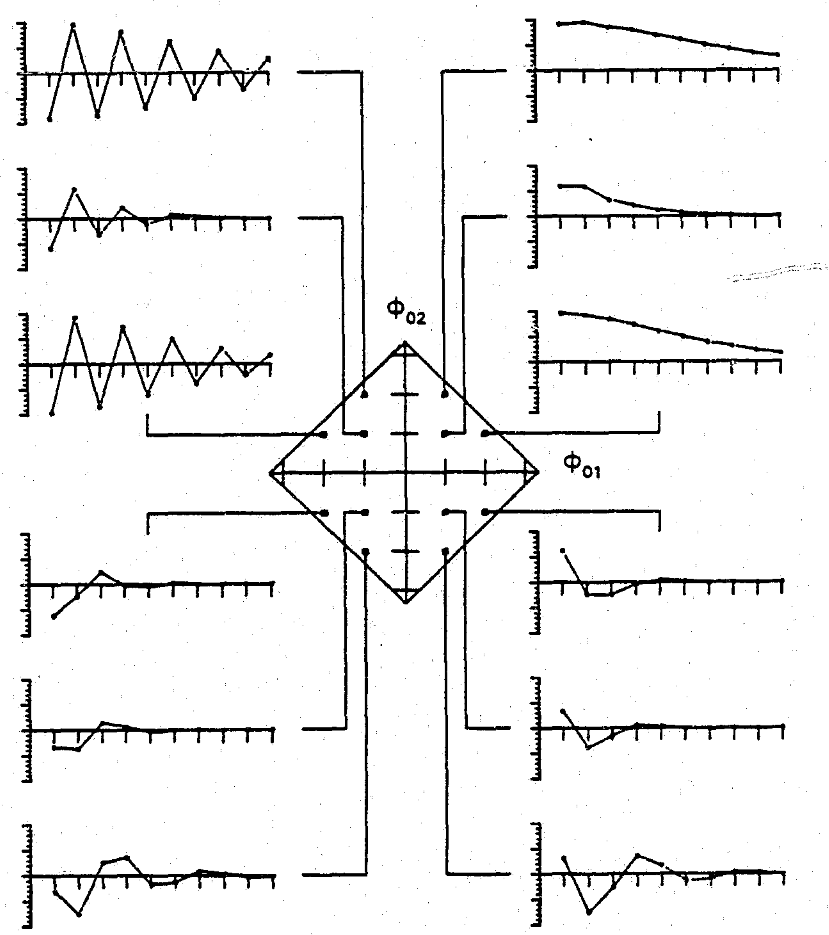


Figure 5-8(c). The Expected Sample Autocorrelation Functions E($\hat{\rho}_k$), k=1,2,...,10 for the Purely Spatial AR(2) Models of the 1X25 Line System.

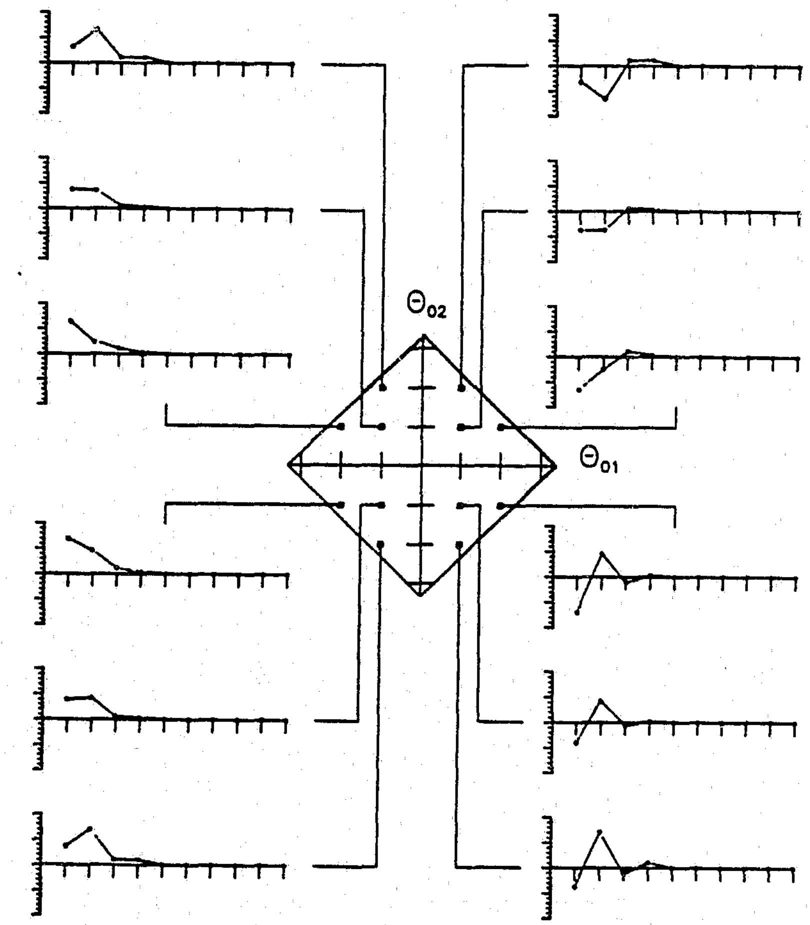


Figure 5-8(d). The Expected Sample Autocorrelation Functions E($\hat{\rho}_k$), k=1,2,...,10 for the Purely Spatial MA(2) Models of the 1X25 Line System.

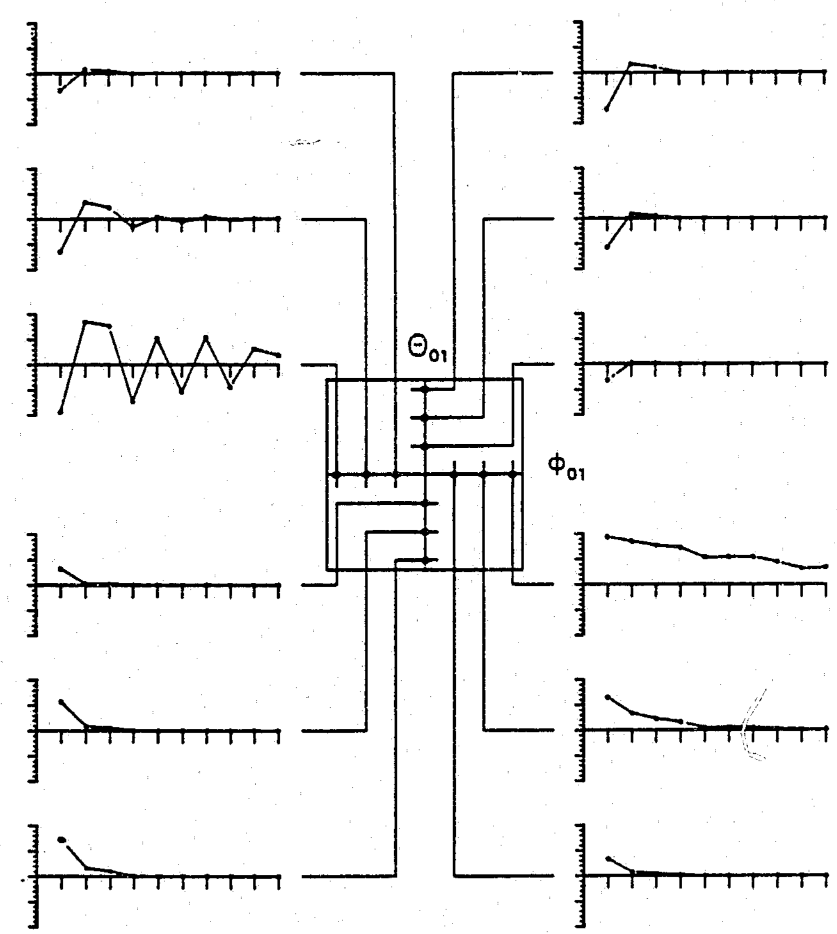


Figure 5-9(a). The Expected Sample Autocorrelation Functions $E(\hat{\rho}_k)$, $k=1,2,\ldots,10$ for the Purely Spatial AR(1) and MA(1) Models of the 5X5 Square REgular Grid System.

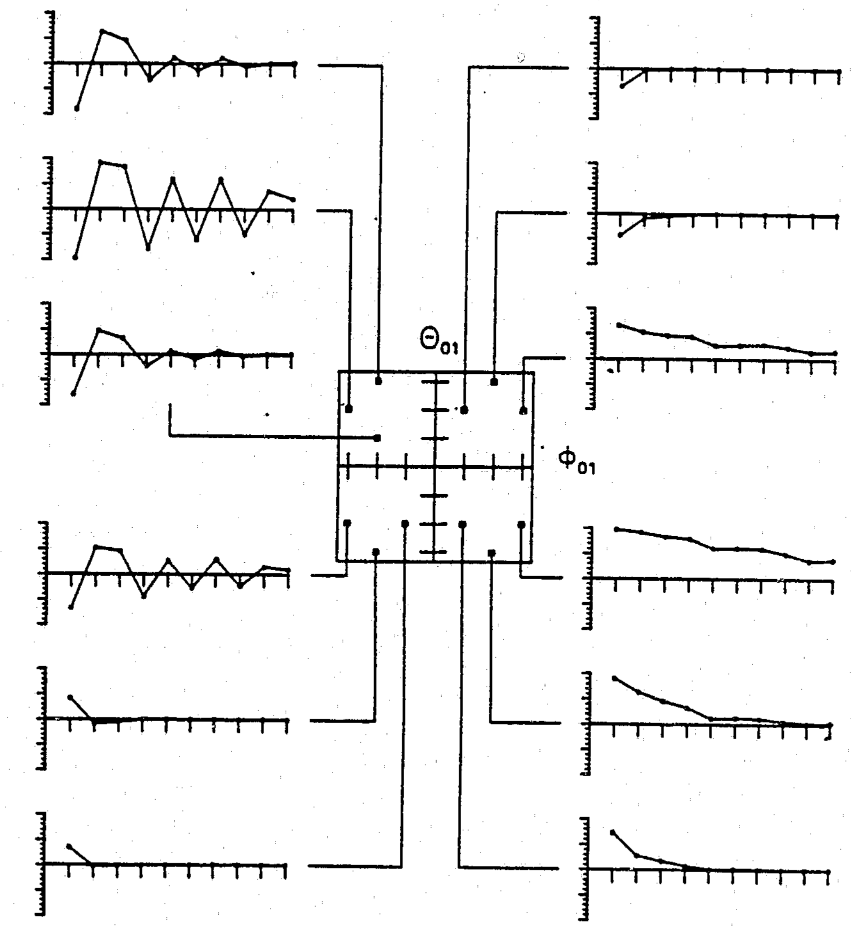


Figure 5-9(b) The Expected Sample Autocorrelation Functions $E(\hat{\rho}_k)$, $k=1,2,\ldots,10$ for the Purely Spatial ARMA(1,1) Models of the 5X5 Square Regular Grid System.

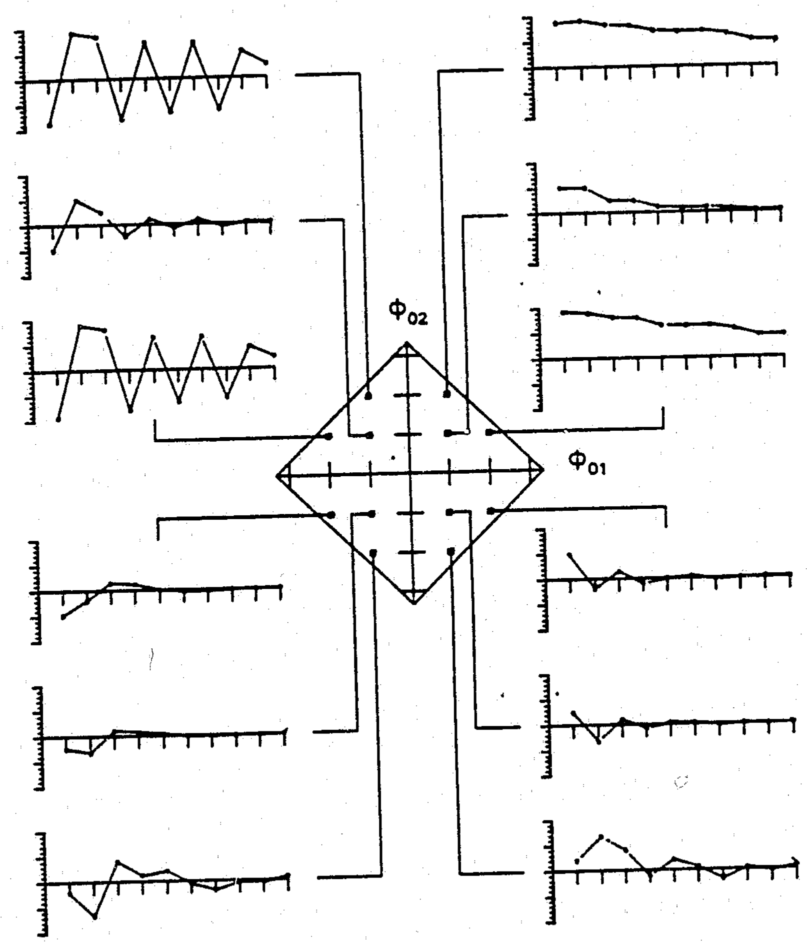


Figure 5-9(c). The Expected Sample Autocorrelation Functions $E(\hat{\rho}_k)$, $k=1,2,\ldots,10$ for the Purely Spatial AR(2) Models of the 5X5 Square Regular Grid System.

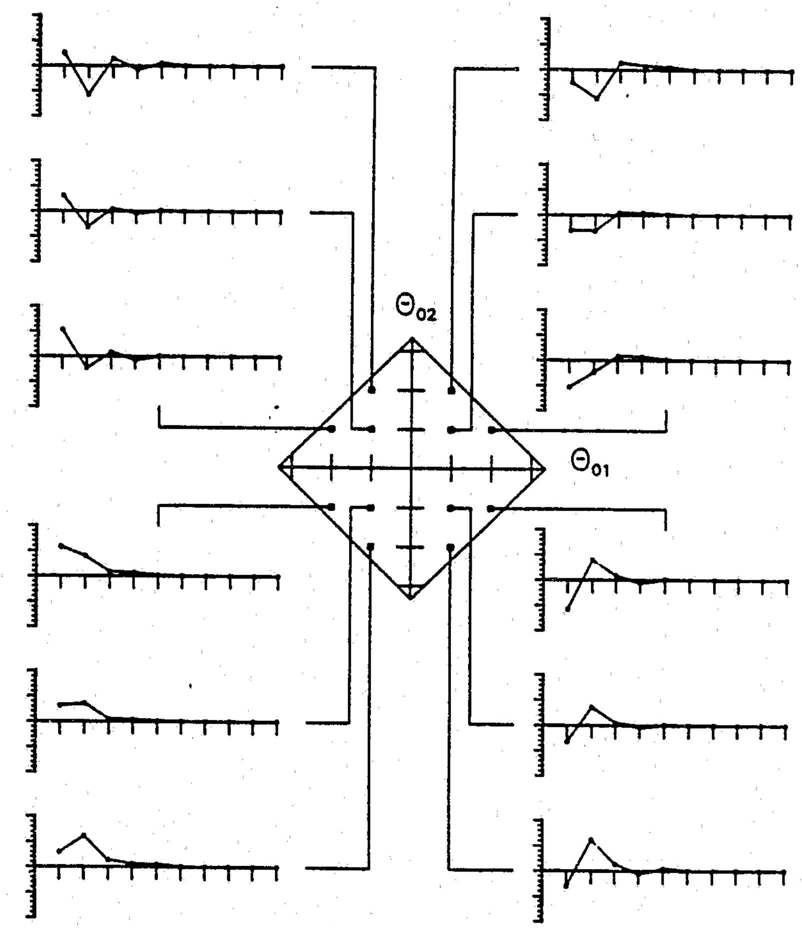


Figure 5-9(d). The Expected Sample Autocorrelation Functions $E(\hat{\rho}_k)$, $k=1,2,\ldots,10$ for the Purely Spatial MA(2) Models of the 5X5 Square Regular Grid System.

5.3 The Initial Estimates of the Low Order Purely Spatial Model Parameters

If the candidate model selected is nonlinear in the parameters, initial estimates of the model parameters are needed to initiate any estimation procedure. A set of good initial estimates will make their computational effort reasonably efficient.

5.3.1 The Initial Estimates of the MA(2) Model

Substituting B(0) and A(0) in Equation (5-27) with 0 and $\sum_{\ell=1}^{2} \theta_{0\ell} W^{(\ell)}$, respectively, we get the M_{ks} matrix for the MA(2) model as,

$$M_{ks} = (I - \sum_{\ell=1}^{2} \theta_{0\ell} W^{(\ell)})' W^{(k)'} W^{(s)} (I - \sum_{\ell=1}^{2} \theta_{0\ell} W^{(\ell)}).$$

Equation (5-28) is then applied to compute the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the purely spatial MA(2) model. In the Figure 5-10(a), the contours of the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ are plotted in every 0.1 interval for the 5×5 regular grid system. The surfaces of $(E(\hat{\rho}_1), \phi_{01}, \phi_{02})$ and $(E(\hat{\rho}_2), \phi_{01}, \phi_{02})$ of the purely spatial MA(2) model are plotted in Figures 5-10 (b) and (c), respectively. THe levels of the square boundary edges, that are served as the reference level, are set to zero for these three dimension plots. To obtain the initial estimates for the purely spatial MA(2) model, we compute the purely spatial autocorrelation functions $\hat{\rho}_2$, s=1,2,... first, identify the MA(2) model as the candidate model by the pattern comparison, and then read the initial estimates from Figure 5-10(a). For example, we have the process that has been

identified to be MA(2) process with $\hat{\rho}_1 = 0.5$, $\hat{\rho}_2 = 0.4$, in Figure 5-10(a), the value $(\theta_{01}, \theta_{02}) = (0.5, -0.22)$ is read to be the initial estimate for $(\theta_{01}, \theta_{02})$.

5.3.2 The Initial Estimates of the ARMA(1,1) Model

Substituting B(0) and A(0) in Equation (5-27) with $\phi_{01} W^{(1)}$ and $\theta_{01} W^{(1)}$, respectively, we get the M_{ks} matrix for the ARMA(1,1) model as,

$$M_{ks} = [(I - \phi_{01} W^{(1)})^{-1} (I - \theta_{01} W^{(1)})]' W^{(k)'} W^{(s)}$$
$$[(I - \phi_{01} W^{(1)})^{-1} (I - \theta_{01} W^{(1)})].$$

Equation (5-28) is then applied to compute the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the purely spatial ARMA(1,1) model. In the Figure 5-11(a), the contours of the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the 5×5 regular grid system are plotted in every 0.1 interval. The surfaces of $(E(\hat{\rho}_1), \phi_{01}, \theta_{01})$, $(E(\hat{\rho}_2), \phi_{01}, \theta_{01})$ are plotted in Figure 5-11 (b) and (c) respectively. To obtain the initial estimates for the purely spatial ARMA(1,1) model, we locate $(\hat{\rho}_{01}, \hat{\rho}_{02})$ position in Figure 5-11(a) to read the initial estimate of (ϕ_{01}, θ_{01}) from the coordinates, where $\hat{\rho}_{01}$, $\hat{\rho}_{02}$ are computed in the identification stage. For example, we have a process that has been identified to be the ARMA(1,1) process with $\hat{\rho}_1 = -0.5$, $\hat{\rho}_2 = 0.4$, in Figure 5-11(a), the value $(\phi_{01}, \theta_{01}) = (-0.9, -0.7)$ is read to be the initial estimate.

5.3.3 The Initial Estimates of the AR(2) Models

Substituting B(0) and A(0) in Equation (5-27) with $\sum_{\ell=1}^{2} \phi_{0\ell} W^{(\ell)}$

and 0, respectively, we have the M_{ks} matrix for the AR(2) model as,

$$M_{ks} = (I - \sum_{\ell=1}^{2} \phi_{0\ell} W^{(\ell)})^{-1} W^{(k)} W^{(s)} (I - \sum_{\ell=1}^{2} \phi_{0\ell} W^{(\ell)}).$$

Equation (5-28) is then applied to compute the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the purely spatial AR(2) model. In the Figure 5-12(a), the contours of the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the 5×5 regular grid system are plotted in every 0.1 interval. The surfaces of $(E(\hat{\rho}_1), \phi_{01}, \phi_{02})$, $E(\hat{\rho}_2), \phi_{01}, \phi_{02})$ are plotted in Figure 5-12 (b) and (c), respectively. To obtain the initial estimates for the purely spatial AR(2) model, we locate $(\hat{\rho}_{01}, \hat{\rho}_{02})$ position in Figure 5-12(a) to read the initial estimate of (ϕ_{01}, ϕ_{02}) from the coordinates. For example, given that $(\hat{\rho}_1, \hat{\rho}_2) = (-0.5, 0.4)$, in Figure 5-12(a), the value $(\phi_{01}, \phi_{02}) = (-0.35, 0.20)$ is read to be the initial estimate.

In previous sections the contour charts have been given with the three dimension plots of the $E(\hat{\rho}_1)$ surfaces and $E(\hat{\rho}_2)$ surfaces. Comparing the surfaces of $(E(\hat{\rho}_1), \theta_{01}, \theta_{02})$, $(E(\hat{\rho}_1), \phi_{01}, \theta_{01})$, $(E(\hat{\rho}_1), \phi_{01}, \phi_{02})$, we see that these surfaces share the same property that values of $E(\hat{\rho}_1)$ go smoothly from very positive (+1) to very negative (-1). However, the surfaces of $E(\hat{\rho}_2)$ do not share this property. The surfaces of $(E(\hat{\rho}_2), \theta_{01}, \theta_{02})$ and $(E(\hat{\rho}_2), \phi_{01}, \phi_{02})$ have the $E(\hat{\rho}_2)$ values go smoothly from very positive to very negative, while the surface of $(E(\hat{\rho}_2), \phi_{01}, \theta_{01})$ is symmetric about $\phi_{01} = \theta_{01}$ axis, it goes down and up again along any axis that is perpendicular to the $\phi_{01} = \theta_{01}$ axis. The similarity of $E(\hat{\rho}_1)$, $E(\hat{\rho}_2)$ surfaces of

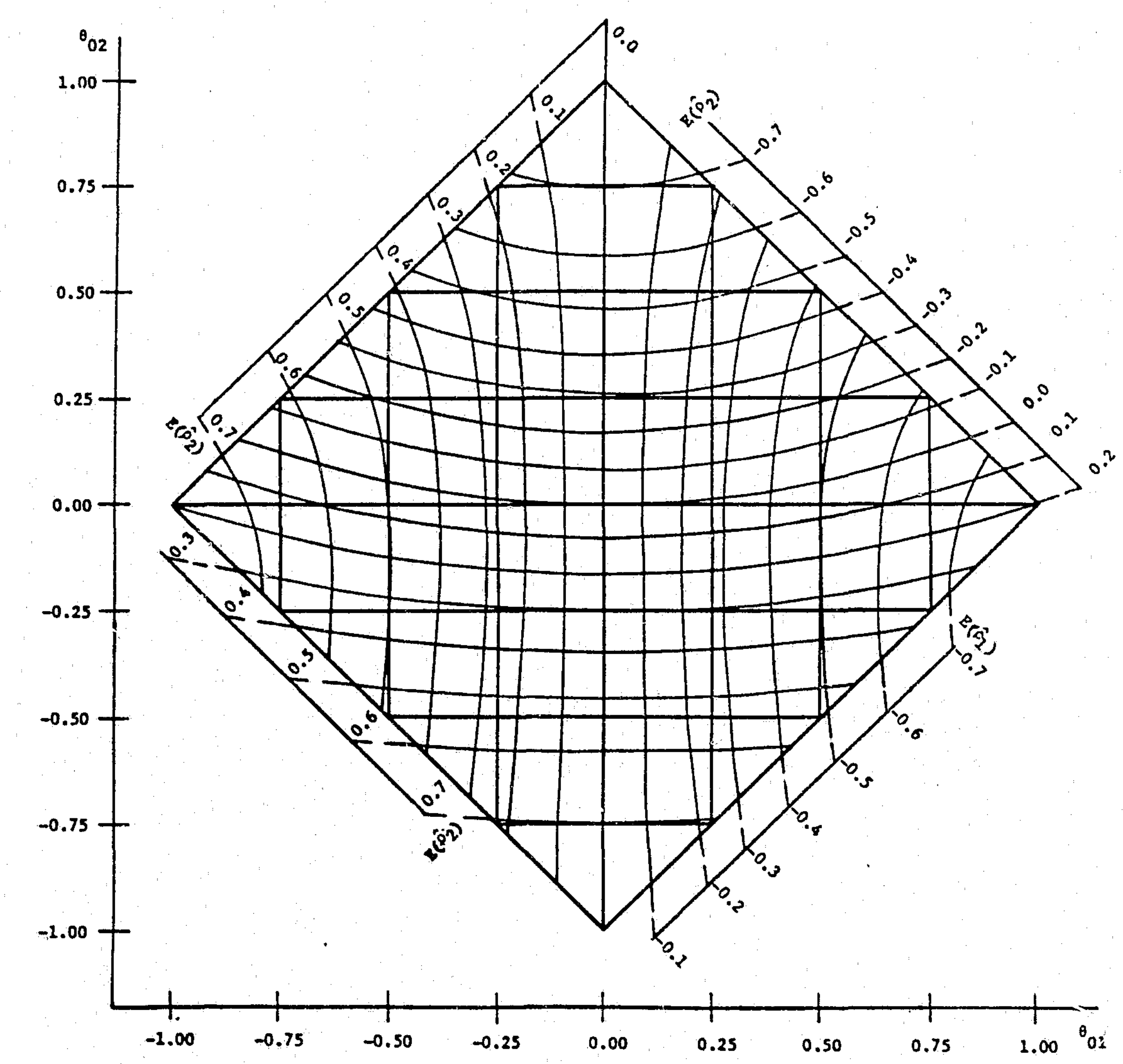


Figure 5-10(a). The Contours of the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the Purely Spatial MA(2) Models of the 5X5 Square Regular Grid System.

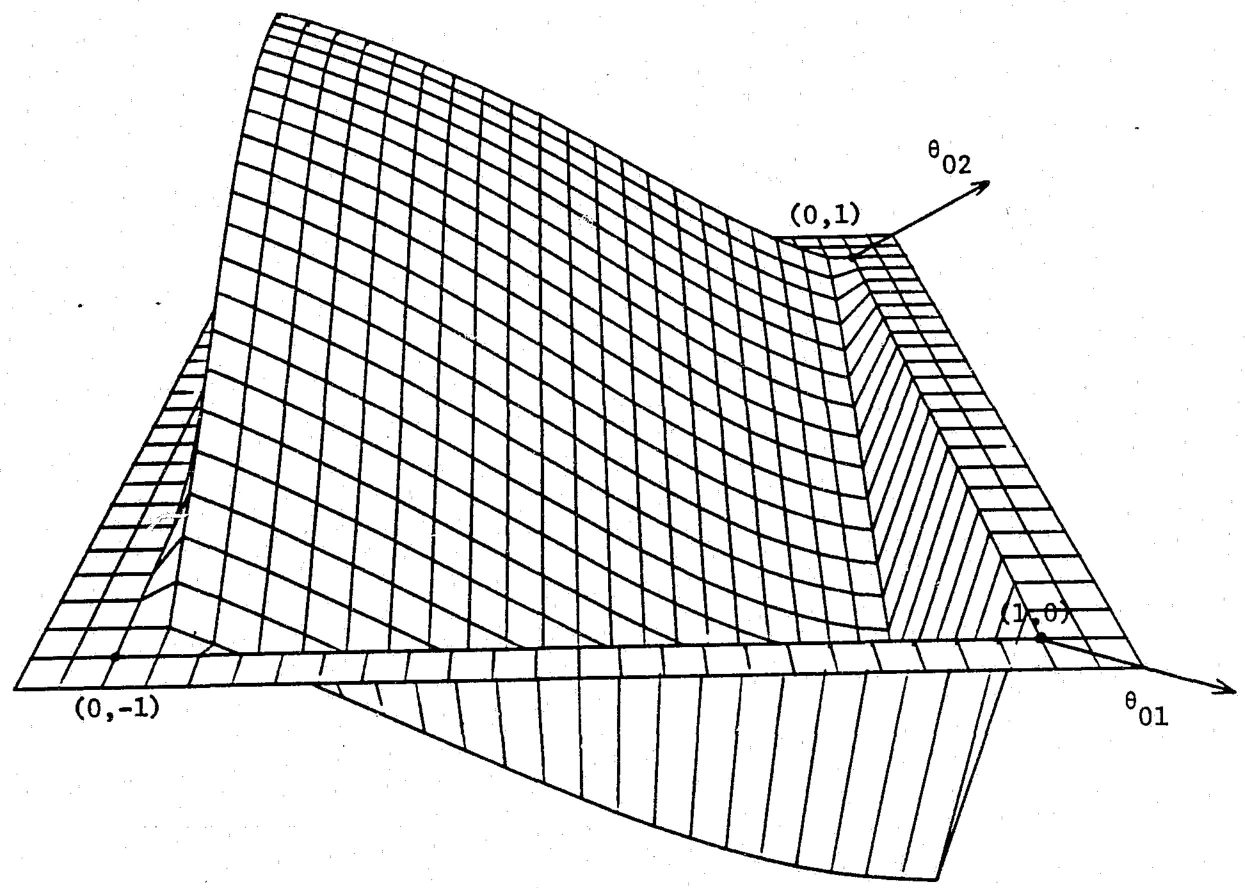


Figure 5-10(b). The $(E(\hat{\rho}_1), \theta_{01}, \theta_{02})$ Surface of the Purely Spatial MA(2) Model of the 5X5 Square Regular Grid System.

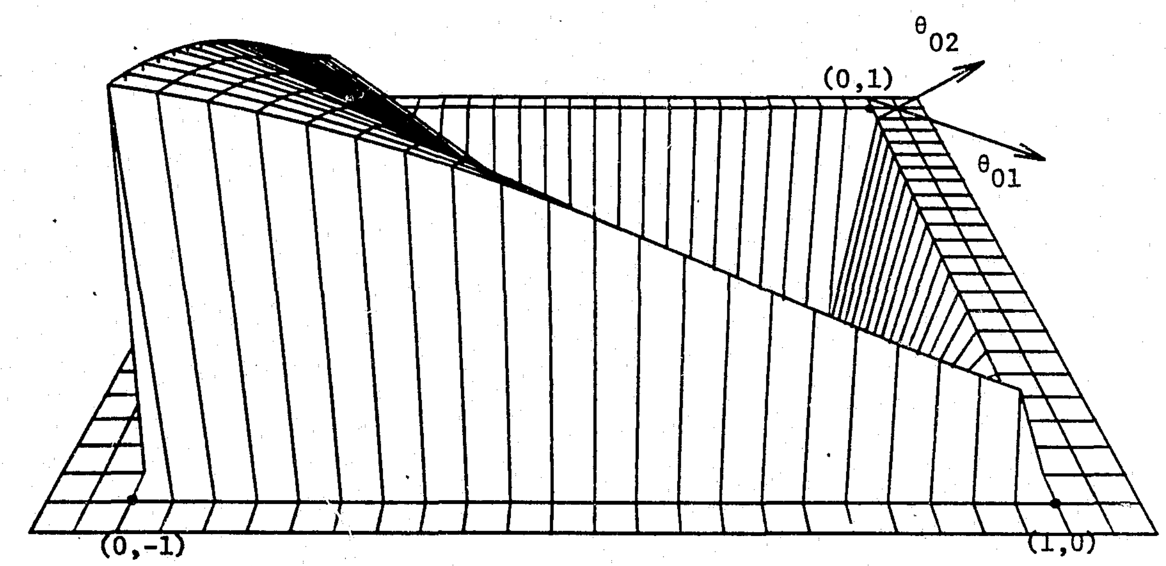


Figure 5-10(c). The $(E(\hat{\rho}_2), \theta_{01}, \theta_{02})$ Surface of the Purely Spatial MA(2) Model of the 5X5 Square Regular Grid System.

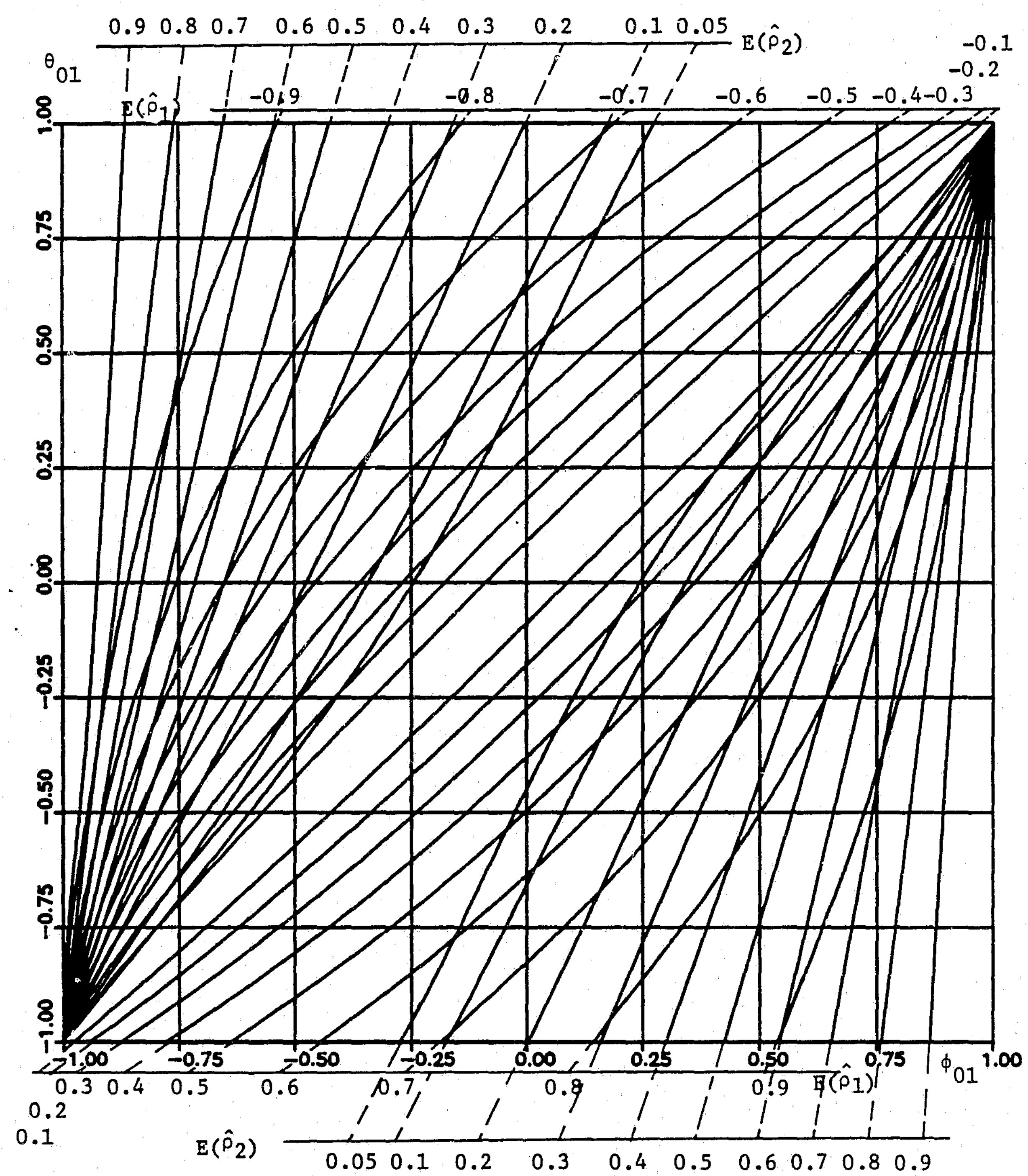


Figure 5-11(a). The Contours of the $E(\hat{\rho}1)$ and $E(\hat{\rho}2)$ for the Purely Spatial ARMA(1,1) Models of the 5X5 Square Regular Grid System.

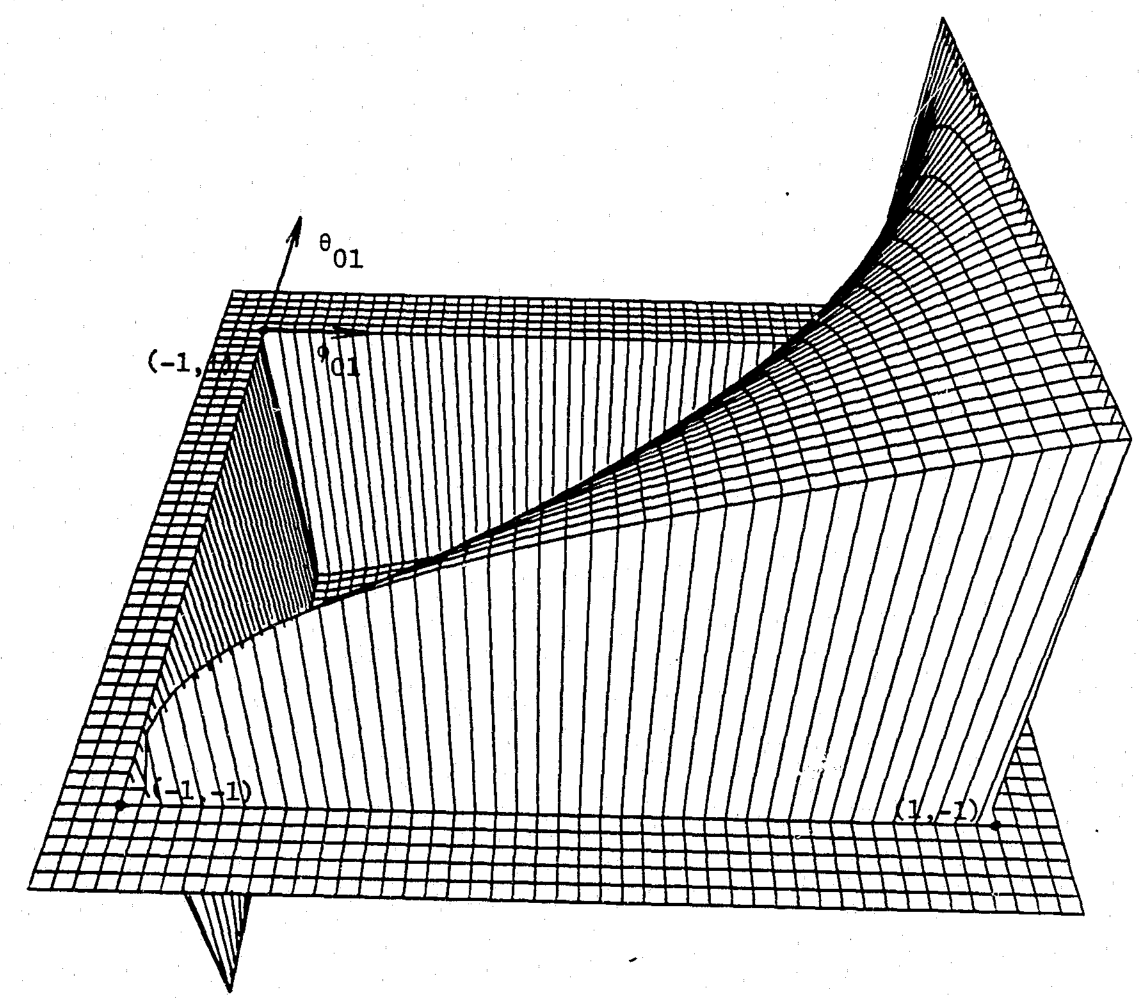


Figure 5-11(b). The $(E(\hat{\rho}1), \phi_{01}, \theta_{01})$ Surface of the Purely Spatial ARMA(1,1) Model of the 5X5 Square Regular Grid System.

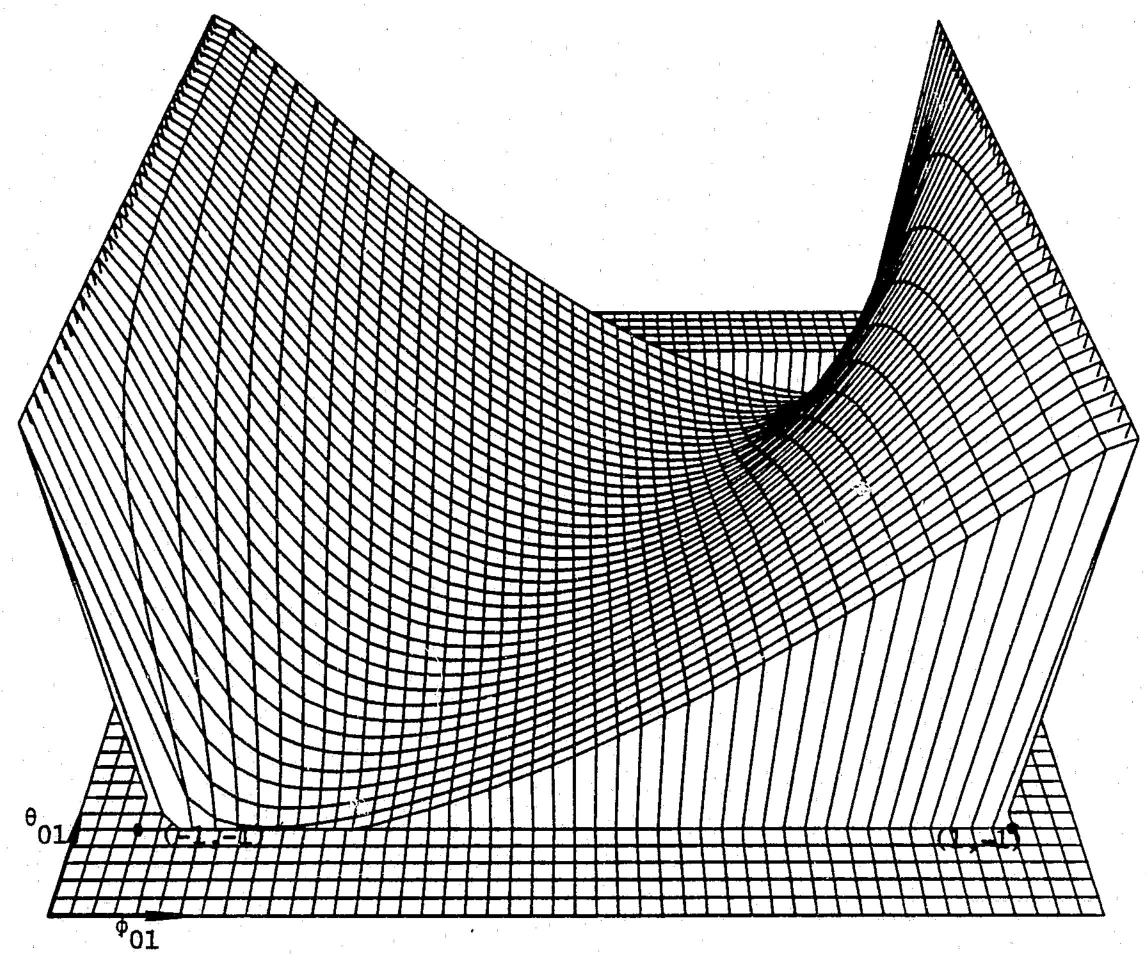


Figure 5-11(c). The $(E(\hat{\rho}2),\ \phi_{01},\ \theta_{01})$ Surface of the Purely Spatial ARMA(1,1) Models of the 5X5 Square Regular Grid System.

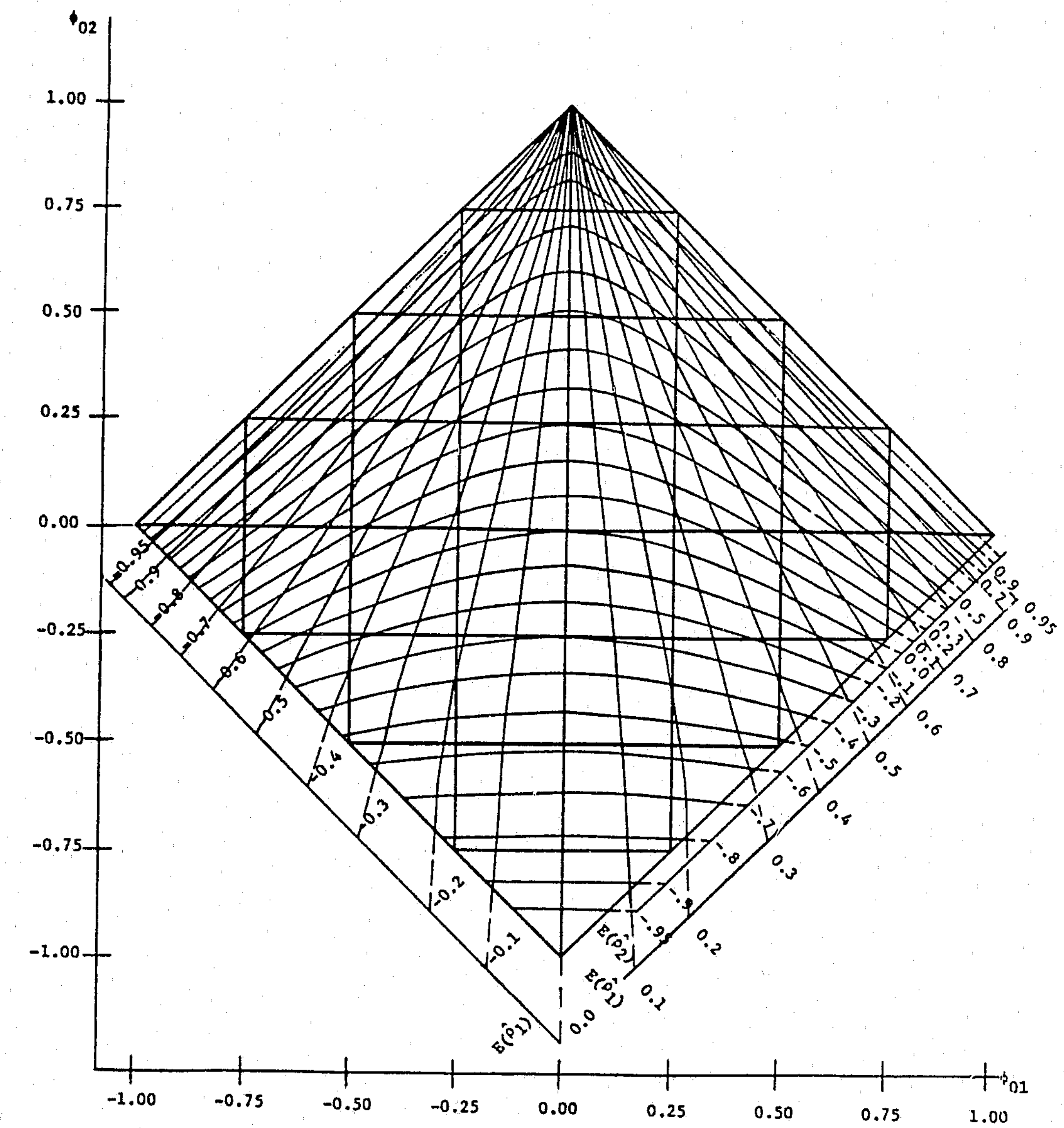


Figure 5-12(a). The Contours of the $E(\hat{\rho}_1)$ and $E(\hat{\rho}_2)$ for the Purely Spatial AR(2) Models of the 5X5 Square Regular Grid System.

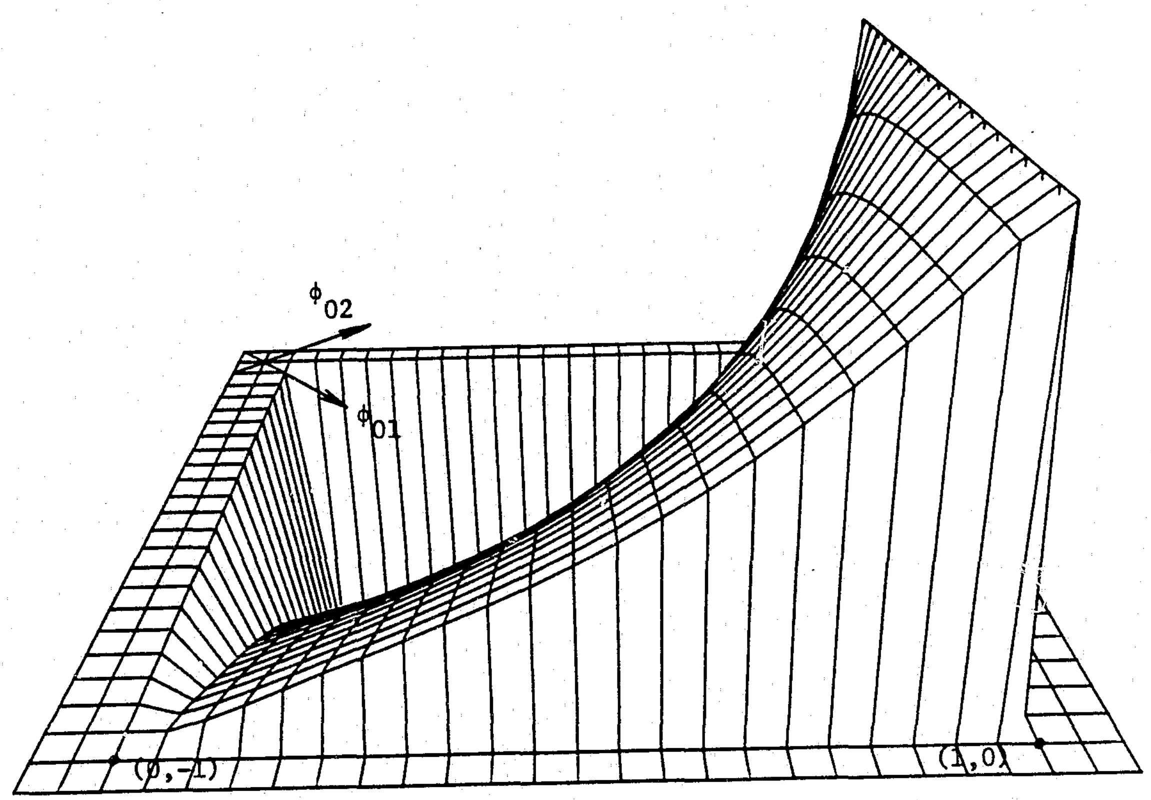


Figure 5-12(b). The $(E(\hat{\rho}_1), \phi_{01}, \phi_{02})$ Surface of the Purely Spatial AR(2) Models of the 5X5 Square Regular Grid System.

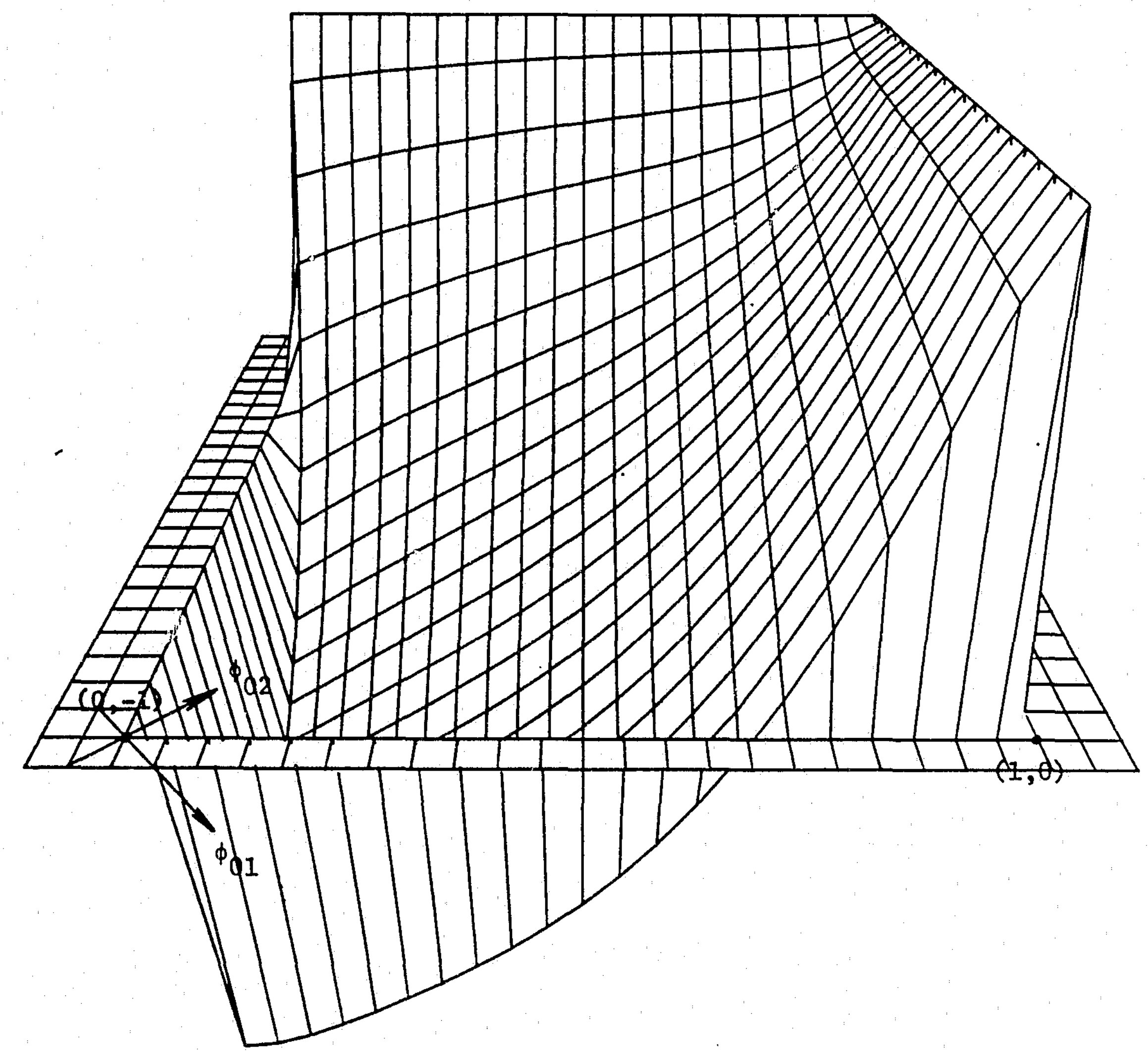


Figure 5-12(c). The $(E(\hat{\rho}_2), \phi_{01}, \phi_{02})$ Surface of the Purely Spatial AR(2) Models of the 5X5 Square Regular Grid System.

the purely spatial AR(2) model to the corresponding surfaces of the purely spatial MA(2) model reveals the fact that it is difficult to tell the AR(2) candidate model from the MA(2) candidate model if we should select the candidate model base on $\hat{\rho}_1$ and $\hat{\rho}_2$ only. And it is not so difficult to tell the ARMA(1,1) candidate model from the AR(2) candidate model or the MA(2) candidate model base on the information contained in $\hat{\rho}_1$ and $\hat{\rho}_2$.

It should be noted that for practical purposes, the plots of the 5×5 regular grid system that are applied to help in the pattern recognition and initial estimation are capable of representing most of the neighbor structures. This is especially appropriate when the location number exceeds 24, since the boundary effect is negligible. See Deutsch and Pfeifer [1980e].

5.4 The Estimation Procedures of the Purely Spatial Models

In this section, we will develop the procedures to get M.L. and L.S. estimates for the parameters of the candidate spatial model. The likelihood function for the purely spatial process is developed in Section 5.4.1 and the estimation procedures for AR(λ_0), MA(m_0) and ARMA(λ_0,m_0) under the assumption of $G = \sigma^2 I$ will be given in Sections 5.4.2, 5.4.3 and 5.4.4, respectively. In Section 5.4.5 joint confidence regions for the model parameters are developed. The estimation procedures under the general G assumption are covered in Section 5.4.6 and in Section 5.4.7 the estimation procedures when G is unknown are discussed.

5.4.1 M.L. Estimation

Given the purely spatial ARMA(λ_0,m_0) model,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T.$$

where

$$\sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \quad \text{and} \quad \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \tag{5-31}$$

are power convergent and

$$\underset{\sim}{\varepsilon}_t \sim \text{NID}(0,G),$$

the distribution function for $\underset{\sim}{\varepsilon}_t$, t=1,2,...,T is

$$f(\underset{\sim}{\varepsilon}_1,\underset{\sim}{\varepsilon}_2),\ldots,\underset{\sim}{\varepsilon}_T|G) = (2TL)^{-LN\cdot T/2}|G|^{-T/2}\exp\{-1/2 \sum_{t=1}^{T} \underset{\sim}{\varepsilon}_t' G^{-1} \underset{\sim}{\varepsilon}_t\} \tag{5-32}$$

and $\underset{\sim}{Z}_t = M\underset{\sim}{\varepsilon}_t$ with

$$M = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)}). \tag{5-33}$$

We have the distribution for $\underset{\sim}{Z}_t$, t=1,2,...,T as

$$f(\underset{\sim}{Z}_1,\underset{\sim}{Z}_2,\ldots,\underset{\sim}{Z}_T|\underset{\sim}{\phi},\underset{\sim}{\theta},G)$$

$$= (2\pi)^{-LN\cdot T/2}|G|^{-T/2}|M|^{-T}\exp\{-1/2\sum_{t=1}^{T}\underset{\sim}{Z}_t'(M^{-1})'G^{-1}M^{-1}\underset{\sim}{Z}_t\}.$$

Then the likelihood function for given observations $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_T$ is

$$L(\underset{\sim}{\phi},\underset{\sim}{\theta}|\underset{\sim}{Z}_1,\underset{\sim}{Z}_2,\ldots,\underset{\sim}{Z}_T,G) = (2\pi)^{-LN\cdot T/2}|G|^{-T/2}|M|^{-T}\exp\{-s(\underset{\sim}{\phi},\underset{\sim}{\theta})\} \quad (5\text{-}34)$$

where

$$S(\underset{\sim}{\phi},\underset{\sim}{\theta}) = \frac{1}{2}\sum_{t=1}^{T}\underset{\sim}{Z}_t(M^{-1})'G^{-1}M^{-1}\underset{\sim}{Z}_t$$

$$= \frac{T}{2}\,Tr(G^{-1}\hat{G}(\underset{\sim}{\phi},\underset{\sim}{\theta})),$$

$$\hat{G}(\underset{\sim}{\phi},\underset{\sim}{\theta}) = \frac{1}{T}\sum_{t=1}^{T}[M^{-1}\underset{\sim}{Z}_t][M^{-1}\underset{\sim}{Z}_t]'.$$

Taking the natural log and dropping the constant term involving 2π gives the log likelihood to be

$$\ell n\, L(\underset{\sim}{\phi},\underset{\sim}{\theta}|\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_T) \propto -\frac{T}{2}\,\ell n|G| - T\ell n|M| - S(\underset{\sim}{\phi},\underset{\sim}{\theta}). \quad (5\text{-}35)$$

In Equation (5-35), the first term, $-\frac{T}{2}\,\ell n|G|$, is not a function of

model parameter, and is not important in the searching procedure if G is known. Thus to obtain M.L. estimates we need to maximize

$$-\,T\,\ell n|M| - S(\underset{\sim}{\phi},\underset{\sim}{\theta}).$$

Given the true model parameters $(\underset{\sim}{\phi},\underset{\sim}{\theta})$ and $(\underset{\sim}{Z}_1',\ldots,\underset{\sim}{Z}_T')'$, $\ell n|M|$ is constant. Since $|T\ell n|M|| = T\cdot|\ell n|M||$, so when T increases $|T\ell n|M||$ will increase according to the same ratio. Also, the third term, $S(\underset{\sim}{\phi},\underset{\sim}{\theta}) = LN\cdot T\cdot\sigma^2$, will increase according to the same ratio. For moderate size systems and moderate T, typically $LN\cdot\sigma^2 > |\ell n|M||$ and $T\cdot LN\cdot\sigma^2 \gg T\cdot|\ell n|M||$, thus typically $S(\underset{\sim}{\phi},\underset{\sim}{\theta})$ will dominate the second term $T\cdot\ell n|M|$, and the M.L. estimates will be approximately equal to the L.S. estimates.

5.4.2 The Estimation Procedures of the Purely Spatial AR(λ_0) Processes

To obtain the log likelihood equation for the AR(λ_0) model, we set $\theta_{0\ell} = 0$ for $\ell=1,2,\ldots,m_0$ in Equation (5-35). This yields,

$$\ell n\, L(\underset{\sim}{\phi}|\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_T) \propto -(\frac{T}{2})\ell n|G| - T\ell n|M| - S(\underset{\sim}{\phi})$$

where

$$S(\underset{\sim}{\phi}) = \frac{1}{2}\sum_{t=1}^{T}\underset{\sim}{Z}_t'(I - \sum_{\ell=1}^{\lambda_0}\phi_{0\ell}W^{(\ell)})'(I - \sum_{\ell=1}^{\lambda_0}\phi_{0\ell}W^{(\ell)})\underset{\sim}{Z}_t, \text{ and}$$

$$M = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})^{-1} \qquad (5\text{-}36)$$

To get the M.L. estimator for $\underset{\sim}{\phi}$'s in AR(λ_0) model, we take the derivative of this log likelihood, set it equal to zero and solve for $\underset{\sim}{\phi}$. Thus,

$$-T \frac{\partial}{\partial \underset{\sim}{\phi}} \ln|M| - \frac{\partial}{\partial \underset{\sim}{\phi}} S(\underset{\sim}{\phi}) = 0 \qquad (5\text{-}37)$$

where

$$-\frac{\partial}{\partial \underset{\sim}{\phi}} \ln|M| = \left[\frac{\partial}{\partial \phi_i} \ln|I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}|\right] \qquad i=1,2,\ldots,\lambda_0$$

$$-\frac{\partial}{\partial \phi_i} \ln|M| = |I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}|^{-1} \frac{\partial}{\partial \phi_i} |I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}| \qquad (5\text{-}38)$$

$$= |I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}|^{-1}\, t_r[M^c W'^{(i)}], \text{ and}$$

$$\frac{\partial}{\partial \underset{\sim}{\phi}} S(\underset{\sim}{\phi}) = \left[\frac{\partial S(\underset{\sim}{\phi})}{\partial \phi_i}\right] \qquad i=1,2,\ldots,\lambda_0$$

$$\frac{\partial S(\underset{\sim}{\phi})}{\partial \phi_i} = -\frac{1}{2} \sum_{t=1}^{T} \underset{\sim}{Z}_t'(t)[W^{(i)} + W'^{(i)} - W'^{(i)} \sum_{\ell \neq i}^{\lambda_0} \phi_{0\ell} W^{(\ell)}$$

$$- \sum_{\ell \neq i}^{\lambda_0} \phi_{0\ell} W'^{(\ell)} W^{(i)} - 2\phi_{0i} W'^{(i)} W^{(i)}] .$$

In the above formulation M^c is the matrix of co-factors of the matrix M.

To solve these equations is not an easy task since the term $|I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}|$ in Equation (5-38) contains the terms of $\underset{\sim}{\phi}$'s up to approximately the LN power. If we drop the first term in Equation (5-38) and maximize $- S(\underset{\sim}{\phi})$, we get the L.S. estimates, which can be easily solved and expressed in closed form,

$$\hat{\underset{\sim}{\phi}} = (X X)^{-1} X \underset{\sim}{Y} \qquad (5\text{-}39)$$

where

$$X = [\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_{\lambda 0}]$$

$$\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{Z}(1) \\ \underset{\sim}{Z}(2) \\ \vdots \\ \underset{\sim}{Z}(T) \end{bmatrix} \qquad \underset{\sim}{X}_i = \begin{bmatrix} W^{(i)}\underset{\sim}{Z}(1) \\ W^{(i)}\underset{\sim}{Z}(2) \\ \vdots \\ W^{(i)}\underset{\sim}{Z}(T) \end{bmatrix}$$

5.4.3 The Estimation Procedures for the Purely Spatial MA(m_0) Processes

To obtain the log likelihood for MA(m_0) model, we set $\phi_{0\ell} = 0$, $\ell=1,2,\ldots,\lambda_0$ in Equation (5-35). Thus,

$$\ln L(\underset{\sim}{\theta}|\underset{\sim}{Z}_1,\underset{\sim}{Z}_2,\ldots,\underset{\sim}{Z}_T) = -(\tfrac{T}{2}) \ln|G| + T\ln|M| - S(\underset{\sim}{\theta}),$$

where

$$S(\underset{\sim}{\theta}) = \frac{1}{2}\sum_{t=1}^{T} \underset{\sim}{Z}_t'[(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})^{-1}]'(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})^{-1}\underset{\sim}{Z}_t,$$

and

$$M = I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)}.$$

Because of the non-linearity in the parameters, we can't obtain a closed form for the M.L. estimators or L.S. estimators. Thus, search over the parameter space will be needed to get the M.L. estimates, which maximize the $\ln L(\underset{\sim}{\theta}|\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_T)$, or the L.S. estimates, which maximize the $S(\underset{\sim}{\theta})$.

5.4.4 The Estimation Procedures for the Purely Spatial ARMA(λ_0,m_0) Models

As in the case of the MA(m_0) process because of the non-linear nature in the $\underset{\sim}{\theta}$'s parameters, we can't obtain a closed form solution for the M.L. estimators or L.S. estimators for the ARMA(λ_0,m_0) process. However, we can reduce the searching effort to get the L.S. estimators by exploiting the linear nature in the $\underset{\sim}{\phi}$'s parameters. Given $\underset{\sim}{\theta}$'s, we can transform the model into the linear model form,

$$\underset{\sim}{Y}(t) = X(t)\underset{\sim}{\phi} + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T$$

where

$$\underset{\sim}{Y}(t) = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell})^{-1}\underset{\sim}{Z}_t$$

$$X(t) = [\underset{\sim}{X}_1(t),\underset{\sim}{X}_2(t),\ldots,\underset{\sim}{X}_{\lambda 0}(t)] \tag{5-40}$$

$$\underset{\sim}{X}_i(t) = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell})^{-1}W^{(i)}\underset{\sim}{Z}_t$$

The L.S. estimator under the assumption of $\underset{\sim}{\varepsilon}_t \sim NID(0,\sigma^2 I)$ is,

$$\hat{\underset{\sim}{\phi}} = (X'X)^{-1}X'\underset{\sim}{Y} \tag{5-41}$$

where

$$X = \begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(T) \end{bmatrix} \qquad \underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{Y}(1) \\ \underset{\sim}{Y}(2) \\ \vdots \\ \underset{\sim}{Y}(T) \end{bmatrix}$$

5.4.5 The Confidence Regions

In the previous sections the procedures to get the point estimates for the parameters of $AR(\lambda_0)$, $MA(m_0)$ and $ARMA(\lambda_0,m_0)$ models under the assumption of $G = \sigma^2 I$ have been detailed. In this section, we will find the variance-covariance matrix for the estimated parameters. This will be an approximate variance-covariance matrix and will be used to construct the approximate confidence region for both M.L. estimators and L.S. estimators.

According to the work done by Whittle [1953], the variance-covariance matrix of the parameter estimators $V(\hat{\beta})$, $\hat{\beta} = [\hat{\phi}', \hat{\theta}']$, is,

$$V_{ij}(\hat{\beta}) = \left\{-E\left[\left(\frac{\partial \ln L}{\partial \beta_i}\right) \cdot \left(\frac{\partial \ln L}{\partial \beta_j}\right)\right]\right\}^{-1} \tag{5-42}$$

or equivalently,

$$V_{ij}(\hat{\beta}) = \left\{-E\left[\frac{\partial^2 \ln L}{\partial \beta_i \, \partial \beta_j}\right]\right\}^{-1}, \tag{5-43}$$

where $V_{ij}(\hat{\beta})$ denotes the (i,j) element of $V(\hat{\beta})$. By substituting the log likelihood expression in the Equation (5-42), we have

$$V_{ij}(\hat{\beta}) = \left\{-E\left[\frac{\partial(T \ln|M| + S(\beta))}{\partial \beta_i} \cdot \frac{\partial(T \ln|M| + S(\beta))}{\partial \beta_j}\right]\right\}^{-1}, \tag{5-44}$$

or equivalently

$$V_{ij}(\hat{\beta}) = \left\{E\left[\frac{\partial^2(T \ln|M| + S(\beta))}{\partial \beta_i \partial \beta_j}\right]\right\}^{-1} \tag{5-45}$$

For moderate T and LN, $S(\beta)$ dominates $T \cdot \ln|M|$ in most cases, the $V_{ij}(\beta)$ is relatively constant. Thus

$$V_{ij}(\hat{\beta}) \doteq \left\{E\left[\frac{\partial^2 S(\beta)}{\partial \beta_i \partial \beta_j}\right]\right\}^{-1}. \tag{5-46}$$

A further approximation can be made by using the actually observed sum of squares as $S(\beta)$ in place of the expectation operator. Thus

$$V_{ij}(\hat{\beta}) \doteq \left\{\frac{\partial^2 S(\beta)}{\partial \beta_i \partial \beta_j}\right\}^{-1}. \tag{5-47}$$

For the $AR(\lambda_0)$ model, $S(\beta)$ is quadratic in β over the relevant region and thus the second order derivative of $S(\beta)$ will be constant over this region, and the confidence regions can be directly approximated by $V_{ij}(\hat{\beta})$.

For the $MA(m_0)$ model and the $ARMA(\lambda_0,m_0)$ model, $S(\beta)$ is not strictly quadratic in β. Since for the $MA(m_0)$ and $ARMA(\lambda_0,m_0)$ models, $S(\beta)$ will be approximately quadratic over the relevant region of the parameter space, we will use the same approximation and assume that these derivations are best determined at or near the point $\hat{\beta}$.

To construct the joint confidence region, we use the results found in Drapper Smith [1966]. The 100(1-α)% confidence regions are bounded approximately by the value of $\beta' = (\phi', \theta')$ which solves

$$(\beta-\hat{\beta})' [V(\hat{\beta})]^{-1}(\beta-\hat{\beta}) = k_0 F_{1-\alpha,k_0,N-k_0}$$

where

$$k_0 = \lambda_0 + m_0$$

$$N = LN \cdot T \tag{5-48}$$

5.4.6 Estimation When G Is Known and $G \neq \sigma^2 I$

In previous estimation sections, we have assumed $G = \sigma^2 I$ for simplicity. Now we look at estimation under the more general G assumption. Consider the Spatial ARMA(λ_0,m_0) model,

$$Z_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} Z_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \varepsilon_t + \varepsilon_t$$

where

$$\varepsilon_t \sim NID(0,G), \quad G \neq \sigma^2 I \tag{5-49}$$

G is known and

$$E(\varepsilon_t \varepsilon'_{t+k}) = 0, \quad k \neq 0.$$

Since G is a positive definite matrix of rank LN, so there exists a matrix B of rank LN such that $BB' = G$ (Graybill [1976]). Let $C = B^{-1}$, and pre-multiply Equation (5-49) with C and defining $\varepsilon_C(t) = C\varepsilon_t$ to get

$$CZ_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} CW^{(\ell)} Z_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} CW^{(\ell)} Z_t + \varepsilon_C(t)$$

$$\varepsilon_C(t) \sim NID(0,I) \quad \text{and} \tag{5-50}$$

$$E(\varepsilon_C(t)\varepsilon_C(t+k)) = 0, \quad k \neq 0$$

This form allows the application of the same estimation procedures that has been developed for the $G = \sigma^2 I$. Further, to check the assumption that $\varepsilon_t \sim NID(0,G)$, the components of $\varepsilon_C(t)$ are checked to be independent with variance 1. Therefore all the estimation and diagnostic checking techniques appropriate under the assumption of $G = \sigma^2 I$ can be applied to the transformed model. Therefore in case that G is known and $G \neq \sigma^2 I$, we will transform the original data, and model the transformed data with the same estimation techniques, diagnostic checking that are appropriate for the $G = \sigma^2 I$ case.

5.4.7 Estimation When G is Unknown

Usually G is unknown, and it is natural to assume that $G = \sigma^2 I$ first and then check this assumption at diagnostic checking stage. So the estimation procedure under the $G = \sigma^2 I$ assumption is applied directly to get the parameter estimates $(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}})$ and the $G = \sigma^2 I$ assumption is tested upon the estimated residuals $\hat{\underset{\sim}{\varepsilon}}_t$, $t=1,2,\ldots,T$. Situations may arise when the assumption $G = \sigma^2 I$ is unacceptable, then the assumption G=D is tested, where D is diagonal matrix with all diagonal elements positive. The G=D assumption is equivalent to the assumption that the contemporaneous noise $\underset{\sim}{\varepsilon}_t$ are independent. If G=D hypothesis is not rejected, D is estimated by $\hat{D}$, which is constructed as follows,

$$\hat{D} = [\hat{d}_{ij}] \tag{5-51}$$

where $d_{ij} = \delta_{ij}\hat{G}_{ij}$, $\hat{G}_{ij}$ is the (i,j) element of $\hat{G}$, and

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases}$$

Then $\hat{D}$ is treated as the true D to estimate the model parameters $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$.

If G=D hypothesis is rejected again, it then is appropriate to assume the general G covariance and estimate G with $\hat{G}$, treat $\hat{G}$ as the true covariance to estimate the model parameters $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$.

The estimation procedure for the covariance matrix G unknown

situation is shown in Figure 5-13. This procedure assumes $G = \sigma^2 I$ in the very beginning and performs the estimation under the $G = \sigma^2 I$ assumption. If the residuals pass the $G = \sigma^2 I$ test, the procedure stops. The procedure enters the iterative subroutine if $G = \sigma^2 I$ can't pass. In the iterative subroutine, G=D assumption is tested in every iteration, the estimation procedure under G=D assumption is performed if G=D assumption is accepted, otherwise the estimation procedure under general G assumption is applied. This two-stage iterative procedure may be applied to get the L.S. estimates of $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$ as well as the M.L. estimates of $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$.

An alternative method to get the M.L. estimates can be obtained directly from Equation (5-35). Since G is unknown, the log likelihood function without the term involving 2π is now expressed as,

$$\ln L(\underset{\sim}{\phi}, \underset{\sim}{\theta}, G \mid Z_1, \ldots, \underset{\sim}{Z}_T)$$

$$\alpha - \frac{T}{2} \ln|G| - T\ln|M| - S(\underset{\sim}{\phi}, \underset{\sim}{\theta}, G) \tag{5-52}$$

Since $S(\underset{\sim}{\phi}, \underset{\sim}{\theta}, G) = LN \cdot T$ is constant when G is set to its M.L. estimates $\hat{G}$, so maximizing the log likelihood function is equivalent to minimizing

$$\ln|G| + 2 \ln|M|$$

A search routine is then needed to search through the existence regions of the $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$ space to get the M.L. estimates $(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}})$ such that

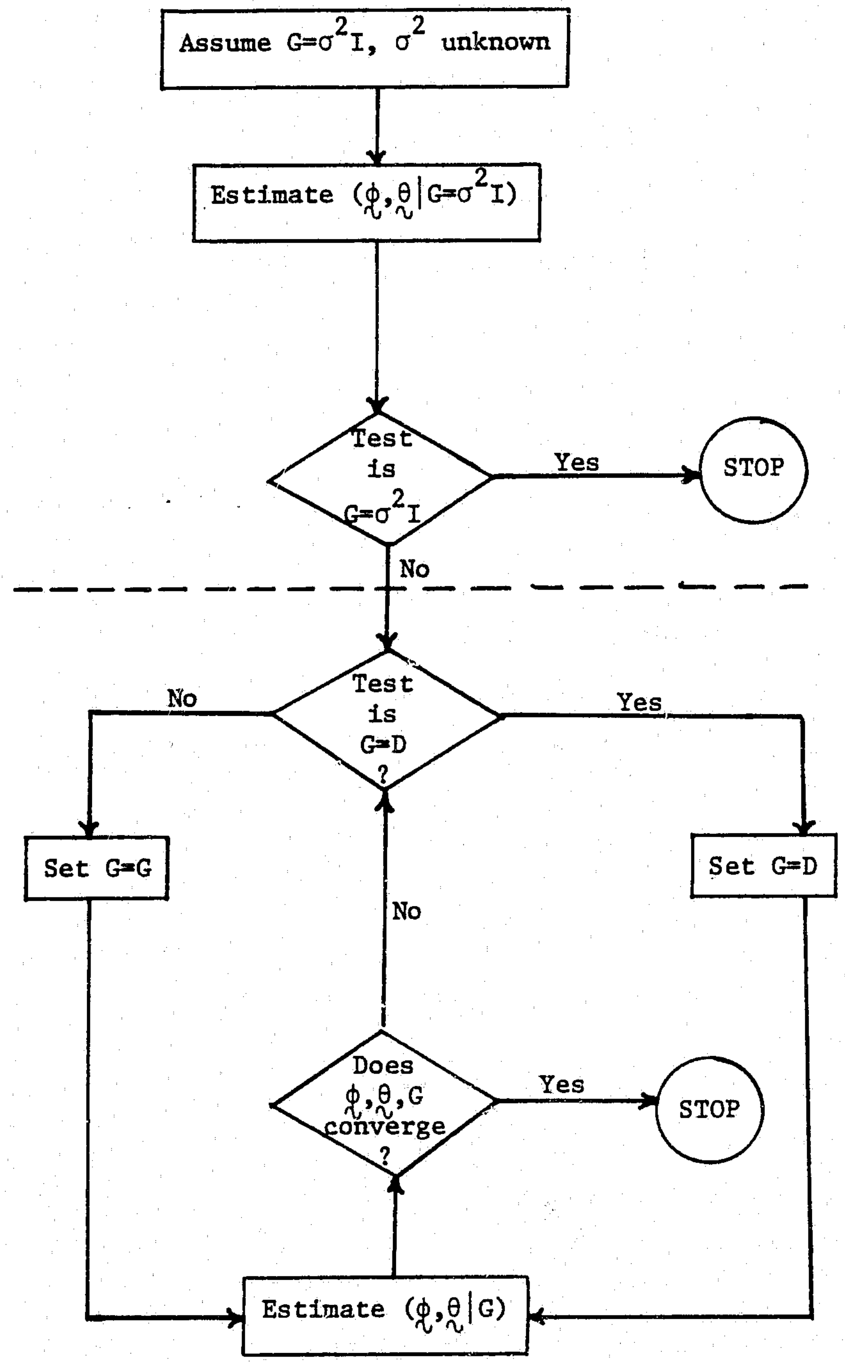


Figure 5-13. The Two-Stage Iteration Scheme for the General G Models (G Unknown)

$$\ln|\hat{G}(\hat{\phi},\hat{\theta})| + 2\ln|M(\hat{\phi},\hat{\theta})| \leq \ln|\hat{G}(\phi,\theta)| + 2\ln|M(\phi,\theta)|$$

and $\hat{G}(\hat{\phi},\hat{\theta})$ is the M.L. estimate of covariance G.

5.5 Diagnostic Checking

After a candidate model is selected and its parameters are estimated, two questions need to be addressed,

1. Do the residuals from the fitted model adhere to the assumptions concerning the properties of the purely spatial ARMA model? and
2. Is this model parsimonious? or equivalently, could the data be adequately represented by a simpler model?

In Section 5.5.1 we will discuss the first question, i.e., procedures to check the white noise assumption and contemporaneous correlation assumption for the unobservable residuals. In Section 5.5.2, we will discuss the second question, which is equivalent to check the significance of model parameters in the purely spatial ARMA(λ_0,m_0) model.

5.5.1 Diagnostic Checking That Is Applied to the Residuals

In order to check the adequacy of the usual assumptions of the residuals is equivalent to testing the following two hypothesis:

1. G is consistent with the assumption under which the estimation is performed.
2. ε_t are white noise.

To test the constency of the assumption about G, the two assumptions stated below may arise:

1. $G = \sigma^2 I$, and

2. $G = D$.

To test the hypothesis $G = \sigma^2 I$, we first construct the matrix $\hat{G} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$, where $\hat{\varepsilon}_t$ are estimated residuals, and then use the following procedure;

1. Calculate

$$W = \frac{|\hat{G}|}{\left[\frac{tr(\hat{G})}{LN}\right]^{LN}} \tag{5-53}$$

2. Find Z such that the probabilities on the righ-hand-side of the equation below is equal to $1-\alpha$.

$$Pr\{-(T-1)\rho \ \ell n\ W \leq Z\} = Pr\{\chi_f^2 \leq Z\} + W_2[Pr\{\chi_{f+4}^2 \leq Z\} - Pr\{\chi_f^2 \leq Z\}] + 0\ ((T-1))^{-3} \tag{5-54}$$

where

$$f = \frac{1}{2} LN\ (LN+1) - 1$$

$$\rho = 1 - \frac{2(LN)^2 + LN + 2}{6LN(T-1)}$$

$$W_2 = \frac{(LN+2)(LN-1)(LN-2)(2LN^3+6LN^2+3LN+2)}{288(LN)^2(T-1)^2\rho^2}$$

3. Reject the hypothesis $G = \sigma^2 I$ if $-(T-1)\rho\ \ell n\ W \geq Z$.

The table of Z values for chosen LN, T and α has been given by Deutsch and Pfeifer [1980]. Here Z values are computed by applying Equations (5-53) and (5-54). The LN, T and α have been tabled for combinations of the following values:

$$LN = 3, 5, 10, 20, 30, 50,$$
$$T = 10, 20, 30, 50, 75, 100, 200,$$
$$\alpha = 0.01, 0.05, 0.10.$$

Since the $Pr\{\chi_f^2 \leq Z\}$ term of Equation (5-54) dominates the other right-hand terms for large T, so the Z can be approximated by the z', a value easily determined from χ^2 tables such that $Pr\{x_f^2 \leq z'\} = 1-\alpha$.

