ESTIMATION FOR THE MULTI-CONSEQUENCE INTERVENTION MODEL

Frank B. Alt, University of Maryland
Stuart J. Deutsch, Georgia Institute of Technology
Jamie J. Goode, Georgia Institute of Technology

Introduction

The interrupted time-series experiment and its statistical inference was first introduced by Box and Tiao [3] specifically for the ARIMA (0,1,1) process. Their work was extended by Glass, Willson, and Gottman [4] to include other types of ARIMA processes. Their model formulations assume that the autoregressive and moving average parameters before the intervention are the same as those afterwards where these parameters describe the correlative structure. In this paper, these models are made more flexible to allow for the consequences of the intervention affecting these parameters and the process level parameters.

Also, maximum likelihood (ML) and iterative conditional lease squares (ICLS) estimation techniques are presented for both sets of process parameters: those describing process level and those describing internal correlative structure. While explicit expressions are developed for the estimates of the level and shift parameters, algorithms are presented for the numerical computation of those parameter estimates describing the correlative structure. The ML estimates can be used to set up an asymptotic likelihood ratio test to investigate the hypothesis that the autoregressive and moving average parameters prior to the intervention are equal to those after the intervention.

These concepts are specifically addressed to the first-order moving average intervention model with an obvious generalization to other model types. An example is included.

Model Description and Properties

We will be primarily concerned with the continuous intervention situation, where the intervention or treatment remains in effect for each time period after its introduction. For example, if we are monitoring the monthly occurrences of homicide for a particular city, an intervention might consist of a gun control law which remains in effect for a relatively long period of time after its introduction. Furthermore, we will assume that the intervention abruptly changes the level of the observations, although other types of level changes can be easily accomodated.

To account for a possible change in level only upon the introduction of an intervention after the n_1th observation, consider the following modification of an MA(1) process:

$$\left.\begin{array}{l} Z_t = \mu + a_t - \theta_1 a_{t-1},\ t=1,\ldots,n_1\ ; \\ Z_t = \mu + \delta + a_t - \theta_1 a_{t-1},\ t=n_1+1,\ldots,n\ , \end{array}\right\} \quad (1)$$

where $n = n_1+n_2$. We will assume $a_t \sim NID(0,\sigma_a^2)$ for $t = 1,\ldots,n$. This single consequence intervention model, denoted MASCI(1), and its statistical analysis via ICLS was considered by Glass, Willson, and Gottman [4]. We will further modify the intervention model of equation (1) to allow for the intervention affecting the process variability as well as the level. This multi-consequence intervention model, denoted MAMCI(1), has the following formulation:

$$\begin{array}{l} Z_t = \mu + a_t - \theta_1 a_{t-1},\ t=1,\ldots,n_1; \\ Z_t = \mu + \delta + a_t - \gamma_1 a_{t-1},\ t=n_1+1,\ldots,n. \end{array} \quad (2)$$

Thus, the model given in equation (2) differs from that presented in equation (1) since γ_1 has replaced θ_1 for $t = n_1+1,\ldots,n$.

Let $\underset{\sim}{Z} = [Z_1,\ldots,Z_{n_1} | Z_{n_1+1},\ldots,Z_n]^t = [\underset{\sim}{Z}_1^t | \underset{\sim}{Z}_2^t]$, where $\underset{\sim}{Z}_1$ is an $(n_1 \times 1)$ vector and $\underset{\sim}{Z}_2$ is an $(n_2 \times 1)$ vector. Then $E(\underset{\sim}{Z})$, denoted by $\mu_{\underset{\sim}{Z}}$, can be written as

$$\mu_{\underset{\sim}{Z}} = \begin{bmatrix} \mu_{\underset{\sim}{Z}_1} \\ \mu_{\underset{\sim}{Z}_2} \end{bmatrix} = \mu \underset{\sim}{j}_n + \underset{\sim}{k}, \quad (3)$$

where $\underset{\sim}{j}_n$ is an $(n \times 1)$ vector of 1's and

$$\underset{\sim}{k} = \begin{bmatrix} \underset{\sim}{0}_{n_1} \\ \underset{\sim}{j}_{n_2} \end{bmatrix}.$$

Thus, the $(n \times 1)$ vector $\underset{\sim}{k}$ has 0's for its first n_1 entries followed by n_2 1's.

