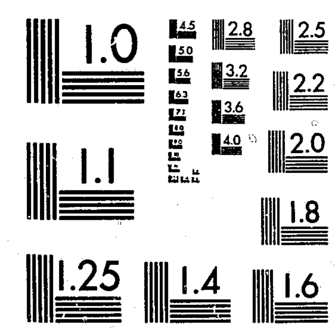

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

7/29/83

88425

Industrial and Systems Engineering
Report Series No. J-79-24
August, 1979

#14 Final Copy

IDENTIFICATION AND MODELING OF DYNAMIC INTERVENTION EFFECTS ASSOCIATED WITH PUBLIC POLICY DECISIONS

by

Stuart Jay Deutsch

Lu Ann Sims*

Philip E. Pfeifer

Georgia Institute of Technology
Atlanta, Georgia 30332

* Project Engineer, Industrial Handling Engineers, Inc., Houston, Texas 77092

78-NI-AX-0040

IDENTIFICATION AND MODELING OF DYNAMIC INTERVENTION EFFECTS ASSOCIATED WITH PUBLIC POLICY DECISIONS

BY

STUART JAY DEUTSCH*
LUANN SIMS**
PHILLIP E. PFEIFER***

SUMMARY

The nature of an evaluation and subsequent policy decisions in intervention analysis impinges directly on the descriptive model form and its associated parameter values. Previous modeling efforts using dynamic intervention modeling have relied on a priori postulations of the model form thus perhaps incorporating an experimentor's subjective biases. An algorithm for the identification of a tentative model form from the data alone is reviewed. Previous intervention analyses not using an identification phase are revisited to illustrate the differences in resulting models and associated policy evaluations. The first example deals with controls applied in the Los Angeles County area circa 1960-1966 to help alleviate a serious pollution problem evidenced by rising levels of ozone [3]. The second illustration is directed to the evaluation of changes of property or life loss associated with motor vehicle crashes after implementation of a traffic safety program [7].

INTRODUCTION

In quasi-experimental designs the data often takes the form of autocorrelated time series. Two methodological approaches to analyzing such data is that proposed by Box and Tiao in 1965 [2] and Box and Tiao in 1975 [3]. These methods although related offer fundamentally different information with regard to measurement of change.

In the former the detection of a shift or change in the level of a nonstationary time series is presented. Here the problem is formulated in a quality control framework in which the historical data y_t, $t=1,2,3...T$ is autocorrelated and described by a model from the autoregressive moving average class [1]. The model is augmented by a parameter, δ, which measures a change in level for the process for new data that is monitored from the process in times $t > T$. From the $(1-\alpha)$ 100% confidence interval for δ, significant changes in the level of the process can be made (e.g., the transition point for the in control to out of control process state identified). Thus the first of two questions of intervention analysis "Did a change take place?" is directly addressed.

*Professor, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332
**Project Engineer, Industrial Handling Engineers, Inc., Houston, Texas 77092
***Research Engineer, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

The second question in intervention analysis is "If a change in the process level has occured, how is this change manifested?" That is, from a policy decision standpoint, a temporary change in which the new system level reverts to the pre-intervention level is quite different than a situation in which the new post-intervention level is maintained over time--a new steady state process level is reached after intervention. The latter Box-Tiao work [3] proposes directly in an intervention framework, dynamic intervention models for analyzing time series data to make inference about the second question in intervention analysis.

Here the t^{th} observation, y_t, is proposed as an additive sum of the dynamic intervention component, D_t, and the noise component, N_t. The noise components are described by the flexible autoregressive moving average model class while the D_t may take any form to reflect the actual time patterned change in the process mean. Several representative examples of D_t for different order dynamics and system delays when the system is subject to an intervention in the form of a pulse or step function were illustrated [3].

The form of D_t impacts directly upon policy inferences. An identification algorithm for the form of D_t utilizing the sequential application of the Box and Tiao (1965) [2] method for detecting a change in the level of a nonstationary series has been suggested [4]. The purpose of this paper is to illustrate the differences in policy inference from intervention modeling when identification procedures are not employed. In the following sections the identification algorithm is overviewed and several examples of intervention modeling are revisited.