To test the hypothesis G=D, we apply the following procedure by Deutsch and Pfeifer [1980e]:

1. Compute

$$V = \frac{|\hat{G}|}{\prod_{i=1}^{LN} \hat{G}_{ii}} \tag{5-55}$$

2. Find v such that

$$Pr\{\chi_f^2 \leq v\} + \frac{r}{m^2}[Pr\{\chi_{f+4}^2 \leq v\} - Pr\{\chi_f^2 \leq v\}] = 1-\alpha \tag{5-56}$$

where

$$f = LN(LN+1)/2$$
$$m = T - (2LN+11)/6$$
$$r = LN\cdot(LN-1)\cdot(2LN^2-2LN-13)/288$$

3. Reject the G=D hypothesis if $-m \ln V \geq v$ at the significant level α.

The v values computed from Equation (5-56) have been tabled by Deutsch and Pfeifer [1980] for combinations of those LN, T and α values that have been tabled for the $G = \sigma^2 I$ test.

In the context of the three stage iterative model building procedure, one of the best way for testing the statistical independence of random process is to check the calculated sample space-time autocorrelation function. If there is no structure detected in the calculated sample space-time autocorrelation functions, the residuals pass the white noise assumption and the fitted model is accepted as adequate. This has been developed by Deutsch and Pfeifer [1981d]. However, they didn't include the purely spatial autocorrelation functions within sample space-time autocorrelation functions, and this noise check didn't have the capability for checking the existence of contemporaneous dependence among residuals.

For the LN dimensional pure white noise process, the $\mathrm{var}(\hat{\rho}_s(k))$ for $k > 0$ is approximately equal to $[LN(T-k)]^{-1}$. From Equation (5-29) of Section 5.2.4, we have

$$\mathrm{var}(\hat{\rho}_s(0)) = \frac{W^{(S)}\cdot(W^{(S)}+W'^{(S)})}{(LN\cdot T)\cdot \mathrm{Tr}(W'^{(S)}W^{(S)})}$$

and this is of order of $(\frac{1}{LN\cdot T})$. So we compute the purely spatial autocorrelation functions $\hat{\rho}_s(0)$, s=1,2,..., to compare the corresponding $\sqrt{\mathrm{var}(\hat{\rho}_s(0))}$ of the white noise to detect the existence of the unexhausted contemporaneous structure. Also we will compute the sample space-time autocorrelation function and compare the absolute value of these sample autocorrelation function with $(\frac{1}{LN\cdot(T-S)})$ to detect the unexhausted temporal structure among the estimated residuals. The fitted model will be accepted as adequate if no structures are detected.

5.5.2 Testing the Significance of STARMA Parameters

To test the significance of the model parameters, two approaches may be used. The first approach is through constructing the joint confidence interval and inferencing from the location of a zero element for the i parameter whether the i^{th} parameter is significant. The second method is by overfitting and testing the significance a particular trailing parameter.

5.5.2.1 The Confidence Region Approach for Joint Inference. To test the hypothesis,

$$H_0: \prod_{i=1}^{k_0} \beta_i = 0 \quad \text{(all the parameters are insignificant)}$$

vs.

$$H_1: \prod_{i=1}^{k_0} \beta_i \neq 0 \quad \text{(all the parameters are significant)}$$

at the significance level of α, the $(1-\alpha)$ joint confidence region is constructed according to the formula given in the Equation (5-48). We accept the hypothesis that the k_0 parameters are insignificant when $\underset{\sim}{0}$ is contained in the $(1-\alpha)$ confidence region. In addition, subsets of parameters can be evaluated. If there is not a plausibility of the i^{th} element in 0 to be $\neq 0$ (the i^{th} zero element is not contained in the $(1-\alpha)\%$ contour) the i^{th} parameter can be individually accepted to be significant.

5.2.2.2 Testing the Significance of the Model Parameters by Overfitting. Based on the "extra sum of squares" principle, Draper and Smith [1966], we compute the sum of squares for both the original candidate model parameters and the model with added parameters. For linear estimation problems with normal errors,

$$\frac{SS(\underset{\sim}{\beta}_{k_0}) - SS(\underset{\sim}{\beta}_{k_0+j})}{SS(\underset{\sim}{\beta}_{k_0+j})}\left[\frac{N-(k_0+j)}{j}\right] \sim F_{j,N-(k_0+j)} \tag{5-57}$$

where $\underset{\sim}{\beta}_{k_0}$ is the parameter vector of the original candidate model and $\underset{\sim}{\beta}_{k_0+j}$ is the parameter vector of the overfitted model which contains all the original parameters. We will reject the hypothesis that some of the added j parameters are significant at the α levels if the left-hand-side of Equation (5-57) is less than $F_{\alpha,j,N-(k_0+j)}$.

The direct extension of a test as above to the nonlinear purely spatial MA and ARMA is not necessarily valid. However, in practice, this test is used since these models are approximately linear over the

region of interest around the point estimates.

5.6 Examples

To illustrate the modeling procedure for purely spatial ARMA(λ_0,m_0) models, two examples are presented, one from hydrology and the second from crimenology. The first example is the heights of the Mohawk River at 6 locations observed twice a year. These data have been modeled by Perry and Aroian [1979] with $G = \sigma^2 I$ assumption. These data also have been modeled using STARIMA models by Deutsch and Pfeifer [1980a] with the general G innovations addressed. The second example is the monthly total number of arrests for assault in 14 areas of Northeast Boston. This set of data now examined in Deutsch and Pfeifer using STARIMA models [1979] and again by Deutsch and Pfeifer [1980a] in which this modeling effort has been extended to include the possibility of a contemporaneously corrected innovation. All these models are not able to describe the contemporaneously correlated structure. In this section, the purely spatial modeling techniques, that have been developed in the previous sections of this chapter, are applied to model the residuals of these two data sets to exhaust the contemporaneous correlations.

5.6.1 The Mohawk River Heights

The heights at six locations along the Mohawk River were observed every six months for the years 1967 to 1976 yielding 20 observations. These observations were recorded in feet above sea level. The series $\underset{\sim}{Z}_t = \underset{\sim}{W}_t - \bar{\underset{\sim}{W}}$, where $\underset{\sim}{W}_t$ is the observations at time t and $\bar{\underset{\sim}{W}}$ is the average, was modeled as a STMA(1_1) process,

$$\underset{\sim}{Z}_t = -\theta_{10}\underset{\sim}{a}_{t-1} - \theta_{11}W^{(1)}\underset{\sim}{a}_{t-1} + \underset{\sim}{a}_t \qquad (5\text{-}58)$$

with

$$\theta_{10} = 0.140$$

$$\hat{\theta}_{11} = -0.066$$

$$|\hat{G}| = 1.095, \qquad \hat{G} = \frac{1}{T}\sum_{t=1}^{T} \hat{\underset{\sim}{a}}_t \hat{\underset{\sim}{a}}_t'.$$

$$\hat{G} = \begin{bmatrix} 0.56 & & & \text{Symmetric} & & \\ 1.82 & 12.42 & & & & \\ -0.05 & 0.16 & 0.47 & & & \\ 1.25 & 11.42 & 0.03 & 14.81 & & \\ 0.25 & 0.96 & -0.07 & 0.68 & 0.40 & \\ 0.96 & 6.52 & 0.07 & 8.30 & 0.72 & 6.22 \end{bmatrix}$$

and $W^{(1)}$ is given as:

$$W^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Applying the purely spatial modeling procedure, we model the residuals $\hat{\underset{\sim}{a}}_t$, t=1,2,...,20 as the purely spatial process to exhaust the contemporaneous correlations.

Computing the sample purely spatial autocorrelation functions of the residuals of the STMA(1_1) model, we get

$$\hat{\rho}_{01} = 0.0687 \quad \text{and} \quad \text{Var}(\hat{\rho}_{01}) = 0.00266$$

Since we have only the 1st order neighbors, so we construct the purely spatial ARMA(1,1) model for these residuals, i.e.,

$$\hat{\underset{\sim}{a}}_t = \phi_{01}W^{(1)}\hat{\underset{\sim}{a}}_t - \theta_{10}W^{(1)}\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \qquad t=1,2,\ldots,20$$

First, the L.S. estimation under the $G = \sigma^2 I$ assumption was performed to give the following parameter estimates,

$$\hat{\phi}_{01} = 1.285$$

$$\hat{\theta}_{10} = 1.100$$

$$\hat{\sigma}^2 = 4.568$$

$$|\hat{G}| = \frac{1}{T}\sum_{i=1}^{T} (\hat{\underset{\sim}{\varepsilon}}_t \hat{\underset{\sim}{\varepsilon}}_t') = 1.093,$$ $\hat{\underset{\sim}{\varepsilon}}_t$ are the estimated residuals,

The procedures that have been described in Section 5.5.1 are applied to test the $G = \sigma^2 I$ assumption and the G=D assumption. To test the $G = \sigma^2 I$ hypothesis, the W-statistics in Equation (5-53), ρ and $-(T-1)\rho\ \ell n\ W$ are computed,

$$W = 0.000117.2$$
$$\rho = 0.883$$
$$-(T-1)\rho \ \ell n \ W = 151.58.$$

$$\hat{G} = \begin{bmatrix} 0.55 & & & & & \\ 1.89 & 12.50 & & \text{Symmetric} & & \\ -0.51 & -2.62 & 0.88 & & & \\ 0.94 & 8.22 & 1.76 & 9.58 & & \\ -0.60 & -4.58 & 0.79 & -3.60 & 2.13 & \\ 0.06 & 0.20 & -3.41 & 1.62 & 0.16 & 1.85 \end{bmatrix}$$

Figure 5-14. The $\hat{G}$ Matrix at $(\phi_{10},\theta_{10}) = (1.285,1.100)$.

The Z value in Equation (5-54) are computed at $\alpha = 0.01$ to give $Z_{\alpha=0.01} = 37.783$. Since $-(T-1)\rho \ \ell n \ W > Z_{\alpha=0.01}$, so the $G = \sigma^2 I$ hypothesis is rejected and G=D hypothesis is tested. Applying Equations (5-55) and (5.56), we obtain,

$$V = 0.0047373$$
$$m = 16.167$$
$$-m \ \ell n \ V = 96.53$$
$$v = 32.747 \quad \text{at} \quad \alpha = 0.01.$$

Since $-m \ \ell n \ V > v$ at $\alpha = 0.01$, so the G=D hypothesis is rejected.

Then the L.S. estimation under $G = \hat{G}$ assumption is executed to give $\hat{\phi}_{01} = 1.285$, $\hat{\theta}_{01} = 1.100$ and $SS = \sum_{t=1}^{20} \hat{\underset{\sim}{\varepsilon}}_t' \hat{G}^{-1} \hat{\underset{\sim}{\varepsilon}}_t = 120$. The model

parameter $(\phi_{01},\theta_{01},G)$ converges to $(\hat{\phi}_{01} = 1.285,\ \hat{\theta}_{01} = 1.100,\ G)$, and the estimation procedure stops. The sample space-time autocorrelation functions of the residuals are computed to check the white noise assumption. In Table 5-3, we see that the sample space-time autocorrelation functions of the residuals do not reveal any inadequacy and this purely spatial ARMA(1,1) model is accepted as adequate.

To see how significant the model parameters (ϕ_{01},θ_{01}) are, we construct the 95% confidence interval, shown in Figure 5-15. The contour is corresponding to the sum of squares SS = 126.

From Figure 5-15, we see that both ϕ_{01}, θ_{01} are significant, and we have the final model as follows,

$$\underset{\sim}{Z}_t = -\theta_{10}\hat{\underset{\sim}{a}}_{t-1} - \hat{\theta}_{11}W^{(1)}\hat{\underset{\sim}{a}}_{t-1} + \hat{\underset{\sim}{a}}_t$$
$$\hat{\underset{\sim}{a}}_t = \hat{\phi}_{01}W^{(1)}\hat{\underset{\sim}{a}}_t - \hat{\underset{\sim}{\theta}}_{01}W^{(1)}\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \qquad t=1,2,3,\ldots,20, \qquad (5\text{-}59)$$

where

$$\hat{\theta}_{10} = 0.140$$
$$\hat{\theta}_{11} = -0.066$$
$$\hat{\phi}_{01} = 1.285$$
$$\hat{\theta}_{01} = 1.100$$

$\underset{\sim}{\varepsilon}_t \sim NID(0,\hat{G})$, $\hat{G}$ is given in Figure 5-14.

Comparing the model, Equation (5-59), with the space-time $STMA(1_1)$ model, Equation (5-58), we see that the space-time model has

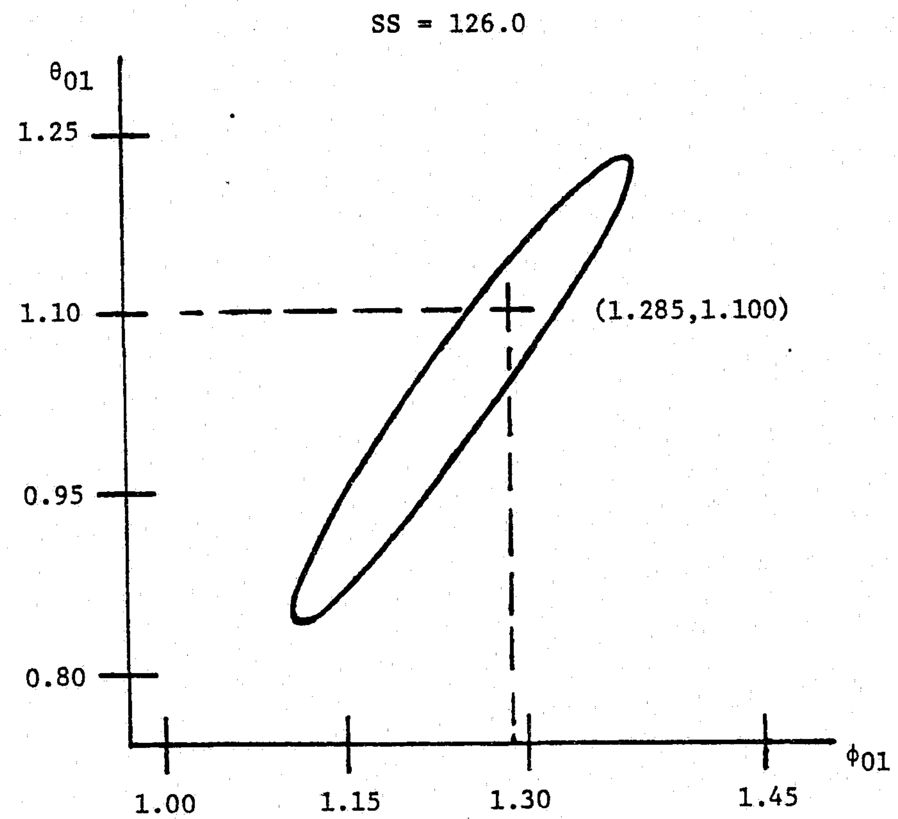


Figure 5-15. The Approximate 95% Confidence Interval of (ϕ_{01},θ_{01}) of the Mohawk River Data.

Table 5-3. The Sample Space-Time Autocorrelation Functions of the Residuals from the Purely Spatial ARMA(1,1) Model of the Mohawk River Heights Data

	The Sample Space-Time Autocorrelations/		The Standardized Sample S-T Autocorrelations	
Space Lag / Time Lag	0	1	0	1
0	1.00	0.02	1.00	0.28
1	0.04	0.04	0.42	0.41
2	0.01	0.00	0.14	0.03
3	-0.00	0.01	-0.03	0.09
4	0.01	-0.02	0.12	-0.23
5	0.06	-0.03	0.59	-0.32
6	-0.19	-0.09	-1.68	-0.76
7	-0.06	-0.01	-0.54	-0.08
8	0.01	0.03	0.10	0.25
9	-0.14	-0.01	-1.04	-0.09
10	-0.04	-0.03	-0.31	-0.26
11	-0.03	0.00	-0.21	0.00
12	-0.27	0.02	-0.16	0.12

the capability to model the spatial-temporal correlations, but it cannot model the contemporaneously correlated structures, and the model, Equation (5-59), has the capability to describe the contemporaneously correlated structures as well as the spatial-temporal correlations. Also the $|\hat{G}| = 1.093$ of model, Equation (5-59), is smaller than the $|\hat{G}| = 1.095$ of the space-time STMA(1_1) model.

5.6.2 The Northeast Boston Assaults Data

As the second application example, we consider the T = 72 monthly observations of the LN = 14 sites Northeast Boston Assault Data. The mean corrected first differences of these data, $\underset{\sim}{Z}(t)$, $t=2,3,\ldots,72$, have been modeled via the STMA(1_1) process, i.e.,

$$\underset{\sim}{Z}_t = -\hat{\theta}_{10}\hat{G}_{1/2}\underset{\sim}{\hat{a}}_{t-1} - \hat{\theta}_{11}W^{(1)}\hat{G}_{1/2}\underset{\sim}{\hat{a}}_{t-1} + \hat{G}_{1/2}\underset{\sim}{\hat{a}}_t \qquad (5\text{-}60)$$

where $\hat{G}_{1/2}\hat{G}'_{1/2} = \hat{G}$, $\hat{G}_{1/2}$ is a lower triangular matrix, and $\hat{G} = \frac{1}{T-1}\sum_{t=2}^{T}\underset{\sim}{\hat{a}}_t\underset{\sim}{\hat{a}}'_t$ is given in Figure 5-16. The M.L. estimates that have been obtained by Deutsch and Pfeifer [1980] are,

$$\hat{\theta}_{10} = 0.861$$

$$\hat{\theta}_{11} = -0.035, \text{ and}$$

$\hat{G}$ is limited in the Figure 5-16.

This space-time STMA(1_1) model have exhausted the spatial-temporal correlations, but it doesn't describe the contemporaneous correlations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.11													
2	-0.19	1.57												
3	0.18	-0.09	1.81											
4	-0.07	0.37	0.34	6.64						Symmetric				
5	-0.42	-0.06	0.61	0.21	3.87									
6	0.36	0.62	-0.11	1.19	0.10	7.17								
7	-0.22	1.14	0.02	1.00	0.10	0.83	14.80							
8	-0.32	0.53	-0.36	1.17	-0.43	0.60	-0.74	5.23						
9	-0.16	0.54	0.00	0.79	0.14	0.27	1.42	0.60	4.57					
10	-0.33	0.04	-0.16	-0.88	-0.22	-0.61	1.63	-0.53	0.59	6.06				
11	0.23	0.19	0.66	1.92	0.30	1.48	1.77	-0.10	-0.19	0.38	14.50			
12	-0.24	0.08	0.62	0.46	1.00	0.56	1.11	0.15	0.03	-0.05	1.49	5.57		
13	-0.31	-0.20	-1.65	-0.05	2.24	0.94	0.77	-0.40	-0.64	1.87	-1.81	1.63	14.90	
14	-0.08	0.16	-0.32	1.20	-0.13	0.10	0.08	-0.37	0.14	0.17	0.79	0.38	1.77	3.14

Figure 5-16. The Sample Covariance Matrix of the Estimated Residuals from the STMA(1_1) Model, Equation (5-60).

To exhaust the contemporaneously correlated structures, the purely spatial modeling procedures are applied to build the purely spatial model. The sample purely spatial autocorrelations and the standardized sample purely spatial autocorrelations are computed to give the following results,

Sample Autocorrelation	Standardized Autocorrelation
$\rho_{01} = 0.098$	2.400
$\rho_{02} = -0.002$	-0.041
$\rho_{03} = 0.063$	1.621

The pattern of expected sample autocorrelation functions for the purely spatial AR(1) model in Figure 5-9 matches the pattern of these sample purely spatial autocorrelation functions best. Reading Figure 5-11(a), we obtain the initial estimate $\phi_{01} = 0.1$. The L.S. estimation under $G = \sigma^2 I$ assumption gives

$$\hat{\underset{\sim}{a}}_t = \hat{\phi}_{01} W^{(1)} \hat{\underset{\sim}{a}}_t + \underset{\sim}{\varepsilon}_t, \quad t=2,3,\ldots,72$$

where

$$\hat{\phi}_{01} = 0.221$$

$$\hat{\sigma}^2 = 0.98155$$

To test the $G = \sigma^2 I$ hypothesis, Equations (5-53) and (5-54) are used to compute the following statistics:

$$W = 0.23382$$

$$\rho = 0.9306$$

$$-(T-1)\rho \ \ell n \ W = 109.31$$

$$Z_{\alpha=0.05} = 123.47$$

Since $-(T-1)\rho \ \ell n \ W < Z_{\alpha=0.05}$, so the $G = \sigma^2 I$ assumption for $\underset{\sim}{\varepsilon}_t$ cannot be rejected. Also the diagnostic checking applied on the estimated residuals $\hat{\underset{\sim}{\varepsilon}}_t$ to check the white noise assumption doesn't reveal any model inadequacies. The sample space-time autocorrelation functions are shown in Table 5-4.

To check the model parsimony, the model parameter ϕ_{01} is tested for significance in the following,

$$H_0: \ \phi_{01} = 0$$

vs.

$$H_1: \ \phi_{01} \neq 0$$

The extra sum of squares $SS_E = 18.339$ with $\hat{\sigma}_\varepsilon^2 = 0.982$, and

$$F = 18.68$$

Performing the F-test, we have

$$F > F_{0.05,1,991} = 3.84,$$

so ϕ_{01} is significant and the following final model is obtained,

Table 5-4. The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Purely Spatial Model, Equation (5-61)

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2	0	1	2
0	1.00	0.03	0.03	1.00	0.08	0.95
1	0.00	0.02	0.02	0.25	0.90	0.76
2	0.04	-0.02	0.02	1.23	-0.85	0.63
3	0.00	-0.01	-0.01	0.12	-0.35	-0.33
4	-0.02	0.03	0.03	-0.64	1.02	1.12
5	-0.06	0.02	0.03	-2.08	0.67	0.93
6	-0.00	-0.07	0.06	-0.05	-2.10	1.99
7	-0.04	0.03	-0.06	-1.37	0.88	-1.89
8	-0.02	0.02	0.02	-0.74	0.65	0.68
9	-0.00	-0.01	0.03	-0.02	-0.53	1.03
10	-0.03	-0.00	0.03	-1.07	-0.11	1.06
11	0.06	0.02	-0.01	1.81	0.65	0.44
12	0.03	0.01	-0.02	0.91	0.28	-0.60
13	-0.02	0.02	0.04	-0.74	0.66	1.37
14	-0.01	-0.02	0.00	-0.34	-0.72	0.16
15	0.03	-0.03	-0.02	0.90	-0.32	-0.77

$$\underset{\sim}{Z}_t = -\hat{\theta}_{10}\hat{G}_{1/2}\hat{\underset{\sim}{a}}_{t-1} - \hat{\theta}_{11}W^{(1)}\hat{G}_{1/2}\hat{\underset{\sim}{a}}_{t-1} + \hat{G}_{1/2}\hat{\underset{\sim}{a}}_t \tag{5-61}$$

$$\hat{\underset{\sim}{a}}_t = \hat{\phi}_{01}W^{(1)}\hat{\underset{\sim}{a}}_t + \underset{\sim}{\varepsilon}_t, \qquad t=2,3,\ldots,72,$$

where

$\hat{\theta}_{10} = 0.861$

$\hat{\theta}_{11} = -0.035$

$\hat{\phi}_{01} = 0.221$

$\hat{G}_{1/2}\hat{G}'_{1/2} = \hat{G}$, $\hat{G}_{1/2}$ is a low triangular matrix and

$\hat{G} = \frac{1}{(T-1)}\sum_{t=2}^{T}\hat{\underset{\sim}{a}}_t\hat{\underset{\sim}{a}}'_t$ is given in Figure 5-16.

The space-time STMA(1_1) model, Equation (5-60), models the spatial-temporal correlated structure of the observed process, but it can't model the contemporaneous spatial structures. By adding the purely spatial model, that has been built for the estimated residuals of the STMA(1_1) model, Equation (5-60), we increase the model capability to be able to describe the purely spatial structures as well as the spatial-temporal correlated structures.

CHAPTER VI

COUPLING AND REPARAMETERIZING MODELS

In this chapter two topics concerning coupling and reparameterizing models are analyzed; coupling and reparameterizing an aggregate (for all t) purely spatial model and the space-time model and coupling the purely spatial models for subsets of t. The purely spatial model,

$$\underset{\sim}{z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{z}_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t \varepsilon S \tag{6-1}$$

where S is a non-empty subset of $U_T = \{1,2,\ldots,T\}$, and observations are available from t=1 to t=T, is referred to as the individual purely spatial model when the temporal index set S contains only one element, and it is referred to as the aggregate purely spatial model when S contains more than one element. In Chapter V, the purely spatial model was the aggregate purely spatial model with $S = U_T$. In this chapter, the aggregate purely spatial model is distinguished from the individual purely spatial models. In Section 6.1 the procedures of coupling and reparameterizing the aggregated purely spatial model with the space-time model are developed. Three types of space-time processes and two modeling sequences are considered. Three space-time process types are: 1. every location has the same univariate ARIMA process, 2. the system follows a STARIMA process, and 3. each location

has a different univariate ARIMA model. The two modeling sequences are: 1. Modeling the observations to be a purely spatial process, then modeling the residuals to be a space-time process. Comparisons of the STARIMA models and the resulting coupled and reparameterized model for each model type are made and the equivalence of these two modeling sequences is discussed. Two application examples are included.

After the coupling procedures for the space-time model and the aggregated purely spatial model, the procedures for coupling the purely spatial models are developed in Section 6.2. The ergodic process is introduced and the modeling procedures for the ergodic process are developed under the homogenity assumption. Since an ergodic process may behave like non-ergodic because of inhomogeneous inputs, in Section 6.3 a procedure is developed to detect the outliers under the ergodic assumption. An example application is included to illustrate these procedures.

6.1 Coupling and Reparameterizing Aggregate Purely Spatial Models and the Space-Time Model

The space-time model has the capability of describing space-time diffusion mechanisms with the exception of the contemporaneously spatial diffusion mechanism. On the other hand, purely spatial models describe only contemporaneously spatial correlated structures without any temporal relations. Once a space-time model has been constructed and its residuals follow a purely spatial model, we may want to couple these models and reparameterize the coupled model to obtain the single system model. In this section, the coupling and reparameterization of the aggregate purely spatial models and space-time models are discussed.

Two modeling sequences may arise: 1. The observations are modeled to be an aggregate purely spatial process and then the residuals are modeled to be a space-time process, or 2. The observations are modeled to be a space-time process and then the residuals are modeled by the aggregate purely spatial process. In the following, the first modeling sequence, i.e., the aggregate purely spatial modeling followed by the space-time modeling, is detailed and the equivalence of two modeling sequences is verified in Section 6.1.1. Comparison of the coupled reparameterized models and the STARIMA models is made in Section 6.1.2. Two application examples are presented in Section 6.1.3.

6.1.1 Space-Time Models with Contemporaneous Terms

In this section it is assumed that the purely spatial model,

$$(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\underset{\sim}{Z}_t = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})\underset{\sim}{a}_t, \quad t \in U_T \qquad (6\text{-}2)$$

has been built and the LN streams of the estimated residuals has been computed. These residuals depending upon their structure, may follow one of three potential models;

1. All LN streams of residuals are described by the same univariate ARIMA model,
2. The LN streams of residuals follow a STARIMA process,

3. The LN streams of residuals are described by different univariate ARIMA models.

6.1.1.1 All LN Streams of Residuals Are of the Same Univariate ARIMA Model.

Here the residuals are assumed to be generated from the same univariate ARIMA(p,d,q) model. Without loss of generality, we will assume d=0 and have the aggregate purely spatial model,

$$\underset{\sim}{a}_t = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\underset{\sim}{Z}_t, \quad t=1,2,..,T, \qquad (6\text{-}3)$$

and

$$\underset{\sim}{a}_t = (I - \sum_{k=1}^{p} \phi_{k0} IB^k)^{-1}(I - \sum_{k=1}^{q} \theta_{k0} IB^k)\underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,T, \qquad (6\text{-}4)$$

for the LN streams of residuals with $\underset{\sim}{\varepsilon}_t \sim \mathrm{NID}(0,\sigma^2 I)$. By substituting for $\underset{\sim}{a}_t$ we obtain the coupled model for the system.

$$(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\underset{\sim}{Z}_t =$$

$$(I - \sum_{k=1}^{p} \phi_{k0} IB^k)^{-1}(I - \sum_{k=1}^{q} \theta_{k0} IB^k)\underset{\sim}{\varepsilon}_t$$

Since

$$(I - \sum_{k=1}^{p} \phi_{k0} I B^k) = (1 - \sum_{k=1}^{p} \phi_{k0} B^k) I,$$

so,

$$(I - \sum_{k=1}^{p} \phi_{k0} I B^k)(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} =$$

$$(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1}(I - \sum_{k=1}^{p} \phi_{k0} I B^k)$$

Multiplying both sides with $(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})$, we get,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \sum_{k=1}^{p} \phi_{k0}(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}) \underset{\sim}{Z}_{t-k}$$

$$- \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t - \sum_{k=1}^{q} \theta_{k0}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)}) \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t. \quad (6\text{-}4)$$

This can be reparameterized to the model,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_0} \phi^*_{k\ell} W^{(\ell)} \underset{\sim}{Z}_{t-k}$$

$$- \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t - \sum_{k=1}^{q} \sum_{\ell=0}^{m_0} \theta^*_{k\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t, \quad (6\text{-}5)$$

where

$$W^{(0)} = I$$

$$\phi^*_{k0} = \phi_{k0}$$

$$\theta^*_{k0} = \theta_{k0}$$

$$\phi^*_{k\ell} = \phi_{k0} \cdot \phi_{0\ell} \qquad k \geq 1,\ \ell \geq 1$$

$$\theta^*_{k\ell} = -\theta_{k0} \cdot \theta_{0\ell}, \text{ and}$$

$$\underset{\sim}{\varepsilon}_t \sim NID(0, \sigma^2 I).$$

<u>6.1.1.2 The LN Streams of Residuals Are from the STARIMA Process</u>. In this section, the residual of the purely spatial model is assumed to be generated by the STARIMA($\underset{\sim}{P}_{\lambda}$,0,$\underset{\sim}{q}_{m}$) process. These residual models are,

$$\begin{cases} \underset{\sim}{a}_t = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1}(I - \sum_{\ell=0}^{\lambda_0} \phi_{0\ell} W^{(\ell)}) \underset{\sim}{Z}_t, \quad t=1,2,\ldots,T \\ \underset{\sim}{a}_t = (I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} B^k)^{-1}(I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{0\ell} W^{(\ell)} B^k) \underset{\sim}{\varepsilon}_t \\ \qquad t=1,2,\ldots,T \end{cases} \quad (6\text{-}6)$$

with $\underset{\sim}{\varepsilon}_t \sim \text{NID}(0,\sigma^2 I)$. Equating the $\underset{\sim}{a}_t$ expressions results in the coupled system model.

$$(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}) \underset{\sim}{Z}_t =$$

$$(I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} B^k)^{-1} (I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell} W^{(\ell)} B^k) \underset{\sim}{\varepsilon}_t$$

or

$$(I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} B^k) (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}) \underset{\sim}{Z}_t$$

$$= (I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell} W^{(\ell)} B^k) \underset{\sim}{\varepsilon}_t .$$

This is equivalent to

$$[(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})$$

$$- \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} B^k (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})] \underset{\sim}{Z}_t$$

$$= (I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell} W^{(\ell)} B^k) \underset{\sim}{\varepsilon}_t .$$

Multiplying both sides with $(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})$ and rearranging the coefficients, we have

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \sum_{k=1}^{p} \Big[\sum_{\ell=0}^{\lambda} \phi_{\ \ell} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})$$

$$\cdot W^{(\ell)} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}) \Big] \underset{\sim}{Z}_{t-k}$$

$$- \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t - \sum_{k=1}^{q} \Big[\sum_{\ell=0}^{m_k} \theta_{k\ell} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)}) W^{(\ell)} \Big]$$

$$\cdot \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t . \qquad (6\text{-}7)$$

Equation (6-7) can be reparameterized as follows;

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t + \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W_{\phi}^{(\ell)} \underset{\sim}{Z}_{t-k}$$

$$- \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell} W_{\theta}^{(\ell)} \underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t \qquad (6\text{-}8)$$

where

$$W_{\phi}^{(h)} = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)})\, W^{(h)} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})$$

$$W_{\theta}^{(h)} = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)})\, W^{(h)}$$

In the isotropic preference STARIMA system since the weight matrices are symmetric, by applying the matrix multiplication, we see that $W^{(\ell)}W^{(h)} \approx \frac{1}{n^{(\ell)}+n^{(h)}} (W^{|\ell+h|} + W^{|\ell-h|})$, where $n^{(\ell)}$ and $n^{(h)}$ denote the average number of the ℓ^{th} and the g^{th} order neighbors, respectively. It should be noted that in the regular grid system, the average number of the arbitrary ℓ^{th} order neighbor is approximately 4, and $W^{(\ell)}W^{(h)}$ is negligible when compared to the identiy matrix I, where $\ell \neq 0$, $h \neq 0$. By omitting all negligible terms in Equation (6-8), we obtain,

$$W_{\phi}^{(h)} \doteqdot W^{(h)}$$
$$W_{\theta}^{(h)} \doteqdot W^{(h)}$$

This approximation holds closely for the low spatial order models, i.e., models with $\lambda_0, m_0 \leq 2$.

6.1.1.3 The LN Streams of Residuals Are from Different ARIMA Processes. The residuals of the aggregate purely spatial model are assumed to be generated from LN different ARIMA(p_i, d_i, q_i) i=1,2,...,LN univariate models. The models that describe the whole system are,

$$\begin{cases} \underset{\sim}{a}_t = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\, \underset{\sim}{Z}_t & t=1,2,\ldots,T, \\ \underset{\sim}{a}_{it} = (1 - \sum_{k=1}^{p_i} {}_i\phi_{k0}\, B^k)^{-1} (1 - \sum_{k=1}^{q_i} {}_i\theta_{k0}\, B^k)\, \varepsilon_{it} & t=1,2,\ldots,T \\ & i=1,2,\ldots,LN, \end{cases} \qquad (6\text{-}9)$$

where

${}_i\phi_{k0}$, $k=1,2,\ldots,p_i$ are AR parameters of the i^{th} residual stream.
${}_i\theta_{k0}$, $k=1,2,\ldots,q_i$ are MA parameters of the i^{th} residual stream.

In order to simplify the expression of Equation (6-9), we define $D_k = [{}_kd_{i,\ell}]$, $F_k = [{}_kf_{i,\ell}]$ below,

$${}_kd_{i,\ell} = \begin{cases} {}_i\phi_{k0} & \text{when } i=\ell \text{ and } 1 \leq k \leq p_i \\ 0 & \text{otherwise} \end{cases}$$
$${}_kf_{i,\ell} = \begin{cases} {}_i\theta_{k0} & \text{when } i=\ell \text{ and } 1 \leq k \leq q_i \\ 0 & \text{otherwise.} \end{cases} \qquad (6\text{-}10)$$

and $D_0 = I$, $F_0 = I$. The matrices D_j and F_k are diagonal which allows the LN different univariate ARIMA models in Equation (6-9) to be written as

$$a_t = (I - \sum_{j=1}^{p} D_j B^j)^{-1} (I - \sum_{k=1}^{q} F_k B^k) \varepsilon_t \tag{6-11}$$

where $p = \max\{p_i,\ i=1,2,\ldots,LN\}$ and $q = \max\{q_i,\ i=1,2,\ldots,LN\}$.

Coupling the aggregate spatial model in Equation (6-9) and the residual model in Equation (6-11) yields,

$$(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\, Z_t =$$

$$(I - \sum_{k=1}^{p} D_k B^k)^{-1} (I - \sum_{k=1}^{q} F_k B^k)\, \varepsilon_t .$$

This is equivalent to,

$$(I - \sum_{k=1}^{p} D_k B^k)\, (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{m_0} \phi_{0\ell} W^{(\ell)}) Z_t =$$

$$(I - \sum_{k=1}^{q} F_k B^k) \varepsilon_t$$

or

$$[(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})$$

$$- (\sum_{k=1}^{p} D_k B^k)\, (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})] Z_t$$

$$= (I - \sum_{k=1}^{q} F_k B^k) \varepsilon_t$$

Multiplying both sides by $(I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})$, we have

$$Z_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} Z_t + \sum_{k=1}^{p} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})\, D_k (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1}$$

$$\cdot (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})\, Z(t-k) - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \varepsilon_t$$

$$- \sum_{k=1}^{q} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})\, F_k\, \varepsilon_{t-k} + \varepsilon_t \tag{6-12}$$

Reparameterizing Equation (6-12) and we get,

$$Z_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} Z_t + \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_0} \phi^*_{0\ell}\, W_\phi(k,\ell) Z_{t-k} - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \varepsilon_t$$

$$- \sum_{k=1}^{q} \sum_{\ell=0}^{m_0} \theta^*_{0\ell}\, W_\theta(k,\ell) \varepsilon_{t-k} + \varepsilon_t \tag{6-13}$$

where

$$\phi^*_{0\ell} = \begin{cases} 1 & \text{when } \ell=0 \\ -\phi_{0\ell} & \text{when } 1 \leq \ell \leq \lambda_0 \end{cases}$$

$$W_\phi(k,h) = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})\, D_k (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} W^{(h)}, \quad k \geq 0, \; h \geq 0$$

$$\theta^*_{0\ell} = \begin{cases} 1 & \text{when } \ell=0 \\ -\theta_{0\ell} & \text{when } 1 \leq \ell \leq m_0 \end{cases}$$

$$W_\theta(k,h) = W^{(h)} F_k, \quad k \geq 0, \quad h \geq 0$$

Note that $W_\phi(0,h) = W^{(h)}$

$W_\theta(0,h) = W^{(h)}$.

6.1.1.4 The Equivalence of Modeling Sequences. Two modeling sequences can be followed to model the same space-time observed process. The observations may be modeled to be an aggregate purely spatial process then the residuals are modeled to be a space-time model, or the observations can be modeled to be a space-time model and then the residuals are modeled to be an aggregate purely spatial model. Models obtained from either sequence are coupled and reparameterized. In

this section, the equivalence of these two modeling sequences is discussed.

In all the three coupled models discussed in Section 6.1.1.1 to 6.1.1.3 coupling sequence employed followed block diagram in Figure 6-1, where

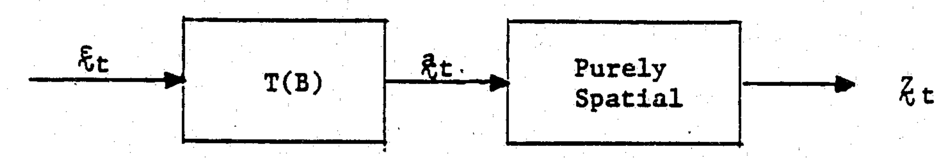


Figure 6-1. Coupling the Purely Spatial Observation Model and the Space-Time Residual Model

T(B) denotes a transfer function involving the temporal back shift operator B. In the coupled model 6-5, the case where the residuals are described by the same ARMA process, $T(B) = (I - \sum_{k=1}^{q} \theta_{k0} I B^k)^{-1} (I - \sum_{k=1}^{p} \phi_{k0} I B^k)$, in the coupled model 6-8, the case where the residuals are described by the STARMA process,

$$T(B) = (I - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell} W^{(\ell)} B^k)^{-1} \quad (I - \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} B^k),$$

and in the coupled model 6-13, in which the residuals are from different ARMA processes, $T(B) = (I - \sum_{k=1}^{p} F_k B^k)^{-1} (I - \sum_{k=1}^{p} D_k B^k)$. Changing this modeling sequence to the alternative sequence, we have the modeling sequence shown in Figure 6-2.

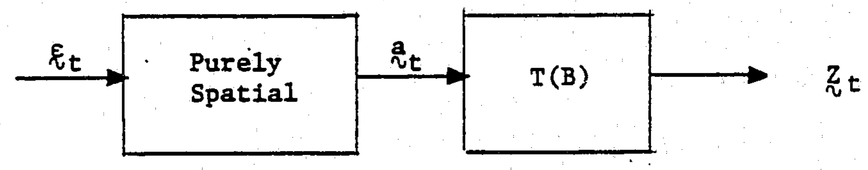


Figure 6-2. Coupling the Space-Time Observation Model and the Purely Spatial Residual Model

For the sequence, let $T(B)=(I-\sum_{k=1}^{q}\theta_{k0}IB^k)^{-1}(I-\sum_{k=1}^{p}\phi_{k0}IB^k)$, i.e., $\mathbf{Z}_t$ is modeled as LN same ARMA(p,q) processes and the residuals are modeled as a purely spatial model. In this case the two system models are;

$$\mathbf{a}_t = (I - \sum_{k=1}^{q} \theta_{k0}\, IB^k)^{-1}\,(I - \sum_{k=1}^{p} \phi_{k0}\, IB^k)\mathbf{Z}_t \tag{6-14}$$

$$\mathbf{a}_t = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}\, W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)})\boldsymbol{\varepsilon}_t \tag{6-15}$$

$t=1,2,\ldots,T.$

Equating Equations (6-14) and (6-15), we obtain the coupled system model,

$$(I - \sum_{k=1}^{q} \theta_{k0}\, IB^k)^{-1}\,(I - \sum_{k=1}^{p} \phi_{k0}\, IB^k)\mathbf{Z}_t$$
$$= (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}\, W^{(\ell)})^{-1}\,(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)})\boldsymbol{\varepsilon}_t .$$

This is equivalent to

$$\mathbf{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}\, W^{(\ell)}\mathbf{Z}_t + \sum_{k=1}^{p}\left[\phi_{k0}(I - \sum_{\ell=1}^{\lambda_0}\phi_{0\ell}\, W^{(\ell)})\right]\mathbf{Z}_{t-k}$$
$$- \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)}\boldsymbol{\varepsilon}_t - \sum_{k=1}^{q}\left[\theta_{k0}(I - \sum_{\ell=1}^{m_0}\theta_{0\ell}\, W^{(\ell)})\right]\boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t \tag{6-16}$$

Reparameterizing Equation (6-16), and we get,

$$\mathbf{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}\, W^{(\ell)}\mathbf{Z}_t + \sum_{k=1}^{p}\left[\sum_{\ell=0}^{\lambda_0}\phi^{*}_{k\ell}\, W^{(\ell)}\right]\mathbf{Z}_{t-k}$$
$$- \sum_{\ell=1}^{m_0} \theta_{0\ell}\, W^{(\ell)}\boldsymbol{\varepsilon}_t - \sum_{k=1}^{q}\left[\sum_{\ell=0}^{m_0}\theta^{*}_{k\ell}\, W^{(\ell)}\right]\boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t \tag{6-17}$$

where

$$\phi^{*}_{k0} = \phi_{k0} \qquad k \geq 1$$
$$\phi^{*}_{k\ell} = -\phi_{k0}\cdot\phi_{0\ell} \qquad k \geq 1,\ \ell \geq 1$$
$$\theta^{*}_{k0} = \theta_{k0} \qquad k \geq 1$$
$$\theta^{*}_{k\ell} = -\theta_{k0}\cdot\theta_{0\ell} \qquad k \geq 1,\ \ell \geq 1 .$$

Comparing Equations (6-5) and (6-17), we see that they are exactly the same. This is because the transfer function of the same LN ARMA models T(B) is interchangeable with the transfer function of the purely spatial model,

$$T(B)T(\text{purely spatial}) = T(\text{purely spatial})T(B). \qquad (6\text{-}18)$$

This property comes from the fact that identity matrix I is interchangeable with any matrix, and Equation (6-18) is true among the purely spatial transfer functions and the transfer function of STARIMA($p_{\underset{\sim}{\lambda}}$,d,$q_{\underset{\sim}{m}}$) with $\underset{\sim}{\lambda} = (0,0,0,\ldots,0)$ and $\underset{\sim}{m} = (0,0,\ldots,0)$, $\underset{\sim}{\lambda}$ and $\underset{\sim}{m}$ are p dimensions vector and q dimensions vector, respectively.

The STARIMA model is a more general case than LN streams of univariate ARIMA models. The STARMA system with isotropic preference has the property that if the location i is an ℓ^{th} order neighbor of the location j, then the location j is an ℓ^{th} order neighbor of the location i. In such isotropic systems, the following relationship holds,

$$\begin{aligned} W^{(\ell_1)}W^{(\ell_2)} &\approx (n^{(\ell_1)} + n^{(\ell_2)})^{-1}(W^{|\ell_1+\ell_2|} + W^{|\ell_2-\ell_2|}) \\ &= W^{(\ell_2)}W^{(\ell_1)}, \end{aligned} \qquad (6\text{-}19)$$

When the weight matrices $W^{(\ell_1)}$, $W^{(\ell_2)}$ are symmetric. Thus,

$$(I - \sum_{k=1}^{p} \phi_{k\ell}W^{(\ell)}B^k)^{-1}(I - \sum_{k=1}^{q}\sum_{\ell=0}^{m_k} \theta_{k\ell}W^{(\ell)}B^k)(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}$$

$$(I - \sum_{\ell=1}^{m_0} \phi_{0\ell}W^{(\ell)}) = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})$$

$$(I - \sum_{k=1}^{q}\sum_{\ell=0}^{\lambda_k} \phi_{k\ell}W^{(\ell)}B^k)^{-1}(I - \sum_{k=1}^{q}\sum_{\ell=0}^{m_k} \theta_{k\ell}W^{(\ell)}B^k), \qquad (6\text{-}20)$$

where $n^{(\ell_i)}$ is the average number of the ℓ_i^{th} neighbors in the system $\phi_{k\ell}$, $k \geq 1$ are in the stationary region, $\phi_{k\ell}$, $k \geq 1$ are in the invertible region, and $\phi_{0\ell}$, $\theta_{0\ell}$ are in the existence region. The transfer functions of the STARIMA process and the purely spatial model are,

$$T(B) = (I - \sum_{k=1}^{p} \phi_{k\ell}W^{(\ell)}B^k)^{-1}(I - \sum_{k=1}^{q}\sum_{=0}^{m_k} \theta_{k\ell}W^{(\ell)}B^k),$$

and

$$T(\text{Purely Spatial}) = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)}),$$

respectively. Equation (6-20) shows the relationship

$$T(B)\cdot T(\text{Purely Spatial}) = T(\text{Purely Spatial})\cdot T(B).$$

Therefore the exchange property, Equation (6-18), holds approximately for the system of equal preference.

In the STARMA system of two parallel preferential directions the weight matrices are symmetric, too. So Equation (6-19) and Equation (6-20) hold, and the exchange property, Equation (6-18) also holds. Similarly, in the STARMA system of one-direction preference, we have,

$$W^{(\ell_1)}W^{(\ell_2)} = W^{(\ell_2)}W^{(\ell_1)} = 0,$$

and Equations (6-20) and (6-18) hold.

For the case where the residuals follow LN different univariate ARIMA models, we have the following:

1. If the system is isotropic or preferential in two parallel directions and it is large enough to make the boundary effect negligible, i.e., $W^{(\ell)}$ is approximately symmetric, then the exchange property in Equation (6-18) holds because the relation $W^{(\ell)}D = DW^{(\ell)}$ holds when $W^{(\ell)}$ is symmetric and D is diagonal.
2. If the system is preferential in one direction, the $W^{(\ell)}$'s are not symmetric, and the exchange property in Equation (6-18) doesn't hold.

In the situations that the interchange property in Equation (6-18) holds, these two modeling sequences are equivalent and the transfer functions, i.e., T(B) and T(Purely Spatial), are independent.

In the situations that the interchange property does not hold, these two modeling sequences are not equivalent, transfer functions T(B) and T(Purely Spatial) are dependent. This means that the overall transfer function of this process is decomposed into two different modeling sequences. However, since they are decompositions of the same overall transfer function, so they have the same description ability even though the model interpretations of different modeling sequences are different. Since the overall transfer function can't be decomposed into sequence independent transfer functions when T(B) and T(Purely Spatial) are dependent, therefore sequential modeling is not recommended and the overall transfer function should be obtained from simultaneous estimation of the space-time model with contemporaneous terms.

6.1.2 Comparison of the Reparameterized Models and the STARIMA Models

The reparameterized models are expressed in Equation (6-5), Equation (6-8) and Equation (6-13). Equation (6-5) is the reparameterized model of the purely spatial model and LN same univariate ARIMA models. Equation (6-8) is that of the purely spatial model and the STARIMA model. Equation (6-13) is the reparameterized model of the purely spatial model and LN different univariate ARIMA models. Comparing Equations (6-5), (6-8) and (6-13) with the STARIMA($p_{\underset{\sim}{\lambda}}$,0,$q_{\underset{\sim}{m}}$) model,

$$\underset{\sim}{Z}_t = \sum_{k=1}^{p}\sum_{\ell=0}^{\lambda_k} \phi_{k\ell}W^{(\ell)}\underset{\sim}{Z}_{t-k} - \sum_{k=1}^{q}\sum_{\ell=0}^{m_k} \theta_{k\ell}W^{(\ell)}\underset{\sim}{\varepsilon}_{t-k} + \underset{\sim}{\varepsilon}_t ,$$

we see that all these coupled models still keep the purely spatial terms unchanged and the STARIMA model doesn't have any terms to present the 0^{th} lag dependence between observations and unobservable errors. Looking at the temporal lag terms of the coupled models, we see that except the LN same univariate ARIMA case, all the weight matrix of non-zero time lags are transformed. In Equation (6-8), the weight matrix of the non-zero lag terms are transformed by the matrix that contain purely spatial model parameters only. The weight matrices $W_\phi^{(h)}$, $W_\theta^{(h)}$ in Equation (6-8) are independent of the temporal lag, while the transformed weight matrix $W_\phi(k,h)$, $W_\theta(k,h)$ in Equation (6-13) are temporal lag k dependent, i.e., $W_\phi(k_1,h) \neq W_\phi(k_2,h)$, $W_\theta(k_1,h) \neq W_\theta(k_2,h)$ for different temporal lags k_1 and k_2. Also, for practical purposes, $W_\phi^{(h)} \doteq W^{(h)} \doteq W_\theta^{(h)}$ in Equation (6-8) for the low order spatial system.

We have already seen that all the coupled models keep the same zero^{th} lag terms, while the STARIMA models do not contain any zero^{th} lag terms. There are situations in which the influence of one location on another is so quick that there is no lead time before this influence is reached. If we employ the STARIMA models to model such quick influence-spreading processes, we can't exhaust the whole structure and the resulting model will usually have the residuals distributed as NID(0,G). This can be seen more clearly in the coupled process represented in Equation (6-5) or (6-8). If the residuals from the observation model of Equation (6-6) are not modeled by the purely spatial model, the resulting STARIMA model will have the residuals $\underset{\sim}{a}_t$ distributed as NID(0,G) with

$G = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})(I - \sum_{\ell=1}^{m_0} \theta_{0\ell}W^{(\ell)})'[(I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell}W^{(\ell)})^{-1}]'\sigma^2$ even when the true residual (after being fitted by a spatial model) $\underset{\sim}{\varepsilon}_t$ is distributed as $NID(0,\sigma^2 I)$. So by coupling the purely spatial model with the space-time models, we add in the zero^{th} terms to generalize the capability of describing the instantaneous influence mechanism with the contemporaneous temrs. Thereby unconfounding the purely spatial structure from the noise covariance.

Since the most obvious change by coupling and reparameterizing these models lies in the insertion of purely spatial terms, so the models presented in Equations (6-8) and (6-13) will be referred to as the STARIMA model with the contemporaneous terms and the LN different univariate ARIMA models with the contemporaneous terms. They will be classified as space, space-time models denoted by $[ARMA(\lambda_0,m_0) + STARIMA(\underset{\sim}{p_\lambda},d,\underset{\sim}{q_m})]$ and $[ARMA(\lambda_0,m_0) + ARIMA(p,d,q)^{LN}]$, respectively. The coupled model in Equation (6-5) is a special case of Equation (6-8) and can be denoted by $[ARMA(\lambda_0,m_0) + STARIMA(\underset{\sim}{p_0},0,\underset{\sim}{q_0})]$, here $\underset{\sim}{\lambda}$ and $\underset{\sim}{m}$ are set to the null vector $\underset{\sim}{0}$.

6.1.3 Example Applications

In this section, two examples, i.e., the Mohawk River Heights Data and the Northeast Boston Crime Data, are given to illustrate the coupling procedures. The STARIMA models have been built by Deutsch and Pfeiffer [1980] and the purely spatial models have been constructed in Chapter V. The forecasts, based on the STARIMA models and the coupled models, are also computed and compared to illustrate the practical consequences of ignoring the contemporaneously correlated structure.

6.1.3.1 The Models of the Mohawk River Heights Data. In Chapter V we had the STARIMA model for the observations and the purely spatial model for the residuals as listed below,

The STARIMA model for the observations:

$$\underset{\sim}{Z}_t = -\hat{\theta}_{10}a_{t-1} - \hat{\theta}_{11}W^{(1)}\underset{\sim}{a}_{t-1} + \underset{\sim}{a}_t, \text{ and}$$

the purely spatial model for the residuals:

$$\underset{\sim}{a}_t = \hat{\phi}_{01}W^{(1)}\underset{\sim}{a}_t - \hat{\theta}_{01}W^{(1)}\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,20.$$

where

$\hat{\theta}_{10} = 0.140$

$\hat{\theta}_{11} = -0.066$

$\hat{\phi}_{01} = 1.285$

$\hat{\theta}_{01} = 1.100$

$\underset{\sim}{\varepsilon}_t \sim NID(0,\hat{G})$, $\hat{G}$ is given in Figure 5-13.

The observations are modeled to be STMA(1_1) process and the residuals

are modeled as purely spatial ARMA(1,1) model. According to the equivalent discussion in Section 6.1.1.4, this modeling sequence is equivalent to the other alternatives. Coupling and reparameterizing the models, we have the space, space-time [ARMA(1,1) + STARIMA(0,0,1)] model as follows,

$$\underset{\sim}{Z}_t = {}^c\phi_{01}W^{(1)}\underset{\sim}{Z}_t - {}^c\theta_{01}W^{(1)}\underset{\sim}{\varepsilon}_t - {}^c\theta_{10}\underset{\sim}{\varepsilon}_{t-1} - {}^c\theta_{11}W^{(1)}\underset{\sim}{\varepsilon}_{t-1} + \underset{\sim}{\varepsilon}_t, \quad (6\text{-}21)$$

with the initial guess values of the coupled model parameters being,

${}^c\phi_{01} = 1.285,$

${}^c\theta_{01} = 1.100,$

${}^c\theta_{10} = 0.140,$

${}^c\theta_{11} = -0.22.$

The M.L. estimation, which minimize $|G|$, gives

		95% CI
${}^c\hat{\phi}_{01}$	= 1.236	(1.228, 1.245)
${}^c\hat{\theta}_{01}$	= 0.382	(0.366, 0.397)
${}^c\hat{\theta}_{10}$	= 0.146	(0.102, 0.190)
${}^c\hat{\theta}_{11}$	= -0.120	(-0.136, -0.103)
$\lvert {}^c\hat{G} \rvert$	= 1.063	

and ${}^{c}\hat{G}$ as

$$
{}^{c}\hat{G} = \begin{bmatrix} 0.554 & & & & & \\ 1.513 & 10.160 & & \text{Symmetric} & & \\ -1.909 & -10.140 & 10.880 & & & \\ 0.750 & 6.138 & -6.299 & 8.079 & & \\ -1.252 & -9.554 & 9.250 & -8.831 & 11.880 & \\ 0.247 & 1.123 & -1.110 & 1.502 & -1.926 & 1.585 \end{bmatrix}
$$

The sample space-time autocorrelation functions and the sample space-time partial autocorrelation functions are listed in Table 6-1. No additional patterns are seen, so the reparameterized space, space-time [ARMA(1,1) + STARIMA(0,0,1)] model is accepted as adequate.

To test the general G assumption, we test the hypothesis,

$$
H_0: \; G = D
$$

vs.

$$
H_1: \; G \neq D
$$

where D is arbitrary diagonal matrix with positive diagonal elements. Following the testing procedures described in Section 5.5.1, we have,

$$
V = \frac{|\hat{G}|}{\prod_{i=1}^{LN} \hat{G}_{ii}} = 0.0001141, \text{ where } \hat{G}_{ii} \text{ is the (i,i) element of } \hat{G},
$$

Table 6-1. The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Reparameterized Model for the Mohawk River Heights

	Space-Time Autocorrelations		Standardized S-T Autocorrelations	
Space Lag / Time Lag	0	1	0	1
0	1.00	0.03	1.00	0.32
1	-0.03	-0.06	-0.26	-0.58
2	-0.05	0.01	-0.51	0.13
3	-0.23	0.08	-2.18	0.79
4	-0.10	0.15	0.94	1.37
5	-0.10	-0.28	-0.84	-2.47
6	-0.84	-0.09	-0.71	-0.72
7	-0.19	-0.13	-1.59	-1.06
8	0.21	0.13	-1.65	1.04
9	0.13	0.06	0.97	0.43
10	-0.09	0.01	-0.61	0.09
11	-0.27	-0.08	-1.77	-0.54
12	-0.09	0.12	-0.55	0.71
13	-0.15	0.22	-0.81	1.19

$$m = T - \frac{2LN+11}{6} = 16.17, \quad T = 20,\ LN = 6,$$

$$-m\ \ell n\ V = 146.768$$

$$f = 21$$

Since $-m\ \ell n\ V = 146.768 > \chi^2_{0.01,f} = 38.93$. So the null hypothesis H_0: G=D is rejected. Since $G \neq D$, so G can't be $\sigma^2 I$ and the contemporaneous white noise are correlated. It should be noted that in the Section 6.1.2, it has been pointed out that without the capability of modeling the purely spatial structure, the contemporaneous structure is confounded with the noise structure. Unconfounding doesn't guarantee the simplification of the covariance G to give $G = \sigma^2 I$ or G=D. However, if the process noises are distributed as $N(0,\sigma^2 I)$, then the unconfounding of the purely spatial structure and the noise structure gives $G = \sigma^2 I$.

Once the process is modeled, it will be employed to build the forecasting model for the purpose of process forecasting and/or process control. The forecast functions for both the coupled model and the STARMA model are constructed by taking conditional expectations at time T. The resulting forecast functions are,

$$\begin{aligned} {}^c\hat{\underset{\sim}{Z}}_T(1) &= (I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}(-{}^c\hat{\theta}_{10}I - {}^c\hat{\theta}_{11}W^{(1)})({}^cp')^{-1}\hat{\underset{\sim}{\varepsilon}}_T \\ {}^c\hat{\underset{\sim}{Z}}_T(\ell) &= 0 \qquad\qquad \ell \geq 2 \end{aligned} \tag{6-22}$$

$$\begin{aligned} \hat{\underset{\sim}{Z}}_T(1) &= (-\hat{\theta}_{10}I - \hat{\theta}_{11}W^{(1)})(p')^{-1}\hat{\underset{\sim}{\varepsilon}}_T \\ \hat{\underset{\sim}{Z}}_T(\ell) &= 0 \qquad\qquad \ell \geq 2 \end{aligned} \tag{6-23}$$

where

T is the last time index of the observations that were used to build the model,

$\underset{\sim}{Z}_T(\ell)$ is the ℓ-step ahead forecast at time point T,

$pp^1 = \hat{G}$ and ${}^cp{}^cp^1 = {}^c\hat{G}$.

The forecasting function contained in Equation (6-22), is based on the coupled [ARMA(0,1) + STARMA($0,1_1$)] model and the forecasting function contained in Equation (6-23), is based on the STARMA($0,1_1$) model. The primary difference in these forecasting functions is the contemporaneous spatial information that is contained in the $(I - \hat{\phi}^c_{01}W^{(1)})^{-1}$ matrix, which is numerically computed and listed below.

$$(I - {}^c\hat{\phi}_{01}W^{(1)})^{-1} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1.24 & 1 & 0 & 0 & 0 & 0 \\ 1.53 & 1.74 & 1 & 0 & 0 & 0 \\ 1.89 & 1.53 & 1.24 & 1 & 0 & 0 \\ 2.34 & 1.89 & 1.53 & 1.24 & 1 & 0 \\ 2.89 & 2.34 & 1.89 & 1.53 & 1.24 & 1 \end{bmatrix}$$

This matrix is lower triangular with all non-zero elements positive and

$e(i_1,j) > e(i_2,j)$, $i_1 > i_2$, where $e(i,j)$ is the (i,j) element of $(I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}$. The $e(i,j)$ elements represent the instantaneous influence from location j to location i. It can be seen that location 1 has the strongest influence to the other locations, and that location 2 has the second strongest influence. The downstream locations have less influence to their downstream locations than their upstream locations. This corresponds to the fact that the stream flows down the river and the heights at downstream locations will have no influence on the heights at the upstream locations. Also since the influence accumulates along the downstream direction, the influence from some upstream location, say location j, is stronger than the influence from any downstream location of the location j.

Since the process is a space-time moving average process, the point forecasts ${}^c\hat{\underset{\sim}{Z}}_T(\ell)$ and $\hat{\underset{\sim}{Z}}_T(\ell)$ and of the same zero value for $\ell \geq 2$, i.e., ${}^c\hat{\underset{\sim}{Z}}_T(\ell L = \hat{\underset{\sim}{Z}}_T(\ell) = 0$ for $\ell \geq 2$. However, the 1-step ahead forecasts are different and the variances of these forecasts as well as the interval forecasts are different. The variance of ${}^c\hat{\underset{\sim}{Z}}_T(\ell)$ is the conditional expectation of $({}^c\underset{\sim}{Z}_{T+\ell}\, {}^c\underset{\sim}{Z}'_{T+\ell})$ given that observations $\underset{\sim}{Z}_t$ are realized for $t \leq T$, so,

$$\begin{aligned} \mathrm{Var}({}^c\hat{\underset{\sim}{Z}}_T(1)) &= (I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}(1+{}^c\hat{\theta}_{10}^2)\,{}^c\hat{G}(I - {}^c\phi_{01}W^{(1)})^{-1'}, \quad \ell=1 \\ \mathrm{Var}({}^c\hat{\underset{\sim}{Z}}_T(\ell)) &= (I - {}^c\phi_{01}W^{(1)})^{-1}\,{}^c\hat{G}(I - {}^c\phi_{01}W^{(1)})^{-1'}, \quad \ell \geq 2. \end{aligned} \tag{6-24}$$

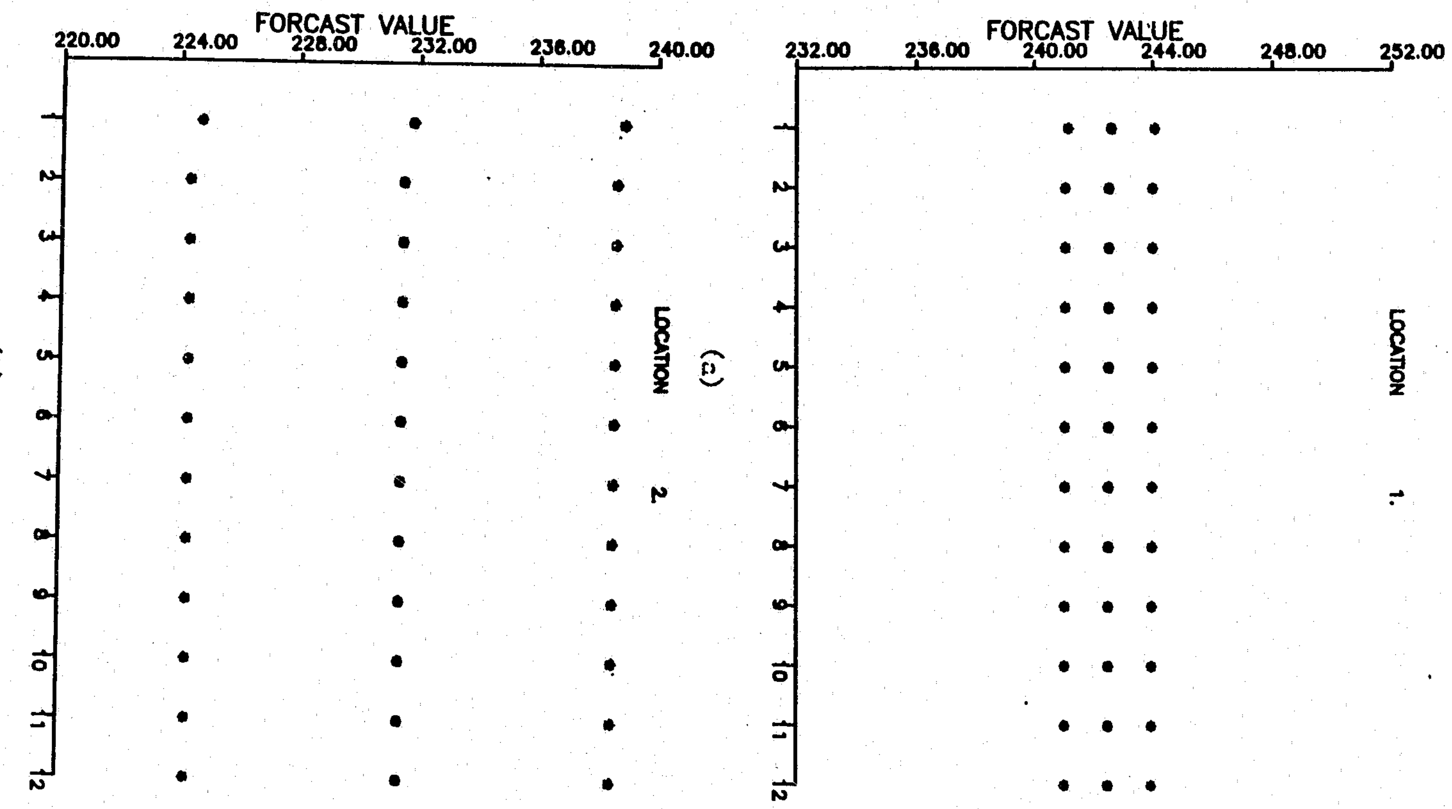
Similarly, the variance of $Z_T(\ell)$ is

$$\begin{aligned} \mathrm{Var}(\hat{\underset{\sim}{Z}}_T(1)) &= \hat{G} & \ell=1, \\ \mathrm{Var}(\hat{\underset{\sim}{Z}}_T(\ell)) &= (1+\hat{\theta}_{10}^2)\hat{G}, & \ell \geq 2. \end{aligned} \tag{6-25}$$

The point forecasts as well as the interval forecasts are computed and plotted in Figures 6-3(a) - (f) for location 1 to location 6 for both models.

From these plots, we see that the point forecasts ${}^c\hat{\underset{\sim}{Z}}_T(\ell)$ and $\hat{\underset{\sim}{Z}}_T(\ell)$ do not differ too much for all these 6 locations since they are all unbiased forecasts. While theinterval forecasts differ the most at location 6, which is the last downstream location. Since the contemporaneous influence from the upstream locations to the downstream locations accumulates, so it is reasonable that the forecasts from the forecasting model with the contemporaneous spatial structure will differ most from the corresponding forecasts that are from the forecasting model without the contemporaneous spatial structure at the last downstream location. The interval forecasts of the model always, Equation (6-22),lie inside the interval forecasts of the model, Equation (6-23), reflecting the improved model structure due to the addition of the contemporaneous spatial information. Thus, the precision of the forecasts are significantly improved.

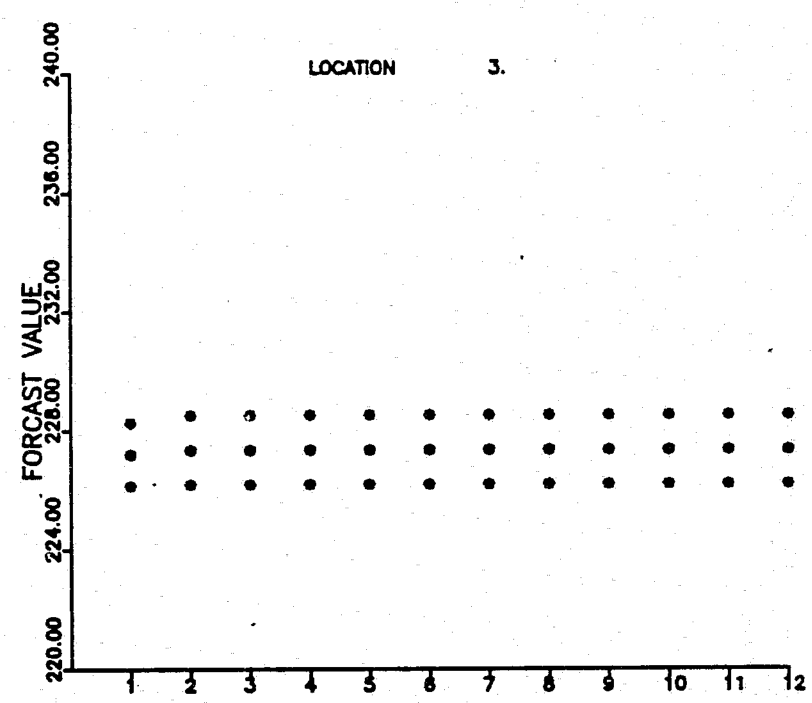
<u>6.1.3.2 The Models of Northeast Boston Assault Arrests</u>. In Chapter V we have the STARIMA model for the observations of assault arrests in Boston and the purely spatial model for the residuals of this model as,



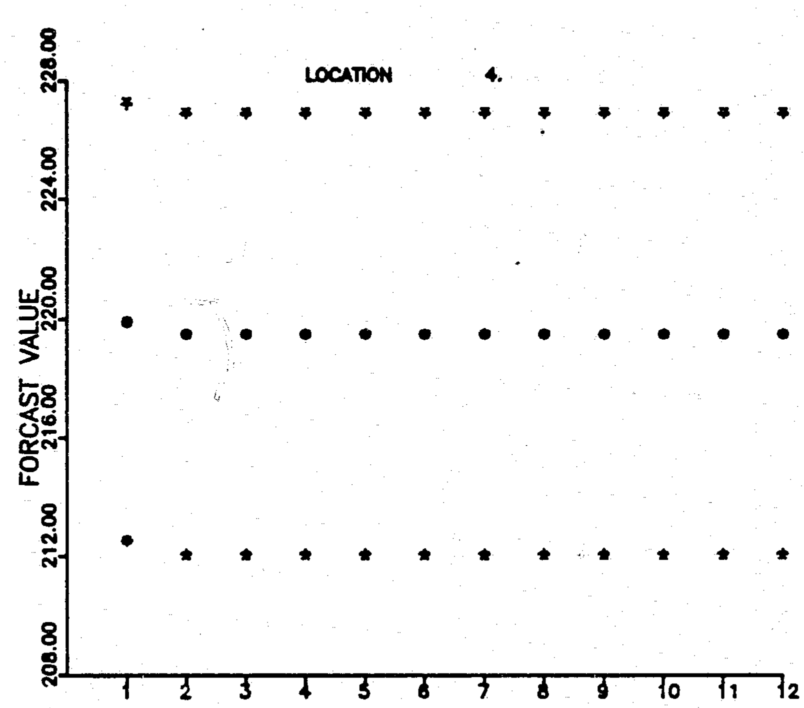
Notations

*: Point Forecasts of the Model with Contemporaneous Terms.
✲: Point Forecasts of the Model without Contemporaneous Terms.
×: 95% Confidence Intervals of the Model with Contemporaneous Terms.
+: 95% Confidence Intervals of the Model without Contemporaneous Terms.

Figure 6-3. The Point Forecasts and the 95% Confidence Interval Forecasts for the Mohawk River Heights.

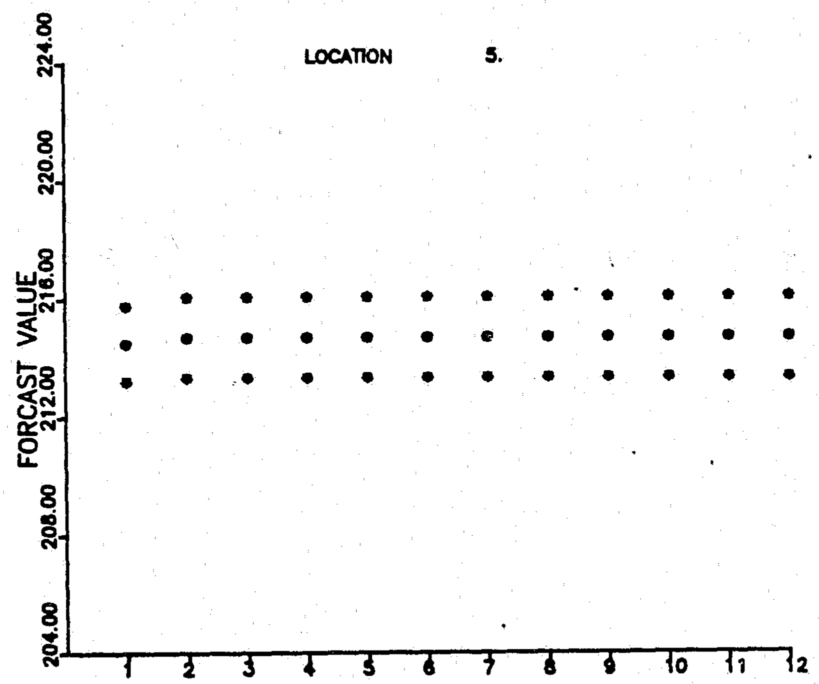


(c)

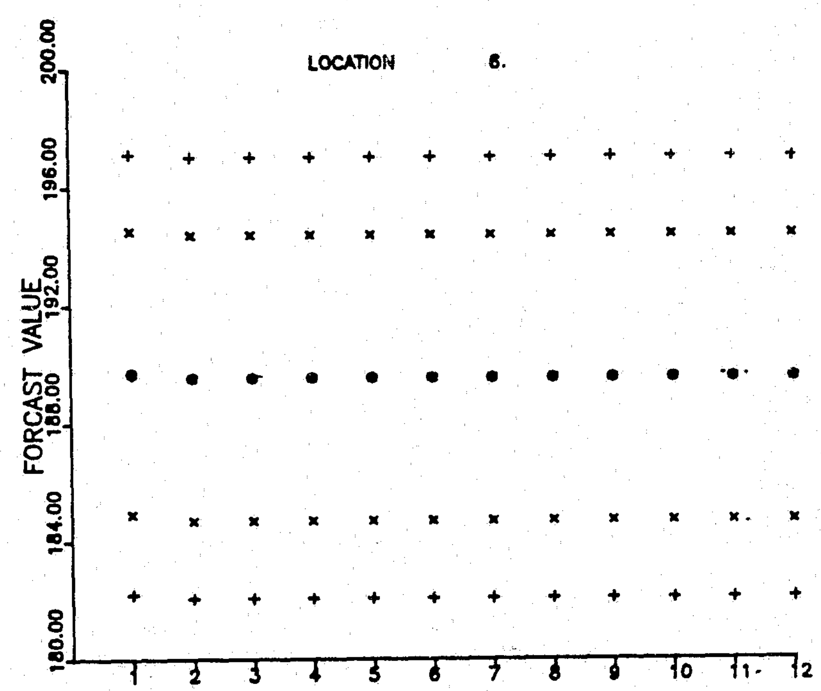


(d)

Figure 6-3. (Cont'd)



(e)



(f)

Figure 6-3. (Cont'd)

the STARIMA model for the observations:

$$\begin{cases} \mathbf{Z}_t = \mathbf{y}_t - \mathbf{y}_{t-1} \\ \mathbf{Z}_t = -\hat{\theta}_{10}\mathbf{a}_{t-1} - \hat{\theta}_{11}W^{(1)}\mathbf{a}_{t-1} + \mathbf{a}_t, \text{ and} \end{cases} \tag{6-26}$$

the purely spatial model for the residuals:

$$\mathbf{a}_t = \hat{\phi}_{01}W^{(1)}\mathbf{a}_t + \boldsymbol{\varepsilon}_t$$

where

$\mathbf{y}_t$ is the monthly observations,

$\mathbf{Z}_t$ is the differenced data, and

$\hat{\theta}_{10} = 0.861$

$\hat{\theta}_{11} = -0.035$

$\hat{\phi}_{01} = 0.2213$

$\boldsymbol{\varepsilon}_t \sim NID(0,\hat{G})$ and $\hat{G}$ is given in Figure 5-15.

Coupling and reparameterizing these models, we have the space, space-time [ARMA(1,0) + STARIMA(0,1,1)] model as follows,

$$\begin{cases} \mathbf{Z}_t = y_t - y_{t-1} \\ \mathbf{Z}_t = {}^c\phi_{01}\mathbf{Z}_t - {}^c\theta_{10}\boldsymbol{\varepsilon}_{t-1} - {}^c\theta_{11}W^{(1)}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t, \end{cases} \tag{6-27}$$

with the initial guesses of the model parameters, ${}^{c}\phi_{01} = 0.2213$, ${}^{c}\theta_{10} = 0.861$, ${}^{c}\theta_{11} = -0.035$. The M.L. estimation with $G = \sigma^2 I$ assumption gives, ${}^{c}\hat{\phi}_{01} = 0.2394$, ${}^{c}\hat{\theta}_{10} = 0.8162$, ${}^{c}\hat{\theta}_{11} = -0.0748$. Then the following hypotheses concerning the sphericity of G are subjected to test. The testing procedures described in Section 5.5.1 are used to test the following hypotheses,

1. H_0: $G = D$

   vs

   H_1: $G \neq D$

   Applying the approximate χ^2 test, Equation (5-55), we have $-m \ln V = 108.45$, f=91 and $-m \ln V < \chi^2_{0.10,f} = 108.67$, so the null hypothesis H_0: G=D can't be rejected at significant level $\alpha = 0.10$.

2. H_0: $G = \sigma^2 I$

   vs.

   H_1: $G \neq \sigma^2 I$

   Applying the approximate χ^2 test, Equation (5-50), we have $-(T-1)\rho \ln W = 383.45$, f = 104 and $-(T-1)\rho \ln W > \chi^2_{0.01,f} = 129.48$, so H_0: $G = \sigma^2 I$ is rejected at $\alpha = 0.01$.

So G=D assumption is appropriate and the G=D M.L. estimation is performed. The positive square roots of the diagonal elements of D are listed in Figure 6-4. The G=D M.L. estimation gives the model,

$$[d_{ii}] = \begin{bmatrix} 1.070 \\ 1.276 \\ 1.365 \\ 2.540 \\ 2.004 \\ 2.653 \\ 3.836 \\ 2.342 \\ 2.108 \\ 2.389 \\ 3.777 \\ 2.261 \\ 3.767 \\ 1.705 \end{bmatrix}$$

Figure 6-4. The Positive Square Roots of the Diagonal Elements of the Diagonal Covariance Matrix for the Reparameterized Model of the Northeast Boston Assault Data

$$\mathbf{Z}_t = {}^c\phi_{01}\mathbf{Z}_t - {}^c\theta_{10}D^{1/2}\boldsymbol{\varepsilon}_t - {}^c\theta_{11}W^{(1)}D^{1/2}\boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_t, \qquad (6\text{-}28)$$

where $D^{1/2}$ is the diagonal matrix such that $D^{1/2}D^{1/2} = D$,

$$\begin{aligned} {}^c\hat{\phi}_{01} &= 0.1087 \\ {}^c\hat{\theta}_{10} &= 0.848 \\ {}^c\hat{\theta}_{11} &= -0.029 \\ \hat{\sigma}^2 &= 0.9913. \end{aligned}$$

The extra sum of squares and F-statistics associated with ${}^c\hat{\theta}_{11}$ are,

$$SS_E({}^c\hat{\theta}_{11} | {}^c\hat{\phi}_{01}, {}^c\hat{\theta}_{10}, \hat{D}) = 1.817$$

$$\text{F-statistics} = 1.848 = F_{0.17,1,1991}.$$

So the null hypothesis H_0: $\hat{\theta}_{11} = 0$ is rejected at significant level $\alpha = 0.17$.

