THE UNIVERSITY OF ALBERTA

INTERVENTION EFFECT ANALYSIS IN TIME SERIES PROCESSES

by

LOUISE A. JENSEN

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1989

ISBN 0-315-52861-3

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: . . . . LOUISE A. JENSEN . . . .

TITLE OF THESIS: . . INTERVENTION EFFECT ANALYSIS IN TIME SERIES

. . . . PROCESSES . . . .

DEGREE: . . DOCTOR OF PHILOSOPHY . . . .

YEAR THIS DEGREE GRANTED: . . 1989 . . . .

. . . . . Louise A. Jensen . . . . . .
(Student's signature)

. . . . . 2303 - 117 A Street . . . . . .
(Student's permanent address)
. . . . . Edmonton, Alberta T6J 4X6 . . . . .

Date: March 20, 1989

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled INTERVENTION EFFECT ANALYSIS IN TIME SERIES PROCESSES . . . .

. . . .

. . . .

submitted by . . . . LOUISE A. JENSEN . . . .

in partial fulfillment of the requirements for the degree

of . . . . DOCTOR OF PHILOSOPHY . . . .

in EDUCATIONAL PSYCHOLOGY

Thomas O. Maguire . . . . .
(Supervisor)

S. Hunka . . . .

C.B. Hazlett . . . .

. . . .

. . . .

Date: March 20, 1989 . . .

ABSTRACT

Much of statistical methodology is concerned with models in which the observations are assumed to vary independently. Randomization of the experimental design is introduced to validate analysis conducted as if the observations were independent. However, a great deal of data occur in the form of time series where observations are dependent and where the nature of this dependence is of interest. The technique available for the analysis of such series of dependent observations is called time series analysis. Time series models are techniques that allow the researcher to identify the structure of a time series and to determine if a discrete intervention accounts for a statistically significant change in the level of the series without artificial experimental conditions.

Monte Carlo studies were used to analyze issues in time series procedures with small data sets. Five Autoregressive-Integrated Moving Average (ARIMA) models with 20 and 40 data points were generated; a constant intervention effect was added to each time series; values of the correlation and intervention parameters were varied. The size of the intervention effect and the bias in intervention effect estimates were calculated for the true and misidentified ARIMA models. A second set of Monte Carlo simulations was used to investigate the procedures used in the model identification stage of time series analysis. ARIMA model identification is a crucial step in the assessment of intervention effects in interrupted time series experiments.

The results indicate that correctly identified ARIMA models gave fairly accurate estimates of the intervention effect. However, the

length of the time series realization plays a crucial role in determining the accuracy of estimates examined in this investigation. The magnitude of the standard errors, the inaccuracy of the estimated standard errors, inflation of Type I error rates, and lack of power, are quite severe in short time series realization. Also, extreme serial dependence magnifies the problems observed in estimation procedures of the autocorrelation function as well as the intervention component. Intervention effect estimates were inaccurate when the ARIMA model was inadequately differenced, having a detrimental effect on power.

Time series analysis techniques provide the tools for analyzing unique behavioral fluctuations through time and a framework for predicting future changes in the individual. The inherent limitations in the statistical procedures will be helpful in applying time series analysis techniques to research problems. The application of time series analysis to clinical research may provide a scientist-practitioner model for developing knowledge.

ACKNOWLEDGEMENTS

I am fortunate to have had the opportunity to know and work with Dr. T.O. Maguire, a most effective teacher — his valuable support, insight, and guidance contributed greatly to this study. My professional and personal respect for him continues to grow. His interest in time series analysis and faith in my ability were appreciated.

To the other members of my committee, Dr. S.M. Hunka, Dr. C.B. Hazlett, Dr. G.M. Kysela, and Dr. W. Muir, I am indebted for their assistance and understanding. To Dr. S.M. Hunka for [illegible] his insights and patience, to Dr. C.B. Hazlett for his appraisal and attention to detail, as well as being a role model as a social scientist, to Dr. G.M. Kysela for his valuable feedback, and to Dr. W. Muir for his enthusiasm and interest, my gratitude is expressed. Serving as a committee member is only part of each individual's contribution to my experience as a graduate student.

A special thanks is due to Dr. D. Harley, Division of Educational Research Services, who provided the expertise in programming. I gratefully acknowledge the computer funds made available for conducting the simulation procedures.

I am particularly grateful to Tom Hall, whose typing talents contributed to the appearance and accuracy of the manuscript.

Finally, I thank my friends and colleagues in the Faculty of Nursing. Their support provided the sustaining environment from which accomplishment is made possible and without which, accomplishment would mean little.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

The science of nursing becomes explicit when research is used to guide and enhance practice. In generating knowledge for practice, the discipline of nursing borrows and blends methods of inquiry from those of related disciplines. The present study falls within this tradition. It is an investigation into the utility of the interrupted time series design for the field of nursing. The interrupted time series design involves a series of observations taken at regular intervals, during which an interruption or treatment is administered. The purpose of the interrupted time series experiment is to determine if the treatment had an impact on the measures or observations. The interrupted time series design has been widely used as a research paradigm by social scientists in such areas as educational psychology (Berryman & Cooper, 1982; Kratochwill, 1978), behavioral psychology (Barlow & Hersen, 1973, 1984; Kazdin, 1984), psychology (Gregson, 1987; Larsen, 1987; Nurius, 1983), sociology (Berman, Meyer, & Coats, 1984; Blose & Holder, 1987; Calsyn, 1977), economics (Dalton & Todor, 1984), and law (Glass, 1968; Glass, Tiao, & Maguire, 1971). Campbell and Stanley (1966) and Cook and Campbell (1979) have described some of the benefits that this design has as an alternative to the "experimental-control" design. Among these benefits are: time series quasi-experiments allow researchers to meaningfully interpret data that are collected in the absence of rigorous control over variables necessary in true experimental designs or when the feasibility of comparison groups is questionable; time series experiments permit hypothesis testing of treatment effects in studies involving only a single subject or unit of observation; interrupted time series experiments provide information concerning the nature of the intervention effect over a period of time.

Technically speaking, a time series is a sequence of time-ordered observations (Y_t) of some underlying process; the interval between Y_t and Y_{t-1} and the sources of data are assumed to be fixed and constant. Examination of unique fluctuations in phenomena tracked through time can be modeled using stochastic processes. A stochastic process describes an underlying process of unobserved errors which make the observed time series unpredictable. Thus a stochastic process may be defined as a collection of random variables which are ordered in time according to certain probabilistic laws. Inferential statistics assume unique fluctuations to be random error and may hide the significance of individual differences in predicting behavior. Human behavior cannot always be explained by a generalized mean observation. If sequences of observations proceed through time according to probabilistic laws, then the future evolution of a behavioral series may be predicted by knowledge of its past values if the series is stable and can be modeled accurately.

In time series designs, a series of measures is assessed on a single variable over a period of time prior to some intervention. The same variable is then measured over time subsequent to the intervention. The hypothesis under consideration concerns the impact of the intervention, which is evaluated by comparing the pre-intervention time series with the post-intervention time series. The research design is conceptually simple, and researchers may be tempted to use a t-test to compare the mean of measures collected before the intervention with the mean of those collected after the intervention. However, due to the non-independence of the measures assessed over time, the results of such a procedure can be fallible. While ordinary least squares regression estimates of time series parameters are not biased per se, the estimates of standard deviations and hence, of significance tests are biased whenever error terms are correlated. Thus, when naturally occurring events are observed repeatedly over time, events closer to each other in time tend to be more correlated with each other than with events further removed in time. Since time is the independent variable of an ordinary least squares regression, it follows that the error terms of consecutive observations may be correlated. Consequently, time series analysis offers a method of incorporating this correlation into the model, allowing for the calculation of unbiased estimates of standard deviations in intervention effect estimation. Models which take the interdependence of measures into account are necessary in order to draw inferences with any degree of confidence on the basis of data collected in this type of study.

In addition to the statistical problems associated with a simple comparison of pre- and post-intervention means, there are logical problems with the procedure. For example, a time series process that follows a steady upward trend will result in a post-intervention mean that is substantially larger than the pre-intervention mean. A conclusion that the intervention is responsible for the difference would be illogical however, because the post-intervention mean would be greater in the absence of the intervention as a result of the upward trend. In other instances, the equality of pre- and post-intervention means may lead the researcher to the false conclusion that the intervention had no impact. This situation may occur if a time series process follows an upward trend prior to the intervention, and the intervention results in a downward trend during the post-intervention phase. This dramatic intervention effect would not be evident if the researcher simply compared pre- and post-intervention means.

In summary, an interrupted time series analysis assesses the magnitude and statistical significance of change in the time series following an intervention (Cook & Campbell, 1979). This change is of two types, deterministic and stochastic. The deterministic component describes the systematic behavior of a time series and is not dependent on error. The stochastic component, the error structure in a time series, accounts for the unpredictability of change through time. Statistical analysis of a time series creates a model of the structure of the systematic stochastic component. Departures from this model behave like independent random events. The remaining error allows for calculation of an unbiased estimate of the standard deviation. As a result, more accurate inferences can be drawn from significance tests of a parameter representing the change associated with the interruption in the time series.

In contrast to the time series design, the classical experimental approach is costly and if properly used, may ignore the changes that can occur over the normal course of time or changes prompted by the act of measurement. Unfortunately, either the assumptions of the statistical models for "true experiments" are too rigid for use in most clinical paradigms or these assumptions are violated, with possible resulting false conclusions. Interrupted time series experimentation is proposed as a viable approach for clinical nursing research. Time series analysis techniques provide the tools for analyzing behavioral fluctuations through time and a framework for predicting future changes in the individual. This basic ideographic technique has many implications for clinical research.

Classical time series analysis procedures have not been applied in nursing research, yet observations over time are often basic features of clinical studies. Questions related to the outcome of this investigation are: When is time series analysis appropriate to test for the presence of intervention effects? What time series models fit the data in nursing? Can changes of the order found in nursing research be detected by interrupted time series experiments? What is the form of change? In this investigation, the appropriate procedures for the analysis of data from interrupted time series experiments are assessed, and potential difficulties pointed out that may be encountered in applying these procedures to "real life" data sets.

Objectives of the Study

The purpose of this study was to explore intervention effect analysis in short time series processes and subsequent application of the interrupted time series experiment in the context of clinical nursing research. There are a variety of practical issues that are important in the utilization of the interrupted time series experiment. Issues in the application of time series analysis procedures to small samples under a variety of conditions were explored.

The objectives of the study were to:

1. examine time series experiments in comparison to traditional experimental designs employed in clinical research;
2. investigate intervention effect estimation in short time series processes; and
3. explicate the issues in the application of interrupted time series experiments in the context of clinical nursing research.

The discussion in subsequent chapters deals with the basic statistical procedures necessary to model time series processes and to test for the effect of interventions. Additionally, the benefits and limitations encountered in the application of time series designs and analysis are explored, specifically in the context of clinical research. Statistical inferences regarding intervention effect estimation in interrupted time series experiments were investigated empirically via computer simulations.

Significance of the Study

The use of interrupted time series experiments has many implications for nursing practice. The concept of uniqueness of the individual is not only allowed to exist as a basis for nursing care but in fact is central to most developing nursing theories; thus uniqueness provides a framework for assessing change, not as a deviation from an aggregate mean, but as an alteration in a consistent pattern. In time series data sets, successive values are assumed to be related to each other so that in this respect the model is a plausible representation of reality. In the clinical research context, this provides a tool for assessing change statistically, so that decisions can be based upon objective criteria.

The goals of measuring change are to determine whether there is an intervention effect, and what is the functional form of that effect.

A time series approach to clinical research takes advantage of the sequence of the nursing process. Time series analysis techniques could provide a ' ~is for computer-assisted clinical decision-making (Metzger & Schult.)82). Time series analysis offers nurse researchers a technique that is adaptable to clinical research because intervention strategies can be evaluated in a normal environment without artificial experimental conditions. The interrupted time series experiment provides techniques that allow the researcher both to describe the structure of processes and to determine if a discrete intervention accounts for a statistically significant change in the series.

The results of clinical research add to nursing knowledge, providing a basis for clinical practice. How applicable are research findings to the clinical decision-making process? How may research generate appropriate knowledge for clinical utilization? Nursing research efforts have not exploited the full range of potential procedures for empirical investigation. For the development of a scientist-practitioner model of process research, these are concerns warranting attention. Nursing is concerned with individual as well as aggregate responses to health problems. In addition, there is an expressed commitment to understanding the whole person and thos• factors that influence health status. The study of an individual within an intensive, longitudinal perspective, maximizing internal and external validity, may assist in bridging the research-practice gap (Meier & Pugh, 1986). The time series approach is well suited to the humanistic validity-seeking method of nursing research. To address the issues of interrupted time series experiments, it is important to begin with the basic underlying statis-tical procedures that are necessary to model time series processes and to test for the effect of interventions.

CHAPTER II

Time Series Processes

The subject of the present investigation impinges on a broad range of theoretical and empirical references. To provide clarity in the review of relevant literature, the content is divided into three sections. An overview of time series processes is presented in the first section. The second section is devoted to time series model building, while the third section is a presentation of the interrupted time series experiment.

Overview of Time Series Processes

Time series data occur in many fields of study providing opportunity for the application of an experimental or quasi-experimental time series design. These designs and threats to their validity are discussed extensively in Campbell and Stanley (1966), Cook and Campbell (1979), and Glass, Willson, and Gottman (1975). Sources of invalidity in time series experiments are: an event extraneous to the intervention but coincident with it may produce an alteration of the series; interventions may come about as reactions to past or impending changes in the time series; false attribution of an effect to an intervention may occur when in fact it is due to the intervention plus a previous intervention; change in the method of observing the outcome variable may cause an abrupt change in a time series; random variation in a time series may be misinterpreted as the effect of an intervention; composition of the experimental unit may change over time; events unrelated to the intervention may cause the time series to change abruptly at the point of intervention; and results may be inappropriately generalized from the observed time series to subjects other than the ones involved in the experiment. According to McDowall, McCleary, Meidinger, and Hay (1980), an interrupted time series quasi-experiment can be diagramed as

. . . O O O O O O O O X O O O O O O O O O . . .

where O represents the observation point and X denotes the intervention. Time series analysis has several strengths (Velicer & McDonald, 1984). First, time series designs can be employed in clinical situations where traditional between-subject designs are difficult to implement. Second, time series designs are appropriate for dealing with effects of change. Third, time series designs permit the study of the pattern of intervention effects through the evaluation of change in the level, or in the slope, or both (Jones, Vaught, & Weinrott, 1977). The evaluation of only simple mean changes before and after intervention could obscure other intervention effects.

The analysis of time series data requires different considerations than generally encountered in traditional data analysis procedures. The distinguishing aspect of the structure of time series data is the dependency among observations. Most inferential statistical models are based on the premise that observations are independent, or uncorrelated, with each other. This basic assumption is seldom met when data are collected on the same experimental unit across time (Jones, Weinrott, & Vaught, 1978). Usually, observations are related to other observations in close temporal proximity and relatively independent from more distant observations. To date, the most promising statistical procedure which accounts for the serial dependency between repeated measures is that proposed by Box and Jenkins (1970, 1976). The way in which the problem of serial correlation is resolved is to empirically model the autocorrelation of the observations, and then to test for the presence of an intervention effect while controlling for the autocorrelation.

Autocorrelation and Partial Autocorrelation

The concept of autocorrelation is used to describe the type of dependence among the observations of a time series process and is central to the discrimination among time series models. The autocorrelation is defined as the correlation between all pairs of observations separated by a fixed number of points in the time series (Anderson, 1976). For an observed series of length n, the estimated lag k autocorrelation coefficient (ACF) is given by

$$r_k = \frac{\left(\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+1} - \bar{Y})\right)}{n - k} \div \frac{\left(\sum_{t=1}^{n} (Y_t - \bar{Y})^2\right)}{n}$$

where Y_t is the observation at time t, Y_{t+1} represents the observation at time t+1, and $\bar{Y}$ is the mean of all observations. For a given lag k, however, variance (denominator of the equation) is estimated over all n observations, while covariance (numerator of the equation) is estimated over only n-k pairs of observations. Therefore, the ACF(k) is not quite the same as the familiar Pearson product-moment correlation coefficient between time series observations. The plot of r_k as a function of the lag k is called the correlogram of the series. The autocorrelations provide the key by which the model underlying an observed series is identified. In practice, to obtain a useful estimate of the autocor relation function, Box and Jenkins (1970) recommend at least 50 observ tions (n) with k not larger than n/4.

The partial autocorrelation (PACF) is a measure of correlation between time series observations k units apart, after the correlation of intermediate lags has been controlled or partialed out (McCleary & Hay, 1980). Unlike the ACF, the algebraic formula for the PACF(k) is inconvenient; it is usually estimated from a solution of the Yule-Walker equation system (Box & Jenkins, 1976). The solution gives the values of

$$PACF(1) = ACF(1)$$

$$PACF(2) = ACF(2) - [ACF(1)]^2 \ / \ 1 - [ACF(1)]^2$$

and so forth. Thus, the expected partial autocorrelation is a function of the expected autocorrelation.

Stationarity and Nonstationarity

Among the assumptions that are made about the nature of the underlying process when estimating a statistical model of a time series process, are the conditions of stationarity. Loosely put, stationarity means that specific characteristics of the underlying time series process remain stable over time (Gottman, 1981). A finite set of parameters must be estimated to determine the model believed to have generated an observed time series realization; consequently, if the process was not consistent over time, it would be impossible to apply the same model to different portions of the time series. One condition of stationarity is that the mean and variance of a time series process do not change with historical time. Therefore, a stationary model is one in which the time series remains in equilibrium around a constant mean level with uniform variability over time. Another condition of stationarity is that the autocovariance of a time series process is independent of historical time. Thus the covariance is determined by the relative lag of the time points, irrespective of the function of the time series under consideration (Gottman, 1981; Gregson, 1983).

Time series data must conform to the conditions of stationarity in order to properly model the time series process. In actual practice, data sets often do not conform to these requirements. A realization of a time series may exhibit one or more of the following forms of nonstationarity: (1) a change in the level or slope of a series over time, (2) periodicity (seasonal component), (3) nonconstant variance, and (4) a shift in the autocovariance structure of the time series (Gottman, 1981). If it is determined that the underlying process is not stationary, the researcher should attempt to either model the nonstationarity, or transform the data so that the observations conform to the conditions of stationarity. In general, the autocorrelations of a stationary time series process lie on the interval -1 to +1, and will approach zero after a relatively small number of lags. In contrast, a nonstationary series will result in autocorrelations which very slowly approach zero as the number of lags increases.

Nonstationarity in the time series may be considered to be either a deterministic or stochastic process. Deterministic behavior can be expressed as a fixed function of time; future time points are completely determined by past observations. Stochastic behavior can be expressed as a random process operating through time; observations are only partially determined by previous occurrences (McCleary & Hay, 1980). One procedure utilized in the analysis of nonstationary series involves modeling the nonstationary components of the series and subtracting these components from the original data set. Assuming that the nonstationarity has been accurately modeled, the removal of these components will result in a set of residuals conforming to the conditions of stationarity. The residuals may then be modeled as a stationary time series. The nonstationary components may be modeled via ordinary least squares fitting procedures (Gottman, 1981).

An alternate method for analyzing series that are nonstationary with respect to level involves a transformation of the data referred to as differencing. The first differencing of a time series is defined as

$$\nabla Y_t = Y_t - Y_{t-1}$$

where ∇ is used to indicate that the time series has been differenced; all observations are subtracted from the immediately preceding observation. Thus, if a time series is nonstationary in level, it will oscillate around a mean level for a time and then drop or rise to a new temporary level. First differencing produces a stationary series. If a series is nonstationary in slope, it will drift in one direction for a time and then temporarily shift direction for a time. Second differencing is necessary to produce a stationary series in this case (Box & Jenkins, 1970, 1976). In practical applications of time series analysis, it is rarely necessary to difference beyond the second order (McCleary & Hay, 1980). Gottman (1981) cautions that over-differencing, for example, differencing a White Noise series, actually introduces dependency in the data set.

McCleary and Hay (1980) discuss the potential problems with the linear regression model for trend, as well as emphasizing the differences between deterministic trend and stochastic drift. The modeling of trend via ordinary least squares regression analysis assumes an underlying deterministic process. Differencing on the other hand, assumes an underlying stochastic process that is free to vary in a probabilistic manner. Box and Jenkins (1976) also advocate that the ordinary least squares regression modeling procedure may be considered to be appropriate only when it can be assumed that the nonstationarity is of a deterministic nature. They suggest that the issue of deterministic trend versus stochastic drift is really the issue of fitting versus modeling a time series.

To make the description of time series analysis simpler, many authors have introduced a new mathematical operation known as the backward shift operation. The backward shift operator (B), is an operator which shifts the time series backward one point in time. Thus, the notation $B(Y_t)$ refers to Y_{t-1}. The representation of the differencing operator can be simplified by using the backward shift notation. First differencing can be represented as:

$$\nabla Y_t = Y_t - Y_{t-1}$$

or $$= Y_t - B(Y_t)$$

or $$= (1-B)Y_t$$

The backward shift operator also has the property of invertibility, so that $B^{-1}B$ equals one (McCleary & Hay, 1980).

Time Series Models

A model is a set of assumptions made about the mathematical process that may have generated the data (Gottman, 1981). The time series process model is a class of combined Autoregressive-Moving Average models (ARIMA) of Box and Jenkins (1976) and Box and Tiao (1965, 1975). ARIMA represents Autoregressive Integrated Moving Average, after the three components of the general ARIMA model. Conceptually, in Autoregressive processes, each data point is a function of the preceding value; each value is correlated with all preceding values plus a random component. In Moving Average processes, each data point is a function of the averaged current random shocks plus one or more previous shocks; each value is a weighted average of the most recent random shocks (Harvey, 1981). The general ARIMA model can be summarized with the equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

where Y_t is the present output or observation, a_t is the present input or random shock, ϕ is the parameter that reflects the influence of preceding outputs, and θ is the parameter that reflects the influence of preceding inputs. Time series (ARIMA) models are built empirically from the data. The data are closely examined to determine the parameters that represent the series of observations. The analysis focuses on identifying one of two systematic stochastic processes: the Autoregressive (AR) and Integrated Moving Average (IMA) components (Glass, Willson, & Gottman, 1975). These are represented by the structural parameters p and q respectively, in the ARIMA model. ARIMA (p, d, q) models describe a time series as the realization of a stochastic or "noise" process: the relationship between random shocks and the time series. The integer values of p, d, and q are specified through statistical analysis called identification and refer to the number of terms in the AR or IMA processes. The concept of stationarity is associated with the structural parameter d. The structural parameter d indicates the number of times a series has to be differenced before it is made stationary. Differencing amounts to subtracting the first observation of the series from the second observation, and so on.

The concept of autoregression is associated with the structural parameter p. The structural parameter p indicates the autoregressive order of the ARIMA model or when the time series are characterized by a direct relationship between adjacent observations. The parameter p thus denotes the number of autoregressive structures in the model: the number of past observations required to predict the current observation.

The concept of moving average is associated with the structural parameter q which indicates the moving average order of an ARIMA model. The value of q exceeds zero when the time series is characterized by the persistence or legacy of a random shock from one observation to the next. In a Moving Average process, the random shock persists for no more than q observations and then its effect on the series is gone. Glass, Willson, and Gottman (1975) reported the five most frequently occurring models in behavioral data, in descending order, to be: Integrated Moving Average, Autoregressive, White Noise, Moving Average, and Nonstationary series.

Autoregressive Models

Autoregressive time series models are an extension of common regression models in that they predict observations in a time series from a previous set of observations in the series. A first-order Autoregressive (AR) model is specified as

$$Y_t = \phi_1 Y_{t-1} + a_t$$

where ϕ is the autoregressive coefficient that minimizes the squared error, $\sum_{t=1}^{n} a_t^2$, and a_t has a mean of zero, a variance of σ^2, and is uncorrelated with Y_t (Glass, Willson, & Gottman, 1975). The observations of time series processes are usually represented as deviations from the mean observation. The absolute value of ϕ_1 must be less than one if a first-order Autoregressive process is to be stationary. The variance of an AR process is:

$$\text{Var } Y_t = \sigma^2 / 1 - \phi^2$$

If ϕ_1 is greater than one, a given observation will be more strongly related to distant observations than to those that are in close temporal proximity. The autocorrelation function (ACF) of the Autoregressive process is expected to decay exponentially and can be specified as

$$ACF_{(k)} = \phi_1^k$$

where k is the number of lags. As the expected partial autocorrelation (PACF) is a function of the expected ACF, the PACF(1) is non-zero, while PACF(2) and all successive lags are expected to be zero for the Autoregressive process. Therefore, Autoregression refers to a stochastic behavior in which a random shock has an exponentially diminishing impact over time. The relative simplicity of parameter estimation in Autoregressive models has led some researchers (Gottman, 1981) to recommend their use almost exclusively. However, the consideration of only Autoregressive models may result in a rather large number of autoregressive parameters. Consequently, other researchers prefer ARIMA models, since they are more parsimonious (Blumberg, 1984; Gorsuch, 1983; Harrop & Velicer, 1985; Velicer & McDonald, 1984).

Moving Average Models

As Autoregressive processes are characterized by a dependency between observations that decays exponentially as the number of observations increases, Moving Average processes are characterized by a dependency between observations that are separated by a finite number of time points. Observations separated by more than q points in time are independent from each other. While Autoregressive processes are modeled in terms of previous observations, Moving Average processes are modeled in terms of previous error terms (a_t), referred to as random shocks. The principle of the model is that an observation Y_t is a function of the current random shock (a_t), and a portion of a fixed number of previous random shocks. A first-order Moving Average (MA) model is specified as

$$Y_t = a_t - \Theta_1 a_{t-1}$$

where Θ is the moving average coefficient that minimizes the squared error, $\sum_{t=1}^{n} a_t^{\,2}$, and a_t, white noise, is assumed to have a mean of zero, a constant variance of σ^2, and is uncorrelated with Y_t (Glass, Willson, & Gottman, 1975). The variance of an MA process is:

$$\text{Var } Y_t = \sigma^2(1 + \Theta^2)$$

The absolute value of Θ_1 must be less than one if a first-order Moving Average process is to be stationary, the bounds of invertibility. While the name is different, the bounds of invertibility play much the same role as the bounds of stationarity for autoregressive parameters. In practice, when ϕ_p or Θ_q parameters exceed the bounds of stationarity or the bounds of invertibility, the series is either nonstationary and requires differencing, or it was differenced too many times. The autocorrelation function (ACF) of the first-order Moving Average process is expected to truncate after a single lag and can be specified as:

$$ACF_{(1)} = \frac{-\theta_1}{1 + \theta_1^2}$$

The partial autocorrelations (PACF) are expected to be non-zero, with successive lags of the partial autocorrelation function growing smaller and smaller in absolute value. Therefore, Moving Average refers to stochastic behavior which is a weighted summation of random shocks that truncate to zero after q lags.

<u>Duality of AR and MA processes</u>. There is a fundamental duality between Autoregressive and Moving Average models. A stationary first-order Autoregressive process can be represented as an infinite order Moving Average process. Similarly, a first-order Moving Average process can be expressed as an infinite order Autoregressive process. The practical significance of this duality is the flexibility offered in modeling time series data; one can adequately model a Moving Average process with an Autoregressive model AR(p), where p is relatively large. The autocorrelation function of an Autoregressive process decays exponentially over time, while that of a Moving Average process truncates after lag q. Logically, the potential for modeling the dependency of an Autoregressive process with a MA(q) model would be to specify a large value of q.

By continuing to go backward in time using the first-order Autoregressive model to minus infinity:

$$Y_t = \phi_1 Y_{t-1} + a_t$$

or

$$Y_t = \phi_1(\phi_1 Y_{t-2} + a_{t-1}) + a_t$$

or

$$Y_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \cdots$$

results in

$$Y_t = \sum_{i=0}^{\infty} \phi_1^i a_{t-i}$$

Thus the first-order Autoregressive process, AR(1), can be written as an MA(∞) model. This relationship is the case only when $|\phi_1^k|$ is less than one, so that ϕ_1^k approaches zero as k approaches infinity for the series to converge to a finite limit (McCleary & Hay, 1980).

Now the first-order Moving Average process, MA(1) can sometimes be written as an AR(∞) model. The first-order Moving Average model:

$$Y_t = a_t - \theta a_{t-1}$$

also implies

$$Y_{t-1} = a_{t-1} - \theta a_{t-2}$$

Solving for a_{t-1} and substituting the result in t first equation results in:

$$Y_t = a_t - \theta(Y_{t-1} + \theta a_{t-2})$$

or

$$Y_t = -\theta Y_{t-1} + a_t - \theta^2 a_{t-2}$$

Rewriting the MA(1) and substituting for a_{t-2} gives:

$$Y_{t-2} = a_{t-2} - \theta a_{t-3}$$

or

$$Y_t = -\theta Y_{t-1} + a_t - \theta^2(Y_{t-2} + \theta a_{t-3})$$

or

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} + a_t - \theta^3 a_{t-3}$$

results in

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \theta^3 Y_{t-3} - \cdots$$

with repeated substitutions, which is an AR(∞) process. This relationship only holds if $|\Theta|$ is less than one, which is known as the invertibility condition. In general, the invertibility condition of the MA(q) process is similar to the stationarity condition of the AR(p) process. As Gottman (1981) points out, in practice, data generated by an MA(1) process could be estimated using an Autoregressive process of high order, as Θ to some power will eventually become zero. Nonetheless, the most parsimonious model is still the MA(1), with only one parameter. Moreover, the variance of the time series will converge to a finite limit only if the conditions of stationarity and invertibility are satisfied. Consequently, any stationary time series process can be modeled with an infinite moving average and approximated by a finite MA(q) model. However, this representation is not always the most efficient model. Autoregressive models are clearer, more flexible, and easier to handle than Moving Average models. In fact, Gottman (1981) indicates that at the price of using models with larger numbers of parameters, the more tractable Autoregressive model can generally be used.

Random Walk and Other Integrated Processes

A random walk is a stochastic process wherein successive random shocks accumulate or integrate over time, hence an integrated process. The random variate (Y_t) can make wide swings from its expected level; it drifts, and if only a short realization of the process is available, one might conclude that the process follows a trend. Processes thought of as random walks are encountered in the social sciences, which have random shocks varying in sign (positive or negative) as well as in size

(McCleary & Hay, 1980). Because a random walk observation is the sum of all past random shocks, the integrated process or Nonstationary (NS) series can be represented as:

$$Y_t = Y_{t-1} + \theta_0 + a_t$$

The series, which follows a linear trend, is simply differenced which transforms the random walk into a "white noise" process. White Noise (WN), the residual or error term of the prediction, is the unsystematic part of the stochastic process. With an integrated model, the best prediction of the current time series observation (Y_t) is the preceding observation (Y_{t-1}) and a constant, where θ_0 equals zero. This process can also be represented as an AR(1) process with ϕ equal to one.

The most common nonstationary process is the Integrated Moving Average model. This Integrated Moving Average (IMA) model is depicted as:

$$Y_t = Y_{t-1} - \theta_1 a_{t-1} + a_t$$

The Integrated Moving Average process will evidence nonstationarity; it will "wander away" from any given level for long periods of time rather than oscillating around a single level (Glass, Willson, & Gottman, 1975). The autocorrelations of the non-differenced data do not die out to zero either exponentially or abruptly. After first differences of the data, a Moving Average structure is apparent as the lag 1 autocorrelation is non-zero, but the autocorrelations for successive lags are essentially zero.

Time Series Model Building

Box and Jenkins (1976) propose an ARIMA model building strategy which is an iterative strategy consisting of identification, estimation, and diagnosis. The model building strategy is outlined in Figure 1. The goal of the identification process is to determine an ARIMA model which parsimoniously describes the data set. The pattern of autocorrelations and partial autocorrelations are determined and inspected. Once the correct model is chosen, parameter estimation is performed. Finally, the adequacy of one model is assessed in the diagnosis stage, by fitting the model to the data to determine if any dependency remains.

ARIMA model identification is an important key to the analysis of time series data and is crucial for testing the hypothesis of an intervention effect. The practical basis for selecting one tentative model over another is the pattern of autocorrelation found in the ACFs and PACFs estimated from the time series realization. Subjective judgment is involved at this stage of model building, in that no precise objective method is available for determining the best values of p, d, and q in the ARIMA (p, d, q) model (McCleary & Hay, 1980). The autocorrelation and partial autocorrelation functions are examined to determine the dependency of time series observations. The general properties of the autocorrelation and partial correlation function are summarized as:

Process	ACF	PACF
AR(p)	Decays after lag p	Truncates after lag p
MA(q)	Truncates after lag q	Decays after lag q

Figure 1. The ARIMA Model Building Strategy

(1) IDENTIFICATION

↓

(2) ESTIMATION: Parameter estimates must be statistically significant and must lie within the bounds of stationarity-invertibility. If either criterion is not met, return to IDENTIFICATION. If both are met, proceed to DIAGNOSIS.

↓

(3) DIAGNOSIS: Model residuals must be white noise as judged by two criteria. First, the residual ACF must have no spikes at key lags. Second, the Q-statistic must not be significant. If either criterion is not met, return to IDENTIFICATION. If both are met, accept the model.

↓

(4) USE THE MODEL: After a tentative model has been accepted, it may be used for intervention effect assessment.

The first consideration in model identification is the stationarity of the time series process. A tentative identification of the differencing parameter (d) is determined. If the specified value of the differencing parameter is too large, dependencies among the observations are introduced. The problem of over-differencing can be avoided by applying the difference operator only when the estimated autocorrelation function unambiguously demonstrates that the time series process is nonstationary. Box and Jenkins (1976) as well as McCleary and Hay (1980) contend that application of time series analyses rarely requires differencing beyond the second order.

The actual autocorrelation function of a time series process is never known, thus a finite realization of the time series process is used to estimate the true ACF and PACF. Ambiguity in identification can be lessened by placing confidence bands around the ACFs and PACFs. For the ACF, standard errors (SE) of the ACF(k) are estimated by:

$$SE(ACF_k) = \sqrt{1/n\ [1 + 2 \sum_{i=1}^{k} (ACF_i)^2]}$$

and for the PACF, standard errors of the PACF(k) are estimated by:

$$SE(PACF_k) = \sqrt{1/n}$$

Approximate regions of nominal acceptance or rejection can be formed around the ACF and PACF using the values of plus or minus two standard errors. If the ACF and PACF fall within this region, they are considered to be not significantly different from zero. The width of the confidence bands are directly related to the number of observations in the time series process (Box & Jenkins, 1976). Certainty in the identification of the time series process is increased as the number of observations in the realization becomes greater.

Next, after having identified an ARIMA (p, d, q) model for the time series realization, the ϕ_p and Θ_q parameters of the model are estimated. Estimation procedures that converge to a minimum residual sum of squares $(\sum_{t=1}^{n} a_t^2)$ of an ARIMA model are used. All parameter estimates must be within the bounds of stationarity-invertibility, as well as being statistically significant.