Let $\Sigma_{\underset{\sim}{Z}}$ denote the $(n \times n)$ variance-covariance matrix of $\underset{\sim}{Z}$. Then

$$\Sigma_{\underset{\sim}{Z}} = \begin{bmatrix} B_{\underset{\sim}{Z}_1} & B_{21}^t \\ B_{21} & B_{\underset{\sim}{Z}_2} \end{bmatrix}, \quad (4)$$

where

$$B_{Z_1} = \begin{bmatrix} (1+\theta_1^2) & -\theta_1 & \cdots & 0 & 0 \\ -\theta_1 & (1+\theta_1^2) & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & (1+\theta_1^2) & -\theta_1 \\ 0 & 0 & \cdots & -\theta_1 & (1+\theta_1^2) \end{bmatrix}$$

$$B_{Z_2} = \begin{bmatrix} (1+\gamma_1^2) & -\gamma_1 & \cdots & 0 & 0 \\ -\gamma_1 & (1+\gamma_1^2) & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & (1+\gamma_1^2) & -\gamma_1 \\ 0 & 0 & \cdots & -\gamma_1 & (1+\gamma_1^2) \end{bmatrix}$$

and B_{21} is a $(n_2 x n_1)$ matrix all of whose entries are zero except for the element in the northeast corner which is $-\gamma_1$. Furthermore, since $Z = Ca+\mu_Z$, where C is an $[nx(n+1)]$ matrix, we see that Z is distributed as an n-variate normal. The above results can be summarized by saying that for a MAMCI(1) process

$$Z \sim N_n(\mu_Z, \Sigma_Z) \quad , \tag{5}$$

where μ_Z and Σ_Z are presented in equations (3) and (4), respectively.

Iterative, Conditional Least Squares Estimation

Although policy makers are primarily concerned with the estimation of μ and δ for the intervention models, we shall see that least squares estimates of both of these parameters are directly dependent upon the values of the moving-average parameters. The basic idea, which is an extension of that employed by Box and Tiao [3], is to transform the n original observations to another set of variables amenable to linear statistical model analysis. For these transformed variables, we employ an iterative technique of searching on the moving-average parameters until those values are found which minimize the residual sum of squares.

Before finding the necessary transformation, recall that the model $Y = X\beta+a$ with $a \sim N_n(0,\sigma^2 I)$, describes the classic normal linear regression model, details of which can be found in Goldberger [5]. In our case, Y is an (nx1) vector as is a, X is an (nx2) matrix, and $\beta = [\mu,\delta]^t$.

Let $z_1, z_2, \ldots, z_n$ be n successive observations generated from the MAMCI(1) model stated in equation (2). In order to transform the z_t's to y_t's, which are in statistical linear model form, we let $a_o = 0$, $y_1 = z_1$, and $y_t = z_t+\theta_1 y_{t-1}$, for $t = 2,\ldots,n_1$, while $y_t = z_t+\gamma_1 y_{t-1}$, for $t = n_1+1,\ldots,n$. Thus, the transformed variables can be expressed as

$$y_t = (1+\theta_1+\ldots+\theta_1^{t-1})\mu+a_t, \tag{6}$$

for $t = 1,\ldots,n_1$, while

$$\left.\begin{aligned} y_t = {} & [1+\ldots+\gamma_1^{t-n_1}(1+\theta_1+\ldots+\theta_1^{n_1-1})]\,\mu \\ & + (1+\gamma_1+\ldots+\gamma_1^{t-(n_1+1)})\,\delta+a_t \end{aligned}\right\} \tag{7}$$

for $t = n_1+1,\ldots,n$. Equations (6) and (7) have the matrix representation $Y = X\beta+a$, where

$$X = \begin{bmatrix} 1 & 0 \\ 1+\theta_1 & 0 \\ \vdots & \vdots \\ 1+\theta_1+\ldots+\theta_1^{n_1-1} & 0 \\ \hline 1+\gamma_1(1+\theta_1+\ldots+\theta_1^{n_1-1}) & 1 \\ 1+\gamma_1+\gamma_1^2(1+\theta_1+\ldots+\theta_1^{n_1-1}) & 1+\gamma_1 \\ \vdots & \vdots \\ 1+\ldots+\gamma_1^{n_2}(1+\theta_1+\ldots+\theta_1^{n_1-1}) & 1+\gamma_1+\ldots\gamma_1^{n_2-1} \end{bmatrix} \tag{8}$$