To confirm the assumption that the covariance matrix G=D is equivalent to test the covariance of $\boldsymbol{\varepsilon}_t$ in Equation (6-28) that $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \sigma^2 I$. Testing the hypothesis

$$\begin{aligned} H_0&: \ E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \sigma^2 I \\ &\text{vs.} \\ H_1&: \ E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') \neq \sigma^2 I \end{aligned}$$

we have $-(T-1)\rho \ \ln W = 103.65 = \chi^2_{0.49,104}$. The null assumption that $E(\varepsilon_t\varepsilon_t') = \sigma^2 I$ is accepted and G=D assumption is justified. The sample space-time autocorrelation functions are listed in Table 6-2. This disgnostic check doesn't reveal any further inadequacy, so the model, Equation (6-28), is accepted as adequate.

In addition to the space-time structure, the coupled model, Equation (6-28), has the capability of modeling the purely spatial structure and unconfounding the purely spatial structure from the contemporaneous noise structure. The space-time model, Equation (6-26), can't model the purely spatial structure and the purely spatial structure is confounded with the contemporaneous noise structure. The physical interpretations are different, the coupled model distinguishes the purely spatial structure from the contemporaneous noise structure while the space-time model doesn't have the capability to extract the information of purely spatial structure and the estimated noise covariance represents the confounded white noise structure and the pure spatial structure. In this example, we have the purely spatial ARMA(1,0) process structure and independent noise structure with different variances for different locations, i.e., G=D. Without the inclusion of the purely spatial terms, in Equation (6-26), the noise structure is misunderstood as general G, since the purely spatial ARMA(1,0) process structure is confounded with the G=D noise structure to give the general G noise covariance. Because the purely spatial structure is embedded in the estimated noise structure, the interpretation of these two models are different, the descriptive capabilities of these two models are the same, however, since both models include

Table 6-2. The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Model, Equation (6-28)

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
0	0.02	0.01	-0.03	0.84	0.58	-1.21
1	0.01	0.03	0.02	0.42	1.23	0.87
2	0.04	-0.01	0.01	1.35	-0.37	0.53
3	0.00	-0.00	-0.01	0.10	-0.07	-0.35
4	-0.01	0.03	0.04	-0.37	1.05	1.20
5	-0.06	0.02	0.03	-1.86	0.62	1.04
6	-0.00	-0.07	0.06	-0.00	-2.09	2.07
7	-0.04	0.02	-0.06	-1.26	0.87	-1.84
8	-0.02	0.02	0.03	-0.74	0.63	0.88
9	-0.00	-0.01	0.03	-0.07	-0.45	1.02
10	-0.03	-0.00	0.04	-1.05	-0.02	1.17
11	0.06	0.03	0.01	1.81	0.85	0.31
12	0.03	0.00	-0.02	0.89	0.20	-0.65
13	-0.02	0.02	0.05	-0.67	0.64	1.44
14	-0.01	-0.03	0.00	-0.33	-0.93	0.20
15	0.02	-0.02	-0.02	0.79	-0.61	-0.80

the purely spatial structure.

The coupled model is employed to build the following point forecasting function by taking conditional expectations at time T to yield,

$$\hat{\underset{\sim}{Z}}_T(\ell) = \begin{cases} (I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}[-{}^c\hat{\theta}_{10}(D)^{1/2}\underset{\sim}{\varepsilon}_T - {}^c\hat{\theta}_{11}(D^{1/2})\underset{\sim}{\varepsilon}_T], & \ell=1 \\ 0 & , \ell\geq 2 \end{cases}$$

$$\hat{\underset{\sim}{Y}}_T(\ell) = \begin{cases} \underset{\sim}{Y}_T + \underset{\sim}{Z}_T(1) & , \ell=1 \\ y_T(\ell-1) + \underset{\sim}{Z}_T(\ell) & , \ell\geq 2\ . \end{cases} \tag{6-29}$$

The forecasting function developed from the STARIMA(0,1,1) model is,

$$\hat{\underset{\sim}{Z}}_T(\ell) = \begin{cases} -\hat{\theta}_{10}(G_{1/2})\underset{\sim}{\varepsilon}_T - \hat{\theta}_{11}W^{(1)}(G'_{1/2})\underset{\sim}{\varepsilon}_T & , \ell=1 \\ 0 & , \ell\geq 2 \end{cases}$$

$$\hat{\underset{\sim}{Y}}_T(\ell) = \begin{cases} y_T - \hat{\theta}_{10}(G'_{1/2})\underset{\sim}{\varepsilon}_T - \hat{\theta}_{11}W^{(1)}(G'_{1/2})\underset{\sim}{\varepsilon}_T, & \ell=1 \\ y_T(\ell-1) + \underset{\sim}{Z}_T(\ell) & , \ell\geq 2. \end{cases} \tag{6-30}$$

Comparing these results, we see that only the forecast function, Equation (6-29), is influended by the contemporaneous spatial structure.

The variances of the point forecasts for both models are given below in Equation (6-31) and for the coupled model and Equation (6-32) for the STARMA model,

$$\mathrm{Var}(\underset{\sim}{Z}_T(\ell)) = \begin{cases} ((I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}D(I - {}^c\hat{\phi}_{01}W^{(1)})'^{-1}\sigma^2, & \ell=1 \\ W_B W_B' \sigma^2 + W_c W_c' \sigma^2, & \ell \geq 2 \end{cases} \tag{6-31}$$

$$\mathrm{Var}(\underset{\sim}{y}_T(\ell)) = \begin{cases} \mathrm{Var}(\underset{\sim}{Z}_T(1)), & \ell=1 \\ \mathrm{Var}(\underset{\sim}{y}_T(\ell-1)) + \mathrm{Var}(\underset{\sim}{Z}_T(\ell)) + \mathrm{cov}_T(\ell) \\ \quad + (\mathrm{cov}_T(\ell))', & \ell \geq 2 \end{cases}$$

where

$$\mathrm{COV}_T(\ell) = \begin{cases} 0, & \ell \leq 1 \\ -W_c W_B' \sigma^2, & \ell \geq 2 \end{cases}$$

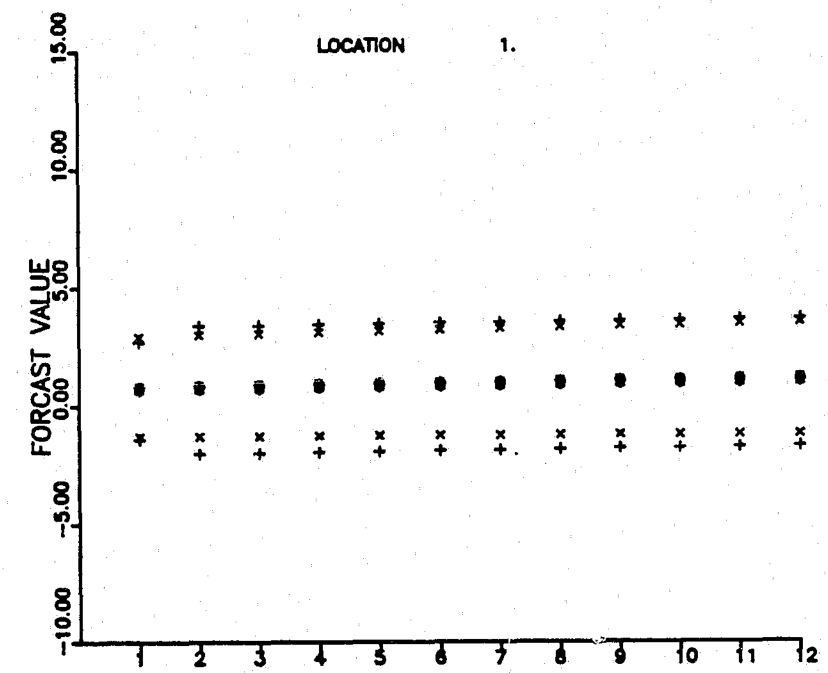
$$W_B = (I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}(D^{1/2})({}^c\hat{\theta}_{10}I + {}^c\hat{\theta}_{11}W^{(1)})$$

$$W_c = (I - {}^c\hat{\phi}_{01}W^{(1)})^{-1}(D^{1/2})$$

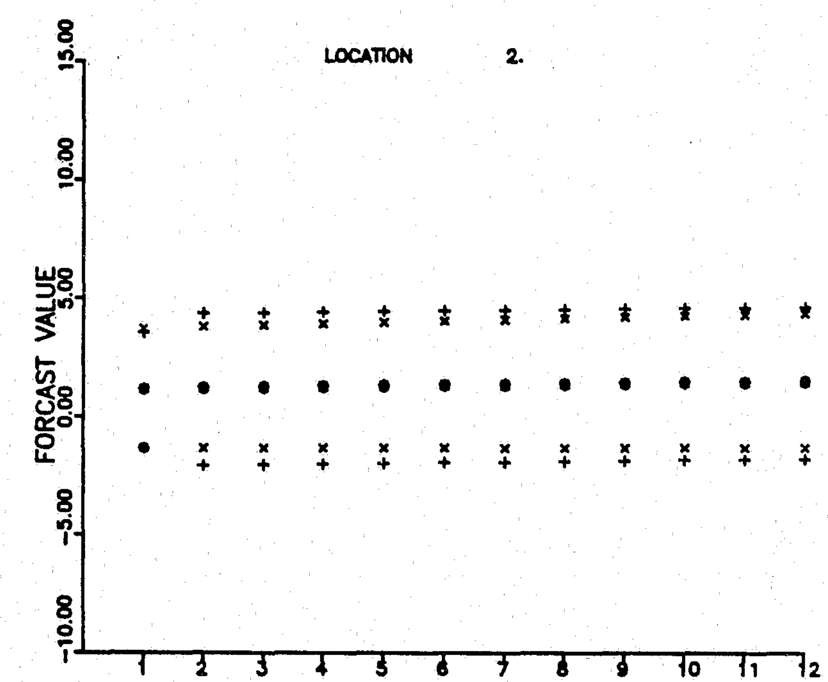
$$\mathrm{Var}(\underset{\sim}{Z}_T(\ell)) = \hat{G} + \sum_{k=1}^{\ell-1} \Lambda(K)G\Lambda(K)', \quad \ell=1 \tag{6-32}$$

$$\Lambda(K) = 0.861I - 0.035W^{(1)}, \quad \ell \geq 2$$

The point forecasts $\underset{\sim}{y}_T(\ell)$ as well as the interval forecasts are computed at T = 72 for ℓ=1,2,...,12 and are plotted in Figures 6-5(a) - (n) for location 1 to location 14, respectively. From these plots, we see that although the point forecasts of both models are close, because they are all unbiased, the 95% confidence intervals of the forecasts of the coupled model always lie inside the corresponding intervals of the STARMA model without contemporaneous terms. The inclusion of the

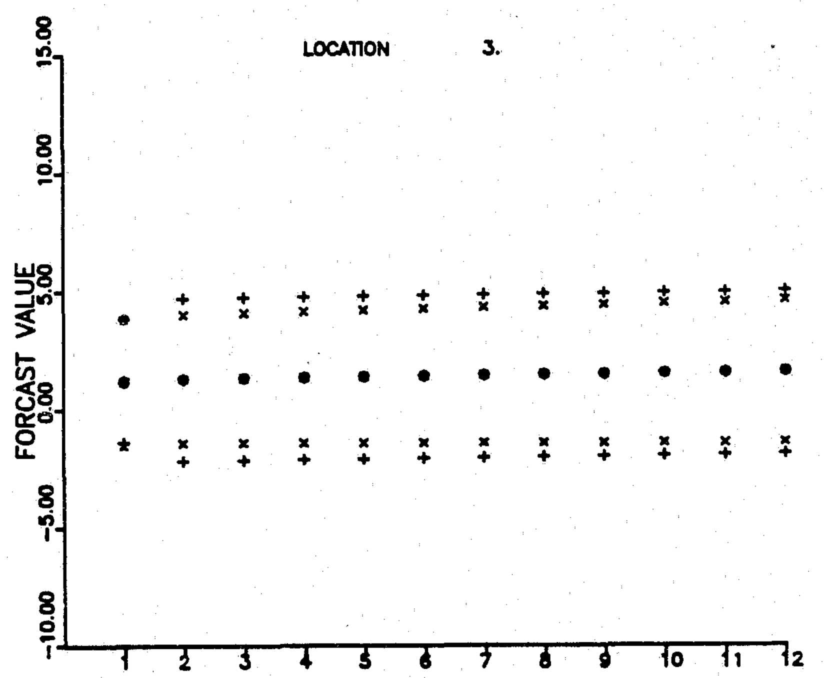


(a)

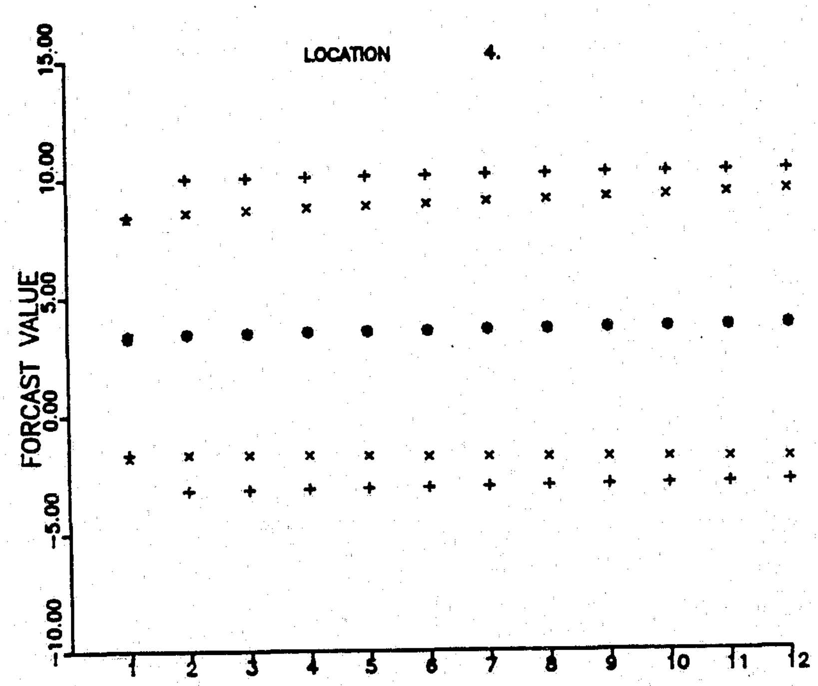


(b)

Figure 6-5. The Point Forecasts and the 95% Confidence Interval Forecasts for the Northeast Boston Assault Arrests.

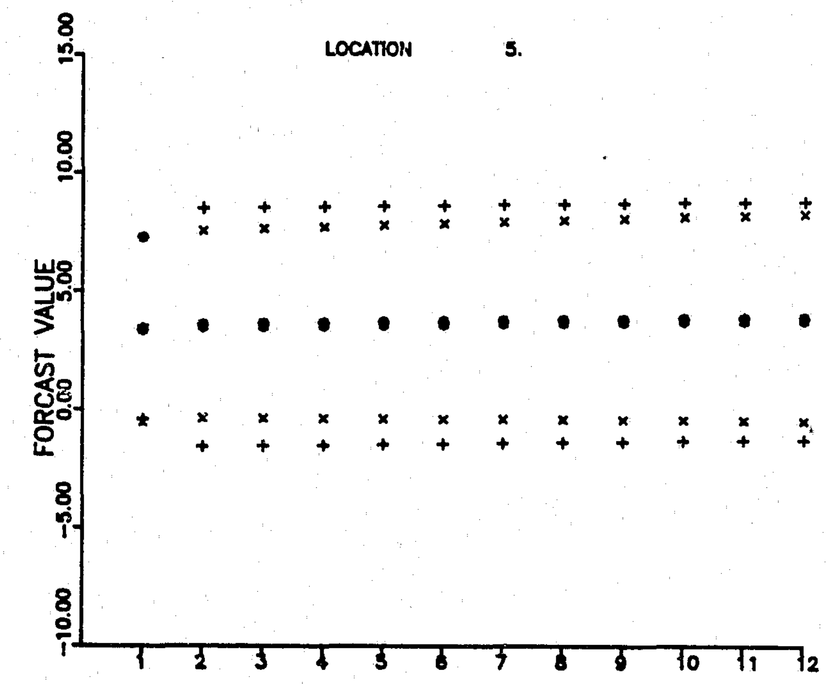


(c)

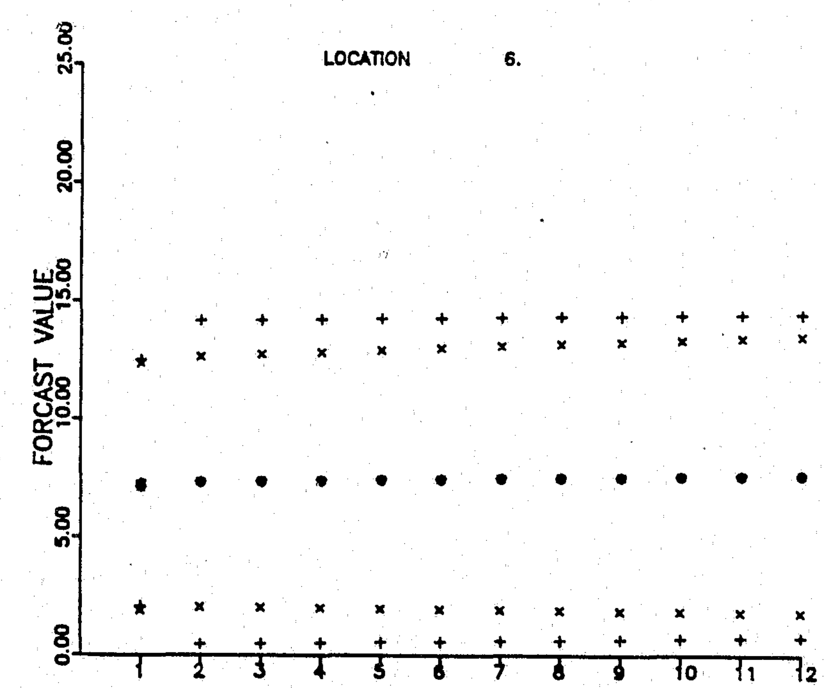


(d)

Figure 6-5. (Cont'd)

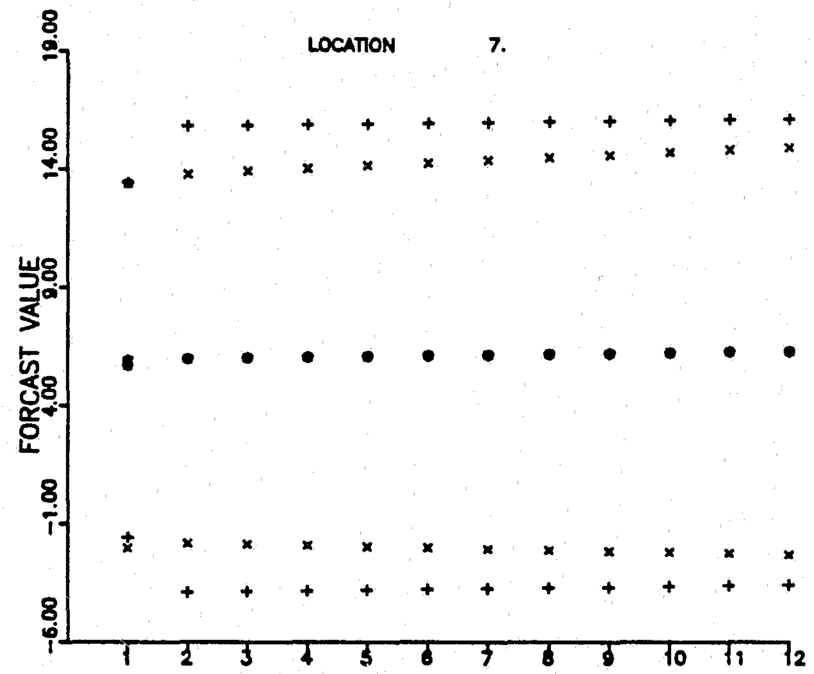


(e)

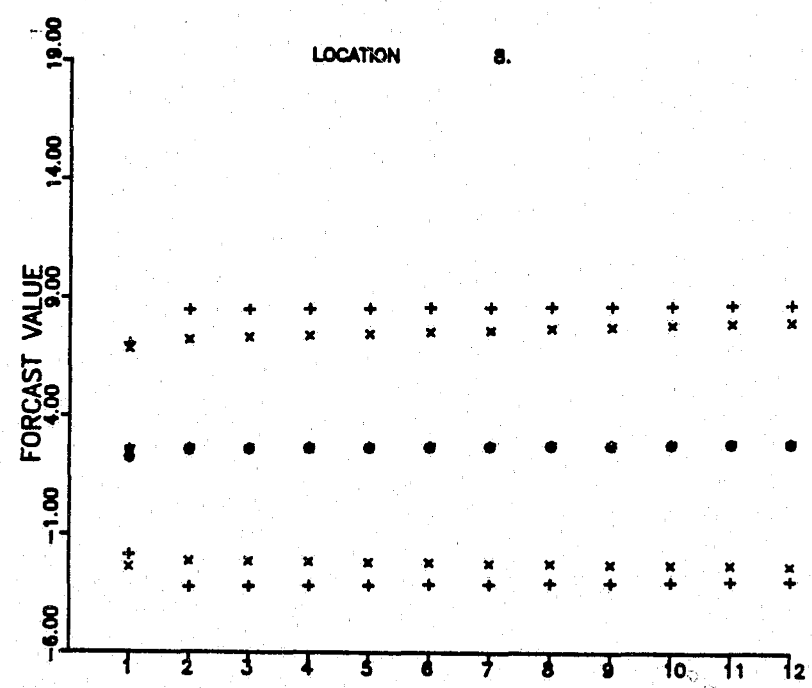


(f)

Figure 6-5. (Cont'd)

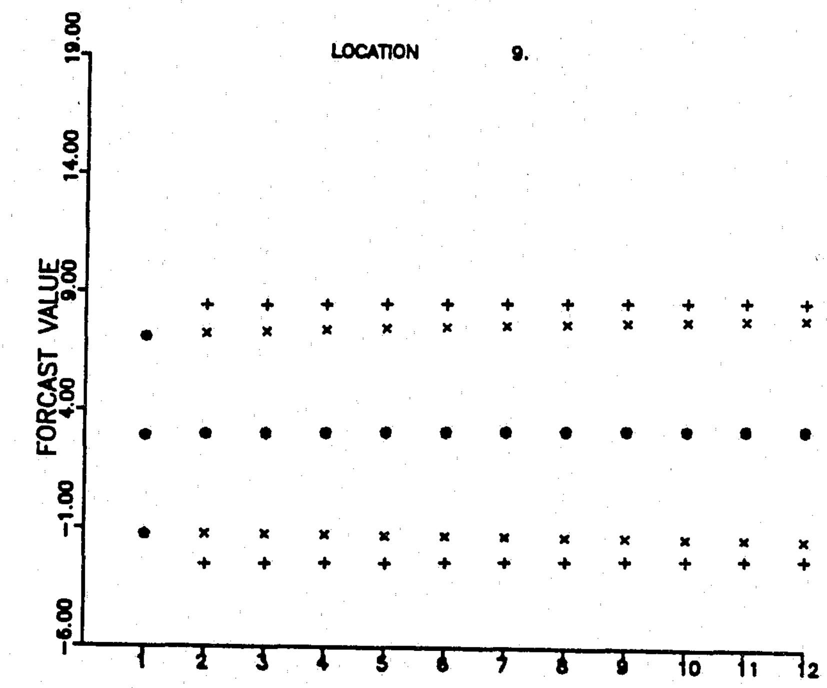


(g)

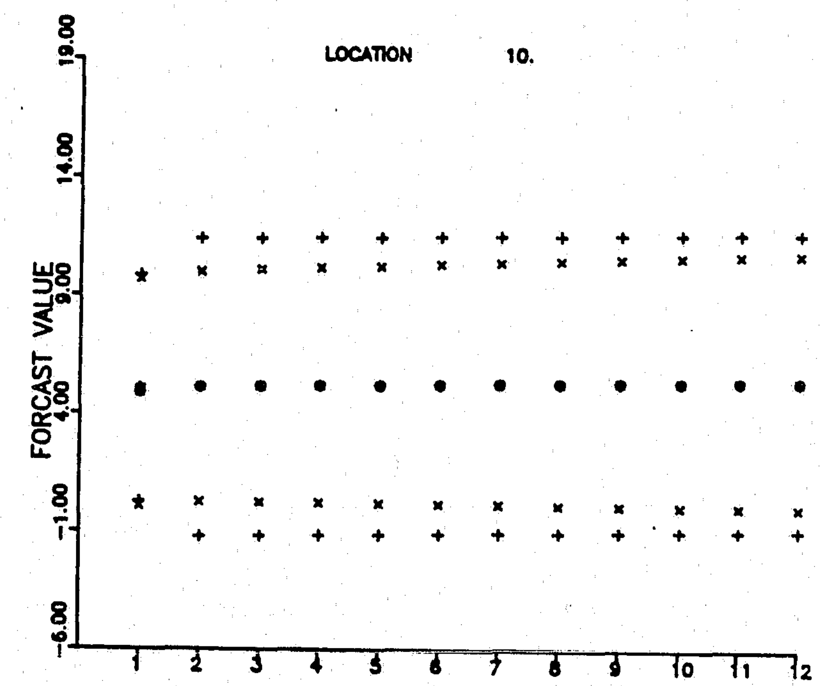


(h)

Figure 6-5. (Cont'd)

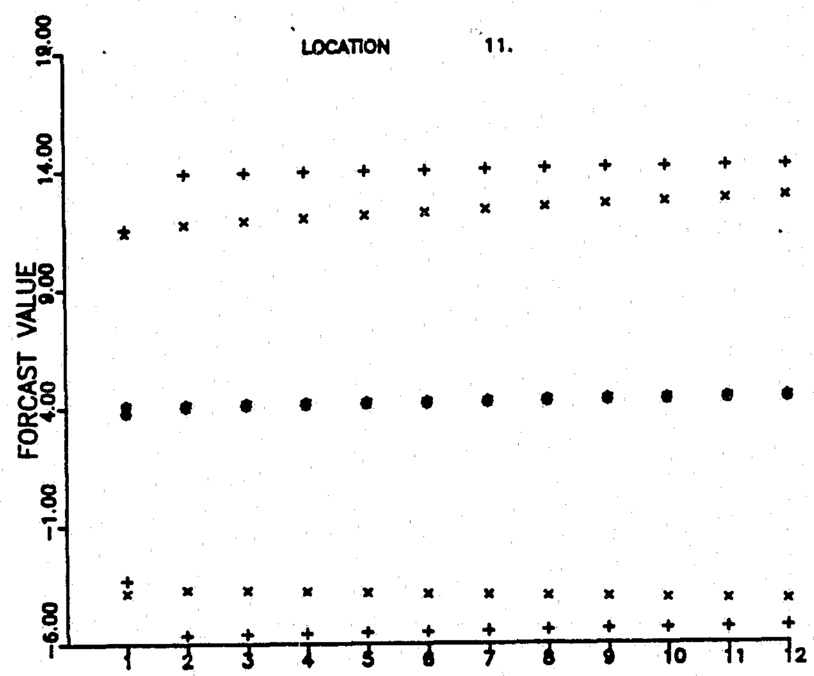


(i)

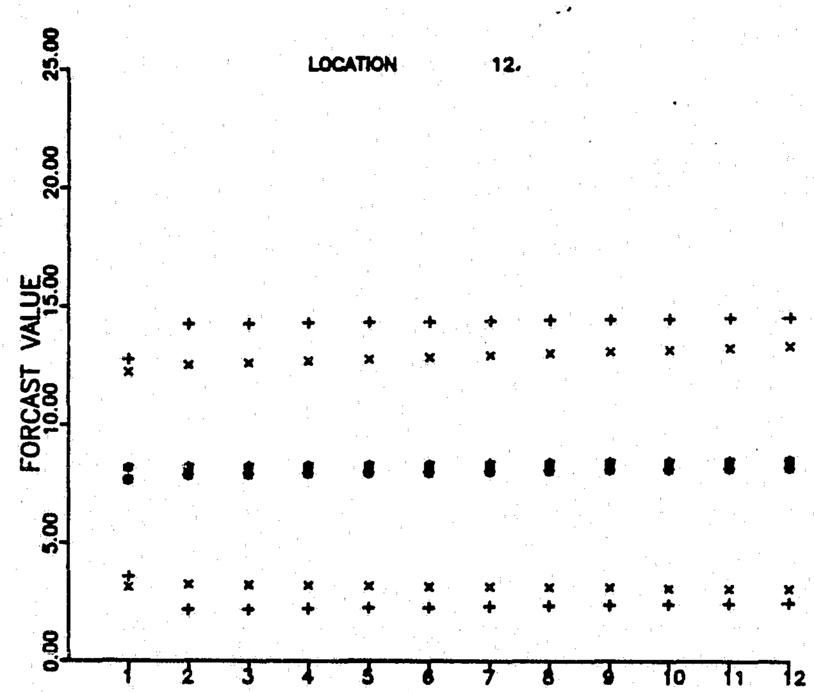


(j)

Figure 6-5. (Cont'd)

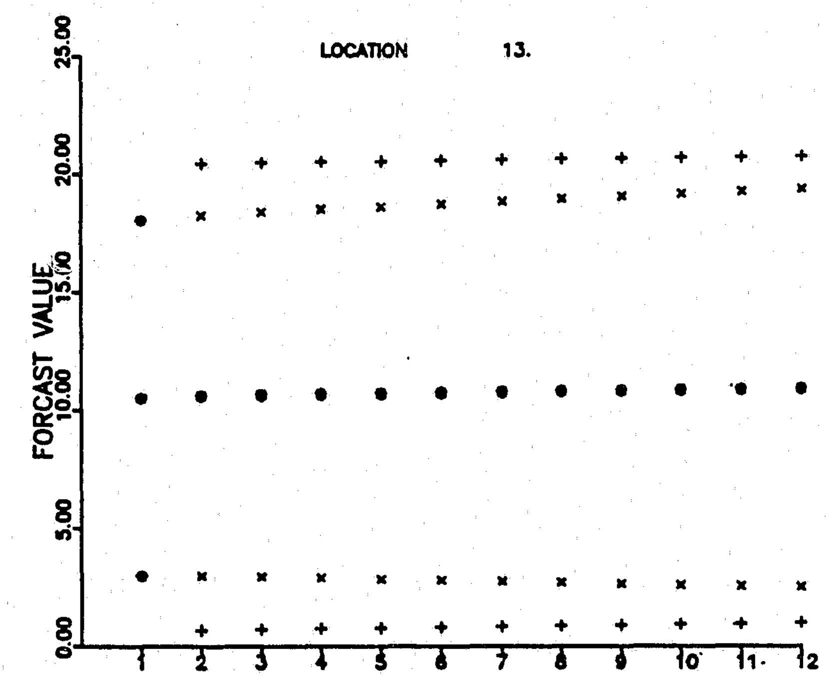


(k)

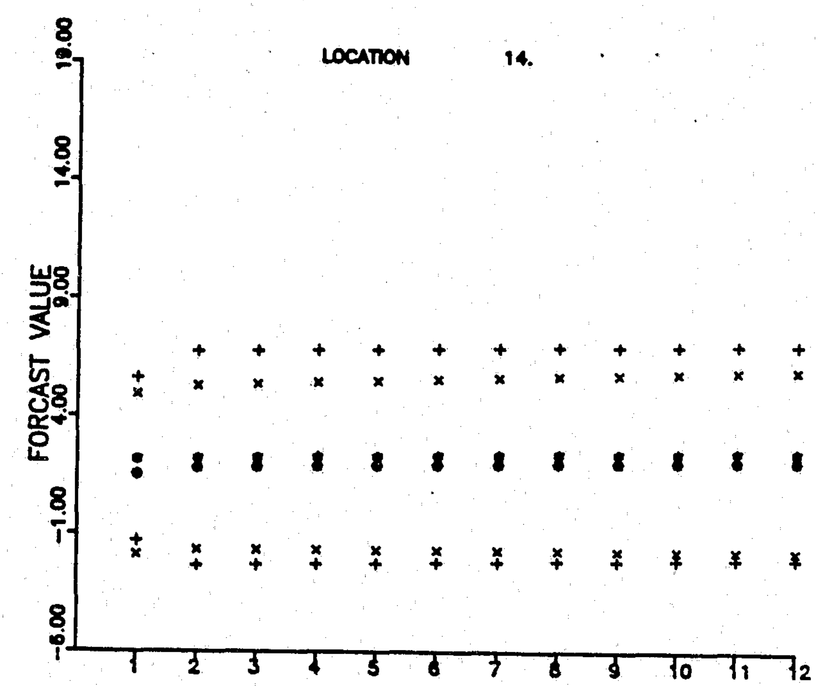


(ℓ)

Figure 6-5. (Cont'd)



(m)



(n)

Figure 6-5. (Cont'd)

contemporaneous spatial information refines the model and improves the forecasting precision.

6.2 Coupling Purely Spatial Models

The purely spatial model doesn't contain any information of temporal correlations among process observations but only contains information about contemporaneous spatial structure. Thus the initial value problem doesn't arise in the purely spatial model estimation procedures. The purely spatial model can be constructed period by period to extract the information of the contemporaneous spatial structures contained in each individual observation period. It also can be constructed for a set of observation periods, that are not necessarily consecutive, to extract the "averaged" information of the contemporaneous spatial correlations for the modeled periods. A system is said to be ergodic if every individual purely spatial model contains the information of the same contemporaneous spatial structures that are contained in the aggregate purely spatial model.

The ergodic process is defined in Section 6.2.1. Here the necessary and sufficient condition for a process to be ergodic are discussed under the assumption of noise homogeneity, and the behaviour of the coupled models are detailed for the ergodic processes as well as the mixed processes. In Section 6.2.2, the procedures of coupling purely spatial models are addressed. Two methods for testing ergodic property under the homogeneity assumption are proposed and these two testing methods are implemented in coupling procedures.

6.2.1 The Ergodic Process

In this section it is assumed that the system contains LN locatins and the observations are available from time t=1 to t=T and the following aggregate and/or individual purely spatial models describe the system,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t \varepsilon U_T, \qquad (6\text{-}33)$$

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0^a} \phi_{0\ell}^a W^{(\ell)} \underset{\sim}{Z}_t - \sum_{\ell=1}^{m_0^a} \theta_{0\ell}^a W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t \varepsilon S_a, \qquad (6\text{-}34)$$

and

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0^b} \phi_{0\ell}^b W^{(\ell)} \underset{\sim}{Z}_t - \sum_{\ell=1}^{m_0^b} \theta_{0\ell}^b W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t \varepsilon S_b, \qquad (6\text{-}35)$$

where

$$S_a \cup S_b = U_T, \quad S_a \cap S_b = \phi, \quad \underset{\sim}{\varepsilon}_t \sim \mathrm{NID}(0, \sigma^2 I),$$

$$\lambda_0,\ \lambda_0^a,\ \lambda_0^b,\ m_0,\ m_0^a,\ m_0^b,\ \phi_{0\ell},\ \phi_{0\ell}^a,\ \phi_{0\ell}^b,\ \theta_{0\ell},\ \theta_{0\ell}^a,\ \theta_{0\ell}^b$$

are model parameters. If the observations $\{\underset{\sim}{Z}_t,\ t \varepsilon S_a\}$ share the same contemporaneous spatial structure with the observations $\{\underset{\sim}{Z}_t,\ t \varepsilon S_b\}$, then the observations $\{\underset{\sim}{Z}_t,\ t \varepsilon U_T\}$ inherit the same structure, and the purely spatial models in Equations (6-33), (6-34) and (6-35) contain the same contemporaneous spatial relations. Therefore, this system is ergodic.

It should be noted that under the homogeneity assumption, the necessary and sufficient condition for the process $\{\underset{\sim}{Z}_t, t\varepsilon U_T\}$ to have ergodic property is that the process parameters do not change during the observation periods $t\varepsilon U_T$, i.e., $\lambda_0 = \lambda_0^a = \lambda_0^b$, $m_0 = m_0^a = m_0^b$, $\phi_{0\ell} = \phi_{0\ell}^a = \phi_{0\ell}^b$, $\theta_{0\ell} = \theta_{0\ell}^a = \theta_{0\ell}^b$ for any non-empty subset S_a, S_b such that $S_a \cdot S_b = \phi$, $S_a \cup S_b = U_T$. Given that the noises $\underset{\sim}{\varepsilon}_t$ are distributed the same for all $t\varepsilon U_T$, if the process parameters do not change, then the distribution of $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$ is the same as that of $\{\underset{\sim}{Z}_t, t\varepsilon U_T\}$, and the observed process is ergodic. Therefore, the unchanged purely spatial parameter condition is the sufficient condition. On the other thand, given that $\{\underset{\sim}{Z}_t, t\varepsilon U_T\}$ is an ergodic process, then $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$ are of the same distribution, i.e.,

$$E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')_{t\varepsilon S_a} = E(\underset{\sim}{Z}_t\underset{\sim}{Z}_t')_{t\varepsilon S_b},$$

or

$$P_a \textstyle\sum_a P_a' = P_b \sum_b P_b'$$

where

$\sum_a$, $\sum_b$ are the covariance matrix of $\{\underset{\sim}{\varepsilon}_t, t\varepsilon S_a\}$ and $\{\underset{\sim}{\varepsilon}_t, t\varepsilon S_b\}$, respectively, and

P_a, P_b are the transfer matrix of the $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$ process, respectively.

Since the noise is distributed homogeneously, so $\sum_a = \sum_t$ and we have $P_a = P_b$ or the process parameters unchanged because S_a and S_b are arbitrary non-empty set such that $S_a \cap S_b = \phi$ and $S_a \cup S_b = U_T$, and U_T may contain arbitrarily many elements, i.e., T may be arbitrarily large.

If the observations $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and the observations $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$ do not share the same contemporaneous spatial structure, then the observations $\{\underset{\sim}{Z}_t, t\varepsilon U_T\}$ still inherit the individual structure but result in a mixed pattern or a compromise between the structure contained in $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and the structure contained in $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$. The overall purely spatial model, Equation (6-33), will thus contain the "averaged" information.

When coupling a set of individual or aggregate purely spatial models, we always obtain an aggregate overall model no matter if the process is ergodic or not. But if the process is not ergodic, then there are infinite such processes that share the same aggregate overall model, since the aggregate overall model contains the "averaged" structures of the coupled models. However, if the ergodic property is imposed, then there is only a set instead of infinite sets of ergodic individual or aggregate processes to be coupled to give the aggregate overall model, since the overall model contains the "averaged" structures of the coupled models that are of the same spatial structure.

6.2.2 The Ergodic Modeling Procedures

It has been made clear in the last section that if the system is ergodic, then the coupled model, Equation (6-33), contains the more precisely estimated information of the purely spatial structure than that contained in the models, Equations (6-34) and (6-35), but the coupled model mixed the information of the purely spatial structures that are contained in the models in Equation (6-34) and (6-35), respectively when they are not ergodic. The quality of the estimated informations of the purely spatial process is not improved except that the ergodic property is held.

In the models, Equations (6-33), (6-34) and (6-35), if $\lambda_0^a = \lambda_0^b = \lambda_0$, $m_0^a = m_0^b = m_0$, $\phi_{0\ell}^a = \phi_{0\ell}^b = \phi_{0\ell}$, $\theta_{0\ell}^a = \theta_{0\ell}^b = \theta_{0\ell}$, absolutely Equations (6-33), (6-34) and (6-35) contain the information of the same structures. However, due to the limited observations and the random nature of stochastic process, this can hardly happen even if the observations $\{Z_t, t\varepsilon S_a\}$ and $\{Z_t, t\varepsilon S_b\}$ are from the same process. The statistical testing procedures are then needed to test the homogeneity between the structures contained in any two arbitrary disjoint observation sets. Two methods, that are the confidence interval method and the χ^2 test method, are proposed for testing the equivalence of models in Equations (6-33), (6-34) and (6-35).

1. The confidence interval method: This method includes the following steps,

   (i) Construct the $100(1-\alpha)\%$ confidence regions R^a and R^b of $(\underset{\sim}{\phi}^a,\underset{\sim}{\theta}^a)$ and $(\underset{\sim}{\phi}^b,\underset{\sim}{\theta}^b)$, respectively.

   (ii) Reject the ergodic hypothesis at significant level α if $R^a \cap R^b = \phi$, and accept the ergodic hypothesis if $R^a \cap R^b \neq \phi$.

Here the confidence regions R^a and R^b can be constructed according to Equation (5-47). This method is especially useful when $\lambda_0 + m_0 \leq 2$, since it is easy to figure out $R^a \cap R^b = \phi$ or not once R^a and R^b are plotted in one or two dimensions.

2. The χ^2 test method: This method computes the W statistics (see Wilks [1938]),

$$W = LN\cdot T\cdot([\ell n\ \hat{\sigma}^2]_{(\underset{\sim}{\phi}^a,\underset{\sim}{\theta}^a)=(\underset{\sim}{\phi}^b,\underset{\sim}{\theta}^b)} - [\ell n\ \hat{\sigma}^2]_{(\underset{\sim}{\phi}^a,\underset{\sim}{\theta}^a)\neq(\underset{\sim}{\phi}^b,\underset{\sim}{\theta}^b)}) \qquad (6\text{-}36)$$

The ergodic hypothesis is rejected when $W > \chi^2_{\alpha,(\lambda_0+m_0)}$ or accepted when $W \leq \sigma^2_{\alpha,(\lambda_0+m_0)}$.

Similar to the fact in set theory that $A \cap B = \phi$, $B \cap C = \phi$ is true does not imply that $A \cap C = \phi$ is true, situations may arise that observations $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$ are ergodic, $\{\underset{\sim}{Z}_t, t\varepsilon S_b\}$ and $\{\underset{\sim}{Z}_t, t\varepsilon S_c\}$ are ergodic, but $\{\underset{\sim}{Z}_t, t\varepsilon S_a\}$ and $\{\underset{\sim}{Z}_t, t\varepsilon S_c\}$ are not ergodic, where $S_a \cap S_b = \phi$, $S_a \cap S_c = \phi$, $S_b \cap S_c = \phi$ are assumed. Based on this consideration and the fact that the individual purely spatial process is the elementary process for testing the ergodic property over the observed periods, the following two alternative schemes are proposed for testing the ergodic property of the observed process.

1. Based on the confidence interval method, this scheme 1 builds T individual purely spatial models. If these T individual models share some common confidence interval, the observed process is accepted as ergodic. Otherwise, the observed process is not accepted as ergodic.
2. Based on the χ^2 test method, this scheme 2 tests the ergodic property for every combination of two individual purely spatial model. If any test fails the ergodic property, the observed processed is rejected to be ergodic. This scheme needs (T)(T-1) tests to accept the hypothesis that the observed process is ergodic.

These T individual purely spatial models are coupled to obtain the aggregate purely spatial model for the observed process if the system is accepted as ergodic. The procedures for building the aggregated purely spatial model have been developed in Chapter V.

The procedures for modeling the aggregated purely spatial models for the observed process are summarized in Figure 6-6.

The construction of purely spatial models exhausts the contemporaneously correlated structures, but nothing has been done for the spatial-temporal correlations. The residuals from the purely spatial model may be modeled as the space-time process. Two situations may arise: 1. The system is not ergodic and invidividual purely spatial models are employed to compute the estimated residuals; 2. The system is ergodic and the aggregated ergodic purely spatial model is built and employed to compute the estimated residuals. For the first

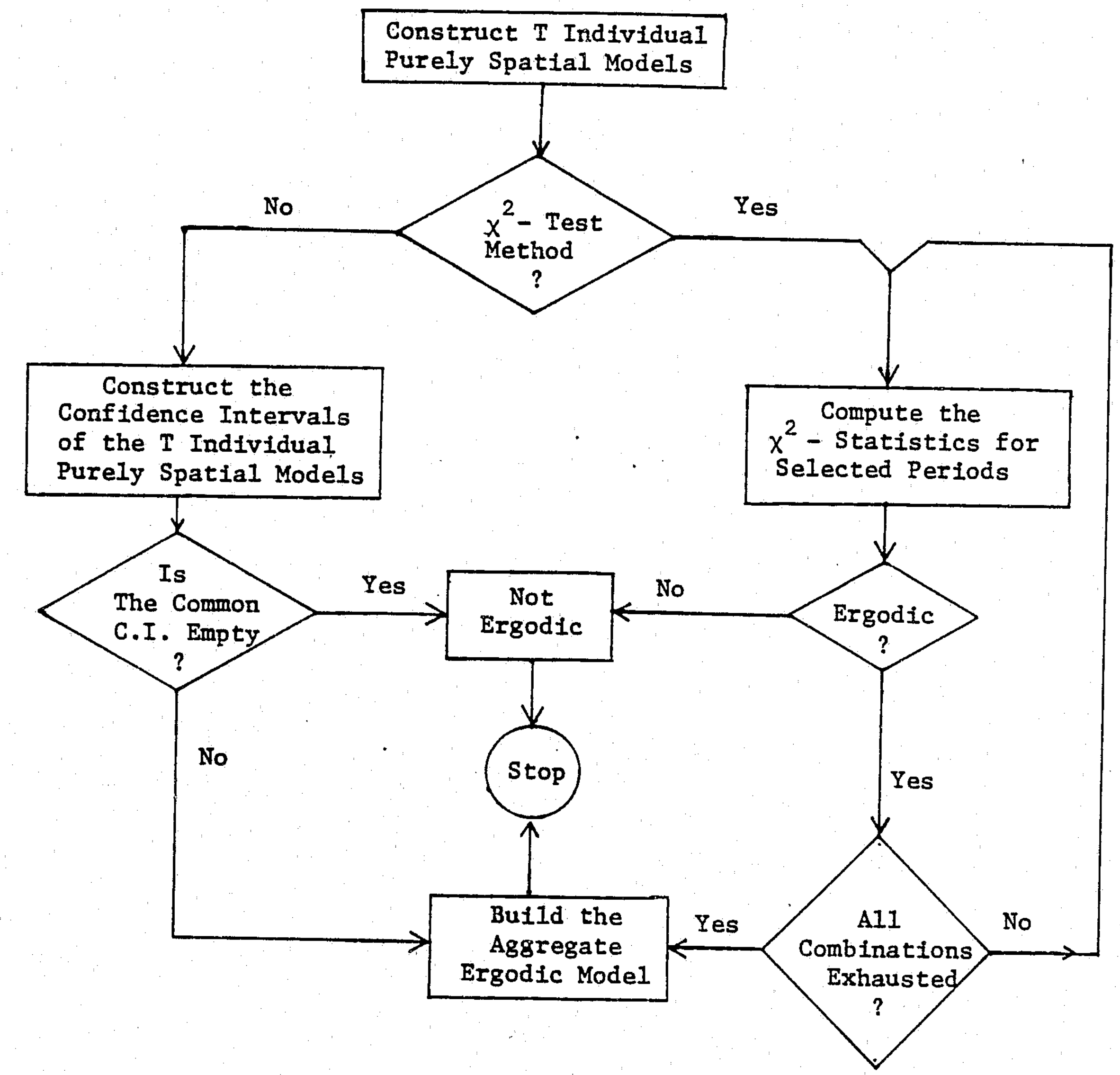


Figure 6-6. The Procedures of Aggregate Ergodic Model Building.

situation, the residual models are constructed and the modeling procedures stop at the multiplicative form of individual purely spatial models and the space-time models. For the second situation, the coupling and reparameterizing procedures, that are detailed in Section 6.1, are applied to construct the space, space-time [ARMA [ARMA(λ_0, m_0) + STARIMA$(\underset{\sim}{p}_\lambda, d, \underset{\sim}{q}_m)$] model or the space, space-time [ARMA(λ_0, m_0) + ARMA $(p,d,q)^{LN}$] model.

6.3 The Ergodic Systems with Outliers

In the models, Equations (6-33), (6-34) and (6-35), it has been assumed that $\underset{\sim}{\varepsilon}_t$, $t \varepsilon U_T$, are identically distributed. This assumption may not be always true, situations may arise that some locatable outliers, that comes from unrecognizable sources, are input to the system and follow the transfer process that the white noises follow. If such outliers exist and are not corrected, then the ergodic property cannot be detected even the true process parameters do not change. The informations of the contemporaneous purely spatial structures that are extracted from the aggregated model as well as from the individual models might be very misleading.

The process with the inputed outliers is formulated as,

$$\underset{\sim}{Z}_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} \underset{\sim}{Z}_t - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} (\underset{\sim}{\varepsilon}_t + \underset{\sim}{\delta}_t) + \underset{\sim}{\varepsilon}_t + \underset{\sim}{\delta}_t \qquad (6\text{-}37)$$

where $\underset{\sim}{\delta}_t$ is the outliers input at time t, $t \varepsilon U_T$. Given the model parameters $\phi_{0\ell}$'s and $\theta_{0\ell}$'s, the M.L. estimates of $\underset{\sim}{\delta}_t$ is

$$\hat{\underset{\sim}{\delta}}_t = (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)}) \underset{\sim}{Z}_t. \qquad (6\text{-}38)$$

It should be noted that the $\hat{\underset{\sim}{\delta}}_t$ is estimated every period, so all the degrees of freedom are used up, and no degrees of freedom will be estimating the noise variance, $\hat{\sigma}^2$. Since $\underset{\sim}{\varepsilon}_t$ and $\underset{\sim}{\delta}_t$ are confounded, so if all the $\hat{\underset{\sim}{\delta}}_t$ are accepted as significant, we have $\hat{\sigma}^2 = 0$ situation.

The process described in Equation (6-37) can be described alternatively in terms of outlier outputs $(\underset{\sim}{\delta})_t$ as follows,

$$\underset{\sim}{Z}_t - (\underset{\sim}{\delta})_t = \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)} (\underset{\sim}{Z}_t - (\underset{\sim}{\delta})_t) - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t. \qquad (6\text{-}39)$$

It should be noted that $(\underset{\sim}{\delta}_t)$, the outlier output, can be expressed in terms of $\underset{\sim}{\delta}_t$, the outlier input, and vice versa.

$$(\underset{\sim}{\delta})_t = (I - \sum_{\ell=1}^{\lambda_0} \phi_{0\ell} W^{(\ell)})^{-1} (I - \sum_{\ell=1}^{m_0} \theta_{0\ell} W^{(\ell)}) \underset{\sim}{\delta}_t \qquad (6\text{-}40)$$

These two modeling alternatives, Equations (6-38) and (6-39), are equivalent but stand at different viewpoints. The formulation, Equation (6-38), finds the roots of process outliers in the input source, and the realized effect of the outlier inputs are influenced by the environments. The formulation, Equation (6-39), detects the process outliers as the realization of outliers themselves and these realizations are not influenced by the environments any more, they just appear as outliers that are realized and imposed on the output observations. In the following analysis, it is based on the formulation in Equation (6-38), however, it should be reminded that this analysis can be based on the formulation in Equation (6-39) as well.

In the following sections, the procedure for detecting the outliers based on the unchanged purely spatial structure assumption is developed in Section 6.3.1, here the modeling sequence for the ergodic process with outliers are discussed in detail. In Section 6.3.2, an example of building the ergodic process by correcting the outliers are given to illustrate the modeling procedures described in Section 6.3.1. The same set of observations are also modeled without the outlier corrections in Section 6.3.2 to serve as a comparison to the outlier corrected ergodic model.

6.3.1 Detecting the Outliers and Modeling the Ergodic Process

A procedure for detecting the outliers is developed and shown in Figure 6-7. Following this procedure, the significant level α for testing the significance of estimated outliers and $\underset{\sim}{\delta}_t = 0$, $t \epsilon U_T$ are set in the very beginning. Then the estimation of the aggregate ergodic purely spatial model are performed to obtain the model parameter estimates $\hat{\phi}_{0\ell}$, $\ell=1,2,\ldots,\lambda_0$, $\hat{\theta}_{0\ell}$, $\ell=1,2,\ldots,m_0$ and $\hat{\sigma}^2$. Three alternatives may be used here for estimating the aggregate ergodic purely spatial model parameters: 1. Build the overall aggregate purely spatial model and accept the model parameters as the aggregate ergodic purely spatial model parameters; 2. Build T individual purely

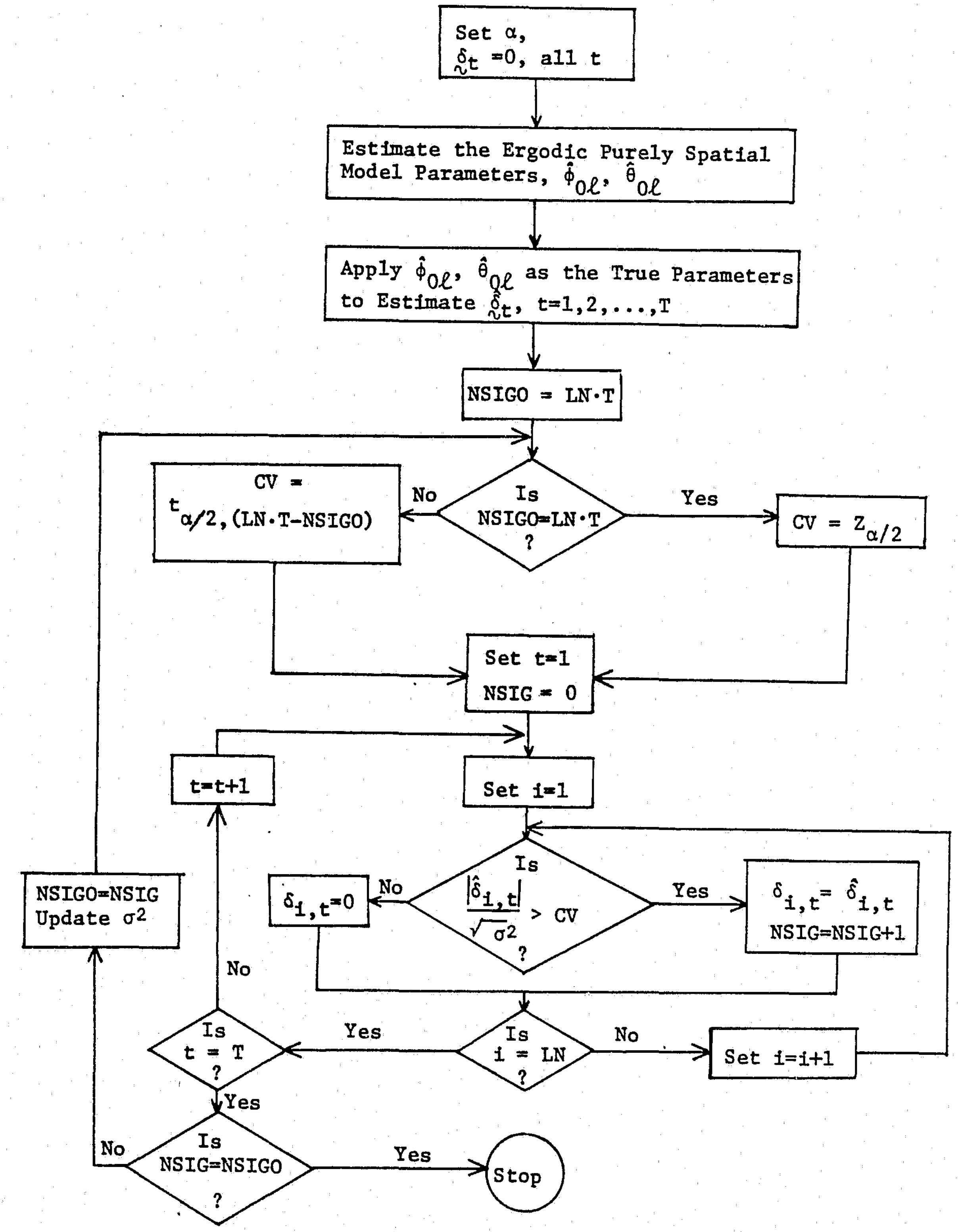


Figure 6-7. The Procedure of Detecting the Outlier Inputs of the Ergodic System.

spatial models and choose the aggregate ergodic purely spatial model parameter estimates from the common confidence region, that is shared by most of the confidence intervals of the individual purely spatial models; 3. Build T individual purely spatial models and then build the aggregated purely spatial models for those periods that share the most popular common confidence intervals, the aggregated model parameters are accepted as the parameters of the overall aggregate ergodic purely spatial models. In most of the cases, the third alternative is recommended, since the estimates obtained here comes from only those periods that are more justifiable to be from the no outliers ergodic processes, and the "averaged" parameter values are estimated to give the aggregate ergodic model parameters. The second alternative allows the property that the more unlike the individual model to the other models, the more contribution it will have in the aggregate ergodic model parameter values, this is contradictive to the intuition that similar individual periods should have more contribution in estimating the aggregate ergodic model parameters. The first alternative is not so attractive in the respect that the non-ergodic individual periods still give contributions to the parameter estimates of the ergodic process. After the parameters are estimated, the estimated $\hat{\phi}_{0\ell}$, $\hat{\theta}_{0\ell}$ are then applied as the true model parameters in Equation (6-38) to estimate $\hat{\underset{\sim}{\delta}}_t$, t=1,2,...,T. Since the estimates of $\hat{\underset{\sim}{\delta}}_t$ have used up all the available degree of freedom so the residual mean square of the overall aggregated purely spatial model, i.e., $\hat{\sigma}^2$, is used as the initial estimate of residual mean square of the model, Equation (7-39). After setting NSIGO = LN•T, the procedure

enters an iterative routine that will exit when the solution converges. Here NSIGO is used to keep record of the number of significant $\hat{\delta}_{i,t}$ of last iteration, and (LN•T - NSIGO) is the degree of freedom that the sum of squares of residuals has. So if NSIGO = LN•T, (LN•T - NSIGO) = 0. no degree of freedom is available for $\hat{\sigma}^2$, thus CV, which is used as the critical value for testing significance of standardized $\hat{\delta}_{i,t}$, is set to $Z_{\alpha/2}$ of normal distribution when NSIGO = LN•T, and CV is set to $t_{\alpha/2,(LN\cdot T-NSIGO)}$ of the student-t distribution when NSIGO < LN•T. The iterative subroutine updates the $\hat{\sigma}^2$ every iteration, use the statistics $|\hat{\delta}_{i,t}/\sqrt{\hat{\sigma}^2}\,|$ to test the significance of $\hat{\delta}_{i,t}$, i.e., the significance of outliers at time t in location i. This procedure exits the subroutine when the number of significant outliers converges. Equation (6-40) is applied to compute the estimated outlier output, $(\hat{\underset{\sim}{\delta}})_t$, from the estimated outlier input, $\hat{\underset{\sim}{\delta}}_t$, and the observations are corrected. Then the procedures described in Figure 6-6 are applied to model the corrected data. If the corrected observations show ergodic property, the aggregate purely spatial model and the space-time residual model are constructed, and the coupling and reparameterizing procedures, that are described in Section 6.1, are applied to construct the space, space-time models. If the corrected observations do not show ergodic property, the space-time model for the residuals is built and the modeling effort stops at the multiplicative form of individual purely spatial models and the space-time model.

6.3.2 An Example Application

In this section eleven periods, t=61 to 71, of the first differenced Northeast Boston Assault Arrests are modeled to illustrate the modeling procedure shown in Figure 6-7 and reveal the needs of correcting the outliers that may mask the purely spatial information. In Section 6.3.2.1, the individual purely spatial models are built to test the ergodic hypotheses, the outliers are then estimated and corrected. The aggregate purely spatial model as well as the space-time model are built for the outlier corrected data to illustrate the ergodic modeling procedure described in Section 6.2. In Section 6.3.2.2, the same set of data are modeled without outlier correction to see the masking effect of the outliers and to serve as a contrast to the appropriate modeling procedures that includes the ergodic tests and the outlier detection, outlier correction steps.

6.3.2.1 Outlier Detection and Ergodic Model Building. In this section, the eleven periods (t=61 to 71) of the Northeast Boston Assault Arrest Data are first modeled as eleven individual purely spatial models. The joint confidence intervals of these individual purely spatial models are plotted and the test of the ergodic property is performed for every consecutive two periods. Then procedures for detecting the noise outliers and building the ergodic models are applied.

Following the purely spatial model building procedure, i.e., the identification, estimation and diagnostic checking, we have the eleven individual purely spatial ARMA models listed in Table 6-3. The 95% C.I.'s and 40% C.I.'s of these 11 individual purely spatial models are plotted in Figure 6-8. In Figure 6-8(b), it is seen that there are two groups of individual models, i.e., S_1 = {t=63, 64, 65, 71} and

Table 6-3. The Individual Purely Spatial Model of the Differenced Northeast Boston Assault Arrests Data

Time	Model Class	ϕ_{01}	ϕ_{02}	ϕ_{03}	$\hat{\sigma}^2$
61	AR(2)	-1.06	-0.19	-	2.62
62	AR(2)	-0.70	-0.34	-	1.68
63	AR(2)	1.23	-0.54	-	1.27
64	AR(3)	0.05	0.10	-1.01	2.62
65	-	-	-	-	3.44
66	AR(1)	-1.12	-	-	1.79
67	AR(2)	-0.28	-0.54	-	0.96
68	AR(2)	-0.92	-0.51	-	1.79
69	AR(2)	-0.81	-0.39	-	1.34
70	AR(1)	-0.72	-	-	1.68
71	AR(2)	0.31	-0.43	-	2.29

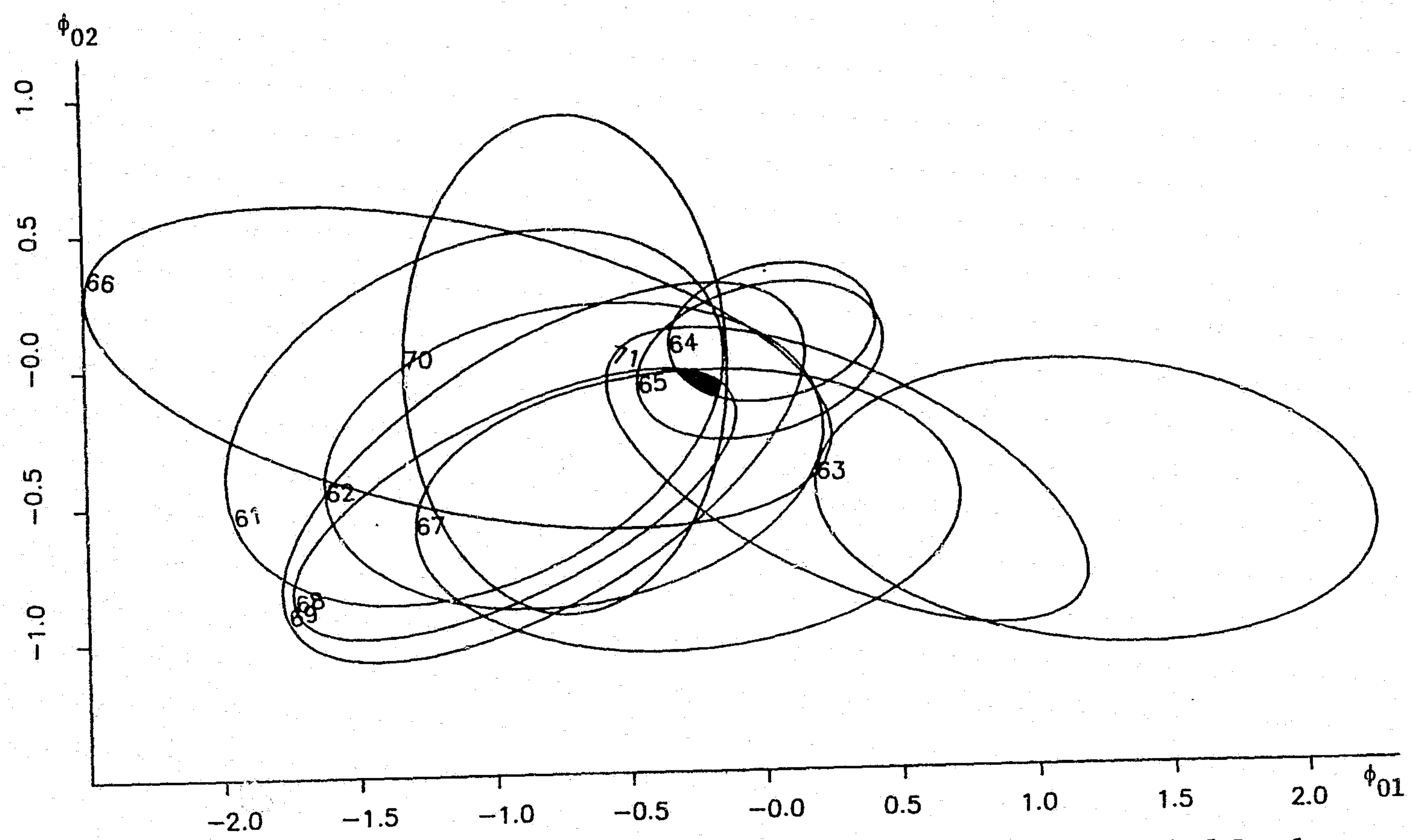


Figure 6-8(a). The 95% Confidence Intervals for the Model Parameters of the Individual Purely Spatial Models (Uncorrected Data).

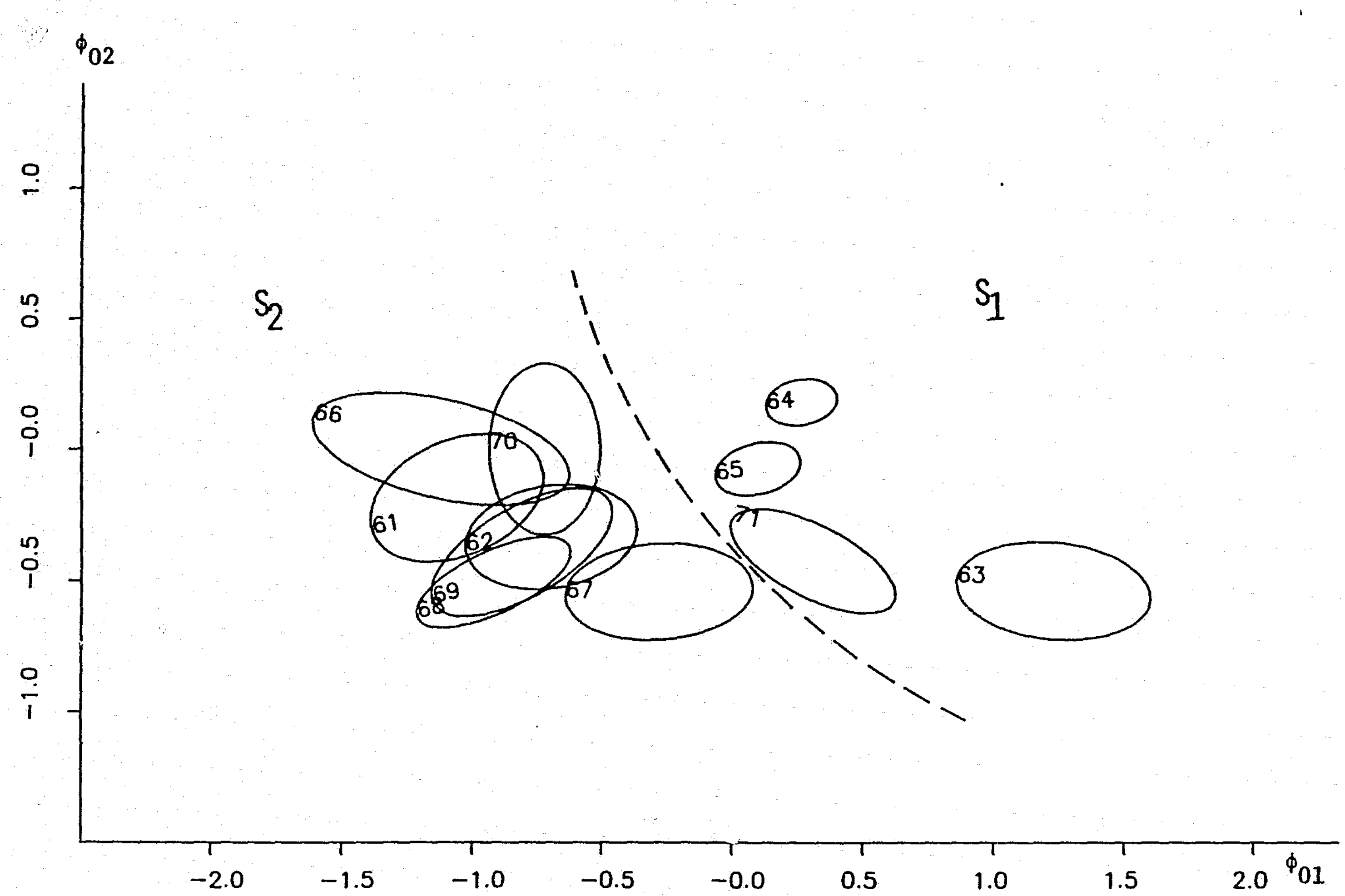


Figure 6-8(b). The 40% Confidence Intervals for the Model Parameters of the Individual Purely Spatial Models (Uncorrected Data).

S_2 = {t=61, 62, 66, 67, 68, 69, 70}, they may contain different purely spatial information. In Figure 6-8(a), we see that 10 out of these 11 models (except t=63) share the darkened region as their common 95% C.I.'s.

Table 6-4 contains the results of ergodic property tests for the consecutive two observation periods. In this table α represents the significant level for these ergodic property tests, the ergodic assumption will be accepted for all the α values that are less than or equal to the values given in the table. It should be noted that the information contained in the confidence interval plots of Figure 6-8 are adequate to decide the ergodic property of the process at the given significant level α, but the information contained in Table 6-4 is not adequate to make the ergodic process conclusion, since it takes all 110 combinations of χ^2 tests to complete the test of ergodic property for eleven periods. However, it is seen that at α = 0.01 all these tests pass and the ergodic assumption for the consecutive two observations are accepted. These tests are consistent with the information of the ergodic property that are contained in the 95% confidence interval plots in Figure 6-8.

In Figure 6- 8 (a), we see that 10 individual purely spatial models out of 11 models, except t=63, share the common darkened region in their 95% confidence interval. In Figure 6-8 (b), we see two model groups, i.e., S_1 and S_2. These two groups may contain different purely spatial information. Since the input outliers may mask the purely spatial information, one of these two model groups may contain the ergodic purely spatial models, while the other model group con-

Table 6-4. The Summarized Results of the Tests of Ergodic Property for Two Consecutive Observations

Time Index	Aggregated Model Parameters ϕ_{01}	ϕ_{02}	χ^2 Statistics	α
61.62	-0.89	-0.27	0.230	0.891
62.63	1.25	-0.36	8.507	0.014
63.64	0.33	0.06	5.464	0.065
64.65	0.21	0.07	0.967	0.616
65.66	0.04	-0.12	2.210	0.331
66.67	-0.45	-0.35	3.291	0.193
67.68	-0.71	-0.47	1.609	0.447
68.69	-0.85	-0.46	0.089	0.956
69.70	-0.34	0.14	0.367	0.832
70.71	-0.40	-0.13	3.428	0.180

tains the masked purely spatial information. Comparing these two groups in Figures 6-8 (a) and (b), we may expect that models in the S_1 group contain the masked purely spatial information. In the following, the procedure shown in Figure 6-7 is applied to estimate the input outliers. The output outliers are computed by applying Equation (6-40) and the observations are corrected. Then the procedures of aggregate ergodic model building, that are shown in Figure 6-6, are applied to model the corrected observations.

Applying the procedures described in Figure 6-7, using the observations $\underset{\sim}{Z}_t$, $t \varepsilon S_2$ to build the aggregate purely spatial model to obtain the estimated ergodic process parameters (the method 3 described in Section 6.3.1) we have $\hat{\phi}_{01} = -1.0432$, $\hat{\phi}_{02} = -0.2273$ and the estimates of the input outliers $\hat{\underset{\sim}{\delta}}_t$ that are listed in Table 6-5. The output outliers for the observation corrections are evaluated according to Equation (6-40), i.e.,

$$(\hat{\underset{\sim}{\delta}})_t = (I - \sum_{\ell=1}^{2} \hat{\phi}_{0\ell} W^{(\ell)})^{-1} \hat{\underset{\sim}{\delta}}_t \qquad (6\text{-}41)$$

where

$\hat{\phi}_{01} = -1.043$, $\hat{\phi}_{02} = -0.227$

$\hat{\underset{\sim}{\delta}}_t$ is the estimated input outlier, and

$(\hat{\underset{\sim}{\delta}})_t$ is the output outliers that are used in data correction. The evaluated $(\hat{\underset{\sim}{\delta}})_t$ values are listed in Table 6-6. It should be noted that in Table 6-5, we see that 14 out of 17 detected noise outliers are found in $\{\underset{\sim}{Z}_t,\ t \varepsilon S_1\}$.

Table 6-5. The Estimated Input Outliers of Northeast Boston Assault Arrests

	Locations													
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	-4.0	0	0	0	3.2	0	0
64	-6.9	-5.2	-5.7	0	0	0	0	0	0	0	0	0	3.5	4.4
65	0	0	3.1	0	0	0	0	0	0	0	0	-3.6	-4.6	-5.1
66	0	0	0	0	3.2	0	0	0	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68	-3.9	0	0	0	0	0	0	0	0	0	0	0	0	0
69	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	-3.2	0	0	0	0	0	0	0	0	0
71	3.6	3.3	0	0	0	0	0	0	0	-3.3	0	0	0	0

Table 6-6. The Observation Correction Values for the Northeast Boston Assault Arrests

	Locations													
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	0.4	-1.1	-0.3	1.7	-4.7	-0.6	5.8	-11.3	4.8	0.4	-7.8	14.2	-4.9	0
64	0.4	7.3	-8.6	4.2	-2.3	2.0	-0.8	3.2	-2.6	-6.1	9.1	-14.4	20.3	-0.4
65	-3.9	1.6	6.2	-1.3	1.5	-1.6	-0.3	-2.5	3.2	5.2	-2.1	2.3	-20.2	-1.1
66	-1.3	1.7	1.6	-2.4	8.5	-1.8	-1.0	-3.2	1.3	0	0	-0.2	0.2	-0.8
67	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68	-6.9	3.3	2.0	0	0.2	-0.6	0.2	0	0	0	0	0	0	-0.1
69	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70	-2.4	-0.6	4.5	2.0	2.3	-12	3.4	-2.3	1.5	-1.5	0.7	-0.5	0.8	-1.0
71	2.9	3.4	-1.4	-1.0	-0.4	0.3	1.7	0.5	4.6	-13.7	9.8	-6.3	7.0	-3.6

After correction for outliers, the corrected data are modeled by applying the procedures described in Figure 6-6. Following the purely spatial ARMA model building procedure, we have the 11 individual purely spatial models that are listed in Table 6-7. Six of these eleven purely spatial models are purely spatial AR(2) models and the remaining five are purely spatial AR(1) models. Comparing the model classes listed in Table 6-7 with those listed in Table 6-3, we see that the model forms as well as the model parameters are more consistent in Table 6-7 than those listed in Table 6-3. The 95% and 40% confidence intervals of these individual purely spatial model parameters are plotted in Figure 6-9. In Figure 6-9(a), we see that all these models share the marked parameter region in their 95% C.I.'s. From Figure 6-9(b), it is even clearer that these purely spatial models are from the same purely spatial process, i.e., the corrected observations are from an ergodic process. Comparing the confidence intervals in Figure 6-9 with those in Figure 6-8, we see that the outlier corrected data are ergodic, while the uncorrected data are not, since in Figure 6-9 the 95% confidence intervals share the same marked region as their common confidence interval, while no such common confidence region exist in Figure 6-8.

These individual purely spatial models show ergodic property, the aggregated purely spatial model is expected to contain the same purely spatial structures as those contained in the individual models. To construct the aggregate purely spatial model, we compute the purely spatial autocorrelation functions,

Table 6-7. The Individual Purely Spatial Model of the Outliers Corrected Northeast Boston Assault Arrests

Time	Model Class	ϕ_{01}	ϕ_{02}	$\hat{\sigma}^2$
61	AR(2)	-1.06	-0.19	1.082
62	AR(2)	-0.70	-0.34	1.628
63	AR(2)	-1.07	-0.39	1.377
64	AR(1)	-1.10	-	1.683
65	AR(2)	-0.92	0.14	1.864
66	AR(1)	-0.80	-	1.060
67	AR(2)	-0.28	-0.54	0.958
68	AR(1)	-0.92	-	1.375
69	AR(2)	-0.81	-0.39	1.338
70	AR(1)	-0.56	-	1.055
71	AR(1)	-0.88	-	1.457

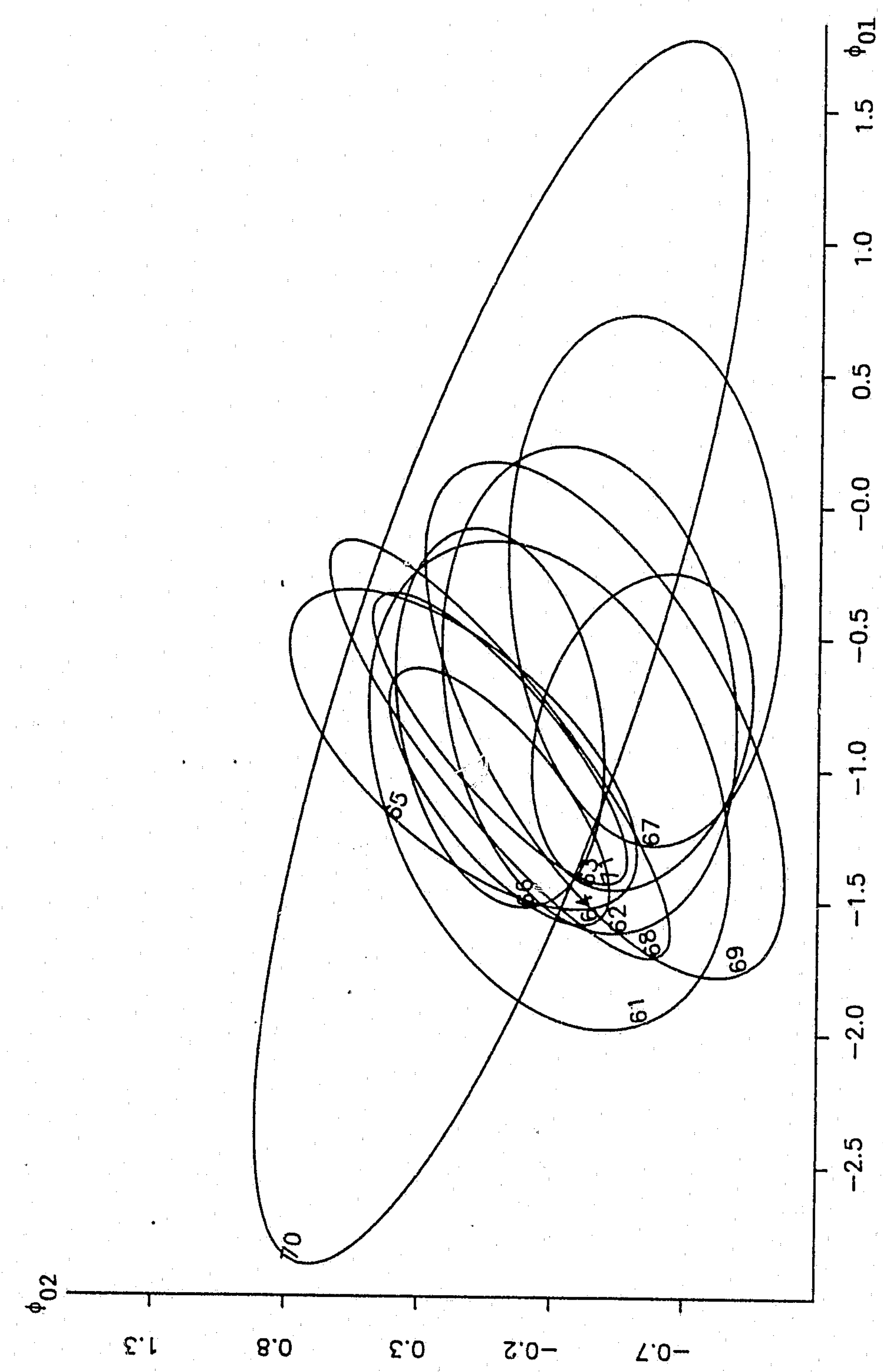


Figure 6-9(a). The 95% Confidence Intervals for the Model Parameters of the Individual Purely Spatial Models (Outlier Corrected).