The last stage of the model building process involves evaluating the adequacy of the tentative model. In the diagnosis stage, residuals $(\hat{a}_t)$ of a time series model are assessed by analyzing the estimated autocorrelation function. McCleary and Hay (1980) outline two criteria for evaluating the adequacy of an ARIMA model. First, there should be no dependency between the estimated autocorrelation at the first or second lag. Box and Jenkins (1976) point out that the approximate standard errors of the estimated autocorrelations of the residuals $(1/\sqrt{n})$ tend to be inflated at low lags. Thus, discrepancies from the expected autocorrelation of zero at lags 1 and 2 should, for diagnostic purposes, be considered significant if slightly less than the confidence band of two standard errors in magnitude. Second, the residuals of the tentative model must be distributed as white noise. This diagnostic check considers an entire set of autocorrelations simultaneously to evaluate whether the set of estimated autocorrelations are different from zero. Box and Pierce (1970) suggest using the Q-statistic given by

$$Q = n \sum_{t=1}^{k} (ACF_t)^2$$

which is distributed approximately as Chi-square with k-p-q degrees of freedom, where n represents the number of observations used to estimate the autocorrelation function, k is the number of lags used for calculating the estimated ACF, p indicates the autoregressive order, and q indicates the moving average order. When using the Q-statistic to detect serial dependency of the residuals, McCleary and Hay (1980) recommend setting the value of k between 20 and 30 lags. A large value of k lacks power in rejecting the null hypothesis of independent observations, whereas a value of k less than 20 will tend to be over-sensitive and lead to rejections of the null hypothesis even when the residuals are distributed as white noise.

A variety of other residual checks may be useful in diagnosing the estimated model (Box & Jenkins, 1976). Inspecting a plot of the residual series and a plot of the predicted values versus the observed values are invaluable for assessing the fit and adequacy of the model. Another procedure is referred to as over-fitting. The basic concept of model over-fitting is the attempt to find a better fitting model by adding parameters to the tentative ARIMA model. A statistic, R^2, can be computed which describes the amount of variance in the time series accounted for by the ARIMA model (McCleary & Hay, 1980). This statistic is analogous to the percentage of variance accounted for by a regression analysis. Another related measure of the goodness-of-fit of the ARIMA model is the residual mean square, represented by:

$$\mathrm{RMS} = 1/n\sqrt{\sum_{t=1}^{n} n\hat{a}_t^{2}}$$

Models that have a smaller residual mean square are better fitting than those models with a larger value. Additional parameters should be selected on the basis of knowledge concerning possible sources of dependency that may not have been adequately modeled.

Simonton (1977), Algina and Swaminathan (1977, 1979), Gottman (1981), Glass (1980, 1984), Velicer and Harrop (1983), Velicer and McDonald (1984), and Harrop and Velicer (1985) have discussed difficulties in ARIMA model identification. There is a lack of precision in the ARIMA model building procedure which is apparent in the estimation of the standard errors of the autocorrelation function in the use of the Q-statistic, and in the evaluation of the goodness-of-fit of a model. There is also a great degree of subjective judgment involved in the interpretation of information gained from time series data analysis, as many properties of the estimators used in ARIMA modeling are not precisely defined. Interest is developing in providing information upon which judgments can be made to produce useful time series models.

Interrupted Time Series Experiments

The time series quasi-experiment was originally proposed as a means of assessing the impact of a discrete social intervention on behavioral processes (Campbell & Stanley, 1966). The analysis of the time series quasi-experiment almost inevitably takes the researcher into a consideration of ARIMA models and strategies (Cook & Campbell, 1979). The interrupted time series experiment must address the threats to validity (internal, external, statistical conclusion, and construct). The use of the statistical ARIMA model involves an adequately designed quasi-experiment. Intervention assessment is concerned with the effect of an event (changes in states) and requires the onset of an event be specified *a priori*. A null hypothesis that an event caused a change in some behavior can be tested only if the time of the event is known *a priori*.

A complex model is fit which involves not only ARIMA components, but also parameters that describe the intervention. The residual values are then used to test the parameters. In interrupted time series analysis, the impact of some intervention or interruption in the series is represented by:

ARIMA model (Y_t) = noise(a_t) + intervention component (I_t)

The research question asks: (1) does the addition of the intervention component to the model significantly increase the model's predictive power, (2) what causal inference concerning the interrelationship of separate time series processes can be drawn, or (3) what is the nature of an intervention effect?

The most common method used for the analysis of interrupted time series experiments was developed by Box and Tiao (1965) and discussed by Glass, Willson, and Gottman (1975). This method involves the simultaneous estimation of the intervention component and the parameter of an ARIMA model using nonlinear estimation procedures. Gottman (1981) proposes another procedure which involves reducing the time series realization to a white noise process by removing the dependency of the observations with an Autoregressive model. The residuals of the AR model are then used to assess the intervention effect using ordinary least squares procedures. This alternative is relatively uninvestigated and not often utilized. Box and Tiao (1975) propose another procedure for assessing the effect of interventions. McCain and McCleary (1979) give an introduction to the Box and Tiao (1975) dynamic method of modeling interventions of which the original Box and Tiao (1965) model is a special case. More detailed presentations are found in McCleary and Hay (1980). This approach to modeling the intervention effect provides a more global technique for incorporating intervention effects. The one type of intervention effect considered previously was that of an abrupt constant change in level, where a constant value, ω, was added to each post-intervention point. The development of the Box and Tiao (1975) dynamic method offers greater flexibility in evaluating a wide variety of intervention effects.

The time series process is simply assumed to be the outcome of two components: (1) the stochastic process of ARIMA (p, d, q) model, and (2) the deterministic effect of an intervention component. Three different types of intervention effects can be assessed using the interrupted time series model (Glass, Willson, & Gottman, 1975). First, the simplest type of intervention effect is that of an abrupt, constant change in level. In this case the intervention component is represented as

$$I_t = \omega I_t$$

where $I_t=0$ prior to the intervention and $I_t=1$ after the intervention, with ω being the intervention effect. Second, a gradual, constant change can be evaluated by modifying the intervention component as

$$I_t = \delta Y_{t-1} + \omega$$ [illegible]

where the parameter δ is in the interval of -1 to +1, and estimates the rate at which the intervention effect approaches the asymptote, or the change in level of the time series. Thus, this model in which δ is not equal to zero implies a gradual change in level in which the time series remains relatively stable throughout the post-intervention phase. In the extreme case where $\delta=1$, the level continues to increase at a continuous rate instead of eventually reaching a constant level. When $\delta=0$, the intervention component reduces to the model representing an abrupt, constant change. Third, an abrupt, temporary change can be modeled by:

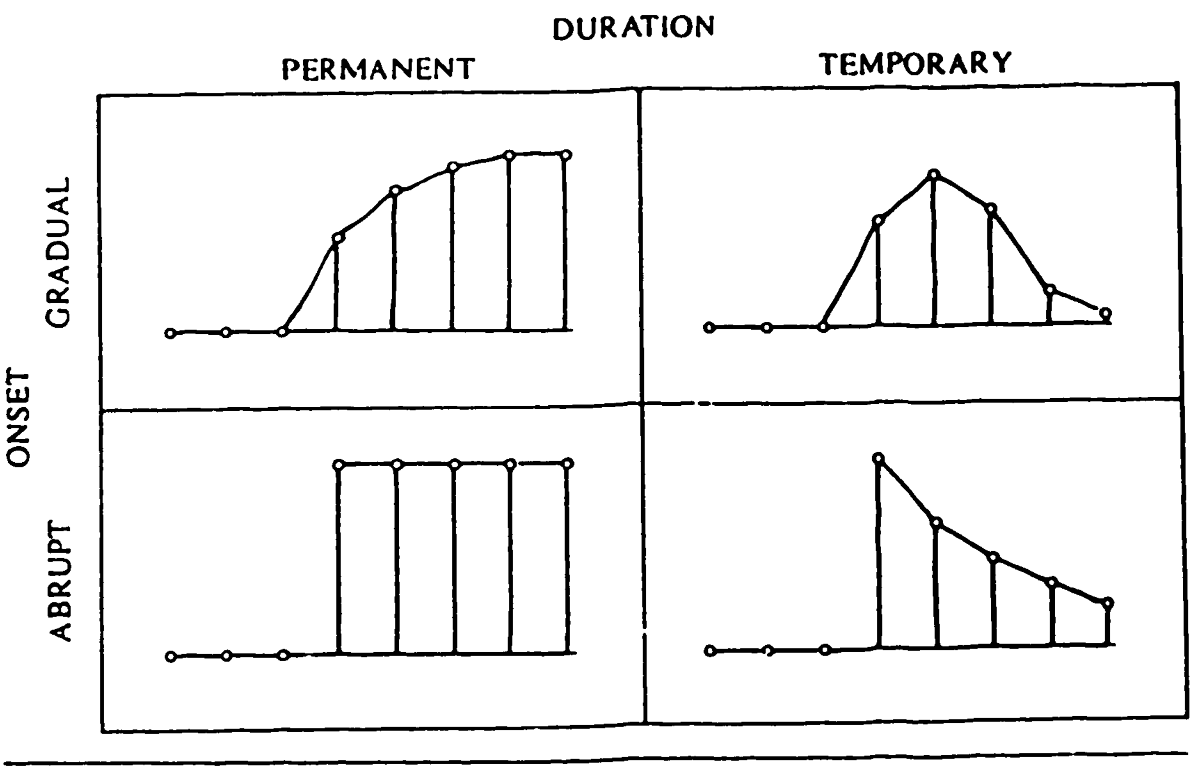
$$I_t = \delta Y_{t-1} + \omega(1-Y_{t-1})I_t$$

This form of the intervention component represents an abrupt change with a magnitude of ω at the point of intervention. The intervention effect then decays at a rate determined by δ. These three forms of intervention effects illustrated in Figure 2 probably represent the most commonly encountered changes in the social sciences (McCleary & Hay, 1980).

The general procedure of evaluating the effect of an intervention begins with identifying an ARIMA (p, d, q) model according to the general model building process. Caution is advised since the intervention effect can sometimes change the nature of the ARIMA process from pre-intervention to post-intervention data. There is an underlying assumption that the stochastic process of the time series realization is equivalent before and after intervention. Some authors suggest explicitly modeling the two distinct time series processes if the intervention alters the nature of the time series process (Stoline, Huitema, & Mitchell, 1980), or relying on the pre-intervention data to identify the ARIMA model (McCleary & Hay, 1980), and then fit the entire model including the intervention component. Assuming that the most appropriate ARIMA (p, d, q) model has been identified, the intervention component is added and all of the parameters of the full model are estimated. The intervention effect is evaluated using the estimate for the parameter ω. In the analysis of interrupted time series experiments, the actual test of the basic hypothesis is relatively simple. Unfortunately, difficulties often arise when attempting to model the dependency in data sets.

Figure 2. Patterns of Intervention Effects

From Applied Time Series Analysis for the Social Sciences (p. 172) by R. McCleary and R.A. Hay, Jr., 1980, Beverly Hills, California: Sage Publications.



A general overview of time series methods and the wide variety of analytic approaches may be found in Anderson (1977), Makridakis (1976, 1978), and Newbold (1981). Gottman and Glass (1978) give an overview and rationale to time series analysis; Stoline, Huitema, and Mitchell (1980) consider the case of a change in parameter values for an Autoregressive model. From a methodological perspective, the primary references for the evaluation of interventions are Box and Tiao (1965, 1975) for the dynamic intervention model, and Glass (1972) for the popular extension of the model based on Box and Tiao (1965). Glass and several others expanded the technique (Maguire & Glass, 1967; Glass, 1968; Glass, Tiao, & Maguire, 1971; Glass, 1972), which was later integrated into a general account of the time series experiment and its analysis by Glass, Willson, and Gottman (1975). There are numerous examples of studies employing Glass's approach to evaluating interventions (Blose & Holder, 1987; Bowie & Prothero, 1981; Dalton & Todor, 1984; Hamilton & Waldman, 1983; Rotton & Fry, 1985). Generally, in many of these studies, no discussion of the model chosen or how well the model fits the data is given, no attempt to rule out the threat of history to the validity of the experiment is made, and in some cases the number of data points is very small.

Other authors have investigated the assumptions of the ARIMA models and the estimation of an intervention effect in interrupted time series designs. Padia (1975), using Monte Carlo simulation techniques, found that failure to difference a non-stationary time series results in a gross underestimation of Type I error and power; whereas, over-differencing a stationary time series results in large over-estimates of Type I error and power. Marquis (1983) also found estimates of Type I error considerably higher than the nominal level when the ARIMA model is inadequately differenced. White (1985) also found similar results in intervention effect estimation in short time series realization with the Autoregressive model, as did Padia (1975) and Marquis (1983). Further clarification of the consequences of violating the assumptions of ARIMA models when assessing intervention effects is important in the application of interrupted time series procedures.

Human behavior, if viewed as a sum of deterministic and stochas processes, can be described by a probabilistic model. With the use time series analysis in quasi-experimental designs, change can detected and evaluated in a data set over time, as well as the nat and process of treatment effects can be detected and described. relevant issues in the application of interrupted time series desi are discussed in the following sections. The final section outlines problems deemed necessary to investigate in the assessment of interv tion effects in short time series processes.

Utility of the Interrupted Time Series Design

Time series analysis defines and describes the sequence of ra variables through the building of a model which describes or pred interdependent phenomena. The interrupted time series design car beneficial in a variety of areas of behavioral science research. Fi in an experimental approach, subjects are typically divided equivalent experimental and control groups and compared after experimental group receives the treatment. True experiments em randomization in the experimental design to validate data anal conducted as if observations were independent. Many research quest are impossible to investigate within the structure of traditi experimental designs. Limitations in true experimental designs conducting meaningful clinical research have been discussed Kratochwill (1979), Kazdin (1982), and Barlow and Hersen (1984).

traditional experimental designs, large numbers of homogeneous subjects are required. In some situations, group comparisons are therefore not feasible or meaningful due to discrepancies between populations. For example, children with cerebral palsy are a heterogeneous population for which a wide variety of treatments are recommended. In other instances, it may be unethical to withhold a beneficial treatment from a sample in order to determine the effect of an intervention, as in clinical drug studies. Time series quasi-experiments can provide a strategy for exploring data that are collected in the absence of rigorous experimental controls (Barlow & Hayes, 1979).

Additionally, classical repeated measures designs assume a constant covariance structure. When the assignment of a treatment cannot be randomized, serial correlation becomes important. Behavioral data are believed to be highly autocorrelated, although there has been a recent attempt to suggest that the presence of serial correlation is not statistically significant according to traditional tables (Huitema, 1985). However, successive observations are correlated in any cyclical data, such as hormonal cycles and circadian rhythms. Time series analysis evaluates changes in slope and level in data while taking into account the serial dependency in the data. Removal of autocorrelation between data points therefore results in an analysis that separates chance fluctuations from intervention effects.

Time series experiments also permit the assessment of changes in the behavior of individuals. Hypothesis testing of intervention effects in studies involving only a single subject or unit of observation can be undertaken. Hartman, Gottman, Jones, Gardner, Kazdin, and Vaught (1980) provide an overview of the time series method in the context of single-subject designs. Kazdin (1978, 1981, 1982, 1984, 1986) provides several overviews of single-case research designs. The interrupted time series design affords an empirical approach to clinical practice which can test the significance of a difference between treatments for an individual. Individual differences produce variability in the size of a measurable effect and uncertainty as to the occurrence of a qualitative effect. Multivariate statistical models can overlook this fact, leaving the impression that statistical variability is due to external random shocks of unknown causes. In single-case experiments, the hypothesis tested is differences between treatments for an individual, rather than differences in mean effects for groups. With single-case designs, one knows a little, but what is known is dependable and reliable; however, there is also the chance of missing influential treatments. By minimizing the probability of a Type I error, the probability of a Type II error is increased. Baer (1977) has argued that minimizing Type I errors in detecting intervention effects will lead to identification of a few variables whose effects are consistent and potent across a wide range of conditions. Time series designs may assist in making inferences and the ability to discriminate between intervention and error effects.

Lastly, the interrupted time series paradigm also affords the opportunity to assess the impact of an intervention. The impact of an intervention on human behavior is complex in nature and unlikely to be consistent over time. The interrupted time series design, most importantly, provides longitudinal information about the impact of the intervention. Traditional experimental designs generally assess the impact of a treatment at a single time point after the occurrence of the intervention. It is often of interest to evaluate the immediacy, duration, and pattern over time of the intervention effect by examining the post-intervention data. The time series design provides such a method which is appropriate to the complexity found in the effects of interventions with human beings (Glass, Willson, & Gottman, 1975). For example, electrocardiogram changes may be evaluated following the ingestion of ice water by coronary patients at three, ten, and twenty-five minute intervals post-ingestion to determine the relationship of time and volume. Visual inspection alone may overlook reliable but weak changes in such a series with a potential unstable baseline, intrasubject variability, and small effect size.

Thus, one reason to collect time series data is to try to discover systematic patterns in the series so a mathematical model can be built to explain the past behavior of the series. Another important reason for doing time series analysis is to predict future values of the series. The parameters of the model that explained the past values may also predict future behavior patterns. The ability to make such predictions, for example, regarding headache sequences, is obviously important. A final reason for utilizing time series data is to evaluate the effect of some treatment or event that intervenes and changes the behavior of a series. Crisis intervention and the prevention of institutionalization, drug therapy and the decrease in blood pressure, behavior modification and the change in compliance, are examples of interrupted time series. These have in common an hypothesized interruption in their usual pattern after the specific time when some outside event occurred. A primary emphasis in clinical research is therefore testing the significance of a difference between treatments for individuals, or the measurement of change.

Process and Measurement of Change in Clinical Research

A goal of clinical research is to identify, describe, explain, and predict the effects of processes that bring about therapeutic change over an entire course of treatment (Greenburg, 1986). Research questions are often: Can patterns of change be reliably identified? Are these patterns of change related to the outcome? The basis of determining therapeutic change is dependent upon several research design issues. First, the treatments included in a study need to be representative of those used in practice. Second, the results should reflect what would actually happen in practice, not under ideal, experimental conditions when treatments are monitored and delivered with special care. Third, the value of knowing the relative efficacy of alternative treatment techniques, the difference in the range of effects produced, and the process through which such effects are achieved, should encompass critical theoretical and practical questions (Kazdin, 1986); for example, what warrants the label "improved" and when is improvement clinically significant? Fourth, ethical considerations (such as withholding a treatment from a control group or the administration of an ineffective treatment) may limit the choice of research designs. A related issue is that in reality, the nature of any treatment is often tailored to the individual; in short, how effective would a specific treatment be for a specific individual?

It is important to determine whether therapeutic changes are maintained and whether they surpass the gains that may be associated with the passage of time without the treatment. Additionally, it is important to identify the form of change associated with the treatment.

Time series designs offer a method which enables examination of various aspects of treatment effects without altering treatments in a major way (Hayes, 1981). In a single pretest-posttest design, the effectiveness of treatment is assessed just once after implementation. This strategy precludes an analysis of effects during or after treatment. A time series design allows for a number of repeated observations of a dependent variable over time. Therefore, the form of the intervention effect as well as the statistical significance of the effect can be assessed (Edgington, 1967; Shine & Bower, 1971; Namboordiri, 1972; Glass, Willson & Gottman, 1975).

Group comparison designs may, and typically do, test global treatments against no treatment. Although experimental designs can answer different questions, this strategy obviates an analysis of the specific mechanisms of change. In order to validly attribute change to the treatment, a change in behavior needs to be shown to co-vary in a lawful way with changes in the treatment. There are at least three patterns of change in intervention effects: (1) an abrupt, constant change, (2) a gradual, permanent change, or (3) an abrupt, temporary change. These patterns of change may be masked in the statistical analysis required by experimental designs. Statistical analysis may average out this pattern of change yielding non-significant findings. Time series designs highlight the patterns of change through repeated measurement and also offer a test to detect differences in the individual or unit of observation. Time series analysis therefore allows the experimental isolation of the effects of the treatment from the effects of other factors which may be simultaneously acting upon the subject.

Another critical issue is the extent to which a study can detect differences; that is, the power of the design (Cohen, 1977). If effect size is likely to be small, the size of the sample needs to be increased commensurately. Consideration of effect size, sample size, and alpha can increase the precision of the test. However, if the sample size is large enough, statistical significance can be achieved, yet there may only be a small effect. Various statistical models for analyzing change are available. Vitaliano (1982) examined several statistical procedures for analyzing repeated measures designs. One such method, raw gain, change, or difference scores, formed by subtracting pre-test scores from post-test scores, can lead to erroneous conclusions as such scores are systematically related to any random error of measurement (Cronbach & Furby, 1970; Burckhardt, Goodwin, & Prescott, 1982). Consequently, gain scores have a low reliability which can lead to a loss of power. Alternatives when studying change over time are: repeated measures analysis of variance (Winer, 1971); trend-analysis; analysis of covariance on post-test scores, with the pre-test variable treated as the covariate; factorial analysis of variance with blocking on pre-test scores; regression analysis; and time series analysis (Cook & Campbell, 1979).

The linear statistical models assume that correlations or covariances between all pairs of repeated measures are equal. This is seldom the case when the same subject is measured at different time points. Violation of the assumption may inflate the observed F-values and t-values, thereby contributing to an inappropriate rejection of a null hypothesis; increasing the absolute values of the correlations leads to an increase in the deviation of the actual significance from the nominal level. Another important assumption underlying the appropriate use of linear statistical models is that the residuals of the model are independent. Frequently, time series data are misanalyzed using ordinary least squares regression when the assumption of nonindependence of observations does not hold (Hibbs, 1974). Ignoring a correlated error structure can lead to over- or underestimates of effects depending on the nature of the design and the error structure. Traditional regression solutions ignore the problem of autocorrelation and also face the problem of increasing variance of the forecast error and diverging confidence bands around the parameter estimates (Box & Jenkins, 1970). Therefore, these solutions are acceptable only very near the center of the baseline period and very near the intervention point (Kazdin, 1984). The intent of the interrupted time series design is to study the effect of an intervention while controlling for threats to validity such as maturational trends or statistical regression; time series analysis procedures incorporate serial correlation into the statistical model for determining parameter estimates. Time series analysis can thereby provide confidence in the conclusions drawn about the intervention effect.

<u>Levels of Inference in Time Series Analysis</u>

Changes in a time series which coincide with the occurrence of an intervention are presumed to be the effects of the intervention. This causal claim may be invalid. The paradigmatic assertion in causal relationships is that the manipulation of a cause will result in the manipulation of an effect (Cook & Campbell, 1979). To infer effects, comparison is needed. Random assignment of people to treatments makes the expected value of the correlation between treatment and outside background variables equal to zero. Thus, in the long run, the only systematic difference (*a priori*) between randomly assigned groups should be the treatment. Quasi-experiments have treatments, outcome measures, and experimental units, but do not use random assignment to create comparison from which change is inferred. Therefore, the irrelevant causal forces hidden within random assignment need to be made explicit by separating effects due to treatment from those due to the initial noncomparability between average experimental units. Experiments probe but do not establish causal hypotheses; alternative hypotheses are successfully eliminated. Generalizations are made from particular observations to scientific propositions. If the data fit the pattern, theory is supported, that is, no other current theory accounts for the data pattern observed, but the theory is not proven to be true. The problem is that data can fit two inconsistent theories equally well. Because theories are fallible, all point null hypotheses are false (Meehl, 1978; Serlin & Lapsley, 1985). Threats to valid inference making are found in statistical conclusion validity and internal validity. To test causal hypotheses, internal validity is the *sine qua non* of causal inference.

Use of statistics helps to portray the uncertainty of data. Hypothesis testing compares experimental effects against alternative explanations; in most behavioral sciences, hypothesis testing is not decision-making. Randomness is the model to which data are compared. Random generation supplies rival hypotheses; if observed data are different from those random data, theory is supported. Data can then be used in theory construction and evaluation. Theory is created from and grounded in data (Glaser & Strauss, 1967; Glaser, 1978). For deductive inquiry, when the theory is strong, the data can be tested against a point hypothesis which derives from the theory. Weak theories can be assessed against no differences hypotheses in an attempt to assess questions like: Is there an effect (relationship)? Is this relationship (effect) due to the treatment or is individual variation a plausible explanation? Statistical validity first establishes whether two variables co-vary. Is there causal relationship and in what direction? If causal relationship is found, what are the hypotheses involved in the relationship? Given a probable relationship of one construct to another construct, how generalizable is it across people, settings, and times? Interrupted time series designs are concerned with effects of a treatment inferred from comparing measures of performance taken at many time intervals before and after a treatment. Consequently, the statistical significance test only indicates that there is an observed relationship. The importance of the results from the interrupted time series experiment awaits representative studies. Practical significance of the results from the interrupted time series experiment is gained through generalizability studies.

The adequacy of inference in research is a function of how the data are produced, as well as how they are analyzed (Cohen & Cohen, 1983). As Cook and Campbell (1979) state, the goal is to rule out as many threats to validity as possible and reduce the number of possible alternative interpretations of the data. Inferential statistics assist in pattern detection in the data by excluding hypotheses of random fluctuations. Statistics are used to discern the uncertainty of the effect, whereas the design of the experiment assists in revealing the source of the effect. The greater the overlap of the sampling distributions, the lower the power with respect to that alternative. The level of statistical significance is set by the researcher. Therefore, the researcher needs to ask: what are the risks in decisions based on the experimental evidence? The choice of the region for rejection of the null hypothesis will ultimately affect the statistical power of the results. Time series designs and analysis can provide an a priori basis for directional hypothesis testing. Inference in time series analysis is made to the process that may have generated the sample series.

The vast majority of studies evaluating change infer treatment efficacy on the basis of statistical comparisons between treatment conditions. These comparisons have limited utility in that they are based on the average improvement score for all subjects and need not apply to any one individual. It is possible to detect a change, have a return to normal functioning, yet the change may not be statistically reliable as it falls within the margin of measurement error. Since variability within the group is treated as error to be minimized, clinical significance cannot be inferred even if statistically significant effects are obtained (Jacobsen, Follette, & Revenstorf, 1984). However, tests of significance in time series experiments can also impose a criterion for determining a treatment effect which may have little clinical relevance. How much change should there be to be considered improved?

Consequently, in interrupted time series experiments, it may not only be important to determine that change occurred, or that the treatment was effective, but also to show that some clear practical benefit is associated with the change. Statistical significance is a function of the size of effect being measured, amount of variance already present in the population, size of the sample, and level of significance chosen. All claims of change need to be compared against alternative explanations. Change [illegible]es multiple perspectives, individual, family, and society, as well as multiple criteria for defining the magnitude and importance of change. Determination of effect size can lead to a rule for making decisions regarding clinical judgments of improvement. What is done a priori is in reference to the effect size that is clinically significant. Measures of clinical significance have been suggested such as the extent to which the subject returns to normative values, the magnitude and range of change, and the degree to which change is perceived as significant to others (Kazdin & Wilson, 1978; Hugdahl & Ost, 1981). Effect size is therefore determined scientifically, clinically, and practically (Nelson, Rosenthal, & Rosnow, 1986).

Time series procedures can assist in generating appropriate knowledge for clinical decision-making. The clinician translates problems into questions related to theory and tries to derive solutions from theory; the ultimate goal is to develop knowledge and produce evidence to validate clinical practice. In returning to the basic issue of analyzing data statistically rather than clinically, with respect to the question of, did a treatment change a person's behavior from level a to level b, it comes down to choosing a correct unit of analysis to provide statistical inference (Jacobson, Follette, & Revensto.f, 1984). Time series analysis procedures can provide more statistical power for separating real intervention effects from other trends, thus increasing confidence in the results.

Issues in Clinical Nursing Research

A crucial question in nursing is whether the science as practiced furthers the field of interest. Has nursing inquiry idolized experimental designs and inferential statistics at the expense of more representative paradigms? What are appropriate paramorphic models of clinical realities? Measuring changes in behavior occurring along a delineated time interval is a concern of nursing research. Consequently, issues related to the nature of the change to be detected require exploration. What is an appropriate index of change due to an intervention? What degree of change warrants clinical significance? Are strategies required which go beyond statistical evaluation of group differences? Are adequate research methods open to practitioners? Has clinical research failed to influence clinical practice?

How can the results of clinical research add to nursing knowledge, and provide a basis for clinical practice? Donaldson and Crawley (1978) identified three themes of the nursing perspective: (1) concern with principles and laws that govern life processes, well-being, and optimum functioning of human beings, (2) concern with the patterning of human behavior in interaction with the environment in critical life situations, and (3) concern with the processes by which positive changes in health status are affected. These are the boundaries of the domain of nursing and the basis for nursing research. The unique perspective of nursing is most evident in the research questions posed. Exemplary investigations reported in a recent issue of Nursing Research are: (1) patient management of pain medication after cardiac surgery (King, Norsen, Robertson & Hicks, 1987), (2) predictors of dyspnea intensity in asthma (Janson-Bjerklie, Ruma, Stulbarg & Carrieri, 1987), (3) occurrences of symptoms in expectant fathers (Strickland, 1987), and (4) serum indirect bilirubin levels and meconium passage in early fed normal newborns (Boyer & Vidyasagar, 1987).

Nurse researchers often evaluate nursing interventions through the use of pre-existing or intact groups. A recent survey found that 27 percent of published nursing research was experimental in design, with two-thirds of such studies being quasi-experiments (Jacobsen & Meininger, 1985). Design issues can lead to problems in making correct inferences about the effects of interventions on outcomes. Cook and Reichardt (1976) discuss the importance of multiple analytic strategies when analyzing quasi-experimental data to avoid threats to statistical conclusion validity. Such techniques adjust for the initial nonequivalence between groups by using analysis of covariance with the pre-test scores entered as the covariate. However, the assumption of the covariate being specified and measured without error is typically not met, necessitating reliability adjustments to pre-test values. Despite the use of more sophisticated research designs and statistical techniques in nursing research, samples continue to be primarily nonprobability samples. Of the 90 sample plans of a random selection of research in <u>Nursing Research</u>, <u>Research in Nursing and Health</u>, and <u>The Western Journal of Nursing Research</u> (1980-1986), 74 related to applied research. Random samples were used in only eight of these studies. Such convenience sampling makes it difficult to effect generalization to the population.

Brown, Tanner and Padrick (1984) analyzed a sample of 137 research articles from <u>Nursing Research</u>, <u>International Journal of Nursing</u>

Studies, Research in Nursing and Health, and The Western Journal of Nursing Research, from 1952 to 1980. The following statements were included among their conclusions: clinical studies have increased, tendency to select relatively small samples by nonprobability, sampling unit tends to be an individual, more effort needed in data interpretation, experimental designs are often unethical or not feasible, insufficient knowledge to design appropriate experimental interventions, more longitudinal designs with repeated measures should be employed to study change processes and outcomes. Jacobsen and Meininger (1985) extended the Brown, Tanner and Padrick (1984) study by providing more detail and additional data points to examine trends. They did not include the International Journal of Nursing Studies. The issues identified included: lack of emphasis on random sampling and replication, lack of longitudinal and case-control designs, and lack of methods for reducing bias and strengthening causal interpretations from observational designs.

To generalize from these articles, it would seem that the major methodological issues confronting nursing research are statistical analysis and psychometric properties of data gathering devices. Norbeck (1987) defends the empirical approach and challenges researchers to design studies to: (1) identify and describe variables of interest, (2) establish relationships among variables and control for competing hypotheses, (3) test and define causal models, and (4) test applications of theory to previous findings. To date, there is minimal evidence to suggest that research findings are directly influencing the quality of patient care.

The traditional group comparison method of research has not readily been adaptable to applied nursing research. Achieving sufficient sample sizes (in context of the presumed effect size, randomized conditions, and known/expected heterogeneity of the subjects) is often difficult. Moreover, the ethical considerations of withholding treatment from a control group have also inhibited clinical research in nursing. Time series analysis has been suggested as a valuable approach to evaluating the effectiveness of nursing interventions (Metzger & Schultz, 1982; Jirovec, 1986), although the technique has not been applied in nursing studies. Time series analysis allows: (1) the establishment of the principal characteristics of a time series, (2) the determination of the nature of the system assumed to be generating a time series, (3) the forecasting of future values of a series, and (4) the specification of the relationships among time series.

Given nursing research objectives, it is important to develop an approach that detects if significant changes occurred and rules out competing hypotheses and threats to internal validity. The time series approach accounts for autocorrelation in nursing data to produce an unbiased estimate of error variance. The objective is to identify the ARIMA model which most accurately represents the stochastic process underlying each series, and to adequately account for autocorrelation in the series, resulting in white noise residuals that can be used for testing intervention effects.

Time series analysis allows for clinical experimentation without some of the difficulties inherent in experimental models. Random assignment to experimental and control groups is not an issue. The treatment effect can be quantitatively analyzed in terms of change within an individual through intrasubject design research. The individualistic focus of nursing is maintained, yet the change is analyzed statistically so that research consumers can base their decisions regarding the usefulness of a particular intervention on objective criteria.

The longitudinal perspective of serial data takes advantage of and parallels the natural clinical decision-making process (Hayes, 1981). Time series analysis lends itself to variables that can be repeatedly measured in the same subject or group of subjects; for example, response to illness, frequency of urinary incontinence, medication required per specified time period. Time series analysis may bring research to the practitioner level and may therefore be beneficial to nursing in establishing an empirically based practice. Rival alternative hypotheses are eliminated logically and sequentially: those remaining are supported by the data. Internal validity is strengthened through comparison with all variables of interest and possible explanations. Therefore, time series analysis assists in developing hypotheses, interpreting, and applying generic principles to subsequent subjects.

Therefore, as a human science, nursing takes into account various characteristics of the human realm. The process nature of human phenomena signifies continual growth and re-patterning over time. This implies the use of longitudinal time series designs to fully explicate the phenomena of interest (Metzger & Schultz, 1982). A feasible approach to clinical investigations is single subject research (Holm, 1983) as the use of an experimental design or a large sample size does not ensure that a study is relevant to nursing in the real world. Nursing science needs to develop methods of validating the theoretical constructs of nursing practice. Research methods are selected to reduce measurement error and increase the obability of significant findings. The results of clinical research add to the body of nursing knowledge, providing a basis for clinical practice. The inherent limitations of traditional experimental research procedures for yielding generalizations beyond the particular experimental paradigm have been expressed by Petrinovich (1979). However, time series analysis does not remedy the issue of correlation in errors due to inadequately collected data. It is only on the basis of theory that one can decide on an appropriate hypothesis to be tested, on a correct method of statistical analysis, and on whether the experimental results can be generalized to a population of interest (Serlin, 1987).

Silva (1986) reported that only nine out of 62 nursing studies address validation. In clinical nursing research, the true experimental design required for causal inference is difficult to implement: the dependent variable is not inert and can change over time regardless of treatment, there is a lack of precise measuring instruments, random assignment is often difficult to achieve with humans, control over variables is difficult to accomplish and often an irrelevant design to use for theory and practice, it is impossible to isolate phenomena under investigation from all extraneous sources, and there is a lack of explicit theories for determining effect size. Methods to evaluate theoretical constructs of nursing practice are required to provide evidence of concept identity, confirmation of concept linkages, explanation of prediction, and prescription.

In summary, serial dependencies within data sets are of particular concern in clinical investigations involving repeated measurements.

Such instances include many nursing intervention procedures such as regularizing or stabilizing interventions. There are other clinical interventions which involve more discrete outcomes which can be investigated either by time series or true experimentation, and there are others which must be subjected to true experimentation if an ethical and scientific approach is maintained, such as in developing the state of clinical investigation.