The elements of X^tX will be denoted by c_{11}, c_{12}, and c_{22}, where $c_{12} = c_{21}$. After much tedious algebra, it can be shown that

$$\left.\begin{aligned} c_{11} = {} & (1-\theta_1)^{-2}[n_1-(1-\theta_1)^{-1}(2\theta_1)(1-\theta_1^{n_1})] \\ & + (1-\theta_1)^{-2}[(1-\theta_1^2)^{-1}\theta_1^2(1-\theta_1^{2n_1})] \\ & + (1-\gamma_1)^{-2}[n_2-(1-\gamma_1)^{-1}(2\gamma_1)(1-\gamma_1^{n_2})] \\ & + (1-\gamma_1)^{-2}[(1-\gamma_1^2)^{-1}\gamma_1^2(1-\gamma_1^{2n_2})] \\ & + (1-\theta_1)^{-2}(1-\gamma_1^2)^{-1}(1-\theta_1^{n_1})^2\gamma_1^2(1-\gamma_1^{2n_2}) \\ & + 2(1-\theta_1)^{-1}(1-\gamma_1)^{-1}(1-\theta_1^{n_1})\,q_1 \,, \end{aligned}\right\} \tag{9}$$

$$c_{22} = (1-\gamma_1)^{-2}(1-\gamma_1^2)^{-1}\,q_2 \,, \tag{10}$$

and

$$\left.\begin{aligned} c_{12} = {} & (1-\gamma_1)^{-2}[n_2-2(1-\gamma_1)^{-1}\gamma_1(1-\gamma_1^{n_2})] \\ & + (1-\gamma_1)^{-2}[(1-\gamma_1^2)^{-1}\gamma_1^2(1-\gamma_1^{2n_2})] \\ & + (1-\theta_1)^{-1}(1-\gamma_1)^{-1}(1-\theta_1^{n_1})\,q_1 \,. \end{aligned}\right\} \tag{11}$$

Note that $q_1 = (1-\gamma_1)^{-1}\gamma_1(1-\gamma_1^{n_2})-(1-\gamma_1^2)^{-1}\gamma_1^2(1-\gamma_1^{2n_2})$, and

$$q_2 = n_2(1-\gamma_1^2)-2\gamma_1(1+\gamma_1)(1-\gamma_1^{n_2})+\gamma_1^2(1-\gamma_1^{2n_2}) .$$

Let s_{1Y} and s_{2Y} denote the elements of $X^t Y$. Then

$$\left.\begin{aligned} s_{1Y} &= (1-\theta_1)^{-1}(n_1\bar{y}_{n_1} - \sum_{i=1}^{n_1} \theta_1^i y_i) \\ &+ (1-\gamma_1)^{-1}(n_2\bar{y}_{n_2} - \sum_{i=1}^{n_2} \gamma_1^i y_{n_1+i}) \\ &+ (1-\theta_1)^{-1}(1-\theta_1^{n_1}) \sum_{i=1}^{n_2} \gamma_1^i y_{n_1+i} , \end{aligned}\right\} \quad (12)$$

and

$$s_{2Y} = (1-\gamma_1)^{-1}(n_2\bar{y}_{n_2} - \sum_{i=1}^{n_2} \gamma_1^i y_{n_1+i}), \quad (13)$$

where $\bar{y}_{n_1} = n_1^{-1} \sum_{i=1}^{n_1} y_i$ and $\bar{y}_{n_2} = n_2^{-1} \sum_{i=1}^{n_1} y_{n_1+i}$. It follows from linear model theory that