AN IDENTIFICATION ALGORITHM

The shift detection method of Box and Tiao [2] can be used to identify an appropriate dynamic model component D_t by estimating the shift parameter δ. This means that one must think of delta not strictly as a parameter, but as a time-dependent variable in the model, denoted by δ_t. In the following paragraphs, an algorithm for identifying forms of intervention effects utilizing successive estimates of δ_t over time is overviewed.

To estimate delta at any point in time, we must first specify that time, which will be denoted by T. This is equivalent to specifying n_1, the number of observations before T and n_2, the number of observations after time T, where $n_1 + n_2 = N$ and N is the total number of observations used in the estimator. The quantities n_1 and n_2 enter directly into the estimator as do the observations associated with n_1 and n_2. Thus, by specifying n_1 and n_2, we effectively specify which observations will enter into the estimator as historical (before T) data and those which will be used as "future" (after T) data. Of course, the estimator, δ_t, is also functionally dependent on the form of the ARIMA model noise term and the associated moving average and/or autoregressive parameter values [5,6]. For our purposes, we will denote the ARIMA model form as M and the parameters from this model collectively as Λ. Thus we can write,

$$\hat{\delta}_t = f(n_1,n_2,Y,M,\Lambda).$$

In sequential estimation, we must monitor and correct for significant shifts if the sequential plot of $\hat{\delta}_t$ is to mimic the dynamics of intervention. To translate what it means to use past significant values of delta in the actual estimation procedure without correction, we refer to the functional description of $\hat{\delta}_t$ given previously. We know that if we estimate δ_t at time T, $\hat{\delta}_T$, with $n_2 = 1$, we are using the n_1 observations before time T and an additional observation, y_{n_1+1}. If we let Y_{n_1} denote the observations occurring before time T, we can write:

$$\hat{\delta}_T = f(n_1,n_2,Y_{n_1},y_{n_1+1},M,\Lambda).$$

Suppose we estimate δ_T and find that it is significant, i.e., not statistically equal to zero. Then, in effect we are saying that there has been an intervention effect at time T which is evidenced by the change from the previous n_1 observations to the observation y_{n_1+1}. If we move ahead and stand at time T+1, again letting $n_2 = 1$, we thus have $n_1^* = n_1+1$, where n_1 was the prevous value of n_1 used to estimate δ_t at time T. Now, an additional observation y_{n_1+2} or $y_{n_1^*+1}$ will be used, and y_{n_1+1} will be grouped with the n_1 observations (Y_{n_1}) to form the set of n_1^* observations. However, since we have already concluded that an intervention effect of significant magnitude occurred at time T, the set of n_1^* observations which we are comparing the $y_{n_1^*+1}$ observation to is not internally consistent. That is, the n_1^* set of observations does not represent a single population. To form an historically consistent population we subtract the previous significant value of δ_t, $\hat{\delta}_T$, from the y_{n_1+1} observation to account for the difference in level with the previous observations. We can then proceed to estimate δ_{T+1}. At this point in time, we can write,

$$\hat{\delta}_{T+1} = f(n_1,N_2,Y_{n_1^*},y_{n_1+2},M,\Lambda)$$

where $Y_{n_1^*}$ includes the original n_1 observations and the corrected n_1+1^{st} observation. Of course, if $\hat{\delta}$ was not statistically significant, then no adjustment is needed since δ_T is effectively zero and there is already consistency between the n_1 observation and the n_1+1^{st} observation. For subsequent estimates of δ_t, we use the same procedure as we move ahead to time T+2, T+3, T+4, etc. The estimation procedure is depicted in flow chart form in Figure 1.

Two points should be considered with respect to the ability of the pattern produced to reliably mimic the dynamic behavior. First, the pattern produced will be fuzzy if in fact the shift parameter estimator is insensitive in the sense of requiring large amounts of post-intervention data (large n_2 values) in order to assess significant changes. This however is not generally the case. In a previous effort, [5] a sensitivity study was conducted of the shift detection method. It was found that shifts as small as two percent of the initial level of the time series could reliably be detected under a wide range of all other factors for $n_2 = 1$.