Problems Associated with the Analysis of Time Series Data

From a review of the literature, there has been an increase in the use of time series analysis in the social sciences. However, there appears to be a tendency to oversimplify the application or to use the technique inappropriately. Guidelines for the application of time series analysis are being developed. First, a sufficient number of data points for analysis should be obtained. A minimum of 50 observations has been suggested to adequately analyze time series data (McCleary & Hay, 1980). However, Gottman (1981) addresses the issue of the number of data points required for use in time series analysis. He found that using five baseline and five intervention points had a likelihood of Type I error actually less than five percent. It is the avoidance of Type II errors which requires more data points (Gottman, 1981). Second, the time series observations should be equally spaced, discrete, and interval level. Third, the time points should be sensitive to the intervention effects they intend to measure. A time series should be a true representation of the underlying process. Fourth, the structure of the time series needs to be identified in order to determine stationarity. Fifth, the degree of autocorrelation in the series should be determined. Apart from the conceptual assumption underlying behavioral investigations that performance at a given time (t) is a function of previous time points (t-1 to t-n) and the present time, there is the issue of the size of the autocorrelation which is necessary to assume dependence of data points. The preferred conservative procedure is to assume that any degree of autocorrelation is sufficient to affect future data points (Sharpley, 1986). Sixth, a tentative ARIMA model for the time series realization should be determined and its parameters estimated. A question of practical importance thus appears: What are the consequences of model misidentification? Of vital importance is estimating the integrative component (d value) correctly by checking that the autocorrelations are not more than twice the standard errors (Sharpley, 1986). Time series are also not free of a diverging confidence interval around the parameter estimates. The variance of error increases as time elapses from the last observation (Box & Jenkins, 1970). Seventh, the goodness of fit of the ARIMA model should be determined by minimizing the value of the residual mean square. Eighth, the pattern of change in the intervention effect should be identified and subsequently tested for statistical significance.

Often in the behavioral sciences, the complexity of the time intervention model is not known and fewer observations than 50 are available for study. What are the consequences of choosing the wrong model? Padia (1975) studied nine ARIMA models with various misspecifications and found that a failure to difference a nonstationary series leads to an underestimation of Type I errors and power. Unnecessary differencing tends to increase the residual variance of the random shock component achieved by ARIMA modeling. Also, as some ARIMA models are similar and may well be approximated by other models, simply misspecifying the model form may not always be very serious. However, the exact effects of misidentification are not clear, especially if the estimation procedure used finds the most likely values of the parameters for a given set of data and a given ARIMA model. Therefore, a wrong ARIMA model can over- or underestimate the statistical significance of interventions. A model should have an empirical basis as well as being parsimonious. Over-modeling may be statistically adequate but can lead to attributing variance to the random component instead of to the intervention component. Estimation of the intervention effect is a function of the ARIMA model structure.

Another question in intervention effect analysis is whether there is an appropriate time to intervene, given the choice. Glass, Willson, and Gottman (1975) indicate that choice of time for intervention does affect the precision of estimates and the power to detect real differences from zero in the intervention effect. It has been shown for the Integrated Moving Average model, that shorter confidence interval about the intervention effect occurs when the number of pre- and post-intervention observations are equal. For the Autoregressive model, as ϕ departs from zero, the number of post-interventions should be increased as the number of pre-interventions is decreased. Also, beyond 90 to 100 data points, extending the number of observations of a time series does not improve the estimation of the intervention component (Box & Jenkins, 1965).

Serial dependency as a property of behavioral data is being recognized as having an important influence on the judgment of graphical data. Jones and his associates argue that serial dependency cannot be appraised visually (Jones, Vaught, & Weinrott, 1977). Application of time series analysis as an alternative to the unreliable visual analysis is advocated by Jones, Weinrott, and Vaught (1978) and supported by De Prospero and Cohen (1979). The average magnitude of the errors and the bias in prediction was smaller for the time series analysis as compared to those made using the visual analysis techniques (Horne, Yong, & Ware, 1982).

Thus, even though time series analysis offers the benefit of providing a statistically powerful test for detecting intervention effects, guidelines are required about the structure and length of the data. More information is required regarding the accuracy of intervention effect estimates with limited data points and misidentified ARIMA models. Also, the Type I error rates and statistical power need to be examined as they are affected by the ARIMA model and the number of observations in a time series realization.

Statement of the Problem

Time series experimentation is fundamental to questions posed in social science research, that is, the study of change over time. Clinical data are generally represented by limited numbers of observations, non-random samples, non-random assignment to groups, heterogeneity within groups, plus the overall randomness in human response. Inferences based on estimated parameters from traditional statistical models can be biased. Violations in estimation occur because data (error process) are serially dependent (autocorrelated). As a consequence of autocorrelation, the error variance is underestimated, leading to narrow confidence bands, overestimation of parameters, and erroneously concluding that an intervention exerted a significant influence. What degree of autocorrelation is present in the time series data? The stochastic time series processes include this formalized relationship between error terms in ARIMA model identification. The ARIMA model building process can be hampered by serial correlation in the data. Assuming the correct ARIMA model, it is not the parameter estimates which may be biased by autocorrelation, but rather the estimate of the variance. With positive correlation the model can appear to fit the data better than it actually does (Gottman, 1981). The stationarity conditions of ARIMA processes are a prerequisite for modeling the autocorrelation of the time series realization, thus careful attention must be devoted to this assumption. What are the ramifications of this assumption?

In the study of change, the researcher wishes to detect small but significant effects which are buried within noisy data. Visual inspection alone may overlook small effects that are present and obscured by uncontrolled error. Exclusive use of clinical criteria for detecting change presents problems. Change by one clinician may not be defined by another in the same way. Other errors occur as initially small effects could combine to produce large effects in subsequent research and small intervention effects may be important in the future or show large effects in replication series, as in clinical drug trials. Inferential statistical tests complement other approaches by increasing the reliability of estimates. In other words, time series designs can assist in making judgements regarding the validity of the results, in that random fluctuations are not interpreted as intervention effects.

Interrupted time series analysis offers the flexibility of modeling almost any type of intervention effect. What is the accuracy and sensitivity of interrupted time series analysis? Are time series analysis estimates biased? Circumstances under which the application of these procedures is appropriate are under investigation. A fundamental question concerns the number of data points necessary to accurately estimate the effect of an intervention. The number of data points is a practical issue of power; more observations lead to better ARIMA model identification, which provides a better fit of the ARIMA model to the data, which in turn makes small effects easier to detect. The ARIMA model must first be identified on the basis of the estimated autocorrelation and partial autocorrelation functions. The observed realization of the time series process must be long enough to appropriately identify the p, d and q parameters. The question of sufficient length has not been answered in the literature. Furthermore, it is not known how severely the misidentification of the ARIMA model will distort the test of intervention effects. Is it possible that similar time series processes to the true ARIMA model will result in a negligible effect on the test of intervention effects?

Other issues of importance involve the consequences of violating the assumptions underlying time series processes. In practical applications, the theoretical assumptions of a statistical procedure are rarely fulfilled completely. Therefore, there are a variety of practical issues important to researchers who employ the interrupted time series quasi-experimental design. What is the robustness of time series procedures? What are the problems of applying these procedures to real clinical data sets? To address these issues, small sample properties of interrupted time series procedures will be examined empirically under a variety of conditions by conducting Monte Carlo simulations. Various research questions were investigated. With respect to detection of intervention effects: what is the accuracy of the estimates of the intervention effect when present? What is the actual Type I error rate when no intervention effect is present? What is the power to detect a non-zero intervention effect? What is the accuracy of intervention effect estimation for correctly identified and misidentified ARIMA models? With respect to time series model identification: What is the accuracy of parameter estimates? What is the sampling variability of the autocorrelation and partial autocorrelation functions? What is the power to detect a non-zero parameter? How do the error rates compare with nominal levels? In particular, the following discussion deals with the potential limitations and problems encountered in the application of time series analysis to small samples and the extent to which these potential problems may affect statistical inferences.

CHAPTER IV

Method of Investigation

The purpose of this investigation was to examine the use of Box-Tiao-Jenkins intervention analysis (1965, 1975, 1976) with small data sets of twenty and forty observations. The method of investigation was conducted as two sets of interrelated Monte Carlo simulations. Monte Carlo methods are computer-assisted simulation methods designed to obtain solutions to statistical problems by using random procedures and samples of random numbers. Statistical distributions are simulared with random numbers and violations of assumptions behind statistical tests of significance are introduced to study the consequences. The first set of Monte Carlo simulations investigated the estimation of the intervention component and the small sample properties of the test statistic for the analysis of intervention effects prescribed by Box and Tiao (1965, 1975). The second set of Monte Carlo simulations examined the small sample properties of the autocorrelation and partial autocorrelation functions, as they are utilized in the identification of ARIMA (p, d, q) time series models. The imp ance of the model identification stage of interrupted time series analysis cannot be underestimated, since it is a necessary prerequisite to the test of intervention effects.

ARIMA Model Representations

The ARIMA processes can be represented as:

(1) Autoregressive Model or AR(1)

$$Y_t = \phi_1 Y_{t-1} + a_t$$

(2) Moving Average Model or MA(1)

$$Y_t = a_t - \Theta_1 a_{t-1}$$

(3) Integrated Moving Average Model or IMA(1,1)

$$Y_t = Y_{t-1} - \Theta_1 a_{t-1} + a_t$$

(4) Nonstationary Model or NS

$$Y_t = Y_{t-1} + a_t$$

(5) White Noise Model or WN

$$Y_t = a_t$$

In each of these models, t refers to a measurement taken at time t. The value a_t is random error, and is assumed to be distributed normally and independently with a mean of zero and a constant variance. The coefficients ϕ and Θ are measures of the serial correlation in the data. The Autoregressive model represents a process in which the observation at time t is a function of the previous observation t-1. A Moving Average model represents a process in which an observation is a function of the previous random shock. The Nonstationary model represents an integrated process in which an observation at time t is the sum of the previous observation and a random shock. If the processes are represented solely as a function of random error, the Autoregressive model becomes the sum of an infinite number of random shocks of past time periods, each modified by a power of ϕ. The Moving Average model is a function of a finite number of previous random shocks, and the Integrated process becomes the sum of an infinite number of random shocks.

A simplified representation of an ARIMA model is ARIMA (p, d, q), where p is the number of autoregressive parameters, d is the order of differencing required to produce stationarity, and q is the number of moving average parameters. The order of differencing refers to the power (exponent) of B (backward shift operator) in the ARIMA model. The backward shift operator B can be represented as:

$$B^1Y_t = Y_{t-1} \; , \; B^2Y_t = Y_{t-2} \; \cdots \quad .$$

Thus, ARIMA (1 0 0) represents an Autoregressive model, ARIMA (0 0 1) represents a Moving Average model, ARIMA (0 1 1) represents an Integrated Moving Average model, ARIMA (0 1 0) represents a Nonstationary model, and ARIMA (0 0 0) represents White Noise.

The intervention component can be represented by

$$\frac{\omega}{1-\delta B} I_t$$

where I_t is a step variable equal to zero before the intervention, and one after the intervention. The parameter δ varies from zero to one and represents the way in which a change in level occurs. In this study, δ is zero, indicating an abrupt, permanent change with ω representing the magnitude of the change. Therefore, for the simple ARIMA intervention model, three parameters (ω, ϕ, θ) can be varied in addition to changing the number of observations. A summary of ARIMA models investigated and the parameters varied is presented in Figure 3.

Figure 3. Summary of ARIMA Processes Investigated

General Model

$$Y_t = \omega I_t + \frac{(1 - \theta B)}{(1 - \phi B)(1 - B)^d} a_t$$

MODELS INVESTIGATED

ARIMA (1 0 0) or AR(1)	Autoregressive
ARIMA (0 0 1) or MA(1)	Moving Average
ARIMA (0 1 1) or IMA(1,1)	Integrated Moving Average
ARIMA (0 1 0) or NS	Nonstationary
ARIMA (0 0 0) or WN	White Noise

Parameters Varied	Values Assumed			
Length of series or n	20	40		
Autoregressive Parameter or ϕ	0.2	0.5	0.8	
Moving Average Parameter or θ	0.2	0.5	0.8	
Order of Differencing or d	0	1		
Intervention Effect or ω	0.0	0.2	0.5	0.8

Placement of Intervention	Unbalanced n.1		Balanced n.2	
n=20	5	15	10	10
n=40	10	30	20	20

Strategies to Assess Intervention Effects

The strategies used to investigate intervention effects in short time series processes are presented in the following sections.

ARIMA Processes

The ARIMA processes investigated were limited to the five most commonly encountered in the social sciences as reported by Glass, Willson, and Gottman (1975). The values of ϕ and θ were limited to the interval (0,1), since a Moving Average model with small negative values can be approximated by an Autoregressive model with positive parameter values. The selected values for the correlation parameters were 0.2, 0.5, and 0.8. From a cursory review of the literature, the selected correlation parameter values are consistent with low to high values reported in the social sciences (Cook & Campbell, 1979; Jones, Vaught, & Weinrott, 1977). Specific questions addressed were: (1) What effect does ARIMA model identification have on intervention effects? (2) Do parameters change as a function of the intervention? (3) What are the consequences of choosing the wrong ARIMA process? (4) Which ARIMA model provides a good fit of data from time series interventions, or is all this complexity really necessary? (5) What ARIMA model can be used to generate effect sizes for changes in the level occurring as a result of an intervention?

Intervention Component

When delta (δ) equals zero, the intervention effect is represented by omega, which measures a change in the level of the time series. The change may be viewed as a change in the mean. Because serial correlation affects the variance, the size of an intervention effect that can be detected will vary with the magnitude of ϕ and Θ. Therefore, in order to compare the power of these methods across ARIMA models and various misspecifications, the intervention effect was calculated to take into consideration the differences in the variance. Omega (ω) was multiplied by the standard deviation of the observed time series (Y_t) to obtain a true value for the intervention component. Power depends on the Type I error rate, the variance in samples, and the size of the effect (Cohen, 1969). An effect size is represented as the difference between the two means divided by their common standard deviation. Four values for ω (0.0, 0.2, 0.5, or 0.8) were used to represent an effect size of zero, small, medium, and large. The intervention component was added in a balanced and unbalanced location to the time series realization. Specific questions addressed were: (1) What is the accuracy of intervention effect estimation with ARIMA models? (2) Is the empirical Type I error rate of intervention effect estimate greater than the nominal error rate? (3) Does the size of the intervention affect Type I error and power rates? (4) What is the power of intervention detection with the various ARIMA models? (5) Does the choice of time at which to intervene affect the precision of estimates and the power to detect real differences from zero in the intervention effect? (6) Does power to detect an intervention increase with sample size? (7) What effect do ARIMA models have on the accuracy of the standard error of intervention effect estimates?

Length of Time Series Realization

Fifty data points are considered by many researchers to be an absolute minimum for using time series analysis (Box & Jenkins, 1976; Gottman & Glass, 1978). In this study, data sets of twenty and forty points were examined. This length of series was considered a representative number of observations found in situations typically available in applied or clinical investigations (Barlow & Hersen, 1984). The effects of the varying number of short data points on parameter estimation, Type I error and power rates, and standard error estimates were determined. Specific questions addressed were: (1) What is the number of data points required for a baseline time series? (2) What is the number of data points necessary to detect intervention effects? (3) Does the percentage of null hypotheses rejections increase with the number of data points? (4) Does the standard error of estimates decrease with the increased number of data points? (5) Does the estimated autocorrelation function improve with the increased number of observations?

Power of the Test Statistic

Most tests of the null hypotheses require that certain assumptions be met if the results of the data analysis are to be meaningfully interpreted. The decision whether to accept or reject the null hypothesis is based upon a consideration of how probable it is that observed differences are due to chance alone. The stricter the criterion used for rejecting a null hypothesis, the greater the probability will be for accepting a false null hypothesis. A Type I error is rejecting a true null hypothesis. This would occur when the data indicate a statistically significant result, when in fact, there is no difference. The probability of making a Type I error (alpha) is often set at 0.01 or 0.05 (nominal value). The exact probability (true value) from tables and computer printouts, may be less than or greater than 0.01 or 0.05, depending upon how well assumptions are met. Monte Carlo methods were used to test the statistical characteristics of small samples. The consequences of violating the assumptions behind statistical tests of significance were studied by simulating statistical distributions with random numbers and introducing violations of assumptions into the procedure to study the consequences. Specific questions addressed were: (1) What is the Type II error rate involved for the various ARIMA models? (2) Is the empirical Type I error rate of intervention effect estimates greater than the nominal error rate for shorter length series? (3) Is there an underestimation of the standard error of the intervention effect which would increase Type I errors? (4) What is the power to detect true differences from the null hypothesis? (5) What is the relationship between ARIMA processes, number of data points, alpha level, and power?

Strategies to Assess ARIMA Model Identification

The strategies used to investigate ARIMA model identification are presented in the following section. The processes investigated were limited to the five previously discussed ARIMA models. The selected values for the correlation parameters were 0.2, 0.5, and 0.8. The amount and direction of bias in significance tests dictated by the nature and degree of dependency among the residuals were analyzed in the sample data series. The data under analysis were used to explore the issue of identification of the correct ARIMA model to fit the data so that subsequent tests for level changes in the series can be made. Specific questions addressed were: (1) What is the structure of serial dependency in the data? (2) Is there a significant level of autocorrelation present in the data? (3) What is the extent of bias in the estimates of the autocorrelation and partial autocorrelation coefficients with short time series? (4) What length of time series realization is necessary to ensure a reasonably high likelihood of an appropriately identified ARIMA model? (5) What risk is taken by processing data by methods which do not take the presence of autocorrelation into account when testing for change? (6) How much will traditional significance tests be inflated when serial correlation is not taken into account?

Simulation Procedures

The steps of the Monte Carlo simulations, along with the computer routines used in each step, are presented in the following sections.

Time Series Generation

For each specific condition in the two sets of Monte Carlo simulations, 500 time series realizations were randomly generated according to a given ARIMA (p, d, q) process using the TIME99 DERS program (Appendix A). The TIME99 program, based on IMSL routines, was programmed by D. Harley, Division of Educational Research Services, University of Alberta. The data generating program allows the user to specify the autoregressive, moving average, and white noise parameters of the time series process, as well as the degree of differencing, and length of the time series realization. An intervention effect of an abrupt, permanent change in level was imposed by adding a constant to the appropriate post-intervention time points. The residuals (a_t) are distributed normally and independently with a mean of zero and a variance of one. Because of the impact of starting conditions on the ARIMA series generated (Anderson, 1979), a specified number of data points were discarded from the time series.

Parameter Estimation

The time series realizations, correctly identified and misidentified, were submitted to the TIME02 DERS program (Appendix B) for estimation of the parameters via a least squares normal theory analysis. Box and Tiao (1965, 1975) demonstrated that a test for ARIMA intervention model parameters being significantly different from zero is provided by dividing the parameter estimates by its standard error; this ratio follows a t-distribution with n_1+n_2-k degrees of freedom, where k is the number of parameters estimated, and n_1 and n_2 . numbers of observations pre- and post-intervention, respectively.

For time series model identification, the TIME01 DERS program (Appendix C) was used to estimate autocorrelation and partial autocorrelation functions, as well as the chi-square statistic with k-p-q degrees of freedom. TIME01 and TIME02 are a pair of programs developed by C.P. Bower, W.L. Padia, and G.V. Glass (1974) at the Laboratory of Educational Research, University of Colorado and modified by D. Harley, Division of Educational Research Services, University of Alberta.

Type I Error and Power Estimates

SPSSx routines were used to calculate means and standard deviations of parameter estimates, as well as the region of nominal acceptance or rejection. Frequencies were accumulated for significant parameter values. Either Type I error or power rates were calculated from these frequencies.

Intervention Effect Estimation

This set of Monte Carlo simulations was designed to investigate the small sample properties of the test statistic of intervention effect proposed by Box and Tiao (1965, 1975). The specific issues examined were: (1) the estimation and accuracy of the intervention effect, (2) the distribution of the test statistic, (3) the statistical power of the test statistic, and (4) the accuracy of the estimates of standard error.

The form of the intervention effect considered is an abrupt, permanent change in level of a stationary time series process. This intervention effect was chosen for investigation on the basis of two considerations. First, an abrupt permanent change in level is the most common form of impact assessment in social science research applications (McCleary & Hay, 1980). Second, it is important to gain an understanding of the sampling properties of a straightforward intervention model before attempting to study more complicated intervention components involving additional parameters. In this study, the time of intervention was known and its presence was constant after its introduction.

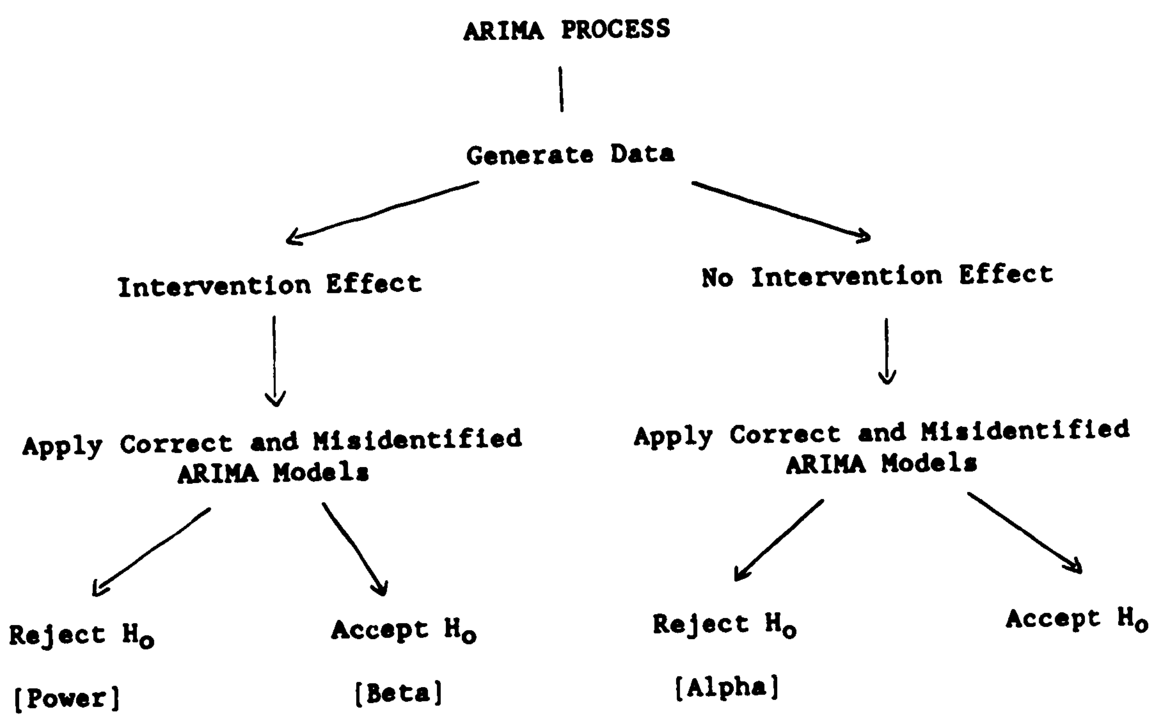
The scope of the study was also limited to the ARIMA models having at most one non-seasonal autoregressive and/or moving average parameter;

the order of differencing was no greater than one. Again, it is important to begin the systematic investigation of the intervention analysis procedure with the more basic time series processes. Thus, the full model investigated was:

$$Y_t = \frac{\omega}{1-\delta B} I_t + \frac{(1 - \theta B)}{(1 - \phi B)(1-B)} a_t$$

The design of the Monte Carlo experiment can be thought of as a completely crossed factorial experimental design. The factors that were systematically manipulated were: (1) the magnitude of the autocorrelation parameter (ϕ or θ = 0.2, 0.5, or 0.8), (2) the magnitude of the intervention parameter (ω = 0.0, 0.2, 0.5, or 0.8), (3) the placement of the intervention (balanced or unbalanced design), and (4) the length of the time series realization (n = 20 or 40 time points). The variance of the time series process is a function of both the white noise variance of the series and the autocorrelation of the series. The intervention parameter was adjusted by the standard deviation of the time series realization to produce the generated true value. For each of the five ARIMA models under consideration, 500 time series realizations were generated by the TIME99 DERS program. Each series was identified as the correct model and as each of the four misidentified models. For each correct ARIMA model and misidentification, the parameters were estimated and the test statistic calculated with the TIME02 DERS program. Estimates of the intervention component and its standard error were obtained. Conditions in which the ARIMA model did not include an intervention effect were also examined. SPSSx routines were employed to test for the normality of the distribution of the test statistic, to generate descriptive statistics of all the estimated parameters, and to calculate the percentage of statistically significant rejections of the null hypothesis. A summary of this set of simulations used to examine intervention effect estimation is illustrated in Figure 4.

Figure 4. Intervention Effect Estimation



Time Series Model Estimation

This set of Monte Carlo simulations was designed to investigate the sampling variability of the autocorrelation and partial autocorrelation functions under a variety of conditions. The bias of the estimates was considered given the parameters of the ARIMA process and the length of the time series realization. The specific issues examined were: (1) the sampling properties of the autocorrelation and partial autocorrelation functions, (2) the discrepancy between the estimated and empirical standard errors of the autocorrelation functions, (3) the Type I error rate and power of the statistical test for a non-zero autocorrelation coefficient, (4) the Type I error rate and power of the statistical test for a non-zero partial autocorrelation coefficient, and (5) the usefulness of the chi-square distribution with df=k-p-q, as an unbiased estimator of White Noise residuals.

For each condition examined, 500 data sets were generated by the TIME99 DERS program according to the parameters specified in the condition. The discrepancy between the mean of the 500 parameter estimates and the true parameter was used to assess the degree of bias in the various conditions. The sampling variability of the estimates was measured by computing the standard deviation of the 500 parameter estimates. This measure was considered the empirical estimate of the standard error of the autocorrelation and partial autocorrelation coefficients.

The five simple ARIMA (p, d, q) processes were considered, with three different parameter values examined for each of the five models. The AR(1) and MA(1) processes were generated with ϕ and Θ values of 0.2, 0.5, or 0.8, respectively. The order of differencing was no greater than one. The second factor varied was the number of time points in the data set. Each replication consisted of either 20 or 40 time points. The length of the realization was intended to examine small samples as most researchers consider 50 data points as an adequate number. Thus a total of 22 ARIMA processes were examined.

The estimates of the autocorrelation and partial autocorrelation functions, as well as the estimated standard error and Q-statistic, were computed with the TIME01 DERS program. SPSSx routines were employed to estimate the mean and standard deviation of the estimates. The mean values of the estimated standard errors were then compared with the empirical standard errors. Regions of nominal acceptance or rejection were constructed to investigate the Type I error rate and power of testing the null hypothesis.

CHAPTER V

Presentation and Discussion of Findings

The overall objective of the study was to investigate the estimation of intervention effects in short time series processes. The findings are discussed in two sections. In the first section, the intervention effect estimation in time series processes is presented, while in the second section time series model identification is discussed. Tables delineating the results are grouped at the end of relevant sections.

Intervention Effect Estimation in Time Series Processes

Monte Carlo simulations were used to investigate the small sample properties of the intervention procedure developed by Box and Tiao (1965, 1975). Five simple ARIMA models were investigated: a first-order Autoregressive model, a first-order Moving Average model, an Integrated Moving Average model, the Nonstationary model, and White Noise. Three characteristics of the time series process were systematically manipulated: (1) the magnitude of the parameter measuring serial correlation in the data (ϕ or θ = 0.2, 0.5, or 0.8); the Nonstationary model does not contain explicitly a parameter measuring serial correlation but is equivalent to an Autoregressive model with the correlation parameter ϕ equal to one, (2) the length of the time series (n = 20 or 40), and (3) the magnitude of the intervention parameter (ω = 0.2, 0.5, or 0.8) so as to represent small, medium, and large effects; in order to assess Type I error rates, a zero effect was also investigated. The intervention effect was added at the midpoint (balanced design) or at the uneven point (unbalanced design) of the time series process. Additionally, the value of the intervention parameter was multiplied by the standard deviation of the observed time series to obtain a true value for the intervention component. The white noise variance of the generated data sets was fixed at a constant value of one.

Least squares estimates of the intervention parameter were obtained for 500 replications of each condition. Regions of nominal acceptance or rejection were constructed around the estimates based on the t-distribution and approximation of the standard error of the estimator. The proportion of replications that resulted in the rejection of the null hypothesis, $H_0: \mu_1 - \mu_2 = 0$ (where μ_1 and μ_2 represent the pre- and post-intervention means of the time series process), was calculated for each condition. This measure provides an estimate of the empirical Type I error rates for conditions in which the null hypothesis is true, and of the power of the test statistic for conditions in which the intervention component is different from zero.

The primary concern in this study was the estimation of the intervention effect. Three issues were of interest: (1) the accuracy of the estimates of the intervention effect when present, (2) Type I error rates when no intervention effect was present, and (3) the power to detect a non-zero intervention effect. The findings presented discuss the intervention effect estimation and the time series model. In each case, the results are discussed for the correctly identified model and then for the misidentifications.

Estimates of Intervention Effect

For each specified model, the mean value of the intervention effect was calculated over 500 replications. The results of this set of simulations are presented in Tables 1 to 48, giving estimates of the intervention effect for both correctly identified and misidentified models. The true value of the intervention effect parameter is also presented for the purpose of comparison.

True model: Autoregressive processes. The correctly identified Autoregressive series with twenty observations gave intervention effect estimates that tended to be slightly higher or lower than the true value. The differences ranged from 0.00 to 0.076. The discrepancy between the true value and the estimated value increased as ϕ increased; often another model gave more accurate estimates than the true model, although the differences were very slight. Tables 1 and 2 show the estimated intervention effect for the Autoregressive model with twenty data points. There is little change over the discrepancy between the estimates and the true value for the estimated intervention effect for the Autoregressive model with forty data points. Tables 3 and 4 show the estimated intervention effects for the Autoregressive model with forty data points.

With twenty observations, misidentifying the Autoregressive series as a Moving Average model or White Noise did not give estimates that varied from the true value by any great difference than the correct model. In general, the estimates given by these two misidentifications were similar to those given by the correct model. Misidentification as the Nonstationary model gave estimates that were generally higher, while misidentification as the Integrated Moving Average model gave estimates that became increasingly discrepant as ϕ increased. With forty observations, the assumed Moving Average and White Noise models again gave estimates similar to the correct model. The Integrated Moving Average and Nonstationary models gave estimates that were generally less than the true value at all levels of ϕ. Therefore, the discrepancy in the estimates appears larger for the assumed Nonstationary and Integrated Moving Average Models as compared to the other misidentified models. This is more notable at low values of ϕ and for the larger number of observations. Generally, as the length of the series increased, the discrepancy in the intervention effect estimates decreased in the non-differenced models. Also, as ϕ increased, the discrepancy in the intervention effect estimates increased for all models, except for the Nonstationary model.

True model: Moving Average processes. For the correctly identified Moving Average model, the discrepancy between the intervention effect estimated value and the true value ranged from 0.001 to 0.036. There was a slight decrease in the discrepancy as the series length increased from twenty to forty data points and as Θ increased. Tables 5 to 8 show the estimated intervention effect for the Moving Average model with twenty and forty data points, respectively.

When the number of observations was twenty, the intervention effect estimates given by the assumed Autoregressive, White Noise, and Integrated Moving Average models were quite close to the correct model. The Nonstationary model generally gave the most discrepant estimates which increased slightly with increasing Θ. In the series with forty data points, the Autoregressive, White Noise, and Integrated Moving Average models gave estimates quite similar to the true model. The Nonstationary model again, had the largest discrepancy in the intervention effect estimates. As Θ increased, the discrepancy in the intervention effect estimates became larger in the differenced models, but decreased in the non-differenced models.

True model: Integrated Moving Average processes. For the correctly identified Integrated Moving Average model, the discrepancy between the intervention effect estimates given by the correctly identified model and the true value ranged from [illegible]00 to 0.07, with the largest differences occurring at a Θ of [illegible]. With forty data points, the difference ranged from 0.001 to 0.04, with the largest differences occurring at a Θ of 0.2. Tables 9 to 12 show the estimated intervention effect for the Integrated Moving Average model with twenty and forty data points, respectively.

With series of length twenty, misidentification of the Integrated Moving Average series as the Autoregressive, Moving Average, and White Noise models gave estimates that were nearly similar to the correct model. Misidentification as the Moving Average or White Noise model gave intervention effect estimates that tended to be slightly higher than the true value. The pattern of over and underestimation with the series length of forty in the Nonstationary series gave estimates generally lower than the true value. For the larger values of ω, the Autoregressive model overestimated the intervention effect. The Moving Average and White Noise models generally overestimated the intervention effect, although the difference decreased as Θ increased. The discrepancy in intervention effect estimates in the Moving Average model was relatively large. As Θ increased, the discrepancy in intervention effect estimates in the Autoregressive, Moving Average, and White Noise series decreased.

True model: Nonstationary processes. The results for the Nonstationary model with twenty observations indicated a difference between the true intervention effect and estimated values ranging from 0.03 to 0.16. With forty observations, the intervention effect estimates for the correctly identified model were more accurate, ranging from 0.04 to 0.11. Tables 13 and 14 show the estimated intervention effect for the Nonstationary model with twenty and forty data points, respectively.

The pattern of over and underestimation of the intervention effect is similar in the Nonstationary model to that found in the Integrated Moving Average model. For series with twenty data points, the Autoregressive and Integrated Moving Average models tend to slightly overestimate the intervention effect as compared to the true model. With the White Noise misidentification, slight underestimation of the intervention effect occurs. With forty observations, the Integrated Moving Average model gave intervention effect estimates slightly higher than the true model. The discrepancy in estimates of the intervention effect for the Moving Average and White Noise models increased for the longer series.

True model: White Noise. The discrepancy between the estimated intervention effect and the true value for the correctly identified White Noise model with twenty observations ranged from 0.000 to 0.025. With forty data points, the largest discrepancy was 0.019. Tables 15 and 16 show the estimated intervention effect for the White Noise model with twenty and forty data points, respectively.

The intervention effect estimates given by the misidentified Autoregressive and Moving Average models were similar to the true model regardless of the length of the series. The discrepancy in the estimates was greatest for the assume. Nonstationary model. For the other models, the discrepancy in the estimates of the intervention effect decreased as the length of the series increased.

Time series model effect on intervention effect estimates. In summary, selecting a model with the incorrect order of differencing and with forty observations led to estimates of the intervention effect that were either too large (not differencing when needed) or too small (differencing when not needed). With the incorrect order of differencing, the discrepancy in the estimates of the intervention effect was large. Assuming a White Noise model for a true Autoregressive or Moving Average series did not appear to give biased estimates of the intervention effect. This is the result one would expect from a theoretical perspective. However, the White Noise estimates for a model that requires differencing are not as accurate; that is, slightly underestimated for the shorter series and slightly overestimated for the longer series. For the non-differenced models, misidentification as a Nonstationary model usually led to estimates of the intervention effect that were discrepant. Misidentification as an Integrated Moving Average model gave estimates that were slightly less discrepant. In the case of misidentifying a differenced model as Moving Average or White Noise, the obtained estimates for the intervention effect were usually slightly lower than the true model. In general, the Autoregressive identifications are more accurate in intervention effect estimation with the longer series, regardless of the true model. The discrepancy in intervention effect estimates is somewhat smaller in time series realizations of greater length and balanced design. Therefore, in terms of error of estimation, the worst problem of misidentification occurs with incorrect differencing. When the correct time series model is specified, the procedure seems to give intervention effect estimates that are quite close to the true value; the Autoregressive model tended to give intervention effect estimates that were the least discrepant from the true value.