$$\hat{\mu} = c^{11} s_{1Y} + c^{12} s_{2Y} \quad (14)$$

and

$$\hat{\delta} = c^{12} s_{1Y} + c^{22} s_{2Y} , \quad (15)$$

for fixed θ_1 and γ_1 where c^{ij} denote the elements of $(X^t X)^{-1}$. Extending the ad hoc procedure of Glass, Willison, and Gottman to the multi-consequence model, we let $\hat{a} = y - X\hat{\beta}$, where the $\hat{a}$ vector is contingent upon particular values of $\hat{\mu}$ and $\hat{\delta}$ which in turn are contingent upon values of θ_1 and γ_1. Let $S_*(\theta_1,\gamma_1)$ be the sum of squared residuals or estimated erros for particular values of θ_1, γ_1, $\hat{\mu}$, and $\hat{\delta}$. That is,

$$S_*(\theta_1,\gamma_1) = \sum_{t=1}^{n} \hat{a}_t^2 = \hat{a}^t\hat{a} = (y - X\hat{\beta})^t(y - X\hat{\beta}) . \quad (16)$$

Minimizing $S_*(\theta_1,\gamma_1)$ is equivalent to minimizing $\hat{\sigma}_a^2 = \hat{a}^t\hat{a}/(n-2)$. The search for the minimizing (θ_1,γ_1) pair can be restricted to the open unit square, that is, $(\theta_1,\gamma_1)\varepsilon\{(x_1,x_2):0<x_i<1, i=1,2\}$. The output format assoicated with the search can be set up in table fashion with the following column headings: $\theta_1, \gamma_1, \hat{\mu}, \hat{\delta}, \hat{\sigma}_a^2$. After that (θ_1,γ_1) is selected with minimizes $\hat{\sigma}_a^2$, confidence intervals can be constructed or tests of hypotheses can be performed for μ or δ by making use of the fact that $(\hat{\mu}-\mu)/\hat{\sigma}_a(c_{11})^{1/2}$ and $(\hat{\delta}-\delta)/\hat{\sigma}_a(c_{22})^{1/2}$ are each distributed as pseudo Student-t random variables with n-2 degrees of freedom. The "pseudo" prefix is necessitated by the fact that both ratios depend on the nuisance parameters (θ_1,γ_1). Furthermore, keep in mind that the true confidence region for (μ,δ) is elliptical in nature. Thus, any confidence interval for μ or δ alone is merely a marginal one and the confidence levels should be adjusted accordingly.

Note that this lease squares estimation approach was iterative in that it searched on (θ_1,γ_1) and conditional in that we set $a_0=0$. For this reason, it was designated iterative, conditional least squares.

Maximum Likelihood Estimation of μ and δ

In this section, we will obtain closed form expressions for the maximum likelihood estimates of μ and δ where these estimates are functions of the moving average parameters. As such, they are designated conditional maximum likelihood estimates.

Let $\underset{\sim}{z} = [z_1,\ldots,z_{n_1},z_{n_1+1},\ldots,z_n]^t$ be a sample of n observations generated from a MAMCI(1) model and let $\underset{\sim}{Z}$ be the (nx1) random vector associated with the vector of sample observations. Also, let $\underset{\sim}{a} = [a_0,a_1,\ldots,a_n]^t$ be an ((n+1)x1) random vector where a_t-NID$(0,\sigma_a^2)$. Thus, the joint distribution of $\underset{\sim}{a}$ equals

$$f(\underset{\sim}{a}^t;\sigma_a^2) = (2\pi\sigma_a^2)^{-(n+1)/2} \exp\{-\underset{\sim}{a}^t\underset{\sim}{a}/2\sigma_a^2\} . \quad (17)$$

Since $\underset{\sim}{Z} = C\underset{\sim}{a} + \underset{\sim}{\mu}_Z$, it follows that $\underset{\sim}{Z}$-$N_n(\underset{\sim}{\mu}_Z, \sigma_a^2 CC^t)$. But, $\sigma_a^2 CC^t = \Sigma_Z$ where Σ_Z is presented in equation (4). Let $\Sigma_Z = \sigma_a^2 M^{-1}$. Thus,

$$f_{Z^t}(\underset{\sim}{z}^t;\underset{\sim}{\xi}^t) = (2\pi\sigma_a^2)^{-n/2}|M|^{1/2}\exp\{-Q(\mu,\delta)\}, \quad (18)$$

where $\underset{\sim}{\xi}^t = (\mu,\delta,\theta_1,\gamma_1,\sigma_a^2)$ and

$$Q(\mu,\delta) = (\underset{\sim}{z}-\mu\underset{\sim}{j}_n-\delta\underset{\sim}{k})^t M(\underset{\sim}{z}-\mu\underset{\sim}{j}_n-\delta\underset{\sim}{k})/2\sigma_a^2 . \quad (19)$$

In the logarithm of the likelihood function associated with equation (18), μ and δ appear only in the quadratic from $Q(\mu,\delta)$. Let $Q^*(\mu,\delta) = -2\sigma_a^2 Q(\mu,\delta)$.