A second consideration that will distort the sequential pattern of $\hat{\delta}_t$ is the α level of significance chosen. Larger values of α will force significance of a given estimate and therefore necessitate correction of the level of the y_{n_1+1} observation which impacts upon subsequent estimation. In practice, the

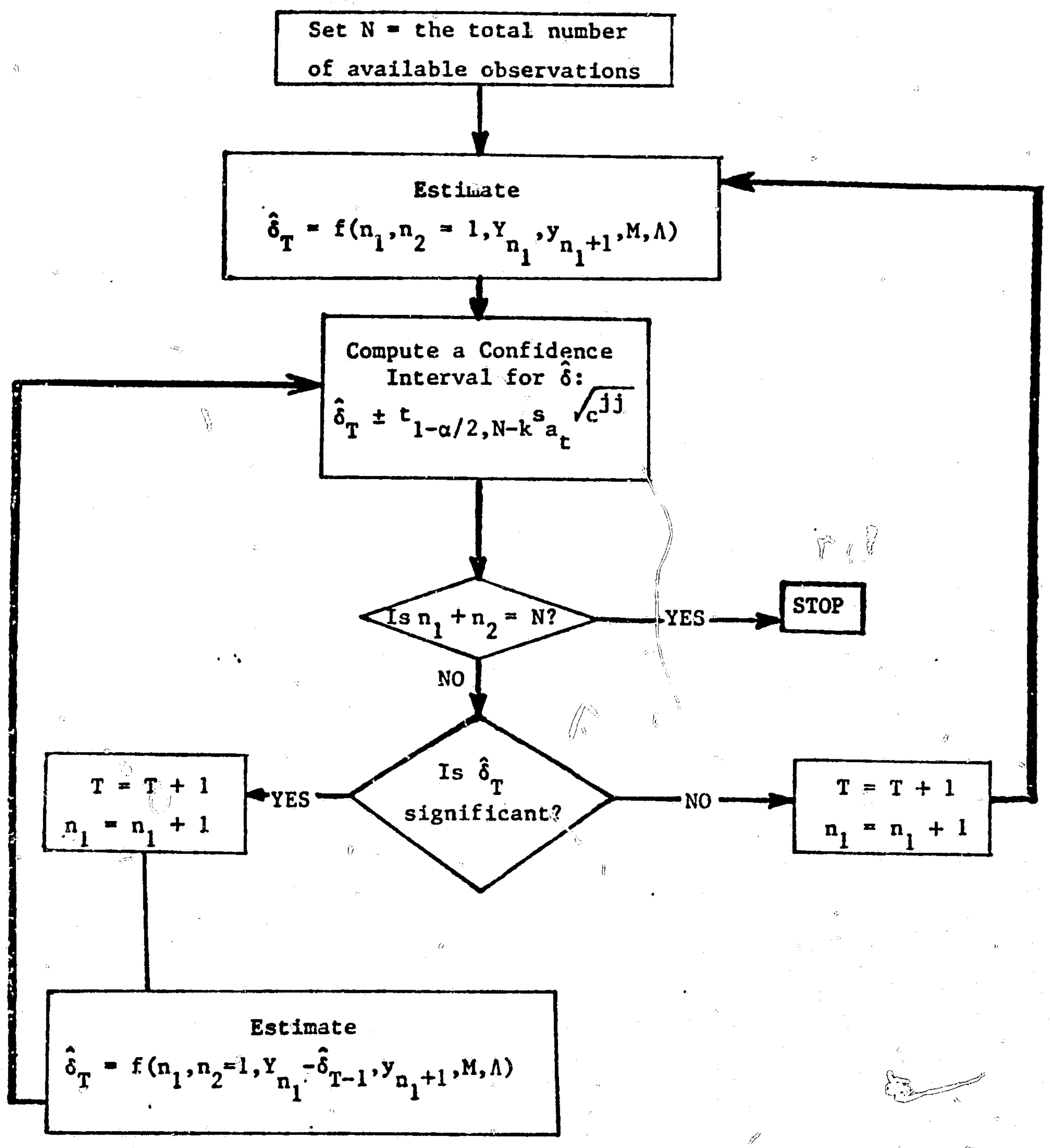

Figure 1. Flowchart for Identification Algorithm

algorithm should be applied to the data set while iteratively reducing the α level. The pattern least sensitive to a single significance decision should be used for tentative identification of the dynamic component D_t.