Table 1

Estimates of Intervention Effect

True Model: Autoregressive

Series Length = 20.1

Unbalanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
		True Omega			
		0.000	0.197	0.497	0.794
AR ϕ=0.2	AR	0.008	0.138	0.468	0.797
	MA	0.020	0.129	0.470	0.804
	IMA	0.004	0.155	0.438	0.803
	NS	-0.067	0.197	0.497	0.844
	WN	0.007	0.138	0.471	0.787
		True Omega			
AR ϕ=0.5		0.000	0.214	0.533	0.861
	AR	0.059	0.265	0.508	0.861
	MA	0.040	0.256	0.499	0.853
	IMA	0.051	0.253	0.454	0.853
	NS	0.113	0.278	0.485	0.882
	WN	0.052	0.271	0.539	0.853
		True Omega			
AR ϕ=0.8		0.000	0.268	0.663	1.054
	AR	-0.049	0.305	0.623	1.078
	MA	0.018	0.208	0.556	1.118
	IMA	-0.019	0.337	0.643	1.114
	NS	-0.018	0.310	0.660	1.085
	WN	0.011	0.201	0.563	1.134

Table 2

Estimates of Intervention Effect

True Model: Autoregressive

Series Length = 20.2

Balanced Design

True Model	Assumed Model	Zero	Small	Medium	Large
			Effect Size		
AR			True Omega		
ϕ=0.2		0.000	0.196	0.495	0.801
	AR	-0.002	0.183	0.525	0.765
	MA	-0.006	0.172	0.526	0.767
	IMA	-0.003	0.209	0.527	0.765
	NS	-0.010	0.262	0.590	0.802
	WN	0.002	0.180	0.517	0.766
AR			True Omega		
ϕ=0.5		0.000	0.213	0.544	0.851
	AR	-0.011	0.287	0.550	0.927
	MA	-0.022	0.261	0.531	0.903
	IMA	-0.062	0.263	0.547	0.948
	NS	-0.094	0.191	0.544	0.974
	WN	-0.013	0.267	0.544	0.901
AR			True Omega		
ϕ=0.8		0.000	0.264	0.653	1.065
	AR	0.025	0.270	0.683	1.047
	MA	0.091	0.163	0.706	1.115
	IMA	0.030	0.268	0.691	1.039
	NS	0.009	0.307	0.677	1.044
	WN	0.098	0.137	0.727	1.111

Table 3

Estimates of Intervention Effect

True Model: Autoregressive

Series Length = 40.1

Unbalanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
AR ϕ=0.2		True Omega			
		0.000	0.202	0.502	0.814
	AR	0.011	0.190	0.508	0.764
	MA	0.013	0.187	0.509	0.768
	IMA	0.006	0.188	0.517	0.747
	NS	0.055	0.239	0.515	0.842
	WN	0.009	0.184	0.506	0.766
AR ϕ=0.5		True Omega			
		0.000	0.222	0.554	0.895
	AR	0.008	0.207	0.517	0.918
	MA	0.007	0.209	0.509	0.921
	IMA	-0.009	0.232	0.542	0.882
	NS	0.034	0.222	0.546	0.856
	WN	0.010	0.209	0.505	0.932
AR ϕ=0.8		True Omega			
		0.000	0.291	0.731	1.184
	AR	0.055	0.248	0.749	1.119
	MA	0.103	0.304	0.775	1.163
	IMA	-0.020	0.229	0.681	1.158
	NS	-0.029	0.232	0.671	1.144
	WN	0.122	0.326	0.780	1.182

Table 4

Estimates of Intervention Effect

True Model: Autoregressive

Series Length = 40.2

Balanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
			True Omega		
AR ϕ=0.2		0.000	0.203	0.500	0.804
	AR	0.018	0.194	0.484	0.800
	MA	0.016	0.194	0.483	0.798
	IMA	0.027	0.178	0.480	0.792
	NS	-0.043	0.201	0.475	0.764
	WN	0.020	0.194	0.485	0.796
			True Omega		
AR ϕ=0.5		0.000	0.221	0.557	0.895
	AR	-0.030	0.190	0.573	0.911
	MA	-0.034	0.181	0.568	0.897
	IMA	-0.044	0.254	0.556	0.883
	NS	-0.006	0.222	0.578	0.903
	WN	-0.034	0.181	0.571	0.903
			True Omega		
AR ϕ=0.8		0.000	0.294	0.731	1.195
	AR	0.025	0.273	0.696	1.173
	MA	0.011	0.256	0.775	1.279
	IMA	-0.009	0.300	0.647	1.156
	NS	-0.003	0.318	0.657	1.162
	WN	0.005	0.267	0.799	1.300

Table 5

Estimates of Intervention Effect

True Model: Mo ing Average

Series Length = 20.1

Unbalanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
MA θ=0.2		True Omega			
		0.000	0.206	0.508	0.815
	MA	-0.003	0.195	0.507	0.793
	AR	-0.014	0.194	0.506	0.796
	IMA	-0.010	0.193	0.496	0.795
	NS	0.003	0.105	0.422	0.862
	WN	-0.016	0.192	0.499	0.797
MA θ=0.5		True Omega			
		0.000	0.224	0.561	0.898
	MA	0.019	0.226	0.532	0.905
	AR	0.018	0.220	0.538	0.916
	IMA	0.024	0.218	0.540	0.909
	NS	0.066	0.142	0.638	0.852
	WN	0.024	0.218	0.540	0.909
MA θ=0.8		True Omega			
		0.000	0.261	0.644	1.032
	MA	0.002	0.297	0.636	1.025
	AR	-0.006	0.286	0.643	1.032
	IMA	-0.006	0.286	0.648	1.025
	NS	-0.012	0.267	0.723	0.964
	WN	-0.006	0.286	0.648	1.025

Table 6

Estimates of Intervention Effect

True Model: Moving Average

Series Length = 20.2

Balanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
		True Omega			
MA θ=0.2		0.000	0.205	0.505	0.805
	MA	0.007	0.197	0.519	0.774
	AR	0.008	0.190	0.517	0.780
	IMA	0.008	0.186	0.520	0.783
	NS	-0.001	0.097	0.495	0.839
	WN	0.012	0.188	0.515	0.784
		True Omega			
MA θ=0.5		0.000	0.225	0.561	0.907
	MA	0.010	0.212	0.533	0.941
	AR	0.010	0.211	0.547	0.933
	IMA	0.021	0.213	0.544	0.930
	NS	0.088	0.246	0.671	0.857
	WN	0.017	0.213	0.550	0.930
		True Omega			
MA θ=0.8		0.000	0.260	0.639	1.042
	MA	0.008	0.242	0.643	1.040
	AR	0.011	0.248	0.646	1.035
	IMA	0.009	0.256	0.648	1.029
	NS	0.007	0.375	0.685	0.904
	WN	0.009	0.256	0.648	1.029

Table 7

Estimates of Intervention Effect

True Model: Moving Average

Series Length = 40.1

Unbalanced Design

True Model	Assumed Model	Effect Size			
		Zero	Small	Medium	Large
		True Omega			
MA θ=0.2		0.000	0.203	0.507	0.821
	MA	0.011	0.201	0.525	0.786
	AR	0.011	0.200	0.526	0.784
	IMA	0.014	0.199	0.529	0.782
	NS	0.019	0.113	0.735	0.772
	WN	0.012	0.198	0.528	0.783
		True Omega			
MA θ=0.5		0.000	0.224	0.559	0.901
	MA	-0.001	0.233	0.549	0.908
	AR	0.004	0.232	0.552	0.901
	IMA	0.001	0.231	0.553	0.897
	NS	-0.123	0.212	0.570	0.772
	WN	0.001	0.231	0.553	0.897
		True Omega			
MA θ=0.8		0.000	0.258	0.636	1.012
	MA	-0.002	0.248	0.626	1.016
	AR	0.001	0.256	0.632	1.011
	IMA	0.005	0.254	0.630	1.014
	NS	0.102	0.215	0.621	1.054
	WN	0.005	0.254	0.630	1.014

Table 8

Estimates of Intervention Effect

True Model: Moving Average

Series Length = 40.2

Balanced Design

		Effect Size			
True Model	Assumed Model	Zero	Small	Medium	Large
MA $\theta = 0.2$		True Omega			
		0.000	0.203	0.507	0.814
	MA	-0.038	0.198	0.493	0.802
	AR	-0.036	0.197	0.495	0.802
	IMA	-0.036	0.198	0.496	0.802
	NS	0.006	0.173	0.535	0.689
	WN	-0.036	0.198	0.496	0.800
MA $\theta = 0.5$		True Omega			
		0.000	0.227	0.558	0.900
	MA	-0.001	0.231	0.546	0.916
	AR	-0.002	0.232	0.547	0.917
	IMA	-0.002	0.237	0.551	0.915
	NS	-0.054	0.378	0.726	0.833
	WN	-0.002	0.236	0.550	0.916
MA $\theta = 0.8$		True Omega			
		0.000	0.254	0.648	1.022
	MA	0.003	0.255	0.644	1.025
	AR	0.004	0.257	0.639	1.023
	IMA	0.005	0.260	0.640	1.025
	NS	0.064	0.331	0.662	1.068
	WN	0.005	0.260	0.640	1.026

Table 9

Estimates of Intervention Effect

True Model: Integrated Moving Average

Series Length = 20.1

Unbalanced Design

		Effect Size			
True Model	Assumed Model	Zero	Small	Medium	Large
			True Omega		
IMA θ=0.2		0.000	0.253	0.629	0.992
	IMA	0.037	0.278	0.651	1.007
	MA	0.028	0.282	0.683	1.008
	AR	0.082	0.301	0.652	0.997
	WN	0.093	0.310	0.622	1.006
	NS	0.105	0.337	0.634	0.974
			True Omega		
IMA θ=0.5		0.000	0.242	0.603	0.975
	IMA	-0.044	0.263	0.592	0.944
	MA	-0.007	0.248	0.611	0.959
	AR	-0.037	0.247	0.606	0.988
	WN	-0.036	0.238	0.609	0.984
	NS	-0.005	0.248	0.552	0.980
			True Omega		
IMA θ=0.8		0.000	0.258	0.651	1.030
	IMA	0.038	0.249	0.667	0.987
	MA	0.035	0.309	0.664	0.973
	AR	0.036	0.264	0.641	0.992
	WN	0.039	0.267	0.648	0.993
	NS	0.045	0.276	0.615	0.880

Table 10

Estimates of Intervention Effect

True Model: Integrated Moving Average

Series Length = 20.2

Balanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
IMA θ=0.2	True Omega	0.000	0.250	0.631	1.000
	IMA	-0.070	0.198	0.594	0.997
	MA	-0.053	0.221	0.604	0.996
	AR	-0.049	0.221	0.599	0.985
	WN	-0.047	0.228	0.614	1.002
	NS	-0.048	0.207	0.565	1.057
IMA θ=0.5	True Omega	0.000	0.240	0.620	0.983
	IMA	0.036	0.240	0.687	0.943
	MA	0.034	0.262	0.685	0.971
	AR	0.041	0.270	0.679	0.985
	WN	0.036	0.273	0.669	1.001
	NS	-0.029	0.316	0.646	0.983
IMA θ=0.8	True Omega	0.000	0.260	0.653	1.035
	IMA	-0.012	0.228	0.659	1.066
	MA	-0.008	0.242	0.659	1.055
	AR	-0.001	0.241	0.666	1.055
	WN	0.001	0.245	0.663	1.045
	NS	0.085	0.273	0.628	1.011

Table 11

Estimates of Intervention Effect

True Model: Integrated Moving Average

Series Length = 40.1

Unbalanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
IMA θ=0.2	True Omega	0.000	0.256	0.638	1.026
	IMA	-0.034	0.223	0.602	0.994
	MA	-0.005	0.226	0.640	1.042
	AR	-0.019	0.207	0.622	1.011
	WN	-0.017	0.215	0.636	1.001
	NS	-0.013	0.213	0.555	0.958
IMA θ=0.5	True Omega	0.000	0.243	0.606	0.975
	IMA	-0.004	0.231	0.584	1.001
	MA	-0.031	0.216	0.586	0.977
	AR	-0.006	0.232	0.581	0.984
	WN	-0.006	0.235	0.582	0.984
	NS	0.021	0.148	0.587	1.013
IMA θ=0.8	True Omega	0.000	0.259	0.655	1.034
	IMA	0.012	0.251	0.649	1.035
	MA	0.028	0.244	0.637	1.031
	AR	0.015	0.249	0.648	1.035
	WN	0.012	0.250	0.649	1.035
	NS	0.022	0.104	0.662	1.067

Table 12

Estimates of Intervention Effect

True Model: Integrated Moving Average

Series Length = 40.2

Balanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
IMA θ=0.2			True Omega		
		0.000	0.255	0.640	1.015
	IMA	0.008	0.289	0.626	0.974
	MA	-0.014	0.282	0.641	1.018
	AR	-0.016	0.298	0.672	1.011
	WN	-0.019	0.302	0.677	1.022
	NS	0.011	0.264	0.628	0.954
IMA θ=0.5			True Omega		
		0.000	0.246	0.611	0.978
	IMA	-0.026	0.238	0.627	0.973
	MA	-0.007	0.220	0.620	0.972
	AR	-0.013	0.240	0.617	0.978
	WN	-0.015	0.244	0.615	0.976
	NS	0.000	0.173	0.611	1.079
IMA θ=0.8			True Omega		
		0.000	0.259	0.648	1.051
	IMA	0.012	0.265	0.657	1.049
	MA	0.012	0.245	0.656	1.050
	AR	0.010	0.264	0.657	1.050
	WN	0.009	0.265	0.656	1.049
	NS	-0.019	0.365	0.882	1.070

Table 13

Estimates of Intervention Effect

True Model: Nonstationary

Series Length = 20.1

Unbalanced Design

True Model	Assumed Model	Effect Size: Zero	Effect Size: Small	Effect Size: Medium	Effect Size: Large
		True Omega			
NS		0.000	0.269	0.682	1.095
	NS	0.163	0.222	0.578	1.167
	MA	-0.015	0.309	0.623	1.117
	AR	0.035	0.268	0.587	1.180
	IMA	0.127	0.274	0.648	1.138
	WN	-0.002	0.280	0.650	1.178

Series Length = 20.2

Balanced Design

True Model	Assumed Model	Zero	Small	Medium	Large
		True Omega			
NS		0.000	0.274	0.671	1.084
	NS	-0.029	0.398	0.618	0.992
	MA	0.002	0.354	0.714	1.036
	AR	0.017	0.327	0.733	1.013
	IMA	0.038	0.306	0.646	1.043
	WN	0.018	0.360	0.768	1.032

Table 14

Estimates of Intervention Effect

True Model: Nonstationary

Series Length - 40.1

Unbalanced Design

True Model	Assumed Model	Effect Size			
		Zero	Small	Medium	Large
		True Omega			
NS		0.000	0.275	0.699	1.110
	NS	0.095	0.236	0.616	1.106
	MA	0.020	0.238	0.692	1.109
	AR	0.030	0.249	0.662	1.096
	IMA	0.048	0.271	0.705	1.105
	WN	0.020	0.252	0.659	1.097

Series Length - 40.2

Balanced Design

True Model	Assumed Model	Zero	Small	Medium	Large
		True Omega			
NS		0.000	0.277	0.690	1.114
	NS	-0.112	0.231	0.643	1.109
	MA	0.002	0.212	0.692	1.096
	AR	-0.031	0.201	0.702	1.113
	IMA	-0.068	0.184	0.686	1.165
	WN	-0.013	0.190	0.704	1.102

Table 15

Estimates of Intervention Effect

True Model: White Noise

Series Length = 20.1

Unbalanced Design

True Model	Assumed Model	Zero	Effect Small	Size Medium	Large
			True Omega		
WN		0.000	0.200	0.494	0.798
	WN	0.011	0.175	0.501	0.797
	MA	0.000	0.176	0.510	0.812
	AR	0.009	0.174	0.506	0.801
	IMA	-0.011	0.181	0.505	0.801
	NS	-0.037	0.181	0.392	0.833

Series Length = 20.2

Balanced Design

			True Omega		
WN		0.000	0.198	0.490	0.799
	WN	-0.009	0.198	0.478	0.779
	MA	-0.006	0.196	0.484	0.769
	AR	-0.006	0.199	0.480	0.777
	IMA	-0.011	0.214	0.480	0.765
	NS	0.009	0.320	0.613	0.808

Table 16

Estimates of Intervention Effect

True Model: White Noise

Series Length - 40.1

Unbalanced Design

True Model	Assumed Model	Effect Size: Zero	Small	Medium	Large
		True Omega			
WN		0.000	0.199	0.498	0.784
	WN	0.009	0.208	0.496	0.783
	MA	0.008	0.209	0.502	0.784
	AR	0.009	0.209	0.500	0.783
	IMA	0.011	0.212	0.506	0.787
	NS	0.023	0.274	0.488	0.765

Series Length - 40.2

Balanced Design

		True Omega			
WN		0.000	0.198	0.493	0.791
	WN	-0.018	0.197	0.474	0.778
	MA	-0.019	0.196	0.473	0.778
	AR	-0.017	0.196	0.473	0.778
	IMA	-0.011	0.191	0.481	0.773
	NS	-0.028	0.180	0.538	0.815

Estimates of Type I Error for Intervention Effect

The Type I error rate was determined by the proportion of intervention effects found to be significant when in fact, the intervention effect was zero. The test for significance was done at the 0.05 level using a two-tail test with appropriate degrees of freedom for the number of observations. The results are presented in Tables 17 to 21, which give the estimates of Type I error for both correctly identified and misidentified ARIMA models. The estimates for the twenty data point series and the forty data point series appear in the same table.

True model: Autoregressive processes. An examination of the Type I error rate shows that for all conditions the empirical rate of rejection of the null hypothesis is greater than the nominal error rate. Table 17 gives the estimate of Type I error rates for the Autoregressive model. The inflation of Type I error is somewhat smaller for time series realizations of forty observations. For twenty observations, the Type I error rate fluctuated between 0.08 and 0.22, depending on the value of ϕ. For the longer series, the Type I error rate ranged from 0.08 to 0.16. Furthermore, the inflation of Type I error becomes more severe as the value of the autoregressive parameter increases. Misidentifying the Autoregressive series as either Integrated Moving Average or Nonstationary models, resulted in Type I error rates that were approximately the same as or slightly lower than the rates obtained for the true model. However, misidentifying the series as White Noise or the Moving Average model led to Type I error rates that were higher than the true model, increasing as ϕ and the length of the series increased. For the other assumed models, error rates decreased slightly with the longer series.

<u>True model: Moving Average processes</u>. In the Moving Average model with twenty observations, the Type I error rate appeared to decrease as θ increased, moving from 0.37 to 0.31 with the unbalanced design and 0.36 to 0.21 with the balanced design. With the longer series, the error rate appeared to increase as θ increased. At a θ of 0.8, the error rates increased as the series length increased. Otherwise, the inflation of Type I error for other values of θ is somewhat smaller for balanced time series realizations of greater length. Estimates of Type [illegible] error rates for the Moving Average model are presented in Table 18. The issue of inflated Ty[illegible] error rates will be discussed in subsequent sections when possible [illegible]ions are considered. The Moving Average model is the only model [illegible] which misidentifications had a consistent effect on Type I error. In all cases, the misidentifications yielded relatively low Type I error rates; these rates ranged from 0.00 to 0.06, and were lower than the rates obtained for the true model. Therefore, misidentifying a Moving Average model appears to have a rather conservative effect. Except for the assumed Nonstationary model, all Type I error rates decreased as θ increased.

<u>True model: Integrated Moving Average processes</u>. In the Integrated Moving Average model with twenty and forty observations, the Type I error rate fluctuated between 0.00 and 0.14, with the rate decreasing as the length of the series and value of θ increased. The rate was fairly constant between the unbalanced and balanced designs. The Type I error rate estimates for the Integrated Moving Average model are presented in Table 19. As indicated by the values in Table 19, misidentifying the Integrated Moving Average model as a Moving Average model resulted in higher Type I error rates than were obtained for the true model; these

rates ranged from 0.04 to 0.32. For the other assumed models, the Type I error rate decreased somewhat as Θ increased and were less for the balanced design and longer series. For most conditions, the empirical rate of rejection of the null hypothesis is greater than the nominal error rate.

True model: Nonstationary processes. The Nonstationary model with both twenty and forty data points had estimated Type I error rates near the nominal error rate of 0.05. The estimated Type I error rates for the Nonstationary model are presented in Table 20. The Nonstationary model misidentified as White Noise and Integrated Moving Average models had inflated error rates, while the misidentified Autoregressive model had rates only slightly higher than the true model. The estimated Type I error rates were overall, slightly higher for the balanced design and longer series. In the case of twenty observations with the Moving Average unbalanced design model, the estimated Type I error rate was 0.33; this rate decreased to nominal error level in the longer series.

True model: White Noise. The White Noise model with both twenty and forty observations had estimated Type I error rates near the nominal rate of 0.05, similar to the Nonstationary model. The estimated Type I error rates for the White Noise model are presented in Table 21. Once again, the error rates for the assumed Moving Average model were high, ranging from 0.10 to 0.31. The misidentified Integrated Moving Average and Nonstationary models have error rates somewhat closer to the nominal error rate. Little change in error rates occurred between length or design of the series. The assumed Autoregressive model had inflated error rates of 0.09 and 0.11 for the unbalanced and balanced series of twenty observations, respectively.

<u>Time series model effect on Type I error estimates</u>. In summary, as with estimation of intervention effects, the models group together by whether or not they require differencing. The effect of misidentification as a differenced model generally yielded relatively lower Type I error rates. Alternatively, higher Type I error rates occurred for misidentifications in which a non-differenced model was assumed for the Nonstationary or Integrated Moving Average model or the comparable Autoregressive model with a ϕ value of 0.8. Misidentifying a series as an Autoregressive model yielded Type I error rates consistently close to the nominal error rate. The Moving Average model had inflated Type I error rates for both the true and misidentified cases. The inflation of Type I error is somewhat smaller for time series realizations of greater length and balanced design. Furthermore, the inflation becomes more severe as the value of the autoregressive or moving average parameter increases. Therefore, the difficulty encountered was a consistently higher empirical Type I error rate than the nominal level for time series of length twenty and even for some time series realizations of forty data points.

Table 17

Estimated Type I Error Rates for Intervention Effect

True Model: Autoregressive

Nominal Alpha = 0.05

Number of Replications = 500

True Model	Assumed Model	Type I Error n=20.1 Unbalanced Design	Type I Error n=20.2 Balanced Design	Type I Error n=40.1 Unbalanced Design	Type I Error n=40.2 Balanced Design
AR $\phi=0.2$	AR	.08	.11	.08	.08
	MA	.23	.24	.09	.10
	IMA	.09	.10	.11	.11
	NS	.04	.03	.06	.03
	WN	.09	.10	.11	.11
AR $\phi=0.5$	AR	.12	.16	.12	.13
	MA	.22	.20	.17	.17
	IMA	.19	.26	.19	.18
	NS	.06	.06	.07	.04
	WN	.20	.27	.25	.28
AR $\phi=0.8$	AR	.19	.22	.16	.16
	MA	.43	.36	.39	.42
	IMA	.21	.25	.10	.09
	NS	.04	.06	.04	.04
	WN	.43	.46	.51	.54

Table 18

Estimated Type I Error Rates for Intervention Effect

True Model: Moving Average

Nominal Alpha = 0.05

Number of Replications = 500

True Model	Assumed Model	Type I Error n=20.1 Unbalanced Design	Type I Error n=20.2 Balanced Design	Type I Error n=40.1 Unbalanced Design	Type I Error n=40.2 Balanced Design
MA θ=0.2	MA	.37	.36	.15	.18
	AR	.06	.05	.04	.07
	IMA	.02	.01	.01	.01
	NS	.04	.05	.06	.05
	WN	.01	.01	.01	.02
MA θ=0.5	MA	.41	.37	.31	.29
	AR	.02	.02	.01	.01
	IMA	.00	.01	.00	.00
	NS	.05	.05	.04	.06
	WN	.00	.00	.00	.00
MA θ=0.8	MA	.31	.21	.44	.31
	AR	.01	.00	.00	.00
	IMA	.00	.00	.00	.00
	NS	.05	.04	.04	.05
	WN	.00	.00	.00	.00

Table 19

Estimated Type I Error Rates for Intervention Effect

True Model: Integrated Moving Average

Nominal Alpha = 0.05

Number of Replications = 500

True Model	Assumed Model	Type I Error n=20.1 Unbalanced Design	Type I Error n=20.2 Balanced Design	Type I Error n=40.1 Unbalanced Design	Type I Error n=40.2 Balanced Design
IMA θ=0.2	IMA	.14	.14	.12	.11
	MA	.32	.10	.05	.04
	AR	.09	.06	.05	.04
	WN	.12	.11	.10	.11
	NS	.06	.05	.04	.03
IMA θ=0.5	IMA	.07	.07	.04	.03
	MA	.31	.20	.05	.05
	AR	.04	.04	.02	.01
	WN	.04	.04	.04	.03
	NS	.04	.07	.05	.06
IMA θ=0.8	IMA	.02	.01	.00	.00
	MA	.32	.19	.15	.13
	AR	.02	.00	.00	.00
	WN	.01	.00	.00	.00
	NS	.05	.06	.05	.04

Table 20

Estimated Type I Error Rates for Intervention Effect

True Model: Nonstationary

Nominal Alpha = 0.05

Number of Replications = 500

		Type I Error	Type I Error	Type I Error	Type I Error
True Model	Assumed Model	n=20.1 Unbalanced Design	n=20.2 Balanced Design	n=40.1 Unbalanced Design	n=40.2 Balanced Design
NS					
	NS	.03	.04	.05	.06
	MA	.33	.08	.05	.07
	AR	.07	.09	.05	.07
	IMA	.18	.20	.18	.19
	WN	.15	.16	.14	.17

Table 21

Estimated Type I Error Rates for Intervention Effect

True Model: White Noise

Nominal Alpha = 0.05

Number of Replications = 500

True Model	Assumed Model	Type I Error n=20.1 Unbalanced Design	Type I Error n=20.2 Balanced Design	Type I Error n=40.1 Unbalanced Design	Type I Error n=40.2 Balanced Design
WN					
	WN	.04	.06	.05	.06
	MA	.31	.29	.10	.11
	AR	.09	.11	.06	.07
	IMA	.04	.07	.05	.05
	NS	.04	.04	.05	.05

<u>Estimates of Power for Intervention Effect</u>

Power was determined by the proportion of intervention effects found to be significant when in fact, the intervention effect did not equal zero. The test of significance was done at the 0.05 level with a two-tail test using the appropriate degrees of freedom for the number of observations. The results are presented in Tables 22 to 29, which give the estimates of power for both correctly identified and misidentified ARIMA models. As mentioned previously, the white noise variance of the generated data sets was fixed at 1.0 in order to facilitate the comparison of conditions in which the correlation parameter differed. The results indicate the percentage of rejections of the null hypothesis, $H_o: \mu 1 - \mu 2 = 0$, for situations in which the intervention component is equal to 0.2, 0.5, or 0.8. Obviously, the power of the test statistic increases in proportion to the Type I error rate and as the magnitude of the intervention parameter becomes larger.

<u>True model: Autoregressive processes</u>. The impact of the autoregressive parameter on the power of the test statistic is of interest. The power of the test statistic for a large effect size diminishes as the autoregressive parameter becomes larger. The estimated power ratio for the Autoregressive model with misidentifications are presented in Tables 22 and 23, illustrating the importance of serial correlation in determining the probability of correctly rejecting the null hypothesis at the 0.05 level. The length of the time series realization is also important in determining the power of the test statistic. Forty data points is apparently not a sufficient length to assume certainty in rejecting the null hypothesis when an intervention effect is in fact present. It can be seen that there is very little power (only about

one-third of the test statistics fall in the region of rejection) for an Autoregressive time series realization of forty data points and a large autoregressive parameter of 0.8. Even a large intervention effect size results in only rejecting the null hypothesis in 34 percent of the replications at the 0.05 level. Therefore, as the magnitude of the autoregressive parameter increases, the ability to detect an intervention effect diminishes.

Misidentifying the Autoregressive model as a differenced model, either Nonstationary or Integrated Moving Average, resulted in lower power in all cases. Misidentifying the Autoregressive model as either a Moving Average or White Noise model resulted in power estimates that were slightly higher than those of the true model. This result is consistent with the higher Type I error rates obtained for these same misidentifications in the Autoregressive model when the intervention effect was zero.

True model: Moving Average processes. Generally, the power to detect small, medium, and large effects increased as the moving average parameter increased and as the series length increased with a balanced design. The estimated power rates for the Moving Average model with misidentifications are presented in Tables 24 and 25. The Moving Average model had consistent effects across all misidentifications as in the situation of Type I error. Misidentifications led to power values that were lower than those obtained for the true model. Generally, the power of all misidentified models decreased as θ increased. Also, misidentifying the Moving Average model as a differenced model resulted in a substantial reduction in power across all values of θ and ω. Misidentifying the Moving Average model as either White Noise or Autoregressive models also reduced power, although the power of the assumed Autoregressive model was usually greater than the power of the assumed White Noise model. Therefore, misidentifying a Moving Average model resulted in generally lower power estimates.

True model: Integrated Moving Average processes. The power of the true Integrated Moving Average model to detect non-zero intervention effects decreased as the serial correlation increased. The power values increased with intervention effect size and length of series. The estimated power rates for the Integrated Moving Average model are presented in Tables 26 and 27. The power values are very low for small and medium effect sizes but increase substantially for a large effect size at $\Theta = 0.8$. As anticipated from the higher Type I error for misidentifying the Integrated Moving Average model as a non-differenced model, these same misidentifications resulted in higher estimates of power. For a small intervention effect size, these values of power were closer to the true model for the misidentified Autoregressive and White Noise series. In the forty observation series, power estimates of the misidentified Autoregressive series are slightly lower than the true model. In the twenty observation series, the misidentified Moving Average series has power estimates consistently higher than the true model, which increase as the value of Θ increases. For the other misidentified models, power values decreased as the serial correlation increased. Misidentification of the true Integrated Moving Average as a Nonstationary model resulted in power values that were fairly constant across all values of the correlation parameter. However, as Θ increased, the power of the true model increased; consequently, the power estimates between the two models diverged. Generally, power estimates increased in all cases as the length of the series increased.

True model: Nonstationary processes. The power rates for the Nonstationary model are fairly constant across intervention effect size and series length, with the values being consistently low. The estimated power rates for the Nonstationary model are presented in Table 28 for both series lengths. Misidentification as the other models resulted in generally higher power values in all cases than the true model. In general, power appeared to be higher for the longer series and large effect size.

True model: White Noise. The power values for the true White Noise model tend to increase as the series length and effect size increases. The estimated power rates for the White Noise model are presented in Table 29 for both series lengths. Misidentifying White Noise as a differenced model substantially reduces power, which parallels the results obtained for the true Autoregressive and Moving Average models. The reduction in power is most apparent in the misidentified Nonstationary model at a large effect size. However, misidentifying White Noise as a non-differenced Autoregressive or Moving Average model resulted in power rates similar to, yet slightly higher than, the true model.

Time series model effect on power estimates. In summary, the effects of the type of time series model on power parallel the effects found for Type I error. For those cases in which lower Type I error rates were found, lower power rates were also found, and for higher Type I error rates, higher power rates were found. A similar pattern was detected with differenced models, generally resulting in lower power than the true model. When the assumed model was a non-differenced model applied to a true model that was differenced, the misidentification resulted in higher power rates in most cases, but slightly lower power rates occurred for a large effect size in the true Integrated Moving Average model. The true Autoregressive model resulted in estimated power rates which diminished as the autoregressive parameter became larger. With the true as well as the assumed Moving Average model, power estimates were considerably higher and generally tended to increase as the correlation parameter and series length increased. Misidentification of the true Moving Average or Autoregressive models as a differenced series resulted in lower power estimates. Generally, the power of the test statistic to detect a non-zero intervention effect is less than desirable for many of the cases studied.