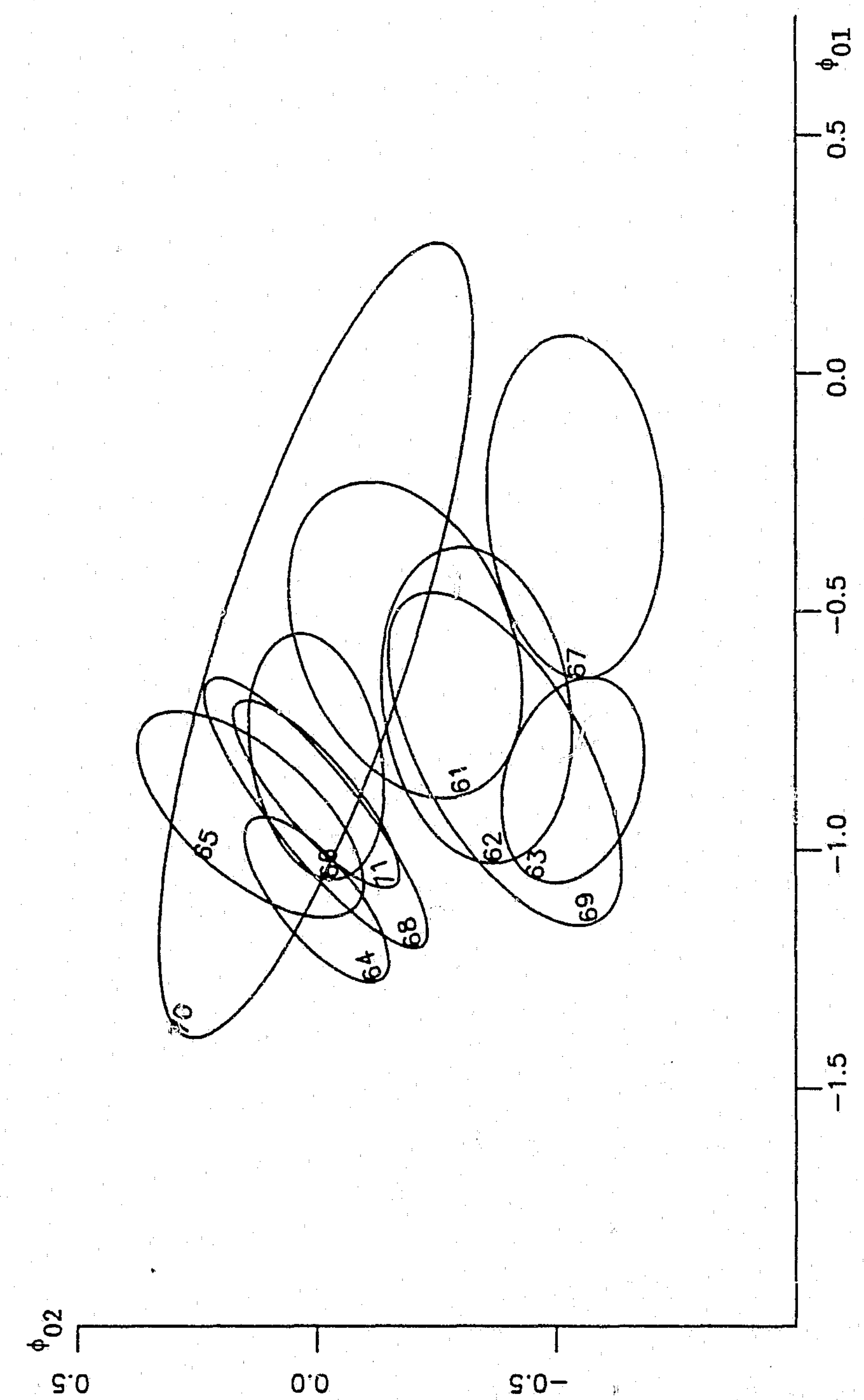


Figure 6-9(b). The 40% Confidence Intervals for the Model Parameters of the Individual Purely Spatial Models (Outlier Corrected).

$$\hat{\rho}_{01} = -0.605,\ \hat{\rho}_{02} = 0.045,\ \hat{\rho}_{03} = 0.307,$$

and the standardized purely spatial autocorrelation functions,

$$(\hat{\rho}_{01}) = -7.501,\ (\hat{\rho}_{02}) = 0.558,\ (\hat{\rho}_{03}) = 3.812.$$

The pattern of purely spatial autocorrelation functions suggests the purely spatial AR(2) candidate model with $\phi_{01} = -.80$, $\phi_{02} = -1.5$. The M.L. estimation gives $\hat{\phi}_{01} = -1.08$, $\hat{\phi}_{02} = -0.02$. The extra sums of squares associated with $\hat{\phi}_{01}$ and $\hat{\phi}_{02}$ are 270.8 and 18.38, respectively, the F-statistics associated with $\hat{\phi}_{01}$ and $\hat{\phi}_{02}$ are 188.2 and 12.6, respectively, and ϕ_{01}, ϕ_{02} are significant at significant level $\alpha = 0.01$. The sample space-time autocorrelation functions and the sample space-time partial autocorrelation functions, that are listed in Table 6-8, suggest the STMA(10) model for the residuals. The M.L. estimation gives

$$\hat{\theta}_{10} = 0.254,\ 95\%\ \text{C.I.}\ (0.092, 0.416)$$

The sample space-time autocorrelation functions of the residuals, that are listed in Table 6-9, do not show any model inadequacies and the model, Equation (6-42)

$$\begin{cases} \underset{\sim}{Z}_t = \hat{\phi}_{01} W^{(1)} \underset{\sim}{Z}_t + \hat{\phi}_{02} W^{(2)} \underset{\sim}{Z}_T + \underset{\sim}{a}_t \\ \underset{\sim}{a}_t = -\theta_{10} \underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t \end{cases} \qquad (6\text{-}42)$$

Table 6-8(a). The Sample Space-Time Autocorrelation Functions of the Residuals of the Aggregated Purely Spatial Model of the Outlier Corrected Data

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
0	1.00	0.07	0.00	1.00	0.89	0.00
1	-0.30	-0.08	-0.00	-3.25	-1.03	-0.08
2	0.12	0.05	0.06	1.21	0.58	0.68
3	0.05	-0.00	0.06	0.47	-0.02	0.57
4	0.11	0.19	-0.10	0.88	1.61	-0.88
5	-0.13	-0.02	0.20	-1.00	-0.15	1.52

Table 6-8(b). The Sample Space-Time Autocorrelation Functions of the Residuals of the Aggregated Purely Spatial Model of the Outlier Corrected Data

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
1	-0.30	0.09	0.01	-3.64	1.08	0.17
2	0.02	0.06	0.11	0.32	0.71	1.28
3	0.10	-0.05	0.18	1.09	-0.57	1.93
4	0.17	0.22	-0.23	1.75	2.24	-2.35
5	-0.12	0.24	0.12	-1.10	2.27	1.12

Table 6-9. The Sample Space-Time Autocorrelation Functions of the Residuals of the Model, Equation (6-46)

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
0	1.00	0.02	0.01	1.00	0.33	0.16
1	-0.03	0.03	0.04	-0.38	0.38	0.45
2	0.12	0.07	0.09	1.20	0.75	0.89
3	0.07	0.04	0.05	0.70	0.38	0.50
4	0.08	0.14	-0.03	0.76	1.18	-0.27
5	-0.06	0.01	0.10	-0.45	0.11	0.79

is accepted as adequate. Reparameterizing the models in Equation (6-42) and performing the M.L. estimation, we obtain the model, Equation (6-43),

$$\underset{\sim}{Z}_t = \hat{\phi}_{01}W^{(1)}\underset{\sim}{Z}_t + \hat{\phi}_{02}W^{(2)}\underset{\sim}{Z}_t - \hat{\theta}_{10}\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t \tag{6-43}$$

where

$$\hat{\phi}_{01} = -1.086$$

$$\hat{\phi}_{02} = -0.206$$

$$\hat{\theta}_{10} = 0.255$$

$$\hat{\sigma}^2 = 1.352$$

The sample space-time autocorrelation functions of the residuals, that are listed in Table 6-10, show adequacy and the model, Equation (6-43), is accepted as adequate. The model described by Equation (6-43) contains purely spatial structures that are contained in the aggregate purely spatial model as well as the individual purely spatial models.

6.3.2.2 Consequences of Ignoring the Homogeneity Assumption. In the last section we have the Northeast Boston Assault Arrests model by applying the procedures in Figure 6-7. It has been concluded that the process is an ergodic process with the estimated input outliers listed in Table 6-5. An interesting problem may arise that what will happen in the modeling if the input outliers are not corrected. In this section, we will ignore the presence of outliers which negate

Table 6-10. The Sample Space-Time Autocorrelation Functions of the Residuals of the Coupled Model for the Outlier Corrected Data

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelation

Space Lag / Time Lag	0	1	2	0	1	2
0	1.00	0.03	0.01	1.00	0.38	0.14
1	-0.03	0.04	0.04	-0.39	0.46	0.52
2	0.14	0.09	0.10	1.47	0.94	1.07
3	0.10	0.05	0.07	0.96	0.54	0.69
4	0.14	0.22	-0.05	1.22	1.88	-0.42
5	-0.11	0.02	0.19	-0.82	0.20	1.46

the homogeneity assumption, build the models for the uncorrected observations and compare the resulting model to the outlier corrected model.

Ignoring that the ergodic property doesn't hold in Figure 6-8, we build the aggregate purely spatial model for the eleven observation periods. The purely spatial autocorrelation functions and the standardized purely spatial autocorrelation functions were computed. Since for the purely spatial autocorrelation functions, $\hat{\rho}_{01} = -0.2463$, $\hat{\rho}_{02} = -0.2069$, $\hat{\rho}_{03} = 0.1418$ with corresponding standardized values of $(\hat{\rho}_{01}) = -2.19$, $(\hat{\rho}_{02}) = -1.90$, $(\hat{\rho}_{03}) = 1.346$, purely a spatial AR(2) model with negative ϕ_{01}, ϕ_{02} values is suggested. The initial guesses $(\phi_{01},\phi_{02}) = (-.25,-.20)$ are read from the contours of Figure 5-10(a) in Section 5.3. The M.L. estimation gives

	95% C.I.	SS_E	F-Stat.	α
$\hat{\phi}_{01} = -0.165$	(-0.381, 0.051)	5.537	2.267	0.134
$\hat{\phi}_{02} = -0.126$	(-0.291, 0.039)	7.721	3.161	0.077

$\hat{\sigma}^2 = 2.4924$

where

SS_E is the extra sum of squares given that the other parameters are in the fitted model, and

α is the significant level.

It should be noted that $(\hat{\phi}_{01},\hat{\phi}_{02}) = (-0.165,-0.126)$ fall in the rectangular region $\{(\phi_{01},\phi_{02})\ |-0.4 \le \phi_{01} \le -0.1,\ -0.3 \le \phi_{02} \le -0.05\}$

that contains the darkened region in Figure 6-8, as might be expected. The information of the purely spatial structure contained in this aggregate purely spatial model reflects the averaged structure that is contained in the individual purely spatial models.

The sample space-time autocorrelation functions and the sample space-time partial autocorrelation functions of the residuals from this aggregated purely spatial model were computed and are listed in Table 6-11. The cut-off behavior in the sample space-time autocorrelations and the tail-off behavior in the sample space-time partial autocorrelations suggest the STMA(1_0) model to be the potential candidate. M.L. estimation gives,

$\hat{\theta}_{10} = 0.807$ 95% C.I. = (0.702, 0.911)

$\hat{\sigma}^2 = 1.521$

The sample space-time autocorrelation function of these residuals, listed in Table 6-12, does not reveal any unexhausted structure and this model

$$\underset{\sim}{Z}_t = \hat{\phi}_{01}W^{(1)}\underset{\sim}{Z}_t + \phi_{02}W^{(2)}\underset{\sim}{Z}_t + \underset{\sim}{a}_t$$

$$\underset{\sim}{a}_t = -\hat{\theta}_{10}\underset{\sim}{\varepsilon}_t + \underset{\sim}{\varepsilon}_t, \quad t=1,2,\ldots,11 \qquad (6\text{-}44)$$

where

Table 6-11(a). The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Aggregated Purely Spatial Model

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
0	1.00	0.03	0.00	1.00	0.39	0.05
1	-0.53	0.05	0.00	-5.68	0.63	0.05
2	0.04	-0.10	-0.01	0.49	-1.04	-0.14
3	-0.07	-0.04	-0.10	-0.72	-0.41	-0.93
4	0.21	0.11	0.09	1.81	0.98	0.75
5	-0.24	-0.00	-0.01	-1.83	-0.00	-0.07

Table 6-11(b). The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Aggregated Purely Spatial Model

The Sample Space-Time Partial Autocorrelations/
The Standardized Sample S-T Partials

Space Lag / Time Lag	0	1	2 /	0	1	2
1	-0.53	0.16	0.04	-6.35	1.90	0.57
2	-0.35	0.05	0.04	-3.99	0.59	0.48
3	-0.36	-0.07	-0.09	-3.81	-0.80	-0.96
4	0.02	0.15	-0.11	0.24	1.48	-1.10
5	-0.20	0.19	-0.10	-1.91	1.79	-0.94

Table 6-12. The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Model, Equation (6-44)

The Sample Space-Time Autocorrelation/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
0	1.00	-0.05	0.00	1.00	0.64	0.10
1	-0.04	0.05	0.02	-0.45	0.59	0.22
2	0.02	-0.08	-0.05	0.21	-0.87	-0.49
3	-0.01	-0.06	-0.09	-0.10	-0.57	-0.84
4	0.11	0.05	0.03	0.99	0.44	0.27
5	-0.07	-0.00	0.00	-0.53	-0.03	0.06

$$\hat{\phi}_{01} = -0.165,\ \hat{\phi}_{02} = -0.126,$$

$$\hat{\theta}_{10} = 0.807,\ \ \hat{\sigma}^2 = 1.521,$$

is accepted as adequate. The purely spatial AR(2) model and the space-time STMA(1_0) model is then coupled and reparameterized to give,

$$\underset{\sim}{Z}_t = \hat{\phi}_{01}W^{(1)}\underset{\sim}{Z}_t + \hat{\phi}_{02}W^{(2)}\underset{\sim}{Z}_t - \hat{\theta}_{10}\underset{\sim}{\varepsilon}_{t-1} + \underset{\sim}{\varepsilon}_t,$$

$$t=61,62,\ldots,71 \qquad (6\text{-}45)$$

where

$$\hat{\phi}_{01} = -0.083$$

$$\hat{\phi}_{02} = -0.105$$

$$\hat{\theta}_{10} = 0.802$$

$$\hat{\sigma}^2 = 1.475$$

The extra sum of squares associated with ϕ_{01} is 0.838 and the hypothesis H_0: $\phi_{01} = 0$ is accepted for $\alpha = 0.45$. The extra sum of squares associated with ϕ_{02} is 3.074 and the hypothesis H_0: $\phi_{02} = 0$ is accepted for $\alpha = 0.65$. The extra sum of squares associated with θ_{10} is 156.86 and the hypothesis H_0: $\theta_{10} = 0$ is rejected. The sample space-time autocorrelation functions, that have been computed and listed in Table 6-13, show adequacy of this model.

CONTINUED

5 OF 8

Table 6-13. The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelation Functions of the Residuals of the Coupled Model, Equation (6-42)

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2 /	0	1	2
0	1.00	-0.07	0.00	1.00	0.98	0.03
1	-0.06	0.06	0.02	-0.63	0.71	0.22
2	0.03	-0.09	-0.04	0.37	-0.96	-0.47
3	-0.01	-0.07	-0.13	-0.09	-0.72	-1.21
4	0.17	0.05	0.05	1.50	0.48	0.44
5	-0.13	-0.00	-0.00	-1.00	-0.02	-0.02

Since ϕ_{01} and ϕ_{02} are non-significant at significant level $\alpha = 0.45$ and $\alpha = 0.15$, respectively, so the STMA model, that is obtained by dropping the ϕ_{01}, ϕ_{02} terms in the model, Equation (7-42), is constructed. The M.L. estimation gives

$$\underset{\sim}{Z}_t = -\hat{\theta}_{10}\underset{\sim}{\varepsilon}_{t-1} + \underset{\sim}{\varepsilon}_t \qquad (6\text{-}46)$$

where

$$\hat{\theta}_{10} = 0.7863, \ \hat{\sigma}^2 = 1.501$$

The sample space-time autocorrelation functions, that are listed in Table 6-14, do not reveal any significant structure and this model is adequate.

Comparing the model in Equation (6-43), which is built for the outlier corrected observations, with the model in Equation (6-46), which is built for the uncorrected data, we see that the purely spatial structures are not masked and are significant in the model of the outlier connected observations. On the other hand, the purely spatial structure is masked by the outliers, and ϕ_{01}, ϕ_{02} are nonsignificant in the model of the unconnected data. It should be noted that the estimated residual variance of the outlier corrected model is 1.352 which is smaller than that of the uncorrected model 1.501.

An alternative way to see the inadequacy of the model which ignores the homogeneity assumption which causes a masking of the spatial structure is to build the STMA(1_0) model for the outlier

Table 6-14. The Sample Space-Time Autocorrelation Functions and the Standardized Sample Space-Time Autocorrelations of the Residuals of the STMA Model, Equation (6-43)

The Sample Space-Time Autocorrelations/
The Standardized Sample S-T Autocorrelations

Space Lag / Time Lag	0	1	2	/ 0	1	2
0	1.00	-0.14	-0.13	1.00	-1.79	-1.62
1	-0.06	0.09	0.05	-0.64	0.99	0.55
2	0.05	-0.03	-0.00	0.56	-0.30	-0.07
3	0.01	-0.02	-0.06	0.11	-0.23	-0.63
4	0.10	0.05	0.02	0.88	0.43	0.20
5	-0.07	-0.00	0.00	-0.58	-0.04	0.02

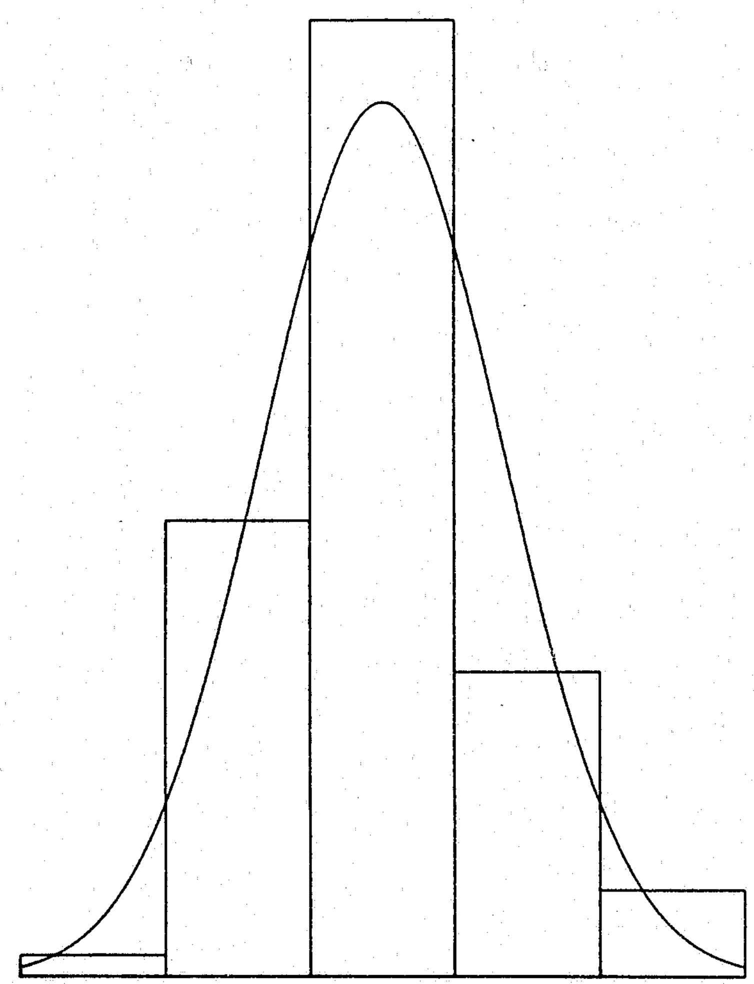


Figure 6-10(a). The Histogram of the Residuals of the model for the Uncorrected Data.

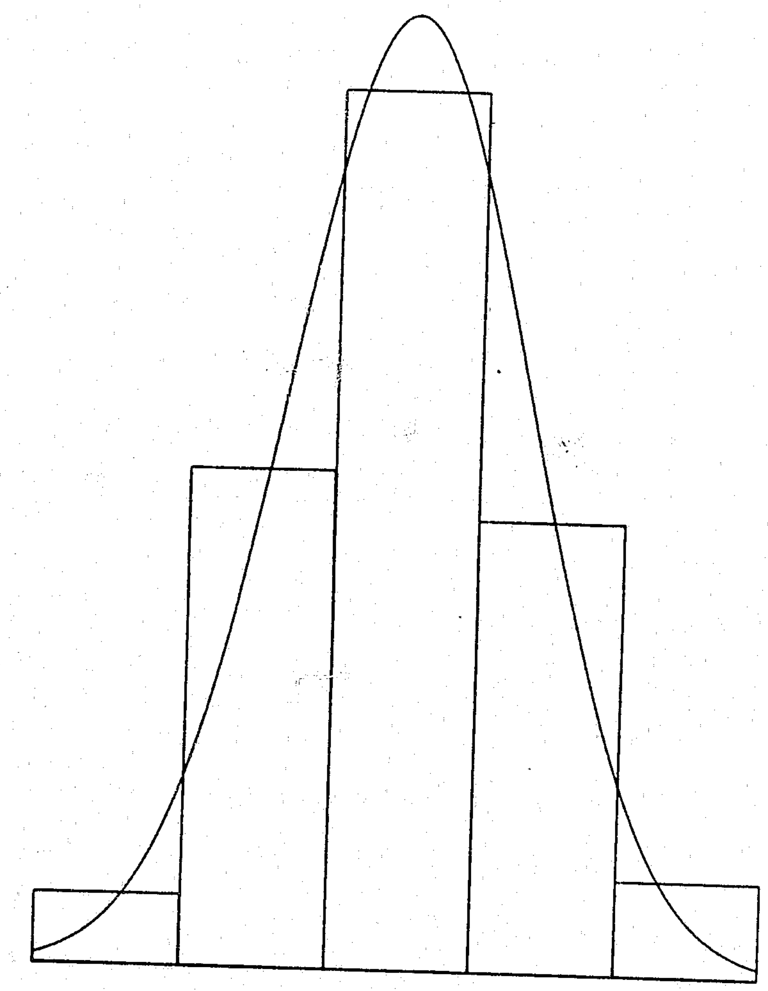


Figure 6-10(b). The Histogram of the Residuals of the Model for the Outlier Corrected Data.

corrected data. This results in,

$$\underset{\sim}{Z}_t = -\hat{\theta}_{10}\underset{\sim}{\varepsilon}_{t-1} + \underset{\sim}{\varepsilon}_t \qquad (6\text{-}47)$$

where

$$\hat{\theta}_{10} = 0.265 \qquad 95\%\ \text{C.I.}\ (0.093,\ 0.438)$$

The sample space-time autocorrelation functions of these residuals, that are listed in Table 6-15 and clearly shows the needs of purely spatial AR terms. Comparing the sample space-time autocorrelations listed in Table 6-15 with those listed in Table 6-14, we see that by removing the masking effect of the outliers, makes the purely spatial structure significant, and the purely spatial terms are needed to build the adequate model.

Since the outliers are confounded with the noise, so it is expected that the normality assumption for the residuals will be satisfied better for those that are from the outlier correct model. The five-block histogram for the residuals from the model, Equation (6-46), which is for the uncorrected data, is plotted in Figure 6-10(a). The histogram for the residuals of the model, Equation (6-43), which is built for the outlier corrected data, is plotted in Figure 6-10(b). Comparing these two plots, we may make the judgement that the normal assumption is satisfied better in Figure 6-10(a), which is built for the outlier corrected data. Performing the χ^2 goodness-of-fit test confirms this visual judgement since we have $\chi^2 = 4.49$, $\alpha = 0.2129$,

Table 6-15. The Sample Space-Time Autocorrelation Functions of the Residuals of the STMA(1_0) Model, Equation (6-47)

Space Lag / Time Lag	The Sample Space-Time Autocorrelations / 0	1	2	The Standardized Sample S-T Autocorrelations / 0	1	2
0	1.00	-0.60	0.06	1.00	-7.47	0.77
1	-0.03	0.01	0.08	-0.34	0.15	0.72
2	0.08	-0.04	0.05	0.82	-0.48	0.51
3	0.15	-0.11	0.03	1.42	-1.05	0.36
4	-0.08	0.10	-0.11	-0.52	0.90	-0.95
5	-0.03	0.02	-0.02	-0.27	0.19	-0.22

for the residuals of uncorrected data and $\chi^2 = 0.53$, $\alpha = 0.9103$ for the residuals of outlier corrected data, where α is the significant level. The residuals of the outlier corrected model satisfies the normality assumption better.

Based on the model in Equation (6-46), which has been built for the differenced observations without corrections, the forecast function and the variance of these forecasts are obtained as the conditional expectation values of mean and variance for the differenced observations $\underset{\sim}{Z}_t$ as well as the original observations $\underset{\sim}{y}_t$. They are,

$$\hat{\underset{\sim}{Z}}_T(\ell) = \begin{cases} -\hat{\theta}_{10}\hat{\underset{\sim}{\varepsilon}}_T & , \ell=1 \\ 0 & , \ell\geq 2 \end{cases}$$

$$\hat{y}_T(\ell) = \begin{cases} y_T + \hat{\underset{\sim}{Z}}_T(1) & , \ell=1 \\ \hat{y}_T(\ell-1) + \hat{\underset{\sim}{Z}}_T(\ell) & , \ell\geq 2 \end{cases} \qquad (6\text{-}48)$$

$$\mathrm{Var}(\hat{\underset{\sim}{Z}}_T(\ell)) = \begin{cases} I\sigma^2 & , \ell=1 \\ (1+\hat{\sigma}_{10}^2)\hat{\sigma}^2 & , \ell\geq 2 \end{cases}$$

$$\mathrm{Var}(\hat{y}_T(\ell)) = \begin{cases} \hat{\sigma}^2 & , \ell=1 \\ \mathrm{Var}(\hat{\underset{\sim}{y}}_T(\ell-1)) + \mathrm{Var}(\underset{\sim}{Z}_T(\ell)) - 2D\hat{\theta}_{10}, & \ell\geq 2 \end{cases}$$

where

$$\hat{\theta}_{10} = 0.7863,\ \hat{\sigma}^2 = 1.501.$$

Based on the model in Equation (6-43), which has been built for the differenced observations with outlier corrections, the forecast function and the variance of these forecasts are,

$$\underset{\sim}{Z}_T(\ell) = \begin{cases} -\hat{\theta}_{10}(I - \hat{\phi}_{01}W^{(1)} - \hat{\phi}_{02}W^{(2)})^{-1}\hat{\underset{\sim}{\varepsilon}}_T & \ell=1 \\ 0 & \ell\geq 2 \end{cases} \qquad (6\text{-}49)$$

$$y_T(\ell) = \begin{cases} y_T + \underset{\sim}{Z}_T(1) & \ell=1 \\ y_T(\ell-1) + \underset{\sim}{Z}_T(\ell) & \ell\geq 2 \end{cases}$$

$$\mathrm{Var}(\underset{\sim}{Z}_T(\ell)) = \begin{cases} [(I-\hat{\phi}_{01}W^{(1)}-\hat{\phi}_{02}W^{(2)})(I-\hat{\phi}_{01}W^{(1)}-\phi_{01}W^{(2)})^{-1}]\hat{\sigma}^2, & \ell=1 \\ W_B W_B'\hat{\sigma}^2 + W_c W_c'\hat{\sigma}^2 & , \ell\geq 2 \end{cases} \qquad (6\text{-}50)$$

$$\mathrm{Var}(y_T(\ell)) = \begin{cases} \mathrm{Var}(\underset{\sim}{Z}_T(1)), & \\ \mathrm{Var}(y_T(\ell-1) + \mathrm{Var}(\underset{\sim}{Z}_T(\ell)) + \mathrm{cov}_T(\ell) + \mathrm{cov}_T(\ell)', & \ell\geq 2 \end{cases}$$

where

$$\mathrm{cov}_T(\ell) = \begin{cases} 0 & \ell\leq 1 \\ -W_c W_B'\hat{\sigma}^2, & \ell\geq 2 \end{cases}$$

$$W_B = \hat{\theta}_{10}(I - \hat{\phi}_{01}W^{(1)} - \hat{\phi}_{02}W^{(2)})^{-1}$$

$$W_c = (I - \hat{\phi}_{01}W^{(1)} - \hat{\phi}_{02}W^{(2)})^{-1}$$

$$\hat{\phi}_{01} = -1.086$$

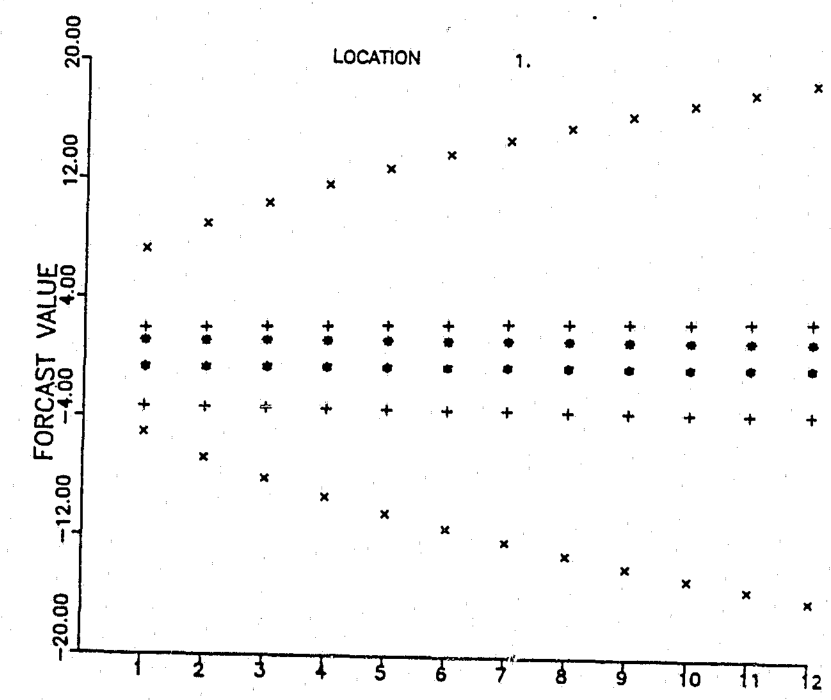
$$\hat{\phi}_{02} = -0.206$$

$$\hat{\theta}_{10} = 0.255$$

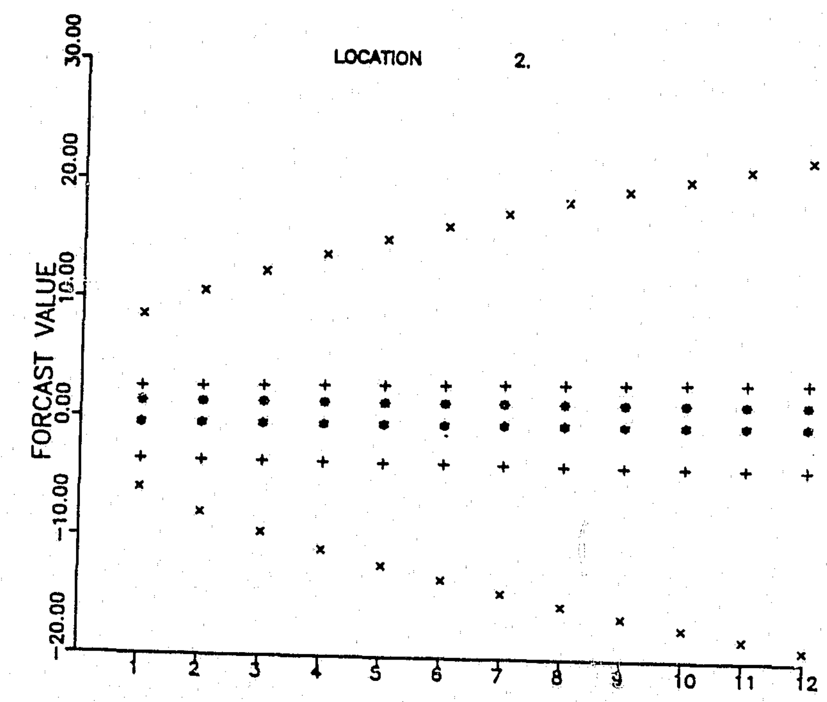
$$\hat{\sigma}^2 = 1.352$$

The point forecasts as well as the 95% confidence interval forecasts of y_T at T=71 are computed for the -step ahead forecasts, =1,2,..,12, by applying the forecasting function in Equations (6-49) and (6-50). These forecasts are plotted in Figure 6-11 for each of the 14 locations for both the outlier corrected and uncorrected model. The point forecasts for these two models are different in all 14 locations. The interval forecasts of the outlier corrected model always contains the interval forecasts of the uncorrected model. The 95% confidence interval of the outlier corrected model is larger than that of the uncorrected model because the correctly described purely spatial structure in Equation (6-49) inflates the forecasting variance.

Given the general space-time model with contemporaneous structure in Equation (6-51)



(a)



(b)

Figure 6-11. The Point Forecasts and the 95% Confidence Interval Forecasts Computed from Equations 6-49 and 6-50.

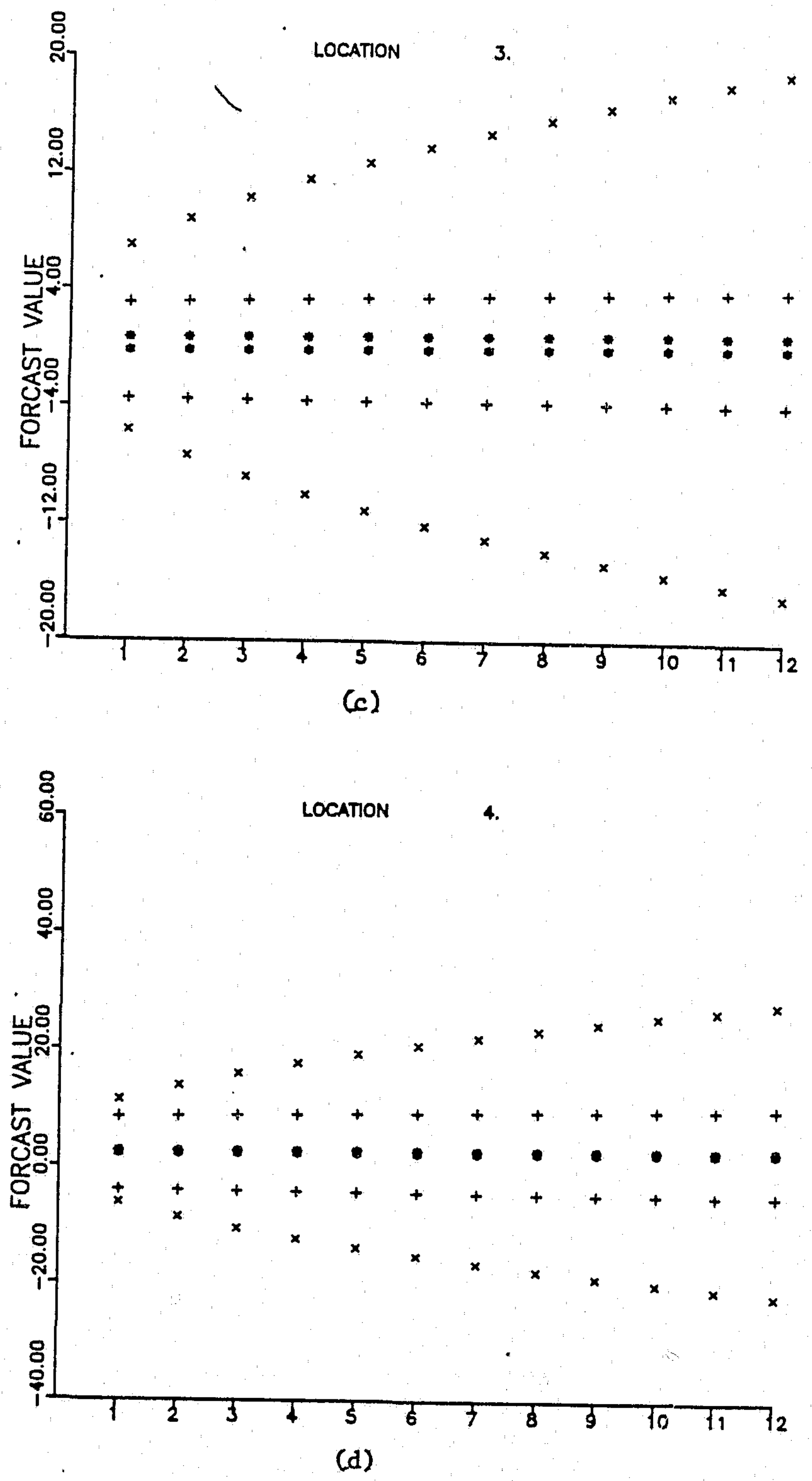


Figure 6-11. (Cont'd)

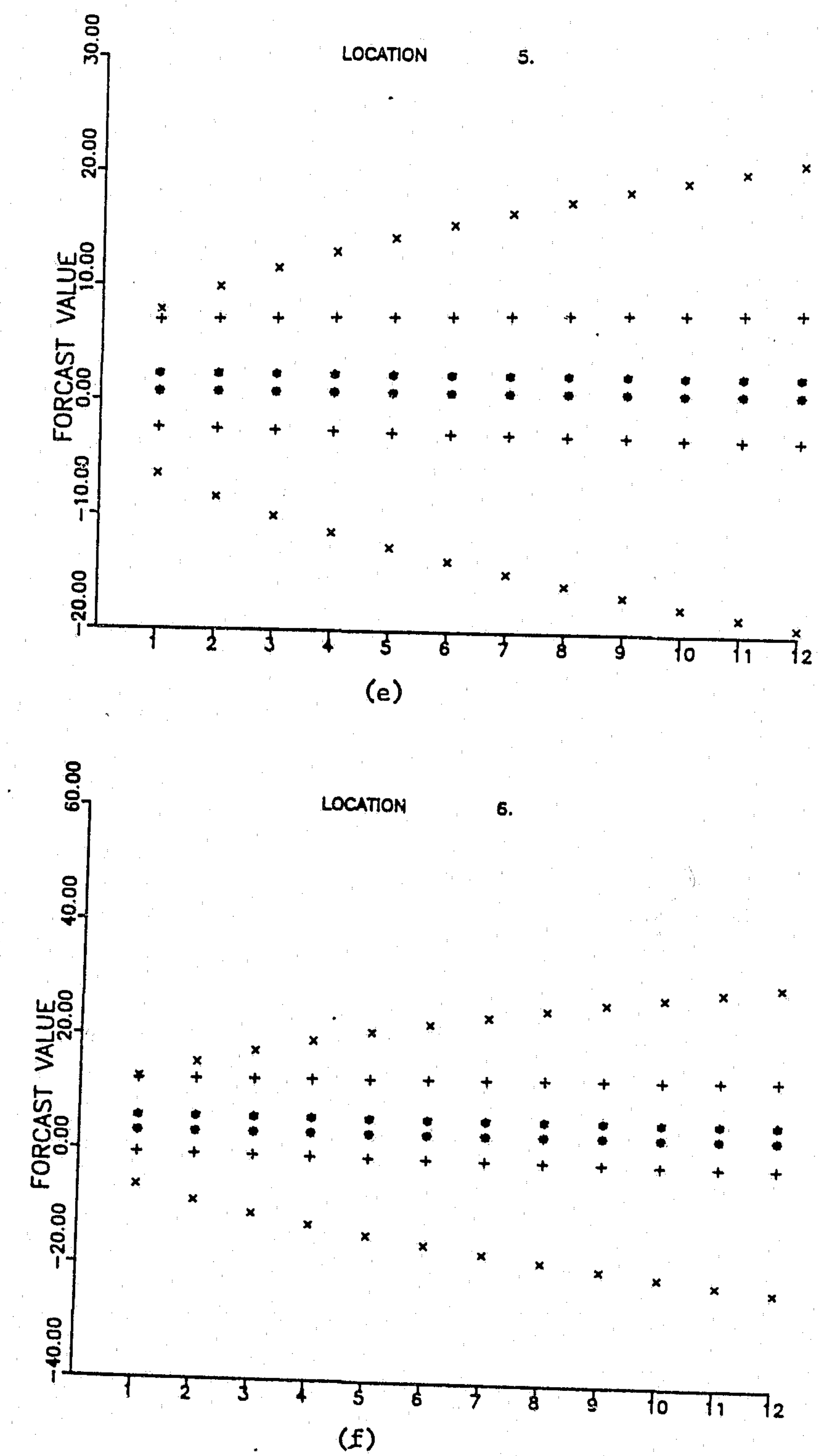
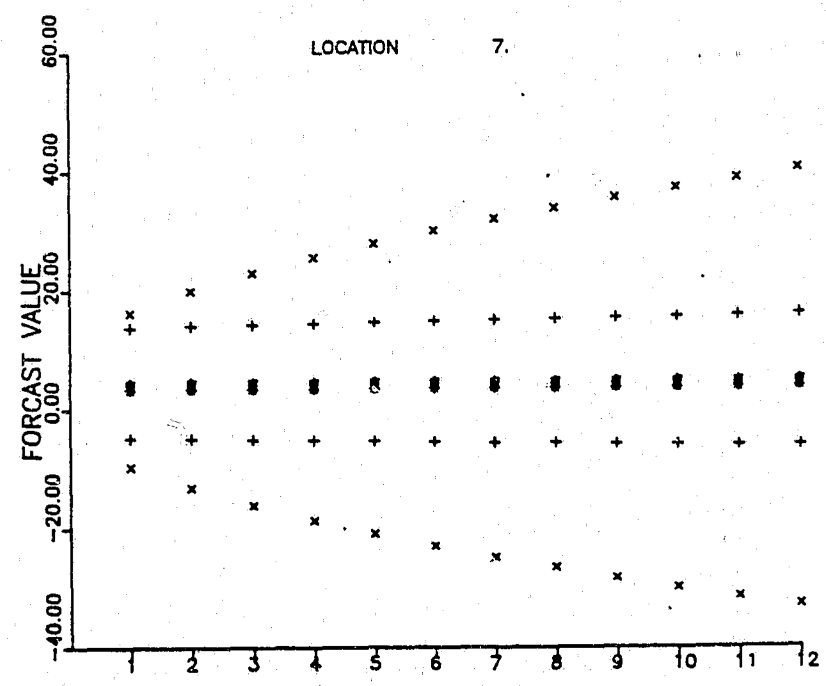
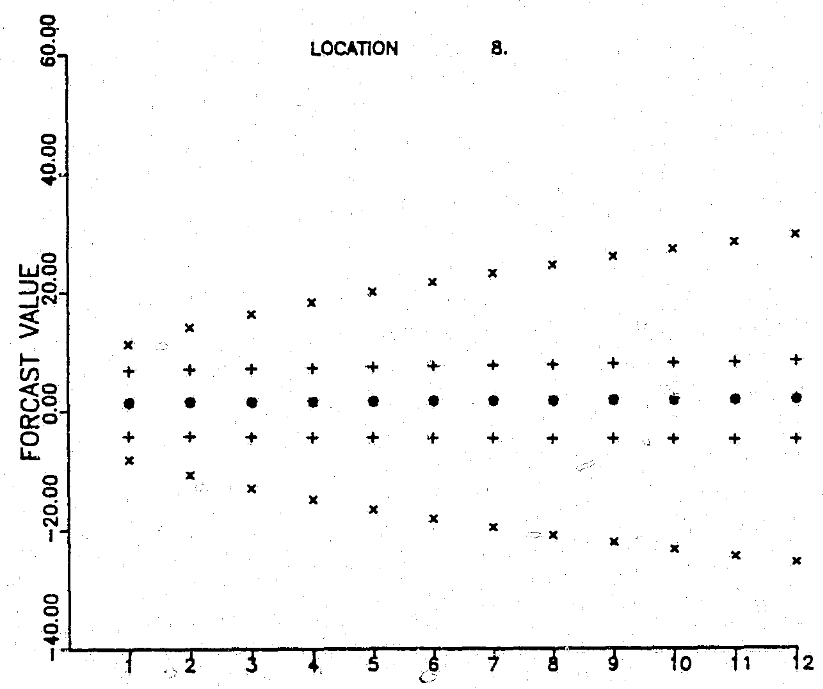


Figure 6-11. (Cont'd)

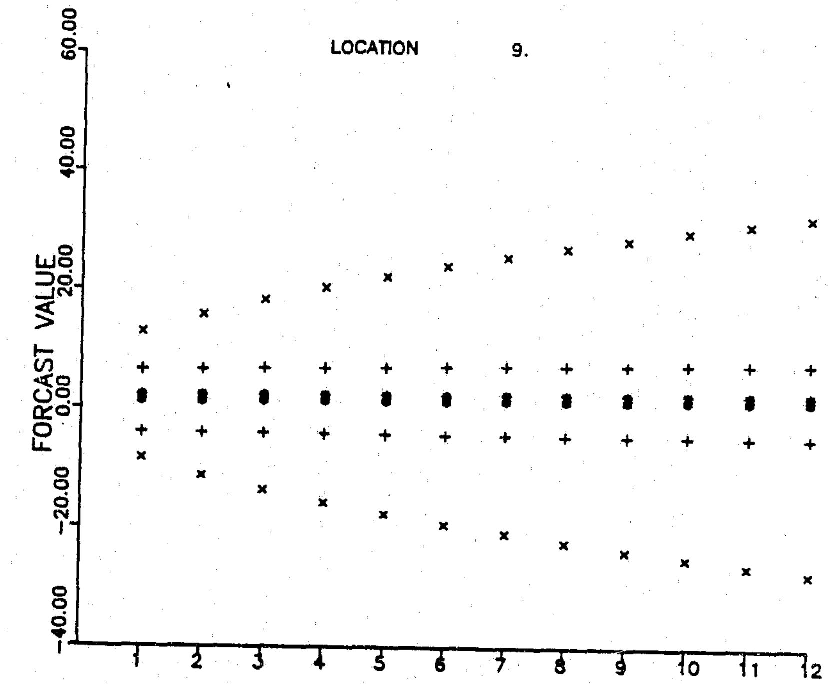


(g)

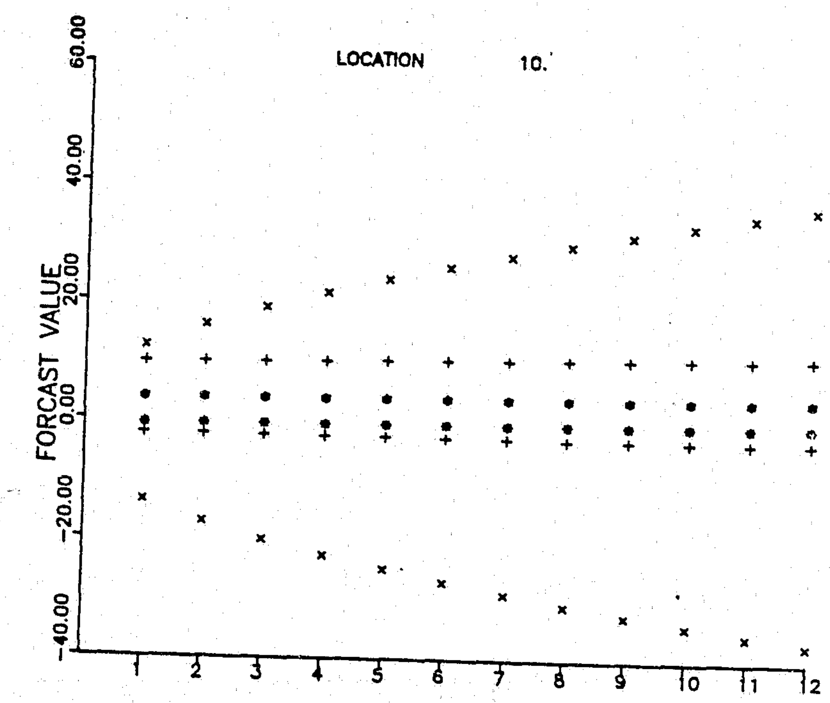


(h)

Figure 6-11. (Cont'd)

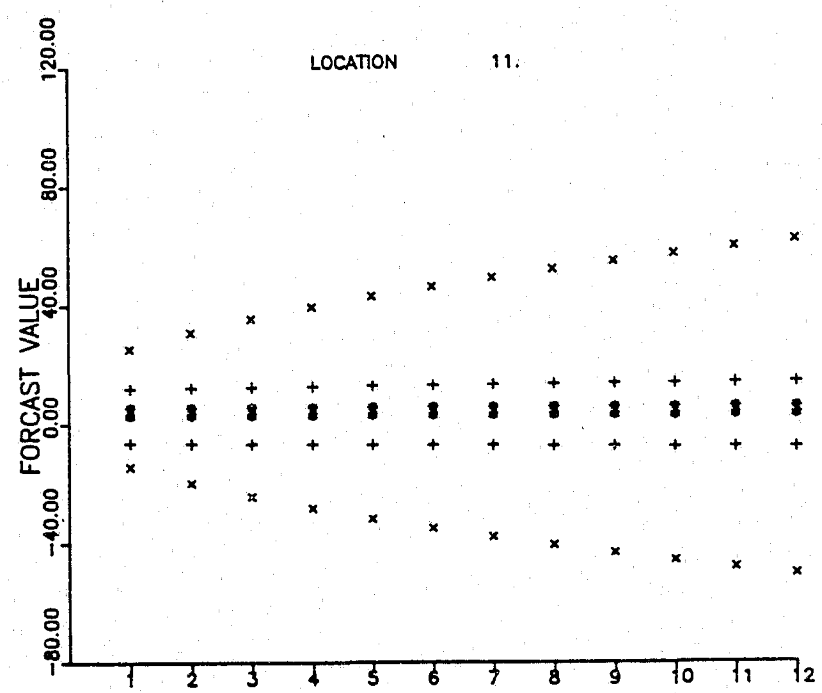


(i)

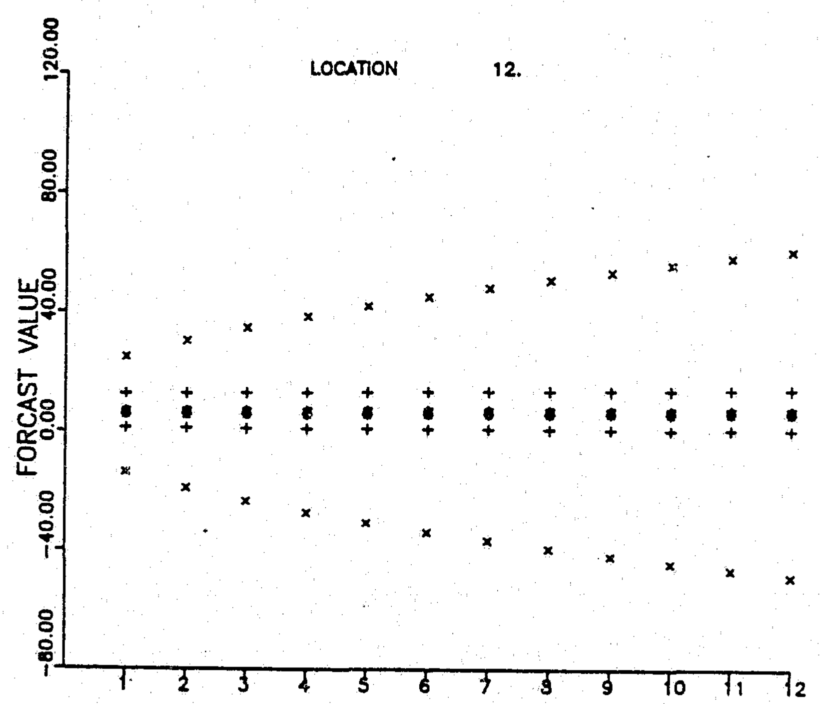


(j)

Figure 6-11. (Cont'd)

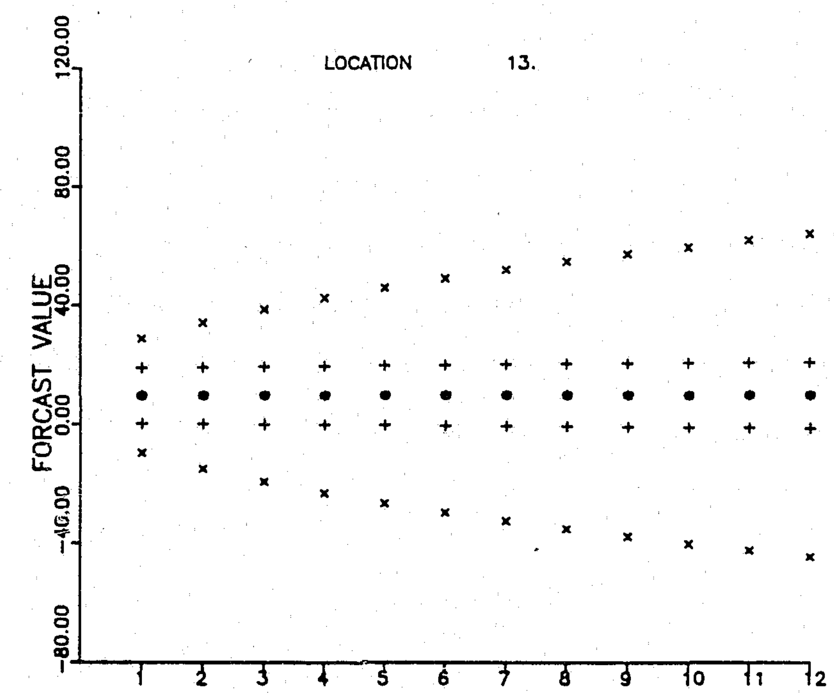


(k)

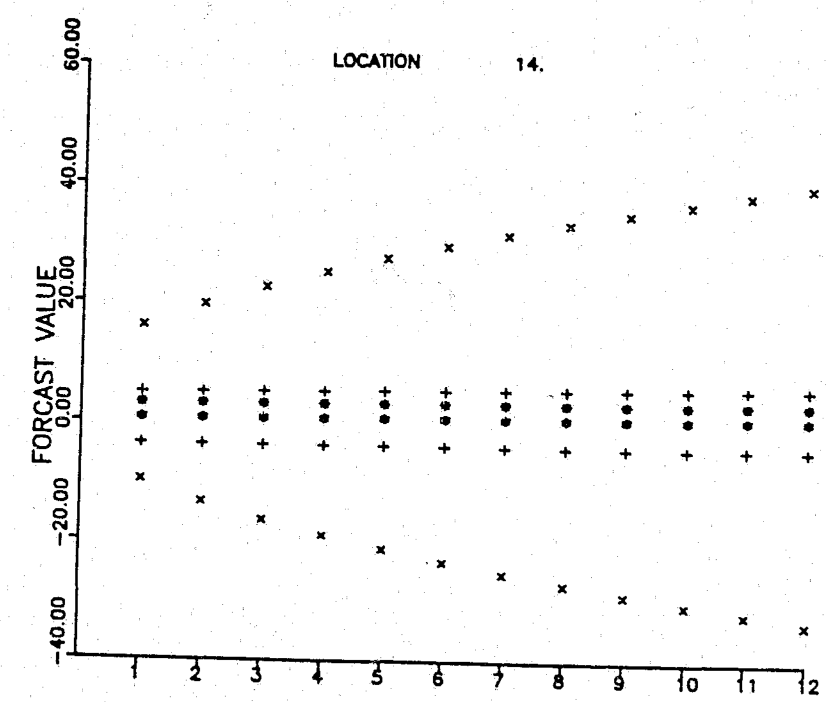


(ℓ)

Figure 6-11. (Cont'd)



(m)



(n)

Figure 6-11. (Cont'd)

$$Z_t = B(0)Z_t - A(0)\varepsilon_t + \sum_{k=1}^{p} B(k)Z_{t-k} - \sum_{k=1}^{q} A(k)\varepsilon_{t-k} + \varepsilon_t, \quad (6\text{-}51)$$

the space-time components describe the spatial-temporal correlations are contained in the forecasting functioning, but the contemporaneous terms, $B(0)Z_t$ and $A(0)\varepsilon_t$, which describe the purely spatial correlations are not realized in the forecasting function. On the other hand, the general space-time process

$$Z_t = \sum_{k=1}^{p} B(k)Z_{t-k} - \sum_{k=1}^{q} A(k)\varepsilon_{t-k} + \varepsilon_t \quad (6\text{-}52)$$

has the same terms contribute to the forecasting function. Thus, the general space-time model with contemporaneous structure has two more terms that have descriptive capability but do not enter the forecasting function and these two added terms inflates the forecasting confidence interval. As a simple example, let us consider the following two models,

$$Z_t = \varepsilon_t \quad (6\text{-}53)$$

and

$$Z_t = B(0)\varepsilon_t + \varepsilon_t \quad (6\text{-}54)$$

In Equation (6-54), B(0) is assumed to satisfy the existence conditions therefore Z_t can be expressed alternatively as,

$$Z_t=(I+B(0)+B^2(0)+\ldots)\varepsilon_t. \quad (6\text{-}55)$$

Here we see that Z_t is an infinite sum of all errors ε_t over space for each location at a given t which is contributed through B(0). Thus the variance of Z_t is inflated accordingly.

It should be noted that the outlier corrected model is a space-time model with contemporaneous spatial structure, while the uncorrelated model is only a space-time model in which the purely spatial structure was masked causing an inflation in σ^2. They are of different descriptive capability. The outlier corrected model describes the process correctly. The larger forecast confidence interval of the outlier corrected model reflects the additional terms needed in the model past the space-time terms. Thus, the smaller confidence interval for the non-outlier corrected model reflects an over estimation of the true α level. The desired $(1-\alpha)\%$ confidence interval for the uncorrected model actually represents the correct $(1-\alpha^*)\%$ confidence interval where $\alpha^* > \alpha$.

CHAPTER VII

MULTIVARIATE STARMA MODELING

In the STARMA model only one type of observations is considered (eg. one attribute). Occasionally more than one category of observations is available for each observation period on every location and there can be some space-time neighboring structure between different categories. If the structure among different categories of observations is significant in the data generating process, then the descriptive ability of models that also characterize the between-category structure will be better than those that don't have this capability.

In this chapter we generalize the STARMA model, to the Multivariate STARMA model (MULSTARMA). In Section 7.1, the model formulation is given. The stationary and invertibility conditions are developed in Section 7.2. Alternative MULSTARMA forms are described in Section 7.3. In Section 7.4, the multivariate space-time autocorrelation function is defined and its statistical properties are derived. The multivariate space-time partial autocorrelation function is derived in Section 7.5, and the computationally efficient schemes are addressed. The model identification techniques that are based on the results obtained in Section 7.4 and 7.5 are contained in Section 7.6. The encoded multivariate space-time autocorrelation functions and the encoded multivariate space-time partial autocorrelation functions are also introduced here to assist in identification. The following

section contains the parameter estimation procedures. This includes conditional M.L. estimation procedures for both situations when G is known or unknown, where G is the covariance matrix of the noise. Also the situations under which the computation effort needed in estimation can be reduced are discussed. In Section 7.8, the model diagnostic checking procedures are described. Then in Section 7.9, Cleveland Crime Data is used to illustrate the MULSTARMA model building procedure and the resulting model's use in forecasting and intervention analysis.

7.1 The Multivariate STARMA Model Class

The multivariate STARMA model is an extension of univariate STARMA model into the multi-category observation domain. It is also considered to be an extension of multivariate ARMA model into the spatial domain. The multivariate STARMA model has the capability to describe the spatially, temporally as well as the inter-category correlated structures. The multivariate STARMA model formulation is introduced in Section 7.1.1. This formulation is related to other special subsets of the multivariate STARMA model class in Section 7.1.2. In Section 7.1.3, its physical interpretation is illustrated in block diagrams for simple systems.

7.1.1 Model Formulation

Consider the general multivariate Autoregressive Moving Average Model,

$$\underset{\sim}{Z}^h(t) = \sum_{g=1}^{\zeta}\sum_{k=1}^{p^{hg}} B^{hg}(k)\underset{\sim}{Z}^g(t-k) - \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}} A^{hg}(k)\underset{\sim}{\varepsilon}^g(t-k) + \underset{\sim}{\varepsilon}^h(t) \quad (7\text{-}1)$$

and $\underset{\sim}{\varepsilon}^h(t)$ is normally distributed with

$$E(\underset{\sim}{\varepsilon}^h(t)\underset{\sim}{\varepsilon}^g(t+k)') = \begin{cases} G^{hg} & \text{when } k=0 \quad h,g=1,2,\ldots,\zeta \\ 0 & \text{otherwise} \end{cases}$$

$$h=1,2,\ldots,\zeta;\ t=1,2,\ldots,T.$$

where $\underset{\sim}{Z}^h(t)$ and $\underset{\sim}{\varepsilon}^h(t)$ are the h-category observation vector and h-category noise vector at time t, respectively. $B^{hg}(k)$ and $A^{hg}(k)$, that are category dependent as well as temporal lag dependent, are LN×LN coefficient matrix for time lag k.

It should be noted that in the previous chapters, the temporal index t is represented as a subscript. Since the subscripts and superscripts are much more complicated in this chapter than those in previous chapters, the temporal index t is placed in the parenthesis in the rest of this chapter.

Let,

$$\begin{aligned}
\underset{\sim}{Z}(t)' &= (\underset{\sim}{Z}^1(t)', \underset{\sim}{Z}^2(t)',\ldots,\underset{\sim}{Z}^\zeta(t)') \\
\underset{\sim}{\varepsilon}(t)' &= (\underset{\sim}{\varepsilon}^1(t)', \underset{\sim}{\varepsilon}^2(t)',\ldots,\underset{\sim}{\varepsilon}^\zeta(t)') \\
\underset{\sim}{Z}' &= (\underset{\sim}{Z}(1)', \underset{\sim}{Z}(2)',\ldots,\underset{\sim}{Z}(T)') \\
\underset{\sim}{\varepsilon}' &= (\underset{\sim}{\varepsilon}(1)', \underset{\sim}{\varepsilon}(2)',\ldots,\underset{\sim}{\varepsilon}(T)') \\
P_{max} &= \max\{p^{hg}|h,g=1,2,\ldots,\zeta\} \\
q_{max} &= \max\{g^{hg}|h,g=1,2,\ldots,\zeta\}
\end{aligned} \quad (7\text{-}2)$$

Define $A(k) = (A^{hg}(k))$, $B(k) = [B^{hg}(k)]$, $G = [G^{hg}]$, i.e., A(k) is the $[(\zeta\cdot LN)\times(\zeta\cdot LN)]$ coefficient matrix which has $A^{hg}(k)$ submatrix at its (h,g) block of size $(\zeta\cdot\zeta)$. B(k) is the $[(\zeta\cdot LN)\times(\zeta\cdot LN)]$ coefficient matrix which has $B^{hg}(k)$ submatrix at its (h,g) block, and G has G^{hg} at the (h,g) block.

Model, Equation (7-1), which will be referred to as the General Multivariate ARMA model, can be rewritten in terms of the definitions in Equation (7-2) as

$$\underset{\sim}{Z}(t) = \sum_{k=1}^{P_{max}} B(k)\underset{\sim}{Z}(t-k) - \sum_{k=1}^{q_{max}} A(k)\underset{\sim}{\varepsilon}(t-k) + \underset{\sim}{\varepsilon}(t) \quad (7\text{-}3)$$

$$t=1,2,\ldots,T$$

Two models that are related to the General Multivariate ARMA model are the General Multivariate AR model and the General Multivariate MA model.

By setting $A(k) = 0$ in Model, Equation (7-1), we have the General Multivariate AR model.

$$\underset{\sim}{Z}(t) = \sum_{k=1}^{p_{max}} B(k)\underset{\sim}{Z}(t-k) + \underset{\sim}{\varepsilon}(t) \tag{7-4}$$

$$t=1,2,\ldots,T$$

By setting $B(K) = 0$ in Model, Equation (7-1), we have the General Multivariate MA model as

$$\underset{\sim}{Z}(t) = -\sum_{k=1}^{q_{max}} A(k)\underset{\sim}{\varepsilon}(t-k) + \underset{\sim}{\varepsilon}(t) \tag{7-5}$$

$$t=1,2,\ldots,T$$

When the coefficient matrix $B^{hg}(k)$ and $A^{hg}(k)$ in Model, Equation (7-1), can be decomposed and expressed in terms of the weight matrix $w^{(\ell)}$, $\ell=0,1,\ldots,$ as

$$B^{hg}(k) = \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg}\, w^{(\ell)},\quad A^{hg}(k) = \sum_{\ell=0}^{m_k^{hg}} \theta_{k\ell}^{hg}\, w^{(\ell)}$$

We get the Multivariate STARMA model

$$\underset{\sim}{Z}^h(t) = \sum_{g=1}^{\zeta}\sum_{k=1}^{p^{hg}}\sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)}\underset{\sim}{Z}^g(t-k) - \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}}\sum_{\ell=0}^{m_k^{hg}} \theta_{k\ell}^{hg}\, w^{(\ell)}\underset{\sim}{\varepsilon}^g(t-k) \tag{7-6}$$

$$+\ \underset{\sim}{\varepsilon}^h(t)$$

where $\underset{\sim}{\varepsilon}^h(t)$ is normally distributed and

$$E(\underset{\sim}{\varepsilon}^h(t)\underset{\sim}{\varepsilon}^g(t+k)) = \begin{cases} G^{hg}, & k=0 \\ 0, & \text{otherwise} \end{cases}$$

$$h=1,2,\ldots,\zeta;\ t=1,2,\ldots,T$$

This model will be referred to as the MULSTARMA $(\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m})$ model with $\underset{\sim}{p}$, $\underset{\sim}{q}$, $\underset{\sim}{\lambda}$, $\underset{\sim}{m}$ vectors defined in the following

$$\begin{aligned}
\underset{\sim}{p} &= (p^{11},p^{12},\ldots,p^{1\zeta},p^{21},\ldots,p^{2\zeta},\ldots,p^{\zeta1},\ldots,p^{\zeta\zeta}) \\
\underset{\sim}{q} &= (q^{11},q^{12},\ldots,q^{1\zeta},q^{21},\ldots,q^{2\zeta},\ldots,q^{\zeta1},\ldots,q^{\zeta\zeta}) \\
\underset{\sim}{\lambda} &= (\underset{\sim}{\lambda}^{11},\underset{\sim}{\lambda}^{12},\ldots,\underset{\sim}{\lambda}^{1\zeta},\underset{\sim}{\lambda}^{21},\ldots,\underset{\sim}{\lambda}^{2\zeta},\ldots,\underset{\sim}{\lambda}^{\zeta1},\ldots,\underset{\sim}{\lambda}^{\zeta\zeta}) \\
\underset{\sim}{\lambda}^{ij} &= (\lambda_1^{ij},\lambda_2^{ij},\ldots,\lambda_{p^{ij}}^{ij}) \\
\underset{\sim}{m} &= (\underset{\sim}{m}^{11},\underset{\sim}{m}^{12},\ldots,\underset{\sim}{m}^{1\zeta},\underset{\sim}{m}^{21},\ldots,\underset{\sim}{m}^{2\zeta},\ldots,\underset{\sim}{m}^{\zeta1},\ldots,\underset{\sim}{m}^{\zeta\zeta}) \\
\underset{\sim}{m}^{ij} &= (m_1^{ij},m_2^{ij},\ldots,m_{q^{ij}}^{ij})
\end{aligned} \tag{7-7}$$

The elementary elements in the above definition are p^{ij}, q^{ij}, λ_k^{ij} and m_k^{ij}, the superscript ij indicates that the parameter is associated with the influence of the category j observations on the category i observation, and the subscript k indicates the temporal lag. Referencing to Equation (7-6), we see that p^{ij} is the maximum temporal lag of influence that the category j observations have on the category i observations, q^{ij} is the maximum temporal lag influence that the category j noise have on the category i observations, λ_k^{ij} is the maximum spatial lag influence that the category j observations have on the category i observations at temporal lag k and m_k^{ij} is the maximum spatial lag influence that the category j noise has on the category i observations at temporal lag k. For example, in Equation (7-6), ϕ_k^{ij} is the autoregressive parameter that measures the strength of influence that the category j observations at some location have on the category i observations of the ℓ^{th} order neighbors at temporal lag k.

$\underset{\sim}{p}$ and $\underset{\sim}{q}$ are $(1\times\zeta^2)$ row vectors. $\underset{\sim}{\lambda}^{ij}$ are $(1\times p^{ij})$ row vectors, $\underset{\sim}{\lambda}$ is $(1\times \sum_{i=1}^{\zeta}\sum_{j=1}^{\zeta} p^{ij})$ row vector. $\underset{\sim}{m}^{ij}$ is $(1\times q^{ij})$ row vector and $\underset{\sim}{m}$ is $(1\times \sum_{i=1}^{\zeta}\sum_{j=1}^{\zeta} q^{ij})$ row vector.

In the rest of this chapter, we will always define the parameter vectors $\underset{\sim}{\phi}$, $\underset{\sim}{\theta}$ in such order as following

$$\underset{\sim}{\phi}' = (\underset{\sim}{\phi}^{11'},\underset{\sim}{\phi}^{12'},\ldots,\underset{\sim}{\phi}^{1\zeta'},\underset{\sim}{\phi}^{21'},\ldots,\underset{\sim}{\phi}^{\zeta\zeta'}) \text{ with}$$
$$\underset{\sim}{\phi}^{ij'} = (\phi_{10}^{ij},\phi_{11}^{ij},\ldots,\phi_{1\lambda_1^{ij}}^{ij},\phi_{20}^{ij},\ldots,\phi_{p^{ij}0}^{ij},\ldots,\phi_{p^{ij}\lambda_{p^{ij}}^{ij}}^{ij})$$
$$\underset{\sim}{\phi}^{i'} = (\underset{\sim}{\phi}^{i1'},\underset{\sim}{\phi}^{i2'},\ldots,\underset{\sim}{\phi}^{i\zeta'}),\ i=1,2,\ldots,\zeta \qquad (7\text{-}8)$$

$$\underset{\sim}{\theta}' = (\underset{\sim}{\theta}^{11'},\underset{\sim}{\theta}^{12'},\ldots,\underset{\sim}{\theta}^{1\zeta'},\underset{\sim}{\theta}^{21'},\ldots,\underset{\sim}{\theta}^{\zeta\zeta'}) \text{ with}$$
$$\underset{\sim}{\theta}^{ij'} = (\theta_{10}^{ij},\theta_{11}^{ij},\ldots,\theta_{1\lambda_1^{ij}}^{ij},\theta_{20}^{ij},\ldots,\theta_{q^{ij}1}^{ij},\ldots,\theta_{q^{ij}m_{q^{ij}}^{ij}}^{ij}), \text{ and}$$
$$\underset{\sim}{\theta}^{i'} = (\underset{\sim}{\theta}^{i2'},\underset{\sim}{\theta}^{i2'},\ldots,\underset{\sim}{\theta}^{i\zeta'}),\ i=1,2,\ldots,\zeta$$

where $\underset{\sim}{\phi}$ is a $(\sum_{i=1}^{\zeta}\sum_{j=1}^{\zeta}\sum_{k=1}^{p^{ij}}(\lambda_k^{ij}+1))$ column vector, $\underset{\sim}{\phi}^{ij}$ a $\sum_{k=1}^{p^{ij}}(\lambda_k^{ij}+1)$ column vectors; $\underset{\sim}{\theta}$ is a $(\sum_{i=1}^{\zeta}\sum_{j=1}^{\zeta}\sum_{k=1}^{q^{ij}}(m_k^{ij}+1))$ column vector, $\underset{\sim}{\theta}^{ij}$ is $\sum_{k=1}^{q^{ij}}(m_k^{ij}+1)$ column vector.

7.1.2 Other Special Subsets of the MULSTARMA

In the MULSTARMA$(\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m})$ model formulation, Equation (7-6), two kinds of specifications can be distinguished: the model parameter specifications and the system specifications. The model parameter specifications include $\underset{\sim}{p}$, $\underset{\sim}{q}$, $\underset{\sim}{\lambda}$ and $\underset{\sim}{m}$, that specify the model subclass of the MULSTARMA models. By setting these parameters to some specified values, we obtain a model subclass. The system specifications include the number of observation categories and the location number LN. The system specifications are specified by the available observations themself, not by the model builder. To model the observed process, the system specifications should match the observed system from which the observations are obtained. In the following, subsets of special interests are introduced. The model subsets, that can be obtained by specifying the model parameter specifications, are presented in Section 7.1.2.1, and the model subsets, that can be obtained by changing the system specifications, are presented in Section 7.1.2.2.

7.1.2.1 Parameter Simplification. In the univariate space-time models, the STAR model formulation can be obtained from the STARMA model by setting the moving average order to zero, and the STMA model formulation can be obtained by setting the autoregressive order to zero. Special subsets of the MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) can be similarly obtained. In this section those models that can be obtained from the MULSTARMA model by setting elements in the autoregressive order vector $\underset{\sim}{p}$ and the moving average order vector $\underset{\sim}{q}$ to zero are presented.

Setting $\underset{\sim}{q} = \underset{\sim}{0}$ in the MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) model, Equation (7-6), we have

$$\underset{\sim}{z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{z}^g(t-k) + \underset{\sim}{\varepsilon}^h(t) \tag{7-9}$$

where $\underset{\sim}{\varepsilon}^h(t)$ is normally distributed with

$$E(\underset{\sim}{\varepsilon}^h(t)\underset{\sim}{\varepsilon}^g(t+k)') = \begin{cases} G^{hg} & k=0 \\ 0 & \text{otherwise} \end{cases}$$

$$h=1,2,\ldots,\zeta;\ t=1,2,\ldots,T$$

This model will be referred to as MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$) model with $\underset{\sim}{p}$, $\underset{\sim}{\lambda}$ defined in Equation (7-7).

Setting $\underset{\sim}{p} = 0$ in the MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) model, we have,

$$\underset{\sim}{z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{q^{hg}} \sum_{\ell=0}^{m_k^{hg}} \theta_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{\varepsilon}^g(t-k) + \underset{\sim}{\varepsilon}^h(t) \tag{7-10}$$

$\underset{\sim}{\varepsilon}^h(t)$ is normally distributed wtih

$$E(\underset{\sim}{\varepsilon}^h(t)\underset{\sim}{\varepsilon}^g(t+k)') = \begin{cases} G^{hg} & k=0 \\ 0 & \text{otherwise} \end{cases}$$

$$h=1,2,\ldots,\ ;\ t=1,2,\ldots,T$$

This model is referred to as the MULSTMA($\zeta,\underset{\sim}{q},\underset{\sim}{m}$) model, with $\underset{\sim}{q}$, $\underset{\sim}{m}$ defined in Equation (7-7).

In the MULSTARMA, MULSTAR and MULSTMA models, the current h-category observations are expressed as weighted sum of previous observations, previous errors of all categories and the current errors of h-category. These models are inter-category dependent. Such inter-category dependence can be removed by setting $p^{hg} = 0$, $q^{hg} = 0$ where $h \neq g$. For example, setting $p^{hg} = 0$, $q^{hg} = 0$, $h \neq g$ in the MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) model, we have

$$\underset{\sim}{Z}^h(t) = \sum_{k=1}^{p^{hh}} \sum_{\ell=0}^{\lambda_k^{hh}} \phi_{k\ell}^{hh} W^{(\ell)} \underset{\sim}{Z}^h(t-k) - \sum_{k=1}^{q^{hh}} \sum_{\ell=0}^{m_k^{hh}} \theta_{k\ell}^{hh} W^{(\ell)} \underset{\sim}{\varepsilon}^h(t-k) + \underset{\sim}{\varepsilon}^h(t),$$

where $\underset{\sim}{\varepsilon}^h(t)$ is normally distributed, $\underset{\sim}{\varepsilon}^h \sim \mathrm{NID}(0, G^{hh})$,

$$h=1,2,\ldots,\ ;\ t=1,2,\ldots,T.$$

In the above formulation, it is seen that $\underset{\sim}{Z}^h(t)$ doesn't receive any contributions from observations or errors of the other categories. The above formulation contains ζ independent STARMA models. Similarly, by setting $p^{hg} = 0$, $h \neq g$, in the MULSTAR model formulation, we obtain the MULSTAR model without inter-category dependence, which contains ζ independent STAR models. By setting $q^{hg} = 0$, $h \neq g$, in the MULSTMA model formulation, we have the MULSTMA model without inter-category dependence, which contains independent STMA models.

7.1.2.2 System Simplification. In the univariate STARMA model, if the location number is reduced to one, i.e., LN=1, the formulation of univariate ARMA model is obtained. In this section the models that are obtained by reducing the system specifications are presented. The system specifications include the number of observation categories ζ and the number of locations LN.

By setting the category number ζ=1 and canceling the dummy superscript, we collapse the MULSTARMA(ζ,$\underset{\sim}{p}$,$\underset{\sim}{q}$,$\underset{\sim}{\lambda}$,$\underset{\sim}{m}$) model into the univariate STARMA model, i.e.,

$$\underset{\sim}{Z}(t) = \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} W^{(\ell)} \underset{\sim}{Z}(t-k) - \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell} W^{(\ell)} \underset{\sim}{\varepsilon}(t-k) + \underset{\sim}{\varepsilon}(t)$$

$$t=1,2,\ldots,T \tag{7-11}$$

$\underset{\sim}{\varepsilon}(t)$ is normally distributed with

$$E(\underset{\sim}{\varepsilon}(t)\underset{\sim}{\varepsilon}(t+k)') = \begin{cases} G & k=0 \\ 0 & \text{otherwise} \end{cases}$$

One step further, by setting $\lambda_k = 0$, all k, and imposing $W^{(0)} = I$ by definition, we have

$$\underset{\sim}{Z}(t) = \sum_{k=1}^{p} \phi_{k0}\, \underset{\sim}{Z}(t-k) - \sum_{k=1}^{q} \theta_{k0}\, \underset{\sim}{\varepsilon}(t-k) + \underset{\sim}{\varepsilon}(t).$$

This can be expressed alternatively as,

$$\underset{\sim}{Z}_i(t) = \sum_{k=1}^{p} \phi_{k0}\, Z_i(t-k) - \sum_{k=1}^{q} \theta_{k0}\, \underset{\sim}{\varepsilon}_i(t-k) + \varepsilon_i(t) \tag{7-12}$$

where the subscript i denotes the location index, and i=1,2,...,LN; t=1,2,...,T.

The above model formulation contains LN univeriate ARMA(p,q) processes. Thus the univariate STARMA model class, that describes only one category, is a subclass of the MULSTARMA model class, and the univariate ARMA model class, that describes only one location, is a subclass of the STARMA model class.

Setting the location number LN=1 in the MULSTARMA model and cancelling the dummy neighbor structure subscripts, we obtain the multivariate ARMA model, i.e.,

$$Z^h(t) = \sum_{k=1}^{p} \sum_{q=1}^{\zeta} \phi_k^{hg} Z^q(t-k) - \sum_{k=1}^{g} \sum_{g=1}^{\zeta} \theta_k^{hg} \varepsilon^g(t-k) + \varepsilon^h(t). \quad (7\text{-}13)$$

This multivariate ARMA model has been proposed by Box and Tiao [1981].

In summary, the multivariate ARMA model describes the inter-category correlations, but not the space-time correlations. The univariate STARMA model describes the space-time correlations, but not the inter-category correlations. Only the MULSTARMA model has the capability of describing the inter-category, space-time correlations.

7.1.3 Physical Interpretation of the MULSTARMA Model

The MULSTARMA model can be decomposed into elementary components, that transform the noise input or the feedback from previous observations to give contributions to the current output observations. These components are identifiable and the MULSTARMA model can be represented by a diagram that contains each identifiable filter component. Figure 7-1 shows the configuration of the elementary components of which the outputs are added to give the observation of Category C at location i for the ζ-category, LN-location MULSTARMA

process. The filters in Figure 7-1 represent the model components in a one-to-one corresponding relationship as follows;

$$T(i,j;\ g,c|\theta) = \sum_{k=1}^{q} \sum_{\ell=0}^{m_k} \theta_{k\ell}^{cg} W_{ij}^{(\ell)} B^k,$$

$$T(i,j;\ g,c|\phi) = \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell}^{cg} W_{ij}^{(\ell)} B^k.$$

The $T(j,i;\ g,c|\theta)$ and $T(j,i;\ g,c|\phi)$ filters describe the influences from the noise and observations of location j, category g to the observation of category c at location i. The "add" operator, that sums up the inputs to give the observation output, performs the addition operation. The MULSTARMA process of ζ-category and LN-location system consists of LN$\cdot\zeta$ such elementary construction units. Each elementary construction unit describes this transformation-then-summation mechanism for some category observation at some location. All these LN$\cdot\zeta$ elementary construction units are connected together to form an $\zeta\cdot$LN input-ports, LN output-ports of an open loop system for the STARMA model. The system inputs are current white noise and the system outputs are current observations. Cancelling the $T(j,i;\ g,c|\theta)$ filters, we obtain the construction unit for the MULSTAR process that is a simplified model of the MULSTARMA model. Connecting all the construction units of the LN-locations, ζ-categories, we obtain the open loop system for the MULSTAR model. Similarly, cancelling the $T(j,i;\ g,c|\phi)$ filters, we obtain the construction unit for the MULSTMA process that is another simplified model of the MULSTARMA model. Connecting all the construction units of the LN-locations, ζ-categories, we obtain the open loop system for the MULSTMA model.