EXAMPLES

To illustrate the importance of using an identification procedure we revisit data previously modeled by Box and Tiao [3]. In light of increasing awareness of the harmful effects of atmospheric pollutants on human, animal and plant life, Los Angeles County has taken several measures intended to reduce the abnormal amount of substances which can produce dangerous pollutants when subjected to the high intensity sunlight in the Los Angeles vicinity. In early 1960, Rule 63 was passed which reduced the allowable proportion of reactive hydrocarbons in locally sold gasoline. Also in 1960, the Golden State Freeway was opened, which diverted traffic from the downtown area. An additional anti-pollution regulation was effective from 1966 onwards, which required engine design changes in new cars. The federal measure might be expected to reduce the production of ozone in Los Angeles. The two regulations, along with the opening of the Golden State Freeway, will hereinafter be referred to as interventions. Figure 2 illustrates this Los Angeles ozone data.

The Box-Tiao analysis was conducted in three segments. First, a noise component, N_t in the form of an ARIMA model was identified from the historical data prior to the first intervention date. Secondly, the form of the dynamic component, D_t, was postulated to reflect the modeler's expectation of the intervention effect. Lastly, the parameters of both components were simultaneously estimated and the residuals of the overall model checked.

The time series model identified was:

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

where N_t is the noise or underlying noise structure for the intervention model and a_t is white noise or an uncorrelated, normally distributed series with mean 0 and variance equal to σ^2.

Because of the summer-winter atmospheric temperature inversion differential and the difference in the intensity and duration of sunlight for the summer and winter, the effect of the intervention was specified to be different in the summer months (roughly June-October) and winter months. Therefore, the following model was postulated for the entire series from January 1955 to December 1972:

$$y_t = \omega_{01}\xi_{t1} + \omega_{02}\frac{\xi_{t2}}{(1-B^{12})} + \omega_{03}\frac{\xi_{t3}}{(1-B^{12})} + \frac{(1-\theta_1 B)(1-\theta_{12}B^{12})}{(1-B^{12})}a_t$$

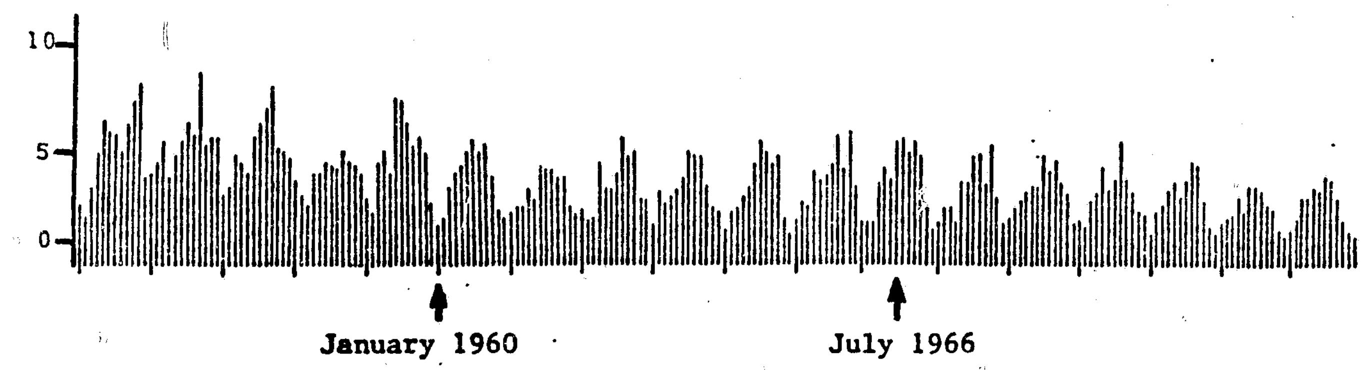

Figure 2. Los Angeles Photochemical Smog Data (Ozone in Parts Per Million)