Table 22

Estimated Power Rates for Intervention Effect

True Model: Autoregressive

Unbalanced Design n = 20.1

Balanced Design n = 20.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=20.1	n=20.2	n=20.1	n=20.2	n=20.1	n=20.2
AR ϕ=0.2	AR	.11	.12	.17	.20	.33	.35
	MA	.25	.23	.31	.34	.46	.43
	IMA	.11	.11	.16	.22	.33	.37
	NS	.04	.06	.06	.07	.10	.09
	WN	.11	.11	.16	.23	.34	.38
AR ϕ=0.5	AR	.14	.18	.18	.27	.31	.34
	MA	.26	.23	.28	.34	.37	.40
	IMA	.21	.25	.23	.31	.34	.47
	NS	.07	.04	.09	.06	.10	.10
	WN	.20	.27	.26	.36	.37	.45
AR ϕ=0.8	AR	.22	.20	.23	.26	.30	.33
	MA	.45	.33	.41	.40	.48	.46
	IMA	.25	.25	.26	.29	.32	.36
	NS	.08	.06	.14	.09	.16	.16
	WN	.45	.42	.41	.46	.48	.53

Table 23

Estimated Power Rates for Intervention Effect

True Model: Autoregressive

Unbalanced Design n = 40.1

Balanced De gn n = 40.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=40.1	n=40.2	n=40.1	n=40.2	n=40.1	n=40.2
AR ϕ=0.2	AR	.10	.13	.24	.26	.41	.56
	MA	.13	.14	.25	.28	.43	.58
	IMA	.12	.14	.25	.31	.44	.61
	NS	.05	.03	.07	.06	.10	.11
	WN	.14	.16	.28	.35	.50	.66
AR ϕ=0.5	AR	.16	.14	.19	.25	.35	.43
	MA	.19	.19	.26	.31	.43	.53
	IMA	.20	.19	.21	.25	.31	.38
	NS	.05	.05	.10	.09	.14	.17
	WN	.28	.27	.35	.43	.52	.62
AR ϕ=0.8	AR	.16	.19	.23	.23	.34	.34
	MA	.37	.42	.43	.50	.51	.54
	IMA	.11	.10	.14	.16	.26	.26
	NS	.06	.06	.10	.09	.20	.20
	WN	.45	.51	.52	.58	.57	.63

Table 24

Estimated Power Rates for Intervention Effect

True Model: Moving Average

Unbalanced Design n = 20.1

Balanced Design n = 20.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=20.1	n=20.2	n=20.1	n=20.2	n=20.1	n=20.2
MA θ=0.2	MA	.39	.41	.51	.56	.67	.72
	AR	.09	.09	.21	.29	.40	.51
	IMA	.02	.04	.10	.13	.22	.33
	NS	.05	.05	.06	.06	.09	.07
	WN	.02	.04	.10	.13	.22	.34
MA θ=0.5	MA	.51	.47	.75	.78	.92	.96
	AR	.05	.04	.20	.30	.59	.76
	IMA	.01	.00	.02	.04	.17	.27
	NS	.04	.03	.05	.07	.10	.08
	WN	.01	.00	.03	.04	.17	.28
MA θ=0	MA	.60	.50	.90	.95	.99	.99
	AR	.02	.01	.25	.32	.68	.85
	IMA	.00	.00	.01	.01	.11	.17
	NS	.06	.04	.06	.06	.06	.06
	WN	.00	.00	.01	.01	.12	.17

Table 25

Estimated Power Rates for Intervention Effect

True Model: Moving Average

Unbalanced Design n = 40.1

Balanced Design n = 40.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=40.1	n=40.2	n=40.1	n=40.2	n=40.1	n=40.2
MA θ=0.2	MA	.24	.24	.52	.57	.79	.89
	AR	.13	.12	.40	.45	.72	.84
	IMA	[illegible]	.03	.23	.26	.51	.70
	NS	[illegible]	.06	.09	.09	.08	.08
	WN	[illegible]	.04	.25	.27	.53	.72
MA θ=0.5	MA	.49	.50	.84	.91	.99	.99
	AR	.07	.09	.52	.62	.93	.99
	IMA	.00	.00	.08	.17	.58	.84
	NS	.05	.05	.05	.09	.06	.08
	WN	.00	.00	.10	.20	.61	.87
MA θ=0.8	MA	.80	.89	.99	1.00	1.00	1.00
	AR	.01	.01	.63	.86	.99	1.00
	IMA	.00	.00	.03	.07	.72	.92
	NS	.05	.06	.05	.06	.08	.07
	WN	.00	.00	.03	.08	.65	.94

Table 26

Estimated Power Rates for Intervention Effect

True Model: Integrated Moving Average

Unbalanced Design n = 20.1

Balanced Design n = 20.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=20.1	n=20.2	n=20.1	n=20.2	n=20.1	n=20.2
IMA θ=0.2	IMA	.15	.12	.24	.25	.33	.43
	MA	.34	.10	.42	.16	.57	.34
	AR	.07	.06	.15	.11	.24	.27
	WN	.11	.10	.20	.19	.32	.40
	NS	.05	.04	.06	.05	.12	.10
IMA θ=0.5	IMA	.08	.08	.13	.20	.26	.37
	MA	.33	.16	.44	.33	.53	.42
	AR	.05	.05	.12	.14	.21	.30
	WN	.04	.06	.12	.16	.27	.37
	NS	.06	.05	.07	.07	.08	.09
IMA θ=0.8	IMA	.03	.03	.07	.10	.21	.30
	MA	.39	.25	.57	.47	.66	.66
	AR	.01	.01	.07	.09	.23	.31
	WN	.01	.00	.06	.07	.20	.31
	NS	.05	.04	.03	.06	.07	.08

Table 27

Estimated Power Rates for Intervention Effect

True Model: Integrated Moving Average

Unbalanced Design n = 40.1

Balanced Design n = 40.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=40.1	n=40.2	n=40.1	n=40.2	n=40.1	n=40.2
IMA θ=0.2	IMA	.13	.19	.29	.36	.46	.54
	MA	.05	.09	.19	.22	.44	.50
	AR	.05	.10	.19	.20	.37	.43
	WN	.12	.21	.31	.41	.50	.65
	NS	.05	.06	.06	.07	.11	.10
IMA θ=0.5	IMA	.04	.06	.19	.28	.57	.68
	MA	.06	.06	.16	.21	.39	.48
	AR	.03	.04	.16	.22	.46	.57
	WN	.05	.06	.20	.30	.58	.72
	NS	.06	.05	.06	.07	.09	.13
IMA θ=0.8	IMA	.01	.01	.09	.15	.60	.83
	MA	.22	.21	.42	.42	.59	.67
	AR	.01	.01	.12	.18	.58	.81
	WN	.01	.01	.09	.16	.62	.85
	NS	.07	.09	.07	.09	.10	.10

Table 28

Estimated Power Rates for Intervention Effect

True Model: Nonstationary

Unbalanced Design n = 20.1

Balanced Design n = 20.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=20.1	n=20.2	n=20.1	n=20.2	n=20.1	n=20.2
NS							
	NS	.06	.06	.08	.06	.10	.10
	MA	.41	.12	.45	.19	.60	.28
	AR	.10	.12	.26	.17	.26	.23
	IMA	.20	.22	.27	.28	.40	.40
	WN	.18	.21	.23	.30	.37	.40

Unbalanced Design n = 40.1

Balanced Design n = 40.2

True Model	Assumed Model	n=40.1	n=40.2	n=40.1	n=40.2	n=40.1	n=40.2
NS							
	NS	.06	.04	.08	.06	.13	.10
	MA	.11	.08	.24	.27	.49	.54
	AR	.11	.05	.15	.18	.30	.37
	IMA	.20	.18	.30	.33	.46	.55
	WN	.21	.18	.28	.39	.53	.60

Table 29

Estimated Power Rates for Intervention Effect

True Model: White Noise

Unbalanced Design n = 20.1

Balanced Design n = 20.2

Number of Replications = 500

True Model	Assumed Model	Power Small Effect		Power Medium Effect		Power Large Effect	
		n=20.1	n=20.2	n=20.1	n=20.2	n=20.1	n=20.2
WN							
	WN	.15	.09	.16	.17	.29	.37
	MA	.31	.34	.40	.42	.57	.60
	AR	.10	.13	.21	.24	.38	.42
	IMA	.06	.10	.17	.16	.28	.36
	NS	.05	.05	.06	.05	.09	.07

Unbalanced Design n = 40.1

Balanced Design n = 40.2

True Model	Assumed Model	n=40.1	n=40.2	n=40.1	n=40.2	n=40.1	n=40.2
WN							
	WN	.10	.09	.26	.31	.54	.68
	MA	.16	.16	.36	.36	.61	.71
	AR	.13	.11	.30	.33	.58	.68
	IMA	.09	.08	.25	.29	.52	.65
	NS	.07	.06	.08	.08	.08	.09

Accuracy of Intervention Effect Estimates

The final results presented in this section concern the accuracy of the estimated standard error of the intervention parameter. The results presented in Tables 30 to 45 are a comparison of the mean estimated standard error of the intervention parameter over 500 replications and the empirical measure of the standard error obtained by calculating the standard deviation of the 500 estimates of the intervention parameter, for both true and assumed time series models. The standard deviation of 500 parameter estimates provides an empirical measure of the standard error in the absence of a theoretically derived expression for an exact estimate of the standard error. Comparison of the estimated and empirical standard error serves to reveal bias in the parameter estimation procedures which in turn assists in the explanation of anomalies in the Type I error and power analysis. For example, in Type I error analysis, the estimated standard error is used; if it is too large, the power may be low. As would be expected, the intervention component does not influence the magnitude of either the estimated or empirical standard error, and thus the values for conditions with the intervention parameter equal to zero will be considered.

True model: Autoregressive processes. For the true Autoregressive model, it can be seen in Tables 30 to 33 that the mean estimated standard error is substantially larger than the empirically obtained standard error, especially when ϕ equals 0.2 in the longer series. The empirical standard error increased as the autoregressive parameter became larger, and as the length of the realization becomes shorter, while the estimated standard error increased as the autoregressive parameter became larger, and as the length of the realization becomes longer. Misidentifying the Autoregressive model as White Noise or Moving Average resulted in increasing Type I error as the autoregressive parameter increased. The discrepancy between the estimated and empirically obtained standard error reflects this increasing Type I error rate. On the other hand, misidentifying an Autoregressive model as a model with differencing resulted in decreased power. For the assumed Nonstationary model, the error estimates of intervention effects were always greater than that for the true Autoregressive model.

True model: Moving Average processes. For the true Moving Average model, the mean estimated standard error is consistently smaller than the empirically obtained standard error. This discrepancy becomes greater as the series length increases and as the moving average parameter increases. These results are presented in Tables 34 to 37. The underestimation of the standard error of the intervention component is the most reasonable explanation for the constant inflation of Type I error as reported. The size of the error in estimation of the intervention effect decreases slightly with increasing series length and value of the moving average parameter. Recall that no Moving Average model misidentification had serious consequences for Type I error, but did lead to reduced power in most cases. The exception was in the case in which the Autoregressive model often had about the same power as the true Moving Average model in the forty observation series. The overestimation of the estimated standard error above the empirically estimated standard error can be seen for the assumed Autoregressive model.

True model: Integrated Moving Average processes. For the true Integrated Moving Average model, there is a small discrepancy between the mean estimated standard error and the empirically estimated standard error. The empirical standard error decreased slightly with longer series length and increased value of the correlation parameter. The estimated standard error increased slightly with an increased correlation parameter value and decreased slightly with the longer length series. The results are presented in Tables 38 to 41. The difficulty in misidentification of the Integrated Moving Average model was an inflated Type I error rate when the assumed model was Moving Average, Autoregressive, or White Noise. These assumed models had similar empirical standard errors as the true Integrated Moving Average model; however, the assumed Moving Average model had a lower estimated standard error and the Autoregressive model had a consistently higher estimated standard error. Misidentifying Integrated Moving Average series as a Nonstationary model did not greatly affect Type I error, but did lead to reduced power. The error of intervention effect estimation was consistently higher for the assumed Nonstationary model than the true model.

True model: Nonstationary processes. For the true Nonstationary series, there is little discrepancy between the estimated standard error and empirical standard error and demonstrated little variation between lengths of the series. The results are presented in Tables 42 and 43. Substantial Type I error occurred when the true Nonstationary model assumed a Moving Average or White Noise model. For these assumed models, as well as the assumed Integrated Moving Average model, the empirical standard error is higher than the estimated standard error, but still lower than the true model in all cases. The assumed Autoregressive model had an estimated standard error slightly higher than the empirical standard error of the intervention effect.

True model: White Noise. For the true White Noise model there is little difference between the estimated standard error and empirical standard error with only a slight decrease occurring with the increased length of the series, as reflected by the small discrepancy reported between the nominal and empirical Type I error rates. The results are presented in Tables 44 and 45. Misidentifying White Noise as a Nonstationary model resulted in a substantial loss of power. The assumed Nonstationary model had a consistently higher level of error in both the mean estimated standard error and empirically estimated standard error of the intervention effect. Misidentifying White Noise as Moving Average or Autoregressive models led to slightly greater Type I errors. Again the pattern of discrepancy between the estimated and empirical standard errors for these models was seen as before. For the assumed Moving Average model, the empirical standard error is lower than the estimated standard error, while for the assumed Autoregressive model the estimated standard error is higher than the empirical standard error. The error estimates of the intervention effect for the assumed Integrated Moving Average model were about the same as for the true model.

Time series model effect on accuracy of standard error estimates. In summary, for the correctly identified Moving Average model the underestimation of the standard error of the intervention component explains the persistent inflation of Type I error as reported. Bias in the least squares estimation procedure was profoundly attenuated in Moving Average time series processes. The least squares procedure for estimating the approximate standard error overestimates the actual magnitude of the autocorrelation function. This finding is discussed in the next section. Whereas, for the Autoregressive model, the standard error was consistently overestimated. The size of the error in estimation of intervention effects generally decreases with increasing series length. For the true differenced models and White Noise, there is little difference between the mean estimated standard error and empirically estimated standard error of intervention effects. For the correctly identified models, there was little difference in the size of the standard error of estimation between the series of twenty observations and those of forty observations, except for those differences in the Moving Average model at large values of the correlation parameter.

Table 30

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Autoregressive

Unbalanced Design n = 20.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
AR ϕ=0.2	AR	1.270	0.617	1.239	0.626	1.253	0.649	1.245	0.682
	MA	0.161	0.638	0.158	0.639	0.160	0.663	0.158	0.711
	IMA	0.515	0.735	0.504	0.733	0.509	0.749	0.505	0.776
	NS	1.284	1.249	1.252	1.291	1.267	1.259	1.258	1.324
	WN	0.512	0.614	0.502	0.617	0.507	0.621	0.503	0.683
AR ϕ=0.5	AR	1.125	0.834	1.137	0.818	1.116	0.894	1.122	0.856
	MA	0.208	0.836	0.201	0.865	0.201	0.894	0.205	0.590
	IMA	0.538	1.013	0.537	1.012	0.531	1.106	0.536	1.045
	NS	1.136	1.125	1.149	1.168	1.128	1.219	1.134	1.148
	WN	0.536	0.833	0.535	0.818	0.529	0.878	0.534	0.875
AR ϕ=0.8	AR	1.013	1.177	1.028	1.194	1.033	1.136	1.040	1.152
	MA	0.276	1.506	0.288	1.449	0.287	1.361	0.289	1.391
	IMA	0.604	1.219	0.619	1.268	0.622	1.266	0.620	1.241
	NS	1.021	1.046	1.036	1.092	1.042	1.061	1.050	1.075
	WN	0.602	1.629	0.617	1.556	0.563	0.620	0.618	1.496

Table 31

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Autoregressive

Balanced Design n = 20.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
AR ϕ=0.2	AR	1.267	0.545	1.239	0.567	1.244	0.562	1.266	0.558
	MA	0.136	0.564	0.133	0.545	0.138	0.562	0.137	0.558
	IMA	0.444	0.610	0.435	0.619	0.440	0.665	0.445	0.698
	NS	1.280	1.216	1.252	1.352	1.257	1.324	1.278	1.258
	WN	0.441	0.547	0.432	0.528	0.438	0.537	0.442	0.547
AR ϕ=0.5	AR	1.132	0.817	1.113	0.794	1.135	0.847	1.124	0.798
	MA	0.169	0.815	0.166	0.762	0.170	0.827	0.169	0.776
	IMA	0.463	1.113	0.457	1.025	0.463	1.011	0.459	0.965
	NS	1.143	1.173	1.123	1.157	1.145	1.126	1.135	1.140
	WN	0.460	0.824	0.454	0.769	0.461	0.838	0.456	0.779
AR ϕ=0.8	AR	1.035	1.177	1.043	1.115	1.027	1.187	1.031	1.194
	MA	0.226	1.354	0.233	1.252	0.218	1.379	0.229	1.418
	IMA	0.515	1.244	0.542	1.240	0.508	1.306	0.520	1.297
	NS	1.042	1.063	1.052	1.061	1.035	1.072	1.039	1.060
	WN	0.512	1.471	0.521	1.359	0.505	1.502	0.517	1.536

Table 32

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: Autoregressive

Unbalanced Design n = 40.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
AR ϕ=0.2	AR	1.262	0.436	1.258	0.455	1.262	0.448	1.277	0.449
	MA	0.085	0.434	0.087	0.453	0.084	0.447	0.085	0.449
	IMA	0.372	0.527	0.372	0.530	0.371	0.526	0.376	0.532
	NS	1.280	1.368	1.276	1.295	1.281	1.275	1.296	1.275
	WN	0.366	0.436	0.366	0.452	0.364	0.443	0.369	0.451
AR ϕ=0.5	AR	1.126	0.696	1.134	0.678	1.121	0.675	1.129	0.700
	MA	0.117	0.690	0.117	0.683	0.120	0.673	0.122	0.681
	IMA	0.402	0.996	0.402	0.931	0.401	0.934	0.406	0.944
	NS	1.142	1.200	1.150	1.186	1.137	1.180	1.145	1.227
	WN	0.395	0.706	0.396	0.701	0.394	0.697	0.399	0.710
AR ϕ=0.8	AR	1.035	1.025	1.025	1.043	1.029	1.030	1.033	1.068
	MA	0.211	1.245	0.202	1.255	0.208	1.244	0.217	1.216
	IMA	0.512	1.115	0.502	1.113	0.504	1.096	0.514	1.136
	NS	1.048	1.051	1.038	1.073	1.042	1.061	1.047	1.128
	WN	0.504	1.351	0.494	1.366	0.497	1.368	0.507	1.318

Table 33

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Autoregressive

Balanced Design n = 40.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
AR ϕ=0.2	AR	1.257	0.398	1.275	0.394	1.259	0.388	1.259	0.412
	MA	0.071	0.397	0.072	0.396	0.071	0.388	0.072	0.412
	IMA	0.333	0.476	0.326	0.458	0.323	0.473	0.324	0.491
	NS	1.272	1.241	1.291	1.274	1.275	1.311	1.275	1.334
	WN	0.313	0.398	0.318	0.393	0.315	0.392	0.316	0.420
AR ϕ=0.5	AR	1.137	0.625	1.129	0.602	1.124	0.609	1.132	0.672
	MA	0.100	0.618	0.100	0.594	0.102	0.609	0.102	0.633
	IMA	0.352	0.922	0.349	0.873	0.352	0.938	0.353	0.928
	NS	1.151	1.217	1.143	1.128	1.138	1.199	1.146	1.163
	WN	0.343	0.632	0.341	0.609	0.343	0.618	0.344	0.644
AR ϕ=0.8	AR	1.045	1.020	1.030	1.011	1.028	1.018	1.031	1.017
	MA	0.174	1.231	0.172	1.234	0.167	1.220	0.177	1.311
	IMA	0.444	1.118	0.434	1.065	0.431	1.109	0.441	1.102
	NS	1.057	1.071	1.042	1.030	1.040	1.034	1.042	1.069
	WN	0.434	1.334	0.424	1.320	0.421	1.306	0.432	1.411

Table 34

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: Moving Average

Unbalanced Design n = 20.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
MA θ=0.2	MA	0.129	0.449	0.130	0.475	0.131	0.471	0.130	0.484
	AR	1.525	0.418	1.555	0.45[illegible]	1.516	0.438	1.533	0.458
	IMA	0.531	0.438	0.540	0.454	0.532	0.460	0.534	0.461
	NS	1.542	1.517	1.572	1.592	1.533	1.576	1.550	1.523
	WN	0.528	0.421	0.537	0.453	0.529	0.436	0.531	0.456
MA θ=0.5	MA	0.120	0.331	0.119	0.338	0.118	0.367	0.120	0.386
	AR	1.805	0.318	1.812	0.317	1.818	0.352	1.797	0.360
	IMA	0.591	0.352	0.591	0.345	0.593	0.365	0.593	0.385
	NS	1.825	1.857	1.833	1.790	1.839	1.778	1.817	1.991
	WN	0.588	0.351	0.588	0.345	0.590	0.364	0.590	0.386
MA θ=0.8	MA	0.125	0.258	0.126	0.276	0.124	0.303	0.124	0.332
	AR	2.109	0.257	2.155	0.278	2.138	0.305	2.135	0.327
	IMA	0.679	0.327	0.690	0.334	0.683	0.358	0.683	0.380
	NS	2.134	2.242	2.180	2.253	2.163	2.175	2.159	0.239
	WN	0.676	0.325	0.687	0.333	0.679	0.357	0.680	0.379

Table 35

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: Moving Average

Balanced Design n = 20.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
MA θ=0.2	MA	0.109	0.377	0.111	0.379	0.111	0.411	0.110	0.423
	AR	1.497	0.371	1.537	0.371	1.506	0.403	1.506	0.402
	IMA	0.453	0.379	0.465	0.372	0.458	0.416	0.457	0.429
	NS	1.512	1.603	1.553	1.664	1.522	1.530	1.522	1.509
	WN	0.450	0.374	0.462	0.368	0.455	0.404	0.454	0.396
MA θ=0.5	MA	0.105	0.272	0.104	0.268	0.104	0.290	0.105	0.321
	AR	1.807	0.278	1.825	0.261	1.811	0.289	1.828	0.313
	IMA	0.515	0.318	0.517	0.284	0.515	0.345	0.520	0.331
	NS	1.826	1.933	1.844	1.794	1.829	1.990	1.848	1.947
	WN	0.512	0.304	0.514	0.284	0.511	0.313	0.516	0.330
MA θ=0.8	MA	0.109	0.188	0.109	0.196	0.109	0.221	0.111	0.260
	AR	2.136	0.187	2.167	0.193	2.114	0.227	2.154	0.261
	IMA	0.593	0.243	0.598	0.245	0.588	0.269	0.600	0.296
	NS	2.159	2.224	2.190	2.223	2.136	2.211	2.176	2.256
	WN	0.589	0.242	0.594	0.244	0.585	0.268	0.596	0.296

Table 36

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Moving Average

Unbalanced Design n = 40.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
MA θ=0.2	MA	0.062	0.297	0.062	0.308	0.061	0.306	0.062	0.306
	AR	1.548	0.294	1.533	0.307	1.529	0.304	1.549	0.303
	IMA	0.382	0.299	0.379	0.313	0.379	0.307	0.384	0.307
	NS	1.571	1.574	1.555	1.516	1.551	1.565	1.571	1.618
	WN	0.375	0.298	0.372	0.311	0.372	0.306	0.377	0.304
MA θ=0.5	MA	0.052	0.207	0.051	0.206	0.052	0.215	0.051	0.246
	AR	1.843	0.206	1.817	0.209	1.821	0.213	1.834	0.244
	IMA	0.425	0.214	0.421	0.225	0.420	0.225	0.423	0.256
	NS	1.870	1.793	1.844	1.923	1.848	1.834	1.861	1.865
	WN	0.416	0.213	0.412	0.224	0.412	0.223	0.415	0.255
MA θ=0.8	MA	0.047	0.138	0.048	0.141	0.048	0.149	0.048	0.195
	AR	2.123	0.140	2.161	0.144	2.127	0.157	2.111	0.192
	IMA	0.477	0.170	0.486	0.179	0.479	0.189	0.477	0.212
	NS	2.155	2.159	2.193	2.211	2.159	2.119	2.141	0.206
	WN	0.467	0.168	0.476	0.176	0.470	0.186	0.467	0.211

Table 37

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: Moving Average

Balanced Design n = 40.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
MA θ=0.2	MA	0.053	0.260	0.052	0.254	0.052	0.267	0.053	0.274
	AR	1.544	0.261	1.531	0.251	1.526	0.265	1.530	0.270
	IMA	0.333	0.265	0.330	0.254	0.331	0.274	0.332	0.283
	NS	1.564	1.614	1.551	1.619	1.546	1.640	1.550	1.619
	WN	0.324	0.264	0.322	0.254	0.322	0.265	0.323	0.270
MA θ=0.5	MA	0.044	0.160	0.045	0.175	0.044	0.185	0.045	0.200
	AR	1.830	0.159	1.853	0.175	1.816	0.183	1.834	0.199
	IMA	0.367	0.169	0.373	0.184	0.366	0.193	0.368	0.209
	NS	1.854	1.834	1.877	1.858	1.839	1.965	1.858	1.854
	WN	0.358	0.168	0.363	0.183	0.357	0.192	0.359	0.208
MA θ=0.8	MA	0.042	0.093	0.042	0.094	0.042	0.126	0.042	0.169
	AR	2.141	0.099	2.122	0.106	2.177	0.136	2.138	0.171
	IMA	0.420	0.128	0.417	0.130	0.426	0.161	0.419	0.188
	NS	2.168	2.154	2.149	2.204	2.205	2.219	2.165	2.127
	WN	0.409	0.127	0.407	0.128	0.415	0.159	0.409	0.186

Table 38

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Integrated Moving Average

Unbalanced Design n = 20.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
IMA θ=0.2	IMA	0.637	0.972	0.648	0.992	0.643	0.969	0.633	0.987
	MA	0.204	0.884	0.217	0.842	0.212	0.866	0.205	0.885
	AR	1.405	0.867	1.409	0.819	1.400	0.798	1.392	0.839
	WN	0.634	0.832	0.645	0.780	0.640	0.799	0.630	0.882
	NS	1.419	1.539	1.424	1.491	1.415	1.447	1.407	1.512
IMA θ=0.5	IMA	0.632	0.788	0.630	0.722	0.626	0.679	0.636	0.726
	MA	0.180	0.744	0.176	0.741	0.176	0.777	0.176	0.767
	AR	1.529	0.642	1.537	0.598	1.514	0.636	1.531	0.633
	WN	0.630	0.610	0.627	0.566	0.624	0.618	0.632	0.603
	NS	1.546	1.550	1.553	1.570	1.530	1.647	1.548	1.555
IMA θ=0.8	IMA	0.698	0.545	0.680	0.583	0.685	0.593	0.677	0.556
	MA	0.168	0.635	0.165	0.677	0.165	0.671	0.165	0.645
	AR	1.801	0.482	1.737	0.471	1.762	0.531	1.740	0.510
	WN	0.695	0.477	0.676	0.459	0.681	0.504	0.674	0.500
	NS	1.821	1.824	1.756	1.818	1.782	1.691	1.760	1.841

Table 39

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: Integrated Moving Average

Balanced Design n = 20.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
IMA θ=0.2	IMA	0.552	0.862	0.557	0.820	0.560	0.938	0.554	0.982
	MA	0.179	0.649	0.181	0.629	0.180	0.659	0.179	0.692
	AR	1.397	0.715	1.407	0.696	1.406	0.741	1.396	0.788
	WN	0.548	0.686	0.553	0.652	0.556	0.700	0.550	0.717
	NS	1.411	1.395	1.420	1.375	1.420	1.429	1.410	1.416
IMA θ=0.5	IMA	0.548	0.717	0.543	0.627	0.561	0.699	0.555	0.695
	MA	0.151	0.579	0.154	0.555	0.154	0.571	0.156	0.591
	AR	1.532	0.557	1.511	0.513	1.567	0.546	1.545	0.555
	WN	0.544	0.506	0.539	0.504	0.557	0.506	0.552	0.520
	NS	1.547	1.683	1.526	1.592	1.582	1.587	1.560	1.614
IMA θ=0.8	IMA	0.590	0.425	0.596	0.505	0.598	0.578	0.592	0.530
	MA	0.145	0.446	0.146	0.432	0.146	0.467	0.145	0.525
	AR	1.734	0.384	1.750	0.380	1.758	0.397	1.744	0.471
	WN	0.586	0.351	0.592	0.345	0.594	0.378	0.588	0.432
	NS	1.752	1.792	1.768	1.737	1.775	1.838	1.762	1.779

Table 40

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Integrated Moving Average

Unbalanced Design n = 40.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
IMA θ=0.2	IMA	0.474	0.795	0.474	0.767	0.471	0.829	0.474	0.798
	MA	0.114	0.546	0.113	0.533	0.111	0.580	0.111	0.568
	AR	1.411	0.569	1.415	0.565	1.406	0.595	1.417	0.583
	WN	0.465	0.554	0.465	0.541	0.463	0.574	0.466	0.576
	NS	1.431	1.401	1.435	1.396	1.426	1.452	1.437	1.447
IMA θ=0.5	IMA	0.450	0.401	0.453	0.406	0.452	0.444	0.455	0.452
	MA	0.082	0.479	0.083	0.482	0.084	0.473	0.084	0.503
	AR	1.532	0.405	1.538	0.392	1.528	0.424	1.545	0.430
	WN	0.442	0.396	0.444	0.384	0.443	0.414	0.446	0.418
	NS	1.554	1.564	1.560	1.566	1.550	1.559	1.567	1.565
IMA θ=0.8	IMA	0.493	0.262	0.487	0.262	0.492	0.280	0.485	0.294
	MA	0.069	0.365	0.068	0.334	0.068	0.381	0.067	0.381
	AR	1.782	0.270	1.757	0.263	1.785	0.283	1.754	0.296
	WN	0.484	0.259	0.477	0.257	0.483	0.278	0.476	0.291
	NS	1.808	1.940	1.783	1.830	1.810	1.789	1.780	1.777

Table 41

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Integrated Moving Average

Balanced Design n = 40.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
IMA θ=0.2	IMA	0.409	0.709	0.408	0.727	0.412	0.734	0.409	0.812
	MA	0.094	0.473	0.096	0.487	0.094	0.499	0.095	0.506
	AR	1.405	0.497	1.403	0.534	1.412	0.530	1.405	0.542
	WN	0.399	0.493	0.398	0.537	0.402	0.523	0.399	0.536
	NS	1.422	1.388	1.420	1.414	1.430	1.455	1.423	1.480
IMA θ=0.5	IMA	0.398	0.374	0.401	0.422	0.398	0.384	0.398	0.412
	MA	0.070	0.400	0.072	0.393	0.071	0.400	0.073	0.392
	AR	1.550	0.346	1.561	0.351	1.550	0.359	1.541	0.373
	WN	0.388	0.342	0.391	0.342	0.387	0.349	0.388	0.366
	NS	1.569	1.706	1.580	1.628	1.570	1.608	1.560	1.642
IMA θ=0.8	IMA	0.425	0.217	0.426	0.208	0.425	0.214	0.430	0.243
	MA	0.059	0.293	0.060	0.280	0.059	0.302	0.059	0.325
	AR	1.769	0.203	1.763	0.211	1.767	0.217	1.792	0.245
	WN	0.415	0.200	0.415	0.205	0.414	0.212	0.420	0.240
	NS	1.791	1.786	1.786	1.891	1.789	1.780	1.815	1.844

Table 42

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: Nonstationary

Unbalanced Design n = 20.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
NS									
	NS	1.407	1.435	1.381	1.445	1.391	1.465	1.424	1.328
	MA	0.255	0.919	0.243	0.933	0.246	0.951	0.244	0.945
	AR	1.392	0.966	1.367	0.970	1.377	1.062	1.409	0.964
	IMA	0.701	1.103	0.681	1.063	0.686	1.146	0.693	1.025
	WN	0.698	0 941	0.678	0.966	0.683	1.029	0.689	1.017

True Model: Nonstatic . y

Balanced Design n=20.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
NS									
	NS	1.381	1.413	1.388	1.436	1.367	1.344	1.401	1.393
	MA	0.213	0.733	0.212	0.759	0.212	0.756	0.212	0.731
	AR	1.367	0.893	1.376	0.954	1.354	0.891	1.388	0.920
	IMA	0.595	1.067	0.597	1.052	0.585	0.999	0.596	0.992
	WN	0.591	0.880	0.593	0.905	0.581	0.901	0.592	0.846

Table 43

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: Nonstationary

Unbalanced Design n = 40.1

True Model	Assumed Model	$I_t=0.0$ EST	$I_t=0.0$ EMP	$I_t=0.2$ EST	$I_t=0.2$ EMP	$I_t=0.5$ EST	$I_t=0.5$ EMP	$I_t=0.8$ EST	$I_t=0.8$ EMP
NS									
	NS	1.412	1.389	1.383	1.428	1.402	1.425	1.400	1.507
	MA	0.133	0.589	0.134	0.666	0.138	0.613	0.134	0.645
	AR	1.392	0.689	1.364	0.766	1.383	0.708	1.380	0.805
	IMA	0.510	0.935	0.504	1.084	1.203	0.970	0.509	1.049
	WN	0.501	0.680	0.495	0.755	0.505	0.691	0.500	0.729

True Model: Nonstationary

Balanced Design n = 40.2

True Model	Assumed Model	$I_t=0.0$ EST	$I_t=0.0$ EMP	$I_t=0.2$ EST	$I_t=0.2$ EMP	$I_t=0.5$ EST	$I_t=0.5$ EMP	$I_t=0.8$ EST	$I_t=0.8$ EMP
NS									
	NS	1.401	1.469	1.402	1.382	1.402	1.425	1.420	1.389
	MA	0.115	0.543	0.114	0.539	0.112	0.518	0.112	0.528
	AR	1.384	0.671	1.385	0.662	1.385	0.612	1.402	0.633
	IMA	0.447	0.973	0.443	0.928	0.442	0.913	0.446	0.904
	WN	0.436	0.632	0.432	0.627	0.431	0.605	0.435	0.627

Table 44

Mean Estimated Standard Error (EST) and Empirical Standard Error (EMP) of Intervention Effect Estimates

True Model: White Noise

Unbalanced Design n = 20.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
WN									
	WN	0.507	0.506	0.515	0.519	0.508	0.542	0.516	0.511
	MA	0.142	0.541	0.143	0.557	0.142	0.576	0.141	0.550
	AR	1.369	0.516	1.392	0.527	1.363	0.573	1.391	0.522
	IMA	0.510	0.591	0.518	0.584	0.510	0.605	0.518	0.544
	NS	1.385	1.410	1.407	1.439	1.378	1.391	1.407	1.439

True Model: White Noise

Balanced Design n = 20.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
WN									
	WN	0.438	0.474	0.441	0.472	0.438	0.456	0.445	0.484
	MA	0.121	0.492	0.121	0.485	0.119	0.476	0.121	0.500
	AR	1.357	0.481	1.373	0.476	1.374	0.464	1.385	0.496
	IMA	0.441	0.541	0.443	0.570	0.441	0.492	0.448	0.538
	NS	1.370	1.371	1.388	1.402	1.388	1.368	1.400	1.299

Table 45

Mean Estimated Standard Error (EST) and Empirical

Standard Error (EMP) of Intervention

Effect Estimates

True Model: White Noise

Unbalanced Design n = 40.1

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
WN									
	WN	0.363	0.364	0.364	0.374	0.363	0.376	0.358	0.367
	MA	0.072	0.370	0.072	0.377	0.072	0.377	0.071	0.369
	AR	1.376	0.368	1.378	0.378	1.386	0.377	1.365	0.368
	IMA	0.370	0.383	0.371	0.383	0.370	0.401	0.365	0.375
	NS	1.397	1.413	1.398	1.471	1.407	1.496	1.385	1.414

True Model: White Noise

Balanced Design n = 40.2

True Model	Assumed Model	I_t=0.0 EST	I_t=0.0 EMP	I_t=0.2 EST	I_t=0.2 EMP	I_t=0.5 EST	I_t=0.5 EMP	I_t=0.8 EST	I_t=0.8 EMP
WN									
	WN	0.313	0.313	0.314	0.318	0.312	0.312	0.313	0.325
	MA	0.062	0.313	0.061	0.319	0.061	0.314	0.062	0.325
	AR	1.380	0.313	1.383	0.319	1.362	0.315	1.363	0.324
	IMA	0.320	0.329	0.322	0.337	0.320	0.322	0.321	0.357
	NS	1.397	1.419	1.401	1.424	1.380	1.427	1.380	1.438

Accuracy of ARIMA Parameter Estimates

The estimates of ARIMA parameters obtained in short series of twen... forty data points were biased. The bias is, in part, due to the ... nematical intractability of calculating the exact likelihood function. In longer series, this difficulty does not affect the estimates as much as it does in short series. Bias is also introduced by the function used for making the estimates of the correlation parameters (Box & Jenkins, 1976; Osborn, 1982). The least squares estimates of the parameters in the ARIMA intervention model are only approximations to the full maximum likelihood estimates. As a standard procedure, the non-linear optimization algorithm suggested by Box and Jenkins (1976) has been used. However, non-linear optimization algorithms cannot guarantee that a global minimum is achieved; it is assumed that the initial estimates used as input are good enough to bring the optimization sufficiently close to global minimum.

Generally, the method used in this study (least squares normal theory analysis of Box and Jenkins [1970]), gave an overestimate of Θ for the Moving Average model. In the Moving Average series for Θ of 0.2, 0.5, and 0.8 with twenty observations, the mean estimated value of Θ was 0.605, 0.843, and 0.936, respectively. The modal estimated value of Θ was consistently 0.98 for all true parameter values of 0.2, 0.5, and 0.8. For the Moving Average series with forty data points, the parameter values were, agai ., consistently overestimated at all true levels of Θ, with a modal estimated value of 0.98 for all moving average parameter values. On the other hand, the autoregressive parameter was underestimated for all values of ϕ in short series. When the number of observations was increased to forty, the estimated values of the autoregressive parameter were closer to the true values, but still remained smaller. The estimated ARIMA parameter values are presented in Tables 46 and 47 for the Moving Average and Autoregressive models. Each least squares analysis produces a residual error variance by which the maximum likelihood estimates of the parameters may be found. Therefore, the maximum likelihood estimated value of the parameter corresponds to the minimum residual error variance. The standard deviation of the 500 parameter estimates is presented for comparison. Estimates of the error variance were similar for both Moving Average and Autoregressive models.

ARIMA parameter estimation effect on accuracy of intervention effect analysis. In summary, using the least squares procedure for estimation, the estimates of the correlation parameters were often biased, even when the correct model was identified. The direction and amount of bias depended on the actual values of the correlation parameter, the method for calculating the initial values in the estimation procedure, and the actual estimation function used. When considering only those Moving Average series with an estimated correlation parameter value of 0.98, the number of significant intervention effects when ω was zero was very high. The Type I error rate for intervention effect estimates was therefore inflated above the nominal level for a large proportion of Moving Average series with an estimated correlation parameter of 0.98.