By finding $\partial Q^*(\mu,\delta)/\partial\mu$ and $\partial Q^*(\mu,\delta)/\partial\delta$ and setting these partial derivatives equal to zero, we obtain a pair of simultaneous equations, the solutions to which are given below:

$$\hat{\mu} = [(\underset{\sim}{z}^t M\underset{\sim}{j}_n) - \hat{\delta}(\underset{\sim}{k}^t M\underset{\sim}{j}_n)]/\underset{\sim}{j}_n^t M\underset{\sim}{j}_n , \quad (20)$$

and

$$\hat{\delta} = \frac{(k^t M z)(j_n^t M j_n) - (z^t M j_n)(k^t M j_n)}{(k^t M k)(j_n^t M j_n) - (k^t M j_n)^2} \qquad (21)$$

Equations (20) and (21) point out that $\hat{\mu}$ and $\hat{\delta}$ are functions of the moving average parameters θ_1 and γ_1 since they depend on $M = \sigma_a^2 \Sigma_Z^{-1}$. However, these estimates are independent of σ_a^2. Note that the main difficulty in obtaining $\hat{\mu}$ and $\hat{\delta}$, for fixed values of the moving average parameters, is the need to find the inverse of Σ_Z.

Maximum Likelihood Estimation of Moving Average Parameters

The procedure used in this section somewhat parallels that presented by Box and Jenkins [2] who treat the non-intervention moving average models and assume $\mu = 0$. Obviously, their procedure needs to be modified.

From the MAMCI(1) model presented in equation (2), we can write down the following (n+1) equations, where the first equation is introduced for convenience:

$$a_0 = a_0$$
$$a_t = Z_t - \mu + \theta_1 a_{t-1}, t=1,\ldots,n_1$$
$$a_t = Z_t - \mu - \delta + \gamma_1 a_{t-1}, t=n_1+1,\ldots,n.$$

By successive substituion of a_1 for a_2 and so on, we can express a in terms of Z and $a_* = a_0$, where this system of (n+1) equations has the following matrix representation:

$$a = LZ + Xa_* - b\mu - c\delta \; . \qquad (22)$$

L is an [(n+1)xn] matrix, while X, b, and c are [(n+1)x1] vectors. Also, L, X, and b are functions of both θ_1 and γ_1 while c is a function only of γ_1. The specific forms of L, X, b, and c can be found in Alt [1].

In making the transformation $a = L^*[a_* | z^t]^t$, where $L^* = [e_1 | L]$, it is easily seen that $|J| = 1$. By substituting equation (22) into equation (17), we see that the joint distribution of Z and a_* is

$$\left.\begin{aligned} &f_{Z^t, a_*}(z^t, a_*; \xi^t) = \\ &(2\pi\sigma_a^2)^{-(n+1)/2} \exp\{-S(\theta_1,\gamma_1,a_*)/2\sigma_a^2\}, \end{aligned}\right\} \qquad (23)$$

where, if we let $d = b\mu + c\delta$,

$$S(\theta_1,\gamma_1,a_*) = (Lz + Xa_* - d)^t (Lz + Xa_* - d) \; . \qquad (24)$$

Define $\hat{a}_*$ to be the value of a_* which minimizes $S(\theta_1,\gamma_1,a_*)$. By taking the derivative of $S(\theta_1,\gamma_1,a_*)$ with respect to a_* and setting this derivative equal to zero, we find that

$$\hat{a}_* = (-X^t L z + X^t d)/(X^t X). \qquad (25)$$

where

$$X^t X = (1-\theta_1^{2n_1})(1-\theta_1^2)^{-1} + \theta_1^{2n_1}(1-\gamma_1^{2(n_2+1)})(1-\gamma_1^2)^{-1},$$