Figures 7-2 to 7-7 are drawn to illustrate the MULSTARMA model and its collapsed models. Figure 7-2 shows the components diagram for the multivariate MA process of two observation categories systems, i.e., LN = 1, $\zeta = 2$. In the diagram, it is seen that $\varepsilon_1^1(t)$ contributes to $Z_1^1(t)$ as well as $Z_1^2(t)$, and $\varepsilon_1^2(t)$ contributes both to $Z_1^1(t)$. Therefore the inter-category dependence of moving average characteristics is implied. In Figure 7-3, the LN = 1, $\zeta = 2$ multivariate STAR system is illustrated. Here $\varepsilon_1^1(t)$ doesn't have any contribution to $\varepsilon_1^2(t)$, and the $\varepsilon_1^2(t)$ doesn't have any contribution to $Z_1^1(t)$. However, both previous $Z_1^1(t)$ and $Z_1^2(t)$ have contributions to $Z_1^1(t)$ and $Z_1^2(t)$. This diagram doesn't show any direct contribution from $\varepsilon_1^1(t)$ to $Z_1^2(t)$ and from $\varepsilon_1^2(t)$ to $Z_1^2(t)$, but following the AR feedback loops, we see that the past noise of one category, say $\varepsilon_1^1(t)$, does contribute to the observation of the same category at the same time, say past $Z_1^1(t)$, and the contribution of $\varepsilon_1^1(t)$ in $Z_1^1(t)$ is feedback to contribute to both the current $Z_1^1(t)$ and $Z_1^2(t)$. Therefore, implicitly past $\varepsilon_1^1(t)$ have contributions to current $Z_1^2(t)$ through the AR feedback loops. The mechanism discussed above is closely related to the well-known fact that AR models can be expressed alternatively as an infinite order MA process. Imposing Figure 7-2 on Figure 7-3, we obtain Figure 7-4, that contains the diagram for the multivariate ARMA process for LN = 1, $\zeta = 2$.

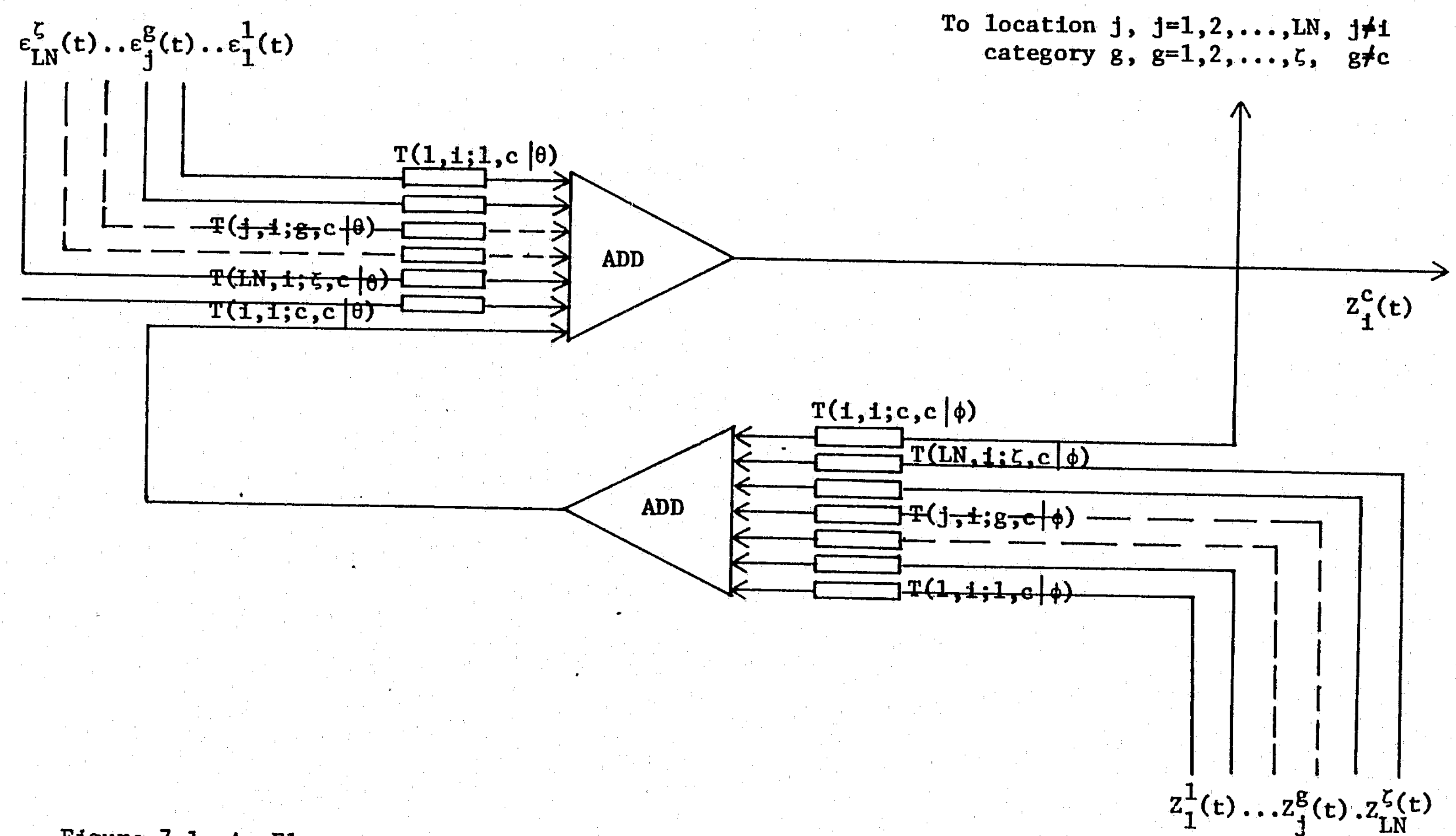


Figure 7-1. An Elementary Component (Category c, Location j) in A MULSTARMA Process of ζ Categories, LN Locations.

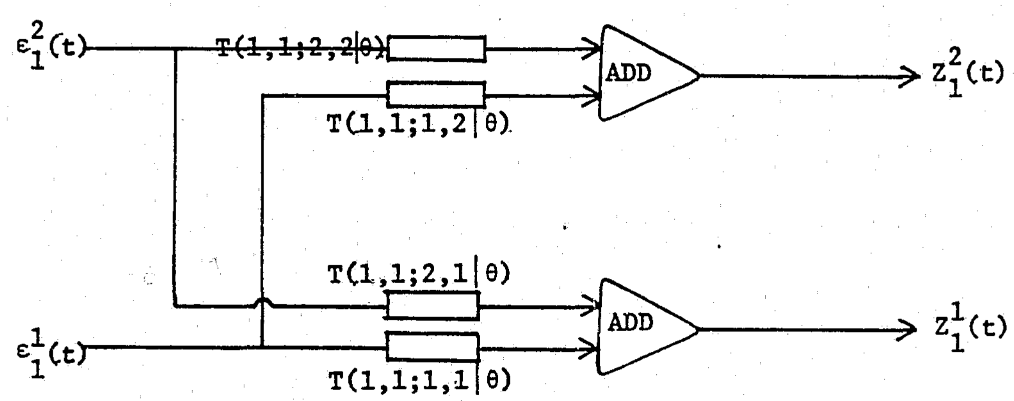


Figure 7-2. The Component Diagram of the Multivariate MA Process with ζ=2.

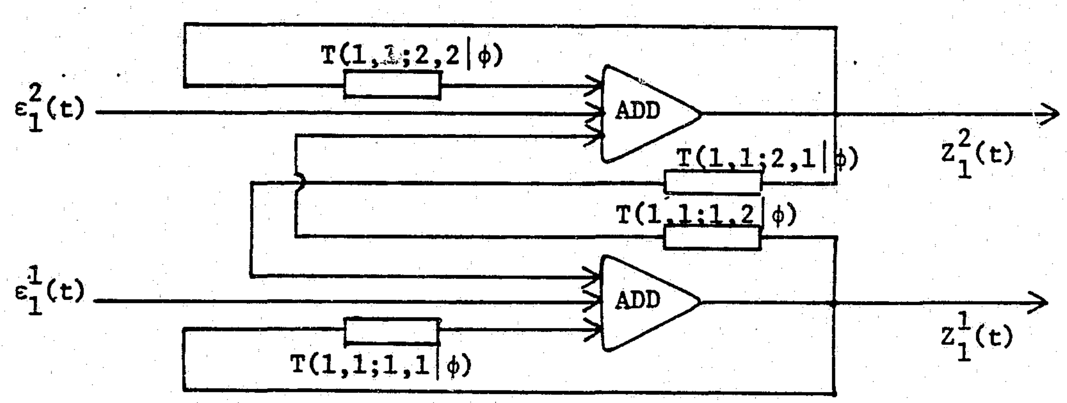


Figure 7-3. The Component Diagram of the Multivariate AR Process with ζ=2.

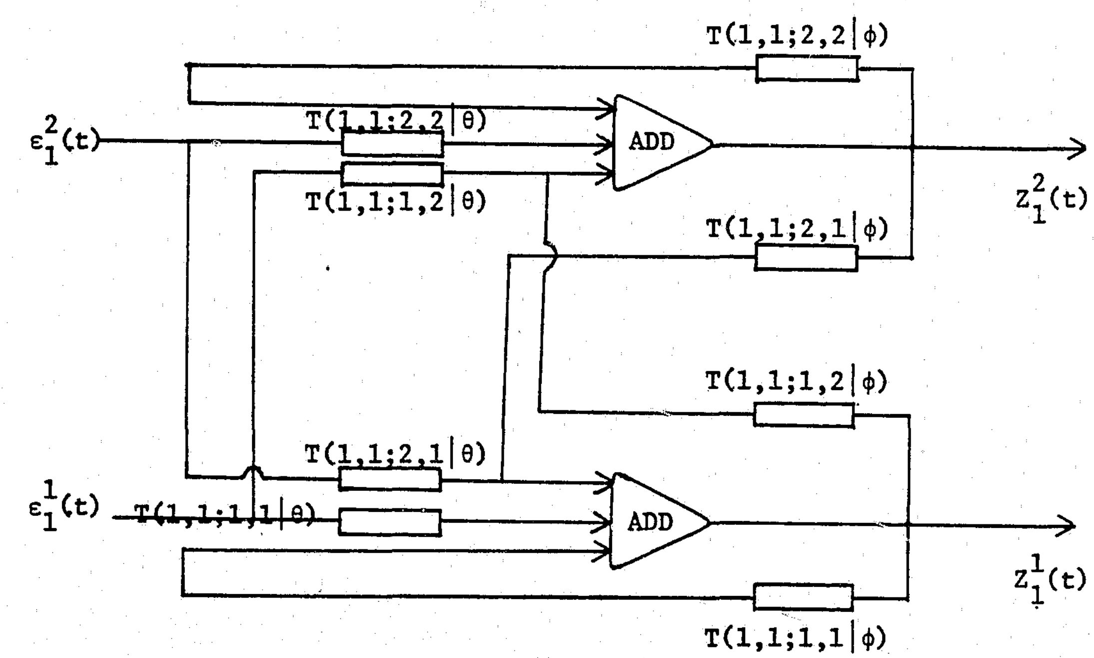


Figure 7-4. The Component Diagram of the Multivariate ARMA Process with ζ=2.

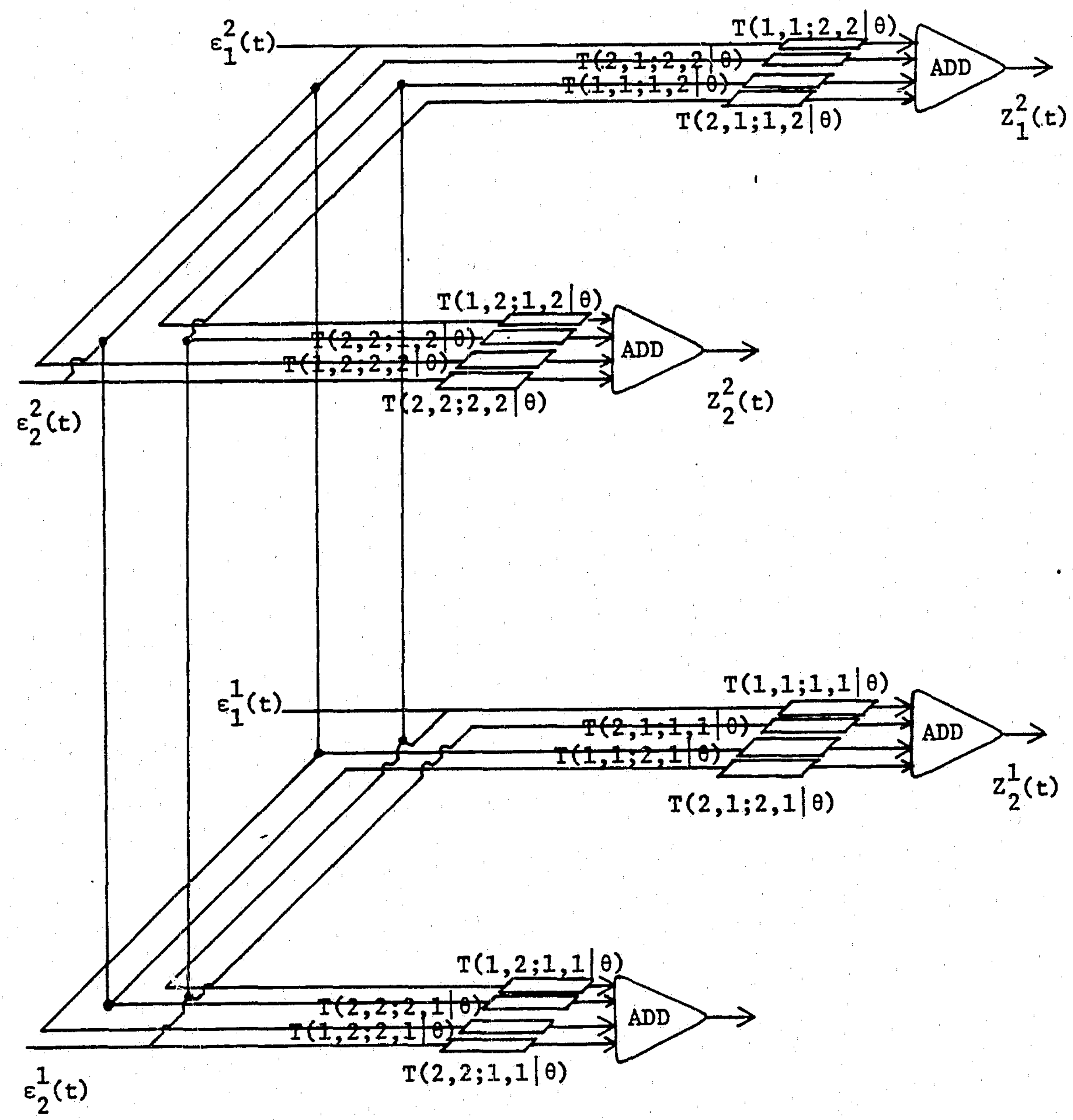


Figure 7-5. The Component Diagram of the MULSTMA Process with ζ=2, LN=2.

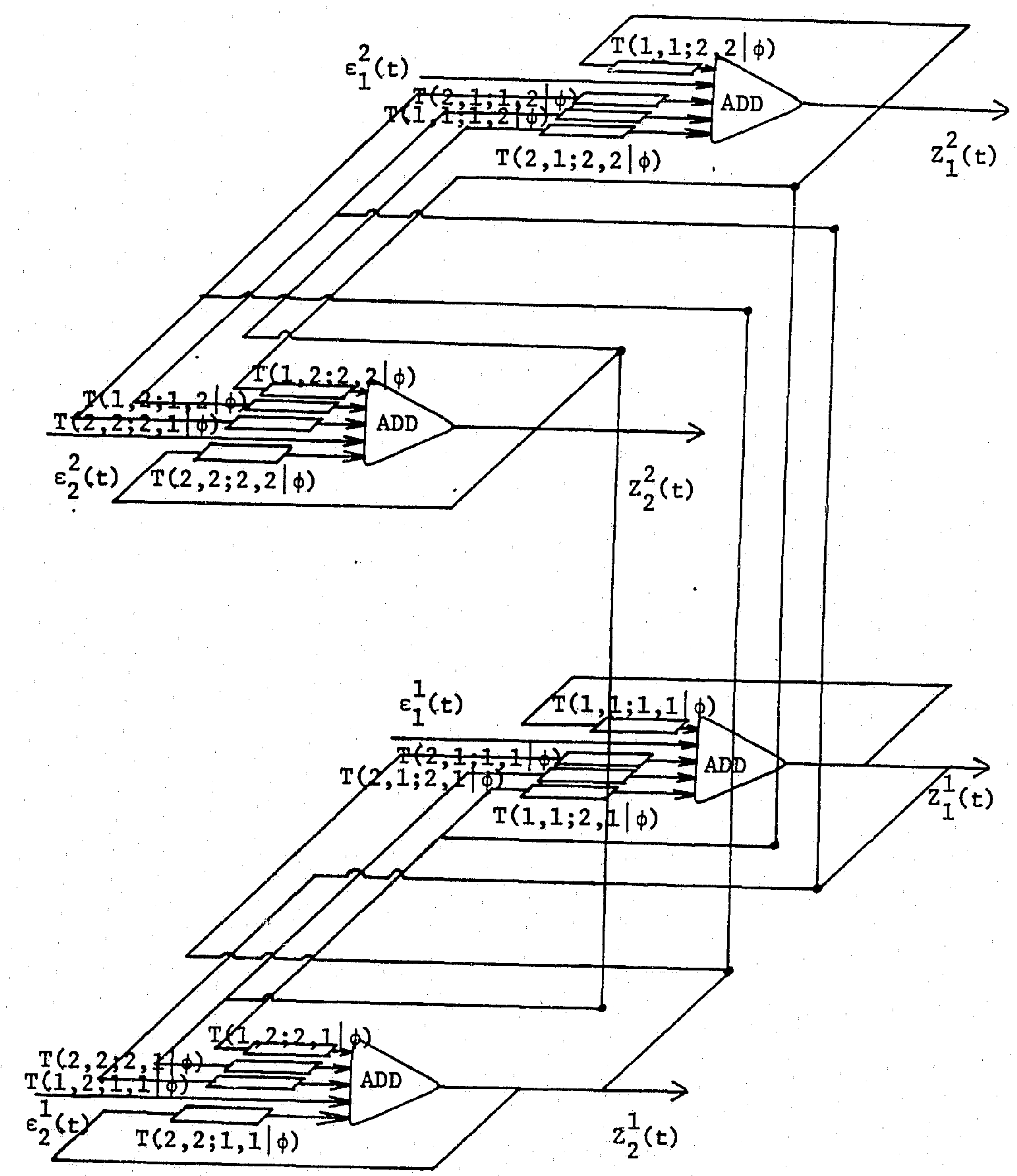


Figure 7-6. The Component Diagram of the MULSTAR Process with ζ=2, LN=2.

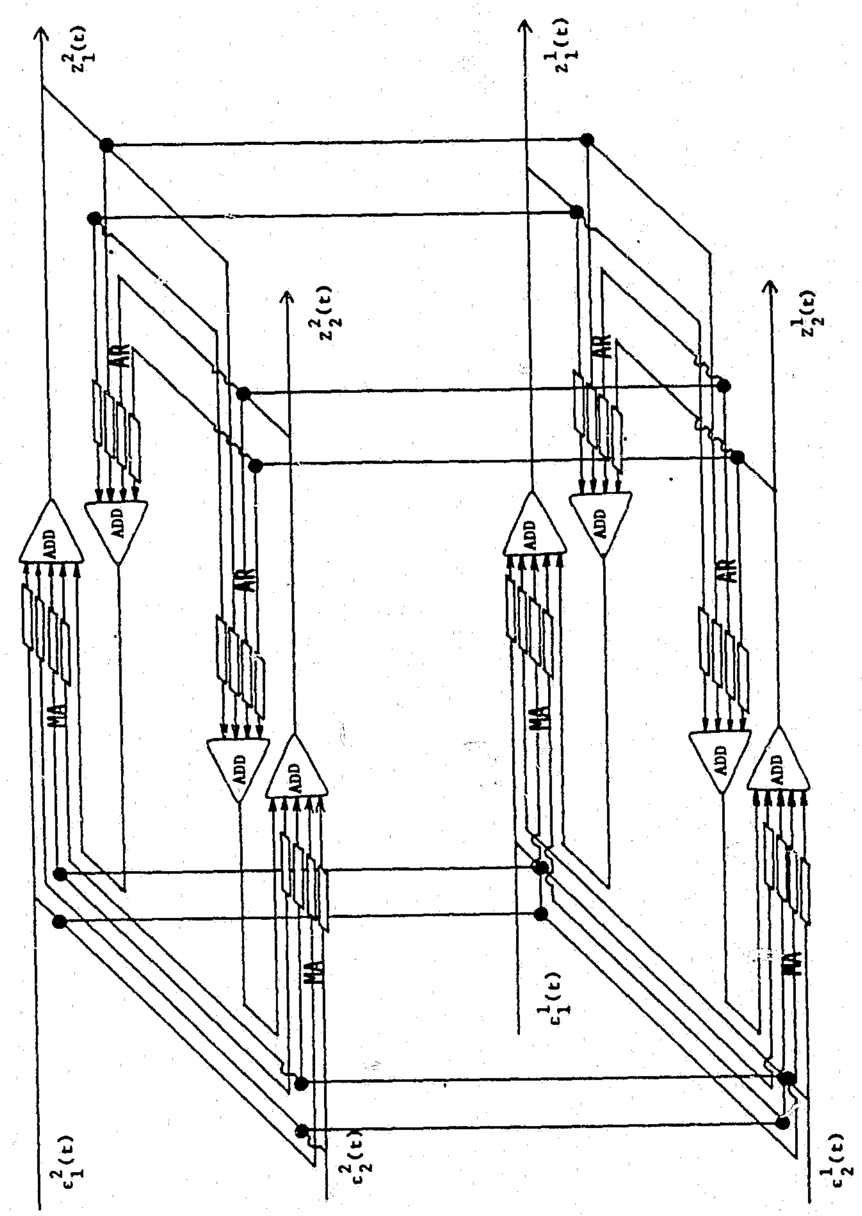


Figure 7-7. The Component Diagram of the MULSTARMA Process with ζ=2, LN=2.

Figure 7-5 shows the MULSTMA process of LN = 2, ζ = 2 system. This diagram contains four MULSTMA construction units. The observations do not feedback. The past noise of each category at each location has contributions to all the current observations. Disconnecting the inter-category filters, or breaking the inter-category dependence, i.e., separating the 2nd floor from the 1st floor, we have two independent STMA process diagrams. Disconnecting the spatial filters, i.e., separating the front wall from the rear wall, we obtain two independent multivariate MA process diagrams. Figure 7-6 contains the diagram for MULSTAR model of LN = 2, ζ = 2 system. This diagram has four MULSTAR construction units. The past observations is feedback to the current process outputs. Similar to that of multivariate AR process in Figure 7-3, the past noise has contributions to all the current observations implicitly. The past noise of either category at either location are input to the addition operator through the AR feedback loops to give contributions to all the current observations. Imposing Figure 7-5 and 7-6 gives Figure 7-7, which is the diagram of the MULSTARMA process of LN = 2, ζ = 2 system. The diagram 7-7 consists of four MULSTARMA construction units, the past noise of either category at each location has explicitly contributions to all the current observations. The past observations are feedback to all construction units.

7.2 Stationary and Invertible Conditions

An important consideration for the MULSTARMA model are the conditions under which the vector process $\underset{\sim}{Z}(t)$ is strictly stationary.

For a strictly stationary process $\underset{\sim}{Z}(t)$, the distributions of $\underset{\sim}{Z}(t_1),\ldots,\underset{\sim}{Z}(t_n)$ and $\underset{\sim}{Z}(t_1+s),\ldots,\underset{\sim}{Z}(t_n+s)$ are the same, this guarantees the property that the covariance of $\underset{\sim}{Z}(t_1)$ and $\underset{\sim}{Z}(t_1+T)$, i.e., $E(\underset{\sim}{Z}(t_1)\underset{\sim}{Z}'(t_1+T))$, depends only on T regardless of the t_1 value, i.e., $E(\underset{\sim}{Z}(t_1)\underset{\sim}{Z}'(t_1+T)) = E(\underset{\sim}{Z}(t_2)\underset{\sim}{Z}'(t_2+T))$ for arbitrary t_1, t_2. The invertibility is the dual of the stationarity. The invertibility property assures that theprocess $\underset{\sim}{Z}(t)$ depends on the latest observations most heavily. In the following, the stationarity regions for the MULSTARMA process are derived in Section 8.2.1. The derivation is followed by the similar derivation of the invertibility regions in Section 8.2.2.

7.2.1 Stationarity Regions

Stationarity implies that for any s and t, the joint density distribution of $\underset{\sim}{Z}(t),\underset{\sim}{Z}(t+1),\ldots,\underset{\sim}{Z}(t)$ and $\underset{\sim}{Z}(t+s),\underset{\sim}{Z}(t+1+s),\ldots,\underset{\sim}{Z}(T+S)$ are the same. In modeling we are particularly concerned with second order stationarity. Under the assumptions of normal distribution and $E(\underset{\sim}{Z}(t)) = \underset{\sim}{0}$, this stationarity means that

$$E(\underset{\sim}{Z}(t)\underset{\sim}{Z}(t+s)') = E(\underset{\sim}{Z}(0)\underset{\sim}{Z}(s)')$$

for all t. This is equivalent to say that the variance-covariance of $\underset{\sim}{Z}(t)$ and $\underset{\sim}{Z}(t+s)$ depends only upon the time difference s, it doesn't depend on time t. Without this stationarity property, the statistical property of the process change with time and the system becomes explosive. The parameter constraints to insure stationarity, form the stationary region.

By applying the result from Hannan [1970] (Equation (7-14)) the stationary conditions for the General Multivariate ARMA Model can be derived. If every u that solves.

$$\det[\mu^{p_{max}} I - \sum_{k=1}^{p_{max}} B(k)\mu^{p_{max}-k}] = 0 \qquad (7\text{-}14)$$

lies inside the unit circle, i.e., $|u| < 1$, then the vector process $\underset{\sim}{Z}(t)$ for the general multivariate ARMA process will be stationary.

Hence these stationary conditions put restrictions on the autoregressive parameter matrices only, and all the General Multivariate MA models are always stationary. Thus the stationary conditions of the MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda}$,m) model will be exactly the same as that of the MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$).

Consider the MULSTAR(ζ,p,λ) model with $\underset{\sim}{p} \le \underset{\sim}{1}$, we have $p_{max} = 1$, and the stationary conditions so that every that solves

$$\det[\mu I - B(1)] = 0 \qquad (7\text{-}15)$$

lies inside the unit circle, i.e., $|\mu| < 1$. Note that the solution μ for $p_{max} = 1$ models is the eigenvalue of coefficient matrix B(1). It is known that the eigenvalue μ of the matrix $B = [b_{ij}]$ is bound within the region that

$$|\mu - b_{ii}| \le \sum_{j=1} b_{ij}, \text{ all } i \qquad (7\text{-}16)$$

Since $B(1) = [B^{hg}(1)]$, $B^{hg}(1) = \sum_{\ell=0}^{\lambda_1^{hg}} \phi_1^{hg} w^{(\ell)}$, where $w^{(\ell)}$ is the scaled weight matrix and $\sum_j w_{ij}^{(\ell)} = 1$ all i, the bounds for the eigenvalue of B(1) are

$$|\mu - \phi_{10}^{hh}| \le |\sum_{g\neq h}^{\zeta} \phi_{10}^{hg}| + \sum_{\ell=1}^{\lambda_1^{h*}} |\sum_{g=1}^{\zeta} \phi_{1\ell}^{hg}|, \quad h=1,2,\ldots,\zeta \tag{7-17}$$

where $\lambda_1^{h*} = \max\{\lambda_1^{hg} | g=1,2,\ldots,\zeta\}$.

Equation (8-17) is equivalent to

$$\phi_{10}^{hh} - |\sum_{g\neq h}^{\zeta} \phi_{10}^{hg}| - \sum_{\ell=1}^{\lambda_1^{h*}} |\sum_{g=1}^{\zeta} \phi_{1\ell}^{hg}| \le \mu \le \phi_{10}^{hh} + |\sum_{g\neq h}^{\zeta} \phi_{10}^{hg}| + \sum_{\ell=1}^{\lambda_1^{h*}} |\sum_{g=1}^{\zeta} \phi_{1\ell}^{hg}|$$

$$h=1,2,\ldots,\zeta \tag{7-18}$$

So the necessary condition for $|\mu| < 1$ becomes that

$$\phi_{10}^{hh} - |\sum_{g\neq h}^{\zeta} \phi_{10}^{hg}| - \sum_{\ell=1}^{\lambda_1^{h*}} |\sum_{g=1}^{\zeta} \phi_{1\ell}^{hg}| > -1 \tag{7-19}$$

and

$$\phi_{10}^{hh} + |\sum_{g\neq h}^{\zeta} \phi_{10}^{hg}| + \sum_{\ell=1}^{\lambda_1^{h*}} |\sum_{g=1}^{\zeta} \phi_{1\ell}^{hg}| < 1$$

$$h=1,2,\ldots,\zeta.$$

It should be noted that the region restricted by Equation (7-19) is the necessary stationary region, which is not necessarily the best stationary region for this MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$) process. The best stationary region is the region that every parameter that fall in this region makes the process stationary and any parameters not contained in this region makes the process nonstationary.

For illustration, let us consider the stationary region for the MULSTAR($2,\underset{\sim}{p},\underset{\sim}{\lambda}$) model with $\underset{\sim}{p} = (1,1,1,1)$, $\underset{\sim}{\lambda} = (0,0,0,0)$. Directly applying Equation (7-19) and setting $\lambda_1^{h*} = 0$ for h=1,2, we get the necessary stationary region in the $\underset{\sim}{\phi}$ parameter space to be,

$$\begin{aligned} \phi_{10}^{11} - |\phi_{10}^{12}| &> -1 \\ \phi_{10}^{11} + |\phi_{10}^{12}| &< 1 \\ \phi_{10}^{22} - |\phi_{10}^{21}| &> -1 \\ \phi_{10}^{22} + |\phi_{10}^{21}| &< 1 \end{aligned} \tag{7-20}$$

Thus the necessary stationary region for this MULSTAR process is restricted to the interior region of the configurations shown in Figure 7-8(a).

As another example, we consider the MULSTAR$(2,\underset{\sim}{p},\underset{\sim}{\lambda})$ model with $\underset{\sim}{p} = (1,1,1,1)$ and $\underset{\sim}{\lambda} = (1,0,0,1)$. Applying Equation (7-19) directly and setting $\lambda_1^{h*} = 1$ for h=1,2, we get the necessary stationary region in the $\underset{\sim}{\phi}$ space as,

$$\begin{aligned} &\phi_{10}^{11} - |\phi_{10}^{12}| - |\phi_{11}^{11}| > -1 \\ &\phi_{10}^{11} + |\phi_{10}^{12}| + |\phi_{11}^{11}| < 1 \\ &\phi_{10}^{22} - |\phi_{10}^{21}| - |\phi_{11}^{22}| > -1 \\ &\phi_{10}^{22} + |\phi_{10}^{21}| + |\phi_{11}^{22}| < 1 \end{aligned} \qquad (7\text{-}21)$$

Thus the necessary stationary region for the MULSTAR$(2,\underset{\sim}{p}=\underset{\sim}{1},\underset{\sim}{\lambda})$ process with $\underset{\sim}{\lambda} = (1,0,0,1)$ is the interior region of the diamond configuration for a ϕ_{11}^{11} and ϕ_{11}^{22} value: These regions are illustrated in Figure 7-8(b).

7.2.2 Invertibility Regions

Invertibility is the dual of stationarity. Stationarity puts restrictions on the autoregressive parameters, while invertibility puts restrictions on the moving average parameters. A non-invertible process implies that the current observation cumulates heavier influence from the earlier observations and errors. This can be seen clearly in the univariate time series. Consider the univariate MA(1) process

$$\underset{\sim}{Z}(t) = -\theta_1 B\underset{\sim}{\varepsilon}(t) + \underset{\sim}{\varepsilon}(t) \qquad (7\text{-}22)$$

$$t=1,2,\dots,T$$

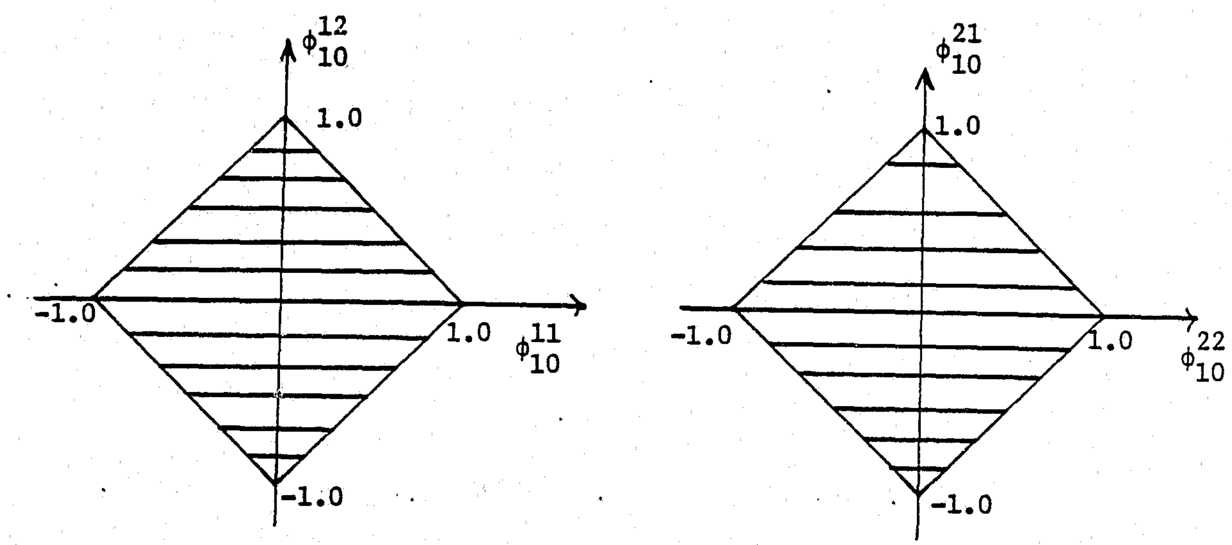


Figure 7-8(a). The Necessary Stationary Region for the MULSTAR$(2,\underset{\sim}{p},\underset{\sim}{\lambda})$ Process with $\underset{\sim}{p}=\underset{\sim}{1}$ and $\underset{\sim}{\lambda} = (1,0,0,1)$

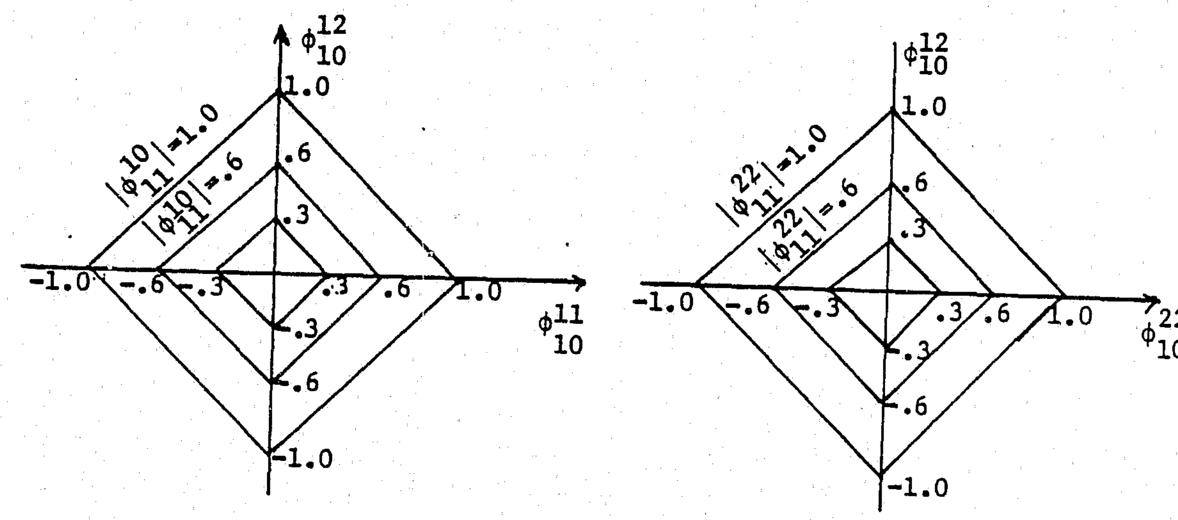


Figure 7-8(b). The Necessary Stationary Region for the MULSTAR$(2,\underset{\sim}{p},\underset{\sim}{\lambda})$ Process with $\underset{\sim}{p}=\underset{\sim}{1}$ and $\underset{\sim}{\lambda}=\underset{\sim}{0}$.

where B is the back shift operator defined as $B\underset{\sim}{\varepsilon}(t) = \underset{\sim}{\varepsilon}(t-1)$. Obviously, when $|\theta| > 1$, the process is not invertible, and $\underset{\sim}{Z}(t)$ will be influenced heavier by the $\theta_1\underset{\sim}{\varepsilon}(t-1)$ term than the current error. Also without invertibility we cannot express $\varepsilon(t)$ in terms of past observations, i.e., $\underset{\sim}{Z}(t-k)$, $k \geq 1$,

$\varepsilon(t) = \underset{T_0\to\infty}{\text{limit}} \sum_{n=0}^{T_0} (\theta_1\ B)^n \underset{\sim}{Z}(t)$, because we can expand $(I-\theta_1\ B)^{-1} = \sum_{n=0} (\theta_1\ B)^n$ only when $|\theta_1| < 1$. Furthermore, even we put on these conditions that $\underset{\sim}{Z}(t) = 0$ for $t \leq 0$ are imposed to make it possible to express $\varepsilon(t)$ in terms of past observations, we have

$$\underset{\sim}{\varepsilon}(t) = \sum_{k=0}^{t-1} \theta_1^k \underset{\sim}{Z}(t-k) \tag{7-23}$$

or

$$\underset{\sim}{Z}(t) = \underset{\sim}{\varepsilon}(t) - \sum_{k=1}^{t-1} \theta_1^k \underset{\sim}{Z}(t-k) \qquad t \geq 2$$

Without imposing invertibility constraints, $|\theta_1|$ can be > 1 and thus $\underset{\sim}{\varepsilon}(t)$ and $\underset{\sim}{Z}(t)$ will cumulate more influence from historically distant observations. Intuitively, more recent observations should have more similarity and influence on the current observations than the distant observations.

Apply the result from Hannan [1970] again, we get the invertible

conditions for the General Multivariate Space-Time Model can be derived. If every μ that solves

$$\det[\mu^{q_{max}} I - \sum_{k=0}^{q_{max}} A(k)\mu^{q_{max}-k}] = 0 \tag{7-24}$$

lies inside the unit circle, then the vector process will be invertible. Hence, the MULSTAR models are always invertible, and the invertible region on the moving average parameter space will be exactly the same for both MULSTMA($\zeta,\underset{\sim}{q},\underset{\sim}{m}$) model and MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) model.

Consider the MULSTMA($\zeta,\underset{\sim}{q},\underset{\sim}{m}$) model with $q \leq 1$. We have $q_{mqx} = 1$ and the invertible conditions is such that every μ that solves

$$\det[\mu I - A(1)] = 0 \tag{7-25}$$

lies inside the unit circle. This is equivalent to say that every eigenvalue of coefficient matrix A(1) lies inside the unit circle. Following the same procedure that we derive the stationary conditions for the MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$) model with $p \leq 1$, we get the necessary conditions for invertibility as,

$$\theta_{10}^{hh} - \Big|\sum_{g\neq h}^{\zeta} \theta_{10}^{hg}\Big| - \sum_{\ell=1}^{m_1^{h*}} \Big|\sum_{g=1}^{\zeta} \theta_{1\ell}^{hg}\Big| > -1$$

$$\theta_{10}^{hh} + \Big|\sum_{g\neq h}^{\zeta} \theta_{10}^{hg}\Big| + \sum_{\ell=1}^{m_1^{h*}} \Big|\sum_{g=1}^{\zeta} \theta_{1\ell}^{hg}\Big| < 1 \qquad (7\text{-}26)$$

$$h=1,2,\ldots,\zeta$$

Note that this is the necessary invertible regions, and not necessarily is the best invertible regions.

For illustration, the invertible region for the MULSTMA(2,$\underset{\sim}{q}$,$\underset{\sim}{m}$) model with $\underset{\sim}{q}$ = (1,1,1,1), $\underset{\sim}{m}$ = (0,0,0,0) is considered. Applying Equation (7-25), we have the necessary invertible region in the parameter space to be,

$$\theta_{10}^{11} - |\theta_{10}^{12}| > -1$$

$$\theta_{10}^{11} + |\theta_{10}^{12}| < 1$$

$$\theta_{10}^{11} + |\theta_{10}^{21}| > -1$$

$$\theta_{10}^{22} + |\theta_{10}^{21}| < 1.$$

Thus the necessary stationary region for this MULSTMA process is the interior region of the configurations shown in Figure 7-9(a).

As another example, lets consider the invertible region of the MULSTMA(2,$\underset{\sim}{q}$,$\underset{\sim}{m}$) with $\underset{\sim}{q}$ = (1,1,1,1) and $\underset{\sim}{m}$ = (1,0,0,1). Applying Equation

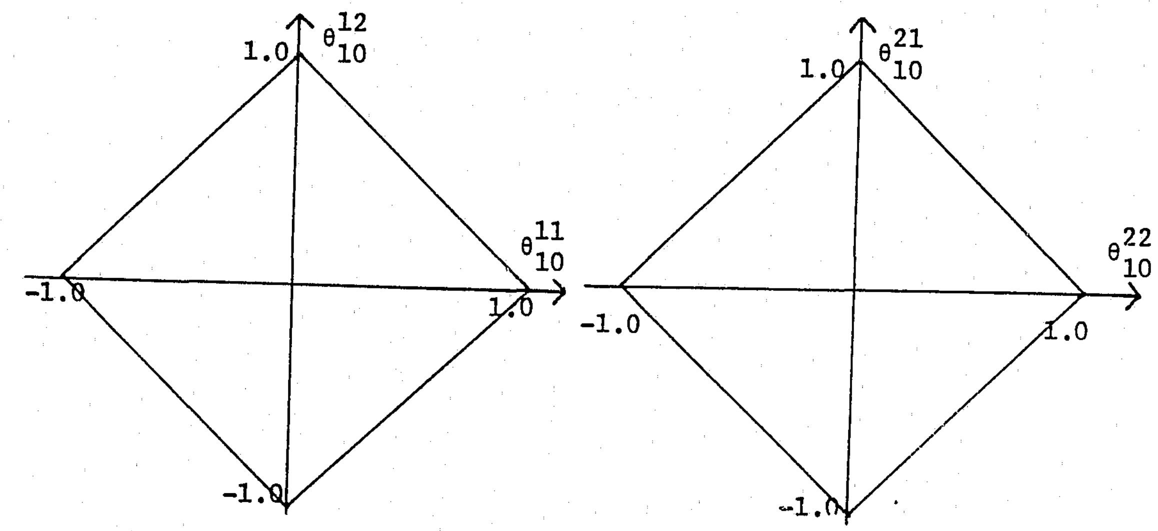


Figure 7-9(a) The Necessary Invertible Region of the MULSTMA(2,$\underset{\sim}{q}$,$\underset{\sim}{m}$) Process with $\underset{\sim}{q}$=1 and $\underset{\sim}{m}$=0

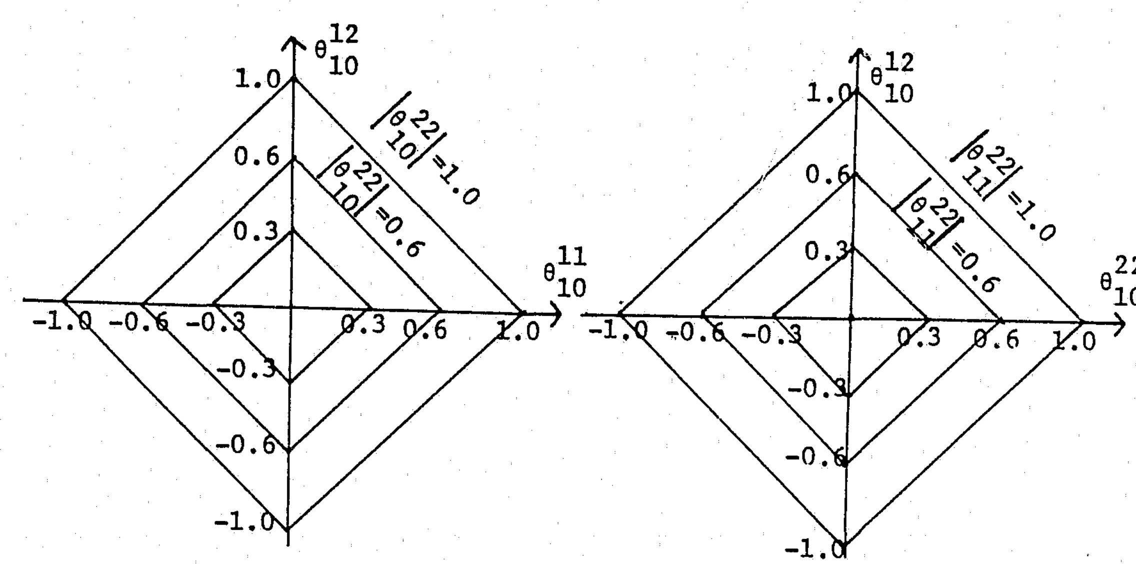


Figure 7-9(b). The Necessary Invertible Region of the MULSTMA(2,$\underset{\sim}{q}$,$\underset{\sim}{m}$) Process with $\underset{\sim}{q}$=1 and $\underset{\sim}{m}$=(1,0,0,1).

(7-25) and setting $m_{11}^{h*} = 1$ for h = 1,2, we obtain the necessary invertible region in the $\underset{\sim}{\theta}$ parameter space to be,

$$\theta_{10}^{11} - |\theta_{10}^{12}| - |\theta_{11}^{11}| > -1$$

$$\theta_{10}^{11} + |\theta_{10}^{12}| + |\theta_{11}^{11}| < 1$$

$$\theta_{10}^{22} - |\theta_{10}^{21}| - |\theta_{11}^{22}| > -1$$

$$\theta_{10}^{22} + |\theta_{10}^{21}| + |\theta_{11}^{22}| < 1.$$

Thus the necessary invertible region is restricted to the interior region of the diamond configuration for a given θ_{11}^{11} and θ_{11}^{22} value. These regions are illustrated in Figure 7-9(b).

7.3 Λ-Weight Representation for the Stationary General Multivariate Space-Time Models

Any model that meets the requirement of stationarity can be expressed alternatively as weighted sum of past and current errors,

$$\underset{\sim}{z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{\infty} \Lambda^{hg}(k)\underset{\sim}{\varepsilon}^g(t-k) + \underset{\sim}{\varepsilon}^h(t) \qquad (7\text{-}27)$$

$$t=1,2,\ldots,T;\ h=1,2,\ldots,\zeta$$

This will be referred to as the Λ-weight representation.

In the following, we will derive the relations between the coefficient matrices of the Λ-weight representation and the coefficient

matrices of the General Multivariate representation. To put the General Multivariate MA model in the Λ-Weight representation representation the coefficients of the same $\varepsilon^q(t-k)$ terms in the Λ-weight representation and the General Multivariate MA model are equated. This yields the relationship,

$$\Lambda^{hg}(k) = \begin{cases} -A^{hg}(k), & k \le q^{hg} \\ 0 & \text{otherwise} \end{cases} \qquad (7\text{-}28)$$

Note that $\Lambda^{hg}(k)$ is null matrix when $k > q^{hg}$.

To put this General Multivariate AR model into the Λ-weight representation is not as straightforward as that of the General Multivariate MA model.

For the General Multivariate AR process,

$$\underset{\sim}{z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} B^{hg}(k)\underset{\sim}{z}^g(t-k)+\underset{\sim}{\varepsilon}^h(t)$$

$$h=1,2,\ldots,\zeta;\ t=1,2,\ldots,T$$

the $\underset{\sim}{z}^g(t-k)$ terms are first substituted for by these equivalent Λ-weight representation. Then the coefficients of the resulting equations are equated with the corresponding coefficients of the Λ-weight representation. The results are as follows,

$$\Lambda^{hg}(1) = B^{hg}(1)$$

$$\Lambda^{hg}(2) = B^{hg}(2) + \sum_{g_2=1}^{\zeta} B^{hg_2}(1)\Lambda^{g_2 g}(1)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\Lambda^{hg}(k) = B^{hg}(k) + \sum_{g_2=1}^{\zeta} \sum_{k_2=1}^{k-1} B^{hg_2}(k_2)\Lambda^{g_2 g}(k-k_2) \quad , k \le p^{hg}$$

$$\Lambda^{hg}(k) = \sum_{g_2=1}^{\zeta} \sum_{k_2=1}^{k-1} B^{hg_2}(k_2)\Lambda^{g_2 g}(k-k_2) \quad , k > p^{hg}$$

$$h=1,2,\ldots,\zeta;\ g=1,2,\ldots, \quad . \tag{7-29}$$

It should be noted that the elements of $\Lambda^{hg}(k)$ will exponentially decay to zero for $k > p^{hg}$ because of the matrix multiplication operation in the recursive formula.

Following the same procedure, the Λ-weight representation for the General Multivariate STARMA Model with the Λ-weights matrix expressed in terms of A's, B's is,

$$\Lambda^{hg}(1) = -A^{hg}(1) + B^{hg}(1)$$

$$\Lambda^{hg}(2) = -A^{hg}(2) + B^{hg}(2) + \sum_{g_2=1}^{\zeta} B^{hg_2}(1)\Lambda^{g_2 g}(1)$$

For $k \le \min\{p^{hg}, q^{hg}\}$

$$\Lambda^{hg}(k) = -A^{hg}(k) + B^{hg}(k) + \sum_{g_2=1}^{\zeta} \left[\sum_{j=1}^{k-1} B^{hg_2}(j)\Lambda^{g_2 g}(k-j)\right]$$

and for $\min\{p^{hg}, q^{hg}\} < k \le \max\{p^{hg}, q^{hg}\}$, we have two situations:

1. If $p^{hg} > q^{hg}$

$$\Lambda^{hg}(k) = B^{hg}(k) + \sum_{g_2=1}^{\zeta} \left[\sum_{j=1}^{k-1} B^{hg_2}(j)\Lambda^{g_2 g}(k-j)\right] \text{, or}$$

2. If $p^{hg} < q^{hg}$

$$\Lambda^{hg}(k) = -A^{hg}(k) + \sum_{g_2=1}^{\zeta} \sum_{j=1}^{p^{hg}} B^{hg_2}(j)\Lambda^{g_2 g}(k-j)$$

For $k > \max\{p^{hg}, q^{hg}\}$

$$\Lambda^{hg}(k) = \sum_{g_2=1}^{\zeta} \sum_{j=1}^{\tilde{p}^{hg}} B^{hg_2}(j)\Lambda^{g_2 g}(k-j) . \tag{7-30}$$

Similar to the MULSTAR case, the elements of $\Lambda^{hg}(k)$ will exponentially decay to zero when $k > \max\{p^{hg}, q^{hg};\ h,g=1,2,\ldots,\zeta\}$.

7.4 Multivariate Space-Time Autocorrelation Function Considerations

In this section, some statistics that will be very useful in identifying the potential candidate models in the model building are addressed. First the covariance matrix $\Gamma^{ij}(u)$ of the observations $\underset{\sim}{Z}^{i}(t)$ and $\underset{\sim}{Z}^{j}(t+u)$ is defined as,

$$\Gamma^{ij}(u) = E(\underset{\sim}{Z}^{j}(t+u)\underset{\sim}{Z}^{i}(t)') \tag{7-31}$$

$$i,j=1,2,\ldots,\zeta;\ u=\ldots,-2,-1,0,1,\ldots$$

The multivariate space-time autocovariance function $\gamma^{sh}_{s\ell}(u)$ is defined as

$$\gamma^{gh}_{s\ell}(u) = \frac{E[(w^{(s)}Z^{g}(t))'(w^{(\ell)}Z^{h}(t+u))]}{LN} \qquad (7\text{-}32)$$

Since

$$\begin{aligned}\gamma^{gh}_{s\ell}(u) &= \frac{E\{Tr[(w^{(s)}Z^{g}(t))'(w^{(\ell)}Z^{h}(t+u))]\}}{LN}\\ &= \frac{E\{Tr[w^{(s)\prime}w^{(\ell)}Z^{h}(t+u)Z^{g}(t)']\}}{LN}\\ &= \frac{Tr\{w^{(s)\prime}w^{(\ell)}E[Z^{h}(t+u)Z^{g}(t)']\}}{LN}\\ &= \frac{Tr(w^{(s)\prime}w^{(\ell)}\Gamma^{gh}(u))}{LN} \qquad (7\text{-}33)\end{aligned}$$

$\gamma^{sh}_{s\ell}(u)$ can be expressed alternatively as

$$\gamma^{gh}_{s\ell}(u) = \frac{(w^{(s)\prime}w^{(\ell)}) \odot \Gamma^{gh}(u)}{LN} \qquad (7\text{-}34)$$

A property for the multivariate space-time autocovariance function $\Gamma^{gh}_{s\ell}(u)$ such defined is that $\Gamma^{gh}_{s\ell}(-u) = \Gamma^{hg}_{s\ell}(u)$. According to the definition of $\Gamma^{ij}(u)$ in Equation (7-31), it can be shown that

$$\Gamma^{ij}(u) = \Gamma^{ji}(-u)'. \qquad (7\text{-}35)$$

So from expression, Equation (7-33), we have

$$\begin{aligned}\gamma^{gh}_{s\ell}(-u) &= \frac{1}{LN}Tr(w^{(s)\prime}w^{(\ell)}\Gamma^{gh}(-u))\\ &= \frac{1}{LN}Tr(w^{(s)\prime}w^{(\ell)}\Gamma^{hg}(u)')\\ &= \frac{1}{LN}Tr(\Gamma^{hg}(u)w^{(\ell)\prime}w^{(s)}) \quad \because Tr(A)=Tr(A')\\ &= \frac{1}{LN}Tr(w^{(\ell)\prime}w^{(s)}\Gamma^{hg}(u))\end{aligned}$$

and we get the relation

$$\gamma^{gh}_{s\ell}(-u) = \gamma^{hg}_{\ell s}(u). \qquad (7\text{-}36)$$

7.4.1 The Multivariate Space-Time Autocorrelation Function

In univariate STARMA models, the space-time autocorrelation function $\rho_{k\ell}(u)$ was defined as (Pfeifer and Deutsch [1980b]).

$$\rho_{k\ell}(u) = \frac{\gamma_{k\ell}(u)}{(\gamma_{kk}(0)\gamma_{\ell\ell}(0))^{1/2}} \qquad (7\text{-}37)$$

For the multivariate STARMA models, there are two alternatives in defining the multivariate space-time autocorrelation function.

$$\rho_{k\ell}^{gh}(u) = \frac{\gamma_{k\ell}^{gh}(u)}{(\gamma_{kk}^{gg}(0)\gamma_{\ell\ell}^{gg}(0)\gamma_{kk}^{hh}(0)\gamma_{\ell\ell}^{hh}(0))^{1/4}} \tag{7-38}$$

$g,h=1,2,\ldots,\zeta$

or

$$\rho_{k\ell}^{gh}(u) = \frac{\gamma_{k\ell}^{gh}(u)}{(\gamma_{kk}^{gh}(0)\gamma_{\ell\ell}^{gh}(0))^{1/2}} \tag{7-39}$$

$g,h=1,2,\ldots,\zeta$

Both definitions collapse to the definition of the univariate space-time autocorrelation function when we set $\zeta = 1$, cancel the dummy superscripts. However, from the physical interpretation of these alternatives of $\rho_{kk}^{gh}(0)$, the first definition (Equation (7-38)) is the appropriate choice. According to the second definition, Equation (7-39), by setting u=0, we have

$$\rho_{kk}^{gh}(0) = 1, \quad g,h=1,2,\ldots,\zeta; \; k=0,1,\ldots \tag{7-40}$$

This means that $\underset{\sim}{Z}^{g}(t)$ and $\underset{\sim}{Z}^{h}(t)$ are perfectly correlated, even the observations are not of the same category which is not reasonable.

Whereas for the first definition,

$$\rho_{kk}^{gh}(0) = \frac{\gamma_{kk}^{gh}(0)}{(\gamma_{kk}^{gg}(0)\gamma_{kk}^{hh}(0))^{1/2}} \tag{7-41}$$

$\rho_{kk}^{gh}(0) = 1$ only when g=h. This observation of any two different categories are perfectly correlated.

The sample multivariate space-time autocovariance function $\hat{\gamma}_{s\ell}^{gh}(u)$ and the sample multivariate space-time autocorrelation function $\hat{\rho}_{s\ell}^{gh}(u)$ are defined analogously to their theoretical corresponding forms. Thus

$$\hat{\gamma}_{s\ell}^{gh}(u,T) = \frac{\sum_{t=1}^{T-u} [(w^{(s)}\underset{\sim}{Z}^{g}(t))'(w^{(\ell)}\underset{\sim}{Z}^{h}(t+u))]}{LN\cdot(T-u)} \tag{7-42}$$

$g,h=1,2,\ldots,\zeta; \; s,\ell,u=0,1,2,\ldots$

and

$$\hat{\rho}_{s\ell}^{gh}(u,T) = \frac{\hat{\gamma}_{s\ell}^{gh}(u)}{[\hat{\gamma}_{ss}^{gg}(0)\cdot\hat{\gamma}_{\ell\ell}^{gg}(0)\cdot\hat{\gamma}_{ss}^{hh}(0)\cdot\hat{\gamma}_{\ell\ell}^{hh}(0)]^{1/4}} \tag{7-43}$$

$g,h=1,2,\ldots, \; ; \; s,\ell,u=0,1,2,\ldots$

7.4.2 Expectation and Variance of the Multivariate Autocorrelation Function

The statistical property of the sample multivariate space-time autocorrelation function for the white noise process is needed for the diagnostic checking stage. In this section the expectation value and variance $\hat{\gamma}^{gh}_{s\ell}(u)$ and $\hat{\rho}^{gh}_{s\ell}(u)$ for the white noise process is derived.

The white noise process is equivalent to a MULSTAR$(\zeta,\underset{\sim}{0},\underset{\sim}{0})$ model with the assumption that $\underset{\sim}{Z}(t)$ are white noise. Thus,

$$\underset{\sim}{Z}^h(t) = \underset{\sim}{\varepsilon}^h(t) \qquad h=1,2,\ldots,\zeta \tag{7-44}$$

where

$$E(\underset{\sim}{\varepsilon}^h(t)\underset{\sim}{\varepsilon}^g(t+k)) = \begin{cases} \sigma_h^2 I & h=g \text{ and } k=0 \\ 0 & \text{otherwise} \end{cases}$$

Applying the result from matrix algebra $T_r(AB) = T_r(BA)$, we have

$$E[(w^{(s)}\underset{\sim}{Z}^g(t))'w^{(\ell)}\underset{\sim}{Z}^h(t+u)] = \begin{cases} \sigma_h^2 \ \mathrm{Tr}(w^{(s)'}w^{(\ell)}, & u=0,\ h=g \\ 0 & ,\ \text{otherwise} \end{cases} \tag{7-45}$$

Also for $h\neq g$, applying the fact that $\underset{\sim}{\varepsilon}^h(t)$ and $\underset{\sim}{\varepsilon}^g(t)$ are statistically independent, we have

$$\begin{aligned} &E[((w^{(s)}\underset{\sim}{Z}^g(t)'w^{(\ell)}\underset{\sim}{Z}^h(t+u))^2] \\ &= E[(w^{(s)}\underset{\sim}{Z}^g(t))'w^{(\ell)}\underset{\sim}{Z}^h(t+u)\cdot(w^{(\ell)}\underset{\sim}{Z}^h(t+u))'w^{(s)}\underset{\sim}{Z}^g(t)] \\ &= E\{\mathrm{Tr}[w^{(s)'}w^{(\ell)}\underset{\sim}{Z}^h(t+u)\underset{\sim}{Z}^h(t+u)'w^{(\ell)'}w^{(s)}\underset{\sim}{Z}^g(t)\underset{\sim}{Z}^g(t)']\} \\ &= \mathrm{Tr}\{w^{(s)'}w^{(\ell)}E[\underset{\sim}{Z}^h(t+u)\underset{\sim}{Z}^h(t+u)']w^{(\ell)'}w^{(s)}E[\underset{\sim}{Z}^g(t)\underset{\sim}{Z}^g(t)']\} \\ &= \sigma_h^2\sigma_g^2 \ \mathrm{Tr}\{w^{(s)'}w^{(\ell)}w^{(\ell)'}w^{(s)}\} \\ &= \sigma_h^2\sigma_g^2 \ \mathrm{Tr}\{w^{(\ell)}w^{(\ell)'}w^{(s)}w^{(s)'}\} \end{aligned} \tag{7-46}$$

Using Equations (7-45) and (7-46), $E(\hat{\gamma}^{gh}_{s\ell}(u))$ and $\mathrm{Var}(\gamma^{gh}_{s\ell}(u))$ for $g\neq h$ are derived as follows,

$$E(\hat{\gamma}^{gh}_{s\ell}(u)) = \begin{cases} \dfrac{\sigma_h^2\mathrm{Tr}(w^{(s)'}w^{(\ell)})}{LN} & u=0,\ g=h \\ 0 & \text{otherwise} \end{cases} \tag{7-47}$$

$$\begin{aligned} \mathrm{Var}(\hat{\gamma}^{gh}_{s\ell}(u)) &= E[(\hat{\gamma}^{gh}_{s\ell}(u,T))^2] \\ &= \frac{E[\sum_{t_1=1}^{T-u}(w^{(s)}\underset{\sim}{Z}^g(t_1))'w^{(\ell)}\underset{\sim}{Z}^h(t_1+u))]\cdot[\sum_{t_2=1}^{T-u}(w^{(\ell)}\underset{\sim}{Z}^h(t_2+u)'w^{(s)}\underset{\sim}{Z}^g(t_2))]}{(LN\cdot(T-u))^2} \\ &= (LN\cdot(T-u))^{-2}E[\sum_{t_1=1}^{T-u}(w^{(s)}\underset{\sim}{Z}^g(t_1))'w^{(\ell)}\underset{\sim}{Z}^h(t_1+u)(w^{(\ell)}\underset{\sim}{Z}^h(t_2+w))'w^{(s)}\underset{\sim}{Z}^g(t_2)] \\ &\quad + \sum_{t_1=1}^{T-u}\sum_{t_2\neq 1}^{T-u}(w^{(s)}\underset{\sim}{Z}^g(t_1))'w^{(\ell)}\underset{\sim}{Z}^h(t_1+u)(w^{(\ell)}\underset{\sim}{Z}^h(t_2+u))'w^{(s)}\underset{\sim}{Z}^g(t_2)] \\ &= (LN\cdot(T-u))^{-2}\ [\sum_{t_1=1}^{T-u}\sigma_g^2\sigma_h^2\cdot\mathrm{Tr}(w^{(\ell)}w^{(\ell)'}w^{(s)}w^{(s)'}) + 0] \end{aligned}$$

So

$$\mathrm{Var}(\hat{\gamma}^{gh}_{s\ell}(u)) = \frac{\sigma^2_g\sigma^2_h \cdot \mathrm{Tr}(w^{(\ell)}w^{(\ell)'}w^{(s)}w^{(s)'})}{LN^2 \cdot (T-u)} \tag{7-48}$$

where

$g \neq h$

To derive the variance of the sample multivariate space-time autocorrelation $\mathrm{Var}(\hat{\rho}^{gh}_{s\ell}(u))$, we follow the customary procedure to get the approximate $\mathrm{Var}(\hat{\rho}^{gh}_{s\ell}(u))$ as expressed in Equation (7-49).

$$\begin{aligned}
\mathrm{Var}(\hat{\rho}^{gh}_{s\ell}(u)) = {} & \frac{\mathrm{Var}(\gamma^{gh}_{s\ell}(u))}{(\gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0))^{1/2}} - \frac{1}{2}(\gamma^{gh}_{s\ell}(u)) \cdot \\
& (\gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0))^{-3/2} \cdot [\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0) \cdot \mathrm{Cov}(\gamma^{gg}_{ss}(0), \gamma^{gh}_{s\ell}(u)) \\
& + \gamma^{gg}_{ss}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0) \cdot \mathrm{Cov}(\gamma^{gg}_{\ell\ell}(0), \gamma^{gh}_{s\ell}(u)) \\
& + \gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{\ell\ell}(0) \cdot \mathrm{Cov}(\gamma^{hh}_{ss}(0), \gamma^{gh}_{s\ell}(u)) + \gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0) \cdot \\
& \quad \mathrm{Cov}(\gamma^{hh}_{\ell\ell}(0)\gamma^{gh}_{s\ell}(u))] + \frac{1}{16}(\gamma^{gh}_{s\ell}(u))^2 \cdot (\gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0))^{-5/2} \\
& \quad [\mathrm{Var}(\gamma^{gg}_{ss}(0))(\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0))^2 + \mathrm{Var}(\gamma^{gg}_{\ell\ell}(0) \cdot (\gamma^{gg}_{ss}(0)\gamma^{hh}_{ss}(0) \cdot \\
& \quad \gamma^{hh}_{\ell\ell}(0))^2 + \mathrm{Var}(\gamma^{hh}_{ss}(0))(\gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{\ell\ell}(0)^2
\end{aligned}$$

$$\begin{aligned}
& + \mathrm{Var}(\gamma^{hh}_{\ell\ell}(0))(\gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0))^2 \\
& + 2\gamma^{gg}_{\ell\ell}(0)\gamma^{gg}_{ss}(0)\gamma^{hh}_{ss}(0)^2\gamma^{hh}_{\ell\ell}(0)^2\mathrm{Cov}(\gamma^{gg}_{\ell\ell}(0), \gamma^{gg}_{ss}(0)) \\
& + 2\gamma^{gg}_{\ell\ell}(0)^2\gamma^{gg}_{ss}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0)^2\mathrm{Cov}(\gamma^{gg}_{ss}(0), \gamma^{hh}_{ss}(0)) \\
& + 2\gamma^{gg}_{\ell\ell}(0)^2\gamma^{gg}_{ss}(0)\gamma^{hh}_{ss}(0)^2\gamma^{hh}_{\ell\ell}(0)\mathrm{Cov}(\gamma^{gg}_{ss}(0), \gamma^{hh}_{\ell\ell}(0)) \\
& + 2\gamma^{gg}_{ss}(0)^2\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0)^2\mathrm{Cov}(\gamma^{gg}_{\ell\ell}(0), \gamma^{hh}_{ss}(0)) \\
& + 2\gamma^{gg}_{ss}(0)^2\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)^2\gamma^{hh}_{\ell\ell}(0)\mathrm{Cov}(\gamma^{gg}_{\ell\ell}(0), \gamma^{hh}_{\ell\ell}(0)) \\
& + 2\gamma^{gg}_{ss}(0)^2\gamma^{gg}_{\ell\ell}(0)^2\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0)\mathrm{Cov}(\gamma^{hh}_{ss}(0), \gamma^{hh}_{\ell\ell}(0)
\end{aligned} \tag{7-49}$$

The expression in Equation (7-49) is tedious to apply, but by using the result in Equation (7-47), i.e., $E(\hat{\gamma}^{gh}_{s\ell}(u)) = 0$, we have,

$$\mathrm{Var}(\hat{\rho}^{gh}_{s\ell}(u)) = \frac{\mathrm{Var}(\gamma^{gh}_{s\ell}(u))}{(\gamma^{gg}_{ss}(0)\gamma^{gg}_{\ell\ell}(0)\gamma^{hh}_{ss}(0)\gamma^{hh}_{\ell\ell}(0))^{1/2}} \tag{7-50}$$

By substituting Equations (7-47) and (7-48) into Equation (7-50), we have the approximate variance of $\hat{\rho}^{gh}_{s}(u)$ as,

$$\mathrm{Var}(\hat{\rho}^{gh}_{s\ell}(u)) \simeq \frac{\mathrm{Tr}(w^{(\ell)}w^{(\ell)'}w^{(s)}w^{(s)'})/(T-u)}{\mathrm{Tr}(w^{(\ell)}w^{(\ell)'}) \cdot \mathrm{Tr}(w^{(s)}w^{(s)'})} \tag{7-51}$$

Setting s=ℓ=0 in the above equation, we have

$$\mathrm{Var}(\hat{\rho}^{gh}_{00}(u)) = \frac{1}{LN\cdot(T-u)} \tag{7-52}$$

Setting ℓ=0 in Equation (7-51), we have

$$\mathrm{Var}(\hat{\rho}^{gh}_{s0}(u)) = \frac{\mathrm{Tr}(w^{(s)}w^{(s)'})}{LN\cdot(T-u)\cdot\mathrm{Tr}(w^{(s)}w^{(s)'})}$$

$$= \frac{1}{LN\cdot(T-u)} \tag{7-53}$$

Therefore when g≠h, we have the variance of $\hat{\rho}^{gh}_{s0}(u)$ to be approximately $(LN\cdot(T-u))^{-1}$.

By setting g=h, the multivariate definition of $\hat{\gamma}^{gh}_{s\ell}(u)$, $\hat{\rho}^{gh}_{s\ell}(u)$ reduces to the univariate definition (see Pfeifer and Deutsch [1981]). Table 7-1 summarizes these results and those developed here.

7.4.3 Theoretical Properties of the Autocorrelation Function for MULSTARMA Processes

The theoretical characteristics of the multivariate autocorrelation function for the MULSTAR, MULSTARMA and MULSTMA models are a primary tool in the identification of a potential candidate model in model building. Since the MULSTAR, MULSTARMA and MULSTMA models are specified models of the General Multivariate STAR, STARMA, STMA model, respectively, in this section we address the properties of the General Multivariate models and apply the results directly to the MULSTAR, MULSTARMA and MULSTMA models.

Table 7-1. Variance of Sample Multivariate Space-Time Autocovariance and Autocorrelation Function of the White Noise Process

Estimator	E{Estimator}	Variance{Estimator}
		h≠g
$\hat{\gamma}^{gh}_{00}(u)$	0	$\frac{\sigma_h^2\sigma_h^2}{LN\cdot(T-u)}$
$\hat{\gamma}^{gh}_{s0}(u)$	0	$\frac{\sigma_g^2\sigma_h^2\cdot\mathrm{Tr}(w^{(s)}w^{(s)'})}{LN^2\cdot(t-u)}$
$\hat{\gamma}^{gh}_{s\ell}(u)$	0	$\frac{\sigma_g^2\sigma_h^2\cdot\mathrm{Tr}(w^{(\ell)}w^{(\ell)'}w^{(s)}w^{(s)'})}{LN^2\cdot(t-u)}$
$\hat{\rho}^{gh}_{00}(u)$	0	$\frac{1}{LN\cdot(t-u)}$
$\hat{\rho}^{gh}_{s0}(u)$	0	$\frac{1}{LN\cdot(t-u)}$
$\hat{\rho}^{gh}_{s\ell}(u)$	0	$\frac{\mathrm{Tr}(w^{(\ell)}w^{(\ell)'}w^{(s)}w^{(s)'})}{(T-u)\cdot\mathrm{Tr}(w^{(\ell)}w^{(\ell)'}\cdot\mathrm{Tr}(w^{(s)}w^{(s)'})}$
		h=g
$\hat{\gamma}^{gh}_{00}(0)$	σ_h^2	$\frac{2\sigma_h^4}{LN\cdot T}$
$\hat{\gamma}^{gh}_{s0}(0)$	0	$\frac{\sigma_h^4\{\mathrm{Tr}(w^{(s)}w^{(s)'})+\mathrm{Tr}(w^{(s)}\ ^{(s)'})\}}{LN^2\cdot T}$
$\hat{\gamma}^{gh}_{ss}(0)$	$\frac{\sigma_h^2}{LN}\mathrm{Tr}(w^{(s)}w^{(s)'})$	$\frac{\sigma_h^4\mathrm{Tr}(w^{(s)}w^{(s)'}w^{(s)}w^{(s)'})}{LN^2\cdot T}$
$\hat{\gamma}^{gh}_{s0}(u)$, u>0	0	$\frac{\sigma_h^4\mathrm{Tr}(w^{(s)}w^{(s)'})}{LN^2\cdot(T-u)}$
$\hat{\rho}^{gh}_{s0}(u)$	0	$\frac{1}{LN(T-u)}$

With Equation (7-34),

$$\gamma_{s\ell}^{gh}(u) = \frac{w^{(s)'}w^{(\ell)} \cdot \Gamma^{gh}(u)}{LN},$$

we need to compute the covariance matrix $\Gamma^{gh}(u)$ of the observations $\underset{\sim}{Z}^{g}(t)$, $\underset{\sim}{Z}^{h}(t+u)$ before computing the multivariate space-time autocovariance function $\gamma_{s\ell}^{gh}(u)$ and the multivariate space-time autocorrelation function $\rho_{s\ell}^{gh}(u)$. The covariance matrix of $\underset{\sim}{Z}^{i}(t)$ and $\underset{\sim}{Z}^{j}(t+u)$ for the General Multivariate models in the Λ-weight representation is derived below.

$$\begin{aligned}
\Gamma^{ij}(u) &= E(\underset{\sim}{Z}^{j}(t+u)\underset{\sim}{Z}^{i}(t)') \\
&= E[(\sum_{g_1=1}^{\zeta}\sum_{k_1=1}^{\infty} \Lambda^{jg_1}(k_1)\underset{\sim}{\varepsilon}^{g_1}(t+u-k_1) + \underset{\sim}{\varepsilon}^{j}(t+u)) \cdot \\
&\quad (\sum_{g_2=1}^{\zeta}\sum_{k_2=1}^{\infty} \Lambda^{ig_2}(k_2)\underset{\sim}{\varepsilon}^{g_2}(t-k_2) + \underset{\sim}{\varepsilon}^{i}(t))']
\end{aligned}$$

$$i,j=1,2,3,\ldots,\zeta;\ u=0,1,2,\ldots,\infty$$

$$u=0 \quad \Gamma^{ij}(u) = G^{ji} + \sum_{g_1=1}^{\zeta}\sum_{g_2=1}^{\zeta}\sum_{k=1}^{\infty} \Lambda^{jg_1}(k)G^{g_1g_2}\Lambda^{ig_2'}(k)$$

$$u=1 \quad \Gamma^{ij}(u) = \sum_{g_1=1}^{\zeta} \Lambda^{jg_1}(1)G^{g_1i} + \sum_{g_1=1}^{\zeta}\sum_{g_2=1}^{\zeta}\sum_{k=1}^{\infty} \Lambda^{jg_1}(k+1)G^{g_1g_2}\Lambda^{ig_2'}(k)$$

$$u=2 \quad \Gamma^{ij}(u) = \sum_{g_1=1}^{\zeta} \Lambda^{jg_1}(2)G^{g_2i} + \sum_{g_1=1}^{\zeta}\sum_{g_2=1}^{\zeta}\sum_{k=1}^{\infty} \Lambda^{jg_1}(k+2)G^{g_1g_2}\Lambda^{ig_2'}(k)$$

$$\vdots \qquad \vdots$$

$$u>0 \quad \Gamma^{ij}(u) = \sum_{g_1=1}^{\zeta} \Lambda^{jg_1}(u)G^{g_1i} + \sum_{g_1=1}^{\zeta}\sum_{g_2=1}^{\zeta}\sum_{k=1}^{\infty} \Lambda^{jg_1}(k+u)G^{g_1g_2}\Lambda^{ig_2'}(k)$$

These above results can be summarized as following,

$$\Gamma^{ij}(u) = \begin{cases} G^{ji} + \sum_{g_1=1}^{\zeta}\sum_{g_2=1}^{\zeta}\sum_{k=1}^{\infty} \Lambda^{jg_1}(k)G^{g_1g_2}\Lambda^{ig_2'}(k), & u=0 \\ \sum_{g_1=1}^{\zeta} \Lambda^{jg_1}(u)G^{g_1i} + \sum_{g_1=1}^{\zeta}\sum_{g_2=1}^{\zeta}\sum_{k=1}^{\infty} \Lambda^{jg_1}(k+u)G^{g_1g_2}\Lambda^{ig_2'}(k), & u>0 \end{cases} \tag{7-54}$$

$$i,j=1,2,\ldots,\zeta.$$

Analogous to the univariate MA models and STMA models, we expect that $\Gamma^{ij}(u)$ will be a null covariance matrix for the General Multivariate STMA model when u is large enough. By substituting Equation (7-28) into Equation (7-54). We get the covariance matrix $\Gamma^{ij}(u)$ of the General Multivariate STMA model,

$$
\Gamma^{ij}(u) = \begin{cases} G^{ji} + \sum_{g_1=1}^{\zeta} \sum_{g_2=1}^{\zeta} \sum_{k=1}^{q^{j*}} A^{jg_1}(k) G^{g_1 g_2} A^{ig_2}(k)', & u=0 \\ \sum_{g_1=1}^{\zeta} - A^{jg_1}(u) + \sum_{g_1=1}^{\zeta} \sum_{g_2=1}^{\zeta} \sum_{k=1}^{q^{j*}-u} A^{jg_1}(k+u) G^{g_1 g_2} A^{ig_2}(k)', & 0<u\leq q^{j*} \\ 0 & u>q^{j*} \end{cases} \tag{7-55}
$$

$$i,j=1,2,\ldots,\zeta$$

where $q^{j*} = \max\{q^{jg} | g=1,2,\ldots,\zeta\}$. To compute the autocovariance function $\gamma_{s\ell}^{ij}(u)$, we substitute Equation (7-55) into Equation (7-34) to get the conclusion that the autocovariance function $\gamma_{s\ell}^{ij}(u)$ cuts off after $u>q^{j*}$, where $q^{j*} = \max\{q^{jg} | g=1,2,\ldots,\zeta\}$.

For the General Multivariate STAR model since the elements of $\Lambda^{hg}(k)$ in the Λ-weight representation exponentially decay to zero when $k > p^{hg}$, will cause the elements of $\Gamma^{ij}(u)$ to also exponentially decay to zero when $u > \max\{p^{jg} | g=1,2,\ldots,\zeta\}$, which in turn causes the autocovariance function $\gamma_{s\ell}^{ij}(u)$ to exponentially decay. Similar consideration gives similar conclusions for the General Multivariate STARMA model. That is, the autocovariance function $\gamma_{s\ell}^{ij}(u)$ will exponentially decay to zero when $u > \max\{p^{jg}, q^{jg} | g=1,2,\ldots,\zeta\}$. Since the MULSTAR, MULSTMA, MULSTARMA models are specific models of the General Multivariate models, the results that we have gotten hold for the MULSTAR, MULSTMA and MULSTARMA models.

7.5 Multivariate Partial Space-Time Autocorrelation Considerations

Another important statistical property that is very helpful in identifying the candidate model is the partial autocorrelation function. The multivariate space-time partial autocorrelation is derived in Section 7.5.1. This analysis is followed by the computation considerations section. Here efficient approaches for computing the multivariate space-time partial autocorrelation functions are proposed which eliminate the need of computing the inverse of large dimensional matrices. This affords computational efficiency since the effort needed to obtain the inverse of a full rank matrix increases exponentially when the dimension of the matrices increases linearly.

7.5.1 The Multivariate Partial Space-Time Autocorrelation Function

Pre-multiplying both sides of the MULSTAR$(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda})$ model,

$$\underset{\sim}{Z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{Z}^g(t-k) + \underset{\sim}{\varepsilon}^h(t)$$

$$h=1,2,\ldots,\zeta.$$

by $[w^{(s)} Z^{\nu}(t-u)]'$ gives

$$[w^{(s)} \underset{\sim}{Z}^{\nu}(t-u)]' \underset{\sim}{Z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} [w^{(s)} \underset{\sim}{Z}^{\nu}(t-u))]' \phi_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{Z}_{\ell}^g(t-k)$$

$$+ [w^{(s)} \underset{\sim}{Z}^{\nu}(t-u)]' \underset{\sim}{\varepsilon}^h(t)$$

Taking the expectation value and dividing both sides by LN gives

$$\gamma_{s0}^{\nu h}(u) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} \gamma_{s\ell}^{\nu g}(u-k) \tag{7-56}$$

Letting $\nu=1,2,\ldots,\zeta$, $h=1,2,\ldots,\zeta$, $u=1,2,\ldots,p^{h\nu}$ and $s=0,1,\ldots,\lambda_u^{h\nu}$, results in the partial autocovariance function,

$$\gamma_{s0}^{\nu h}(u) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} \gamma_{s\ell}^{\nu g}(u-k) \tag{7-57}$$

$$\nu,h=1,2,\ldots,\zeta$$

$$u=1,2,\ldots,p^{\nu h}$$

$$s=0,1,2,\ldots,\lambda_u^{\nu h}$$

This equation consists of $\sum_{\nu=1}^{\zeta} \sum_{h=1}^{\zeta} \sum_{u=1}^{p^{\nu h}} (\lambda_u^{\nu h}+1)$ linearly independent equations from which we can solve for the $\sum_{g=1}^{\zeta} \sum_{h=1}^{\zeta} \sum_{k=1}^{p^{hg}} (\lambda_k^{hg}+1)$ autoregressive parameters.