To assess the bias in the parameter estimation procedure, data sets were examined where the moving average parameter was estimated at 0.98 and where a significant change in level with a zero intervention effect was found. Ten of these Moving Average data sets ($\theta = 0.50$, $\omega = 0$) were chosen from series of both twenty and forty time points. These generated series were resubmitted to TIME02 least squares analysis. The maximum likelihood estimated value of Θ at the true value of 0.50 was compared to the corresponding minimum error variance and least squares estimate of the intervention effect. The intervention parameter was examined for a significant change in level. In the random sample of ten series with the moving average parameter estimated at 0.98 and a significant change in level with a zero intervention effect, a general reduction in Type I errors for intervention effect estimates was found at the true value of Θ for the Moving Average series. This reduction in Type I errors of intervention effect estimates at the true parameter value, was further decreased as the series length increased. For example, out of ten randomly selected series with the moving average parameter of 0.98, only two series had a significant change in level at the true parameter value of 0.50, for the longer series length of forty data points. The number of significant intervention parameters found at the estimated ARIMA parameter values for the true Moving Average series with a Θ of 0.50 are presented in Table 48. The bias in the estimation procedure could therefore also account for the large inflation of Type I error rates found in intervention effect estimates, particularly in the Moving Average time series processes.

Table 46

ARIMA Parameter Estimates

True Model: Moving Average

True Parameter Value	Estimated Parameter Value	Standard Deviation	Error Variance	Mode	Median
		Series Length = 20.1			
Θ = .2	0.619	0.436	0.859	0.980	0.980
Θ = .5	0.874	0.256	0.826	0.980	0.980
Θ = .8	0.957	0.127	0.919	0.980	0.980
		Series Length = 20.2			
Θ = .2	0.605	0.450	0.810	0.980	0.980
Θ = .5	0.843	0.281	0.841	0.980	0.980
Θ = .8	0.936	0.166	0.944	0.980	0.980
		Series Length = 40.1			
Θ = .2	0.337	0.289	0.965	0.980	0.280
Θ = .5	0.699	0.251	0.949	0.980	0.660
Θ = .8	0.941	0.106	0.923	0.980	0.980
		Series Length = 40.2			
Θ = .2	0.338	0.268	0.961	0.980	0.280
Θ = .5	0.695	0.246	0.929	0.980	0.640
Θ = .8	0.911	0.126	0.966	0.980	0.980

Table 47

ARIMA Parameter Estimates

True Model: Autoregressive

True Parameter Value	Estimated Parameter Value	Standard Deviation	Error Variance	Mode	Median
		Series Length = 20.1			
ϕ = .2	0.066	0.229	0.961	0.120	0.080
ϕ = .5	0.322	0.266	0.936	0.460	0.340
ϕ = .8	0.588	0.283	0.903	0.980	0.620
		Series Length = 20.2			
ϕ = .2	0.055	0.233	0.949	0.120	0.080
ϕ = .5	0.299	0.264	0.945	0.300	0.300
ϕ = .8	0.556	0.299	0.928	0.980	0.560
		Series Length = 40.1			
ϕ = .2	0.126	0.167	0.971	0.180	0.140
ϕ = .5	0.404	0.158	0.965	0.320	0.400
ϕ = .8	0.701	0.166	0.977	0.760	0.720
		Series Length = 40.2			
ϕ = .2	0.123	0.156	0.954	0.060	0.120
ϕ = .5	0.395	0.159	0.977	0.380	0.400
ϕ = .8	0.710	0.163	1.004	0.700	0.700

Table 48

Estimated Intervention Effect and ARIMA Parameter

True Model: Moving Average

True Value Θ = 0.50

Series Length = 20.2

		Intervention Effect Estimates Nonsignificant	Significant	Total N
ARIMA Parameter Estimates	Θ = 0.98	216	172*	388
	Θ = all other values	102	10	112
Total	N	318	182	500

* Five of ten random series had a significant change in level at Θ = 0.50

Series Length = 40.2

		Intervention Effect Estimates Nonsignificant	Significant	Total N
ARIMA Parameter Estimates	Θ = 0.98	65	113*	178
	Θ = all other values	278	44	322
Total	N	343	157	500

* Two of ten random series had a significant change in level at Θ = 0.50

Time Series Model Estimation

Monte Carlo methods were used to examine the sampling properties of the autocorrelation functions under a variety of conditions. Five simple ARIMA models were investigated: a first-order Autoregressive model, a first-order Moving Average model, an Integrated Moving Average model, the Nonstationary model, and White Noise. Two characteristics of the time series process were systematically manipulated: (1) the magnitude of the parameter measuring serial correlation in the data (ϕ or θ = 0.2, 0.5, or 0.8), and (2) the length of the time series (n = 20 or 40). The primary interest was in the empirical estimation of the standard errors of the autocorrelation and partial autocorrelation coefficients. The extent of bias in these estimates can be determined theoretically provided that the true parameters of the ARIMA (p, d, q) process are known. Information concerning both the standard error and the bias of the estimates is essential for determining the length of time series realizations that is necessary to ensure a reasonably high likelihood of an appropriately identified ARIMA (p, d, q) model.

Estimates of Correlation Parameters

The mean estimates over 500 replications of the autocorrelation and partial autocorrelation functions for lags 1 through 4 were obtained for each condition. Regions of nominal acceptance or rejection were constructed around the mean value of the autocorrelation coefficient. The proportion of replications for which the estimate was outside of the region of acceptance, that is, the proportion that resulted in the rejection of the null hypothesis, H_0: $\rho_k = 0$ (where ρ represents the correlation rameter of the time series process), were calculated for each condition. This measure provides an estimate of the empirical Type I error rates for conditions in which the null hypothesis is true and of the power of the test statistic for conditions with a non-zero autocorrelation parameter. Estimation of the time series model has implications for the use of the model in the intervention assessment context. The power to detect a significant correlation parameter plays a role in selecting a particular model, and consequently in making misidentifications as well. With respect to the time series models, three issues were investigated: (1) the accuracy of the correlation coefficient estimatio , (2) Type I error rates for conditions in which the true correlation coefficient is equal to zero, and (3) the power to detect a non-zero correlation coefficient. The results presented in Tables 49 to 65 discuss the correlation coefficient estimation for the time series models. The purpose of methods presented in this section is to account for the dependence in serial observations and correct for it so that intervention effects can be estimated and tested with techniques which assume independent observations.

Autoregressive processes. The mean estimates of the autocorrelation function for lags 1 through 4 of the Autoregressive series and the standard deviation of these estimates are presented in Table 49. The true parameter value of each autocorrelation coefficient is also presented for comparison. The mean estimate of the autocorrelation function reflects a considerable bias in the estimator when applied to small sample sizes. The bias is a function of both the true autocorrelation parameter and the number of observations in the series. The bias is always downward, which results in an underestimation of the autocorrelation. The degree of underestimation increases as the autocorrelation parameter becomes larger and as the number of observations become smaller. Furthermore, the magnitude of the bias increases as the lag of the autocorrelation coefficient increases. The pattern of downward bias of the autocorrelation function can be seen at lags 1 through 4. As the number of data points in the realization increases, the mean of the estimated autocorrelations approaches the true parameter values.

The standard deviation of the parameter estimates provides a measure of the empirical standard error. The standard error at lag 1 is a function of the parameter and sample size; the standard error becomes smaller as the autocorrelation parameter and series length become larger. The relative magnitude of the standard error is most variable at larger lags. This result can be explained in that the standard error is a function of all autocorrelations at lags less than the lag being considered. The autocorrelation function of a data set with greater serial dependence will exhibit large autocorrelations at several lags, and consequently, autocorrelations estimates will be more variable. In each case, the estimated standard error decreases as the number of data points increases.

The partial autocorrelation function of lags 2, 3 and 4 of the Autoregressive process are presented in Table 50. The true value of the parameter for all conditions is zero. In all cases, the bias in the estimator results in a mean estimate less than zero. The extent of the bias is minimal for conditions in the longer series of forty observations. The underestimation of the partial autocorrelation coefficient is considerably larger for shorter series of twenty data points. The variability of the partial autocorrelation estimates increases as the length of the realization becomes greater.

Moving Average processes. The mean and standard deviation of the estimates of the autocorrelation and partial autocorrelation functions for the Moving Average series are presented in Tables 51 and 52. The bias in estimation of the autocorrelation function drifts upwards, which results in an overestimation of the autocorrelation. The degree of overestimation increases as the autocorrelation parameter becomes smaller and as the number of observations becomes smaller. The standard error becomes smaller as the autocorrelation parameter and series length become larger. The relative magnitude of the discrepancy in the estimates and standard error is greater at larger lags. The bias in the estimation of the partial autocorrelation function results in a mean estimate less than zero in all cases.

Integrated Moving Average processes. The mean and standard deviation of the estimates of the autocorrelation and partial autocorrelation functions for the Integrated Moving Average series are presented in Tables 53 and 54. Similar results occurred as were reported with the Moving Average processes and will not be discussed further.

Nonstationary processes. The mean and standard deviation of the estimates of the autocorrelation and partial autocorrelation functions for the Nonstationary series are presented in Tables 55 and 56. Similar results occurred as were previously reported with the differenced Moving Average processes and will not be discussed further.

White Noise. The mean and standard deviation of the estimates of the autocorrelation and partial autocorrelation functions for White Noise are presented in Tables 57 and 58. The bias in estimation of the autocorrelation function is always downward, which results in an underestimation of the autocorrelation. The degree of underestimation increases as the number of observations becomes smaller. The bias in the partial autocorrelation function results in a mean estimate less than zero in all cases. The standard error becomes smaller as the series length becomes longer.

Table 49

Mean Estimated Autocorrelation (ACF) and

Empirical Standard Error (SE)

Autoregressive Processes

AR	n	Lag 1 ACF	Lag 1 SE	Lag 2 ACF	Lag 2 SE	Lag 3 ACF	Lag 3 SE	Lag 4 ACF	Lag 4 SE
0.8	TP	0.800	-	0.640	-	0.512	-	0.410	-
	40	0.689	0.117	0.460	0.177	0.294	0.207	0.175	0.219
	20	0.561	0.192	0.284	0.243	0.104	0.238	-0.010	0.219
0.5	TP	0.500	-	0.250	-	0.125	-	0.062	-
	40	0.437	0.141	0.167	0.172	0.043	0.169	-0.014	0.165
	20	0.363	0.200	0.089	0.216	-0.044	0.213	-0.083	0.193
0.2	TP	0.200	-	0.040	-	0.008	-	0.002	-
	40	0.151	0.151	-0.004	0.152	-0.024	0.158	-0.020	0.154
	20	0.102	0.208	-0.036	0.199	-0.053	0.197	-0.054	0.181

Table 50

Mean Estimated Partial Autocorrelation (PACF) and

Empirical Standard Error (SE)

Autoregressive Processes

AR	n	Lag 2 PACF	Lag 2 SE	Lag 3 PACF	Lag 3 SE	Lag 4 PACF	Lag 4 SE	Estimated SE
0.8	TP	0.000	-	0.000	-	0.000	-	-
	40	-0.056	0.139	-0.013	0.066	-0.035	0.107	0.167
	20	-0.109	0.187	-0.045	0.123	-0.052	0.159	0.250
0.5	TP	0.000	-	0.000	-	0.000	-	-
	40	-0.055	0.149	-0.024	0.122	-0.028	0.141	0.167
	20	-0.102	0.189	-0.070	0.175	0.093	0.172	0.250
0.2	TP	0.000	-	0.000	-	0.000	-	-
	40	-0.052	0.150	-0.015	0.156	-0.038	0.149	0.167
	20	-0.094	0.193	-0.047	0.192	-0.087	0.169	0.250

Table 51

Mean Estimated Autocorrelation (ACF) and

Empirical Standard Error (SE)

Moving Average Processes

MA	n	Lag 1 ACF	Lag 1 SE	Lag 2 ACF	Lag 2 SE	Lag 3 ACF	Lag 3 SE	Lag 4 ACF	Lag 4 SE
0.8	TP	-0.488	-	0.000	-	0.000	-	0.000	-
	40	-0.460	0.118	-0.017	0.186	-0.006	0.173	0.005	0.173
	20	-0.445	0.155	-0.014	0.235	0.006	0.228	-0.009	0.221
0.5	TP	-0.400	-	0.000	-	0.000	-	0.000	-
	40	-0.388	0.115	-0.015	0.163	0.001	0.157	-0.007	0.153
	20	-0.371	0.167	-0.017	0.228	-0.013	0.225	-0.012	0.206
0.2	TP	-0.192	-	0.000	-	0.000	-	0.000	-
	40	-0.208	0.149	-0.002	0.158	-0.016	0.156	-0.021	0.155
	20	-0.204	0.202	-0.030	0.211	-0.035	0.196	-0.029	0.195

Table 52

Mean Estimated Partial Autocorrelation (PACF) and

Empirical Standard Error (SE)

Moving Average Processes

MA	n	Lag 2 PACF	Lag 2 SE	Lag 3 PACF	Lag 3 SE	Lag 4 PACF	Lag 4 SE	Estimated SE
0.8	TP	-0.312	-	-0.221	-	-0.165	-	-
	40	-0.311	0.131	-0.194	0.137	0.042	0.162	0.167
	20	-0.298	0.177	-0.143	0.172	0.016	0.213	0.250
0.5	TP	-0.190	-	-0.094	-	-0.047	-	-
	40	-0.213	0.133	-0.097	0.132	0.005	0.149	0.167
	20	-0.216	0.190	-0.091	0.196	-0.016	0.199	0.250
0.2	TP	-0.038	-	-0.008	-	-0.002	-	-
	40	-0.072	0.146	-0.024	0.149	-0.036	0.148	0.167
	20	-0.122	0.198	-0.060	0.188	-0.058	0.182	0.250

Table 53

Mean Estimated Autocorrelation (ACF) and

Empirical Standard Error (SE)

Integrated Moving Average Processes

IMA	n	Lag 1		Lag 2		Lag 3		Lag 4	
		ACF	SE	ACF	SE	ACF	SE	ACF	SE
0.8	TP			-0.488	-	0.000	-	0.000	-
	40	0.020	0.110	-0.440	0.116	0.002	0.177	-0.023	0.171
	20	0.013	0.151	-0.419	0.157	-0.011	0.225	-0.011	0.220
0.5	TP			-0.400	-	0.000	-	0.000	-
	40	0.150	0.113	-0.332	0.129	-0.018	0.168	-0.025	0.171
	20	0.128	0.167	-0.319	0.174	-0.017	0.220	-0.043	0.205
0.2	TP			-0.192	-	0.000	-	0.000	-
	40	0.334	0.119	-0.167	0.161	-0.033	0.170	-0.031	0.167
	20	0.297	0.170	-0.186	0.223	-0.052	0.211	-0.057	0.208

Table 54

Mean Estimated Partial Autocorrelation (PACF) and

Empirical Standard Error (SE)

Integrated Moving Average Processes

IMA	n	Lag 2 PACF	Lag 2 SE	Lag 3 PACF	Lag 3 SE	Lag 4 PACF	Lag 4 SE	Estimated SE
0.8	TP	-0.488	-	-0.312	-	-0.221	-	-
	40	-0.458	0.115	0.028	0.136	-0.228	0.107	0.167
	20	-0.452	0.151	0.012	0.183	-0.204	0.141	0.250
0.5	TP	-0.400	-	-0.190	-	-0.094	-	-
	40	-0.380	0.118	0.123	0.147	-0.148	0.136	0.167
	20	-0.380	0.160	0.092	0.187	-0.165	0.155	0.250
0.2	TP	-0.192	-	-0.038	-	-0.008	-	-
	40	-0.334	0.125	0.183	0.144	-0.057	0.161	0.167
	20	-0.339	0.174	0.117	0.185	-0.111	0.180	0.250

Table 55

Mean Estimated Autocorrelation (ACF) and

Empirical Standard Error (SE)

Nonstationary Processes

NS	n	Lag 1 ACF	Lag 1 SE	Lag 2 ACF	Lag 2 SE	Lag 3 ACF	Lag 3 SE	Lag 4 ACF	Lag 4 SE
0.0	TP			0.000	-	0.000	-	0.000	-
	40	0.458	0.107	-0.049	0.174	-0.045	0.175	-0.051	0.175
	20	0.394	0.160	-0.122	0.221	-0.093	0.210	-0.083	0.206

Table 56

Mean Estimated Partial Autocorrelation (PACF) and

Empirical Standard Error (SE)

Nonstationary Processes

NS	n	Lag 2 PACF	Lag 2 SE	Lag 3 PACF	Lag 3 SE	Lag 4 PACF	Lag 4 SE	Estimated SE
0.0	TP			0.000	-	0.000	-	-
	40	-0.344	0.124	0.185	0.134	-0.017	0.162	0.167
	20	-0.368	0.161	0.127	0.194	-0.082	0.182	0.250

Table 57

Mean Estimated Autocorrelation (ACF) and

Empirical Standard Error (SE)

White Noise

WN	n	Lag 1 ACF	Lag 1 SE	Lag 2 ACF	Lag 2 SE	Lag 3 ACF	Lag 3 SE	Lag 4 ACF	Lag 4 SE
0.0	TP	[illegible]		0.000	-	0.000	-	0.000	-
	40	-0.029	0.150	-0.030	0.148	-0.031	0.144	-0.028	0.140
	20	-0.064	0.205	-0.050	0.205	-0.035	0.203	-0.044	0.185

Table 58

Mean Estimated Partial Autocorrelation (PACF) and

Empirical Standard Error (SE)

White Noise

WN	n	Lag 2 PACF	Lag 2 SE	Lag 3 PACF	Lag 3 SE	Lag 4 PACF	Lag 4 SE	Estimated SE
0.0	TP	0.000	-	0.000	-	0.000	-	-
	40	-0.055	0.149	-0.033	0.145	-0 049	0.133	0.167
	20	-0.101	0.206	-0.051	0.196	-0.080	0.171	0.250

Accuracy of Correlation Parameter Estimates

The results of the comparison of the estimated standard error of the autocorrelation function and empirical standard error assessed as the standard deviation of the estimates over 500 replications are presented in Tables 59 to 63. For the Autoregressive processes, the mean estimated standard error is consistently greater than the empirical standard error. The discrepancy between the estimated and empirical standard errors is largest for the Autoregressive processes with greater values of the correlation parameter. Also, the discrepancy is related to the length of the time series realization. Table 59 illustrates the results at lags 1 through 4, respectively.

For the Moving Average processes, the estimated standard errors at those lags where the true autocorrelation parameter is equal to zero (lags 2, 3, and 4) are relatively close to the empirical estimates of the standard error for the longer series of forty observations. On the other hand, at lag 1 there is a tendency to overestimate the standard error of the autocorrelation coefficient. These results are demonstrated in Table 60.

For the Integrated Moving Average processes, again at those lags where the true autocorrelation parameter is equal to zero, the estimated standard errors are relatively close to the empirical estimates of the standard error for the longer length series. Whereas at lag 1 and lag 2 there is a tendency to overestimate the standard error of the autocorrelation coefficient. These results are shown in Table 61. For the Nonstationary series, parallel results occurred and are demonstrated in Table 62.

For White Noise, the mean estimated standard error is consistently greater than the empirical standard error. The discrepancy is greatest at increasing lags and decreasing series length. Table 63 illustrates these results.

Therefore, the procedure for estimating the approximate standard errors of the autocorrelation function often overestimates the actual magnitude of the standard error. The greater the autocorrelation coeffi ient differs from zero, the greater the standard error estimates deviate from the true standard error. In most conditions, the estimated standard errors are larger than the true values, which may result in a large Type II error rate. Series lengths of forty and less observations provide relatively inaccurate estimates of the standard error of the autocorrelation coefficients.

Table 59

Mean Estimated Standard Error (EST) and

Empirical Standard Error (EMP)

of the Autocorrelation Function

Autoregressive Processes

AR	n	Lag 1 EST	Lag 1 EMP	Lag 2 EST	Lag 2 EMP	Lag 3 EST	Lag 3 EMP	Lag 4 EST	Lag 4 EMP
ϕ 0.8									
	40	0.158	0.117	0.222	0.177	0.246	0.207	0.258	0.175
	20	0.224	0.192	0.290	0.243	0.311	0.238	0.321	0.219
ϕ 0.5									
	40	0.158	0.141	0.188	0.172	0.195	0.169	0.199	0.165
	20	0.224	0.200	0.258	0.216	0.268	0.213	0.276	0.193
ϕ 0.2									
	40	0.158	0.151	[illegible]	0.152	0.168	0.158	0.172	0.154
	20	0.224	0.208	0.235	0.199	0.243	0.197	0.251	0.181

Table 60

Mean Estimated Standard Error (EST) and

Empirical Standard Error (EMP)

of the Autocorrelation Function

Moving Average Processes

MA	n	Lag 1 EST	Lag 1 EMP	Lag 2 EST	Lag 2 EMP	Lag 3 EST	Lag 3 EMP	Lag 4 EST	Lag 4 EMP
θ 0.8									
	40	0.158	0.118	0.190	0.186	0.195	0.173	0.198	0.173
	20	0.224	0.155	0.268	0.235	0.278	0.228	0.286	0.221
θ 0.5									
	40	0.158	0.115	0.182	0.163	0.185	0.157	0.189	0.153
	20	0.224	0.167	0.257	0.228	0.267	0.225	0.276	0.206
θ 0.2									
	40	0.158	0.149	0.168	0.158	0.171	0.156	0.175	0.155
	20	0.224	0.202	0.241	0.211	0.250	0.196	0.257	0.195

Table 61

Mean Estimated Standard Error (EST) and

Empirical Standard Error (EMP)

of the Autocorrelat'on Function

Integrated Moving Average Processes

IMA	n	Lag 1 EST	Lag 1 EMP	Lag 2 EST	Lag 2 EMP	Lag 3 EST	Lag 3 EMP	Lag 4 EST	Lag 4 EMP
θ 0.8									
	40	0.15[illegible]	[illegible]0	0.160	0.116	0.189	0.177	0.193	0.171
	20	0.224	[illegible]1	0.229	0.157	0.268	0.225	0.277	0.220
θ 0.5									
	40	0.158	0.113	0.163	0.129	0.182	0.168	0.185	0.171
	20	0.224	0.167	0.233	0.174	0.259	0.220	0.268	0.205
θ 0.2									
	40	0.158	0.119	0.177	0.161	0.184	0.170	0.188	0.167
	20	0.224	0.170	0.264	0.223	0.264	0.211	0.272	0.208

Table 62

Mean Estimated Standard Error (EST) and

Empirical Standard Error (EMP)

of the Autocorrelation Function

Nonstationary Processes

		Lag 1		Lag 2		Lag 3		Lag 4	
NS	n	EST	EMP	EST	EMP	EST	EMP	EST	EMP
0.0									
	40	0.158	0.107	0.189	0.174	0.194	0.175	0.198	0.175
	20	0.224	0.160	0.260	0.221	0.272	0.210	0.281	0.206

Table 63

Mean Estimated Standard Error (EST) and

Empirical Standard Error (EMP)

of the Autocorrelation Function

White Noise

		Lag 1		Lag 2		Lag 3		Lag 4	
WN	n	EST	EMP	EST	EMP	EST	EMP	EST	EMP
0.0									
	40	0.158	0.150	0.162	0.148	0.165	0.144	0.168	0.140
	20	0.224	0.205	0.233	0.205	0.242	0.203	0.251	0.185

<u>Estimates of Type I Error and Power for Correlation Parameters</u>

The proportion of replications for which the estimate was outside the region of nominal acceptance were calculated, resulting in a Type I error rate for conditions in which the true autocorrelation is equal to zero, and a power estimate for those conditions with a non-zero autocorrelation parameter. The results are presented in Tables 64 to 68. For the Autoregressive processes in Table 64 the autocorrelation function is greater than zero for lags 1 through 4. Thus rejection of the null hypothesis is a measure of the power of the procedure to detect a significant autocorrelation parameter. It can be seen that the power of the test varies with the length of the time series realization and the magnitude of the parameter. The theoretical autocorrelation function for the Autoregressive process is characterized by an exponential rate of decay. As the lag increases, the autocorrelation parameter of an Autoregressive process becomes smaller and more difficult to detect. The proportion of null hypotheses rejected for the test of the partial autocorrelation coefficient are also presented. The coefficients are zero for all cases and the Type I error in these conditions are very close to the nominal value of alpha with a tendency toward a deflated Type I error rate.

Moving Average processes are presented in Table 65. In this situation, the autocorrelation coefficient at lags greater than 1 is equal to zero. The empirical Type I error rates in these cases is relatively close to the nominal alpha level of 0.05. The power to detect true differences from zero at lag 1 is fairly good for the time series realization of length forty and high autocorrelation parameter values. Failure to consider autocorrelation coefficients at lags 2 and beyond as different from zero may result in a misidentification of an Autoregressive process as a Moving Average process.

The Integrated Moving Average processes are illustrated in Table 66. Similar results are found in the Nonstationary processes presented in Table 67. Fairly good power levels are found with series length of forty observations. The Type I error rate for the White Noise process is depicted in Table 68. Empirical Type I error rates for White Noise are similar or slightly lower than the nominal error rate of 0.05. Therefore, for practical implications, the power to detect true differences from zero at each lag 1 through 4 is necessary to form accurate judgments about the form of the autocorrelation and partial autocorrelation functions in order to properly identify a time series process.

A chi-square test (Q-statistic) of lags 1 through 10 was performed to determine whether the series or differenced series is White Noise. Regions of nominal acceptance or rejection were constructed around the correlation parameter estimates based on the chi-square distribution with df = k - p - q, and the approximation of the standard error of the estimator. The proportion of replications that resulted in the rejection of the null hypothesis, H_0:ACF(1) = ... = ACF(k) = 0, were calculated for each condition. The null hypothesis states that the residual autocorrelation function is not different than White Noise. Tables 69 and 70 present the estimates of the empirical Type I error rates for conditions in which the null hypothesis is true. The higher rate of Type I error for longer length series could be explained by the use of only up to lag 10 for estimation of the Q-statistic. For the shorter length series of twenty observations, the empirical alpha level is considerably less than the nominal alpha level. The test statistic is sensitive to the value of k, that is, to the number of lags in the residual autocorrelation function. For a relatively long autocorrelation function of 40 lags, the test statistic is likely to underestimate the serial correlation in the model residuals. The problem here is in that using an autocorrelation function of twenty lags, the null hypothesis might be rejected. For the same set of residuals, using an autocorrelation of 40 lags, the null hypothesis might not be rejected. It is recommended that a minimum of 20 lags be used in calculation of the test statistic. The conditions when bias of the Q-statistic is troublesome occur with the estimation procedure, relatively short time series realizations, and relatively high degrees of serial dependence.

Time Series model effect on correlation parameter estimates. In summary, autocorrelation and partial autocorrelation functions of five different time series processes were examined. For each of the models, the length of the time series realization and the value of the serial dependence of the observation were varied. The approximate identification of an ARIMA (p, d, q) process is difficult for short series. Given the combination of bias and large standard errors of the estimated autocorrelation and partial autocorrelation functions, the accuracy of ARIMA models identified on the basis of forty or less observations is questionable. The power to detect autocorrelation parameters is greater at longer length series. Also, a correlogram should not be computed without regard to intervention effects. The presence of an intervention effect can greatly increase autocorrelation coefficients.

Table 64

Proportion of Null Hypotheses Rejected for

Four Lags of Estimated Autocorrelations (ACF)

and Partial Autocorrelations (PACF)

Autoregressive Processes

Nominal Alpha = 0.05

AR	n	ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
ϕ 0.8	TP	0.800	0.640	0.512	0.410	0.000	0.000	0.000
	40	0.994	0.560	0.208	0.056	0.016	0.002	0.004
	20	0.774	0.114	0.000	0.000	0.016	0.000	0.004
ϕ 0.5	TP	0.500	0.250	0.125	0.062	0.000	0.000	0.000
	40	0.812	0.118	0.028	0.016	0.028	0.004	0.026
	20	0.378	0.022	0.044	0.004	0.020	0.012	0.006
ϕ 0.2	TP	0.200	0.040	0.008	0.002	0.000	0.000	0.000
	40	0.152	0.040	0.038	0.022	0.038	0.026	0.030
	20	0.042	0.018	0.018	0.006	0.012	0.016	0.004

Table 65

Proportion of Null Hypotheses Rejected for

Four Lags of Estimated Autocorrelations (ACF)

and Partial Autocorrelations (PACF)

Moving Average Processes

Nominal Alpha = 0.05

MA	n	ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
θ 0.8	TP	-0.488	0.000	0.000	0.000	-0.312	-0.221	-0.165
	40	0.884	0.044	0.026	0.030	0.426	0.138	0.050
	20	0.526	0.018	0.008	0.002	0.124	0.022	0.026
θ 0.5	TP	-0.400	0.000	0.000	0.000	-0.190	-0.094	-0.047
	40	0.738	0.028	0.020	0.016	0.204	0.032	0.024
	20	0.650	0.022	0.012	0.008	0.058	0.018	0.018
θ 0.2	TP	-0.192	0.000	0.000	0.000	-0.038	-0.008	-0.002
	40	0.250	0.038	0.028	0.020	0.028	0.022	0.030
	20	0.116	0.018	0.010	0.000	0.024	0.012	0.000

Table 66

Proportion of Null Hypotheses Rejected for

Four Lags of Estimated Autocorrelations (ACF)

and Partial Autocorrelations (PACF)

Integrated Moving Average Processes

Nominal Alpha = 0.05

IMA	n	ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
θ 0.8	TP		-0.488	0.000	0.000	-0.488	-0.312	-0.221
	40	0.004	0.854	0.032	0.024	0.860	0.010	0.170
	20	0.008	0.412	0.020	0.008	0.390	0.004	0.022
θ 0.5	TP		-0.400	0.000	0.000	-0.400	-0.190	-0.094
	40	0.062	0.526	0.024	0.028	0.654	0.080	0.086
	20	0.024	0.220	0.008	0.008	0.238	0.010	0.006
θ 0.2	TP		-0.192	0.000	0.000	-0.192	-0.038	-0.008
	40	0.568	0.120	0.036	0.028	0.542	0.142	0.048
	20	0.192	0.090	0.012	0.010	0.192	0.016	0.012

Table 67

Proportion of Null Hypotheses Rejected for

Four Lags of Estimated Autocorrelations (ACF)

and Partial Autocorrelations (PACF)

Nonstationary Processes

Nominal Alpha = 0.05

NS	TP	ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
0.00	TP		0.000	0.000	0.000		0.000	0.000
	40	0.900	0.034	0.024	0.026	0.532	0.138	0.034
	20	0.378	0.030	0.012	0.000	0.214	0.028	0.006

Table 68

Proportion of Null Hypotheses Rejected for

Four Lags of Estimated Autocorrelations (ACF)

and Partial Autocorrelations (PACF)

White Noise

Nominal Alpha = 0.05

WN	n	ACF1	ACF2	ACF3	ACF4	PACF2	PACF3	PACF4
0.00	TP	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	40	0.038	0.026	0.018	0.018	0.032	0.022	0.024
	20	0.044	0.020	0.018	0.008	0.022	0.012	0.004

Table 69

Proportion of Null Hypotheses Rejected for Ten Lags

of Autocorrelation (ACF) Estimates

Nominal Alpha = 0.05

True Value Parameter	n	ARIMA MODEL AR	MA	IMA
0.8				
	40	0.826	0.198	0.142
	20	0.074	0.008	0.000
0.5				
	40	0.266	0.096	0.088
	20	0.002	0.002	0.000
0.2				
	40	0.026	0.050	0.086
	20	0.000	0.000	0.002

Table 70

Proportion of Null Hypotheses Rejected for Ten Lags

of Autocorrelation (ACF) Estimates

Nominal Alpha = 0.05

True Value Parameter	n	ARIMA MODEL NS	WN
0.0			
	40	0.204	0.008
	20	0.000	0.000

CHAPTER VI

Summary and Conclusions

The primary purpose of this study was to explore the use of Box-Jenkins-Tiao interrupted time series models in assessing intervention effects with short time series realizations. Monte Carlo simulations were conducted to provide information with respect to the small sample properties of several estimators used in time series analysis. The first set of Monte Carlo simulations examined the small sample properties of Box and Tiao's (1965, 1975) test statistic for the presence of an intervention effect in a time series process, while the second set of simulations investigated procedures used in the model identification stage of ARIMA (p, d, q) time series analysis. In all of the simulations, the nature of the autocorrelation structure and the length of the time series realization were manipulated.

The discussion in the following sections summarizes the results of the study in terms of the research questions posed and the relationship of these results to previous relevant studies. Limitations of the study as well as suggestions for further investigation in the application of time series analysis in clinical research are proposed.

Intervention Effect Estimation in Time Series Processes

Monte Carlo simulations were used to examine the statistical test of an abrupt, permanent change in level of a stationary time series process as proposed by Box and Tiao (1965, 1975). The criteria employed for assessing the estimation of intervention effects in short time series processes were: (1) Type I error rate, (2) power to detect a non-zero intervention effect, and (3) accuracy of the estimates of the intervention effect. In this study, the estimates of the intervention parameter do not seem to be biased. When the correct ARIMA model was specified, the estimates of the intervention effect deviated only slightly from the true value. The results of the study also indicate that generally, the estimated standard error of the intervention component was considerably smaller than the standard deviation of the 500 estimates of the intervention effect. In addition, the Type I error rate of the test statistic was inflated for most conditions considered, with the inflation being smaller in longer series than in shorter time series realizations. This result can most likely be attributed to the noted underestimation of the standard error of the intervention effect and the use of the least squares estimation procedure. This problem was notably acute for short Moving Average time series. Finally, the power of the test statistic is less than desirable for many conditions that were studied.

Therefore, although the correct ARIMA model works fairly well in the intervention context, the results also indicate that for every model, there is at least one misidentification that leads to serious error. Thus no single model appears adequate for all time series processes. An arbitrary strategy of always fitting a particular kind of ARIMA model will lead to unacceptable Type I error rates in some cases and very reduced power in other situations. For example, always using an Autoregressive model will give reduced power if the true model is Moving Average; always using a Moving Average model will give increased Type I error if the true model is Autoregressive, especially with high values of ϕ; always using White Noise will give high Type I errors when the true model is Autoregressive, or will give a too conservative test if the true model is Moving Average. However, the Autoregressive model tended, overall, to give intervention effect estimates that were fairly close to the true value. Inaccurate estimates were encountered when the model misidentification involved an incorrect order of differencing. In general, misspecifying the order of differencing led to intervention effect estimates that were relatively extreme to the true value. Differencing a time series process that does not require it, gave low Type I error rates, but had a detrimental effect on power. While no single ARIMA model can be fit in all cases, there are models that, at particular values of the correlation coefficient, behave very much like other models either in terms of Type I error or power. In the case of these similar models, misidentifications may not have as serious consequences. Furthermore, problems with the estimation of the intervention effect become more severe as serial correlation increases; the inflation of Type I error becomes greater and there is a large reduction in the statistical power to detect an intervention effect.

The problems encountered in intervention effect estimation with short time series processes of twenty and forty data points were generally consistent with results of three other relevant studies. Padia (1975) used 122 data points and fixed values of the parameters for the misidentified ARIMA models. Padia found inflated Type I error rates for intervention effect estimates when identifying a differenced model as White Noise, misidentifying an Autoregressive model as a Moving Average model, and decreased Type I error rates when misidentifying a true Moving Average series as Autoregressive or White Noise models. Padia also claimed that regardless of the value of differencing, the results of the misidentifications would be the same. For example, misidentification of the Moving Average model as White Noise (d=0) should give the same results as misidentifying the Integrated Moving Average model as the Nonstationary model (d=1). In the present study, the results for Type I error of intervention effect estimates when misidentifying the Moving Average model as White Noise were 0.00 to 0.02, whereas misidentifying the Integrated Moving Average model as the Nonstationary model gave Type I error rates of 0.03 to 0.07.