$$\begin{aligned} X^t d = {} & \mu\theta_1(1-\theta_1)^{-1} \sum_{i=0}^{n_1-1} \theta_1^i (1-\theta_1^{i+1}) \\ & + \delta\theta_1^{n_1}\gamma_1(1-\gamma_1)^{-1} \sum_{i=0}^{n_2-1} \gamma_1^i (1-\gamma_1^{i+1}) \\ & + \mu\theta_1^{n_1}\gamma_1(1-\gamma_1)^{-1} \sum_{i=0}^{n_2-1} \gamma_1^i (1-\gamma_1^{i+1}) \\ & + \mu\theta_1^{n_1}\gamma_1(1-\theta_1^{n_1})(1-\theta_1)^{-1} \sum_{i=0}^{n_2-1} \gamma_1^{2i+1} , \end{aligned}$$

and

$$X^t L z = \sum_{i=1}^{n} k_i z_i \; .$$

The k_i's are such that

$$k_i = \theta_1^i (1-\theta_1^{2(n_1-i)})(1-\theta_1^2)^{-1} + \theta_1^i \theta_1^{2(n_1-i)} (1-\gamma_1^{2(n_2+1)})(1-\gamma_1^2)^{-1} ,$$

for $i = 1,\ldots,n_1-1$,

$$k_{n_1} = \theta_1^{n_1}(1-\gamma_1^{2(n_2+1)})(1-\gamma_1^2)^{-1} ,$$

and

$$k_i = \theta_1^{n_1}\gamma_1^{i-n_1}(1-\gamma_1^{2(n_1+n_2-i+1)})(1-\gamma_1^2)^{-1},$$

for $i = n_1+1,\ldots,n$.

By making use of equation (25), we see that $S(\theta_1,\gamma_1,a_*)$ can be rewritten as

$$S(\theta_1,\gamma_1,a_*) = S(\theta_1,\gamma_1) + (a_* - \hat{a}_*)^2 X^t X, \qquad (26)$$

where

$$S(\theta_1,\gamma_1) = [(Lz + X\hat{a}_*) - d]^t [(Lz + X\hat{a}_*) - d]. \qquad (27)$$

Note that $S(\theta_1,\gamma_1)$ is a function of the observations but not of a_*. Since

$$f_{Z^t,a_*}(z^t, a_*; \xi^t) = f_{Z^t}(z^t;\xi^t) f_{a_*|Z^t}(a^* | z^t; \xi^t) ,$$

it follows from equations (23) and (26) that

$$\left.\begin{aligned} &f_{a_*|Z^t}(a_* | z^t; \xi^t) \\ &= (2\pi\sigma_a^2)^{-1/2} |X^t X|^{1/2} \exp\{-(a_* - \hat{a}_*)^2 (X^t X)/2\sigma_a^2\} \end{aligned}\right\} \qquad (28)$$

and

$$\left.\begin{aligned} &f_{Z^t}(z^t; \xi^t) \\ &= (2\pi\sigma_a^2)^{-n/2} |X^t X|^{-1/2} \exp\{-S(\theta_1,\gamma_1)/2\sigma_a^2\} \; . \end{aligned}\right\} \qquad (29)$$

The following deductions can be made

from the foregoing statements:

(i) That $\hat{a}_*$ is the conditional expection of a_* given z and ξ follows from inspection of equation (28).

(ii) Denote $E(a_*|z^t,\xi^t)$ by $[a_*]$. Thus, $\hat{a}_* = [a_*]$.

Since $a = Lz+Xa_*-d$, it follows that

$[a] = Lz+X[a_*]-d$ and that

$$S(\theta_1,\gamma_1) = \sum_{t=0}^{n}[a_t]^2,$$

where $\hat{a}_*$ is obtained from equation (25).