This system of equations is the multivariate space-time analog of the Yule-Walker equations for univariate time series. As we will state in more detail later, although the system of equations defined by the multivariate autocorrelation function does not have a symmetric coefficient matrix for all the $\underset{\sim}{\phi}$ parameters, for every specified h-

category, the coefficient matrix for $\underset{\sim}{\phi}^h$ parameters is symmetric. This is similar to the Yule-Walker equations for the univariate time series.

Since we have $\gamma_{s\ell}^{gh}(-u) = \gamma_{\ell s}^{hg}(u)$ in Equation (7-36), so in Equation (7-57) we can replace $\gamma_{s\ell}^{\nu g}(u-k)$ with $\gamma_{\ell s}^{g\nu}(k-u)$ when $u-k<0$.

Based on the solution set of the system of Equation (7-57), we will define the Multivariate Space-Time Partial Autocorrelation Function $\underset{\sim}{\phi}(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda})$ as follows,

$$\underset{\sim}{\phi}(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}) = (\phi_{k\lambda_k^{\nu h}}^{\nu h},\ \nu=1,2,\ldots,\zeta,\ h=1,2,\ldots,\zeta,\ k=1,2,\ldots,p^{\nu h}),$$

$\underset{\sim}{\phi}(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda})$ is a $\sum_{\nu} \sum_{h} (p^{\nu h})$ column vector.

Theoretically, we can construct the sequence of Equation (7-57) and compute the associated Multivariate Space-Time Partial Autocorrelation Function $\underset{\sim}{\phi}(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda})$. Assume we have the MULSTAR$(\zeta,\underset{\sim}{p}_T,\underset{\sim}{\lambda}_T)$ as the true model, then the sequence of Multivariate Space-Time Partial Autocorrelation Function vectors will have the following property,

$$\underset{\sim}{\phi}(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}) \begin{cases} = \underset{\sim}{0} & \underset{\sim}{p} > \underset{\sim}{p}_T \text{ and } \underset{\sim}{\lambda} > \underset{\sim}{\lambda}_T \\ \neq \underset{\sim}{0} & \text{otherwise} \end{cases} \tag{7-58}$$

Thus the Multivariate Space-Time Partial Autocorrelation Function $\underset{\sim}{\phi}(\zeta,\underset{\sim}{p},\underset{\sim}{\lambda})$ cuts off after $\underset{\sim}{p}_T$ and $\underset{\sim}{\lambda}_T$ when the true model is MULSTAR$(\zeta,\underset{\sim}{p}_T,\underset{\sim}{\lambda}_T)$. Any invertible MULSTMA$(\zeta,\underset{\sim}{q},\underset{\sim}{m})$ and MULSTARMA$(\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m})$ model can be expressed as infinite MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$) model with the coefficients decaying exponentially to zero after certain tim lag and spatial lag. Therefore, the Multivariate Space-Time Autocorrelation Function vector will not have the cut-off property of the MULSTAR model but instead it will be tail-off.

7.5.2 Computation Considerations

The multivariate partial autocorrelation function, Equation (7-57), consists of ζ independent equation sets. Thus, it can be written as,

$$\left.\begin{array}{ll}
h=1,\ \nu=1,\ u=1,\ s=0 & \gamma^{11}_{00}(1) = \sum\limits_{g=1}^{\zeta}\sum\limits_{k=1}^{p^{1g}}\sum\limits_{\ell=0}^{\lambda^{1g}_k}\phi^{1g}_{k\ell}\gamma^{1g}_{0\ell}(1-k) \\
\vdots\quad\vdots\quad\vdots\quad\vdots & \\
h=1,\ \nu=1,\ u=1,\ s=\lambda^{11}_1 & \gamma^{11}_{\lambda^{11}_1 0}(1) = \sum\limits_{g=1}^{\zeta}\sum\limits_{k=1}^{p^{1g}}\sum\limits_{\ell=0}^{\lambda^{1g}_k}\phi^{1g}_{k\ell}\gamma^{1g}_{\lambda^{11}_1\ell}(1-k) \\
\vdots\quad\vdots\quad\vdots\quad\vdots & \\
h=1,\ \nu=\zeta,\ u=1,\ s=0 & \gamma^{\zeta 1}_{00}(1) = \sum\limits_{g=1}^{\zeta}\sum\limits_{k=1}^{p^{1g}}\sum\limits_{\ell=0}^{\lambda^{1g}_k}\phi^{1g}_{k\ell}\gamma^{\zeta g}_{0\ell}(1-k) \\
\vdots\quad\vdots\quad\vdots\quad\vdots & \\
h=1,\ \nu=\zeta,\ u=1,\ s=\lambda^{1\zeta}_1 & \gamma^{\zeta 1}_{\lambda^{1\zeta}_1 0}(1) = \sum\limits_{g=1}^{\zeta}\sum\limits_{k=1}^{p^{1g}}\sum\limits_{\ell=0}^{\lambda^{1g}_k}\phi^{1g}_{k\ell}\gamma^{\zeta g}_{\lambda^{1\zeta}_1\ell}(1-k) \\
\vdots\quad\vdots\quad\vdots\quad\vdots & \\
h=1,\ \nu=\nu^0,\ u=p^0,\ s=\lambda^0 & \gamma^{\nu_0 1}_{\lambda^0 0}(p^0) = \sum\limits_{g=1}^{\zeta}\sum\limits_{k=1}^{p^{1g}}\sum\limits_{\ell=0}^{\lambda^{1g}_k}\phi^{1g}_{k\ell}\gamma^{\nu_0 g}_{\lambda^0\ell}(1-p^0)
\end{array}\right\} h=1 \qquad (7\text{-}59)$$

where

$$p^0 = p^{1\nu^0}$$

$$\lambda^0 = \lambda^{1\nu^0}_{p1\nu^0}$$

and

$p^{1\nu^0}$ is the maximum time lag for h=1

$\lambda^{1\nu^0}_{p1\nu^0}$ is the maximum spatial lag for h=1 with time lag $p^{1\nu^0}$

ν^0 is the maximum category index number for h=1 with $p^{1\nu^0}$ time lag, $\lambda^{1\nu^0}_{p1\nu^0}$ spatial lag.

This equation also holds for h=2,3,...,ζ, to get the other (ζ-1) equations sets. Each equation set contains $\sum\limits_{g=1}^{\zeta}\sum\limits_{k=1}^{p^{hg}}(1+\lambda^{hg}_k)$ equations to solve for that number parameters. These parameters appear only in that equation set. Therefore the solution of the multivariate partial autocorrelation function can be computed by seqparating the ζ sets and solving within each. This will reduce the task of matrix inverse, which is necessary for computing the solution for the linear equations.

As an example, consider the following example of MULSTAR($2,\underset{\sim}{p},\underset{\sim}{\lambda}$) with $\underset{\sim}{p}$ = (2,1,1,1), λ = (1,0,0,0,1), i.e.,

$$\underset{\sim}{z}^1(t) = +\,\phi^{11}_{10}\underset{\sim}{z}^1(t-1) + \phi^{11}_{11}w^{(1)}\underset{\sim}{z}^1(t-1) + \phi^{12}_{10}\underset{\sim}{z}^{(2)}(t-1)$$

$$+\,\phi^{11}_{20}\underset{\sim}{z}^1(t-2) + \underset{\sim}{\varepsilon}^1(t) \qquad (7\text{-}60)$$

$$\underset{\sim}{z}^2(t) = +\,\phi^{21}_{10}\underset{\sim}{z}^1(t-1) + \phi^{22}_{10}\underset{\sim}{z}^2(t-1) + \phi^{22}_{11}w^{(1)}\underset{\sim}{z}^2(t-1) + \underset{\sim}{\varepsilon}^{(2)}(t) \qquad (7\text{-}61)$$

Premultiplying Equation (7-60) with $\underset{\sim}{z}^{(1)}(t-1)'$, $(w^{(1)}\underset{\sim}{z}^{(1)}(t-1))'$, $\underset{\sim}{z}^{(2)}(t-1)'$ and $\underset{\sim}{z}^{(1)}(t-2)'$, taking the expectation value and dividing both sides by LN we get

$$\begin{aligned}
\gamma^{11}_{00}(1) &= \phi^{11}_{10}\gamma^{11}_{00}(0) + \phi^{11}_{11}\gamma^{11}_{01}(0) + \phi^{12}_{10}\gamma^{12}_{00}(0) + \phi^{11}_{20}\gamma^{11}_{00}(-1)\\
\gamma^{11}_{10}(1) &= \phi^{11}_{10}\gamma^{11}_{10}(0) + \phi^{11}_{11}\gamma^{11}_{11}(0) + \phi^{12}_{10}\gamma^{12}_{10}(0) + \phi^{11}_{20}\gamma^{11}_{10}(-1)\\
\gamma^{21}_{00}(1) &= \phi^{11}_{10}\gamma^{21}_{00}(0) + \phi^{11}_{11}\gamma^{21}_{01}(0) + \phi^{12}_{10}\gamma^{22}_{00}(0) + \phi^{11}_{20}\gamma^{21}_{00}(-1)\\
\gamma^{11}_{00}(2) &= \phi^{11}_{10}\gamma^{11}_{00}(1) + \phi^{11}_{11}\gamma^{11}_{01}(1) + \phi^{12}_{10}\gamma^{12}_{00}(1) + \phi^{11}_{20}\gamma^{21}_{00}(0)
\end{aligned} \qquad (7\text{-}62)$$

Premultiplying Equation (7-61) with $\underset{\sim}{z}^1(t-1)'$, $\underset{\sim}{z}^2(t-1)'$, $(w^{(1)}\underset{\sim}{z}^2(t-1))'$ and taking the expectation value and dividing both sides by LN, we get,

$$\begin{aligned}
\gamma^{12}_{00}(1) &= \phi^{21}_{10}\gamma^{11}_{00}(0) + \phi^{22}_{10}\gamma^{12}_{00}(0) + \phi^{22}_{11}\gamma^{12}_{01}(0)\\
\gamma^{22}_{00}(1) &= \phi^{21}_{10}\gamma^{21}_{00}(0) + \phi^{22}_{10}\gamma^{22}_{00}(0) + \phi^{22}_{11}\gamma^{22}_{01}(0)\\
\gamma^{22}_{10}(1) &= \phi^{21}_{10}\gamma^{21}_{10}(0) + \phi^{22}_{10}\gamma^{22}_{10}(0) + \phi^{22}_{11}\gamma^{22}_{11}(0)
\end{aligned} \qquad (7\text{-}63)$$

Equation Set (7-62) contains 4 equations and will be used to solve for ϕ^{11}_{10}, ϕ^{11}_{11}, ϕ^{12}_{10} and ϕ^{11}_{20} that are not contained in Equation Set (7-63). Also Equation Set (7-63) contains 3 equations and will be used to solve for ϕ^{21}_{10}, ϕ^{22}_{10} and ϕ^{22}_{10} that similarly are not contained in Equation (7-62). We could solve for all the seven ϕ's simultaneously instead of solving for the four ϕ's in Equation (7-62) and the three ϕ's in Equation (7-63) separately. For computation efficiency however we do prefer to solve them separately than simultaneously.

By applying the property $\gamma^{gh}_{s\ell}(-u) = \gamma^{hg}_{\ell s}(u)$ of Equation (7-36), we can see that the coefficient matrix of Equation Sets (7-62) and (7-63) are symmetric. To illustrate that this is not a special case for this specified model, let us add an arbitrary term $\phi^{1h}_{k\ell}w^{(\ell)}\underset{\sim}{z}^h(t-k)$ to model (7-60) with $k,\ell\varepsilon\{0,1,2,\ldots,\}$, $h\varepsilon\{1,2\}$ and derive the equation set that is similar to Equation Set (7-62).

$$
\begin{array}{lllll|l}
\gamma_{00}^{11}(1) = \phi_{10}^{11}\gamma_{00}^{11}(0) & + \phi_{10}^{11}\gamma_{01}^{11}(0) & + \phi_{10}^{11}\gamma_{00}^{12}(0) & + \phi_{20}^{11}\gamma_{00}^{11}(-1) & + & \phi_{k\ell}^{1h}\gamma_{0\ell}^{1h}(1-k) \\
\gamma_{10}^{11}(1) = \phi_{10}^{11}\gamma_{10}^{11}(0) & + \phi_{11}^{11}\gamma_{11}^{11}(0) & + \phi_{10}^{12}\gamma_{10}^{12}(0) & + \phi_{20}^{11}\gamma_{10}^{11}(-1) & + & \phi_{k\ell}^{1h}\gamma_{1\ell}^{1h}(1-k) \\
\gamma_{10}^{21}(1) = \phi_{10}^{11}\gamma_{00}^{21}(0) & + \phi_{11}^{11}\gamma_{01}^{21}(0) & + \phi_{10}^{12}\gamma_{00}^{22}(0) & + \phi_{20}^{11}\gamma_{00}^{21}(-1) & + & \phi_{k\ell}^{2h}\gamma_{0\ell}^{2h}(1-k) \\
\gamma_{00}^{11}(2) = \phi_{10}^{11}\gamma_{00}^{11}(1) & + \phi_{11}^{11}\gamma_{01}^{11}(1) & + \phi_{10}^{12}\gamma_{00}^{12}(1) & + \phi_{20}^{11}\gamma_{00}^{11}(0) & + & \phi_{k\ell}^{1h}\gamma_{0\ell}^{1h}(2-k) \\
\hdashline
\gamma_{\ell 0}^{h1}(k) = \phi_{10}^{11}\gamma_{\ell 0}^{h1}(k-1) & + \phi_{11}^{11}\gamma_{\ell 1}^{h1}(k-1) & + \phi_{10}^{12}\gamma_{\ell 0}^{h2}(k-1) & + \phi_{20}^{11}\gamma_{\ell 0}^{21}(k-2) & + & \phi_{k\ell}^{hh}\gamma_{0\ell}^{hh}(0)
\end{array}
\qquad (7\text{-}64)
$$

We see that, with the property $\gamma_{s\ell}^{gh}(-u) = \gamma_{\ell s}^{hg}(u)$ in mind, the coefficient matrix of Equation Set (7-64) is symmetric. By mathematical induction, we can show that the coefficient matrix for each of these sets of equations are symmetric.

In computing the solution for a sequence value of $\underset{\sim}{p},\underset{\sim}{\lambda}$ for Equation Set (7-57), we do hope that a strictly recursive method similar to that due to Durbin [1960] for univariate time series partial autocorrelation calculation exists. But unfortunately there is no such method available for the multivariate space-time partial autocorrelation function calculation. However, some improvement over the successive solution for Equation (7-57) can be made by using the following result from linear algebra to reduce the matrix inversion task.

Let $R = \left[\begin{array}{c:c} R_{11} & R12 \\ \hdashline R_{21} & r \end{array}\right]$ be a symmtric matrix, and r a scalar. Then

$$
R^{-1} = \left[\begin{array}{c:c}
R_{11}^{-1} + \dfrac{R_{11}^{-1}R_{12}R_{21}R_{11}^{-1}}{\gamma - R_{21}R_{11}^{-1}R_{12}} & \dfrac{-R_{11}^{-1}R_{12}}{\gamma - R_{21}R_{11}^{-1}R_{12}} \\
\hdashline
\dfrac{-R_{12}R_{11}^{-1}}{\gamma - R_{21}R_{11}^{-1}R_{12}} & \dfrac{1}{\gamma - R_{21}R_{11}^{-1}R_{12}}
\end{array}\right] \qquad (7\text{-}65)
$$

It should be noted that the symmetric matrix R is obtained by adding the row vector (R_{21},γ) and the column vector $(R_{21},\gamma)'$ onto the symmetric matrix R_{11}, of which the inverse is available. In the procedure of multivariate space-time partial autocorrelation function computation, the coefficient matrix, that contains multivariate space-time autocorrelations, is increased by adding one row and one column at a time to compute the next partial, and the inverse of the coefficient matrix of last partial is available. Applying the formula in Equation (7-65), that uses the last inverse which is available, we avoid the task of computing the inverse from the very beginning and save computational effort.

7.6 Identification

Given a set of observations from some process, the identification of a candidate model is the first step in building the MULSTARMA model for this process. The cut-off, tail-off properties of the multivariate space-time autocorrelation function have been verified in Section 7.4.3, and the cut-off, tail-off properties of the multivariate space-time partial autocorrelation function have been examined in Section 7.5.1. In this section, these statistical characteristics of the multivariate space-time autocorrelations and the multivariate space-time partial autocorrelations are applied for identification purposes.

The magnitude of the multivariate space-time autocovariance function depends partially on the magnitude of the variances of the errors, while the multivariate space-time autocorrelation function is scaled to get rid of such dependence on the variances of erros, and the multivariate space-time autocorrelation function still retains the cut-off property for the General Multivariate STMA models, the tail-off property for the General Multivariate STARMA, STAR models. So instead of using the multivariate space-time autocovariance function, we will use the multivariate space-time autocorrelation function to identify the candidate model.

Since the theoretical autocovariance and autocorrelation function cannot be available, the sample multivariate space-time autocovariance function defined in Equation (7-42), and the sample multivariate space-time autocorrelation function in Equation (7-43),

$$\hat{\gamma}^{gh}_{s\ell}(u) = \frac{\sum_{t=1}^{T-u} (w^{(s)}\underset{\sim}{z}^{g}(t))'(w^{(\ell)}\underset{\sim}{z}^{h}(t+u))}{LN\cdot(T-u)}$$

$$\hat{\rho}^{gh}_{s\ell}(u) = \frac{\hat{\gamma}^{gh}_{s\ell}(u)}{(\hat{\gamma}^{gg}_{ss}(0)\hat{\gamma}^{gg}_{\ell\ell}(0)\hat{\gamma}^{hh}_{ss}(0)\hat{\gamma}^{hh}_{\ell\ell}(0))^{1/4}}$$

$$g,h=1,2,\ldots,\zeta;\ s,\ell,u=0,1,2,\ldots$$

are used. Therefore, in identification the sample autocovariance function is used to construct the equation sets to solve for the sample Multivariate Space-Time Partial Autocorrelation Function vector.

For convenience, the theoretical properties of the multivariate space-time autocorrelations and the multivariate space-time partial autocorrelations for the statistical properties of the Multivariate STAR, Multivariate STMA, Multivariate STARMA models are summarized in Table 7-2. These cut-off and tail-off properties are used in Pattern matching to the sample estimates in order to choose the potential candidate in the identification stage. It should be noted that for MULSTMA models, we can get the information $q^{j*} = \max\{q^{jg}|g=1,2,\ldots,\zeta\}$; $j=1,2,3,\ldots,\zeta$, from the autocovariance function. But for the individual q^{jg}, we cannot tell anything until estimation is done and the parameter significance test is performed.

In univariate STARMA model building, assuming that up to λ_0 th order neighbors are considered to be potentially significant, (λ_0+1) streams of the sample space-time autocorrelation are computed for candidate model identification. For practical purposes, parsemoneous models are sought so λ_0 should not be too large. Therefor, the examination of (λ_0+1) streams of space-time autocorrelation functions to find the pattern for identification should not be too cumbersome a task. However, in building a MULSTARMA process that contains ζ observation categories, ζ^2 such (λ_0+1) streams of multivariate space-time autocorrelations are computed and examined.

Table 7.2. The Statistical Properties of the Multivariate Space-Time Autocorrelation Functions and the Multivariate Space-Time Partial Autocorrelation Functions

Model Form	Multivariate Space-Time Autocorrelation Function	Multivariate Space-Time Partial Autocorrelation Function
MULSTAR(ζ,$\underset{\sim}{p}$,$\underset{\sim}{\lambda}$)	Tail-off	Cuts off after p time lag and $\underset{\sim}{\lambda}$ spatial lag
MULSTMA(ζ,$\underset{\sim}{q}$,$\underset{\sim}{\lambda}$)	Cuts off after q^{j*} time lag	Tail-off
MULSTARMA(ζ,$\underset{\sim}{p}$,$\underset{\sim}{q}$,$\underset{\sim}{\lambda}$,$\underset{\sim}{m}$)	Tail-off	Tail-off

Examining these ζ^2 (λ_0+1) streams sample multivariate space-time autocorrelation functions in numerical values does become a cumbersome task and needs to be simplified even when is only of moderate size.

To simplify this process, a three step procedure should be followed. The first step, the encoding step, the sample multivariate space-time autocorrelation functions are encoded into three coded classes; the significantly positive, the significantly negative and the insignificant. The second step, the pattern matching step, the patterns of these coded sample multivariate space-time autocorrelation functions are matched to the theoretical patterns of know processes. The third step, the decision step we come back to examine the numerical values before making the final judgement on the viability of the candidate model.

An encoding approach has been proposed by Box and Tiao [1981]. A significantly positive is denoted by a "+" symbol, a significantly negative by a "-" and insignificant by a ".". To encode the computed multivariate space-time autocorrelations, these autocorrelations are standardized by the estimated standard deviations that are computed according to the formulae listed in Table 7-1. Since the standardized multivariate space-time autocorrelations are N(0,1) distributed for the white noise process for an = 0.05 level, the multivariate space-time autocorrelation function is encoded as "+" if its standardized value is greater than 1.96, "-" if less than -1.96 and "." if it is in between. It should be noted that these encoding values simplify the information for model identirication but loses all the necessary information for the initial estimates of the parameters. However, since the MULSTARMA model contains too many parameters to make the effort for initial estimates practical, and the numerical values of the multivariate space-time autocorrelations are still accessible, this loss is not significant.

To illustrate how the encoded symbols will be used for model identification, we examine the encoded multivariate space-time autocorrelation functions and the encoded multivariate space-time partial autocorrelation functions of a $\zeta=2$ system shown in Figure 7-10. The encoded $\hat{\rho}_s^{gh}(k)$ cuts off while the encoded $\hat{\phi}_s^{gh}(k)$ tails off, so this pattern is from the MULSTMA processes. A further step, since $\hat{\rho}_s^{1h}$ cuts off at k=2, and $\hat{\rho}_s^{2h}$ cuts off at k=1, so we can tell from this pattern that $q^{1*} = \max\{q^{11},q^{12}\} = 2$ and $q^{2*} = \max\{q^{21},q^{22}\} = 1$.

7.7 Estimation

After selecting the candidate model in the identification stage, we need to estimate the parameters of the candidate model. In this section, we will give the procedure that leads to the conditional maximum likelihood estimates for MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$), MULSTMA($\zeta,\underset{\sim}{q},\underset{\sim}{m}$) and MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) models. We will first assume that the variance-covariance matrix of the noise G is known, develop the procedure to estimate the model parameters and then we turn to the G unknown situation where we employ the two-stage estimation procedure. We will restrict ourself to the conditional maximum likelihood estimators and conditional least square estimators because of the computation efficiency of these conditional estimators.

g	h	S	1	2	3	4	5	6	7	8	9	10
1	1	0	-	-	•	•	•	•	•	•	•	•
1	1	1	-	-	•	•	•	•	•	•	•	•
1	1	2	+	•	•	•	•	•	•	•	•	•
1	1	3	+	•	•	•	•	•	•	•	•	•
1	2	0	-	+	•	•	•	•	•	•	•	•
1	2	1	-	•	•	•	•	•	•	•	•	•
1	2	2	+	+	•	•	•	•	•	•	•	•
1	2	3	-	•	•	•	•	•	•	•	•	•
2	1	0	+	•	•	•	•	•	•	•	•	•
2	1	1	-	•	•	•	•	•	•	•	•	•
2	1	2	-	•	•	•	•	•	•	•	•	•
2	1	3	-	•	•	•	•	•	•	•	•	•
2	2	0	+	•	•	•	•	•	•	•	•	•
2	2	1	+	•	•	•	•	•	•	•	•	•
2	2	2	-	•	•	•	•	•	•	•	•	•
2	2	3	+	•	•	•	•	•	•	•	•	•

Figure 7-10(a). The Encoded Theoretical Multivariate Space-Time Autocorrelation Functions of MULSTMA($2,\underset{\sim}{q},\underset{\sim}{m}$) Process with $\underset{\sim}{q} = (q^{11} \le 2,\ q^{12} \le 2,\ q^{21} \le 1,\ q^{22} \le 1)$

g	h	S	1	2	3	4	5	6	7	8	9	10
1	1	0	+	+	-	+	-	+	.	.	.	-
1	1	1	-	+	+	-	+	+	.	.	.	.
1	1	2	-	-	+	-	-	+	.	+	.	.
1	1	3	-	+	.	+	.	-	.	.	.	.
1	2	0	+	-	+	-	+	-	.	.	.	.
1	2	1	-	+	.	+	-	+	-	.	.	.
1	2	2	+	.	-	-	+	.	.	.	.	+
1	2	3	+	-	-	.	.	+	.	.	.	.
2	1	0	+	+	.	+	+	-	.	.	.	.
2	1	1	.	-	+	-	+	-	.	.	.	.
2	1	2	+	+	.	+	-	+	.	-	.	.
2	1	3	.	-	-	+	-	+	.	.	.	.
2	2	0	+	+	-	-	+	.	.	.	.	.
2	2	1	+	-	+	.	-	-	.	.	+	.
2	2	2	-	+	-	-	+	.	.	.	.	.
2	2	3	+	+	+	-	+	+	.	.	-	.

Figure 7-10(b). The Encoded Theoretical Multivariate Space-Time Partial Autocorrelation Functions of MULSTMA($2,\tilde{q},\tilde{m}$) process with $\tilde{q} = (q^{11} \leq 2,\ q^{12} \leq 2,\ q^{21} \leq 1,\ q^{22} \leq 1)$

7.7.1 Conditional Maximum Likelihood Estimation with G Known

In this section, the covariance, G, that has been defined in Section 7.1, is assumed to be known. It is well known that the univariate STARMA models are linear in the autoregressive parameters and nonlinear in the moving average parameters. The MULSTARMA models inherit the same characteristics, i.e., linear in the AR components and non-linear in the MA components. Due to their linearity, we can apply the results of linear model theory to reduce the search efforts in estimating the autoregressive parameters. In the following, the conditional maximum likelihood estimation procedures are developed to estimate the model parameters in the order of MULSTAR, MULSTMA and MULSTARMA models.

7.7.1.1 MULSTAR($\zeta,\tilde{p},\tilde{\lambda}$) Model Estimation. Consider the MULSTAR($\zeta,\tilde{p},\tilde{\lambda}$) model

$$\tilde{z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} \sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)} \tilde{z}^g(t-k) + \tilde{\varepsilon}^h(t) \qquad (7\text{-}66)$$

$$h=1,2,\ldots,\zeta;\ t=1,2,\ldots,T$$

where $\tilde{\varepsilon}^h(t)$ is normally distributed with

$$E(\tilde{\varepsilon}^h(t)(\tilde{\varepsilon}^g(t+k))') = \begin{cases} G^{hg} & k=0 \\ 0 & k\neq 0 \end{cases}$$

Denoting

$$\varepsilon(t)' = (\varepsilon^1(t)', \varepsilon^2(t)', \ldots, \varepsilon^\zeta(t)'); \quad \varepsilon' = (\varepsilon(1)', \ldots, \varepsilon(T)')$$

$$Z(t)' = (Z^1(t)', Z^2(t)', \ldots, Z(T)'); \quad Z' = (Z(1)', Z(2)', \ldots, Z(T)')$$

$$\phi' = (\phi^{1'}, \phi^{2'}, \ldots, \phi^{\zeta'})$$

$$\phi^{i'} = (\phi^{ij}_{k\ell}), \; j=1,2,\ldots,\zeta, \; k=1,2,\ldots,p^{ij}, \; \ell=0,1,2,\ldots,\lambda^{ij}_k$$

as stated previously in Equations (7-2) and (7-8) of Section 7.1.1, we rewrite the MULSTAR model to,

$$Z(t) = X(t)\phi + \varepsilon(t), \quad t=1,2,\ldots,T \tag{7-67}$$

where ε (t) is normally distributed with

$$E(\varepsilon(t)\varepsilon(t+k)') = \begin{cases} G & k=0 \\ 0 & \text{otherwise} \end{cases}$$

The matrix X(t) is a matrix where the columns that correspond to $\phi^{hg}_{k\ell}$ defined as

$$\overset{\ell}{N}(h,g)Z(t-k) \tag{7-68}$$

where $\overset{\ell}{N}(h,g)$ is the $\zeta \cdot LN$ null matrix with the (h,g) block of $(LN \times LN)$ size replaced by $w^{(\ell)}$, and $G = [G^{hg}]$ as defined previously.

As an example, consider the MULSTAR$(2,p,\lambda)$ model with $p = (2,1,1,1)$, $\lambda = (1,0,0,0,0)$.

$$Z^1(t) = \phi^{11}_{10}Z^1(t-1) + \phi^{11}_{11}w^{(1)}Z^1(t-1) + \phi^{12}_{10}Z^2(t-1) + \phi^{11}_{20}Z^1(t-2) + \varepsilon^1(t)$$

$$Z^2(t) = \phi^{21}_{10}Z^1(t-1) + \phi^{22}_{10}Z^2(t-1) + \varepsilon^2(t)$$

$$t=1,2,\ldots,T \tag{7-69}$$

which can be rewritten as

$$Z(t) = \phi^{11}_{10}\left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] Z(t-1) + \phi^{11}_{11}\left[\begin{array}{c|c} w^{(1)} & 0 \\ \hline 0 & 0 \end{array}\right] Z(t-1) + \phi^{12}_{10}\left[\begin{array}{c|c} 0 & I \\ \hline 0 & 0 \end{array}\right] Z(t-1)$$

$$+ \phi^{11}_{20}\left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] Z(t-2) + \phi^{21}_{10}\left[\begin{array}{c|c} 0 & 0 \\ \hline I & 0 \end{array}\right] Z(t-1) + \phi^{22}_{10}\left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array}\right] Z(t-1) + \varepsilon(t)$$

$$= \phi^{11}_{10}X_1(t) + \phi^{11}_{11}X_2(t) + \phi^{12}_{10}X_3(t) + \phi^{11}_{20}X_4(t) + \phi^{21}_{10}X_5(t)$$

$$+ \phi^{22}_{10}X_6(t) + \varepsilon(t)$$

$$= X(t)\phi + \varepsilon(t) \tag{7-70}$$

where the column that corresponds to ϕ^{11}_{11} is $\left[\begin{array}{c|c} w^{(1)} & 0 \\ \hline 0 & 0 \end{array}\right] Z(t-1)$.

We have assumed that $\varepsilon(t)$ is normally distributed, and the joint density distribution function of ε is

$$f(\varepsilon|\phi,G) = (2\pi)^{-T\cdot\zeta\cdot LN/2}|G|^{-T/2}\exp\{-1/2\sum_{t=1}^{T}\varepsilon(t)'G^{-1}\varepsilon(t) \tag{7-71}$$

Since the transformation of the Z's

$$\varepsilon(t) = Z(t) - X(t)\phi \qquad t=1,2,\ldots,T \tag{7-72}$$

has unit Jacobian, we have the joint density distribution function of Z conditional on ϕ, G and $Z^h(t)$, $h = 1,2,\ldots,\zeta$, $1-p^{h^*}\leq t\leq 0$, as,

$$f(Z|\phi,G,Z^h(t);\ h = 1,2,\ldots,\zeta,\ 1-p^{h^*}\leq t\leq 0)$$

$$= (2\pi)^{-T\cdot\zeta\cdot LN/2}|G|^{-T/2}\exp\{-1/2\sum_{t=1}^{T}[Z(t)-X(t)\phi)'G(Z(t)-X(t)\phi)] \tag{7-73}$$

Denoting $Z^{*'} = (Z^h(t);\ h = 1,2,\ldots,\zeta$ and $1-p^{h^*}\leq t\leq 0)$, and taking the logarithm of the joint density, we get the log likelihood to be

$$\ell n\, f(Z|\phi,G,Z^*) = -\frac{T}{2}\ell n|G| - \frac{1}{2}SS(\phi,G,Z^*) \tag{7-74}$$

where

$$SS(\phi,G,Z^*) = \sum_{t=1}^{T}[(Z(t)-X(t)\phi)'G(Z(t)-X(t)\phi)]$$

To maximize the likelihood function is exactly the same as to minimize the sum of squares $SS(\phi,G,Z^*)$ for given G,Z^*. So the conditional maximum likelihood estimator is the same as the least sum square when G is given.

Since G is positive definite and symmetric, $G=L'L$, and

$$\frac{\partial^2 SS}{\partial\phi^2}(\phi,G,Z^*) = X'(t)G\,X(t)$$

$$= LX(t)'(LX(t)) \tag{7-75}$$

where $L'L = G$ is positive definite. So set $\frac{\partial SS}{\partial\phi}(\phi,G,Z^*) = 0$ gives the least square estimator as well as the conditional maximum likelihood estimator

$$\hat{\phi} = [\sum_{t=1}^{T}X'(t)G\,X(t)]^{-1}[\sum_{t=1}^{T}X'(t)G\,Z(t)] \tag{7-76}$$

Equation (7-76) is simple in form, but computationally we prefer the matrix $\sum_{t=1}^{T}X'(t)G\,X(t)$, which will be inversed, to be of smaller dimension if possible. When G has the property

$$G^{hg} = \begin{cases} G^{hh} & \text{when } h=g \\ 0 & \text{otherwise} \end{cases}$$

i.e., $\underset{\sim}{\varepsilon}^h(t)$ and $\underset{\sim}{\varepsilon}^g(t)$ are uncorrelated when $h \neq g$, we can decompose Equation (7-76) and compute $\underset{\sim}{\phi}^i$, $i=1,2,\ldots,\zeta$ separately. This will give better computation efficiency.

Denote $X(t) = (X^1(t),X^2(t),\ldots,X^\zeta(t))$, where $X^i(t)$ is the matrix that contains all the columns that corresponds to $\underset{\sim}{\phi}^i$ in corresponding order. We note that $X^i(t)$ contains null row vectors except those from $((i-1)LN+1)^{th}$ to $(i\cdot LN)^{th}$ rows. When G contains null off-diagonal blocks, we have

$$\sum_{t=1}^{T} X'(t)G\,X(t) = \sum_{t=1}^{T} [(X'(t)G\,X(t))_{gh}] \tag{7-77}$$

where $(X'(t)G\,X(t))_{gh}$ is the (g,h) block of $X'(t)G\,X(t)$ matrix, then we have $(X'(t)G\,X(t))_{ii} = X^{i'}(t)G_{ii}X^i(t)$, and $(X'(t)G\,X(t))_{gh}$.

Let $[\sum_{t=1}^{T} X'(t)G\,X(t)]^{-1} = [(X\,G\,X)^{gh}]$ $g,h=1,2,\ldots,\zeta$ where XGX^{gh} is the (g,h) block of the inverse of $\sum_{t=1}^{T} X'(t)GX(t)$. Then we have

$$(XGX)^{gh} = \begin{cases} [\sum_{t=1}^{T} (X^g(t)'GX^g(t))]^{-1} & \text{when } g=h \\ 0 & \text{otherwise} \end{cases} \tag{7-78}$$

Also the vector $\sum_{t=1}^{T} X'(t)GZ(t)$ is resulted in the form

$$\sum_{t=1}^{T} X'(t)G\underset{\sim}{Z}(t) = \begin{bmatrix} \sum_{t=1}^{T} X^1(t)'G_{11}\underset{\sim}{Z}^1(t) \\ \vdots \\ \sum_{t=1}^{T} X^i(t)'G_{ii}\underset{\sim}{Z}^i(t) \\ \vdots \\ \sum_{t=1}^{T} X^\zeta(t)'G_{\zeta\zeta}\underset{\sim}{Z}^\zeta(t) \end{bmatrix} \tag{7-79}$$

So the resulting estimator $\hat{\underset{\sim}{\phi}}' = (\hat{\underset{\sim}{\phi}}^{1'},\ldots,\hat{\underset{\sim}{\phi}}^{i'},\ldots,\hat{\underset{\sim}{\phi}}^{\zeta'})$

$$\hat{\underset{\sim}{\phi}}^i = [\sum_{t=1}^{T} X^i(t)'G_{ii}X^i(t)]^{-1}[\sum_{t=1}^{T} X^i(t)'G_{ii}\underset{\sim}{Z}^i(t)] \tag{7-80}$$

This is exactly the same results as those that we estimate all the ζ-variate separately under the condition that G contains no non-null off-diagonal matrix.

As an example, consider the MULSTAR$(2,\underset{\sim}{p},\underset{\sim}{\lambda})$ model with $\underset{\sim}{p}=(2,1,1,1)$ and $\underset{\sim}{\lambda}=(1,0,0,0,0)$ as we mentioned before, we have

$$\underset{\sim}{\phi}^{1'} = (\underset{\sim}{\phi}_{10}^{11},\underset{\sim}{\phi}_{11}^{11},\underset{\sim}{\phi}_{10}^{12},\underset{\sim}{\phi}_{20}^{11})$$

$$\underset{\sim}{\phi}^{2'} = (\phi_{10}^{21},\phi_{10}^{22})$$

$$X(t) = (X^1(t), X^2(t))$$
$$X^1(t) = (\underset{\sim}{X}{}^{11}_{10}(t), X^{11}_{11}(t), X^{12}_{10}(t), X^{11}_{20}(t))$$
$$X^2(t) = (X^{21}_{10}(t), X^{22}_{10}(t)) \tag{7-81}$$

where

$$X^{11}_{10}(t) = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] \underset{\sim}{Z}(t-1),\quad X^{11}_{11}(t) = \left[\begin{array}{c|c} W^{(1)} & 0 \\ \hline 0 & 0 \end{array}\right] \underset{\sim}{Z}(t-1)$$

$$X^{12}_{10}(t) = \left[\begin{array}{c|c} 0 & I \\ \hline 0 & 0 \end{array}\right] \underset{\sim}{Z}(t-1),\quad X^{11}_{20}(t) = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] \underset{\sim}{Z}(t-2)$$

$$X^{21}_{10}(t) = \left[\begin{array}{c|c} 0 & 0 \\ \hline I & 0 \end{array}\right] \underset{\sim}{Z}(t-1),\quad X^{22}_{10}(t) = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array}\right] \underset{\sim}{Z}(t-1)$$

when $\underset{\sim}{\varepsilon}^1(t)$ and $\underset{\sim}{\varepsilon}^2(t)$ are uncorrelated, i.e.,

$$G = [G_{hg}];\ h,g=1,2, \text{ and } G_{12} = G_{21} = 0.$$

We have

$$\sum_{t=1}^{T} X'(t)GX(t) = \sum_{t=1}^{T} \left[\begin{array}{c|c} (X'(t)GX(t))_1 & (X'(t)GX(t))_{12} \\ \hline (X'(t)GX(t))_2 & (X'(t)GX(t))_{22} \end{array}\right] \tag{7-82}$$

CONTINUED

6 OF 8

where

$(X'(t)GX(t))_{12}$, $(X'(t)GX(t))_{21}$ are null submatrices,

$(X'(t)GX(t))_{11} = [g^1_{ij}(t)] \quad i,j=1,2,3,4$

$(X'(t)GX(t))_{22} = [g^2_{ij}(t)] \quad i,j=1,2$

$$g^1_{11}(t) = \underset{\sim}{Z}^1(t-1)'G_{11}\underset{\sim}{Z}^1(t-1)$$

$$g^1_{12}(t) = \underset{\sim}{Z}^1(t-1)'G_{11}w^{(1)}\underset{\sim}{Z}^1(t-1)$$

$$g^1_{13}(t) = \underset{\sim}{Z}^1(t-1)'G_{11}\underset{\sim}{Z}^2(t-1)$$

$$g^1_{14}(t) = \underset{\sim}{Z}^1(t-1)'G_{11}\underset{\sim}{Z}^1(t-2)$$

$$g^1_{22}(t) = (w^{(1)}\underset{\sim}{Z}^1(t-1)'G_{11}(w^{(1)}\underset{\sim}{Z}^1(t-2))$$

$$g^1_{23}(t) = \underset{\sim}{Z}^1(t-1)'w^{(1)}{}'G_{11}\underset{\sim}{Z}^2(t-1)$$

$$g^1_{24}(t) = \underset{\sim}{Z}^1(t-1)'w^{(1)}{}'G_{11}\underset{\sim}{Z}^1(t-2)$$

$$g^1_{33}(t) = \underset{\sim}{Z}^2(t-1)'G_{11}\underset{\sim}{Z}^2(t-1)$$

$$g^1_{34}(t) = \underset{\sim}{Z}^2(t-1)'G_{11}\underset{\sim}{Z}^1(t-2)$$

$$g^1_{44}(t) = \underset{\sim}{Z}^1(t-2)'G_{11}\underset{\sim}{Z}^1(t-2)$$

$$g^1_{ij}(t) = g^1_{ji}(t)$$

$$g_{11}^2(t) = \tilde{Z}^1(t-1)'G_{22}\tilde{Z}^1(t-1)'$$

$$g_{12}^2(t) = \tilde{Z}^1(t-1)'G_{22}\tilde{Z}^2(t-1)'$$

$$g_{22}^2(t) = \tilde{Z}^2(t-1)'G_{22}\tilde{Z}^2(t-1)'$$

$$g_{21}^2(t) = g_{12}^2(t)$$

$$\sum_{t=1}^{T} X'(t)G\tilde{Z}(t) = \begin{bmatrix} \tilde{Z}^1(t-1)'G_{11}\tilde{Z}^1(t) \\ (w^{(1)}\tilde{Z}^1(t-1)'G_{11}\tilde{Z}^1(t) \\ \tilde{Z}^2(t-1)'G_{11}\tilde{Z}^1(t) \\ \tilde{Z}^1(t-2)'G_{11}\tilde{Z}^1(t) \\ \tilde{Z}^1(t-1)'G_{22}\tilde{Z}^2(t) \\ \tilde{Z}^2(t-1)'G_{22}\tilde{Z}^2(t) \end{bmatrix}$$

We get

$$\hat{\tilde{\phi}}^1 = [X'(t)GX(t)]_{11}^{-1} \begin{bmatrix} \tilde{Z}^1(t-1)'G_{11}\tilde{Z}^1(t) \\ (w^{(1)}\tilde{Z}^1(t-1))'G_{11}\tilde{Z}^1(t) \\ \tilde{Z}^2(t-1)'G_{11}\tilde{Z}^1(t) \\ \tilde{Z}^1(t-2)'G_{11}\tilde{Z}^1(t) \end{bmatrix} \qquad (7\text{-}83)$$

$$\hat{\tilde{\phi}}^2 = [X'(t)GX(t)]_{22}^{-1} \begin{bmatrix} \tilde{Z}^1(t-1)'G_{22}\tilde{Z}^2(t) \\ \tilde{Z}^2(t-1)'G_{22}\tilde{Z}^2(t) \end{bmatrix} \qquad (7\text{-}84)$$

The result is exactly the same as that we get by estimating the first category and the second category separately, i.e., estimating the MULSTAR$(2,\tilde{p},\tilde{\lambda})$, $\tilde{p} = (2,1,0,0)$, $\tilde{\lambda} = (1,0)$ model

$$\tilde{Z}^1(t) = \phi_{10}^{11}\tilde{Z}^1(t-1) + \phi_{11}^{11}w^{(1)}\tilde{Z}^1(t-1) + \phi_{10}^{12}\tilde{Z}^2(t-1) + \phi_{20}^{11}\tilde{Z}^1(t-2) + \tilde{\varepsilon}^1(t)$$

and the MULSTAR$(2,\tilde{p},\tilde{\lambda})$, $\tilde{p} = (0,0,1,1)$, $\tilde{\lambda} = (0,0)$ model

$$\tilde{Z}^2(t) = \phi_{10}^{21}\tilde{Z}^1(t-1) + \phi_{10}^{22}\tilde{Z}^2(t-1) + \tilde{\varepsilon}^2(t)$$

separately.

The above result is intuitively correct even without derivation. Since when $\tilde{\varepsilon}^h(t)$ and $\tilde{\varepsilon}^g(t)$, $g \neq h$ are correlated, all ζ variate parameter estimators will use the knowledge of G^{hg} and will result in different estimates from those gotten by separate estimation. But when $\tilde{\varepsilon}^h(t)$ and $\tilde{\varepsilon}^g(t)$, $h \neq j$, are uncorrelated, the knowledge of $G^{hg} = 0$ is used, so $\hat{\tilde{\phi}}^h$, $\hat{\tilde{\phi}}^g$ are independent and it is natural to get the same result.

In previous discussions, we only assumed that Z^* is specified. Z^* is needed to construct the $X(t)$ matrix. In practice, we set $Z^* = 0$, which is its unconditional mean.

<u>7.7.1.2 MULSTMA($\zeta,\boldsymbol{q},\boldsymbol{m}$) Model Estimation</u>. Consider the MULSTMA($\zeta,\boldsymbol{q},\boldsymbol{m}$) model

$$\boldsymbol{Z}^h(t) = -\sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}}\sum_{\ell=0}^{m_k^{hg}} \sigma_{k\ell}^{hg} w^{(\ell)} \boldsymbol{\varepsilon}^g(t-k) + \boldsymbol{\varepsilon}^h(t), \quad h=1,2,\ldots,\zeta$$

$\boldsymbol{\varepsilon}(t)$ is normally distributed with

$$E(\boldsymbol{\varepsilon}^h(t), \boldsymbol{\varepsilon}^g(t+k)') = \begin{cases} G^{hg} & k=0 \\ 0 & \text{otherwise} \end{cases} \tag{7-85}$$

Denote $\boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}^h(t),\ 1-q^{h^*} \le t \le 0;\ h = 1,2,\ldots,\zeta)$ vector, we have the joint density distribution function of $\boldsymbol{\varepsilon}$

$$f(\boldsymbol{\varepsilon}|\boldsymbol{\sigma},\boldsymbol{Z}^*) = (2\pi)^{-\zeta\cdot LN\cdot T/2}|G|^{-T/2}\exp\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\varepsilon}(t)'G^{-1}\boldsymbol{\varepsilon}(t)\} \tag{7-86}$$

The transformation

$$\boldsymbol{\varepsilon}^h(t) = \boldsymbol{Z}^h(t) + \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}}\sum_{\ell=0}^{m_k^{hg}} \theta_{k\ell}^{hg} w^{(\ell)} \boldsymbol{\varepsilon}^g(t-1), \quad h=1,2,\ldots,\zeta \tag{7-87}$$

has unit Jacobian, so the joint density distribution function of $\boldsymbol{Z}$ conditional on σ,ε^* is

$$f(\boldsymbol{Z}|\boldsymbol{\theta},\boldsymbol{\varepsilon}^*) = (2\pi)^{-\zeta\cdot LN\cdot T/2}|G|^{-T/2}\exp\{-\frac{1}{2}SS(\boldsymbol{\theta},\boldsymbol{\varepsilon}^*)\}$$

where

$$SS(\boldsymbol{\theta},\boldsymbol{\varepsilon}^*) = \sum_{t=1}^{T}\boldsymbol{Z}_c(t)'G^{-1}\boldsymbol{Z}_c(t) \tag{7-88}$$

$$\boldsymbol{Z}_c(t)' = (\boldsymbol{Z}_c^1(t)', \boldsymbol{Z}_c^2(t), \ldots, \boldsymbol{Z}_c^{\zeta}(t)')$$

with

$$\boldsymbol{Z}_c^i = \boldsymbol{Z}^i(t) + \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{ig}}\sum_{\ell=0}^{m_k^{ig}} \theta_{k\ell}^{ig}\boldsymbol{\varepsilon}^g(t-k)$$

Take the logarithm of Equation (7-87), with $\boldsymbol{Z}$ available and we have the logarithm likelihood function conditional on ε^*, $\boldsymbol{Z}$ as

$$\ell n\, f(\boldsymbol{\phi}|\boldsymbol{\varepsilon}^*,\boldsymbol{Z}) = -\zeta\cdot LN\cdot T/2 \cdot \ell n^2\pi - \frac{T}{2}\ell n|G| - \frac{T}{2}SS(\boldsymbol{\sigma},\boldsymbol{\varepsilon}^*) \tag{7-89}$$

When the variance-covariance matrix G is given, we see that to maximize the conditional logarithm likelihood function is equivalent to minimize the weighted conditional sum of squares $SS(\boldsymbol{\sigma},\boldsymbol{\varepsilon}^*)$, which is weighted by G^{-1}.

Because of the nonlinear nature of the MULSTMA model in its moving average parameters, we cannot have closed form expression for the conditional maximum likelihood estimator. To get the conditional maximum likelihood estimate, we search through the $\underset{\sim}{\theta}$'s space and we compute the $\underset{\sim}{\varepsilon}(t)$, $t=1,2,\ldots,T$ by recursively applying Equation and setting $\underset{\sim}{\varepsilon}^*$ to its unconditional mean zero.

<u>7.7.1.3 MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) Model Estimation</u>. Consider the following MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) model.

$$\underset{\sim}{Z}^h(t) = \sum_{g=1}^{\zeta}\sum_{k=1}^{p^{hg}}\sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{Z}(t-k) - \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}}\sum_{\ell=0}^{m_k^{hg}} \sigma_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{\varepsilon}(t-k) + \underset{\sim}{\varepsilon}^h(t) \quad (7\text{-}90)$$

where $\underset{\sim}{\varepsilon}(t)$ is normally distributed with

$$E(\underset{\sim}{\varepsilon}(t)\underset{\sim}{\varepsilon}(t+k)') = \begin{cases} G & k=0 \\ 0 & \text{otherwise} \end{cases}$$

The joint density distribution function for $\underset{\sim}{\varepsilon}$ is

$$f(\underset{\sim}{\varepsilon}|G) = (2\pi)^{-\zeta\cdot LN\cdot T/2}|G|^{-T/2} \exp\{-\frac{1}{2}\sum_{t=1}^{T} \underset{\sim}{\varepsilon}(t)'G^{-1}\underset{\sim}{\varepsilon}(t)\}$$

Since the transformation

$$\underset{\sim}{\varepsilon}^h(t) = \underset{\sim}{Z}^h(t) - \sum_{g=1}^{\zeta}\sum_{k=1}^{p^{hg}}\sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{Z}(t-k)$$

$$+ \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}}\sum_{\ell=0}^{m_k^{hg}} \theta_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{\varepsilon}(t-k) \quad (7\text{-}91)$$

has unity Jacobian, so given $\underset{\sim}{Z}^*$, $\underset{\sim}{\varepsilon}^*$, $\underset{\sim}{\phi}$, $\underset{\sim}{\sigma}$, G, we have the joint density distribution function of Z conditional on $\underset{\sim}{Z}^*$, $\underset{\sim}{\varepsilon}^*$, $\underset{\sim}{\phi}$, $\underset{\sim}{\sigma}$, G as follows, where $\underset{\sim}{Z}^*$ and $\underset{\sim}{\varepsilon}^*$ are the values of $\underset{\sim}{Z}(t)$ and $\underset{\sim}{\varepsilon}(t)$ that are needed to start computing $\underset{\sim}{\varepsilon}^h(1)$, $h=1,2,\ldots,\zeta$, in Equation (7-91)

$$f(\underset{\sim}{Z}|\underset{\sim}{\phi},\underset{\sim}{\theta},\underset{\sim}{Z}^*,\underset{\sim}{\varepsilon}^*,G) = (2\pi)^{-\zeta\cdot LN\cdot T/2}|G|^{-T/2} \exp\{-\frac{1}{2}\sum_{t=1}^{T} \underset{\sim}{Z}_c(t)'G^{-1}\underset{\sim}{Z}_c(t)\}$$

where

$$\underset{\sim}{Z}_c(t)' = (\underset{\sim}{Z}_c^1(t)', \underset{\sim}{Z}_c^2(t)', \ldots, \underset{\sim}{Z}_c^{\zeta}(t)')$$

$$\underset{\sim}{Z}_c^i(t) = \underset{\sim}{Z}^i(t) - \sum_{g=1}^{\zeta}\sum_{k=1}^{p^{hg}}\sum_{\ell=0}^{\lambda_k^{hg}} \phi_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{Z}(t-k)$$

$$+ \sum_{g=1}^{\zeta}\sum_{k=1}^{q^{hg}}\sum_{\ell=0}^{m_k^{hg}} \sigma_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{\varepsilon}(t-k) \quad (7\text{-}92)$$

The logarithm likelihood function of ϕ,σ for a given $\underset{\sim}{G}$, $\underset{\sim}{Z}$, $\underset{\sim}{Z}^*$, $\underset{\sim}{\varepsilon}^*$ can be gotten directly from Equation (7-92)

$$\ln f(\underset{\sim}{\phi},\underset{\sim}{\sigma}|\underset{\sim}{Z},\underset{\sim}{Z}^*,\underset{\sim}{\varepsilon}^*,G) = -\zeta\cdot LN\cdot T/2\cdot\ln(2\pi) - \frac{T}{2}\ln|G| - \frac{1}{2}SS(\underset{\sim}{\phi},\underset{\sim}{\sigma},G,\underset{\sim}{\varepsilon}^*,\underset{\sim}{Z}^*) \qquad (7\text{-}93)$$

where $SS(\underset{\sim}{\phi},\underset{\sim}{\sigma},G,\underset{\sim}{\varepsilon}^*,\underset{\sim}{Z}^*) = \sum_{t=1}^{T} Z_c(t)'G^{-1}Z_c(t)$ is the weighted sum square for given $\underset{\sim}{\phi}$, $\underset{\sim}{\sigma}$, $\underset{\sim}{Z}^*$, $\underset{\sim}{\varepsilon}^*$, G.

From Equation (7-93) we see that the conditional maximum likelihood estimates will be equal to the minimum weighted sum square estimates because to maximize the logarithm likelihood function is exactly the same as to minimize the weighted sum square.

Since the MULSTARMA($\zeta,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) model are of linear nature in autoregressive parameters and of nonlinear nature in moving average parameters. So no closed form expression estimators can be available, and at least we have to search through the $\underset{\sim}{\theta}$ parameter space. Because all the $\underset{\sim}{\varepsilon}(t-k)$'s can be expressed in terms of $\underset{\sim}{Z}(t-k)$, $\underset{\sim}{Z}(t-k-1),\ldots,\underset{\sim}{Z}^*$, and this expression is the linear function of $\underset{\sim}{\phi}$, so with $\underset{\sim}{\theta}$ given we can transform the MULSTARMA model into MULSTAR model, which we have closed form expression for the conditional maximum likelihood estimator. Obviously the transformed MULSTAR model will not have the original weight matrix $w^{(\ell)}$ as its weight matrix and will not necessarily retain the original model parameter $\underset{\sim}{p},\underset{\sim}{\lambda}$. By such transformation, we can only search through the $\underset{\sim}{\theta}$ parameters space rather than the whole $(\underset{\sim}{\phi},\underset{\sim}{\theta})$ space to get the conditional maximum likelihood estimates.

7.7.2 Conditional Maximum Likelihood Estimation When G is Unknown

So far in this chapter, we have assumed that G is known, and the conditional maximum likelihood estimators are exactly the same as the weighted conditional minimum least square estimators, but it is not unusual that G is not known. In such case, we will employ the two stage estimation procedure. At the first stage, we will assume the $G = \sigma^2 I$ and get the first stage estimates, say (ϕ^*,θ^*) for MULSTARMA, (ϕ^*) for MULSTAR and (θ^*) for MULSTMA. Then we compute the estimated residuals recursively and test the hypothesis H_0; $G=\sigma^2 I$. If the null hypothesis is accepted, the first stage estimates will be accepted as the final estimates. If the null hypothesis is rejected, we will use the estimated covariance matrix $G = \sum_{t=1}^{T} \underset{\sim}{\varepsilon}(t)\underset{\sim}{\varepsilon}(t)'$ as the true one and perform the second stage estimation to get the estimates of the second stage and accept it as the final estimates.

7.8 Diagnostic Checking

After the model parameters are estimated, the candidate model is subjected to the test of model adequacy as well as model parsimony, i.e., are examined to see if all the significant structures are included and to see if all the structures included are significant or should be eliminated.

To test the model parsomony is equivalent to test the significance of the individual model parameters that appear in the candidate model. Two approaches for testing the significance of model parameters has been given in Section 4.4.2, the confidence region approach and the overfitting approach. The confidence region approach is model independent once the point estimates of the model parameters and the computed sum of squares surface are available. Similarly the over-fitting approach tests the significance of the model parameter only based on the extra sum of squares and is also model independent. Therefore since these two approaches are model independent, they can be applied to test the significance of the parameters of the MULSTARMA model as well.

To test the model adequacy, the model residuals are examined to see if any structure is left. The sample multivaraite space-time autocorrelation functions of the model residuals are computed and encoded into the ".", "+" and "-" symbols according to the procedure proposed in identification section. The sample multivariate space-time autocorrelations by itself can be applied to check if any unexhausted structures exist. However, the encoded sample multivariate space-time autocorrelation functions can be used by itself and is more practical to be applied to solely examine the possibility unexhausted structure.

7.9 Example

The monthly crime data of murder, rape, robbery and burglary at Cleveland are available from February 1970 to October 1974. The Concentrated Crime Patrol (CCP) was implemented as part of the deterrence, detection and apprehension operating program in ~~May 1973.~~ The Concentrated Crime Patrol involved the addition of 120 rearrangement of patrolmen to the Cleveland police force to be deployed ~~to~~ high crime areas during high crime hours; members of the CCP

patrolled the streets in specially marked impact cars responding to all crime-related requests for service. The CCP was deployed only in the three eastside districts (IV, V and VII) as shown on the Map in Figure 7-11. Districts IV, V and VI are thus referred to as the target area, while district III is denoted as the adjacent area and district I and II are the rest or remaining area. For greater detail of this intervention program see J. S. Dahmann (1975). The data set is used to illustrate the modeling procedure described in this chapter, the scaled crime data is modeled from February 1970 to December 1972. This model is employed to construct the forecasting function as well as to evaluate the effect of CCP on the monthly basis.

The data is listed in Table 7-3 and the mean values and variance by location and crime type are contained in Table 7-4, where

Location 1: the target area,
Location 2: the adjacent area,
Location 3: the rest area,

and the following category index have been assigned for the crimes,

Category 1: Murder,
Category 2: Rape,
Category 3: Robbery,
Category 4: Burglary.

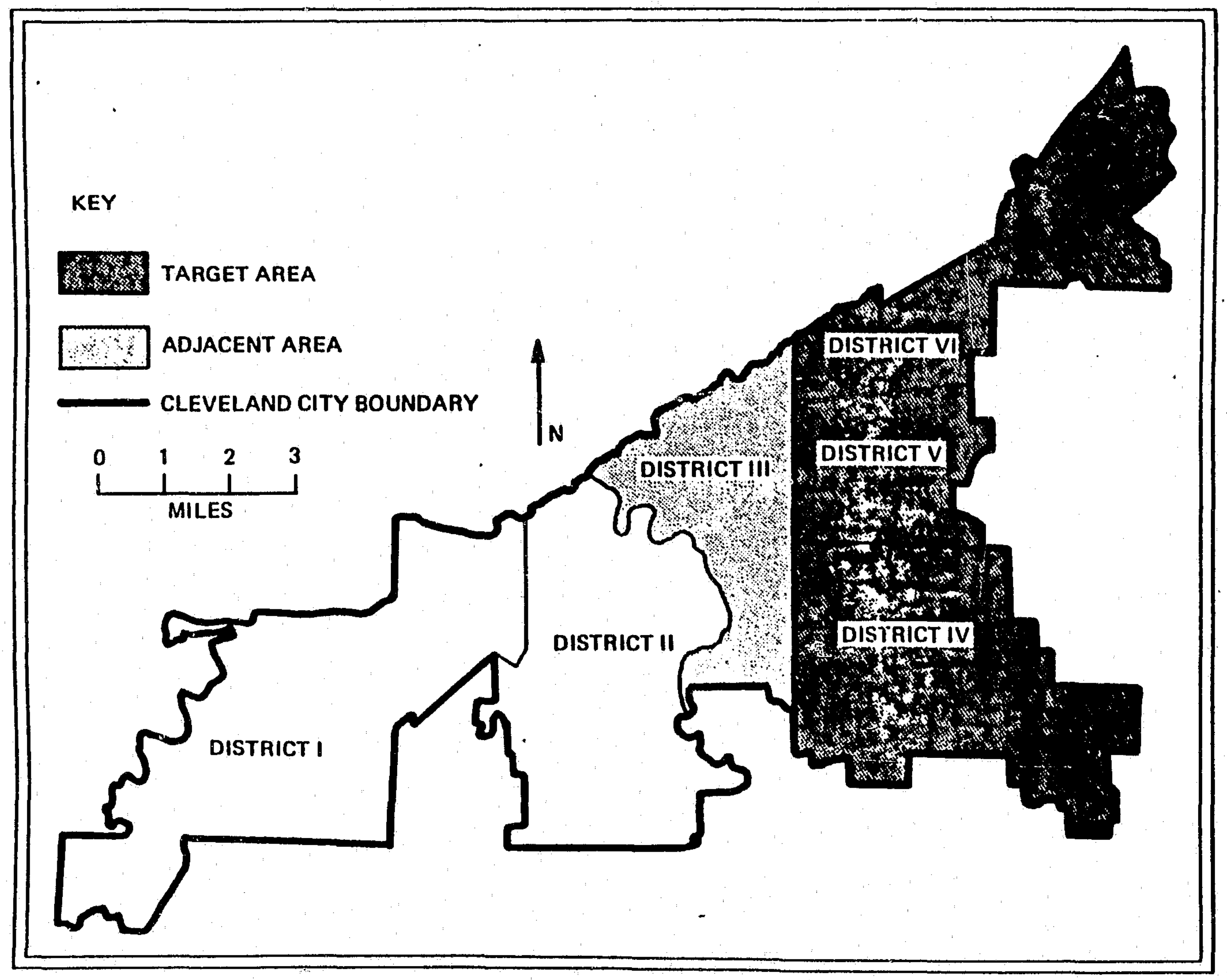


Figure 7-11. CCP Target Areas, Adjacent Areas and Rest Areas.

Table 7-3. The Concentrated Crime Patrol (CCP) Data

t	Category 1			Category 2			Category 3			Category 4		
	Location 1	2	3	Location 1	2	3	Location 1	2	3	Location 1	2	3
1	15	3	3	19	5	2	394	68	52	642	112	164
2	21	2	9	19	3	7	339	85	66	691	124	201
3	22	2	2	19	3	4	326	65	34	702	122	160
4	26	2	0	18	7	2	283	64	42	578	81	140
5	16	1	8	26	1	1	246	59	30	598	99	157
6	13	4	3	13	1	0	307	80	28	557	126	134
7	28	2	2	16	7	2	358	57	37	496	84	131
8	26	2	3	15	7	2	324	72	23	544	71	147
9	21	3	4	19	1	0	379	74	40	703	103	193
10	10	3	4	26	5	1	349	60	34	676	100	184
11	22	2	0	19	8	5	501	68	39	818	100	165
12	14	0	3	17	6	3	401	84	40	789	72	120
13	22	0	2	24	7	2	343	63	40	677	59	167
14	17	0	7	26	11	0	323	67	52	746	76	173
15	21	4	2	22	3	1	359	77	32	655	72	137
16	19	3	7	33	6	3	316	63	36	575	76	176
17	15	2	1	23	3	4	280	73	26	538	66	179
18	18	3	4	19	8	4	343	99	43	611	98	176
19	25	1	6	33	7	1	397	93	40	647	107	184
20	18	1	2	33	9	11	369	96	49	693	111	182
21	24	4	3	30	4	2	459	66	51	838	90	173
22	14	2	7	28	8	1	447	80	47	912	90	226
23	27	3	0	28	4	4	484	94	55	971	84	230
24	23	2	6	27	8	13	391	68	51	647	63	180
25	14	0	3	35	6	7	290	50	72	633	98	214
26	17	4	0	39	3	7	357	58	53	699	124	225
27	22	5	1	22	4	8	296	57	59	531	87	228
28	14	3	2	22	12	3	290	66	49	548	84	177
29	15	2	6	26	6	4	289	73	62	442	79	208
30	32	3	8	25	5	5	285	77	61	422	81	226
31	21	2	3	23	6	13	335	69	52	479	79	186
32	30	4	6	26	5	3	313	68	48	506	70	179
33	21	2	2	33	3	5	396	78	68	644	90	231
34	16	4	3	25	2	7	449	73	87	672	87	263
35	25	0	8	24	3	2	406	70	73	642	85	237
36	21	1	3	30	5	0	399	64	31	453	80	258
37	19	3	1	23	0	9	297	52	19	356	97	260
38	14	1	8	25	2	2	293	42	56	449	80	267
39	15	3	15	16	5	5	285	68	14	393	91	203

Table 7-3. (Cont'd)

t	Category 1			Category 2			Category 3			Category 4		
	Location			Location			Location			Location		
	1	2	3	1	2	3	1	2	3	1	2	3
40	10	0	5	25	2	8	255	58	24	409	68	280
41	17	2	11	20	6	10	190	46	80	361	57	277
42	9	5	5	43	6	0	222	61	53	412	80	266
43	18	4	14	24	4	22	209	55	97	292	63	422
44	8	1	13	25	4	6	232	47	66	325	65	323
45	11	3	8	32	3	6	293	67	25	430	71	257
46	19	1	8	38	5	2	326	66	65	435	67	287
47	23	3	7	24	7	4	321	84	59	429	85	361
48	12	3	2	32	8	8	293	62	56	473	113	280
49	12	4	1	17	10	8	261	53	81	478	89	253
50	9	3	2	40	8	12	323	58	61	565	119	255
51	16	4	3	37	6	12	274	86	92	546	109	294
52	13	3	3	37	3	11	259	69	74	576	99	337
53	16	3	0	31	3	27	281	85	83	605	111	365
54	22	8	3	23	5	7	359	90	84	592	104	378
55	16	7	2	32	7	6	384	111	58	574	94	389
56	21	5	1	36	5	2	306	82	64	605	105	455
57	26	9	4	27	3	9	417	83	224	742	95	486

Table 7-4(a). Mean Levels of the CCP Data (t=1 to 35)

Crime Category	Location 1	2	3
1	20.1	2.3	3.7
2	74.3	5.3	4.0
3	355.0	71.8	47.7
4	643.5	90.0	184.4

Table 7-4(b). Variances of the CCP Data (t=1 to 35)

Crime Category	Location 1	2	3
1	2.7	1.8	6.6
2	36.6	7.0	11.1
3	3746.0	133.0	206.2
4	14770.0	315.6	1147.0

The standard deviations are used to scale the crime data. The distances between centroids of these three areas were measured from the map in Figure 7-11 and are,

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 4.5 & 7.2 \\ 3.5 & 0 & 4.5 \\ 7.2 & 4.5 & 0 \end{bmatrix} \end{array} \quad \text{(in miles)}$$

According to the relative positions and the inverse centroid distances, the weight matrix $w^{(1)}$ and $w^{(2)}$ are assigned. The resulting weight matrices are,

$$w^{(1)} = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 0 & 1 & 0 \\ 7/16 & 0 & 9/16 \\ 0 & 1 & 0 \end{bmatrix} \qquad w^{(2)} = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

7.9.1 MULSTARMA Model Building

To build the MULSTARMA model for the Cleveland Crime Data, the sample multivariate space-time autocorrelation function and the sample multivariate space-time partial autocorrelation function are computed and encoded into dot, plus, minus symbols. The numerical values are listed on Table 7-5 and the encoded symbols are contained in Figure 7-12. In Figure 7-12 and Table 7-5, we see that, $\hat{\rho}_{00}^{24}(k)$, $\hat{\rho}_{10}^{44}(k)$ cuts off at k=2, $\hat{\rho}_{00}^{22}(k)$, $\hat{\rho}_{20}^{22}(k)$, $\hat{\rho}_{00}^{33}(k)$, $\hat{\rho}_{00}^{23}(k)$, $\hat{\rho}_{00}^{44}(k)$, $\hat{\rho}_{00}^{34}$ tails off,

Table 7-5(a). The Sample Multivariate Space-Time Autocorrelations for the Cleveland Crime Data

	$-\hat{\rho}_{50}^{11}(k)$				Standardized $\hat{\rho}_{50}^{11}(k)$		
	Space Lag						
Time Lag	0	1	2				
1	-0.14	0.03	0.07	/	-1.41	0.36	0.72
2	-0.11	0.03	0.10	/	-1.12	0.32	1.00
3	-0.04	0.17	0.09	/	-0.42	1.67	0.88
4	-0.09	0.08	-0.37	/	-0.93	0.80	-3.57
5	0.08	-0.12	0.16	/	0.75	-1.19	1.55
6	0.09	-0.01	0.05	/	0.87	-0.11	0.50
7	-0.13	-0.02	-0.06	/	-1.22	-0.20	-0.57
8	0.03	0.16	-0.00	/	0.35	1.52	-0.02
9	-0.10	0.16	0.00	/	-0.94	1.44	0.00
10	-0.08	-0.19	-0.11	/	-0.70	-1.67	-1.01
11	0.17	-0.17	0.25	/	1.47	-1.49	2.12
12	0.05	0.01	-0.16	/	0.44	0.15	-1.38
13	-0.12	0.14	-0.03	/	-1.05	1.15	-0.31
14	-0.06	0.24	0.16	/	-0.47	1.90	1.31
15	-0.13	0.14	-0.15	/	-1.01	1.11	-1.20
16	0.21	-0.19	0.18	/	1.64	-1.48	1.37
17	0.13	-0.19	0.04	/	1.02	-1.42	0.34
18	-0.35	-0.18	-0.01	/	-2.55	-1.32	-0.10
19	0.05	-0.12	-0.14	/	0.39	-0.88	-1.00
20	-0.11	0.00	-0.31	/	-0.73	0.05	-2.11

Table 7-5(a). (Cont'd)

$-\hat{\rho}^{12}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{12}_{50}(k)$		
1	0.12	-0.02	0.11	/	1.23	-0.20	1.17
2	0.16	-0.05	-0.08	/	1.67	-0.49	-0.86
3	-0.03	-0.00	0.02	/	-0.35	-0.01	0.19
4	-0.12	0.08	0.12	/	-1.21	0.77	1.20
5	-0.05	0.08	-0.02	/	-0.49	0.75	-0.27
6	-0.03	-0.05	-0.17	/	-0.35	-0.52	-1.58
7	-0.06	-0.11	0.00	/	-0.61	-1.02	0.00
8	-0.15	-0.10	-0.01	/	-1.38	-0.94	-0.12
9	-0.08	0.07	-0.06	/	-0.71	0.65	-0.59
10	0.37	0.08	-0.10	/	0.31	0.71	-0.91
11	-0.12	-0.13	0.06	/	-1.02	-1.16	0.54
12	0.10	-0.25	-0.04	/	0.86	-2.13	-0.37
13	-0.01	-0.23	-0.08	/	-0.11	-1.87	-0.68
14	-0.15	-0.08	-0.15	/	-1.25	-0.64	-1.22
15	0.03	0.03	-0.17	/	0.24	0.26	-1.32
16	-0.26	0.06	0.02	/	-2.00	0.65	0.19
17	-0.04	-0.06	0.09	/	-0.32	-0.46	0.66
18	0.20	-0.05	0.01	/	1.43	-0.39	0.05
19	0.00	-0.17	-0.15	/	0.04	-1.18	-1.03
20	-0.29	-0.03	0.14	/	-1.94	-0.24	0.96

Table 7-5(a). (Cont'd)

$\hat{\rho}^{13}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{13}_{50}(k)$		
1	-0.10	0.08	-0.00	/	-1.07	0.84	-0.01
2	0.13	0.00	0.15	/	1.38	0.06	1.50
3	0.12	0.12	0.05	/	1.19	1.21	0.57
4	0.09	0.13	0.12	/	0.87	1.25	1.16
5	-0.00	0.07	0.10	/	-0.04	0.66	1.01
6	-0.21	0.03	0.01	/	-2.01	0.28	0.12
7	-0.08	0.07	0.02	/	-0.79	0.72	0.20
8	-0.08	-0.01	-0.21	/	-0.73	-0.12	-1.93
9	-0.12	-0.22	-0.03	/	-1.12	-1.97	-0.33
10	-0.18	-0.12	0.02	/	-1.58	-1.08	0.24
11	0.01	-0.11	-0.08	/	0.14	-0.97	-0.70
12	0.08	0.02	-0.06	/	0.70	0.17	-0.57
13	0.11	0.06	-0.11	/	0.94	0.53	-0.92
14	0.07	0.01	0.02	/	0.61	0.09	0.17
15	0.10	0.07	-0.13	/	0.82	0.60	-1.06
16	0.00	0.04	0.03	/	0.04	0.36	0.22
17	-0.08	0.03	-0.03	/	-0.61	0.28	-0.24
18	-0.05	0.08	0.06	/	-0.37	0.58	0.47
19	0.06	0.01	0.01	/	0.46	0.12	0.09
20	0.05	-0.15	-0.04	/	0.35	-1.00	-0.32

Table 7-5(a). (Cont'd)

$\hat{\rho}_{50}^{14}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}_{50}^{14}(k)$		
1	-0.04	0.02	-0.00	/	-0.47	0.28	-0.06
2	-0.00	0.02	0.09	/	-0.05	0.25	0.91
3	0.11	0.05	0.13	/	1.13	0.48	1.28
4	0.14	0.10	0.12	/	1.42	0.98	1.22
5	0.06	0.03	-0.04	/	0.65	0.33	-0.43
6	-0.19	0.03	0.04	/	-1.84	0.30	0.45
7	-0.11	0.19	0.07	/	-1.01	1.75	0.68
8	-0.06	-0.04	-0.17	/	-0.61	-0.42	-0.53
9	-0.01	-0.29	0.00	/	-0.82	-2.60	0.03
10	-0.01	-0.21	-0.04	/	-0.13	-1.85	-0.35
11	0.02	-0.08	-0.07	/	0.18	-0.67	-0.60
12	-0.03	0.03	-0.06	/	-0.25	0.71	-0.52
13	-0.09	0.00	-0.14	/	-0.76	0.02	-1.19
14	-0.02	-0.08	0.02	/	-0.21	-0.64	0.18
15	0.08	-0.06	-0.05	/	0.66	-0.51	-0.40
16	0.00	0.05	-0.06	/	0.03	0.40	-0.49
17	0.06	0.17	-0.14	/	0.45	1.26	-1.02
18	0.09	0.13	0.10	/	0.67	0.95	0.73
19	0.15	0.10	0.08	/	1.04	0.72	0.61
20	0.18	-0.05	0.07	/	1.23	-0.37	0.47

Table 7-5(a). (Cont'd)

$\hat{\rho}_{50}^{21}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}_{50}^{21}(k)$		
1	-0.11	0.11	-0.05	/	-1.16	1.12	-0.55
2	0.07	0.13	-0.06	/	0.72	1.30	-0.60
3	0.00	0.00	0.09	/	0.08	0.08	0.91
4	0.21	0.04	0.05	/	2.06	0.43	0.56
5	0.09	0.08	0.12	/	0.87	0.81	1.19
6	0.04	0.03	0.19	/	0.40	0.34	1.78
7	0.08	0.19	-0.02	/	0.73	1.82	-0.22
8	0.18	0.22	0.02	/	1.70	2.01	0.21
9	0.05	-0.16	0.13	/	0.48	-1.48	1.23
10	0.21	0.16	0.08	/	1.84	1.42	0.69
11	0.14	0.13	0.09	/	1.21	1.17	0.82
12	0.07	0.08	0.07	/	0.58	0.68	0.58
13	0.02	-0.03	-0.00	/	0.20	-0.28	-0.01
14	0.23	-0.10	0.03	/	1.89	-0.84	0.26
15	-0.10	-0.20	0.07	/	-0.82	-1.61	0.55
16	0.18	0.20	0.08	/	1.39	1.55	0.62
17	-0.08	0.47	-0.05	/	-0.59	0.34	-0.43
18	-0.01	0.01	-0.08	/	-0.08	0.11	-1.14
19	0.16	0.09	0.27	/	1.14	0.63	1.90
20							

Table 7-5(a). (Cont'd)

$\hat{\rho}_{50}^{23}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}_{50}^{23}(k)$		
1	0.28	-0.03	0.12	/	2.86	-0.31	1.23
2	0.24	0.02	0.15	/	2.47	0.22	1.53
3	0.08	0.11	0.14	/	0.80	1.11	1.44
4	0.16	0.04	0.20	/	1.63	0.44	2.01
5	0.12	0.01	0.06	/	1.17	0.15	0.58
6	-0.00	0.01	0.06	/	-0.05	0.12	0.58
7	0.07	0.12	0.11	/	0.71	1.17	1.07
8	0.01	-0.06	0.23	/	1.04	-0.59	2.12
9	0.13	0.03	0.39	/	1.15	0.26	3.44
10	0.12	0.09	0.27	/	1.08	0.84	2.35
12	-0.00	0.06	0.22	/	-0.01	0.57	1.93
13	-0.11	0.02	0.03	/	-0.97	0.20	0.28
14	-0.21	0.04	0.13	/	-1.73	0.33	1.06
15	-0.16	-0.02	0.14	/	-1.27	-0.16	1.16
16	-0.18	0.11	0.09	/	-1.39	0.85	0.75
17	-0.21	-0.15	-0.14	/	-2.18	-1.10	-1.07
18	-0.15	-0.16	-0.08	/	-1.08	-1.17	-0.61
19	-0.05	0.02	-0.08	/	-0.34	0.15	-0.58
20	-0.21	0.15	-0.04	/	-1.42	1.01	-0.29

Table 7-5(a). (Cont'd)

$\hat{\rho}_{50}^{24}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}_{50}^{24}(k)$		
1	0.07	0.04	0.17	/	0.73	0.50	1.80
2	0.21	0.11	0.15	/	2.14	1.18	1.50
3	0.10	0.09	0.16	/	0.99	0.96	1.62
4	-0.07	-0.00	0.07	/	-0.72	-0.03	0.75
5	-0.06	0.07	-0.02	/	-0.57	0.67	-0.22
6	0.09	0.15	0.04	/	0.88	1.40	0.38
7	0.11	0.09	0.01	/	1.09	0.83	0.17
8	-0.03	-0.01	0.09	/	-0.34	-0.16	0.86
9	-0.00	0.18	0.20	/	-0.04	1.58	1.81
10	-0.06	0.12	0.15	/	-0.59	1.06	1.38
11	-0.10	0.07	0.10	/	-0.88	0.61	0.84
12	-0.08	0.07	0.03	/	-0.73	0.61	0.28
13	-0.20	-0.02	0.06	/	-1.65	-0.16	0.56
14	-0.14	-0.01	0.05	/	-1.11	-0.09	0.42
15	-0.14	0.17	-0.03	/	-1.09	1.33	-0.26
16	-0.31	-0.06	-0.13	/	-2.35	-0.50	-1.02
17	-0.31	-0.33	-0.12	/	-2.32	-2.49	-0.94
18	-0.10	-0.22	-0.09	/	-0.77	-1.58	-0.68
19	-0.16	-0.03	-0.09	/	-1.12	-0.23	-0.67
20	-0.30	-0.12	-0.01	/	-2.02	-0.85	-0.11

Table 7-5(a). (Cont'd)

$\hat{\rho}^{31}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{31}_{50}(k)$		
1	-0.01	0.10	0.09	/	-0.10	1.08	0.95
2	-0.16	-0.00	0.00	/	-1.59	-0.05	0.05
3	-0.01	-0.05	0.02	/	-0.11	-0.48	0.20
4	-0.11	-0.00	-0.10	/	1.10	-0.06	-1.00
5	-0.10	-0.09	0.05	/	-1.00	-0.85	0.47
6	-0.03	-0.03	0.05	/	-0.31	-0.28	0.53
7	0.06	0.12	0.18	/	0.58	1.18	1.72
8	0.21	-0.07	0.03	/	1.89	-0.63	0.28
9	0.31	0.00	0.16	/	2.74	0.04	1.48
10	0.16	0.02	0.04	/	1.39	0.24	0.38
11	0.02	-0.06	0.07	/	0.17	-0.55	0.64
12	0.22	0.32	0.03	/	1.86	2.68	0.29
13	-0.10	0.08	0.00	/	-0.86	0.72	0.06
14	0.03	-0.07	-0.06	/	0.30	-0.61	-0.48
15	0.01	0.09	-0.19	/	0.10	0.75	-1.49
16	-0.12	-0.02	-0.13	/	-0.90	-0.20	-1.00
17	-0.13	0.08	-0.10	/	-0.98	0.64	-0.07
18	-0.05	0.09	0.03	/	-0.40	0.65	0.25
19	0.00	-0.11	0.09	/	0.01	-0.80	0.66
20	0.14	0.04	-0.03	/	0.93	0.26	-0.22