Marquis (1983) used twenty and 50 data points with varying values of the parameters for the misidentified ARIMA models. Marquis also found estimates of the Type I error rates for intervention effects considerably higher than the nominal level when the model is inadequately differenced. Type I error rates reported were consistent with, yet lower than, those of the present study. Misidentifying a true Moving Average series as Autoregressive or White Noise models achieved decreased intervention effect Type I error rates. However, Marquis' Type I error rates when identifying a differenced model as White Noise were nearly half as large as those obtained by Padia (1975). Reduced power was also found in misidentifying the Moving Average model as White Noise or the Nonstationary model. Marquis found very conservative intervention effect Type I error rates (0.00 to 0.01) when misidentifying the Moving Average model as White Noise, whereas misidentifying the Integrated Moving Average as the Nonstationary model gave error rates of 0.04 to 0.05.

White (1985) used 60, 90, 120, or 150 time points with varying values of the parameters for only the Autoregressive time series process. No consequences of testing for an intervention effect when the

ARIMA (p, d, q) model is misidentified were investigated. Similar results in intervention effect estimation problems — the magnitude of standard errors, the accuracy of estimated standard errors, inflation of Type I error rates, and lack of power — were also found with the Autoregressive model in short time series realizations and increased serial correlation. Type I error rates reported over 1000 replications ranged from 0.14 to 0.07 at values of ϕ equal to 0.9 and 0.3, respectively.

In summary, the size of the error in estimation of intervention effects generally decreased with the increased length of time series realizations. The mean estimated standard error was generally smaller than the empirically obtained standard error, which increased the Type I error rate of intervention effect estimates. The inflated Type I error rate in estimation of intervention effects was remarkably predominate with Moving Average series. As tests of significance are based on the estimated standard error, the bias in the least squares procedure for estimating the approximate standard error of the autocorrelation function overestimated the actual magnitude of the standard error, thereby contributing to high Type I error rates and low power. The power of intervention effect detection increased with the number of data points. The placement of the intervention parameter in the time series realization did not greatly affect the problems encountered in estimation with short time series processes. As autocorrelation between data points contributes to the accuracy of interrupted time series analysis, ARIMA model identification is therefore a crucial step in the assessment of intervention effects.

Time Series Model Estimation

Monte Carlo simulations were used to examine the properties of the estimators of autocorrelation and partial autocorrelation in time series processes. The estimated autocorrelation and partial autocorrelation functions provide the basis for model identification, and thus, an understanding of their small sample properties is essential for the meaningful application of interrupted time series analysis. The results of this study emphasize the importance of measuring the time series process over a sufficient number of data points. As demonstrated in this set of simulations, the problems of estimation and discrepancy in the estimates are attenuated as the number of observations in the time series decreases. In comparing the estimated values of the standard error with empirical estimates of the standard error of the autocorrelation and partial autocorrelation functions, a wide discrepancy for many of the conditions investigated was found, with the estimated standard error generally exceeding the empirical standard error. As the tests of significance are based on the estimated standard errors, the empirical Type I error rates and power of the test statistic for short time series realizations tend to be insufficient due to the overestimation of the standard error of the autocorrelation function.

The results indicate that, in fact, it is difficult to identify the correct structure of correlation in the data with as few as twenty observations; the power to detect a significant ARIMA parameter in the short series is quite low. When the number of data points increased to forty, the amount of bias generally decreased and the power increased. These results reinforce previous caveats in the literature that time series of limited length make identification of a correct model difficult (Box & Jenkins, 1976; Glass, Willson & Gottman, 1975; Hartmann, Gottman, Jones, Gardner, Kazdin & Vaught, 1980; Jones, Vaught & Weinrott, 1977; Osborn, 1980). The direction and amount of bias depends on the actual values of the correlation parameter, the method for calculating the initial values in the estimation procedure, and the actual estimation function used. The results of this study generally follow the pattern reported in other studies; an underestimation of ϕ at all values of the correlation coefficient and an overestimation of θ.

In summary, the estimated autocorrelation and partial autocorrelation functions improve with increasing data points; the standard error of estimation decreases with the increased length of time series realizations. Thus, the percentage of null hypotheses rejected increases with a larger number of observations; the power to detect true differences from zero at lag 1 is greater for longer time series, with Type I error rates being closer to the nominal level. In the model identification stage of time series analysis, problems in estimation of the autocorrelation function increase as the serial dependency becomes more severe.

Limitations of the Study

The findings of the present study are limited in several respects. The limitations of the study relate to those parameters that were not manipulated or were limited in the values assumed. The conditions selected for investigation were limited to: the AR(1), MA(1), IMA(1,1), NS, and WN models; an intervention component of an abrupt permanent change in level added at an even or uneven time point; correlation parameter values of 0.2, 0.5, or 0.8; and time series realizations of length twenty or forty observations. Numerous other conditions could have been selected for investigation which may have led to different results. Investigation of other first-order models as well as second-order models, and the use of negative values for the correlation parameters, especially in Moving Average models, could be undertaken.

The practical considerations in using simulation procedures involved conducting 500 replications for each condition. Although the trends in the ARIMA models and misspecifications were consistent, more replications may have given a better indication of the population value. Specifically, more information on the estimates of Type I error for the Autoregressive and Moving Average models with only twenty data points would be beneficial. As well, full maximum likelihood estimation procedures, although costly, may be useful in analyzing a limited set of time series realizations.

Although the significance of the placement of the intervention effect was investigated, another type of model misspecification that warrants attention is the change in the time series process at the point of intervention. The intervention component may simultaneously affect both the level of the time series process and the nature of the interdependence among the observations; however, this study did not investigate this form of model misspecification. An assumption of this study was that the intervention did not change the underlying ARIMA model. Investigation of time series in which the intervention may change the variance of the series, or change the size of the correlation parameter, or change the ARIMA model, is needed to determine how to recognize these changes and how they may affect intervention effect assessment.

Conclusions From Simulation Studies

The primary purpose of this study was to explore the use of Box-Jenkins-Tiao interrupted time series models in assessing intervention effects in small samples. Since the mathematical theory on which the procedures are based assumes large samples, it is difficult to provide definitive guidelines for the application of interrupted time series analysis with limited observations. In addition to providing specific information with respect to the small sample properties of the estimators used in time series analysis, what conclusions can be drawn by considering these two sets of Monte Carlo simulations and what questions arise from these results to guide future research?

Length of Time Series Realizations

An important conclusion concerns the length of the time series realization that is necessary to obtain meaningful results on the basis of the statistical procedures in interrupted time series experiments. The small sample properties of the test statistic are sample dependent, and thus vary according to the autocorrelation structure of a particular data set. This difficulty is accentuated by the large standard error of the estimated autocorrelation coefficients that are based on a small number of data points. In addition, the problems are magnified when analyzing time series data with extreme serial dependence. Therefore, the extensive discrepancy and bias of the small sample estimates limit the usefulness of pilot testing as a method of evaluating the extent of serial dependency in a time series process. Consequently, in applications of time series analysis, the number of observations required is determined in the absence of knowledge concerning the autocorrelation structure of the process.

Based on the findings presented, it is recommended that time series realizations consist of more than forty observations. The length of time series realizations play a critical role in determining the quality of estimates that are obtained when applying interrupted time series procedures. The magnitude of standard errors, inaccuracy of estimated standard errors, inflation of Type I error rates, and lack of power, are less severe for more lengthy time series realizations. Thus, effort should be made to obtain lengthy data sets, and recognize that conclusions based on short time series realizations may be misleading.

Robustness of Time Series Procedures

Additional concerns in the application of time series analysis to actual data sets involves the assumptions of stationarity and the proper identification of an ARIMA (p, d, q) model. The results presented in this study are based on the assumptions of homogeneity of variance and normal distribution of the error term in simulated data sets that are generated according to known stationary ARIMA processes. In actual practice, data sets are not likely to fit an identified ARIMA (p, d, q) process as closely as the generated data sets of the Monte Carlo simulations. Furthermore, in social science research, time series data are unlikely to conform to the assumption of stationarity, and thus, the idiosyncracies of the actual data sets may magnify the concerns identified in the present study. How well these procedures work in short time series where the assumptions of the model are violated requires more investigation.

Serial Dependence in Time Series Processes

Another general conclusion concerns the degree of serial dependence in time series processes. Generally, for all conditions examined in this study, high serial dependence increases the severity of estimation problems in interrupted time series procedures. In the model identification stage of time series analysis, both the bias in the estimation of the autocorrelation function and the overestimation of the standard error of autocorrelation coefficients becomes more severe as the serial dependence becomes more severe. Furthermore, problems in the estimation of the intervention component become more severe as the serial correlation increases; inflation of Type I error increases and the statistical power to detect an intervention effect decreases. It is important for researchers to be aware of the severity of the estimation problems that are encountered when the autocorrelation among data points is large.

In summary, although the statistical analysis of time series processes is a valuable tool, it is important in the application of ARIMA models, to be aware of the limitations and potential problems with the procedures. Two questions should be addressed: (1) Given that some correct ARIMA model identifications will occur as well as some misidentifications, what expectations can be drawn regarding Type I error and power to detect an intervention effect? (2) Given that the correct ARIMA model works well in assessing intervention effects, and given that ARIMA model misidentifications can lead to serious error, what guidelines are available to identify the correct model and rule out unsatisfactory models, in short time series realizations?

Type I Error Rates and Power for Intervention Effect Estimates

In practical applications of time series analysis with short series, it appears that if unsatisfactory ARIMA models can be ruled out, one might expect a slightly higher Type I error rate than the nominal level and somewhat lower power than would be obtained by always using the correct model. Definite statements regarding this conclusion depend upon the likelihood of being both able to rule out ARIMA models with very serious errors and to select models that are correct or close to the true ARIMA model. As noted, selection of the Moving Average model, particularly which short time series, renders highly inflated Type I errors in estimation of the intervention effect. The bias in the least squares analysis procedure results in underestimation of the standard error of the intervention component for Moving Average processes. Application of the Moving Average model to short time series should be done with caution and knowledge of the bias in parameter estimation.

Time Series Model Identification

A useful tool in the ARIMA model identification stage of a time series process is the autocorrelation function. Anderson (1976) suggests that the autocorrelation function is unstable for lags greater than n/4, where n is the number of observations. In the case of short time series realizations with twenty observations, only five values of the autocorrelation function would be used to help identify the model. This appears to be too few to be helpful in model identification. The autocorrelation function is used to determine the adequacy of the selected ARIMA model. The lack of significant correlation in the residuals is generally assumed to be confirming evidence for the fit of an ARIMA model. The usefulness of this tool for short series of twenty and even forty observations is questionable. The findings affirmed that the power to detect a significant ARIMA parameter in short series is low, especially when the values of the autocorrelation coefficient are small and especially for second and higher autocorrelation functions. As the value of the autocorrelation coefficient increases, the power to detect a significant correlation coefficient increases. Can two different ARIMA models for the same time series process have significant ARIMA parameters? There are no clear criteria for selecting the correct model in short time series processes.

Another diagnostic tool for the adequacy of an ARIMA model is the Q-statistic which considers the general residual autocorrelation function taken as a whole. The statistic is sensitive to the value of the number of lags (k) (Davis, Triggs, & Newbold, 1977). Newbold (1981) reported that k must be at least twenty, a value that precludes using it in series of twenty data points. Additionally, in ruling out poor ARIMA model choices, the standard error of estimate tends to reflect the degree of accuracy of the fitted model. As noted, in short time series the approximate standard errors of the autocorrelation function often overestimate the actual magnitude of the standard errors. The greater the autocorrelation coefficient differs from zero, the greater the standard error estimates deviate from the true standard error. Marquis (1983) reported that ARIMA misidentifications with serious consequences generally had an average residual mean square that was significantly larger than that associated with the true ARIMA model. Perhaps with further investigation, the residual mean square may prove to be a diagnostic tool in ruling out unsatisfactory ARIMA models in short time series processes.

Another aid in determining the correct ARIMA model may be to examine the patterns which appear to develop in the data. It was noted from the simulation data that the parameter estimates tended to cluster in patterns based on the order of differencing; the Nonstationary and Integrated Moving Average model estimates tended to group together, while the Moving Average and White Noise model estimates were generally similar. Moreover, the estimates of the Autoregressive model were generally closest to the true model. Similarly, the pattern that occurs in the standard error of estimates may also aid in selecting a correct ARIMA model. Does the Moving Average model consistently underestimate the standard error? For the ARIMA models requiring differencing, which assumed non-differenced model has the highest error of estimate? For a true ARIMA model, which assumed model has the highest average residual mean square?

In summary, with short time series processes it is important to identify the correct ARIMA model. The tools used to identify the correct model need to be adapted and further investigated to be helpful in the identification process of short time series processes. Although methods for the statistical analysis of time series processes are a functional tool, it is important to be aware of the inherent limitations and potential problems in application of interrupted time series procedures.

Application of Interrupted Time Series Analysis

The measurement of the varied phenomena of interest to nurses spans the field of basic, applied, and clinical research as it is germane to nursing practice. The refinement of research techniques to measure the phenomena that are uniquely nursing are crucial issues in nursing research. There is a relationship between research in the field of professional study, the nature of research in that field, and the obligation to ask questions about how to improve practice. Does nursing care make a difference? The questions of intervention research are: Is the intervention effective? When is the intervention most effective? Has a reliable change been produced? The issue is one of separating intervention effects from mere fluctuations. Time series analysis procedures may be a viable strategy for reducing the uncertainty in nursing interventions. An analysis of outcome studies in nursing research could be done to determine what effect time series designs and analysis techniques would have on the results. Specific questions to be addressed are: (1) When is time series analysis appropriate in nursing research? (2) What is the degree of serial dependence in nursing data? (3) Are interrupted time series designs applicable in nursing research? (4) What ARIMA models are applicable in nursing research?

Data evaluation consists of methods that are used to draw conclusions about behavior change. In clinical nursing research, experimental and therapeutic criteria are invoked to evaluate data. The experimental criterion refers to the way in which data are evaluated to determine if an intervention has had a reliable effect on behavior. The therapeutic criterion refers to whether the effects of the intervention are important. Even if the change is reliable and clearly related to the intervention, the change may not be of clinical significance. Statistical analysis provides a quantitative method to determine if a particular intervention effect is reliable and determines whether the change meets the experimental criterion. Research that is focused on the evaluation of nursing interventions enables nursing to advance as a profession by increasing the scientific base for practice. However, individual studies of the effects of nursing interventions on patient outcomes frequently consist of small samples and may not produce statistically significant results. Also, evaluation of intervention effects can be difficult when performance during baseline is often systematically improving. Statistical analysis in single-subject research may provide a valuable supplement to visual inspection about the effects of an intervention. Research-based nursing practice can offer professional nurses the means with which to quantify patient outcomes.

Time series analysis compares data over time for separate phases for an individual or group of subjects. The most basic time series design requires one experimental unit and multiple observations before and after a treatment. Changes in a time series which coincide with the occurrence of an intervention are presumed to be the effects of the intervention. This claim may be invalid; events unrelated to the intervention may cause the series to change abruptly at the point of intervention and/or random variation in a time series may be misinterpreted as the effect of an intervention. Even though the time series design does not eliminate all of the problems of interpreting change, the extended time perspective strengthens the ability to attribute any change to the experimental manipulation. The time series design permits the ruling out of the possibility that the data reflect an unstable measurement made only at two time points.

One major advantage of a time series design over other forms of quasi-experimental analysis is that a maturational trend can be assessed prior to the intervention. Secondly, with time series designs, the possibility of a cyclical trend in the data being interpreted as an intervention effect is reduced. As well, a further strength of time series designs is that they allow assessment of the pretest trend, thereby permitting a check on the plausibility of statistical regression accounting for alternative explanations. The main threat to internal validity with most single time series designs is the effect of history -- the possibility that factors other than the intervention came to influence the dependent variable. A control for history is possible by adding a no-intervention control group. Another threat is instrumentation. In time series designs, it is important to pay attention to the definition of constructs and to possible changes in experimental procedures. Simple selection can also occur in time series designs when the composition of the experimental unit changes abruptly at the time of the intervention. When this happens, it is necessary to determine whether the intervention caused an interruption in the series or whether the interruption was due to different persons being in the pre- and post-intervention series. Finally, time series are subject to many influences of a cyclical nature, including attitudes or variations in performance, and these influences can be interpreted as intervention effects. A lengthy time series is therefore required to control for and to estimate the cyclical pattern.

In terms of construct validity, interpretation of the data for a single time series is not likely to be hindered if the same respondents are repeatedly measured and if they know when the intervention was implemented. Typically there will only be one operationalization of the intervention effect in time series data. Therefore, the intervention needs to be independently measured and tailored to a specific construct. However, there is the difficulty of obtaining multiple measures of an effect, as well as obtaining reasonably reliable measurements and consequently, the time series for each measure requires separate examination. Additionally, to determine external validity, the time series for each experimental unit needs to be examined to determine whether an effect holds across various subgroups. One is generally restricted to the variables in the investigation. Finally, as noted previously, inferential techniques based on the assumption of independent data cannot safely be applied to nonstationary time series. The practical significance of the failure to deal with the dependence among observations is that null hypotheses will be rejected with high power even when the time series evidences no abrupt intervention effect. ARIMA time series analysis procedures help only in ruling out threats to inferences regarding statistical conclusion validity.

In contrast to conventional tests, time series analysis depends upon the serial dependency in the data, adjusts to the specific dependency relationships among data points, and provides separate analysis for changes in light of special characteristics of the data. The time series design provides a method appropriate to the complexity of the effects of interventions with human beings. However, short time series have two undesirable consequences: (1) model-fitting (especially differencing) is performed with less confidence in the adequacy of the model, and (2) statistical tests of intervention effects are less likely to detect real changes. Greater than forty observations seem necessary to decrease Type I error and increase power of intervention effect detection, especially with Moving Average series. In short time series realizations where the usual identification and diagnostic procedures may not work very well, the best strategy may be to fit all simple ARIMA models and select the best one on the basis of the pattern of the parameter estimates, the significant parameters, and the residual mean square. If there is a large residual mean square, the model may be ruled out; if the correlation parameter is not significant, the model is probably not appropriate. Questions to guide the analysis would be: What is the size and pattern of parameter estimates? Do the estimates of the correlation parameter coincide with the expectations of the ARIMA process? Are the estimates of the intervention effect consistent and close in value over other models? Is the intervention effect significant regardless of the ARIMA model?

This suggested method of fitting all simple ARIMA models departs from the proposals of Simonton (1977), Gottman (1981), Velicer and Harrop (1983), and Harrop and Velicer (1985), which bypass the troublesome identification step and apply the Autoregressive model to a wide variety of complex time series designs. When serial correlation exists in the errors, the correlation between observations from the same subject would be expected to decrease as the time separation increases. The simplest model for serial correlation when the repeated observations on a subject are ordered in time (not randomized) is the first-order Autoregressive process. Although the use of only the first-order or higher-order Autoregressive models is potentially more flexible, results of the simulations in this study support the problem of inflated error rates. As well, perhaps the Autoregressive model may not be representative of the time series processes occurring in clinical nursing research, recalling that Glass, Willson, and Gottman (1975) reported the Integrated Moving Average model to be the most common model found in 22 out of 95 social science cases. Identification of the structure of autocorrelation in nursing data is needed to reduce bias in detecting the ARIMA parameters, recalling that as serial dependency increases, it becomes harder to estimate the ARIMA model. Determining the degree of serial correlation in nursing processes would assist in deciding which approach is more feasible: fitting all simple ARIMA models or the sole use of a first-order Autoregressive process.

Despite the utility of the interrupted time series design, a number of problems frequently arise in view of the practical and inferential difficulties in conducting interrupted time series research. Many clinical interventions are not implemented rapidly. A gradual versus an abrupt change may slowly diffuse through a population so that the intervention is better modeled as a diffusion ogive than as a step function. Also, many effects are not instantaneous but occur with unpredictable time delays which may differ among populations from one historical moment to the next. Delayed causation may occur even when a treatment is homogeneously implemented. Finally, many time series are shorter than forty data points, with fifty observations being recommended for statistical analysis (Glass, Willson, & Gottman, 1975; Gottman, 1981; McCleary & Hay, 1980). Situations arise in clinical research when fewer observations than this are possible. However, more observations are achieved than with a single pretest-posttest design.

To illustrate the use of interrupted time series analysis techniques, the results of the simulation studies can be applied to potential studies in clinical nursing research. The status of a patient undergoing intensive post-operative therapy is evaluated by monitoring cardiovascular variables, such as heart rate, associated with the patient. Clinically significant changes in the status of the patient can be indicated by step changes and trends in the monitored series. At the same time, artifacts may be observed with external factors not associated with the status of the patient. A method of detecting artifacts, step changes and slope changes, and a way of distinguishing between them is required. For example, post-cardiac surgery, the variables of heart rate and arterial blood pressure could be monitored at one-minute intervals over ninety data points. Intuitively, the time series realization would most likely fit the Integrated Moving Average model. Slope and step changes could then be distinguished from artifacts.

Another application of time series procedures could be in evaluating electrocardiogram changes following ice water ingestion by coronary patients. Does the time and volume of ice water ingestion have an effect on the electrical conduction of cardiac muscle? Electrocardiogram recordings could be taken at three, ten, and twenty-five minutes post-ingestion. The ARIMA model would be identified and the effects of ice water ingestion detected. The time series design assists in establishing internal validity of these case studies and through replication, external validity can subsequently be established.

Time series analysis could also provide a physiological foundation for establishing the timing as well as the effectiveness of nursing therapies. Is there a physiological basis for the timing of electrolyte replacements? The circadian rhythms of plasma electrolytes could be modeled to determine the effectiveness of nursing therapies on electrolyte balance. For example, the results of such a study may provide guidelines for timing potassium replacement with circadian variation to achieve an optimal patient response.

Numerous examples could be given where the sample in nursing research is not homogeneous and single-subject time series analysis could be applied: clinical drug trials for monitoring drug therapy in a mentally retarded person; evaluating treatment outcomes in the elderly where individual variation increases with age, psychomotor response decreases, and brevity of observations is necessitated. However, if one departs from the straightforward comparisons of changes from one phase to another, interrupted time series analysis is of unknown relevance in assessing the comparative effectiveness of two or more concurrent treatments. These might be examined in a simultaneous treatment design or a multiple-baseline design.

In conclusion, interrupted time series analysis provides a methodology for the quantitative synthesis of intra-subject design research. Time series designs provide the opportunity for concurrent as opposed to retrospective assessment of observations over time. To illustrate, the recording of life events and symptoms over time might detect a time series model of living with uncertainty of a chronic illness. By conducting interrupted time series analysis, the clinical researcher can: improve the quality of visually based judgements, assess more adequately the pattern of treatment effects, and protect against both false acceptance of non-existent effects and the false rejection of existent treatment effects. These benefits should compensate for the effort required to perform the analysis and the precautions needed in the applica of interrupted time series analysis procedures with short time series processes.

References

Algina, J., & Olejnik, S. (1982). Multiple group time-series design: An analysis of data. Evaluation Review, 6(2), 203-232.

Algina, J., & Swaminathan, H. (1977). A procedure for the analysis of time series designs. Journal of Experimental Education, 45(3), 56-61.

Algina, J., & Swaminathan, H. (1979). Alternatives to Simonton's analysis of interrupted and multiple-group time-series designs. Psychological Bulletin, 86(5), 919-926.

Anderson, O. D. (1976). Time series analysis and forecasting: The Box-Jenkins approach. Toronto, Ontario: Butterworth & Company (Canada) Limited.

Anderson, O.D. (1977). A commentary on a survey of time-series. International Statistical Review, 45(3), 273-297.

Anderson, O.D. (1979). On warming-up time series simulations generated by Box-Jenkins models. Journal of the Operational Research Society, 30(6), 587-589.

Anderson, O.D. (Ed.). (1980). Analyzing time-series. Amsterdam: North Holland Publishing Company.

Ansley, C., & Newbold, P. (1980). Finite sample properties of estimators for autoregressive moving average models. Journal of Econometrics, 13(2), 159-183.

Baer, D. M. (1977). "Perhaps it would be better not to know everything." Journal of Applied Behavior Analysis, 10(1), 167-172.

Bakan, D. (1966). The test of significance in psychological research. Psychological Bulletin, 66(6), 423-437.

Barlow, D. H. (1981). On the relation of clinical research to clinical practice: Current issues, new directions. Journal of Consulting and Clinical Psychology, 49(2), 147-155.

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatment design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behavior Analysis, 12(2), 199-210.

Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs. Archives General Psychiatry, 29(9), 319-325.

Barlow, D. H., & Hersen, M. (1984). Single case experimental designs: Strategies for studying behavior change (2nd Ed.). New York: Pergamon Press.

Berman, J. J., Meyer, J., & Coats, G. (1984). Effects of program characteristics on treatment outcome: An interrupted time-series analysis. Journal of Studies on Alcohol, 45(5), 405-410.

Berryman, D., & Cooper, B. (1982). The use of single-subject research methodology in special education. Educational Psychology, 2(3-4), 197-213.

Beutler, L.E., & Hamblin, D.L. (1986). Individualized outcome measures of internal change: Methodological considerations. Journal of Consulting and Clinical Psychology, 54(1), 48-53.

Blose, J. O., & Holder, H. D. (1987). Liquor-by-the-drink and alcohol-related traffic crashes: A natural experiment using time-series analysis. Journal of Studies on Alcohol, 48(1), 52-60.

Blumberg, C. J. (1984). Comments on "A simplified time-series analysis for evaluating treatment interventions." Journal of Applied Behavior Analysis, 17(4), 539-542.

Bowie, C., & Prothero, D. (1981). Finding causes of seasonal diseases using time series analysis. International Journal of Epidemiology, 10(1), 87-92.

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. San Francisco: Holden-Day.

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control (rev. ed.). San Francisco: Holden-Day.

Box, G.E.P., & Pierce, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association, 65(332), 1509-1526.

Box, G. E. P., & Tiao, G. C. (1965). A change in level of a nonstationary time-series. Biometrika, 52(1-2), 181-192.

Box, G.E.P., & Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70(349), 70-92.

Bracht, G.H., & Glass, G.V. (1968). The external validity of experiments. American Educational Research Journal, 5(4), 437-474.

Brown, D.T. (1975). Down with the linear model. American Educational Research Journal, 12(4), 491-505.

Brown, J. S., Tanner, C. A., & Padrick, K. P. (1984). Nursing's search for scientific knowledge. Nursing Research, 33(1), 26-33.

Browning, R. M. (1967). A same-subject design for simultaneous comparison of three reinforcement contingencies. Behavior Research and Therapy, 5(3), 237-243.

Burckhardt, C. S., Goodwin, L. D., & Prescott, P. A. (1982). The measurement of change in nursing research: Statistical considerations. Nursing Research, 31(1), 53-55.

Calsyn, R.J. (1977). Interrupted time series analysis: A research technique for evaluating social programs for the elderly. Journal of Gerontology, 32(1), 89-96.

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company.

Carver, R.P. (1978). The case against statistical significance testing. Harvard Educational Review, 48(3), 378-399.

Catalano, R.A., Dooley, D., & Jackson, R. (1983). Selecting a time-series strategy. Psychological Bulletin, 94(3), 506-523.

Chatfield, C. (1980). The analysis of time series: An introduction. London: Chapman & Hall.

Center, D. B., & Leach, R. A. (1984). The multiple baseline across subjects design: Proposed use in research. Journal of Manipulative and Physiological Therapeutics, 7(4), 231-236.

Center, B., Skiba, R., & Casey, A. (1985-86). A methodology for the quantitative synthesis of intra-subject design research. The Journal of Special Education, 19(4), 387-400.

Chassan, J. B. (1959). On the development of clinical statistical systems for psychiatry. Biometrics, 15(3), 396-404.

Chassan, J. B. (1960). Statistical inference and the single case in clinical design. Psychiatry, 23(2), 173-184.

Chatfield, C. (1975). The analysis of time series: Theory and practice. London: Chapman and Hall.

Clinton, J., Beck, R., Radjenovic, D., Taylor, L., Westlake, S., & Wilson, S. (1986). Time-series designs in clinical nursing research: Human issues. Nursing Research, 35(3), 188-191.

Cobb, E. B. (1985). Planning research studies: An alternative to power analysis. Nursing Research, 34(6), 386-388.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd Ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Collins, F. L., Jr., Baer, R. A., & Blount, R. L. (1985). Single-subject research designs for optometry. American Journal of Optometry and Physiological Optics, 62(8), 516-522.

Connell, P. J., & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. Journal of Speech and Hearing Disorders, 51(3), 214-225.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally College Publishing Company.

Cook, T. D., Gruder, C. L., Hennigan, K. M., & Flay, B. R. (1979). The history of the sleeper effect: Some logical pitfalls in accepting the null hypothesis. Psychological Bulletin, 86(4), 662-679.

Cook, T. D., & Reichardt, C. S. (1976). Statistical analysis of non-equivalent control group designs: A guide to more current literature. Evaluation, 3(1-2), 136-138.

Cook, T. J. (1983). The delayed treatment design - A cautionary note. Evaluation Review, 7(2), 277-280.

Cook, T.J., & Poole, W.K. (1982). Treatment implementation and statistical power: A research note. Evaluation Review, 6(3), 425-430.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"- Or should we? Psychological Bulletin, 74(1), 68-80.

Dalton, D. R., & Todor, W. D. (1984). Unanticipated consequences of union-management cooperation: An interrupted time series analysis. The Journal of Applied Behavioral Science, 20(3), 253-264.

Dar, R. (1987). Another look at Meehl, Lakatos, and the scientific practices of psychologists. American Psychologist, 42(2), 145-151.

Dattilo, J., & Nelson, G. D. (1986). Single-subject evaluation in health education. Health Education Quarterly, 13(3), 249-259.

Davies, N., Triggs, C.M., & Newbold, P. (1977). Significance levels of the Box-Pierce portmanteau statistic in finite samples. Biometrika, 64(3), 517-522.

Demorest, M. E., & Walden, B. E. (1984). Psychometric principles in the selection, interpretation, and evaluation of communication self-assessment inventories. Journal of Speech and Hearing Disorders, 49(3), 226-240.

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. Journal of Applied Behavior Analysis, 12(4), 573-579.

Derrick, T. (1976). The criticism of inferential statistics. Educational Review, 19(1), 35-48.

Deutsch, S. J. (1979). Lies, damn lies, and statistics: A rejoinder to the comment of Hay and McCleary. Evaluation Quarterly, 3(2), 315-328.

Donaldson, S., & Crowley, D. (1978). The discipline of nursing. Nursing Outlook, 26(2), 113-120.

Dow, M. M., White, D. R., & Burton, L. M. (1983). Multivariate modeling with interdependent network data. Behavior Science Research, 17(3-4), 216-245.

Edgington, E. S. (1967). Statistical inference from N=1 experiments. The Journal of Psychology, 65(2), 195-199.

Edgington, E. S. (1972). The design of one-subject experiments for testing hypotheses. The Western Psychologist, 3(1), 33-38.

Edgington, E. S. (1982). Nonparametric tests for single-subject multiple schedule experiments. Behavioral Assessment, 4(1), 83-91.

Edgington, E.S. (1985). Random assignment and experimental research. Educational Administration Quarterly, 21(3), 235-246.

Edgington, E.S. (1987). Randomized single-subject experiments and statistical tests. Journal of Counseling Psychology, 34(4), 437-442.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician, 37(1), 36-48.

Egger, M. J., & Miller, J. R. (1984). Testing for experimental effects in the pretest-posttest design. Nursing Research, 33(5), 306-312.

Feldman, M. J., & Ventura, M. R. (1984). Evaluating change using noninterval data. Nursing Research, 33(2), 182-184.

Furlong, M. J., & Wampold, B. E. (1982). Intervention effects and relative variation as dimensions in experts' use of visual inference. Journal of Applied Behavior Analysis, 15(3), 415-421.

Gentile, J. R., Roden, A. H., & Klein, R. D. (1972). An analysis of variance model for the intrasubject replication design. Journal of Applied Behavior Analysis, 5(2), 193-198.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association, 77(378), 304-314.

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory. Chicago: Aldine.

Glaser, B.G. (1978). Theoretical sensitivity. Mill Valley: Sociology Press.

Glass, G.V. (1968). Analysis of data on the Connecticut speeding crackdown as a time-series quasi-experiment. Law and Society Review, 3(1), 55-76.

Glass, G.V. (1972). Estimating the effects of intervention into a non-stationary time-series. American Educational Research Journal, 9(3), 463-476.

Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions of fixed effects analyses of variance and covariance. Review of Educational Research, 42(3), 237-288.

Glass, G.V., Tiao, G., & Maguire, T.O. (1971). Analysis of data on the revision of German divorce laws as a time-series quasi-experiment. Law and Society Review, 5(4), 539-562.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). Design and analysis of time-series experiments. Boulder, Colorado: Colorado Associated University Press.

Godfrey, L.G. (1979). Testing the adequacy of a time series model. Biometrika, 66(1), 67-72.

Goodwin, L. D. (1984). The use of power estimation in nursing research. Nursing Research, 33(2), 118-120.

Goodwin, L. D. (1984). Increasing efficiency and precision of data analysis: Multivariate vs. univariate statistical techniques. Nursing Research, 33(4), 247-249.

Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. Behavioral Assessment, 5(2), 141-154.

Gottman, J. M. (1973). N-of-one and N-of-two research in psychotherapy. Psychological Bulletin, 80(2), 93-105.

Gottman, J. M. (1981). Time-series analysis. New York: Cambridge University Press.

Gottman, J. M., & Glass, G. U. (1978). Analysis of interrupted time-series experiments. In T.R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change (pp. 197-236). New York: Academic Press.

Gottman, J. M., McFall, R. M., & Barnett, J. T. (1969). Design and analysis of research using time series. Psychological Bulletin, 72(4), 299-306.

Grant, I., Yager, J., Sweetwood, H. L., & Olshen, R. (1982). Life events and symptoms - Fourier analysis of time series from a three-year prospective inquiry. Archives General Psychiatry, 39(5), 598-605.

Graupe, D. (1984). Time series analysis, identification and adaptive filtering. Florida: Robert E. Krieger Publishing Company.

Greenberg, L.S. (1986). Change process research. Journal of Consulting and Clinical Psychology, 54(1), 4-9.

Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. Psychological Bulletin, 82(1), 1-20.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? Psychological Bulletin, 83(2), 314-320.

Gregson, R. A. M. (1983). Time series in psychology. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gregson, R. A. M. (1984). Invariance in time series representations of 2-input 2-output psychological experiments. British Journal of Mathematical and Statistical Psychology, 37(1), 100-121.

Gregson, R.A.M. (1987). The time-series analysis of self-reported headache sequences. Behavior Change, 4(2), 6-13.

Hacker, B. (1980). Single subject research strategies in occupational therapy: Part 1. The American Journal of Occupational Therapy, 34(2), 103-108.

Hacker, B. (1980). Single subject research strategies in occupational therapy: Part 2. The American Journal of Occupational Therapy, 34(3), 169-175.

Hamilton, S. B., & Waldman, D. A. (1983). Self-modification of depression via cognitive-behavioral intervention strategies: A time series analysis. Cognitive Therapy and Research, 7(1), 99-106.

Hannan, E.J. (1979). A note on autoregressive-moving average identification. Biometrika, 66(3), 672-674.