(iii) By comparing equations (18) and (29), we see that

$$|X^tX|^{-1} = |M|$$

and

$$S(\theta_1,\gamma_1) = (z-\mu_z)^t M(z-\mu_z) .$$

Thus, an easy method for finding $|M|$ and evaluating the quadratic form has been provided. Specifically, in order to compute $S(\theta_1,\gamma_1) = \sum_{t=0}^{n}[a_t]^2$, we let $[a_0] = \hat{a}_*$ and recursively calculate the first n_1 $[a_t]$'s from

$$[a_t] = z_t-\hat{\mu}+\theta_1[a_{t-1}] , \qquad (30)$$

for $t = 1,\ldots,n_1$, while the recursive relationship for the last $n_2[a_t]$'s is given by

$$[a_t] = z_t-\hat{\mu}-\hat{\delta}+\gamma_1[a_{t-1}], \qquad (31)$$

for $t = n_1+1,\ldots,n$. The conditional maximum likelihood estimates of μ and δ are given by equations (20) and (21), respectively.

The above results are stated in the following theorem.

Theorem 1: For the MAMCI(1) model, the unconditional likelihood function is given by

$$L(\xi^t|z^t)$$
$$= (2\pi\sigma_a^2)^{-n/2}(X^tX)^{-1/2}\exp\{-\sum_{t=0}^{n}[a_t]^2/2\sigma_a^2\} . \qquad (32)$$

Since X^tX is a scalar, the determinant symbol has been omitted.

Implementing the MLE Procedure

In Theorem 1, a computational form of the likelihood function was given for the MAMCI(1) model. In this section, we present the finer points of implementing the computations.

The problem still remains of finding $\hat{\xi}$ which maximizes $L(\xi^t|z^t)$. Now this maximization procedure can be decomposed as follows:

$$\max_{\xi} L(\xi^t|z^t) = \max_{\theta_1,\gamma_1,\mu,\delta}[\max_{\sigma_a^2} L(\xi^t|z^t)]$$
$$= \max_{\theta_1,\gamma_1}\{\max_{\mu,\delta}[\max_{\sigma_a^2} L(\xi^t|z^t)]\}.$$

Up to now, we have not treated the maximization of L with respect to σ_a^2. However, by finding $\partial \ln L/\partial\sigma_a^2$ and setting this partial derivative equal to zero, we find that

$$\hat{\sigma}_a^2 = \sum_{t=0}^{n}[a_t]^2/n , \qquad (33)$$

which is the maximum likelihood estimate of σ_a^2 for fixed μ,δ,θ_1, and γ_1.

By making use of equation (33) in equation (32), we find that

$$\max_{\xi} L(\xi^t|z^t)$$
$$= \max_{\theta_1,\gamma_1,\mu,\delta}[\max_{\sigma_a^2} L(\xi^t|z^t)]$$
$$= \max_{\theta_1,\gamma_1,\mu,\delta}(2\pi)^{-n/2}(\hat{\sigma}_a^2)^{-n/2}(X^tX)^{-1/2}\exp\{-n/2\}.$$

This last expression is equivalent to

$$\max_{\theta_1,\gamma_1,\mu,\delta}[(\hat{\sigma}_a^2)^{-n/2}(X^tX)^{-1/2}] ,$$

which can be rewritten as

$$\max_{\theta_1,\gamma_1,\mu,\delta}\{\sum_{t=0}^{n}[a_t]^2/n\}^{-n/2}(X^tX)^{-1/2} .$$

In turn, this is equivalent to

$$\min_{\theta_1,\gamma_1}\{\min_{\mu,\delta}[\sum_{t=0}^{n}[a_t]^2/n]^{n/2}(X^tX)^{1/2}\}. \qquad (34)$$

Equation (34) clearly points out the difference between unconditional least squares estimation (UCLSE) and MLE. In UCLSE, one wishes to $\min_{\theta_1,\gamma_1,\mu,\delta}\{\sum_{t=0}^{n}[a_t]^2\}$, which is equivalent to

$$\min_{\theta_1,\gamma_1,\mu,\delta}\{\sum_{t=0}^{n}[a_t]^2/n\}^{n/2} .$$

Thus, UCLSE differs from MLE by the multiplicative effect of $(X^tX)^{1/2}$.

Once that 4-tuple $(\hat{\mu},\hat{\delta},\hat{\theta}_1,\hat{\gamma}_1)$ is found which satisfies equation (34), $\hat{\sigma}_a^2$ is then found using equation (33). The most difficult part of satisfying equation (34) is finding $\hat{\mu}$ and $\hat{\delta}$ since this involves finding M, where $M^{-1} = \Sigma_Z/\sigma_a^2$.