Table 7-5(a). (Cont'd)

$\hat{\rho}^{32}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{32}_{50}(k)$		
1	0.19	0.16	0.20	/	1.95	1.66	2.02
2	0.31	0.20	0.09	/	3.15	2.06	0.93
3	0.02	0.11	0.11	/	0.21	1.11	1.09
4	0.13	0.07	0.03	/	1.30	0.67	0.32
5	0.18	0.13	0.09	/	1.70	1.28	0.91
6	0.13	0.07	-0.04	/	1.22	0.73	-0.40
7	-0.04	0.14	-0.05	/	-0.39	1.31	-0.51
8	0.07	0.12	0.02	/	0.65	1.15	0.25
9	0.05	-0.03	0.02	/	0.46	-0.31	0.23
10	0.14	-0.08	-0.06	/	1.26	-0.72	-0.54
11	-0.04	0.05	-0.09	/	-0.38	0.44	-0.78
12	-0.09	-0.03	0.00	/	-0.75	-0.27	0.04
13	-0.04	-0.02	-0.04	/	-0.39	-0.20	-0.36
14	-0.16	0.05	-0.17	/	-1.31	0.40	-1.36
15	-0.33	-0.13	-0.14	/	-2.57	-1.05	-1.14
16	-0.24	-0.14	-0.23	/	-1.84	-1.10	-1.77
17	-0.33	-0.09	-0.21	/	-2.42	-0.66	-1.61
18	-0.15	0.08	-0.20	/	-1.13	0.57	-1.46
19	-0.23	-0.09	-0.11	/	-1.64	-0.67	-0.76
20	-0.23	-0.21	-0.13	/	-1.56	-1.41	-0.91

Table 7-5(a). (Cont'd)

$\hat{\rho}^{33}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{33}_{50}(k)$		
1	0.47	0.15	0.14	/	4.83	1.54	1.45
2	0.27	0.05	0.02	/	2.71	0.52	0.23
3	0.15	0.11	-0.06	/	1.51	1.16	-0.61
4	0.06	0.06	-0.13	/	0.60	0.62	-1.27
5	-0.08	0.05	-0.15	/	-0.78	0.51	-1.48
6	-0.21	-0.07	-0.17	/	-1.98	-0.70	-1.59
7	-0.20	-0.70	-0.06	/	-1.91	-0.64	-0.57
8	-0.13	-0.14	-0.01	/	-1.19	-1.33	-0.08
9	-0.03	-0.12	0.11	/	-0.26	-1.05	1.01
10	0.14	-0.04	0.19	/	1.29	-0.38	1.64
11	0.25	0.15	0.23	/	2.13	1.30	1.96
12	0.22	0.08	0.27	/	1.90	0.67	2.27
13	-0.01	-0.07	0.17	/	-0.09	-0.58	1.43
14	-0.05	-0.00	0.13	/	-0.45	-0.04	1.06
15	-0.20	0.11	-0.06	/	-1.57	0.90	-0.53
16	-0.29	0.13	-0.16	/	-2.24	0.99	-1.22
17	-0.41	0.06	-0.19	/	-3.05	0.47	-1.43
18	-0.39	-0.09	-0.16	/	-2.82	-0.64	-1.19
19	-0.41	0.03	0.02	/	-2.86	0.24	0.17
20	-0.15	-0.10	0.03	/	-1.01	-0.07	0.21

Table 7-5(a). (Cont'd)

$\hat{\rho}^{34}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{34}_{50}(k)$		
1	0.37	0.03	0.06	/	3.80	0.35	0.69
2	0.33	0.12	-0.00	/	3.30	1.23	-0.04
3	0.30	0.31	-0.08	/	2.97	3.06	-0.81
4	0.15	0.27	-0.17	/	1.49	2.63	-1.70
5	-0.02	0.17	-0.27	/	-0.22	1.69	-2.58
6	-0.13	-0.03	-0.27	/	-1.29	-0.32	-2.58
7	-0.13	0.07	-0.08	/	-0.12	0.71	-0.76
8	-0.00	-0.00	-0.03	/	-0.04	-0.02	-0.31
9	-0.09	-0.18	-0.01	/	-0.80	-1.63	-0.09
10	-0.02	-0.09	0.09	/	-0.24	-0.78	0.77
11	0.04	0.05	0.21	/	0.41	0.42	1.83
12	0.03	-0.07	0.25	/	0.26	-0.66	2.13
13	-0.05	-0.14	0.10	/	-0.43	-1.19	0.85
14	-0.03	0.09	0.09	/	-0.24	0.76	0.72
15	-0.20	0.14	-0.00	/	-1.57	1.10	-0.00
16	-0.29	0.05	-0.07	/	-2.19	0.42	-0.54
17	-0.39	0.03	-0.25	/	-2.91	0.26	-1.87
18	-0.48	-0.12	-0.16	/	-3.42	-0.86	-1.15
19	-0.40	-0.05	0.06	/	-2.83	-0.34	0.46
20	-0.22	-0.04	0.04	/	-1.50	-0.32	0.32

Table 7-5(a). (Cont'd)

$\hat{\rho}^{41}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{41}_{50}(k)$		
1	0.01	-0.03	-0.02	/	0.17	-0.36	-0.28
2	-0.10	-0.05	0.01	/	-1.02	-0.55	0.19
3	-0.12	-0.07	-0.06	/	-1.24	-0.77	-0.58
4	-0.02	0.19	-0.03	/	-0.26	1.89	-0.30
5	0.04	0.12	0.03	/	0.45	1.17	0.30
6	-0.00	-0.00	0.00	/	-0.01	-0.06	0.00
7	0.07	-0.01	0.13	/	0.70	-0.11	1.27
8	0.11	-0.14	0.07	/	0.98	-1.31	0.65
9	0.12	-0.00	0.13	/	1.08	-0.01	1.20
10	-0.00	0.12	0.19	/	-0.04	1.07	1.65
11	-0.08	-0.05	-0.00	/	-0.68	-0.47	-0.02
12	0.13	0.11	-0.03	/	1.12	0.94	-0.28
13	-0.02	0.19	0.13	/	-0.20	1.55	1.13
14	-0.10	-0.09	0.01	/	-0.82	-0.71	0.15
15	-0.13	-0.15	-0.15	/	-1.04	-1.19	-1.16
16	-0.03	-0.06	-0.00	/	-0.26	-0.48	-0.05
17	-0.13	-0.15	-0.06	/	-1.00	-1.16	-0.45
18	0.01	-0.01	-0.03	/	0.13	-0.12	-0.25
19	-0.01	-0.15	0.10	/	-0.08	-1.05	0.69
20	0.18	-0.18	-0.04	/	1.21	-1.25	-0.30

Table 7-5(a). (Cont'd)

$\hat{\rho}^{22}_{50}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}^{22}_{50}(k)$		
1	0.22	0.07	0.31	/	2.24	0.73	3.15
2	0.08	0.22	0.28	/	0.83	2.24	2.84
3	0.24	0.17	0.15	/	2.39	1.72	1.53
4	0.31	0.12	0.15	/	2.99	1.24	1.46
5	0.06	0.13	0.26	/	0.59	1.27	2.51
6	0.26	0.24	0.20	/	2.47	2.28	1.91
7	0.19	-0.00	0.07	/	1.81	-0.02	0.72
8	-0.00	0.03	0.20	/	-0.02	0.31	1.84
9	0.01	0.11	0.06	/	0.17	0.99	0.54
10	0.04	-0.09	0.02	/	0.39	-0.83	0.23
11	-0.08	-0.00	-0.01	/	-0.71	-0.07	-0.09
12	-0.17	0.24	-0.10	/	-1.48	2.01	-0.83
13	-0.21	0.12	-0.08	/	-1.74	1.04	-0.71
14	-0.11	-0.20	-0.09	/	-0.89	-1.59	-0.72
15	-0.31	-0.15	-0.05	/	-2.45	-1.23	-0.45
16	-0.20	-0.10	-0.35	/	-1.54	-0.80	-2.67
17	-0.22	-0.06	-0.20	/	-1.66	-0.49	-1.49
18	-0.29	-0.17	-0.28	/	-2.13	-1.22	-2.03
19	-0.52	-0.10	-0.37	/	-3.63	-0.72	-2.59
20	-0.22	-0.52	-0.31	/	-1.53	-0.35	-2.08

Table 7-5(a). (Cont'd)

$\hat{\rho}^{42}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{42}_{50}(k)$		
1	0.16	-0.05	0.10	/	1.71	-0.55	1.09
2	0.31	-0.08	0.12	/	3.08	-0.88	1.26
3	0.11	-0.23	0.22	/	1.07	-2.32	2.15
4	0.12	-0.15	0.07	/	1.20	-1.53	0.67
5	0.11	0.02	0.08	/	1.12	0.24	0.75
6	-0.01	-0.09	0.09	/	-0.16	-0.83	0.88
7	0.00	-0.24	-0.00	/	0.07	-2.23	-0.06
8	0.10	-0.22	0.08	/	0.92	-2.01	0.74
9	0.08	-0.26	0.02	/	0.78	-2.36	0.19
10	0.16	-0.20	0.02	/	1.42	-1.81	0.17
11	0.19	-0.19	0.02	/	1.64	-1.63	0.23
12	-0.04	-0.42	-0.05	/	-0.36	-3.56	-0.42
13	-0.03	-0.21	-0.12	/	-0.29	-1.72	-1.04
14	-0.05	0.07	-0.16	/	-0.40	0.55	-1.29
15	-0.13	0.05	-0.07	/	-1.01	0.45	-0.54
16	-0.21	-0.18	-0.12	/	-1.60	-1.39	-0.94
17	-0.03	0.00	-0.14	/	-0.22	0.06	-1.04
18	-0.00	0.11	-0.28	/	-0.02	0.81	-2.05
19	-0.14	0.15	-0.22	/	-1.02	1.10	-1.55
20	-0.20	0.14	-0.16	/	-1.35	0.97	-1.07

Table 7-5(a). (Cont'd)

$\hat{\rho}^{43}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{43}_{50}(k)$		
1	0.32	0.03	0.10	/	3.29	0.32	1.03
2	0.12	-0.18	0.01	/	1.25	-1.82	0.12
3	0.07	-0.18	-0.13	/	0.68	-1.79	-1.34
4	-0.04	-0.15	-0.12	/	-0.42	-1.53	-1.24
5	-0.17	-0.18	-0.19	/	-1.65	-1.72	-1.84
6	-0.23	-0.21	-0.17	/	-2.19	-1.96	-1.66
7	-0.10	-0.15	-0.10	/	-1.00	-1.43	-0.93
8	0.07	0.00	-0.03	/	0.65	0.02	-0.30
9	0.21	0.04	0.07	/	1.93	0.42	0.65
10	0.17	-0.16	0.11	/	1.48	-1.38	1.02
11	0.27	-0.26	0.26	/	2.35	-2.25	2.21
12	0.40	-0.14	0.34	/	3.34	-1.20	2.85
13	0.21	-0.09	0.27	/	1.71	-0.70	2.25
14	0.02	-0.09	0.09	/	0.18	-0.76	0.75
15	-0.17	-0.01	-0.03	/	-1.36	-0.09	-0.27
16	-0.13	-0.06	-0.09	/	-0.99	-0.46	-0.70
17	-0.02	-0.04	-0.13	/	-0.15	-0.30	-1.00
18	-0.16	-0.02	-0.12	/	-1.20	-0.19	-0.91
19	-0.24	0.04	-0.06	/	-1.69	0.33	-0.42
20	-0.07	0.19	0.06	/	-0.50	1.28	0.42

Table 7-5(a). (Cont'd)

$\hat{\rho}^{44}_{50}(k)$

	Space Lag				Standardized $\hat{\rho}^{44}_{50}(k)$		
Time Lag	0	1	2				
1	0.56	-0.05	0.01	/	5.72	-0.52	0.15
2	0.22	-0.14	-0.10	/	2.22	-1.40	-1.00
3	0.21	-0.10	-0.12	/	2.10	-0.10	-1.24
4	0.12	-0.03	-0.12	/	1.22	-0.34	-1.18
5	-0.03	-0.18	-0.20	/	-0.30	-1.79	-1.90
6	-0.16	-0.24	-0.21	/	-1.49	-2.30	-1.97
7	-0.04	-0.03	-0.13	/	-0.41	-0.28	-1.27
8	0.80	0.04	-0.10	/	0.72	0.44	-0.94
9	0.04	-0.15	-0.07	/	0.38	-1.32	-0.69
10	-0.03	-0.29	-0.03	/	-0.30	-2.53	-0.33
11	0.03	-0.15	0.15	/	0.30	-1.31	1.28
12	0.05	-0.04	0.29	/	0.41	-0.41	2.41
13	-0.06	-0.06	0.20	/	-0.55	-0.53	1.65
14	-0.15	-0.02	0.09	/	-1.18	-0.19	0.72
15	-0.23	0.08	-0.03	/	-1.81	0.66	-0.24
16	-0.14	0.27	-0.08	/	-1.10	2.09	-0.63
17	-0.10	0.27	-0.10	/	-0.78	2.00	-0.75
18	-0.35	-0.04	-0.10	/	-2.52	-0.30	-0.71
19	-0.28	0.06	-0.03	/	-1.96	0.41	-0.26
20	-0.07	0.32	0.10	/	-0.46	2.15	0.72

Table 7-5(b). The Sample Multivariate Space-Time Partial Autocorrelation Functions for the Cleveland Crime Data

$\hat{\phi}^{11}_{k\ell}$

	Space Lag				Standardized $\hat{\phi}^{11}_{k\ell}$		
Time Lag	0	1	2				
1	-0.17	0.06	0.05	/	-1.73	0.66	0.50
2	-0.06	0.03	0.25	/	-0.63	0.34	2.52
3	-0.13	0.26	0.19	/	-1.36	2.62	1.86
4	-0.06	0.02	-0.08	/	-0.66	0.23	-0.79
5	0.02	-0.02	-0.21	/	0.22	-0.23	-2.04

$\hat{\phi}^{12}_{k}$

	Space Lag				Standardized $\hat{\phi}^{12}_{k\ell}$		
Time Lag	0	1	2				
1	-0.15	0.11	-0.03	/	-1.54	1.18	-0.32
2	0.14	0.12	-0.07	/	1.40	1.20	-0.77
3	0.00	0.16	0.16	/	0.01	1.57	1.65
4	0.14	0.01	0.02	/	1.37	0.10	0.24
5	-0.00	0.15	-0.07	/	-0.00	1.48	-0.74

$\hat{\phi}^{13}_{k\ell}$

	Space Lag				Standardized $\hat{\phi}^{13}_{k\ell}$		
Time Lag	0	1	2				
1	0.01	0.12	0.31	/	0.19	1.30	3.21
2	-0.21	-0.00	-0.06	/	-2.15	-0.01	-0.66
3	0.12	-0.01	0.10	/	1.22	-0.09	0.98
4	-0.00	0.06	-0.16	/	-0.02	0.60	-1.55
5	-0.14	-0.12	0.57	/	-1.40	-1.18	5.47

Table 7-5(b). (Cont'd)

$\hat{\phi}_{k\ell}^{14}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}_{k\ell}^{14}$		
1	0.02	-0.09	-0.25	/	0.29	-0.99	-2.55
2	-0.08	-0.03	0.21	/	-0.79	-0.32	2.10
3	-0.18	-0.24	-0.11	/	-1.77	-2.44	-1.16
4	0.00	0.07	0.20	/	0.07	0.68	1.93
5	0.20	0.02	-0.38	/	1.89	0.26	-3.69

$\hat{\phi}_{k\ell}^{21}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}_{k\ell}^{21}$		
1	0.15	0.02	0.10	/	1.52	0.23	1.09
2	0.11	0.01	-0.06	/	1.11	0.16	-0.64
3	0.08	-0.05	-0.02	/	0.86	-0.54	-0.21
4	-0.07	0.03	0.16	/	-0.70	0.30	1.63
5	-0.06	-0.04	0.04	/	-0.64	-0.45	0.45

$\hat{\phi}_{k\ell}^{22}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}_{k\ell}^{22}$		
1	0.22	0.03	0.30	/	2.29	0.32	3.12
2	0.00	0.20	0.21	/	0.04	2.01	2.10
3	0.20	-0.02	0.05	/	2.00	-0.23	0.51
4	0.07	0.01	0.16	/	0.74	0.13	1.58
5	-0.13	-0.10	-0.28	/	-1.27	-0.96	-2.71

Table 7-5(b). (Cont'd)

$\hat{\phi}_{k\ell}^{23}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}_{k\ell}^{23}$		
1	0.08	0.17	0.20	/	0.81	1.74	2.06
2	0.10	0.01	-0.26	/	1.06	0.18	-2.59
3	-0.25	0.08	-0.10	/	-2.51	0.79	-0.98
4	0.01	-0.11	-0.13	/	0.18	-1.10	-1.27
5	0.00	0.03	-0.26	/	0.04	0.34	-2.50

$\hat{\phi}_{k\ell}^{24}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}_{k\ell}^{24}$		
1	0.09	-0.11	-0.16	/	0.99	-1.11	-1.64
2	0.20	-0.14	0.26	/	2.04	-1.45	2.65
3	-0.08	-0.10	0.22	/	-0.81	-1.07	2.15
4	0.10	-0.05	-0.12	/	0.98	-0.57	-1.24
5	0.02	0.10	0.25	/	0.21	1.03	2.44

$\hat{\phi}_{k\ell}^{31}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}_{k}^{31}$		
1	-0.11	0.06	0.00	/	-1.16	0.70	0.03
2	0.19	0.01	0.21	/	1.94	0.13	2.08
3	-0.00	0.17	0.05	/	-0.40	1.75	0.54
4	0.03	0.10	0.26	/	0.36	0.97	2.58
5	0.00	-0.07	-0.16	/	0.00	-0.72	-1.59

Table 7-5(b). (Cont'd)

$\hat{\phi}_{k\ell}^{32}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{32}$		
1	0.17	-0.04	0.02	/	1.76	-0.46	0.21
2	0.15	-0.02	0.05	/	1.55	-0.26	0.58
3	-0.09	0.11	0.08	/	-0.90	1.14	0.87
4	0.12	-0.06	0.42	/	1.21	-0.58	4.09
5	0.01	-0.27	-0.40	/	0.10	-2.56	-3.80

$\hat{\phi}_{k\ell}^{33}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{33}$		
1	0.39	0.11	0.01	/	4.03	1.20	0.16
2	0.07	0.02	-0.24	/	0.72	0.24	-2.38
3	-0.12	0.03	-0.05	/	-1.18	0.35	-0.55
4	0.06	0.01	-0.04	/	0.61	0.17	-0.46
5	-0.13	0.26	0.06	/	-1.23	2.49	0.65

$\hat{\phi}_{k\ell}^{34}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{34}$		
1	0.09	-0.08	-0.01	/	0.97	-0.85	-0.14
2	-0.21	-0.19	0.09	/	-2.17	-1.96	0.91
3	0.07	-0.06	-0.12	/	0.68	-0.64	-1.20
4	-0.18	-0.10	-0.17	/	-1.77	-1.04	-1.69
5	-0.03	-0.10	-0.33	/	-0.30	-0.98	-3.18

Table 7-5(b). (Cont'd)

$\hat{\phi}_{k\ell}^{41}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{41}$		
1	-0.08	0.08	0.04	/	-0.90	0.90	0.42
2	0.05	0.00	0.13	/	0.58	0.00	1.34
3	0.05	0.09	0.07	/	0.50	0.94	0.70
4	0.00	0.08	0.25	/	0.06	0.83	2.48
5	0.00	-0.04	-0.17	/	0.05	-0.39	-1.67

$\hat{\phi}_{k\ell}^{42}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{42}$		
1	0.00	0.09	0.12	/	0.00	0.93	1.26
2	0.16	0.05	0.04	/	1.59	0.58	0.48
3	-0.10	0.07	0.13	/	-1.05	0.74	1.33
4	-0.07	-0.12	0.24	/	-0.71	-1.18	2.37
5	-0.04	-0.19	-0.40	/	-0.42	-1.84	-3.88

$\hat{\phi}_{k\ell}^{43}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{43}$		
1	0.09	-0.25	0.01	/	0.92	-0.25	0.17
2	0.21	0.21	0.04	/	2.08	2.18	0.45
3	-0.01	0.21	-0.18	/	-0.18	2.05	-1.78
4	-0.04	0.06	-0.16	/	-0.44	0.66	-1.59
5	-0.05	0.25	-0.07	/	-0.51	2.40	-0.68

Table 7-5(b). (Cont'd)

	$\hat{\phi}^{44}_{k\ell}$				Standardized $\hat{\phi}^{44}_{k\ell}$		
	Space Lag						
Time Lag	0	1	2	/			
1	0.52	-0.15	-0.07	/	5.26	-1.59	-0.76
2	-0.31	-0.99	-0.31	/	-3.15	-0.98	-3.10
3	0.23	-0.03	0.08	/	2.29	-0.34	0.82
4	-0.03	-0.10	0.04	/	-0.33	-1.00	0.38
5	0.05	-0.14	-0.16	/	0.49	-1.38	-1.58

g	h	S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.
1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	1	2	.	.	.	-	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	-
g	h	S																				
1	2	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.
1	2	1	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.
1	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
1	3	0	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	3	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
1	4	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	4	1	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.
1	4	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	1	0	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	1	1	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.
2	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	2	0	+	.	+	+	.	+	.	.	.	.	.	.	.	.	-	.	.	-	-	.
2	2	1	.	+	.	.	.	+	.	+	.	.	.	+	.	.	.	.	.	.	.	.
2	2	2	+	+	.	.	+	.	.	.	.	.	.	.	.	.	.	-	.	-	-	-
g	h	S																				
2	3	0	+	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.
2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	3	2	.	.	.	+	.	.	.	+	+	+	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	4	0	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	-
2	4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.
2	4	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 7-12(a). The Encoded Sample Multivaraite Space-Time Autocorrelation Function of the Scaled Cleveland Crime Data

g	h	S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1	0	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.
3	1	1	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.
3	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
3	2	0	.	+	.	.	.	.	.	.	.	.	.	.	.	.	−	.	.	.	.	.
3	2	1	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	2	2	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
3	3	0	+	+	.	.	.	.	.	.	.	.	+	.	.	.	.	−	−	−	−	.
3	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	3	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
3	4	0	+	+	+	.	.	.	.	.	.	.	.	.	.	.	.	−	−	−	−	−
3	4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	4	2	.	.	.	.	−	−	.	.	.	.	.	+	.	.	.	.	.	.	.	.
g	h	S																				
4	1	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
4	2	0	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	2	1	.	.	−	.	.	.	−	−	−	.	.	−	.	.	.	.	.	.	.	.
4	2	2	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	−	.	.
g	h	S																				
4	3	0	+	.	.	.	.	−	.	.	.	.	÷	+	.	.	.	.	.	.	.	.
4	3	1	.	.	−	.	.	.	.	.	.	.	−	.	.	.	.	.	.	.	.	.
4	3	2	.	.	.	.	.	.	.	.	.	.	+	+	+	.	.	.	.	.	.	.
g	h	S																				
4	4	0	+	+	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	−	.	.
4	4	1	.	.	.	.	.	−	.	.	.	−	.	.	.	.	.	+	+	.	.	−
4	4	2	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.

Figure 7-12(a). (Cont'd)

g	h	S	1	2	3	4	5
1	1	0	.	.	.	.	.
1	1	1	.	.	+	.	.
1	1	2	.	+	.	.	−
g	h	S					
1	2	0	.	.	.	.	.
1	2	1	.	.	.	.	.
1	2	2	.	.	.	.	.
g	h	S					
1	3	0	.	−	.	.	.
1	3	1	.	.	.	.	.
1	3	2	+	.	.	.	+
g	h	S					
1	4	0	.	.	.	.	.
1	4	1	.	.	−	.	.
1	4	2	−	+	.	.	−
g	h	S					
2	1	0	.	.	.	.	.
2	1	1	.	.	.	.	.
2	1	2	.	.	.	.	.
g	h	S					
2	2	0	+	.	+	.	.
2	2	1	.	+	.	.	.
2	2	2	+	+	.	.	.
g	h	S					
2	3	0	.	.	−	.	.
2	3	1	.	.	.	.	.
2	3	2	+	−	[illegible]	.	−
g	h	S					
2	4	0	.	.	.	.	.
2	4	1	.	+	.	.	.
2	4	2	.	+	+	.	+

Figure 7-12(b). The Encoded Sample Multivariate Space-Time Partial Autocorrelation Function of the Scaled Cleveland Crime Data

g	h	S	1	2	3	4	5
3	1	0	.	.	.	.	.
3	1	1	.	.	.	.	.
3	1	2	.	+	.	+	.
g	h	S					
3	2	0	.	.	.	.	.
3	2	1	.	.	.	.	-
3	2	2	.	.	.	+	-
g	h	S					
3	3	0	+	.	.	.	.
3	3	1	.	.	.	.	+
3	3	2	.	-	.	.	.
g	h	S					
3	4	0	.	-	.	.	.
3	4	1	.	.	.	.	.
3	4	2	.	.	.	.	-
g	h	S					
4	1	0	.	.	.	.	.
4	1	1	.	.	.	.	-
4	1	2	.	.	.	+	.
g	h	S					
4	2	0	.	.	.	.	.
4	2	1	.	.	.	.	.
4	2	2	.	.	.	+	-
g	h	S					
4	3	0	.	+	.	.	.
4	3	1	.	+	+	.	+
4	3	2	.	.	.	.	.
g	h	S					
4	4	0	+	-	+	.	.
4	4	1	.	.	.	.	.
4	4	2	.	-	.	.	.

Figure 7-12(b). (Cont'd)

$\hat{\phi}^{22}_{k0}$, $\hat{\phi}^{22}_{k2}$, $\hat{\phi}^{33}_{k0}$ cuts off at k=1, $\hat{\phi}^{43}_{k0}$ cuts off at k=2, and $\hat{\phi}^{32}_{k0}$, $\hat{\phi}^{44}_{k0}$, $\hat{\phi}^{32}_{k2}$, $\hat{\phi}^{42}_{k0}$, $\hat{\phi}^{44}_{k2}$ tails off. This pattern suggests the MULSTARMA($4,\underset{\sim}{p},\underset{\sim}{q},\underset{\sim}{\lambda},\underset{\sim}{m}$) where

$$\underset{\sim}{p} = (0,0,0,0,0,1,0,0,0,1,1,0,0,0,2,1),$$

$$\underset{\sim}{q} = (0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,2),$$

$$\underset{\sim}{\lambda} = (2,0,0,0,0,0), \text{ and}$$

$$\underset{\sim}{m} = (0,0,0,0,0,1) \text{ or in difference equation form,}$$

$$\left\{\begin{aligned}
\underset{\sim}{Z}^1_t &= \underset{\sim}{a}^1_t \\
\underset{\sim}{Z}^2_t &= \phi^{22}_{10}\underset{\sim}{Z}^2_{t-1} + \phi^{22}_{12}w^{(2)}\underset{\sim}{Z}^2_{t-1} - \phi^{23}_{20}\underset{\sim}{\varepsilon}^3_{t-2} - \phi^{24}_{20}\underset{\sim}{\varepsilon}^4_{t-2} + \underset{\sim}{\varepsilon}^2_t \\
\underset{\sim}{Z}^3_t &= \phi^{32}_{10}\,\underset{\sim}{Z}^2_{t-1} + \phi^{33}_{10}\underset{\sim}{Z}^3_{t-1} + \underset{\sim}{\varepsilon}^3_t \\
\underset{\sim}{Z}^4_t &= \phi^{44}_{10}\underset{\sim}{Z}^4_{t-1} + \phi^{43}_{20}\underset{\sim}{Z}^3_{t-1} - \phi^{44}_{20}\underset{\sim}{\varepsilon}^4_{t-1} - \phi^{44}_{21}\underset{\sim}{\varepsilon}^4_{t-2} + \underset{\sim}{\varepsilon}^4_t.
\end{aligned}\right. \tag{7-94}$$

The conditional M.L. estimation gives the following results,

Estimate	Variance	t-Statistics
$\phi_{10}^{22} = 0.157$	0.0895	1.7542
$\phi_{12}^{22} = 0.329$	0.1101	2.9882
$\phi_{10}^{32} = 0.184$	0.0887	2.0744
$\phi_{10}^{33} = 0.442$	0.093	4.8948
$\phi_{10}^{44} = 0.507$	0.1060	4.7830
$\phi_{20}^{43} = 0.170$	0.1044	1.6284
$\phi_{20}^{23} = 0.200$	0.1175	1.7021
$\phi_{20}^{24} = 0.229$	0.1263	1.8131
$\phi_{20}^{44} = -0.205$	0.1241	1.6543
$\phi_{21}^{44} = -0.213$	0.1169	1.8212

where $\hat{\sigma}^2 = 0.7868$.

After the parameter estimates were obtained, the sample multivariate space-time autocorrelation functions and the sample multivariate space-time partial autocorrelation functions of the residuals were computed and encoded. The numerical values are listed in Table 7-6, and the encoded symbols are contained in Figure 7-13. No patterns are detected and this model is accepted as adequate.

7.9.2 Using the Model

Once the time series model of the historical data is accepted as adequate, it is ready to be employed. It can be used for forecasting or intervention analysis. New data entering the system may not

Table 7-6(a). The Sample Multivariate Space-Time Autocorrelation Functions for the Residuals of Cleveland Crime Data

$\hat{\rho}_{50}^{11}(k)$

	Space Lag				Standardized $\hat{\rho}_{50}^{11}(k)$		
Time Lag	0	1	2				
1	-0.14	0.03	0.07	/	-1.41	0.36	0.72
2	-0.11	0.03	0.10	/	-1.12	0.31	1.00
3	-0.04	0.17	0.09	/	-0.42	1.67	0.88
4	-0.09	0.08	-0.37	/	-0.93	0.80	-3.57
5	0.08	-0.12	0.16	/	0.75	-1.19	1.55
6	0.09	-0.01	0.05	/	0.87	-0.11	0.50
7	-0.13	-0.02	-0.06	/	-1.22	-0.20	-0.57
8	0.03	0.16	-0.00	/	0.35	1.52	-0.02
9	-0.10	0.16	0.00	/	-0.94	1.44	0.00
10	-0.08	-0.19	-0.11	/	-0.70	-1.67	-1.01
11	0.17	-0.17	0.25	/	1.47	-1.49	2.12
12	0.05	0.01	-0.16	/	0.45	0.15	-1.38
13	-0.12	0.14	-0.03	/	-1.05	1.15	-0.31
14	-0.60	0.24	0.16	/	-0.47	1.90	1.31
15	-1.31	0.14	-0.15	/	-1.01	1.11	-1.20
16	0.21	-0.19	0.18	/	1.64	-1.48	1.37
17	0.13	-0.19	0.04	/	1.02	-1.42	0.34
18	-0.35	-0.18	-0.01	/	-2.55	-1.32	-0.10
19	0.05	-0.12	-0.14	/	0.39	-0.88	-1.00
20	-0.11	0.00	-0.31	/	-0.73	0.05	-2.11

Table 7-6(a). (Cont'd)

$\hat{\rho}^{12}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{12}_{50}(k)$		
1	0.13	-0.04	0.47	/	1.38	-0.44	1.48
2	0.08	-0.04	-0.12	/	0.79	-0.46	-1.26
3	-0.01	-0.02	-0.05	/	-0.12	-0.19	-0.49
4	-0.17	0.07	0.10	/	-1.67	0.68	0.97
5	-0.11	-0.08	-0.00	/	-1.11	-0.07	-0.03
6	-0.42	-0.15	-0.18	/	-0.39	-1.39	-1.68
7	0.00	-0.06	0.02	/	0.06	-0.57	0.18
8	-0.06	-0.05	-0.03	/	-0.55	-0.46	-0.29
9	-0.08	0.09	-0.05	/	-0.74	0.85	-0.47
10	0.06	0.14	0.03	/	0.52	1.26	0.25
11	-0.07	-0.12	0.03	/	-0.64	-1.08	0.30
12	0.15	-0.21	-0.03	/	1.31	-1.81	-0.32
13	-0.07	-0.08	-0.07	/	-0.56	-0.72	-0.57
14	-0.16	-0.03	-0.10	/	-1.32	-0.29	-0.81
15	0.12	0.07	-0.04	/	0.98	0.60	-0.34
16	-0.27	0.08	-0.03	/	-2.07	0.66	-0.27
17	-0.03	-0.17	0.24	/	-0.27	-1.25	1.80
18	0.24	-0.02	-0.05	/	1.71	-0.19	-0.38
19	-0.06	-0.19	-0.25	/	-0.43	-1.37	-1.73
20	-0.27	0.01	0.13	/	-1.87	0.09	0.86

Table 7-6(a). (Cont'd)

$\hat{\rho}^{13}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{13}_{50}(k)$		
1	-0.10	0.04	0.03	/	-1.07	0.50	0.32
2	0.18	-0.01	0.14	/	1.81	-0.13	1.45
3	0.02	0.14	0.00	/	0.25	1.41	0.05
4	0.06	0.09	0.13	/	0.59	0.92	1.29
5	-0.02	-0.00	0.03	/	-0.25	-0.08	0.33
6	-0.21	-0.01	0.00	/	-1.97	-0.16	0.02
7	0.02	0.08	0.05	/	0.25	0.75	0.49
8	-0.04	-0.02	-0.28	/	-0.41	-0.18	-2.58
9	-0.08	-0.18	0.07	/	-0.75	-1.64	0.66
10	-0.14	-0.01	0.05	/	-1.28	-0.12	0.45
11	0.10	-0.10	-0.10	/	0.89	-0.91	-0.91
12	0.13	0.10	-0.03	/	1.08	0.88	-0.29
13	0.05	0.13	-0.07	/	0.44	1.05	-0.63
14	0.04	0.02	0.10	/	0.31	0.20	0.81
15	0.11	0.12	-0.13	/	0.87	0.98	-1.00
16	-0.06	-0.01	0.14	/	-0.49	-0.07	1.05
17	-0.01	-0.02	-0.03	/	-0.10	-0.19	-0.25
18	-0.00	0.10	0.07	/	-0.06	0.72	0.51
19	0.03	-0.00	-0.05	/	0.25	-0.05	-0.35
20	0.04	-0.12	-0.02	/	0.26	-0.84	-0.13

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{14}(k)$

Time Lag	Space Lag				Standardized $\hat{\rho}_{50}^{14}(k)$		
1	-0.06	0.08	-0.00	/	-0.69	0.83	-0.01
2	-0.00	0.01	0.10	/	-0.00	0.16	0.99
3	0.16	0.02	0.12	/	1.64	0.25	1.20
4	0.09	0.12	0.09	/	0.94	1.22	0.95
5	0.01	-0.00	-0.10	/	0.11	-0.00	-1.03
6	-0.22	0.03	0.18	/	-2.09	0.32	1.76
7	0.00	0.19	-0.00	/	0.01	1.78	-0.08
8	-0.02	-0.24	-0.25	/	-0.19	-2.18	-2.30
9	-0.03	-0.25	0.15	/	-0.30	-2.28	1.35
10	-0.04	-0.06	-0.14	/	-0.35	-0.59	-1.22
11	0.01	0.02	-0.05	/	0.16	0.20	-0.45
12	-0.00	0.19	-0.06	/	-0.03	1.62	-0.50
13	-0.09	-0.01	-0.08	/	-0.77	-0.14	-0.71
14	0.06	-0.06	0.13	/	0.47	-0.49	1.07
15	0.05	-0.04	-0.08	/	0.44	-0.34	-0.61
16	-0.08	0.06	-0.03	/	-0.60	0.47	-0.25
17	0.07	0.10	-0.11	/	0.56	0.79	-0.81
18	0.07	0.06	0.21	/	0.51	0.44	1.55
19	0.15	0.08	0.00	/	1.07	0.56	0.04
20	0.20	-0.09	0.08	/	1.38	-0.60	0.56

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{21}(k)$

	Space Lag				Standardized $\hat{\rho}_{50}^{21}(k)$		
Time Lag	0	1	2				
1	-0.08	0.13	-0.04	/	-0.80	1.35	-0.44
2	0.03	0.08	-0.02	/	0.33	0.83	-0.29
3	-0.03	-0.01	0.02	/	-0.31	-0.15	0.24
4	0.22	0.04	0.02	/	2.12	0.46	0.20
5	0.05	0.01	0.02	/	0.56	0.16	0.18
6	0.02	0.00	0.21	/	0.21	0.01	1.97
7	-0.02	0.17	-0.09	/	-0.19	1.59	-0.83
8	0.13	0.27	-0.05	/	1.22	2.44	-0.45
9	0.05	-0.16	0.07	/	0.44	-1.42	0.61
10	0.09	0.09	0.00	/	0.85	0.77	0.03
11	0.15	0.05	0.03	/	1.32	0.50	0.31
12	0.06	0.10	0.04	/	0.57	0.88	0.40
13	-0.00	-0.00	0.01	/	-0.04	-0.01	0.08
14	0.26	-0.02	0.07	/	2.11	-0.23	0.58
15	-0.11	-0.25	0.03	/	-0.91	-1.98	0.29
16	0.25	0.20	0.12	/	1.92	1.53	0.92
17	0.00	0.09	-0.07	/	0.01	0.72	-0.55
18	-0.14	0.07	-0.20	/	-1.01	0.51	-1.47
19	0.03	0.08	0.20	/	0.21	0.55	1.44
20	0.14	-0.03	0.06	/	0.94	-0.25	0.43

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{22}(k)$

	Space Lag				Standardized $\hat{\rho}_{50}^{22}(k)$		
Time Lag	0	1	2				
1	-0.01	0.04	-0.07	/	-0.12	0.40	-0.78
2	-0.10	0.24	0.10	/	-1.08	2.45	1.00
3	0.12	0.03	0.01	/	1.19	0.33	0.13
4	0.17	0.03	-0.04	/	1.72	0.35	-0.41
5	-0.03	0.08	0.09	/	-0.31	0.78	0.85
6	0.17	0.25	-0.02	/	1.64	2.40	-0.20
7	0.11	-0.12	0.01	/	1.02	-1.17	0.13
8	-0.14	0.05	0.10	/	-1.33	0.51	0.93
9	-0.11	0.11	0.04	/	-1.02	1.04	0.35
10	0.13	-0.08	0.02	/	1.20	-0.70	0.20
11	-0.07	-0.08	-0.00	/	-0.65	-0.71	-0.30
12	-0.13	0.24	-0.06	/	-1.12	2.01	-0.54
13	-0.12	0.05	0.00	/	-0.98	0.42	0.04
14	0.04	-0.25	0.14	/	0.38	-2.03	1.14
15	-0.17	-0.05	0.05	/	-1.34	-0.42	0.41
16	-0.14	-0.70	-0.21	/	-1.10	-0.52	-1.63
17	-0.00	-0.08	0.05	/	-0.06	-0.62	-0.38
18	-0.09	-0.11	-0.04	/	-0.63	-0.80	-0.34
19	-0.29	0.09	-0.12	/	-2.06	0.66	-0.86
20	-0.18	0.04	-0.08	/	-1.24	0.30	-0.56

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{23}(k)$

	Space Lag				Standardized $\hat{\rho}_{50}^{23}(k)$		
Time Lag	0	1	2				
1	-0.03	-0.06	-0.01	/	-0.35	-0.67	-0.17
2	0.09	-0.01	0.04	/	0.91	-0.10	0.47
3	-0.07	0.05	0.09	/	-0.73	0.52	0.89
4	0.21	0.04	0.22	/	2.04	0.45	2.20
5	0.07	-0.01	-0.01	/	0.70	-0.11	-0.13
6	-0.10	-0.04	-0.02	/	-0.94	-0.42	-0.20
7	-0.04	0.12	0.03	/	-0.36	1.11	0.30
8	-0.07	-0.09	0.11	/	-0.63	-0.81	1.03
9	0.01	0.04	0.22	/	0.08	0.43	1.98
10	-0.02	0.08	0.08	/	-0.22	0.75	0.69
11	-0.04	0.06	0.14	/	-0.40	0.58	1.23
12	-0.14	0.02	-0.14	/	-1.20	0.20	-1.18
13	-0.13	-0.02	0.15	/	-1.08	-0.16	1.21
14	-0.02	-0.10	0.21	/	-0.20	-0.82	1.69
15	-0.10	0.17	0.13	/	-0.78	1.34	1.01
16	0.11	-0.04	-0.04	/	0.85	-0.30	-0.02
17	-0.13	-0.12	-0.08	/	-1.01	-0.90	-0.61
18	-0.01	-0.11	0.04	/	-0.09	-0.82	0.31
19	0.07	0.06	0.03	/	0.49	0.47	0.24
20	-0.07	0.17	0.04	/	-0.51	1.18	0.31

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{24}(k)$

	Space Lag				Standardized $\hat{\rho}_{50}^{24}(k)$		
Time Lag	0	1	2				
1	-0.11	-0.04	0.02	/	-1.16	-0.40	0.27
2	0.16	0.01	0.05	/	1.66	0.18	0.57
3	-0.01	0.02	0.26	/	-0.14	0.24	2.61
4	-0.07	0.04	0.09	/	-0.71	0.39	0.87
5	-0.07	-0.05	-0.11	/	-0.67	-0.49	-1.10
6	0.08	0.07	0.06	/	0.81	0.65	0.59
7	0.02	0.08	0.01	/	0.24	0.75	0.12
8	-0.15	-0.00	0.10	/	-1.37	-0.01	0.93
9	-0.00	0.24	0.15	/	-0.07	2.12	1.32
10	-0.15	0.05	0.05	/	-1.31	0.43	0.49
11	-0.00	0.10	0.10	/	-0.00	0.91	0.92
12	-0.01	0.02	-0.02	/	-0.14	0.18	-0.18
13	-0.13	-0.07	0.09	/	-1.12	0.60	0.79
14	-0.00	-0.04	0.10	/	-0.02	-0.38	0.79
15	-0.01	0.10	0.07	/	-0.12	0.81	0.58
16	-0.10	-0.05	-0.07	/	-0.77	-0.43	-0.56
17	-0.09	-0.35	-0.03	/	-0.67	-2.62	-0.28
18	0.05	0.12	0.03	/	0.38	-0.87	0.21
19	-0.09	0.00	-0.07	/	-0.63	0.00	-0.48
20	-0.14	-0.04	0.16	/	-0.95	-0.29	1.10

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{31}(k)$

	Space Lag				Standardized $\hat{\rho}_{50}^{31}(k)$		
Time Lag	0	1	2				
1	0.05	0.09	0.12	/	0.53	0.98	1.23
2	-0.18	0.01	-0.02	/	-1.81	0.15	-0.24
3	-0.11	-0.06	0.06	/	-1.11	-0.60	0.64
4	0.16	0.01	-0.17	/	1.60	0.18	-1.71
5	-0.11	-0.09	-0.01	/	-1.08	-0.89	-0.09
6	-0.08	-0.13	-0.02	/	-0.81	-1.27	-0.20
7	-0.06	0.13	0.20	/	-0.63	1.22	1.87
8	0.08	-0.04	-0.07	/	0.71	-0.43	-0.69
9	0.23	-0.04	0.16	/	2.10	-0.35	1.42
10	0.14	0.03	-0.00	/	1.27	0.30	-0.05
11	-0.10	-0.24	0.05	/	-0.85	-2.06	0.48
12	0.30	0.32	0.03	/	2.54	2.72	0.31
13	-0.19	0.15	0.03	/	-1.55	1.28	0.26
14	0.06	-0.09	0.00	/	0.47	-0.73	0.06
15	0.03	0.08	-0.18	/	0.27	0.62	-1.40
16	-0.05	-0.08	-0.14	/	-0.43	-0.61	-1.07
17	-0.12	0.05	0.00	/	-0.93	0.38	0.01
18	-0.10	0.13	-0.06	/	-0.72	0.99	-0.45
19	-0.08	-0.12	0.12	/	-0.62	-0.87	0.83
20	0.14	0.02	-0.15	/	0.94	0.16	-1.05

Table 7-6(a). (Cont'd)

$\hat{\rho}^{32}_{50}(k)$

Time Lag	Space Lag 0	Space Lag 1	Space Lag 2	/	Standardized $\hat{\rho}^{32}_{50}(k)$		
1	0.01	0.07	0.07	/	0.14	0.75	0.70
2	0.00	0.04	-0.05	/	0.00	0.44	-0.56
3	-0.15	0.01	0.02	/	-1.53	0.10	0.20
4	0.04	-0.02	-0.10	/	0.41	-0.21	-0.97
5	0.07	-0.01	0.07	/	0.72	-0.16	0.68
6	0.09	-0.02	-0.03	/	0.87	-0.23	-0.33
7	-0.05	0.05	-0.11	/	-0.53	0.54	-1.04
8	0.21	0.20	0.10	/	1.89	1.84	0.89
9	-0.02	-0.12	0.04	/	-0.21	-1.11	0.42
10	0.21	-0.10	-0.01	/	1.87	-0.94	-0.14
11	0.05	0.15	-0.14	/	0.44	1.34	-1.18
12	-0.04	-0.08	-0.02	/	-0.37	-0.66	-0.16
13	-0.01	-0.06	0.04	/	-0.14	-0.54	0.39
14	-0.06	0.21	-0.24	/	-0.50	1.71	-1.94
15	-0.17	-0.13	0.06	/	-1.34	-1.01	0.49
16	-0.07	-0.12	-0.22	/	-0.53	-0.97	-1.71
17	-0.18	-0.11	-0.00	/	-1.35	-0.80	-0.06
18	0.13	0.25	-0.00	/	0.94	1.83	-0.06
19	-0.03	-0.10	0.06	/	-0.23	-0.70	0.41
20	0.06	-0.06	0.09	/	0.41	-0.44	0.61

Table 7-6(a). (Cont'd)

$\hat{\rho}^{33}_{50}(k)$

Time Lag	Space Lag 0	Space Lag 1	Space Lag 2	/	Standardized $\hat{\rho}^{33}_{50}(k)$		
1	-0.04	0.11	0.08	/	-0.45	1.14	0.80
2	0.07	-0.10	-0.03	/	0.76	-1.08	-0.38
3	-0.03	0.09	-0.06	/	-0.36	0.90	-0.60
4	0.05	-0.01	-0.11	/	0.56	-0.15	-1.06
5	-0.04	0.08	-0.06	/	-0.42	0.75	-0.64
6	-0.21	-0.14	-0.20	/	-1.96	-1.31	-1.88
7	-0.16	0.03	-0.02	/	-1.55	0.34	-0.24
8	-0.07	-0.14	-0.10	/	-0.68	-1.31	-0.91
9	-0.08	-0.12	0.04	/	-0.70	-1.13	0.39
10	0.13	-0.09	0.06	/	1.12	-0.83	0.57
11	0.17	0.20	0.11	/	1.49	1.71	0.96
12	0.25	0.03	0.19	/	2.09	0.30	1.60
13	-0.09	-0.11	0.03	/	-0.74	-0.89	0.31
14	0.04	-0.03	0.14	/	0.34	-0.25	1.15
15	-0.09	0.05	-0.03	/	-0.74	0.43	-0.25
16	-0.01	0.16	-0.09	/	-0.10	1.23	-0.67
17	-0.24	0.12	-0.07	/	-1.77	0.91	-0.56
18	-0.09	-0.12	-0.12	/	-0.70	-0.88	-0.88
19	-0.33	0.06	0.12	/	-2.29	0.47	0.83
20	0.04	-0.16	-0.01	/	0.31	-1.12	-0.08

Table 7-6(a). (Cont'd)

$\hat{\rho}^{34}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{34}_{50}(k)$		
1	0.03	-0.11	0.03	/	0.39	-1.12	0.34
2	0.10	0.05	0.01	/	1.00	0.57	0.18
3	0.15	0.23	-0.05	/	1.47	2.25	-0.57
4	0.03	0.09	-0.07	/	0.33	0.94	-0.76
5	-0.08	0.17	-0.22	/	-0.81	1.63	-2.13
6	-0.22	-0.22	-0.23	/	-2.10	-2.12	-2.15
7	0.13	0.15	0.05	/	1.22	1.40	0.51
8	-0.02	-0.05	-0.10	/	-0.21	-0.53	-0.93
9	-0.00	-0.18	-0.01	/	-0.04	-1.60	-0.08
10	0.03	-0.00	0.04	/	0.29	-0.07	0.42
11	0.13	0.08	0.15	/	1.14	0.73	1.33
12	0.08	-0.07	0.19	/	0.72	-0.64	1.63
13	-0.08	-0.21	-0.03	/	-0.67	-1.78	-0.29
14	0.07	0.13	0.13	/	0.62	1.04	1.10
15	-0.16	0.06	-0.02	/	-1.27	0.47	-0.19
16	-0.03	0.15	0.05	/	-0.29	1.15	0.40
17	-0.22	0.06	-0.29	/	-1.63	0.50	-2.15
18	-0.24	-0.11	-0.07	/	-1.76	-0.83	-0.49
19	-0.17	0.10	0.16	/	-1.23	0.71	1.11
20	-0.00	-0.12	-0.08	/	-0.02	-0.80	-0.58

Table 7-6(a). (Cont'd)

$\hat{\rho}^{41}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{41}_{50}(k)$		
1	0.03	-0.05	-0.08	/	0.36	-0.52	-0.81
2	-0.06	-0.00	0.09	/	-0.60	-0.00	0.91
3	-0.08	-0.16	-0.04	/	-0.85	-1.59	-0.43
4	-0.07	0.14	-0.10	/	-0.72	1.43	-1.00
5	0.06	0.14	0.01	/	0.58	0.37	0.14
6	-0.11	-0.00	-0.13	/	-1.05	-0.02	-1.23
7	-0.02	0.06	0.08	/	-0.26	0.61	0.79
8	0.06	-0.14	0.07	/	0.59	-1.30	0.65
9	0.10	-0.11	0.02	/	0.92	-0.97	0.23
10	0.04	0.14	0.24	/	0.34	1.27	2.12
11	-0.13	-0.11	0.08	/	-1.12	-0.93	0.75
12	0.17	0.04	-0.09	/	1.45	0.35	-0.75
13	-0.04	0.19	0.13	/	-0.36	1.54	1.08
14	-0.02	-0.01	0.13	/	-0.15	-0.09	1.05
15	-0.19	-0.23	-0.22	/	-1.50	-1.81	-1.75
16	0.04	-0.01	-0.02	/	0.35	-0.11	-0.16
17	-0.19	-0.17	-0.05	/	-1.41	-1.25	-0.40
18	-0.04	-0.02	-0.19	/	-0.32	-0.15	-1.36
19	-0.01	0.05	0.22	/	-0.12	0.36	1.58
20	-0.02	-0.29	-0.22	/	-0.15	-1.97	-1.53

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{42}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}_{50}^{42}(k)$		
1	-0.06	-0.00	-0.04	/	-0.63	-0.09	-0.46
2	0.00	-0.09	-0.07	/	0.01	-0.95	-0.75
3	-0.02	-0.17	0.16	/	-0.21	-1.69	1.62
4	0.04	-0.03	-0.01	/	0.40	-0.31	-0.15
5	0.08	0.09	-0.03	/	0.79	0.94	-0.36
6	-0.08	0.00	0.10	/	-0.74	0.07	0.98
7	-0.06	-0.14	-0.09	/	-0.62	-1.33	-0.88
8	0.16	0.01	0.12	/	1.51	0.12	1.08
9	0.06	-0.16	0.03	/	0.60	-1.46	0.27
10	-0.02	-0.15	-0.05	/	-0.20	-1.36	-0.49
11	0.24	0.11	0.06	/	2.09	0.99	0.56
12	-0.03	-0.44	-0.07	/	-0.32	-3.66	-0.60
13	0.06	-0.05	-0.05	/	0.52	-0.45	-0.40
14	-0.10	0.10	-0.26	/	-0.83	0.81	-2.12
15	0.07	0.19	-0.02	/	0.59	1.53	-0.18
16	-0.23	-0.34	-0.14	/	-1.76	-2.63	-1.07
17	0.14	0.09	0.05	/	1.03	0.66	0.36
18	0.17	0.11	-0.15	/	1.21	0.80	-1.07
19	-0.05	-0.00	-0.12	/	-0.35	-0.00	-0.84
20	-0.06	0.22	0.10	/	-0.44	1.50	0.67

Table 7-6(a). (Cont'd)

$\hat{\rho}_{50}^{43}(k)$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\rho}_{50}^{43}(k)$		
1	0.11	0.07	0.04	/	1.12	0.70	0.48
2	-0.03	-0.21	0.07	/	-0.35	-2.17	0.73
3	0.01	-0.13	-0.15	/	0.17	-1.30	-1.53
4	-0.03	-0.03	-0.09	/	-0.36	-0.36	-0.89
5	-0.13	-0.04	-0.22	/	-1.29	-0.44	-2.09
6	-0.23	-0.11	-0.14	/	-0.20	-1.06	-1.31
7	-0.02	-0.07	-0.08	/	-0.19	-0.67	-0.80
8	0.00	0.08	-0.10	/	0.05	0.76	-0.93
9	0.17	0.14	0.01	/	1.52	1.31	0.08
10	-0.01	-0.00	-0.03	/	-0.13	-0.04	-0.30
11	0.13	-0.11	0.20	/	1.14	-1.01	1.73
12	0.31	-0.05	0.19	/	2.62	-0.47	1.63
13	0.14	0.04	0.28	/	1.14	0.39	2.27
14	0.06	-0.10	0.08	/	0.47	-0.80	0.63
15	-0.14	0.08	-0.00	/	-1.12	0.63	-0.05
16	-0.08	-0.09	-0.04	/	-0.60	-0.67	-0.34
17	0.20	-0.02	-0.13	/	1.48	-0.20	-1.01
18	-0.09	0.03	-0.06	/	-0.69	-0.25	-0.47
19	-0.32	-0.10	-0.07	/	-2.22	-0.70	-0.51
20	0.01	0.18	0.01	/	0.13	1.22	0.09

Table 7-6(a). (Cont'd)

$\hat{\rho}^{44}_{50}(k)$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\rho}^{44}_{50}(k)$		
1	0.11	-0.14	0.01	/	1.18	-1.45	0.15
2	0.00	0.01	-0.02	/	0.06	0.13	-0.23
3	0.10	0.02	-0.09	/	1.01	0.23	-0.94
4	-0.00	-0.03	-0.11	/	-0.02	-0.31	-1.11
5	-0.03	-0.11	-0.16	/	-0.36	-1.08	-1.51
6	-0.20	-0.20	-0.15	/	-1.92	-1.91	-1.41
7	-0.01	0.10	-0.11	/	-0.14	0.99	-1.01
8	0.05	0.08	-0.14	/	0.51	0.77	-1.28
9	0.11	-0.07	-0.04	/	1.03	-0.69	-0.40
10	-0.06	-0.16	-0.11	/	-0.58	-1.45	-0.94
11	0.06	0.00	0.14	/	0.52	0.06	1.24
12	0.04	-0.03	0.23	/	0.34	-0.29	1.95
13	-0.04	-0.11	0.13	/	-0.33	-0.95	1.12
14	-0.01	-0.03	0.14	/	-0.11	-0.30	1.17
15	-0.16	0.01	-0.00	/	-1.28	0.11	-0.03
16	-0.08	0.22	-0.10	/	-0.66	1.67	-0.77
17	0.15	0.29	-0.10	/	1.12	2.15	-0.79
18	-0.26	-0.10	0.02	/	-1.86	-0.77	0.17
19	-0.16	0.02	-0.05	/	-1.14	0.14	-0.40
20	0.13	0.29	0.02	/	0.88	1.98	0.15

Table 7-6(b). The Sample Multivariate Space-Time Partial Autocorrelation Function for the Residuals of Cleveland Crime Data

$\hat{\phi}^{11}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{11}_{k\ell}$		
1	-0.16	0.03	0.09	/	-1.66	0.33	0.92
2	-0.08	0.02	0.27	/	-0.82	0.24	2.70
3	-0.14	0.23	0.15	/	-1.39	2.34	1.55
4	-0.05	0.01	-0.12	/	-0.50	0.17	-1.18
5	0.01	0.01	-0.27	/	0.09	0.10	-2.38

$\hat{\phi}^{12}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{12}_{k\ell}$		
1	-0.12	0.13	-0.05	/	-1.21	1.40	-0.54
2	0.08	0.11	-0.00	/	0.85	1.17	-0.04
3	-0.10	0.11	0.03	/	-1.03	1.12	0.34
4	0.16	-0.01	-0.01	/	1.63	-0.17	-0.15
5	-0.00	0.11	-0.18	/	-0.04	1.07	-1.78

$\hat{\phi}^{13}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{13}_{k\ell}$		
1	0.09	0.10	0.35	/	0.95	1.09	3.58
2	-0.26	0.00	-0.13	/	-2.65	0.00	-1.34
3	-0.14	-0.04	0.08	/	-1.43	-0.46	0.84
4	0.04	0.11	-0.18	/	0.39	1.08	-1.81
5	0.01	-0.17	0.59	/	0.10	-1.62	5.60

Table 7-6(b). (Cont'd)

$\hat{\phi}^{14}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{14}_{k\ell}$		
1	0.00	-0.11	-0.38	/	0.06	-1.19	-3.91
2	0.03	0.01	0.18	/	0.35	0.15	1.86
3	-0.03	-0.14	-0.09	/	-0.32	-1.39	-0.93
4	-0.08	0.03	0.11	/	-0.82	0.29	1.14
5	0.10	0.12	-0.35	/	0.99	1.15	-3.34

$\hat{\phi}^{21}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{21}_{k\ell}$		
1	0.12	-0.04	0.13	/	1.25	-0.41	1.38
2	0.09	-0.00	-0.05	/	0.90	-0.07	-0.54
3	0.06	-0.06	-0.06	/	0.63	-0.67	-0.67
4	-0.06	0.01	0.12	/	-0.64	0.16	1.15
5	-0.04	-0.01	0.26	/	-0.42	-0.10	2.50

$\hat{\phi}^{22}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{22}_{k\ell}$		
1	-0.00	0.03	-0.12	/	-0.06	0.37	-1.24
2	-0.06	0.29	0.15	/	-0.67	2.94	1.53
3	0.23	-0.04	0.02	/	2.34	-0.47	0.25
4	0.07	0.03	0.08	/	0.73	0.31	0.82
5	-0.05	-0.01	0.24	/	-0.53	-0.10	2.29

Table 7-6(b). (Cont'd)

$\hat{\phi}^{23}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{23}_{k\ell}$		
1	0.04	0.07	0.23	/	0.43	0.80	2.33
2	-0.02	-0.02	-0.07	/	-0.22	-0.27	-0.76
3	-0.22	0.13	-0.10	/	-2.23	1.20	-1.01
4	-0.00	-0.08	-0.17	/	-0.00	-0.79	-1.68
5	-0.03	-0.03	-0.15	/	-0.36	-0.28	-1.47

$\hat{\phi}^{24}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{24}_{k\ell}$		
1	-0.10	0.00	-0.24	/	-1.09	0.08	-2.46
2	-0.00	-0.06	0.03	/	-0.05	-0.65	0.30
3	0.07	-0.16	0.25	/	0.74	-1.62	2.52
4	0.10	-0.03	0.04	/	0.99	-0.33	0.44
5	0.11	-0.00	0.14	/	1.07	-0.04	1.33

$\hat{\phi}^{31}_{k\ell}$

Time Lag	Space Lag 0	1	2	/	Standardized $\hat{\phi}^{31}_{k\ell}$		
1	-0.09	0.03	0.01	/	-0.97	0.37	0.17
2	0.16	0.02	0.17	/	1.64	0.24	1.76
3	0.02	0.13	0.04	/	0.28	1.34	0.39
4	0.02	0.09	0.22	/	0.19	0.92	2.17
5	0.00	-0.06	0.12	/	0.01	-0.58	1.19

Table 7-6(b). (Cont'd)

$\hat{\phi}_{k\ell}^{32}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{32}$		
1	-0.02	-0.07	-0.01	/	-0.26	-0.79	-0.15
2	0.11	-0.01	0.04	/	1.11	-0.11	0.47
3	-0.08	0.04	0.08	/	-0.81	0.42	0.84
4	0.16	0.01	0.22	/	1.56	0.10	2.12
5	0.07	-0.18	0.33	/	0.72	-1.73	3.21

$\hat{\phi}_{k\ell}^{33}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{33}$		
1	-0.10	0.12	0.11	/	-1.07	1.24	1.13
2	0.07	-0.02	-0.11	/	0.78	-0.25	-1.17
3	-0.12	0.10	-0.22	/	-1.22	1.03	-2.21
4	-0.00	-0.03	-0.20	/	-0.07	-0.35	-1.95
5	-0.04	0.14	-0.07	/	-0.45	1.32	-0.71

$\hat{\phi}_{k\ell}^{34}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{34}$		
1	0.18	-0.00	-0.06	/	1.88	-0.08	-0.62
2	-0.13	-0.24	0.18	/	-1.28	-2.40	1.86
3	0.07	-0.20	0.08	/	0.71	-2.02	0.83
4	-0.03	-0.14	0.00	/	-0.33	-1.38	0.08
5	-0.02	-0.12	-0.05	/	-0.22	-1.21	-0.55

Table 7-6(b). (Cont'd)

$\hat{\phi}_{k\ell}^{41}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{41}$		
1	-0.07	0.07	0.00	/	-0.74	0.79	0.03
2	0.02	-0.00	0.15	/	0.27	-0.09	1.57
3	0.08	0.06	0.06	/	0.84	0.65	0.61
4	0.02	0.08	0.20	/	0.20	0.81	2.00
5	0.00	-0.04	0.07	/	0.01	-0.42	0.66

$\hat{\phi}_{k\ell}^{42}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{42}$		
1	-0.10	-0.03	0.05	/	-1.06	-0.37	0.58
2	0.12	0.00	-0.02	/	1.19	0.02	-0.19
3	-0.04	0.10	0.19	/	-0.41	1.01	1.91
4	0.02	-0.05	0.10	/	0.22	-0.53	1.01
5	-0.03	-0.18	0.19	/	-0.29	-1.77	1.82

$\hat{\phi}_{k\ell}^{43}$

Time Lag	Space Lag 0	1	2		Standardized $\hat{\phi}_{k\ell}^{43}$		
1	0.00	-0.08	0.03	/	0.09	-0.90	0.31
2	0.08	0.06	0.07	/	0.81	0.60	0.72
3	0.09	0.24	-0.13	/	0.90	2.43	-1.28
4	-0.04	0.07	-0.27	/	-0.39	0.68	-2.67
5	-0.04	0.20	-0.12	/	-0.45	1.94	-1.22

Table 7-6(b). (Cont'd)

$\hat{\phi}^{44}_{k\ell}$

	Space Lag				Standardized $\hat{\phi}^{44}_{k\ell}$		
Time Lag	0	1	2				
1	0.11	-0.17	0.03	/	1.13	-1.77	0.32
2	-0.06	0.01	-0.20	/	-0.67	0.10	-2.00
3	0.04	-0.10	0.01	/	0.44	-1.04	0.12
4	-0.00	-0.08	0.13	/	-0.02	-0.84	1.26
5	0.06	-0.10	0.02	/	0.61	-0.98	0.21

g	h	S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.
1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	1	2	.	.	.	-	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	-
g	h	S																				
1	2	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.
1	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
1	3	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	3	2	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
1	4	0	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	4	1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.
1	4	2	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	1	0	.	.	.	+	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.
2	1	1	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.
2	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	2	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.
2	2	1	.	+	.	.	.	+	.	.	.	.	.	+	.	.	.	.	.	.	.	.
2	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	3	0	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	3	2	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
2	4	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	4	1	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	-	.	.	.
2	4	2	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 7-13(a). The Encoded Sample Multivariate Space-Time Autocorrelation Function of the Residuals from the MULSTARIMA Model of the Cleveland Crime Data

g	h	S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1	0	.	.	.	.	.	.	.	.	+	.	.	+	.	.	.	.	.	.	.	.
3	1	1	.	.	.	.	.	.	.	.	.	.	−	+	.	.	.	.	.	.	.	.
3	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
3	2	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
3	3	0	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	−	.
3	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	3	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
g	h	S																				
3	4	0	.	.	.	.	.	−	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	4	1	.	.	+	.	.	−	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	4	2	.	.	.	.	−	−	.	.	.	.	.	.	.	.	.	.	−	.	.	.
g	h	S																				
4	1	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	1	2	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.
g	h	S																				
4	2	0	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.
4	2	1	.	.	.	.	.	.	.	.	.	.	.	−	.	.	.	−	.	.	.	.
4	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	−	.	.	.	.	.	.
g	h	S																				
4	3	0	.	.	.	.	.	−	.	.	.	.	.	+	.	.	.	.	.	.	−	.
4	3	1	.	−	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	3	2	.	.	.	.	−	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.
g	h	S																				
4	4	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.
4	4	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 7-13(a). (Cont'd)

g	h	S	1	2	3	4	5
1	1	0	.	.	.	.	.
1	1	1	.	.	+	.	.
1	1	2	.	+	.	.	−
g	h	S					
1	2	0	.	.	.	.	.
1	2	1	.	.	.	.	.
1	2	2	.	.	.	.	.
g	h	S					
1	3	0	.	−	.	.	.
1	3	1	.	.	.	.	.
1	3	2	+	.	.	.	+
g	h	S					
1	4	0	.	.	.	.	.
1	4	1	.	.	.	.	.
1	4	2	−	.	.	.	−
g	h	S					
2	1	0	.	.	.	.	.
2	1	1	.	.	.	.	.
2	1	2	.	.	.	.	+
g	h	S					
2	2	0	.	.	+	.	.
2	2	1	.	+	.	.	.
2	2	2	.	.	.	.	+
g	h	S					
2	3	0	.	.	−	.	.
2	3	1	.	.	.	.	.
2	3	2	+	.	.	.	.
g	h	S					
2	4	0	.	.	.	.	.
2	4	1	.	.	.	.	.
2	4	2	−	.	+	.	.

Figure 7-13(b). The Encoded Sample Multivariate Space-Time Partial Autocorrelation Function of the Residuals from the MULSTARMA Model of the Cleveland Crime Data

g	h	S	1	2	3	4	5
3	1	0	.	.	.	.	.
3	1	1	.	.	.	.	.
3	1	2	.	.	.	+	.
g	h	S					
3	2	0	.	.	.	.	.
3	2	1	.	.	.	.	.
3	2	2	.	.	.	+	+
g	h	S					
3	3	0	.	.	.	.	.
3	3	1	.	.	.	.	.
3	3	2	.	.	-	.	.
g	h	S					
3	4	0	.	.	.	.	.
3	4	1	.	-	-	.	.
3	4	2	.	.	.	.	.
g	h	S					
4	1	0	.	.	.	.	.
4	1	1	.	.	.	.	.
4	1	2	.	.	.	+	.
g	h	S					
4	2	0	.	.	.	.	.
4	2	1	.	.	.	.	.
4	2	2	.	.	.	.	.
g	h	S					
4	3	0	.	.	.	.	.
4	3	1	.	.	+	.	.
4	3	2	.	.	.	-	-
g	h	S					
4	4	0	.	.	.	.	.
4	4	1	.	.	.	.	.
4	4	2	.	-	.	.	.

Figure 7-13(b). (Cont'd)

compare favorably to forecasted values by chance itself or because the process has been changed after the forecast model has been developed (e.g. by some intervention program). In the former situation, the process is not changed and still can be adequately described by the employed model, while in the latter situation, the process has been changed and the forecasting or historically based model has to be modified to accomodate the change in the process. In the following, from the model, Equation (7-94), the forecasting function is derived in Section 7.9.2.1. The ℓ-step ahead forecasts at T=35 for ℓ=1,2,...,22, the 1-step ahead forecasts at T=35,36,...,56 and their associated 95% confidence intervals are computed and plotted with the corresponding new observed values for each crime and each location. In Section 7.9.2.2, the model, Equation (7-94) is augmented to allow formal intervention analysis in order to evaluate the effect of CCP which was initiated at T=40.

<u>7.9.2.1 Forecasting</u>. Given the general multivariate ARMA model

$$\underset{\sim}{Z}^h(t) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} B^{hg}(k)\underset{\sim}{Z}^g(t-k) - \sum_{g=1}^{\zeta} \sum_{k=1}^{q^{hg}} A^{hg}(k)\underset{\sim}{\varepsilon}^g(t-k) + \underset{\sim}{\varepsilon}^h(t)$$

where

$$h=1,2,\ldots,\zeta.$$

The ℓ-step ahead point forecast of category h at time T is the conditional expectation taken at time T.

$$\hat{Z}^h_T(\ell) = E(Z^h(T+\ell)\,|\,Z(T), Z(T-1), \ldots;\ \varepsilon(T), \varepsilon(T-1), \ldots).$$

So

$$\hat{Z}^h_T(\ell) = \sum_{g=1}^{\zeta} \sum_{k=1}^{p^{hg}} B^{hg}(k)\hat{Z}^g_T(T+\ell-k) - \sum_{g=1}^{\zeta} \sum_{k=1}^{q^{hg}} A^{hg}(k)\hat{\varepsilon}^g(T+\ -k) \qquad (7\text{-}95)$$

where

$\hat{Z}^h_T(T+\ell-k) = Z^h(T+\ell-k)$ is the realized observation if $\ell \le k$,

$\hat{\varepsilon}^g(T+\ell-k)$ is the estimated noise if $\ell \le k$, and

$\hat{\varepsilon}^g(T+\ell-k) = 0$ if $\ell > k$.

To obtain the -step ahead interval forecast, we need to compute $Var(\hat{Z}^h_T(\ell))$. The general multivariate ARMA model can be expressed in Λ-weight representation,

$$\hat{Z}^h_T(\ell) = \sum_{g=1}^{\zeta} \sum_{k=1}^{\infty} \Lambda^{hg}(k)\varepsilon^g(T+\ell-k) + \varepsilon^h(T+\ell), \qquad (7\text{-}96)$$

where

$\Lambda^{hg}(k)$ can be expressed in terms of $B^{hg}(k)$, $A^{hg}(t)$ matrices, the equations that describe this relationship are given in Section 7-3.

Since the variance of the realized values are zeros,

$$Var(\hat{Z}^h_T(\ell)) = G^{hh} + \sum_{k=1}^{\ell-1} \left(\sum_{g=1}^{\zeta} \Lambda^{hg}(k)\right) G \left(\sum_{g=1}^{\zeta} \Lambda^{hg}(k)\right)' \qquad (7\text{-}97)$$

Using the point forecasts and the variance estimate the ℓ-step ahead $100(1-\alpha)\%$ forecast interval for a category h observation is,

$$\hat{Z}^h_T(\ell) \pm Z_{\alpha/2}\sqrt{Var(Z^h_T(\ell))}\ , \qquad (7\text{-}98)$$

where $Z_{\alpha/2}$ is the $\alpha/2$ percentage point of a unit normal distribution (e.g., 1.96 for $\alpha = 0.05$).

By setting

$$B^{hg}(k) = \sum_{\ell=0}^{\lambda^{hg}_k} \phi^{hg}_k w(\ell),$$

$$A^{hg}(k) = \sum_{\ell=0}^{m^{hg}_k} \theta^{hg}_k w(\ell)$$

in Equation (7-95) and (7-98), we obtain the specific forecast functions for the MULSTARMA model. For example, employing the model of the Cleveland CCP Data to build the forecast function, we set,

$$
\begin{aligned}
B^{22}(1) &= \hat{\phi}^{22}_{10} I + \hat{\phi}^{22}_{12} W^{(2)}, \\
B^{32}(1) &= \hat{\phi}^{32}_{10} I, \\
B^{33}(1) &= \hat{\phi}^{33}_{10} I, \\
B^{44}(1) &= \hat{\phi}^{44}_{10} I, \\
B^{43}(2) &= \hat{\phi}^{43}_{20} I, \\
B^{23}(2) &= \hat{\phi}^{23}_{20} I, \\
A^{24}(2) &= \hat{\phi}^{24}_{20} I, \\
A^{44}(2) &= \hat{\theta}^{44}_{20} I + \hat{\theta}^{44}_{21} W^{(1)},
\end{aligned}
$$

wne have the following recursive ℓ-step ahead point forecast function,

$$
\left\{
\begin{aligned}
\hat{\underset{\sim}{Z}}^1_T(\ell) &= \underset{\sim}{0} \\
\hat{\underset{\sim}{Z}}^2_T(\ell) &= (\hat{\phi}^{22}_{10} + \hat{\phi}^{22}_{12} W^{(2)})\hat{\underset{\sim}{Z}}^2_T(\ell-1) - \hat{\theta}^{23}_{20}\hat{\underset{\sim}{\varepsilon}}^3_T(\ell-2) - \hat{\theta}^{24}_{20}\hat{\underset{\sim}{\varepsilon}}^4_T(\ell-2) \\
\hat{\underset{\sim}{Z}}^3_T(\ell) &= \hat{\underset{\sim}{\phi}}^{32}_{10}\underset{\sim}{Z}^2_T(\ell-1) + \hat{\phi}^{33}_{10}\underset{\sim}{Z}^3_T(\ell-1) \\
\hat{\underset{\sim}{Z}}^4_T(\ell) &= \hat{\phi}^{44}_{10}\hat{\underset{\sim}{Z}}^4_T(\ell-1) + \hat{\phi}^{43}_{20}\hat{\underset{\sim}{Z}}^3_T(\ell-2) - \hat{\theta}^{44}_{20}\hat{\underset{\sim}{\varepsilon}}^4_T(\ell-2) - \hat{\theta}^{44}_{21}\hat{\underset{\sim}{\varepsilon}}^4_T(\ -2)
\end{aligned}
\right. \qquad (7\text{-}99)
$$

In the forecast function, Equation (7-99), it is seen that: the category 1 crime, murder, doesn't have any correlated relationship

with the other crimes; the category 2 crime, rape, is influenced by the previous rape occurrences in the same location and the neighboring regions and the category 3 crime, robber, and category 4 crime, burglary in the same location; the category 3 crime, robbery, is correlated to the previous occurrences of robbery and rape in the same location; the category 4 crime, burglary is influenced heavily by the previous occurrences in the same location and neighboring locations burglary and by robbery in the same location.

It should be noted that the forecast function is dependent on the inter-category as well as inter-location correlations. Dropping the inter-category correlated structure, i.e., dropping the $\hat{\phi}^{23}_{20}$, $\hat{\phi}^{24}_{20}$, $\hat{\phi}^{32}_{10}$, $\hat{\phi}^{43}_{20}$ terms, we obtain 4 sets of independent forecast functions that still contain the inter-location structure. Dropping the inter-location structure, that are contained in the $\hat{\phi}^{22}_{12}$, $\hat{\phi}^{44}_{21}$ terms, we get 3 sets of independent forecast functions that still keep the inter-category structure. Dropping all the inter-category and inter-location structure, we obtain $4\times3 = 12$ independent forecast functions that contain no inter-category structure nor inter-location structure. In the appropriate MULSTARMA model, to describe the data, Equation (7-99), murder is independent from the other crimes. Therefore, the forecast function for the category of murder would be the same if the categories were originally treated indpendently or jointly. However, the other three categories if treated independently would not give as refined a forecast function as the simultaneous model approach.

Based on Equations (7-99), (7-97) and (7-98), the standardized N(0,1) point forecasts and the 95% confidence interval forecasts at

T=35 were computed and listed in Table 7-7. Since the standardized expectation value of $\underset{\sim}{Z}_t$ is $\underset{\sim}{0}$, so the point forecasts converge to zero for each category and each location. Conversion to actual units can be accomplished using data from Table 7-4. Also, the process is stationary, so the variance of the point forecast converges and thus the lower bound and upper bound of $100(1-\alpha)\%$ confidence interval forecast for given α converge. In Figure 7-14, the point forecasts, newly realized observations and these 95% confidence interval forecasts are plotted. Here it is seen that the point forecasts and the bounds of the interval forecasts converge quickly. The forecasts do not contain any variations from the zero expectation when the converging rate is negligible, therefore in Table 7-7, these forecasts are omitted except the 22-step ahead forecasts to illustrate the convergent limits.