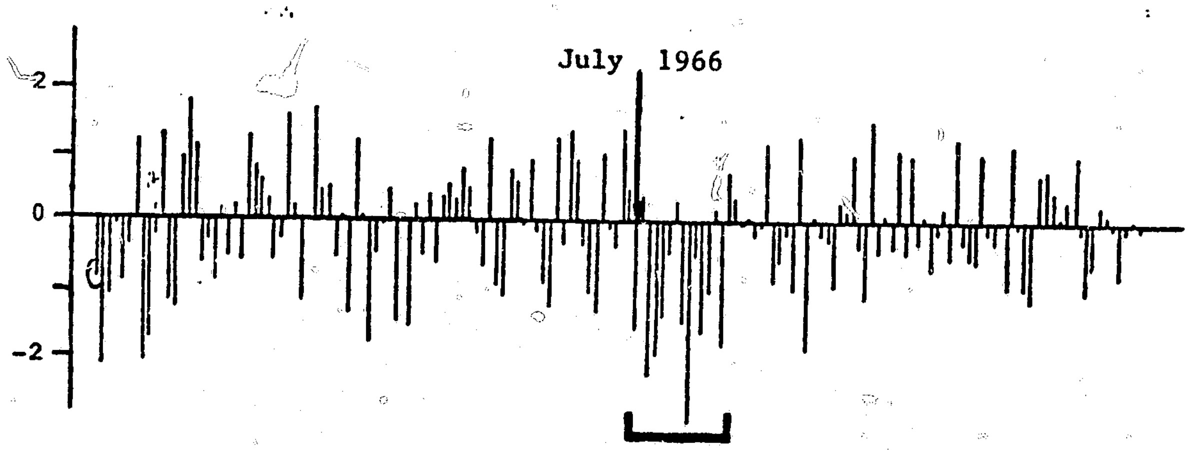

Figure 3. Output from Identification Algorithm $\hat{\delta}$ Versus Time for α = 0.2

where:

$$\xi_{t1} = \begin{cases} 0 & t < \text{January 1960} \\ 1 & t \geq \text{January 1960} \end{cases}$$

$$\xi_{t2} = \begin{cases} 1 & \text{"summer" months June-October beginning 1966} \\ 0 & \text{otherwise} \end{cases}$$

$$\xi_{t3} = \begin{cases} 1 & \text{"winter" months November-May beginning 1966} \\ 0 & \text{otherwise} \end{cases}$$

The estimation results and standard errors for this analysis are:

$\hat{\omega}_{01} = -1.09$	0.13	standard errors
$\hat{\omega}_{02} = -0.25$	0.07	
$\hat{\omega}_{03} = -0.07$	0.06	
$\hat{\theta}_{1} = -0.24$	0.03	
$\hat{\theta}_{12} = 0.55$	0.04	

These results led to the conclusion that a step change of -1.09 units of ozone may be associated with Rule 63. A progressive reduction from year to year of -.25 units in the summer months is expected, and there is little or no change in the winter months.

In the reanalysis three steps are also employed. The first step, the identification of the noise component from the historical data is identical with one exception. The exception is the a priori specification of the intervention date as the point in time in which the process did go out of control. In our case we employ the time point where a measurable shift in the time series data is observed as the point for which a homogeneous historical data series is available for identification of the noise component.

The second step we propose as an alternative is fundamentally different in that we identify the form of the dynamic component from the data. Figure 3 exhibits the resulting pattern of $\hat{\delta}$ versus time for the smog data. It should be noted that only at July 1966 is there an observed significant sequential pattern. This pattern persisted through August 1967 and led to the tentative identification of the dynamic component D_t to be:

$$D_t = \omega B S_t \qquad S_t = \begin{cases} 1 & \text{July 1966—August 1967} \\ 0 & \text{otherwise} \end{cases}$$

The third step is identical to that used by Box and Tiao and consists of simultaneous estimation and diagnostic checking. The ninety-five percent confidence interval for ω was [-1.523,0.727], therefore eliminating any significant dynamic component terms. Thus the overall model for the evaluation of the interventions is,

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

where $\hat{\theta}_1 = 0.718$ and $\hat{\theta}_{12} = 0.579$, which were determined to be jointly significant at the ninety-five percent level.

Comparison of the two resulting models, clearly indicates different conclusions with respect to the value of the interventions in reducing ozone level. Using an identification scheme for the dynamic component, the decision-maker concludes no effect as opposed to having observed a successful policy intervention of reduced ozone level.