Thus, for each (θ_1,γ_1) pair, it becomes necessary to compute another inverse. For a relatively large time series, this exceeds the capacity of core storage. However, simplifications occur by making use of the patterned structure of Σ_Z.

Details of this can be found in Alt [1].

Additional Statistical Inference

Although ML estimates of the model parameters have been obtained, two inferential questions remain unanswered:

(i) Is the pre-intervention moving average parameter (θ_1) significantly different from the post-intervention moving average parameter (γ_1)?

(ii) Is the shift parameter (δ) significantly different form zero?

The first question can be formulated as a hypothesis testing problem. Specifically, we wish to test

$$H_0:\theta_1=\gamma_1 \quad \text{vs.} \quad H_1:\theta\neq\gamma_1. \qquad (35)$$

To test this hypothesis, we employ the asymptotic chi-squared property of the likelihood ratio test. Let $L(\hat{\Omega}|z^t)$ denote the maximum value of the likelihood function using Theorem 1. Let $L(\hat{\Omega}_0|z^t)$ denote the maximum value of the likelihood function using Theorem 1 under constraint that $\theta_1 = \gamma_1$. This is easily obtained. Define

$$\lambda(z) = L(\hat{\Omega}_0|z^t)/L(\hat{\Omega}|z^t).$$

It can be shown that the distribution of $-2\ln\lambda(z)$ converges to a χ_1^2 distribution when the null hypothesis $(\theta_1=\gamma_1)$ is true. Thus, our decision rule is to reject H_0 when

$$-2\ln\lambda(z)>\chi^2_{1,\alpha}\,. \qquad (36)$$

This decision rule can be restated as reject H_0 when

$$n\,\ln(\hat{\sigma}_a^2)_0+(X^tX)_0-n\,\ln(\hat{\sigma}_a^2)-(X^tX)>\chi^2_{1,\alpha} \qquad (37)$$

If the null hypothesis $(\theta_1=\gamma_1)$ is rejected, one could then set up a pseudo t-test for testing $H_0:\delta=0$ vs. $H_1:\delta=0$ as described in the section on ICLSE; if the null hypothesis $(\theta_1=\gamma_1)$ is not rejected, one could set up a pseudo t-test to investigate the significance of δ under the constraint $\theta_1=\delta_1$.

Example

Consider the data reported by Hall et al [6] which 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 baseline period before the commencement of an intervention effect. Beginning on the twenty-first day, the teacher initiated a program of systematic praise for not talking out.

A preliminary statistical analysis of this data was conducted by Glass, Willson, and Gottman [4], who assumed the single consequence model of equation (1) was appropriate. To check the validity of their assumption, we test $H_0:\theta_1=\gamma_1$ using equation (36) and find $-2\ln\lambda(z)=2.08$, which has an observed significance level of approximately 15%. Thus, we adopt model (1) and find that the maximum likelihood estimates are $\hat{\theta}_1=-.25$, $\hat{\mu}=19.26$, and $\hat{\delta}=-14.33$.

Acknowledgements

Professors Alt and Deutsch would like to acknowledge that their work was partially supported under grant number 75NI-99-0091 from the National Institute of Law Enforcement and Criminal Justice. All the authors express their gratitude to V. Venkata Rao for his computer programming assistance.

References

1. Alt, F. B., Economic Aspects of Control Charts for Multivariate Correlated Observations, Ph.D. dissertation, Georgia Institute of Technology, 1977.
2. Box, G. E. P. and Jenkins, G. M. Time-series analysis: forecasting and control. San Francisco: Holden Day, 1970.
3. Box, G. E. P. and Tiao, G. C. A change in level of a non-stationary time-series. Biometrika, 1965, 52, 181-192.
4. Glass, G. V.; Willson, V. L.; and Gottman, J. M. Design and analysis of time-series experiments. Boulder: Colorado Associated University Press, 1975.
5. Goldberger, A. S. Econometric theory, New York: John Wiley & Sons, Inc., 1964.
6. Hall, R. V.; Fox, R.; Willard, D.; Goldsmith, L.; Emerson, M.; Owen, M.; Davis, F.; and Porcia, E. The teacher as observer and experimenter in the modification of disputing and talking-out behaviors. Journal of Applied Behavior Analysis, 1971, 4, 141-149.

END