Instead of computing the ℓ-step ahead forecasts for $\ell=1,2,\ldots,22$, the forecast function can be applied alternatively to compute 1-step ahead forecasts for T=35,36,...,56. It should be noted that the employed model isn't changed but the forecast values are modified by adopting the newly realized values. With shorter forecast horizons and using the newly adopted observations, the forecasts are expected to more closely mimic the new data than for longer forecast horizons which do not adapt the forecasts with new data. The 1-step ahead forecasts at T=35,36,...,56 are computed and plotted in Figure 7-15 with their corresponding realized observations. Comparing Figure 7-14, which is plotted for the ℓ-step ahead forecasts, and Figure 7-15, we see that the shorter length forecast horizons are more adap-

Table 7-7. The Forecasts at T=35 for the CCP Data

h	i	ℓ	$\hat{Z}^h_{i,T}(\ell)$	LB	UB
1	1	1	0.00	-1.74	1.74
		22	0.00	-1.74	1.74
1	2	1	0.00	-1.74	1.74
		22	0.00	-1.74	1.74
1	3	1	0.00	-1.74	1.74
		22	0.00	-1.74	1.74
2	1	1	-0.20	-1.94	1.54
		2	-0.07	-1.92	1.78
		3	-0.04	-2.66	2.59
		4	-0.02	-2.73	2.70
		5	-0.01	-2.75	2.73
		6	-0.00	-2.75	2.74
		22	-0.00	-2.75	2.75
2	2	1	-0.14	-1.88	1.60
		2	-0.02	-1.78	1.74
		3	-0.00	-2.55	2.54
		22	-0.00	-2.56	2.56
2	3	1	-0.11	-1.85	1.63
		2	-0.08	-1.93	1.77
		3	-0.04	-2.66	2.59
		4	-0.02	-2.73	2.70
		5	-0.01	-2.75	2.73
		6	-0.00	-2.75	2.74
		22	-0.00	-2.75	2.75

Table 7-7. (Cont'd)

h i	ℓ	$\hat{Z}^h_{i,T}(\ell)$	LB	UB
3 1	1	0.36	-1.38	2.10
	2	0.12	-1.81	2.05
	3	0.04	-1.93	2.01
	4	0.01	-1.97	1.99
	5	0.00	-1.98	1.99
	22	0.00	-1.99	1.99
3 2	1	-0.23	-1.97	1.51
	2	-0.03	-2.06	1.80
	3	-0.06	-2.03	1.91
	4	-0.03	-2.01	1.95
	5	-0.01	-1.99	1.97
	6	-0.01	-1.99	1.98
	7	-0.00	-1.98	1.98
	22	-0.00	-1.98	1.98
3 3	1	0.67	-1.07	2.41
	2	0.28	-1.65	2.20
	3	0.11	-1.87	2.08
	4	0.04	-1.94	2.02
	5	0.01	-1.97	2.00
	6	0.00	-1.98	1.99
	22	0.00	-1.99	1.99
4 1	1	0.03	-1.70	1.77
	2	0.01	-1.93	1.96
	3	0.01	-2.67	2.70
	4	0.01	-2.84	2.86
	5	0.01	-2.90	2.91
	6	0.01	-2.92	2.93
	7	0.00	-2.93	2.93
	22	0.00	-2.93	2.93

Table 7-7. (Cont'd)

h i	ℓ	$\hat{Z}^h_{i,T}(\ell)$	LB	UB
4 2	1	-0.17	-1.91	1.57
	2	-0.13	-2.08	1.81
	3	-0.10	-2.64	2.45
	4	-0.07	-2.76	2.61
	5	-0.05	-2.79	2.68
	6	-0.04	-2.79	2.71
	7	-0.03	-2.78	2.72
	8	-0.02	-2.78	2.73
	9	-0.02	-2.77	2.74
	10	-0.01	-2.77	2.74
	11	-0.01	-2.76	2.75
	12	-0.01	-2.76	2.75
	13	-0.00	-2.76	2.75
	22	-0.00	-2.76	2.76
4 3	1	1.18	-0.56	2.92
	2	0.86	-1.09	2.81
	3	0.64	-2.05	3.32
	4	0.47	-2.38	3.32
	5	0.35	-2.56	3.25
	6	0.26	-2.67	3.18
	7	0.19	-2.74	3.12
	8	0.14	-2.79	3.07
	9	0.10	-2.83	3.04
	10	0.08	-2.86	3.01
	11	0.06	-2.88	2.99
	12	0.04	-2.89	2.98
	13	0.03	-2.90	2.96
	14	0.02	-2.91	2.96
	15	0.02	-2.92	2.95
	16	0.01	-2.92	2.95
	17	0.01	-2.92	2.94
	18	0.01	-2.93	2.94
	19	0.00	-2.93	2.94
	22	0.00	-2.93	2.94

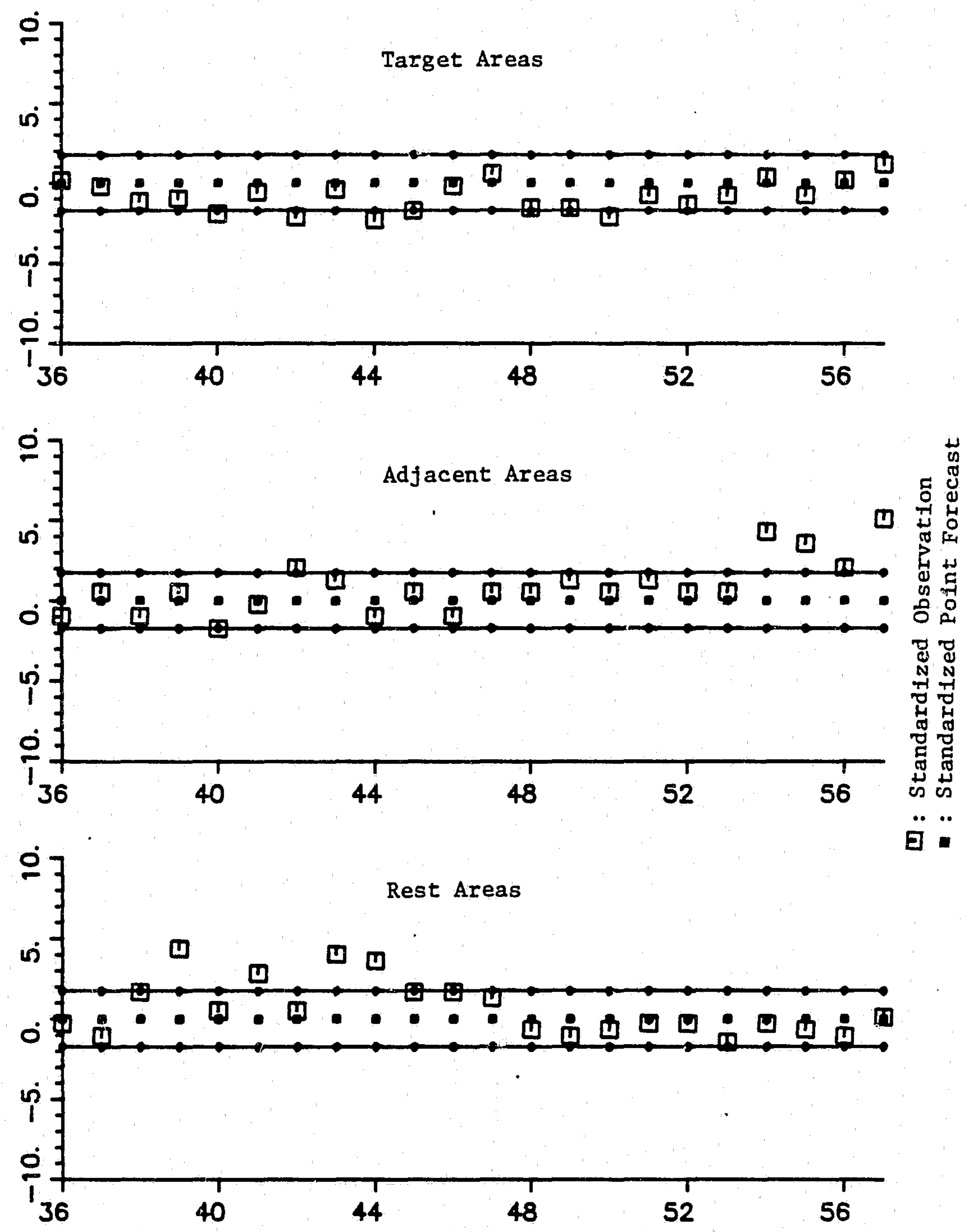


Figure 7-14. Comparison of 22-Step Ahead Forecasts Computed at T=35 and the Observations
(a) Crime Category 1 : Murder

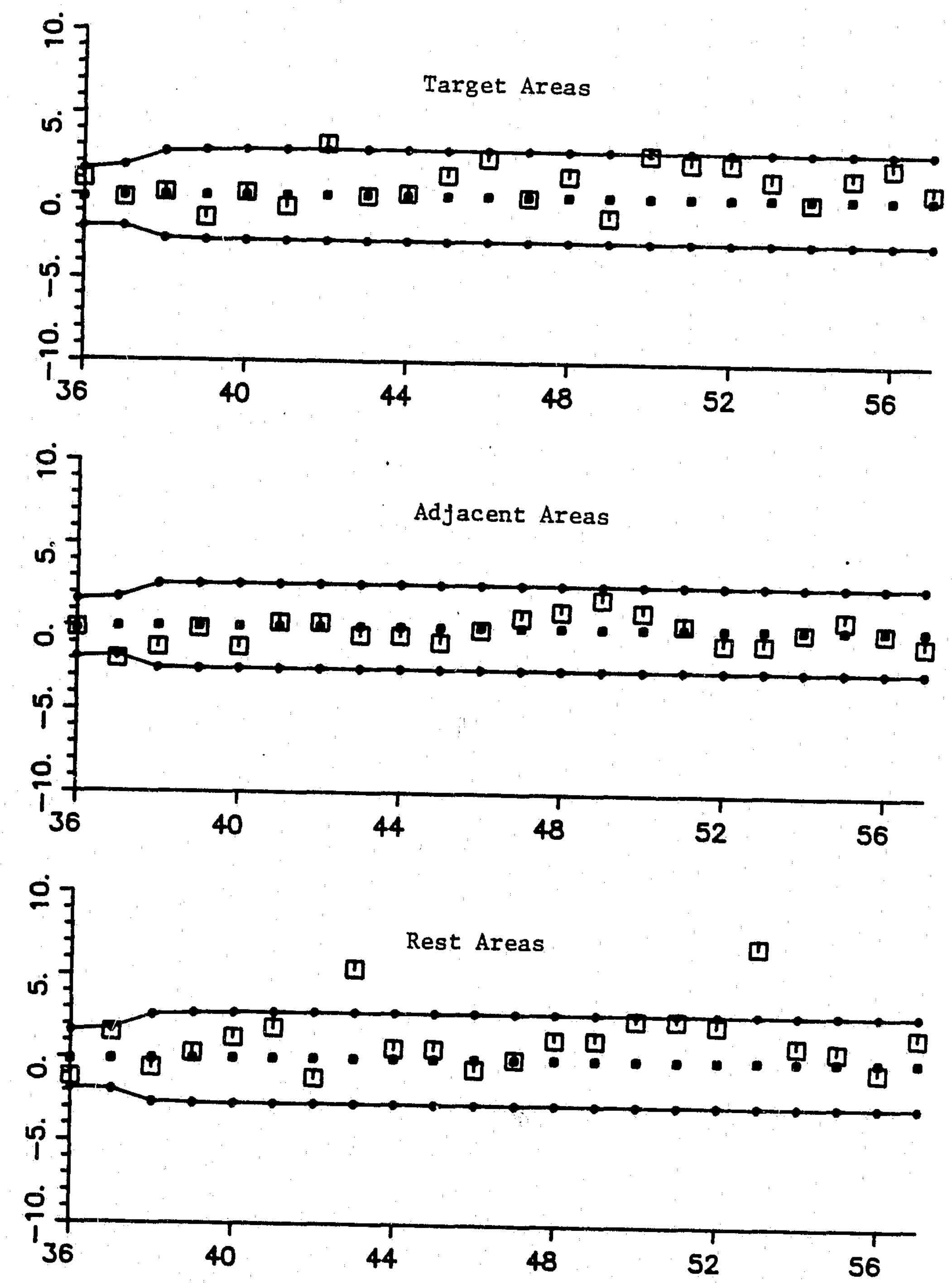


Figure 7-14. (Cont'd)
(b) Crime Category 2 : Rape

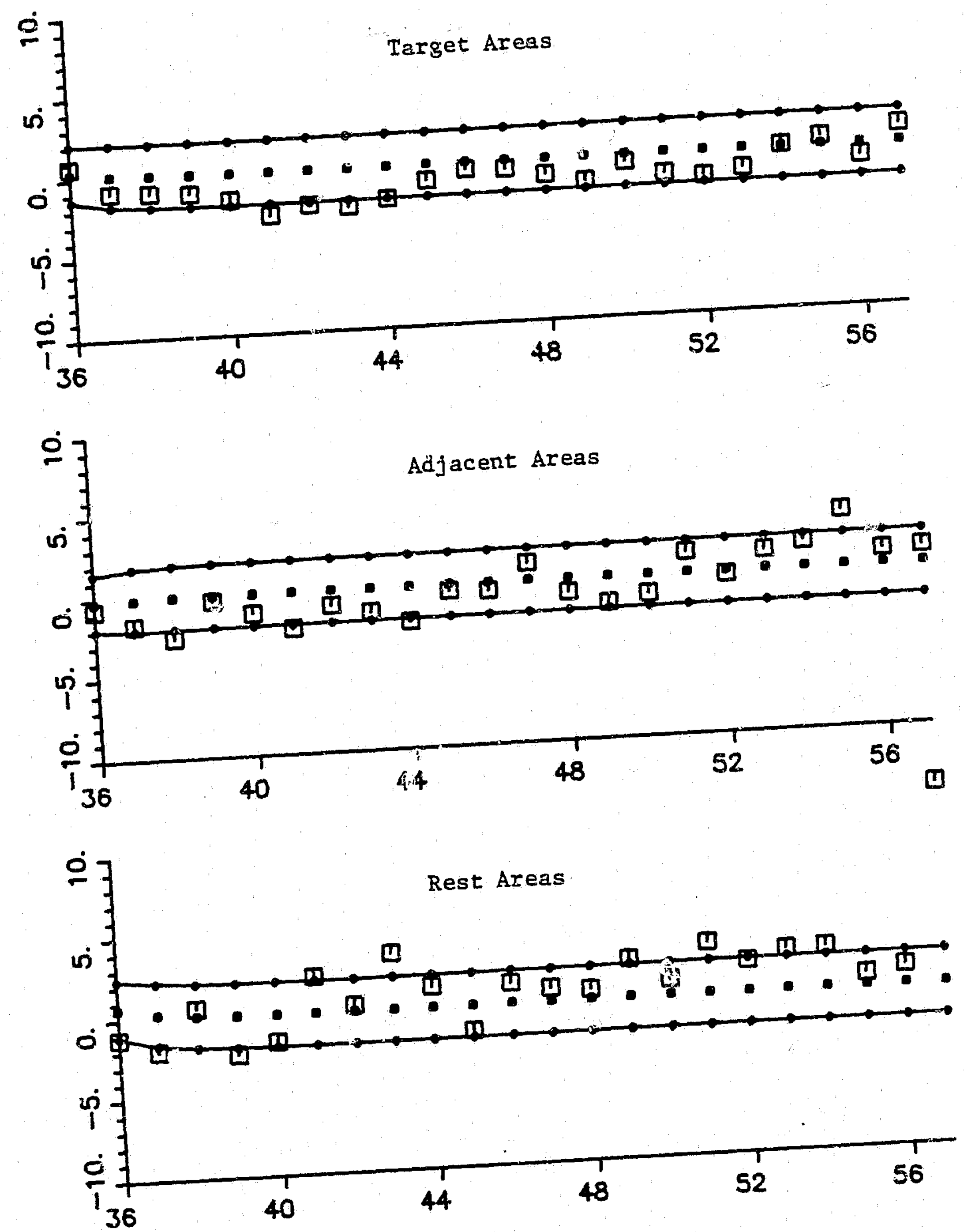


Figure 7-14. (Cont'd)
(c) Crime Category 3 : Robbery

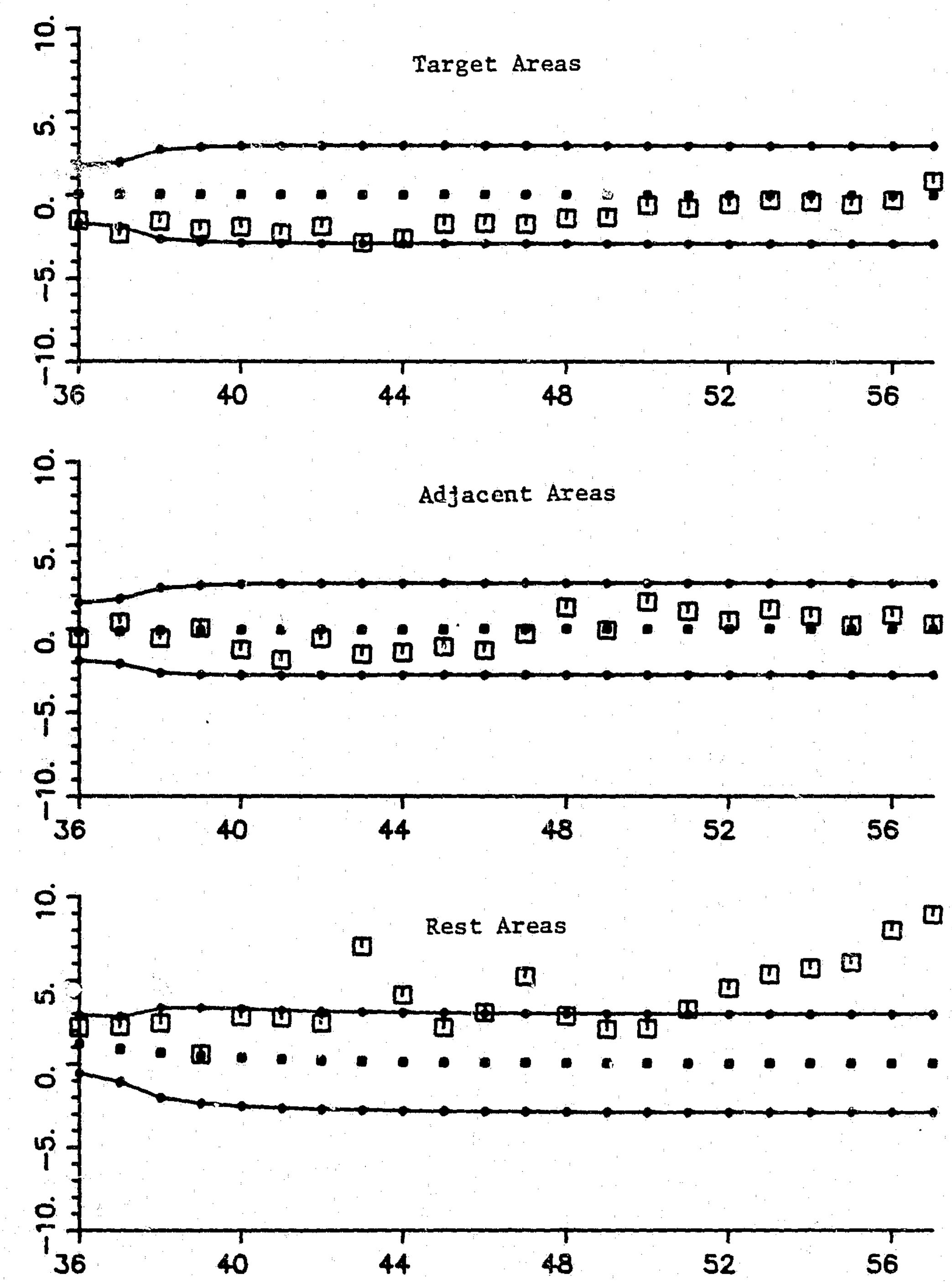


Figure 7-14. (Cont'd)
(d) Crime Category 4 : Burglary

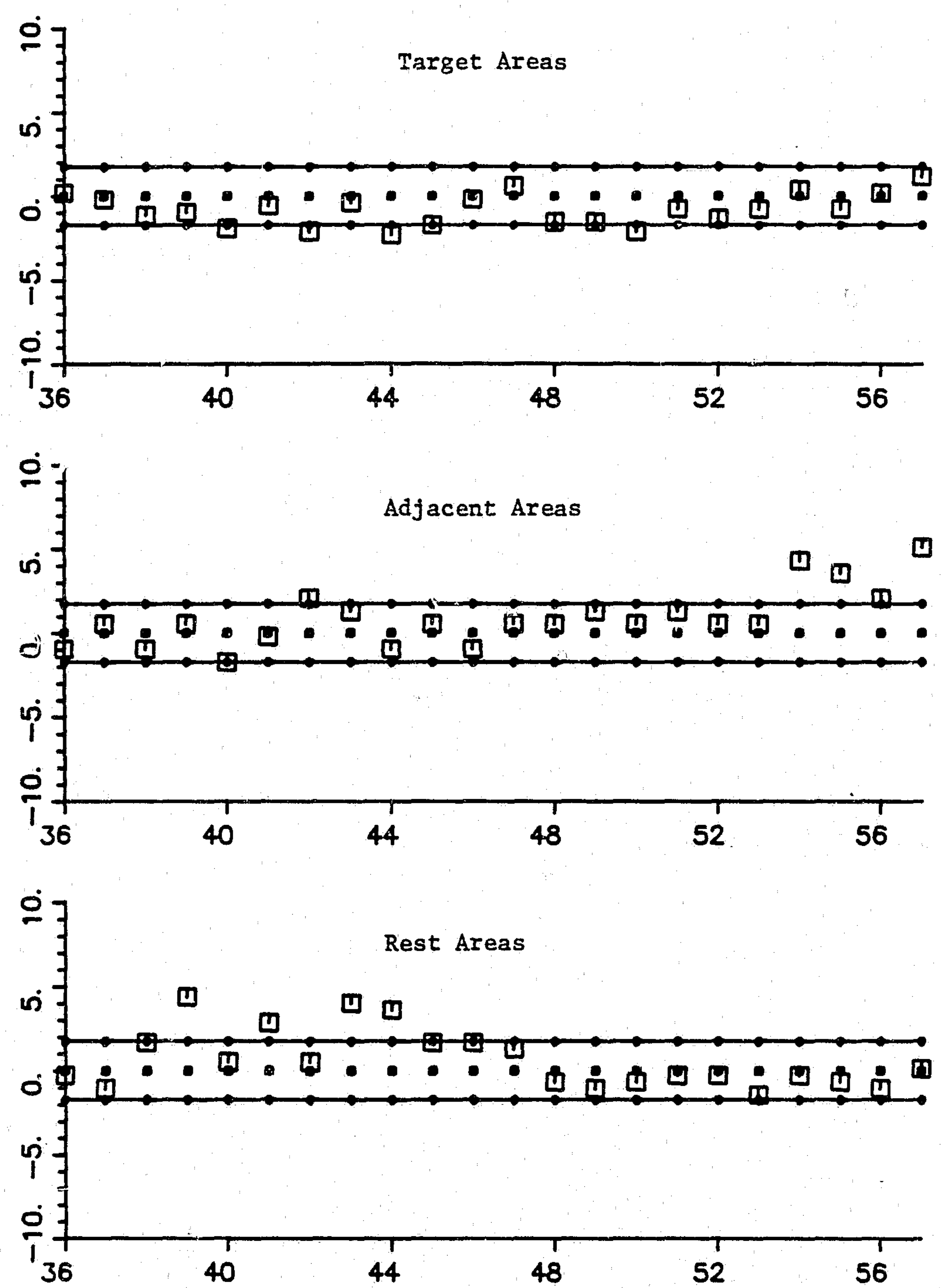


Figure 7-15. Comparison of 1-Step Ahead Partially Updated Forecasts
Computed from T=35 to T=56 and the Observations
(a) Crime Category 1 : Murder

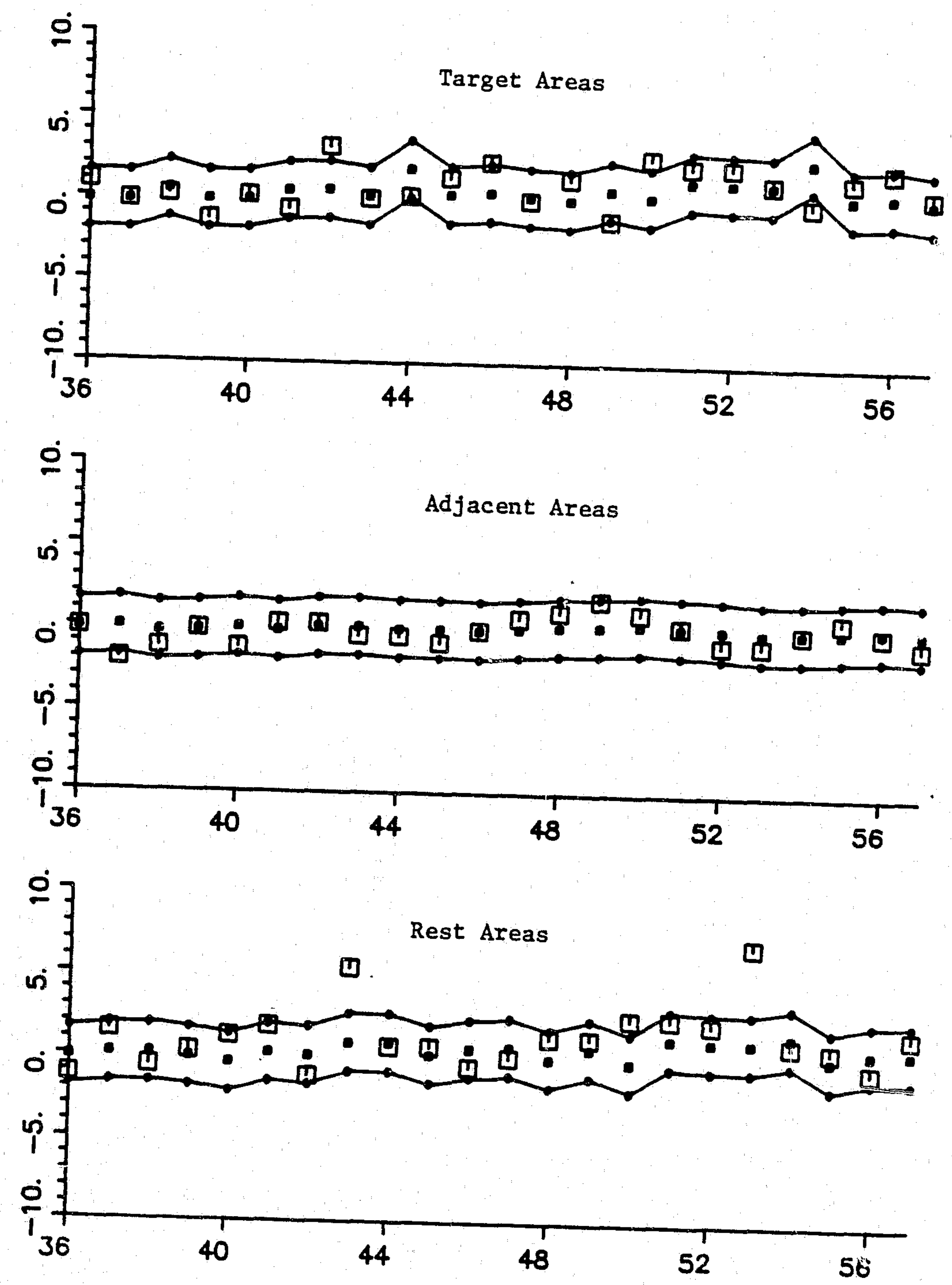


Figure 7-15. (Cont'd)
(b) Crime Category 2 : Rape

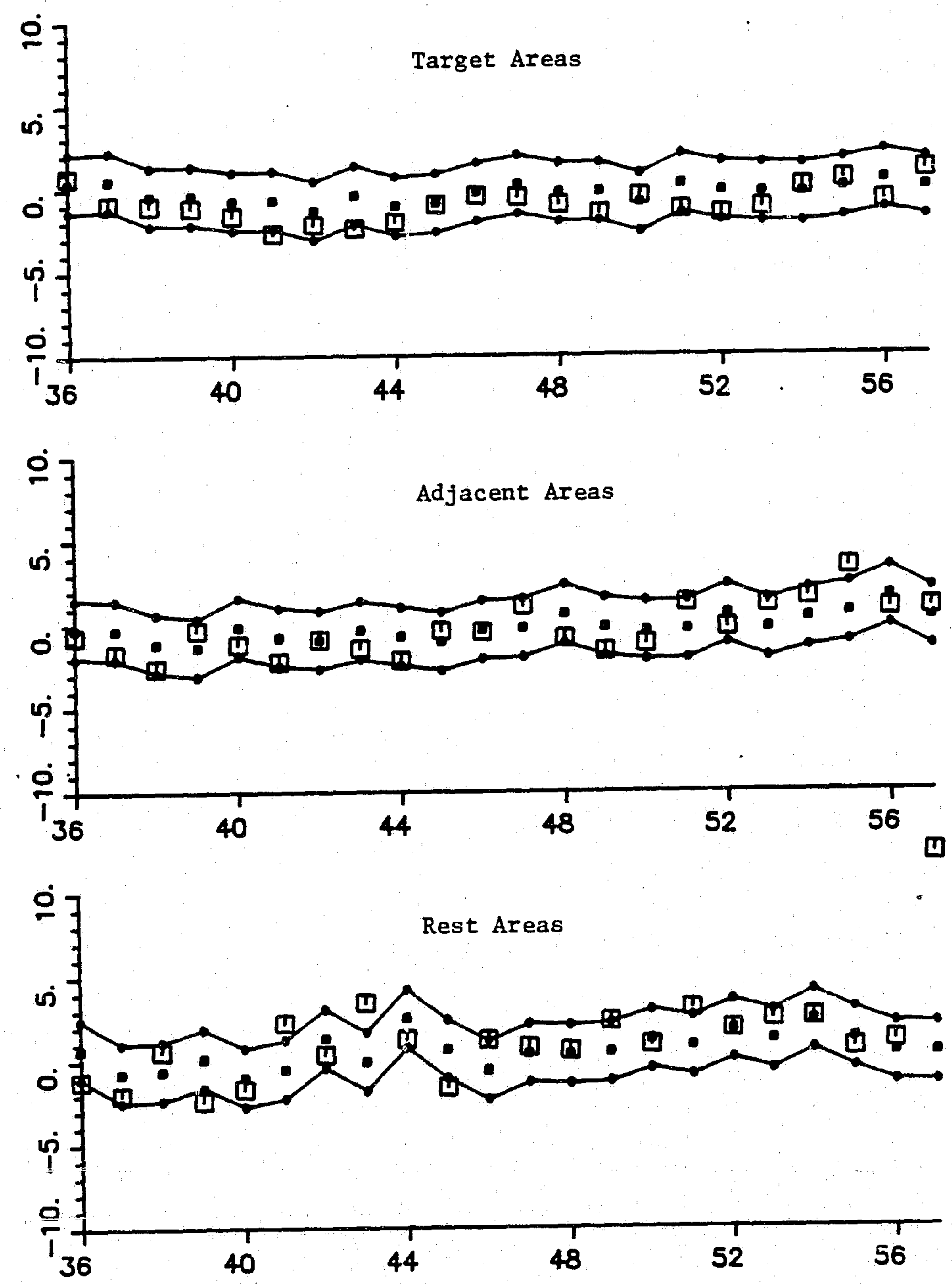


Figure 7-15. (Cont'd)
(c) Crime Category 3 : Robbery

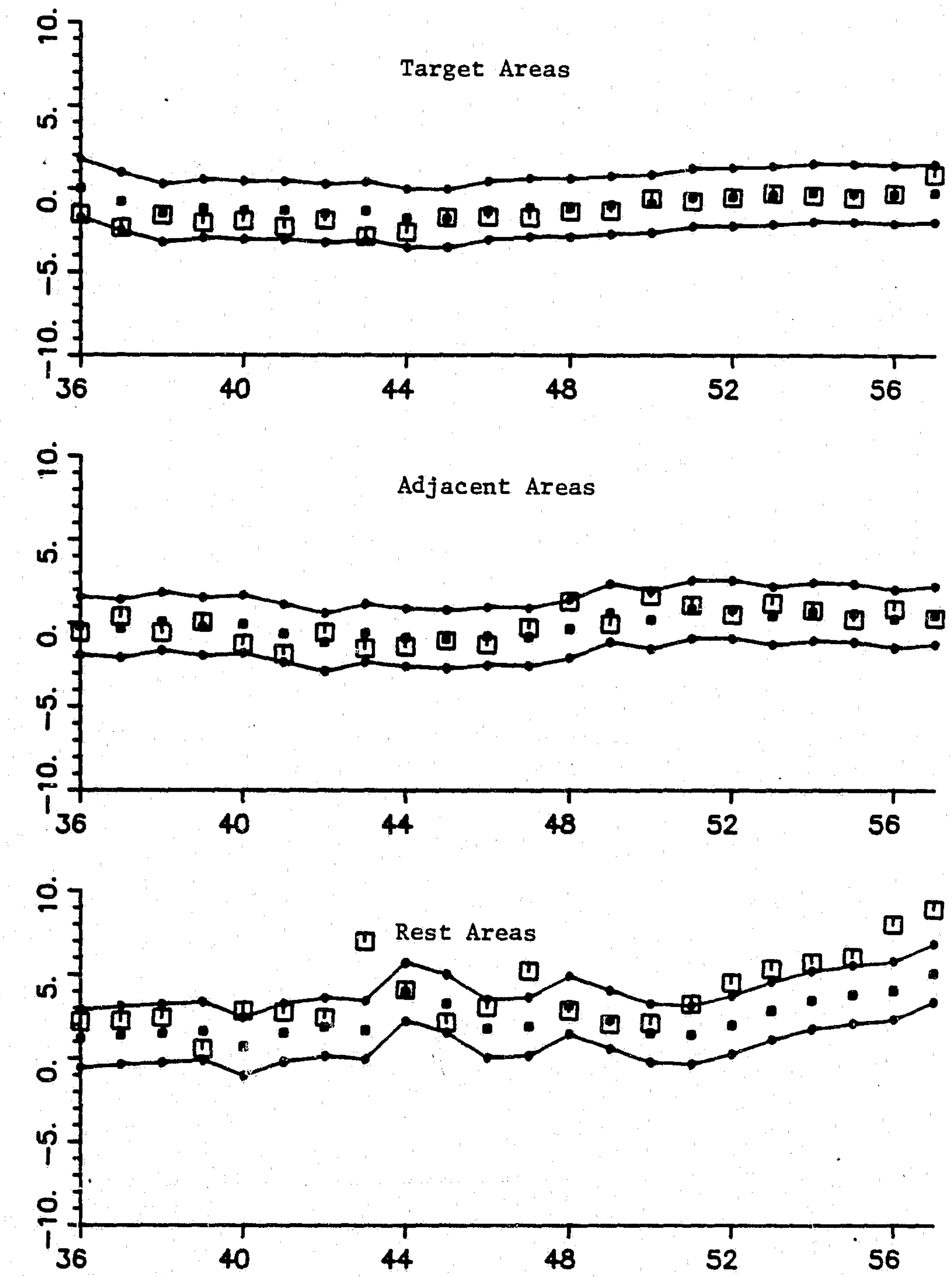


Figure 7-15. (Cont'd)
(d) Crime Category 4 : Burglary

tive. For example, for the category 1 crime, murder, the ℓ-step ahead forecasts are identical to the 1-step ahead forecasts, since the historical process of murder occurrences is a white noise process, which has no structure and therefore newly adapted observations do not have any further information than what was already historically known. However, comparing the ℓ-step ahead forecasts of crime category 4, burglary, in Figure 7-14(d), we see that the 1-step ahead forecasts track the realized observations more closely, because of the adaptation of the new data and its associated information. This is most clearly seen at the rest area. Whereas the observations of burglary crime at the rest area for $t \geq 52$ are clearly out of control with regard to the ℓ-step ahead forecast function while the corresponding 1-step ahead forecast function is still tracking the realized observations within the 95% forecast confidence interval.

7.9.2.2 Preliminary Intervention Analysis. The CCP program initiated at T=40 is considered as an intervention program that was expected to lower the crime occurrences at the target areas and perhaps also reduce the crime occurrences in the adjacent areas. If the underlying process causing crime occurrences is changed by the intervention program then the ℓ-step ahead forecasts should reveal this fact in that the new realized observations should deviate from the forecasts. However, if the process is not changed, then the ℓ-step ahead forecasts should have good forecasting capability for the newly realized observations. In Figure 7-14, it is seen that at the target areas, except category 2 crime, rape, the newly realized observations are consistently lower than the point forecasts, i.e., lower than what

is expected according to the model that describes $T \leq 35$ observations well. Also it is seen that at the rest area, the realized observations are consistently higher than what is expected by the forecast function, Equation (7-99), and the forecast function at T=35 has the best forecast capability for the adjacent areas. Here the largest discrepancy situation is detected at the rest areas for the burglary. To probabilistically assess see how bad this discrepancy is, the probability is, the probability $2 \cdot p_r(D_i^h > |D_i^h|)$ can be computed, where

$$\hat{D}_i^h = (\hat{Z}_{i,T}^h(\ell) Z_{i,T+\ell}^h) / \sqrt{\mathrm{Var}(\hat{Z}_{i,T}^h(\ell))} \sim N(0,1), \qquad (7\text{-}100)$$

and i=3, h=4 are set for the adjacent areas and burglary crime, respectively. Since the forecast function assumes the historical process does not change, the forecast function can be applied to evaluate probabilistically whether the data points entering after t=T represent the process or they are out of control, so the computed probability represents the probability that any new observation would be worse than the new value observed. If this probability is high, the new realization is still consistent with the structure of the historical model. If this probability is low, the new data points do not come from the historical process but rather a different process. The computed probabilities for burglary are,

ℓ=1-11: 0.26, 0.17, 0.19, 0.96, 0.09, 0.10, 0.14, 0.00, 0.01, 0.17, 0.05

12-22: 0.00, 0.06, 0.18, 0.17, 0.03, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00

Clearly the new observations are not consistent with the historical process. Another indication of this can be found by comparing the ℓ-step ahead forecasts contained in Figure 7-14 with the 1-step ahead forecasts contained in Figure 7-15. Again the largest descreponess following a nonrandom pattern occurs in the burglary crime of the rest areas. As seen in Figure 7-15(d), the forecasts for the rest areas, these descreponess goes up rather than fluctuating about a zero level.

To formally model process changes in level and to estimate the amplitude of the intervention effect, the historically based intervention model is augmented to form the intervention model. The intervention model is then applied to evaluate the effects of the CCP on the crime rate dynamically. The intervention model takes the form,

$$\underset{\sim}{Z}^{h}(t) = \sum_{g}\sum_{k}\sum_{\ell} \hat{\phi}^{hg}_{k\ell} w^{(\ell)} \underset{\sim}{Z}^{g}(t-k) - \sum_{g}\sum_{k}\sum_{\ell} \hat{\theta}^{hg}_{k\ell} w^{(\ell)} \underset{\sim}{\hat{\varepsilon}}(t-k) + \delta^{h}(t) + \varepsilon^{h}(t) \qquad (7\text{-}101)$$

where

$h = 1,2,3,4,$

$\hat{\phi}$'s, $\hat{\theta}$'s are parameters of the employed model,

$\underset{\sim}{\hat{\varepsilon}}(t-k)$ is the estimated residual, and

$\underset{\sim}{\delta}^{h}(t)$ is the estimated effect at time t category h.

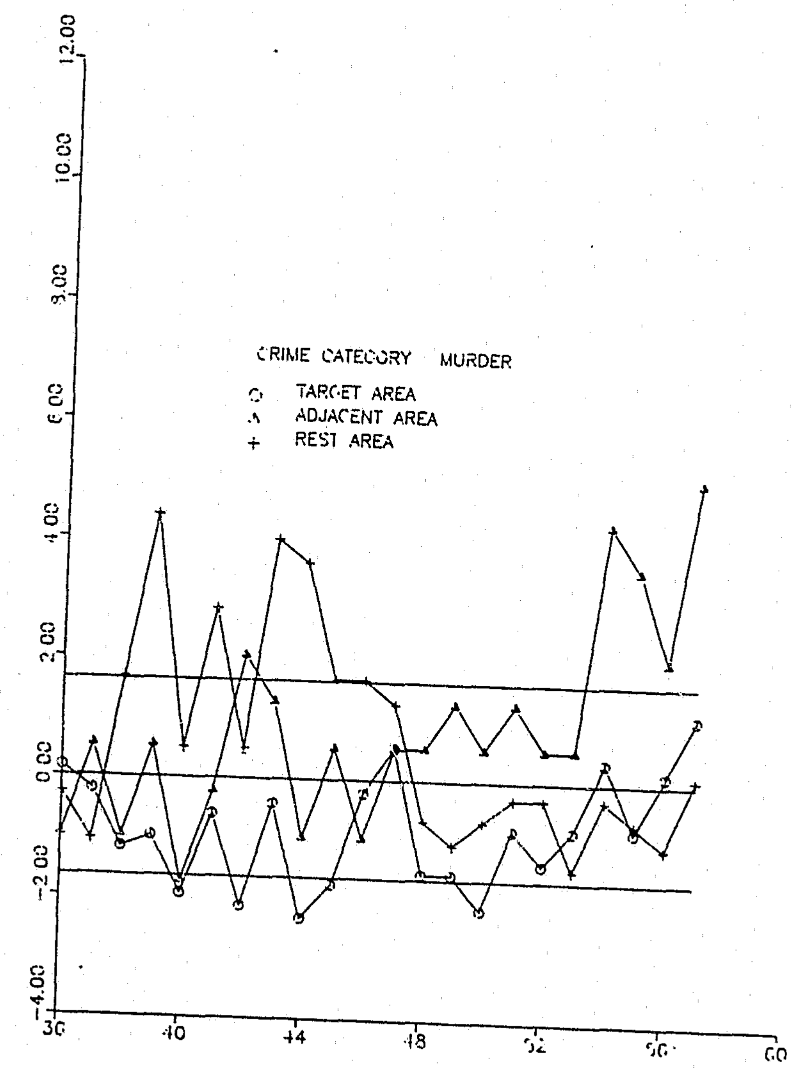


Figure 7-16(a). The Effects of CCP on Murder.

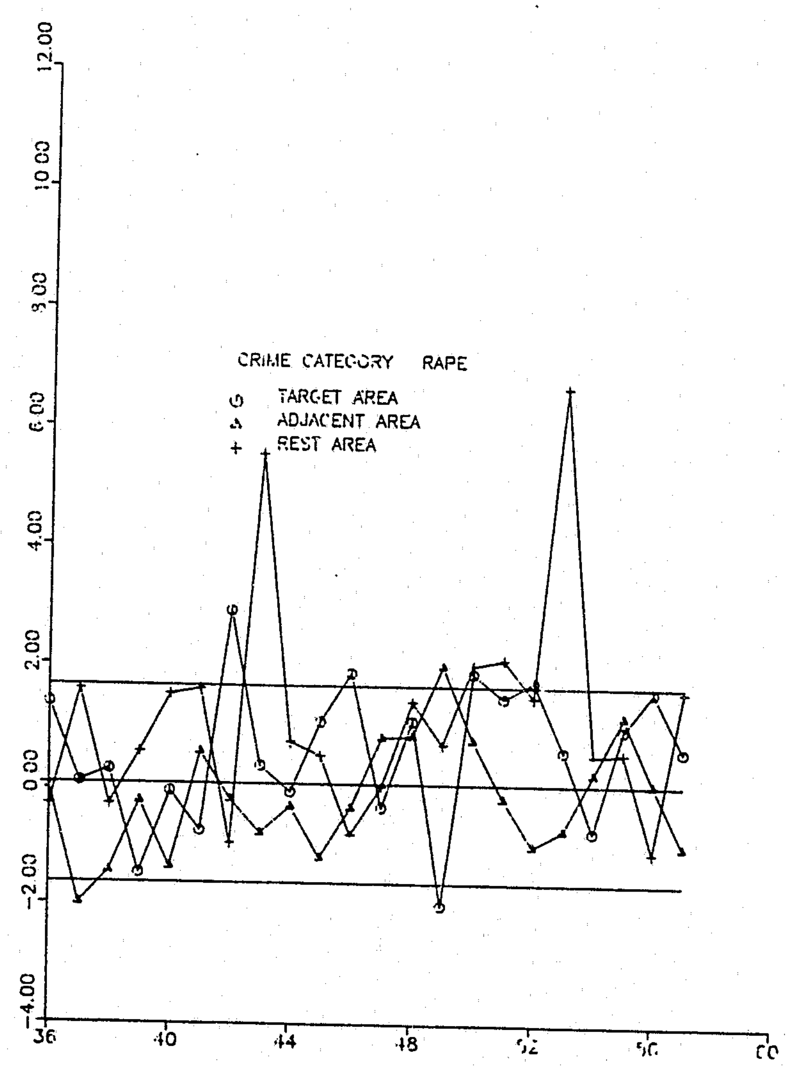


Figure 7-16(b) The Effects of CCP on Rape.

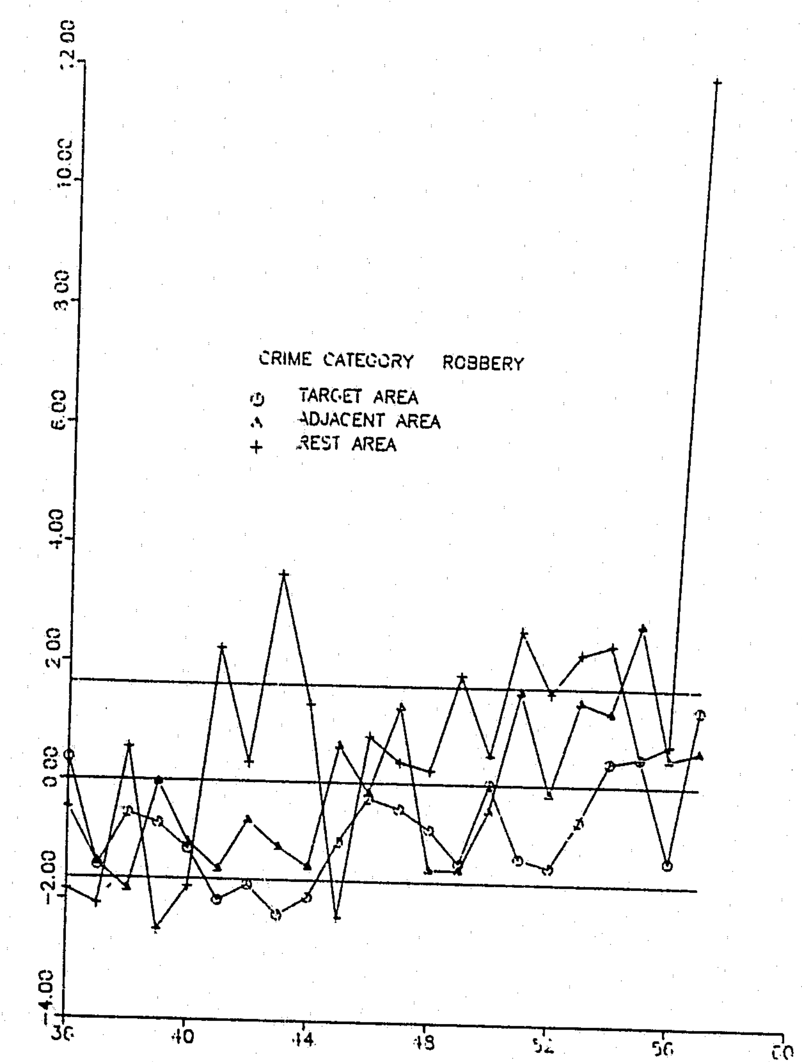


Figure 7-16(c). The Effects of CCP on Robbery.

CONTINUED

7 OF 8

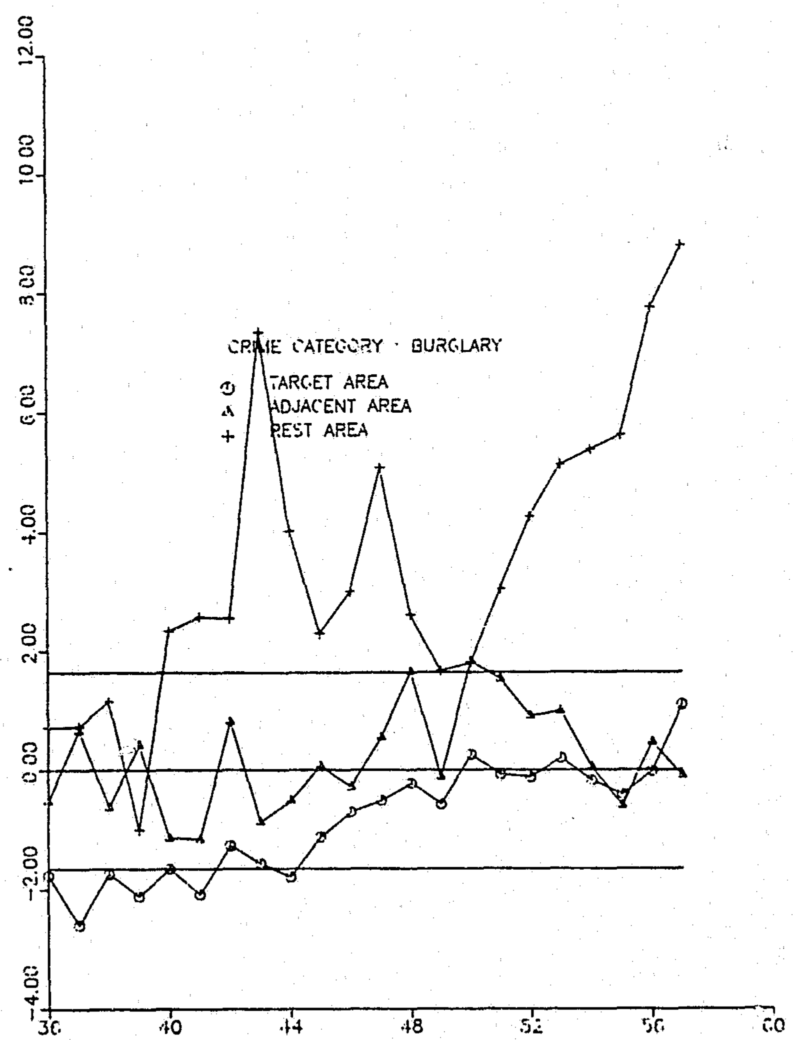


Figure 7-16(d). The Effects of CCP on Burglary.

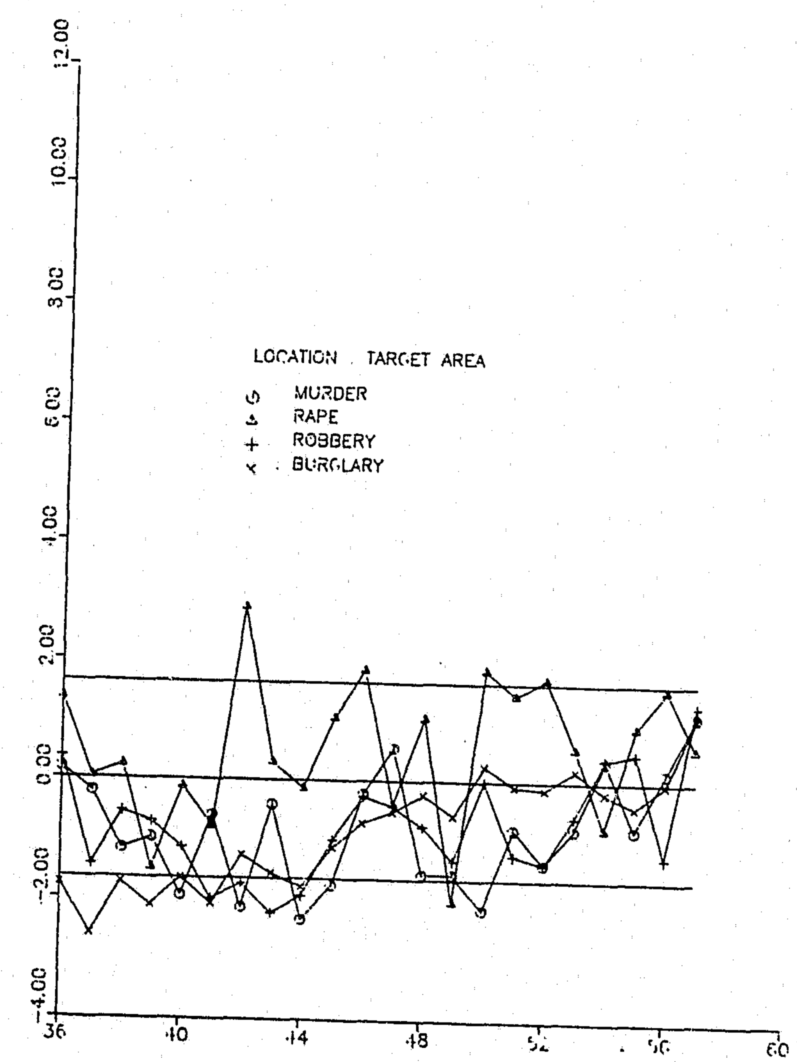


Figure 7-17(a). The Effects of CCP at the Target Area.

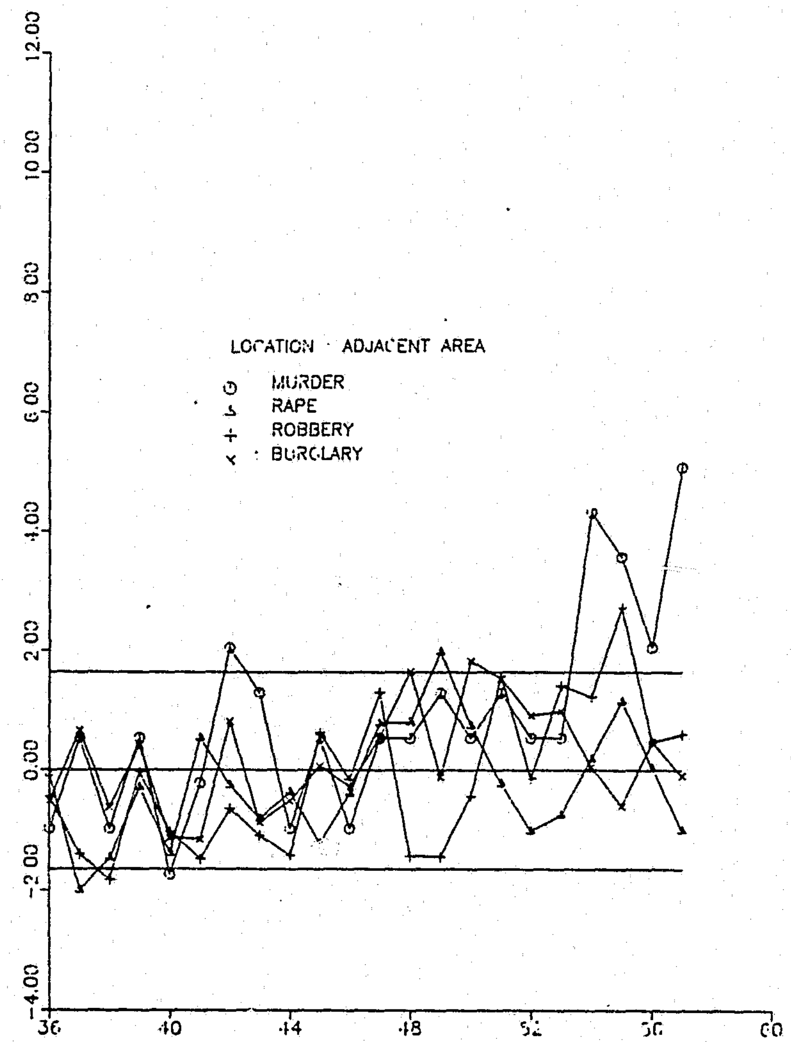


Figure 7-17(b). The Effects of CCP at the Adjacent Area.

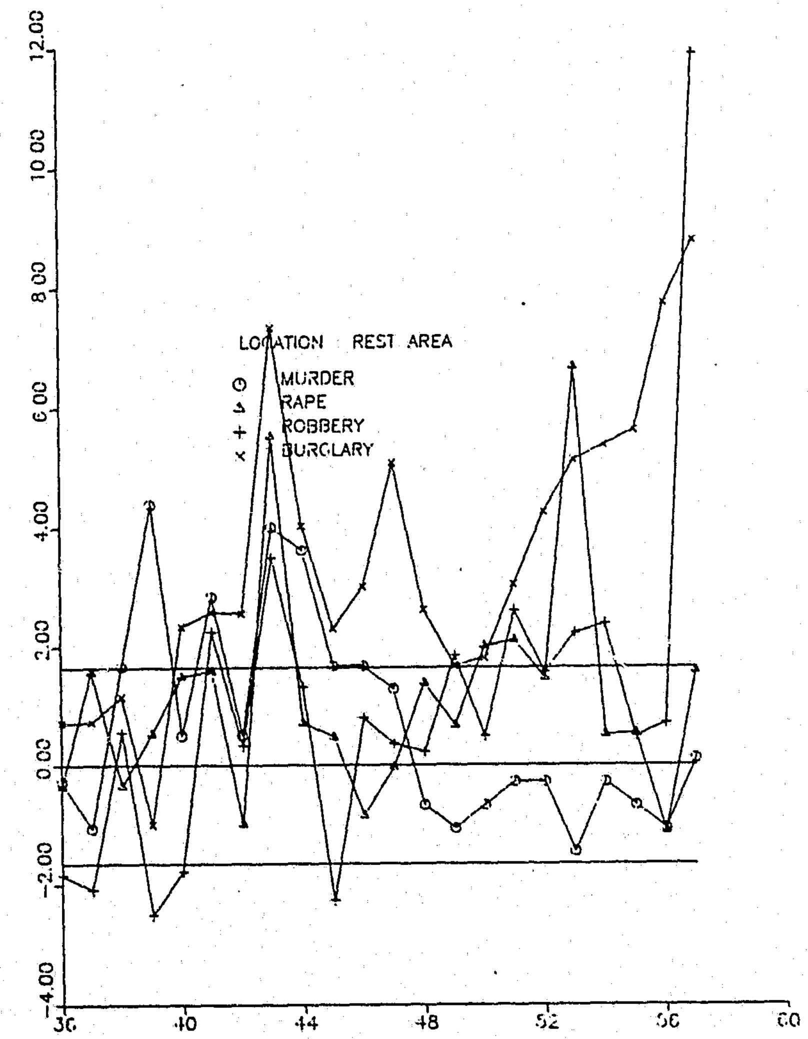


Figure 7-17(c). The Effects of CCP at the Rest Area.

Letting,

$$\underset{\sim}{Y}^{h}(t) = \underset{\sim}{Z}^{h}(t) - \sum_{g}\sum_{k}\sum_{\ell} \hat{\phi}_{k\ell}^{hg} w^{(\ell)} \underset{\sim}{Z}^{g}(t-k) - \sum_{g}\sum_{k}\sum_{\ell} \hat{\theta}_{k\ell}^{hg} w^{(\ell)} \hat{\underset{\sim}{\varepsilon}}(t-k) \quad (7\text{-}102)$$

we have,

$$\underset{\sim}{Y}^{h}(t) = \underset{\sim}{\delta}^{h}(t) + \underset{\sim}{\varepsilon}^{h}(t), \quad (7\text{-}103)$$

and by applying the linear model results, we have

$$\hat{\underset{\sim}{\delta}}^{h} = \underset{\sim}{Y}^{h}(t), \ \mathrm{Var}(\hat{\underset{\sim}{\delta}}^{h}) = \hat{\sigma}^{2} I \quad (7\text{-}104)$$

where $\hat{\sigma}^2$ is the estimated variance of the residuals.

The observations are added to the intervention model, Equation (7-101), one by one chronologically to evaluate the dynamic intervention effect $\underset{\sim}{\delta}_t^h$, h=1,2,3,4 and t=36,37,...,57. The estimated standardized effects are plotted in Figures 7-16 and 7-17. The evaluated effects of all locations are plotted category by category in Figure 7-16 and the evaluated effects of all crimes are plotted location by location in Figure 7-17. For the crime category of murder, the estimated mean shift in the target area are below zero most of the time while the mean shift parameter for the other areas are positive most of the time. This observation suggests that the CCP was effective in lowering the occurrence of murder in the target areas but displacing its occurrence to other areas. For the crime of rape the CCP appears to

have no significant impact since the mean shift estimates for all locations are generally insignificant and randomly vary about a zero value with no overall pattern. The effect of the CCP on the crime of robbery is seen to be consistently negative in the target area and consistently positive in the rest area while little discernable change occurs in the buffer or adjacent area. This suggests a strong tendency to displace robber occurrences rather than deter them. The effect on burglary is even more pronounced than robbery. Figures 7-17(a-c) summarize these dynamic mean shift effects of CCP by locations. The results of this preliminary intervention analysis consistently indicate that the CCP intervention was effective in lowering the crime occurrences in the target area. However, the additional patrol resources in the target area caused an increase in the occurrence of crimes in other areas, particularly those far removed (rest areas) from the target area. Thus, as implemented, these were not general deterrence but more typically displacement. However, since the CCP was primarily based upon differential deployment of existing police resources, the results of this analysis give credence to the favorable input on crime rate of additional police resources.

CHAPTER VIII

SUMMARY CONCLUSION

In this final report, the univariate ARIMA and the STARIMA models were first used to formulate intervention models. These intervention models are developed to have the capability of modeling an environment influenced situation on the non-environment influenced situation, for situations in which, in addition to a change in process level there would also be a covariance change, there are multiple interventions or there is a space-time process. The STARMA model was generalized to include the non-equally preferential structures. The modeling procedures for purely spatial models were developed. Also, the STARMA model was extended to the space, space-time model, that has the capability of capturing the space-time structure and the contemporaneous purely spatial structure. The most generalized MULSTARMA model was constructed to capture the inter-category information as well as the inter-location information.

The specific results and conclusions of this report follows. These results have been grouped into five areas: intervention models, the non-equally preferential model, the purely spatial model, the space, space-time model and the multivariate STARMA model. The section in which the results were first presented is given in parenthesis.

8.1 The Intervention Analysis

1. Based on the univariate ARIMA model class, the multiconsequence intervention was given with two modeling considerations: environment influenced situation and non-environment influenced situation. (2.1)
2. A useful mean shift function takes the form

$$\delta(t) = \begin{cases} 0 & t \leq n_1 \\ \delta k(t) & n_1+1 \leq t, \end{cases}$$

where k(t) is a known numerical value that depends on the model specification and mode parameters. The M.L. estimators of the pre-intervention mean μ and the intrinsic program utility δ were obtained in close form. (2.2.1)
3. The dynamic component identification procedure that gives the unbiased estimates of mean shift function $\delta(t)$ was developed to identify the intervention model specification. (2.2.2)
4. The covariance matrix of the multiconsequence intervention model that is central to the development of the M.L.E. was derived for low-order models. (2.3)
5. The statistics for testing the significance of intrinsic program utility δ, pre-intervention mean value μ, the mean shift function $\delta(t)$ and the change of covariance structure were given and used in estimations and model parsimony considerations. (2.4)

6. The bias in the ability of statistically determining the significance of an intervention program's intrinsic utility due to the misspecification of the mean shift function form and/or the use of a single consequence intervention model from there is changed in the covariance structure was discussed fully. Tables that illustrate representative values of $M = (\sqrt{Var(\hat{\delta})SCI/Var(\hat{\delta})MCI})$ were constructed to illustrate how the standarization of $\hat{\delta}$ in hypothesis testing is affected by ignoring the multiconsequence phenomena. (2.5)

7. The bias of the pre-intervention mean estimator and the program intrinsic utility estimator $\hat{\delta}$ due to misspecification of the model parameters was shown by deriving the formula for $E(\hat{\mu})$ and $E(\hat{\delta})$. Tables of B and D where

$$E\begin{bmatrix}\mu \\ \hat{\delta}\end{bmatrix} = \begin{bmatrix}A & B \\ C & D\end{bmatrix}$$

were computed to illustrate bias tendency. It is concluded that the bias stablizes as n_1 and n_2 increase, where n_1 and n_2 are the pre-intervention and post-intervention sample size. However, increasing sample sizes does not offset the biases induced by parameter misspecification. (2.6)

8. Tables of h, where

$$h = [E(\hat{Var}(\hat{\delta}/\hat{\sigma}_a)_{MISS})/E(\hat{Var}(\hat{\delta}/\hat{\sigma}_a)_{COR})]^{1/2}$$

were given to show the underestimation and overestimation of the significance of δ occur because of the misspecification of model parameters, where MISS and COR denotes misspecified and correct, respectively. Values of h can occur from 0.1 to 2.4. (2.6)

9. The procedures of multiconsequence intervention model building were developed for environment influenced situations as well as the non-environment influenced situations. (2.7).

10. Formulae for computing sample sizes needed for detecting a change in covariance and sample sizes needed for detecing a change in mean level were derived. For a given magnitude change in covariance and/or mean level, a small pre-intervention sample size n_1 requires a larger post-intervention sample size n_2 in order to detect the changes. Also for given n_1, a detection limit for the covariance change as well as mean level change are imposed. (2.8)

11. Principles of optimal economic design for interrupted time series experiments were developed to minimize the total cost while keeping the capability of detecting the specified mean level change and covariance structure change. Typically, the optimal design is robust to the cost coefficient c_1 and c_2, where c_1 denotes the cost per pre-intervention observation and c_2 denotes the cost per post-intervention observation. (2.9)

12. Two substantive examples were presented to illustrate the ARIMA(p,d,q)MCI modeling procedures. In the first example, the newly initiated school poling rewarded students directing in a

desire to modify their inclass behavior in regard to "talkouts". By applying the dynamic component identification procedure, the effect of this education program was determined to be not influenced by the environment and the form of the dynamics were determined to mimic a learning curve type of behavior. It was concluded that the program was effective in reducing "talkouts" by approximately 79%. In the second example, a Gun Control Law was initiated whose purpose was to reduce the murder of gun related crimes. The intrinsic utility of the Gun Control Law was identified by the dynamic identification procedure. From the resulting multi-consequence model the law was determined to be effective in reducing all gun related crimes significantly. (2.10).

13. The space-time single intervention model, (STARIMA)I_m model, for 2N regions was introduced in a general form that has the capability of describing the direct-stimulus-response situation and the indirect-stimulus-response situation for both pulse intervention and step intervention programs. An alternative representation that decomposes the intervention model into two components: the deterministic component and the random component, was obtained. The deterministic component contains no random variables. (3.1, 3.1.1)

14. The space-time multiple intervention model, (STARIMA)I_m, model was introduced as a generalization of (STARIMA)I_m model. When the sequential interventions occur, from the initial realization of the effect of the later intervention, the latter effect is confounded with the former interaction effect. (3.1.2)

15. Two distinct types of spatial-temporal diffusion processes: regenerating type diffusion and relocation type diffusion were introduced as combinations of four distinct diffusion mechanisms: translation, domain-change, growth and contraction. The characterization of the different diffusion types occurring in the environment influence structure was detailed. The stationary (STARMA)I_m model was shown to be always of the regenerating diffusion process. The necessary and sufficient conditions for a homogeneously nonstationary process to be of relocation diffusion type were derived to show that the nonstationary (STARMA)I_m model has the capability of describing the relocation diffusion. (3.2.1, 3.2.2)

16. Three characteristics of a diffusion process: the sphere of influence of the process, the speed of the process and amplitude of the process, were described for the diffusion processes of (STAR)I_m, (STMA)I_m and (STARMA) models. (3.2.3)

17. A system of (11 11) square regions were used to simulate the diffusion processes of (STAR)I_m, (STMA)I_m and (STARMA)I_m model for the regenerating diffusion process and for the relocation diffusion process. The simulated results were plotted in 3-Dimension plots chronologically to illustrate the diffusion phenomena. (3.4)

18. Procedures of modeling space-time intervention processes that include the procedures for building the dynamic model of the intervention effect were given. (3.5)

19. The recursive transformation formula that transforms the $(\text{STARMA})I_m$ model into a linear model was derived for the model parameter known situation. Results from linear model theory was then applied to estimate the pre-intervention mean $\underset{\sim}{\mu}$ and the mean shift vector $\underset{\sim}{\delta}$. A closed form transformation formula for $\text{STARIMA}(1_{\lambda_1},0,1_{m_1})I_m$ model was derived for convenient use. (3.5.1, 3.5.2)

20. To avoid searching the least sum of squares function over a high-dimension parameter space when model parameters are unknown, an efficient approach based on linearlization was developed to search for the least sum of squares iteratively. (3.5.3)

21. The conditional least square estimators $(\underset{\sim}{\psi},\underset{\sim}{\gamma},\underset{\sim}{\delta}|\underset{\sim}{\hat{\phi}},\underset{\sim}{\hat{\sigma}},\underset{\sim}{\hat{\mu}})$ for the multi-consequence space-time intervention model were derived in the linear model form. (3.5.4)

22. A substantive example of air pollution quality control was presented to illustrate the $(\text{STARIMA})I_m$, of modeling procedures. In this example, two interventions occurred in sequence, an engine design change followed by the change in the method of instrument calibration. The effect of engine design change legislation was identified to be influenced by the environment process and the effect of the calibration method change was identified to be non-environment involved. Based on eight year car change over assumption, this analysis concluded that the engine design change legislation significantly reduced the air pollution levels preferentially in the most highly polluted areas while changing the process covariance structure. However, the calibration method change, of which the effect was independent of environment, significantly reduced the pollution readings uniformly at all locations and didn't change the process covariance structure. (3.6)

8.2 The Non-Equally Preferential Model

23. The physical meaning of the relative weights in the weight matrix which determines preferential diffusion were examined. The column scaled weight matrix was determined to have the property that every location gives equally weighted influence to its neighbors of the same order, while the row scaled weight matrix was determined to have the property that every location receives equally weighted influence from those locations that share it as the common neighbor of the same order. (4.1.1)

24. Non-equally preferential weight matrices can't be obtained from linear combinations of equally preferential weight matrices and approaches for the construction of non-equal preference weighted matrix are needed. Two approaches, the strip region approach and the angular region approach, were proposed to construct the non-equally preferential neighbor structures. The angular region approach results in complementary neighbor structures that together resolve to the equally preferential neighbor structure. (4.1.2, 4.1.3)

25. A system of (11×11) locations on the two-dimension regular grids were used to simulate one-direction preferential space-time processes and two-dimension preferential space-time processes for illustrating the relationship between the non-equally preferential

diffusion process and the corresponding weight matrices. The diffusion speed, the influenced regions and the diffusion amplitude were described for low order STAR, STMA and STARMA processes. (4.2)

26. Theoretical diagnostic checking procedures were developed to determine model inadequacies when assuming equal preference structure. These inadequacies are detected from the residual sample space-time autocorrelation functions only when ${}^{N}\Phi_k\Phi_k^{-1}$ and/or ${}^{N}\theta_j\theta_j^{-1}$ are far away from the identity matrix I for some k, j, $1 \leq k \leq p$, $1 \leq j \leq q$. Simulation examples were used to illustrate the theory. (4.3)

27. A procedure that performs the isotropic property test was developed by decomposing the equally preferential neighbor structure into neighbor structures in the preferential directions and in the other directions and then constructing the non-equally preferential model accordingly. A test of model parameter equivalence was then applied to test the isotropic property. (4.4.1)

28. Two procedures for testing the significance of these non-equally were developed to update equal preference models to accomodate non-preferential dependies by exploiting intervention from joint confidence intervals or over fitting. (4.4.2)

29. An extended analysis of the pre-I_1, LA CO Data was presented to illustrate the non-equally preferential diffusion modeling procedures. The preferential neighbor structures were constructed by applying the strip region approach and the angular region approach, and the non-equally preferential models were constructed accordingly.

The ${}^{\|}\Phi_1\Phi_1^{-1}$ matrices were computed for both non-equal preference models to compare with the identity matrix to justify the insensitive diagnostic checking that had been applied on the residuals of the equally preferential models. This resulting preferential model indicated that the coast line perpendicular directions were the directions of preference for the diffusion of the CO particulates in the air. (4.6)

8.3 The Purely Spatial Model

30. The existence conditions were derived for the purely spatial model parameters to guarantee the process existence. The necessary existence conditions of the purely spatial AR(λ_0), MA(m_0), ARMA(λ_0,m_0) models were illustrated. These necessary existence conditions were shown to be also the sufficient existence for regular grid systems with $\lambda_0, m_0 \leq 2$. (5.1)

31. The purely spatial autocovariance function was defined as

$$\gamma_{ks} = E(\hat{\gamma}_{ks}) = \frac{[W^{(k)}\underset{\sim}{Z}_t]'[W^{(s)}\underset{\sim}{Z}_t]}{LN},$$

and the autocorrelation function was defined in terms of the autocovariance function as,

$$\rho_{ks} = \frac{\gamma_{ks}}{[\gamma_{kk}\gamma_{ss}]^{1/2}}$$

The cut-off property of ρ_{ks} was shown for the purely spatial STMA(m_0) model. The tail-off property of ρ_{ks} was shown for the purely spatial STAR(λ_0) model and the STARMA(λ_0,m_0) model. The cut-off, tail-off property was shown to be useful in distinguishing the purely spatial STMA(m_0) models from the other models. (5.2.1, 5.2.2)

32. The partial autocorrelation function set was defined as the set of all the solutions of the appropriate set of the Yule-Walker type equations. The cut-off property of the purely spatial partial autocorrelation functions for the purely spatial STAR(λ_0) model at λ_0 spatial lag, and the tail-off property for the purely spatial STMA(m_0) model and STARMA(λ_0,m_0) model were shown. A numerical example was given to illustrate the cut-off property of the purely spatial partial autocorrelation function for the STAR(1) model, and the computational difficulty in solving the simultaneous quadratic equations was discussed. (5.2.3)

33. Pattern recognition was introduced to help in candidate model identification because of the difficulty in computing the purely spatial partial autocorrelation function sets. The expectated sample autocorrelation function $E(\hat{\rho}_k)$ of the purely spatial AR(1), MA(1), ARMA(1,1), AR(2) and MA(2) models were computed for the 1×25 line systems and 5×5 regular grid systems. Charts were developed to serve as prototypes for comparison in identifying a candidate model. (5.2.4)

34. Charts were constructed to obtain initial estimates of the purely spatial ARMA(1,1), AR(2), MA(2) models to reduce subsequent computation effort in estimation. (5.3.2)

35. The estimation procedures for the purely spatial models were developed for $G = \sigma^2 I$, general G and G unknown situations. (5.4)

36. The residuals computed from a space-time process describing a hydrology process of river flow and from a space-time process describing a criminology process of crime commission were modeled to illustrate the purely spatial modeling procedures. These two examples resulted in models of significant contemporaneously correlated structures past the space-time autocorrelative information. (5.6)

8.4 Space, Space-Time Models

37. The aggregated purely spatial model was coupled with the space-time model and reparameterized to form the space, space-time models. The coupling mechanisms were detailed for three potential space-time models: all LN streams of residuals are of the same univariate ARIMA model, the LN streams of residuals are from the STARIMA process and the LN streams of residuals are from different ARIMA processes. It was shown that the modeling results are independent or approximately independent of modeling sequence for the equally preferential system. (6.1.1)

38. A space, space-time process with $G = \sigma^2 I$ can be mistaken as a space-time process with general G noise structure, which in fact confounds the purely spatial structure described by the contemporaneous terms in the space, space-time model with the process noise covariance. The capability of describing the observed pro-

cess as well as forecasting is the same but the physical interpretations are different. The space, space-time process distinguishes the purely spatial structure from the noise structure while the purely spatial structure and the noise structure stay confounded in the space-time model. (6.1.2)

39. The multiplicative models, that were represented by the space-time observation model and purely spatial residual model, of the Mohawk River Heights and Northeast Boston Assault Arrests were coupled to give the space, space-time models. For the Mohawk River Height Data, the purely spatial structure was confounded with the general G noise structure. However, in the Northeast Boston Assault Arrests example, the coupling and reparameterizing procedure unconfounded the general G noise structure, which resulted from space-time modeling, to give the significant purely spatial correlated structure and a diagonal G=D noise structure. Forecast functions were built and numerically computed for both space, space-time model and general G space-time model and analyzed. Similar point forecast were observed but the space, space-time model had smaller variance of the forecasts than the space-time mode. (6.1.3)

40. The ergodic process was defined as the purely spatial process with the property that any collection of observations shares the same contemporaneous spatial structure contained in the overall observations, $\{Z_t, t \in U_T\}$. (6.2.1)

41. The ergodic modeling procedures were developed. These modeling procedures include the test of the ergodic property. Two methods,

the confidence interval method and the χ^2 test method, were proposed for testing the process ergodic property. For low order purely spatial models, $\lambda_0 + m_0 \leq 2$, the confidence interval method is preferred because it is more efficient in application. (6.2.2)

42. The existence of outliers can mask the ergodic property of a purely spatial ergodic process. These outliers were expressed as environmentally influenced input outliers confounded with the process noise on expressed equivalently as non-environmentally influenced outliers that are confounded with the process observations. (6.3)

43. Sequential estimation of outliers was shown to result in no remaining degrees of freedom in order to estimate the residual noise variance. An iterative procedure was developed to identify the potential outlier or simultaneously estimate the residual noise variance. Outlier correlation may or may not result in an ergodic process. (6.3.1)

44. An example that applied the modeling procedures of the ergodic systems with outliers on the Northeast Boston Assault Arrests for the last eleven observation periods was presented. The outlier corrected aggregate purely spatial model was built to describe the unmasked purely spatial ergodic process. The purely spatial model, that was built without outlier correction, was compared with the ergodic process to see the masking effect due to the presence of outliers. This example illustrated the importance of including the outlier identification procedure in the procedures of the purely spatial model coupling. Forecast models were built based on the

outlier corrected model and uncorrected model. The comparison of these forecasts revealed that the ergodic property modified the point forecasts slightly but the exclusive of the purely spatial components gave rise to larger α value in the $(1-\alpha)$ confidence interval than desired. (6.3.2)

8.5 The Multivariate STARMA Model

45. The multivariate STARMA model was proposed and denoted as the MULSTARMA model. It was shown to collapse to the multivariate ARMA model, STARMA model by system simplification and collapse to the MULSTAR, MULSTMA models by parameter simplification. (7.1.1, 7.1.2)
46. Charts were constructed to illustrate the relationship between the MULSTARMA process and its collapsed subset of models. (7.1.3)
47. Stationary and invertible regions for the low order MULSTARMA models were derived. (7.2)
48. A Λ-weight representation were derived for the stationary general multivariate space-time models, the Λ-weight matrices were computed recursively from the derived recursive formula. The Λ-weight representation expressed the observation of MULSTARMA process as the weighted sum of past process noises of all categories and all locations. (7.3)
49. The multivariate space-time autocovariance function was defined as

$$\gamma_{s\ell}^{gh}(u) = \frac{E[(W^{(s)}\underset{\sim}{z}^{g}(t))\ (W^{(\ell)}\underset{\sim}{z}^{h}(t+u)]}{LN},$$

and the multivariate space-time autocovariance function was shown to cut off after q^{j*} temporal lag for MULSTMA process, and tail off for MULSTAR and MULSTARMA processes. The multivariate autocorrelation function was defined as

$$\rho_{k\ell}^{gh}(u) = \frac{\gamma_{k\ell}^{gh}(u)}{(\gamma_{kk}^{gg}(0)\gamma_{\ell\ell}^{gg}(0)\gamma_{kk}^{hh}(0)\gamma_{\ell\ell}^{hh}(0))^{1/4}}.$$

(7.4.1, 7.4.3)

50. The expectation and variance of the sample multivariate space-time autocorrelation function were derived. The expectation and variance of the sample multivariate space-time autocorrelation function of white noise were tabled for easy use in identifying significant sample multivariate space-time autocorrelations. (7.4.2)
51. The multivariate space-time partial autocorrelation function was defined as the last coefficient in the solution of the appropriate Yule-Walker type equations. This function was shown to cut-off when after $\underset{\sim}{p}$ lags in time and $\underset{\sim}{\lambda}$ lags in space for the MULSTAR($\zeta,\underset{\sim}{p},\underset{\sim}{\lambda}$) model and to tail-off for the MULSTMA model. (7.5.1)
52. The Yule-Walker type equations were shown to consists of ζ independent linear equation sets, and the coefficient matrix of each linear equation set was shown to be symmetric. Based on these facts, a more efficient computation approach was developed to compute the partial autocorrelations. (7.5.2)
53. The cut-off and tail-off statistical properties were summarized

and applied to help identify the candidate MULSTARMA models. (7.6)

54. The linear nature of the autoregressive model parameters was applied to reduce the searching dimension. It was shown that, when inter-category process noises are independent, the estimation of one categories' parameters are independent of the estimation of the parameters of the other categories, and the computation task of autoregressive parameter estimation can be reduced by separately estimating them category by category. (7.7)

55. A substantive example of Cleveland Crime Data, that contained four crime categories and three differently treated regions, was presented to illustrate the MULSTARMA modeling procedures. Based on the pre-intervention MULSTARMA model, the forecast function was built. The ℓ-step ahead forecasts were illustrated. Also based on the pre-intervention MULSTARMA model, a preliminary intervention analysis was performed and determined that the CCP intervention was effective in lowering the crime occurrences in the target areas, but the crime occurrences in the adjacent areas and the rest areas was increased. Since the CCP was primarily based upon differential deployment of existing police resources therefore the crime reduction was not due general deterrence but rather displacement. (7.9)

BIBLIOGRAPHY

Box, G. E. P. and Jenkins, G. M. (1970), Time-Series Analysis : Forecasting and Control, San Francisco : Hloden Day.

Box, G. E. P. and Tiao, G. C. A. (1965), "A Change in Level of a Non-Stationary Time-Series", Bionmetrika, 52, p 181.

Box, G. E. P. and Tiao, G. C. A. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems", Journal of the American Statistical Association, 70, p 70.

Brown, Lawrence A. (1967), Diffusion Processes and Location, Philadelphia : Regional Science Research Institute.

Dahman , J. S. (1975), Examination of Police Patrol Effectiveness High Impact Anti-Crime Program, The Mitre Corporation.

Deutsch, S. J. and F. B. Alt, "The Effect of Massachusett's Gun Control Law on Gun Related Crimes in the city of Boston", Evaluation Quarterly, 1977, 1, p 543.

Deutsch, S. J. and Pfeifer, P. E. (1981d), "Space-Time ARMA Modeling with Contemporaneously Correlated Innovations", Technometrics, Vol. 22, No. 4.

Draper, N. R. and Smith, H. (1966), Applied Regression Analysis, John Wiley & Sons.

Durbin, J. (1960), "The Fitting of Time Series Models", International Statistical Revue, 28, p 233.

Glass, G. V., Wilson, V. L. and Gottman, J. M. (1975), Design and Analysis of Time-Series Experiments, Boulder : Colorado Associated University Press.

Graybill, F. A. (1976), Theory and Application of the Linear Model, North Scituate : Ouxburg Press.

Hall, R. V., R. Fox D. Willard, L. Goldsmith, M. Emerson, H. Owen, F. Davis, and E. Porcia (1971), "The Teacher as Observer and Experimenter in the Modification of Disputing and Talking-out Behaviors", Journal of Applied Behavior Analysis, 4, p 141.

Hannan, E. J. (1970), Multiple Time Series, New York : John Wiley and Sons.

Kendall, M. G., The Advanced Theory of Statistics, Vol. I,II,III, New York : Hanfner.

Pfeifer, P. E. (1979), Spatial-Dynamic Modeling, Ph. D Dissertation, School of Industrial and Systems Engineering, Georgia Institute of Technology.

Pfeifer, P. E. and Deutsch, S. J. (1979), "Comparison of Estimation Procedures for the Parameters of the STAR Model", Communications in Statistics, Vol. B8, No. 2.

Pfeifer, P. E. and Deutsch, S. J. (1980b), "A STARIMA Model Building Procedure with Application to Description and Regional Forecasting", Transactions of the Institute of British Geographers, Vol. 5, No. 3.

Pfeifer, P. E. and Deutsch, S. J. (1980c), "Identification and Interpretation of First-Order Space-Time ARMA Models", Technometrics, Vol. 22, No. 3, p 397.

Pfeifer, P. E. and Deutsch, S. J. (1980e), "Stationary and Invertibility Regions for Low-Order STARMA Models", Communications in Statistics, Vol. B9, No. 5.

Pfeifer, P. E. and Deutsch, S. J. (1981a), "The Variance of the Space-Time Autocorrelation Function", Journal of the Royal Statistical Society, Vol. 43, No. 1.

Pfeifer, P. E. and Deutsch, S. J. (1981c), "Seasonal Space-Time ARMA Modeling", Geographical Analysis, Vol. 13, No. 2.

Tiao, G. C., Box, G. E. P. and Hamming, W. J. (1975), "A Statistical Analysis of the Los Angeles Ambient Carbon Monoxide Date 1955-1972", Journal of the Air Control Association, 25, p 1129.

Tiao, G. C. and Box, G. E. P. (1981), "Modeling Multiple Time Series with Applications", Journal of the American Statistical Association, 76, p 802.

Wilks, S. S. (1938), "The Large Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", Annual Mathematical Statistics, Statistics, 9, p 60.

Whittle, P. (1953), "Estimation and Information in Time Series", Arkiv fur Mathematik, 2, p 423.

END