The second example involving the impact of a traffic safety program must ultimately be assessed in terms of reduced losses associated with motor vehicle crashes. Previous analysis of the traffic loss data was done by Ellingstad and Westra [7]. The thrust of their paper was to explore a variety of procedures for assessing the impact of TSP's on traffic loss or crash time series data. The two primary recommendations of the authors were: (1) Forecast the baseline (pre-TSP) data into the post-TSP period and compare the actual observations when they become available to the statistical forecasts; (2) Obtain a measure of change parameters associated with either drift or a change in level (shift). However, their analysis was not reported as to how to model and therefore did not contain any statistical inference.

Two sets of relevant monthly data were available for the analysis. These sets are: South Dakota total injury accidents and North Dakota injury accidents (NDACC) for the time period January 1969 to December 1974. The implemented traffic safety program was the South Dakota Alcohol Safety Action Project, implemented in January 1972. The set of South Dakota data may be used to test whether the traffic program had an actual effect on overall South Dakota traffic losses, while the North Dakota data will serve as a control for the experiment.

Analysis of the control series which represents the North Dakota data revealed a significant process shift at t=44 or September 1972 which exponentially decayed. Figures 4 and 5 illustrate the data base and the corresponding plot of $\hat{\delta}$ versus time used in identification of the dynamic component respectively. Thus this patterned shift suggested a dynamic component,

$$D_t = \frac{\omega B}{(1-\delta B)} P_t \qquad \text{where } P_t = \begin{cases} 1 & \text{September 1972} \\ 0 & \text{otherwise} \end{cases}$$

The final statistically adequate model result is

$$y_t = \frac{110.354\ B}{(1-0.939\ B)} P_t + \frac{(1-0.582\ B)}{(1-B)} a_t \quad ,$$

which indicates the control series to have exhibited a significant although not permanent increase a few months after implementation of the experimental program.

The treatment series for the parallel South Dakota accident data did not indicate any significant process shift after program implementation, therefore no dynamic component is required. In this present example we see that the

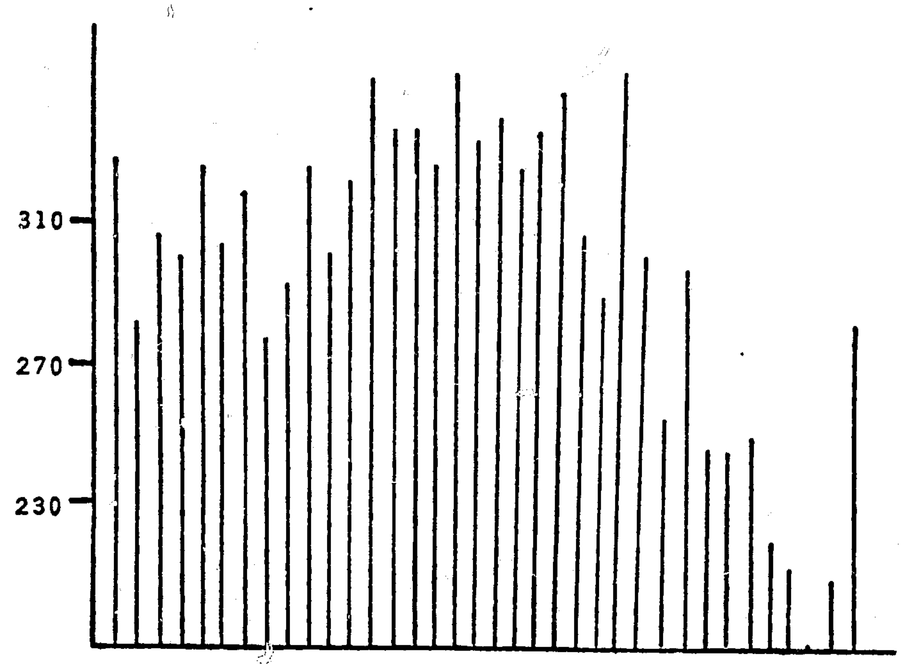

Figure 4. NDACC Data

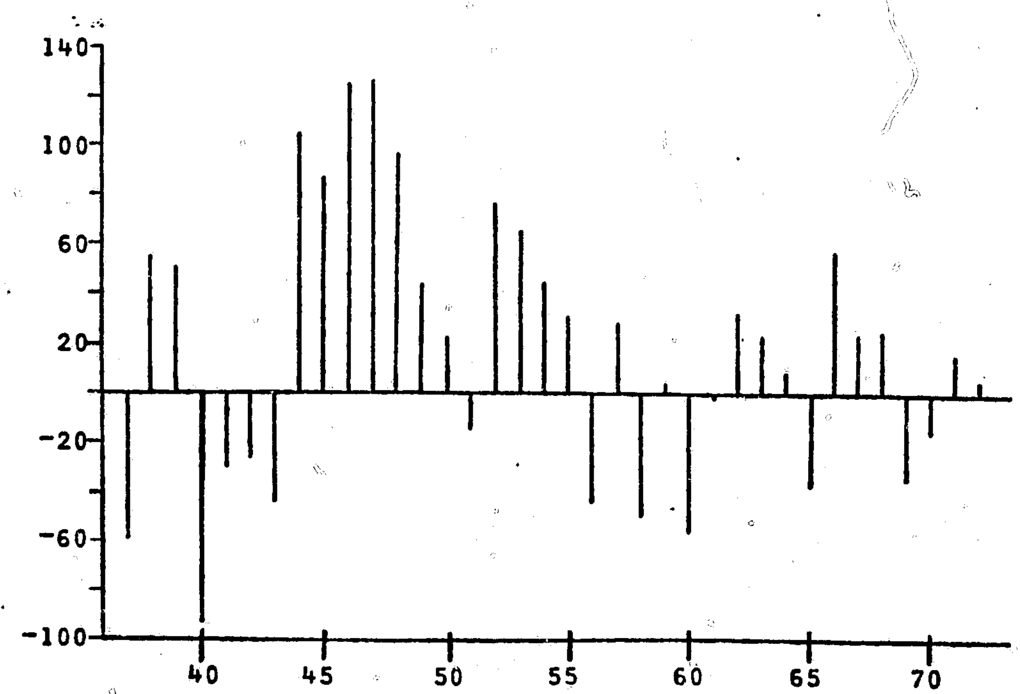

Figure 5. Output from Identification Algorithm for North Dakota Data ($\hat{\delta}$ Versus Time for $\alpha = 0.05$)

identification algorithm allows for development of models for precise statistical inferences with respect to each individual series. However further reconcilliation with an expanded data base would be desirable before tying down the policy decision with respect to the value of the implemented South Dakota accident program.

CONCLUSIONS

An identification algorithm for dynamic intervention models was described and illustrated. The differences that may result in policy decisions when an investigator specifies a priori the effect of the experiment versus allowing the data to "speak for itself" via the identification procedure was also illustrated.

ACKNOWLEDGEMENT

This work was performed under Grant No. 78-NI-AX-0040 from the National Institute of Law Enforcement and Criminal Justice. Points of view or opinions stated are those of the authors and do not necessarily represent the official position or policies of the United States Department of Justice.

REFERENCES

[1] Box, G. E. P. and Jenkins, G. M., Time Series Analysis, Forecasting and Control, San Francisco: Holden-Day, 1970.

[2] Box, G. E. P. and Tiao, G. C., "A Change in Level of a Non-Stationary Time Series," Biometrika, 52 (June 1965), pp. 181-192.

[3] Box, G. E. P. and Tiao, G. C., "Intervention Analysis with Applications to Economic and Environmental Problems," Journal of the American Statistical Association, March 1975, Vol. 70, No. 349, pp. 70-79.

[4] Deutsch, S. J., "The Identification of Dynamic Intervention Models with Application to Gun Control," Industrial and Systems Engineering Report Series, No. J-79-23, Georgia Institute of Technology, Atlanta, Georgia, 1979.

[5] Deutsch, S. J. and Alt, F. B., "Estimation of Shifts in Stochastic Models of Crime Occurrence," Proceedings of the 7th Annual Pittsburgh Conference on Modeling and Simulation, April 26-27, 1976, part 2, pp. 780-785.

[6] Deutsch, S. J. and Alt, F. B., "The Effect of Massachusetts' Gun Control Law on Gun-Related Crimes in the City of Boston," Evaluation Quarterly, Vol. 1, No. 4, November 1977, pp. 543-568.

[7] Ellingstad, V. S. and Westra, D. P., "Evaluation Methodologies for Traffic Safety Programs," Human Factors, Vol. 18, No. 4, 1976, pp. 313-326.

END