Harrop, J. W., & Velicer, W. F. (1985). A comparison of alternative approaches to the analysis of interrupted time-series. Multivariate Behavioral Research, 20(1), 27-44.

Hartmann, D. P. (1974). Forcing square pegs into round holes: Some comments on "An analysis-of-variance model for the intrasubject replication design." Journal of Applied Behavior Analysis, 7(4), 635-638.

Hartmann, D. P., Gottman, J. M., Jones, R. R., Gardner, W., Kazdin, A. E., & Vaught, R. S. (1980). Interrupted time series analysis and its application to behavioral data. Journal of Applied Behavior Analysis, 13(4), 543-559.

Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. Journal of Applied Behavior Analysis, 9(4), 527-532.

Harvey, A.C. (1981). Time series models. New York: John Wiley & Sons.

Harvey, A.C., & Phillips, G.D.A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika, 66(1), 49-58.

Hasza, D. P. (1980). Asymptotic distribution of the sample autocorrelations for an integrated ARMA process. Journal of the American Statistical Association, 75(370), 349-352.

Hay, R. A., Jr., & McCleary, R. (1979). Box-Tiao time series models for impact assessment: A comment on the recent work of Deutsch and Alt. Evaluation Quarterly, 3(2), 277-314.

Hayes, S. C. (1981). Single case experimental design and empirical clinical practice. Journal of Consulting and Clinical Psychology, 49(2), 193-211.

Heath, L., Kendzierski, D., & Borgida, E. (1982). Evaluation of social programs - A multimethodological approach combining a delayed treatment true experiment and multiple time series. Evaluation Review, 6(2), 233-246.

Hibbs, D.A. (1974). Problems of statistical estimation and causal inference in time-series regression models. In H.L. Costner (Ed.), Sociological methodology. San Francisco: Jossey-Bass.

Holm, K. (1983). Single subject research. Nursing Research, 32(4), 253-255.

Horn, W. F., & Heerboth, J. (1982). Single-case experimental designs and program evaluation. Evaluation Review, 6(3), 403-424.

Horne, G. P., Yang, M. C. K., & Ware, W. B. (1982). Time series analysis for single-subject designs. Psychological Bulletin, 91(1), 178-189.

Horowitz, M.J., Marmar, C.R., Weiss, D.S., Kaltreider, N.B., & Wilner, N.R. (1986). Comprehensive analysis of change after brief dynamic psychotherapy. American Journal of Psychiatry, 143(5), 582-589.

Hugdahl, K., & Ost, L. (1981). On the difference between statistical and clinical significance. Behavioral Assessment, 3(3-4), 289-295.

Huitema, B. E. (1985). Autocorrelation in applied behavior analysis: A myth. Behavioral Assessment, 7(2), 107-118.

Jacobsen, B. S., & Meininger, J. C. (1985). The design and methods of published nursing research: 1956-1983. Nursing Research, 34(5), 306-312.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. Behavior Therapy, 15(4), 336-352.

Jirovec, M. M. (1986). Time-series analysis in nursing research: ARIMA modeling. Nursing Research, 35(5), 315-319.

Jones, R. H. (1985). Repeated measures, interventions, and time series analysis. Psychoneuroendocrinology, 10(1), 5-14.

Jones, R. R., Crowell, D. H., & Kapuniai, L. E. (1969). Change detection model for serially correlated data. Psychological Bulletin, 71(5), 352-358.

Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. Journal of Applied Behavior Analysis, 10(1), 151-166.

Jones, R. R., Weinrott, M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis, 11(2), 277-283.

Judd, C.M., & Kenny, D.A. (1981). Estimating the effects of social interventions. New York: Cambridge University Press.

Kaplan, A. (1964). The conduct of inquiry: Methodology for behavioral sciences. San Francisco, California: Chandler Publishing Company.

Kazdin, A. E. (1978). Methodological and interpretive problems of single-case experimental designs. Journal of Consulting and Clinical Psychology, 46(4), 629-642.

Kazdin, A. E. (1981). Drawing valid inferences from case studies. Journal of Consulting and Clinical Psychology, 49(2), 183-192.

Kazdin, A. E. (1982). Single-case research designs - Methods for clinical and applied settings. New York: Oxford University Press.

Kazdin, A. E. (1984). Statistical analyses for single-case experimental designs. In D.H. Barlow & M. Hersen (Eds.), Single case experimental designs (pp. 285-324). New York: Pergamon Press.

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. Journal of Consulting and Clinical Psychology, 54(1), 95-105.

Kazdin, A. E., & Kopel, S. A. (1975). On resolving ambiguities of the multiple-baseline design: Problems and recommendations. Behavior Therapy, 6(5), 601-608.

Kearns, K. P. (1986). Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. Journal of Speech and Hearing Disorders, 51(3), 204-214.

Kelly, F. J., McNeil, K., & Newman, I. (1973). Suggested inferential statistical models for research in behavior modification. The Journal of Experimental Education, 41(4), 54-63.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd Ed.). New York: Holt, Rinehart, & Winston, Inc.

Knapp, T. J. (1983). Behavior analysts' visual appraisal of behavior change in graphic display. Behavioral Assessment, 5(2), 155-164.

Knapp, T. R. (1985). Validity, reliability, and neither. Nursing Research, 34(3), 189-192.

Kratochwill, T. R. (Ed.). (1978). Single subject research - Strategies for evaluating change. New York: Academic Press.

Kratochwill, T., & Brody, G. (1978). Single subject designs: A perspective on the controversy over employing statistical inference and implications for research. Behavior Modification, 2(3), 291-307.

Kratochwill, T., Alden, K., Dawson, D., Panicucci, C., McMurray, P., Hempstead, J., & Levin, J. (1974). A further consideration in the application of an analysis-of-variance model for the intrasubject replication design. Journal of Applied Behavior Analysis, 7(4), 629-633.

Larsen, R.J. (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. Journal of Personality and Social Psychology, 52(6), 1195-1204.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297-303.

Lynn, M. R. (1985). Reliability estimates: Use and disuse. Nursing Research, 34(4), 254-256.

Mace, F. C., Page, T. J., Ivancic, M. T., & O'Brien, S. (1986). Effectiveness of brief time-out with and without contingent delay: A comparative analysis. Journal of Applied Behavior Analysis, 19(1), 79-86.

Maguire, T. O. (1986). Applications of new directions in statistics to educational research. The Alberta Journal of Educational Research, 32(2), 154-171.

Maguire, T. O., & Glass, G. V. (1967). A program for the analysis of certain time-series quasi-experiments. Educational and Psychological Measurement, 27(3), 743-750.

Makridakis, S. (1976). A survey of time series. International Statistical Review, 44(1), 29-70.

Makridakis, S. (1978). Time-series analysis and forecasting: An update and evaluation. International Statistical Review, 46(3), 255-278.

Marquis, J. Goldsberry. (1983). Box-Jenkins-Tiao intervention analysis in short time series. Unpublished Ph.D. dissertation, University of Kansas.

Maxwell, S., & Delaney, H. (1985). Measurement and statistics: An examination of construct validity. Psychological Bulletin, 97(1), 85-93.

Meehl, P. (1978). Theoretical risks, tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46(4), 806-834.

Metzger, B. L., & Schultz, S. (1982). Time series analysis: An alternative for nursing. Nursing Research, 31(6), 375-378.

McCain, L. J., & McCleary, R. (1979). The statistical analysis of the interrupted time-series quasi-experiment. In T.D. Cook & D.T. Campbell (Eds.), Quasi-experimentation: Design and analysis issues for field settings (pp. 233-294). Chicago: Rand McNally College Publishing Company.

McCall, B. B., & Applebaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. Child Development, 44(3), 401-415.

McCauley, R. J., & Swisher, L. (1984). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. Journal of Speech and Hearing Disorders, 49(4), 338-348.

McCleary, R., & Hay, R. A., Jr. (1980). Applied time series analysis for the social sciences. Beverly Hills, California: Sage Publications.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). Interrupted time series analysis. Beverly Hills, California: Sage Publications.

McNamara, J. R., & MacDonough, T. S. (1972). Some methodological considerations in the design and implementation of behavior therapy research. Behavior Therapy, 3(3), 361-378.

McReynolds, L. V., & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part I: Review of the basics of single subject designs. Journal of Speech and Hearing Disorders, 51(3), 194-203.

Meier, P., & Pugh, E. J. (1986). The case study: A viable approach to clinical research. Research in Nursing and Health, 9(3), 195-202.

Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? Journal of Applied Behavior Analysis, 7(4), 647-653.

Namboodiri, N. K. (1972). Experimental designs in which each subject is used repeatedly. Psychological Bulletin, 77(1), 54-64.

Nelson, N., Rosenthal, R., & Rosnow, R.L. (1986). Interpretation of significance level and effect size by researchers. American Psychologist, 41(11), 1299-1301.

Newbold, P. (1981). Some recent developments in time-series analysis. International Statistical Review, 49(1), 53-66.

Norbeck, J. S. (1987). In defense of empiricism. Image: Journal of Nursing Scholarship, 19(1), 28-30.

Nunnally, J.C. (1960). The place of statistics in psychology. Educational and Psychological Measurement, 20(4), 641-650.

Nurius, P. S. (1983). Use of time-series analysis in the evaluation of change due to intervention. The Journal of Applied Behavioral Science, 19(2), 215-228.

Osborn, D.R. (1982). On the criteria functions used for the estimation of moving average processes. Journal of the American Statistical Association, 77(378), 388-392.

Ostrom, C. W., Jr. (1978). Time series analysis: Regression techniques. Beverly Hills, California: Sage Publications.

Ottenbacher, K. J. (1986). Reliability and accuracy of visually analyzing graphed data from single-subject designs. The American Journal of Occupational Therapy, 40(7), 464-469.

Ottenbacher, K., & York, J. (1984). Strategies for evaluating clinical change: Implications for practice and research. American Journal of Occupational Therapy, 38(10), 647-659.

Padia, W.L. (1975). The consequences of model misidentification in the interrupted time-series experiment. Unpublished Ph.D. dissertation, University of Colorado.

Parsonson, B. S., & Baer, D. M. (1978). The analysis and interpretation of graphic data. In T.R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change (pp. 101-165). New York: Academic Press.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd Ed.). New York: Holt, Rinehart, & Winston, Inc.

Pertschuk, M. J., Edwards, N., & Pomerleau, O. F. (1978). A multiple-baseline approach to behavioral intervention in anorexia nervosa. Behavior Therapy, 9(3), 368-376.

Petrinovich, L. (1979). Probabilistic functionalism: A conception of research methods. American Psychologist, 34(5), 373-390.

Pierce, D. A. (1971). Least squares estimation in the regression model with autoregressive-moving average errors. Biometrika, 58(2), 299-312.

Rojahn, J., & Schulze, H. H. (1985). The linear regression line as a judgmental aid in visual analysis of serially dependent A-B time-series data. Journal of Psychopathology and Behavioral Assessment, 7(3), 191-206.

Rotton, J., & Frey, J. (1985). Air pollution, weather, and violent crimes: Concomitant time-series analysis of archival data. Journal of Personality and Social Psychology, 49(5), 1207-1220.

Rozeboom, W.R. (1966). The fallacy of the null-hypothesis significance test. Psychological Bulletin, 57(5), 416-428.

Serlin, R.C. (1987). Hypothesis testing, theory building, and the philosophy of science. Journal of Counseling Psychology, 34(4), 365-371.

Serlin, R.C., & Lapsley, D.K. (1985). Rationality on psychological research: The good-enough principle. American Psychologist, 40(1), 73-83.

Sharpley, C. F. (1986). TMS for ABA - A discussion of some issues in the use of time series analysis with behavioral data. Unpublished manuscript, Monash University, Clayton, Vic, 3168, Australia.

Shine, L. C., & Bower, S. M. (1971). one-way analysis of variance for single-subject designs. Educational and Psychological Measurement, 31(1), 105-113.

Shulman, J.L., Kupst, M.J., & Suran, B.J. (1976). The worship of "p": Significant yet meaningless research results. Bulletin of the Menninger Clinic, 40(2), 134-143.

Siegel, A. F. (1980). Testing for periodicity in a time series. Journal of the American Statistical Association, 75(370), 345-348.

Siegel, G. M., & Spradlin, J. E. (1985). Therapy and research. Journal of Speech and Hearing Disorders, 50(3), 226-230.

Silva, M.C. (1986). Research testing nursing theory: State of the art. Advances in Nursing Science, 9(1), 1-11.

Simonton, D. K. (1977). Cross-sectional time-series experiments: Some suggested statistical analyses. Psychological Bulletin, 84(3), 489-502.

Stake, R. E. (1978). The case study method in social inquiry. Educational Researcher, 7(2), 5-8.

Stanley, B. (1984). A portable method for analyzing time-series data in clinical practice. Perceptual and Motor Skills, 59(3), 915-916.

Stoline, M. R., Huitema, B. E., & Mitchell, B. T. (1980). Intervention time-series model with different pre- and post-intervention first-order autoregressive parameters. Psychological Bulletin, 88(1), 46-53.

Stoodley, K.D.C., & Mirnia, M. (1979). The automatic detection of transients, step changes, and slope changes in the monitoring of medical time-series. The Statistician, 28(3), 163-170.

Swaminathan, H., & Algina, J. (1977). Analysis of quasi-experimental time-series designs. Multivariate Behavioral Research, 12(1), 111-131.

Taylor, R. L., & Adams, G. L. (1982). A review ingle-subject methodologies in applied settings. The Journal applied Behavioral Science, 18(1), 95-103.

Thomson, C., & Baer, D. (1974). A brief report on a comparison of time-sampling procedures. Journal of Applied Behavior Analysis, 7(4), 623-626.

Thoresen, C. E., & Elashoff, J. D. (1974). "An analysis-of-variance model for intrasubject replication design": Some additional comments. Journal of Applied Behavior Analysis, 7(4), 639-641.

Tryon, W. W. (1984). "A simplified time-series analysis for evaluating treatment interventions": A rejoinder to Blumberg. Journal of Applied Behavior Analysis, 17(4), 543-544.

Tryon, W. W. (1982). A simplified time series analysis for evaluating treatment interventions. Journal of Applied Behavior Analysis, 15(3), 423-429.

Velicer, W. F., & Harrop, J. W. (1983). The reliability and accuracy of the time-series model identification. Evaluation Review, 7(4), 551-560.

Velicer, W. F., & McDonald, R. P. (1984). Time-series analysis without model identification. *Multivariate Behavioral Research*, *19*(1), 33-48.

Vitaliano, P. P. (1982). Parametric statistical analysis of repeated measures experiments. *Psychoneuroendocrinology*, *7*(1), 3-13.

White, J.B. (1985). *The assessment of intervention effects in time series processes*. Unpublished Ph.D. dissertation, University of Massachusetts.

Wilson, G. L., & Bornstein, P. H. (1984). Paradoxical procedures and single-case methodology: Review and recommendations. *Journal of Behavior Therapy and Experimental Psychiatry*, *15*(3), 195-203.

Wilson, W., Miller, H., & Lower, J. (196[illegible]). Much ado about the null hypothesis. *Psychological Bulletin*, *67*(3), 188-196.

Winch, R.F., & Campbell, D.T. (1969). Proof? No, evidence? Yes. The significance of tests of significance. *The American Sociologist*, *4*(2), 140-143.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd Ed.). New York: McGraw-Hill.

Wolery, M., & Harris, S. R. (1982). Interpreting results of single-subject research designs. *Physical Therapy*, *62*(4), 445-452.

PPENDIX A

TIME99 PROGRAM

TIME99
Jan 1988
Revised:
999/99-999

UNIVERSITY OF ALBERTA

Division of Educational Research Services

COMPUTER PROGRAM DOCUMENTATION

TITLE:	GENERATE AN ARIMA TIME SERIES
MACHINE:	AMDAHL 5270
LANGUAGE:	FORTRAN IV
PROGRAM TYPE:	COMPLETE
SYSTEM:	TIME
SUBPROGRAMS:	INTERNAL - PVEC7, GENSER, STDDEV XDER:SUB - TITLE, ERRR, FLGCHK, FLGCHR, CHKFMT, PVEC, WARN IMSL: GGUBFS, FTGEN
LIMITS:	See limits from page 2
TIME:	Approximately * Seconds for * Observations

PREPARATION CARDS

Card Seq	Card Type	Columns	Description
1	Title Card	1-80	Any description of the job being run. Must not be left blank.
2	Parameter Card	1-5	Number of cases to generate
		6-10	Length of time series (max.=100)
		11-15	Number of autoregressive parameters (Max.=10)
		16-20	Number of moving average parameters (Max.=10)
		21-25	'1' to input starting values; '0' otherwise
		26-30	Number of parameter sets (Max.=20)
		31-35	'1' to include difference specifications; '0' otherwise
		36-40	Number of initial points to discard
		41-45	'1' to introduce shock; '0' otherwise
		46-50	White noise variance
		51-55	Overall moving average parameter
3	Differencing Parameters	1-80	'1' to take differences; '0' otherwise
4	Shock	1-5	Number of pre-shock points
		6-10	Number of post-shock points
		11-15	Size of shock; 0 to 999
5	Format Card(s)	1-80	F-format for autoregressive and moving average parameters

Card Seq	Card Type	Columns	Description
6	Starting Values	1-80	Starting values in (16F5.3) format
7	Data Cards		Time series parameters as defined by format card
8	End of Data	1-8	$ENDFILE
9	End of Job Card		Blank card to indicate end of job; if more jobs, this card should be a Title Card of next job.

DATA OUTPUT BY PROGRAM

1. Title

2. Summary of input parameters (warning if errors detected)

3. Format for input of autoregressive and moving average parameters.

4. Interpretation of input format (warning if errors detected).

5. Time Series Parameter Set:
   - starting value for the time series
   - autoregressive parameter
   - moving average parameter

6. Confirmation of file output

<u>SAMPLE INPUT</u>

$SIGNON CSID	(SIGNON CARD)
PSWORD	(PASSWORD CARD)
$RUN XDER:PROGRAM PAR-TIME99	(RUN CARD)
Generate Moving Average Series Theta = 0.8	(TITLE CARD)
50 20 0 1 0 1 0 20 1 1.0 0.8	
5 15 0.0	
	(PARAMETER CARDS)
(2F3.1)	(Format Card)
0.00.2	(Parameter Card)
	(Blank Card)
$SIGNOFF	(Signoff Card)

```
TIME99 - GENERATE AN ARIMA TIME SERIES

LAST REVISION: DERS - JAN/88

        GENERATE MOVING AVERAGE SERIES THETA=0.8 OMEGA=0.5 N=20.1

               SUMMARY OF INPUT PARAMETERS

   NUMBER OF CASES TO GENERATE =                  50
   LENGTH OF TIME SERIES =                        20
   NUMBER OF AUTOREGRESSIVE PARAMETERS =           0
   NUMBER OF MOVING AVERAGE PARAMETERS =           1
   OPTION OF INPUTTING STARTING VALUES =           0
   NUMBER OF PARAMETER SETS =                      1
   OPTION OF INCLUDING DIFFERENCE CARDS =          0
   NUMBER OF INITIAL POINTS TO DISCARD =          20
   OPTION OF INTRODUCING A SHOCK =                 1
   WHITE NOISE VARIANCE =                      1.000
   OVERALL MOVING AVERAGE PARAMETER =          0.800

              SHOCK TO BE INTRODUCED INTO THE SERIES

   NUMBER OF PRE-SHOCK DATA POINTS =               5
   NUMBER OF POST-SHOCK DATA POINTS =             15
   SHOCK SCALAR PARAMETER =                  0.50000

FORMAT FOR AUTOREGRESSIVE AND MOVING AVERAGE PARAMETERS:
   (2F3.1)


FORMAT INTERPRETATION

RECORD #  COLUMNS   VARIABLE  DESCRIPTION
    1      1 - 3       1      REAL WITH     1 DECIMAL PLACES
    1      4 - 6       2      REAL WITH     1 DECIMAL PLACES

    2 VARIABLES READ FROM      1 RECORDS.

    PARAMETER SET    1
    *****************
 STARTING VALUES FOR TIME SERIES
    0.88282
 AUTOREGRESSIVE PARAMETERS
    0.0
 MOVING AVERAGE PARAMETERS
    0.80000
      50 CASES OUTPUT TO UNIT 7


 TIME99 NORMALLY TERMINATED WITH A BLANK CARD ON APR 21, 1988
```

SAMPLE OUTPUT

ERROR MESSAGES

1. An error message will be given and the run interrupted if the following parameters are outside their specified range:

Number of cases to generate	1 - 99999
Length of time series	1 - 10
Number of autoregressive parameters	0 - 10
Number of moving average parameters	0 - 10
Option to input starting values	0 - 1
Number of parameter sets	1 - 20
Option to include difference cards	0 - 1
Option of introducing a shock	0 - 1
white noise variance	0 - 99999
overall moving average parameter	0 - 99999

2. **ERROR 1: NUMBER OF PARAMETER CARD ERRORS IS nnn

   The number of errors found on the parameter card is printed. Computation stops.

3. **ERROR 2: FORMAT CARD ERROR

   This error indicates that the Format Card was uninterpretable. Possible causes include unmatched parentheses, invalid char:cters, or invalid character sequences. Computation stops.

APPENDIX B

TIME02 PROGRAM

U N I V E R S I T Y   O F   A L B E R T A

Division of Educational Research Services

Computer Program Documentation

TITLE: TIME SERIES ANALYSIS (PART 2) (GLASS)*
MACHINE: IBM 360/67
LANGUAGE: FORTRAN IV
PROGRAM TYPE: Complete
SUBPROGRAMS: LIMCHK,PSIWGT,RESVAL,SETLIM,SWP,SWPSET,FMTS
(XDER:SUB) TITLE
LIMITS: Total of pre-intervention plus post intervention points must be less than 300.
TIME: Time depends upon options selected.
PROGRAMMER: C.P. Bower, W.L. Padia, G.V. Glass and modified by T.C. Montgomerie
DOCUMENTED: T.C. Montgomerie

Description:

TIME01 and TIME02 are a pair of programs required for carrying out a time series analysis. Given a series of measures prior to an intervention point, and a series of measures after the intervention point, the programs allow an analysis to be made of the effects of the intervention. TIME01 permits the identification of the appropriate model to be used as well as its parameters. The parameters from TIME01 become input to TIME02 for a 'least squares normal theory analysis'.

References:

Bower, C.P., Padia, W.L., & Glass, G.V., TMS: Two Fortran IV Programs for analysis of Time-Series Experiments. Laboratory of Educational Research, University of Colorado, Boulder, Col., October 1974.

NOTE: In the above reference CORREL should be read as TIME01, and TSX should be read as TIME02 in order that the reader may relate the program described in the manual to those named in the DERS library.

*This program was obtained from the Laboratory of Educational Research at the University of Colorado. It was originally written for the CDC 6400 by C.P. Bower, W. Padia, and G.V. Glass.

Preparation of Header Cards for TIME02

Card Seq.	See Notes	Card Type	Cols.	Description
1		Title	1-80	Any description of the job
2	1	Parameters	1- 5	p, the order of the autoregression ($0 \le p \le 3$)
	1		6-10	d, order of differencing ($0 \le d \le 4$)
	1		11-15	q, order of moving average ($0 \le q \le 3$)
			16-20	number of pre-intervention points if column 30 is 0 or 1; if column 30 is 2, then m, the number of design parameters.
			21-25	number of points after intervention if column 30 is 0 or 1; else total number of points if col. 30 is 2
	2		26-30	0 for 2 design parameters (level) 1 for 4 design parameters (level plus drift) 2 for user specified design matrix
			31-35	number of points per cycle if seasonal data; else blank for non-seasonal model
	3		36-40	step increment for ϕ_j or θ_j if other than default values; else leave blank
	4		41-45	if 0, ϕ_j and θ_j will range from lower to upper bound of entire invertibility-stationarity region; 1 user option to limit range
	5		46-50	1 if residual values are to be printed and output on cards for values of ϕ_j and θ_j else leave blank.
			51-55	ϕ_1, if col. 50 is 1; blank if ϕ_1 greater than p
			56-60	ϕ_2, if col. 50 if 1; blank if ϕ_2 greater than p
			61-65	ϕ_3, if col. 50 is 1; blank if ϕ_2 greater than p
			66-70	θ_1, if col. 50 is 1; blank if θ_1 greater than p

Preparation of header Cards for TIME02 Continued:

Card Seq.	See Notes*	Card Type	Cols.	Description
			71-75	θ_2, if col. 50 is 1; blank if θ_2 greater than p
			76-80	θ_3, if col. 50 is 1; blank if θ_3 greater than p
3	6	Upper & Lower Limits	1- 5	Previous run step increment for ϕ_j's
		(only if col. 45 is 1 on Card 2)	6-10	current run step increment for ϕj's
			11-15	ϕ_1 from previous run
			16-20	ϕ_2 from previous run
			21-25	ϕ_3 from previous run
			26-30	previous run step increment for θ_j's
			31-35	current run step increment for θ_j's
			36-40	θ_1 from previous run
			41-45	θ_5 from previous run
			46-50	θ_3 from previous run
4	6	Format for Design Matrix	1-80	use only if col. 30 on Card 2 is a 2. Format is for transpose of N x m design matrix (N=total number of points, and m is number of design parameters.)
5	6	Design Matrix		N x m matrix according to format of Card 4 (use only if card 4 used)
6		Format for Data	1-80	format for data points. Data points must be less than 1000 in absolute value
7		Data Cards		as described in Card 6.
		End of job		A blank card if no further processing; else this should be the Title card of the next job.

User Notes*

1) p,d, and q specify the model for the data from TIME01.

2) 2-design parameter involves level and change in level, while the 4 design parameter model involves level and change in level, as well as drift and change in drift.

3) Default values for step increment are:

p & q	Step Increment
1	.02
2	.10
3	.25
4,5, & 6	.50

4) Default for upper and lower bounds on invertibility-stationarity regions are as follows:

a) p = 1 $-1<\phi_1<1$

b) p = 2 $-1<\phi_2<1$; $\phi_1+\phi_2<1$; $\phi_2-\phi_1<1$

c) p = 3 $-1<\phi_3<1$; $\phi_1+\phi_2+\phi_3<1$; $-\phi_1+\phi_2-\phi_3<1$; $\phi_3(\phi_3-\phi_1)-\phi_2<1$

d) the same are true independently for values of θ and q.

5) Note card output requires CARDS = on $SIGNON card.

6) These cards must be omitted if the noted parameter does not specify their use.

Manipulation Prior to Processing

No DATRAN subroutine is available.

Data Output by TIME02

A. User title on every page.

B. Number of points before and after intervention and degrees of freedom for the design (df = total number of data points minus number of parameters estimated). If user-supplied design matrix, only total number of points and degrees of freedom are printed.

C. Identification of model (p, d, q).

D. Any options selected by user.

E. Input format of data.

F. Design matrix format if specified by user.

Data Output by TIME02

G. Design matrix if specified by user.

H. Pre-intervention data points followed by post-intervention data points, unless design matrix specified by user; in this latter case, all data points are printed as a group.

I. For each increment of ϕ_j and/or θ_j:

1. Current values of ϕ_j and/or θ_j (values of θ_j are incremented before values of ϕ_j and with ϕ_j, ϕ_1 incremented first, then ϕ_2, then ϕ_3, and similarly for θ_j.

2. The residual error variance.

3. The estimate of the true level of the series at time 0 and a *t*-statistic. (If ICODE $\leq$ 1.)

4. The estimate of the change in level, δ, and a *t*-statistic for testing the significance of the difference of the estimate from zero. (If ICODE $\leq$ 1.)

5. The estimate of the deterministic drift component and associated *t*-statistic. (If ICODE = 1.)

6. The estimate of a change in the deterministic drift component and a *t*-statistic. (If ICODE = 1.)

7. If ICODE = 2, estimates of each design parameter and associated *t*-statistics are printed.

J. At the end of all iterations, the minimum error variance is printed with the values of ϕ_j and θ_j at which it was found. If, as occasionally happens, more than one set of values of ϕ_j and θ_j are found for the minimum error variance, up to five sets of values will be printed, and thereafter stopped, although the program counts and prints the total number of such occurrences.

K. If the user requested that residuals be printed and punched for a set of values of ϕ_j and θ_j, these are printed next, exactly as punched (both in 10F8.4 format).

Example of Input Data

```
$SIGNON CSID                                              (SIGNON CARD)
PSWORD                                                    (PASSWORD CARD)
$RUN *DEP:PROGRAM PAR=TIME02                              (RUN CARD)
SAMPLE PROBLEM FOR TIME SERIES2 (PART 2)                  (TITLE CARD)
    0     1            60    60                           (PARAMETER CARD)
(10F3.1)                                                  (FORMAT CARD)
55 56 48 46 56 46 59 60 53 58                             (DATA CARDS)
73 69 72 51 72 69 68 69 79 77
53 63 80 65 78 64 72 77 82 77
35 79 71 73 77 76 83 73 73 91
70 88 88 85 77 63 91 94 72 83
88 79 84 78 75 75 86 79 75 87
66 73 62 27 52 47 65 59 77 47
51 47 49 54 58 56 50 54 45 66
39 51 39 27 39 37 43 41 27 29
27 26 29 31 29 38 37 26 31 45
38 33 33 25 24 29 37 35 32 31
28 40 31 37 34 43 38 33 28 35

                                                          (BLANK CARD)
$SIGNOFF                                                  (SIGNOFF CARD)
```

Example of Output

See pages 9 to 11.

Error Messages

The error messages for TIME02 are self explanatory.

Technical Notes

The computational formulae used in TIME01 and TIME02 can be found in Appendix A of Bower, Padia & Glass.

Suggestions for Modifying Times for Other Installations

1. This program is a modification of the TSX program (Bower, Padia, & Glass) written for a CDC 6400. It is recommended that users of 10 byte/word machines should contact the Laboratory of Educational Research at the University of Colorado for a copy of the program which would be more easily implemented on their machines.

Suggestions for Modifying Times for Other Installations Continued:

2. The one subroutine which will require modifications on all machines not using the MTS operating System is the TITLE Subroutine. The function of the subroutine vital to the program is to check an incoming Title Card to see if it is totally blank, execution is terminated normally if this is the case. If the title card is not blank, the title is stored in the Unlabelled COMMON array, along with the time and the date. The Title, time and date are printed on a new page on the output device, and a normal exit to the calling program is made.

3. The word length of the IBM 360/67 is 4 bytes. If the machine which TIME01 and TIME02 being converted to does not have a 4 byte word size, all format statements containing A4 (or AX, where X is greater than the word length must be changed to the word length of the particular machine.

4. Variable formats are stored in LOGICAL*1 arrays. Machines which are incapable of handling character arrays stored as LOGICAL* data will need to convert the subroutine FMTS. This subroutine sets up the format for printing the headings for each page of iterations over phi and theta and the format for printing the results. The elements of both formats are variable because phi and theta may both range from zero to three and because the number of parameters and t - statistics is variable. In order to assist the programmer who has the task of modifying the programs, the following is a description of the contents of each of the 144 bytes of HFMT, and the 55 bytes of FMT, respectively, the array for the page heading, and the array for the line by line print format.

HFMT

Byte	Condition	Contents
1		(
2 - 13	p = 0	
	p = 1	2X,3HPHI,1X,
	p = 2	5X,3HPHI,4X,
	p = 3	8X,3HPHI,7X
14 - 18	q = 0	
	q ≠ 0	2H ,,
19 - 32	q = 0	
	q = 1	1X,5HTHETA,1X,
	q = 2	4X,5HTHETA,3X,
	q = 3	7X,5HTHETA,6X,

HFMT Continued:

Byte	Condition	Contents
33 - 45		3X,7HERR VAR,
46	icode ≠ 3	3
	icode = 3	
	mdim = 1	1
	mdim = 2	2
	mdim = 3	3
	mdim = 4	4
	mdim = 5	5
47 - 74	icode ≠ 3	X,5HLEVEL,3X,6HT STAT,1X,7HL
	icode = 3	(3X,2HB(,I1,1H),4X
		4X 6HT STAT
75 - 95	icode ≠ 3	EV CHG,3X,6HTSTAT
	icode = 3	
96 -143	icode = 2	,3X,5HDRIFT,3X,6HT STAT,1X,7HDFT CHG,3X,6HT STAT
	icode ≠ 2	
144		)

FMT

Byte	Condition	Contents
1		(
2	p = 0	
	p = 1	1
	p = 2	2
	p = 3	3
3 - 12	p = 0	
	p ≠ 0	(1X,F5.2),
13 - 17	p = 0, q = 0	
	p ≠ 0, and q ≠ 02H	,,
18 ≠	q = 0	
	q = 1	1
	q = 2	2
	q = 3	3
19 - 28	q = 0	
	q ≠ 0	(1X,F5.2),
29 - 55		1X,F9.3,5(1X,F7.2,1X,F8.2))

SAMPLE PROBLEM FOR TIME SERIES (PART 2)

N1 = 60, N2 = 60, DEGREES OF FREEDOM FOR THIS DESIGN = 118

MODEL IDENTIFICATION
    ORDER OF AUTOREGRESSION IS 0
    ORDER OF DIFFERENCING IS 1
    ORDER OF MOVING AVERAGE IS 1

INPUT DATA FORMAT (10F3.1)

PRE-INTERVENTION DATA

55.00000	56.00000	48.00000	46.00000	56.00000	46.00000	59.00000	60.00000	53.00000	58.00000
73.00000	69.00000	72.00000	53.00000	72.00000	69.00000	68.00000	69.00000	79.00000	77.00000
53.00000	63.00000	80.00000	65.00000	78.00000	64.00000	72.00000	77.00000	82.00000	77.00000
35.00000	79.00000	71.00000	73.00000	77.00000	76.00000	83.00000	73.00000	74.00000	91.00000
70.00000	88.00000	88.00000	85.00000	77.00000	63.00000	91.00000	94.00000	72.00000	63.00000
88.00000	78.00000	84.00000	78.00000	75.00000	75.00000	86.00000	73.00000	75.00000	87.00000

POST-INTERVENTION DATA

66.00000	73.00000	62.00000	27.00000	53.00000	47.00000	65.00000	59.00000	77.00000	47.00000
51.00000	47.00000	49.00000	58.00000	58.00000	56.00000	50.00000	54.00000	45.00000	66.00000
39.00000	51.00000	39.00000	27.00000	39.00000	37.00000	43.00000	41.00000	27.00000	24.00000
27.00000	36.00000	24.00000	31.00000	26.00000	38.00000	37.00000	26.00000	31.00000	45.00000
38.00000	32.00000	33.00000	25.00000	24.00000	29.00000	37.00000	35.00000	32.00000	31.00000
28.00000	40.00000	31.00000	37.00000	34.00000	43.00000	38.00000	33.00000	28.00000	35.00000

SAMPLE PROBLEM FOR TIME SERIES (PART 2)

THETA	ERR VAR	LEVEL	T STAT	LEV CHG	T STAT
-0.98	1246.680	99.12	11.21	28.26	3.62
-0.96	1227.682	96.14	9.76	37.63	3.40
-0.94	1139.651	91.34	7.93	36.26	3.15
-0.92	1049.577	85.89	6.76	35.51	2.09
-0.90	963.789	80.52	5.95	32.32	2.39
-0.88	884.667	75.71	5.36	27.70	1.96
-0.86	813.143	71.63	4.92	22.40	1.54
-0.84	749.128	68.24	4.60	16.94	1.14
-0.82	693.152	65.47	4.35	11.61	0.77
-0.80	641.572	63.22	4.16	6.59	0.43
-0.78	596.688	61.39	4.02	1.99	0.11
-0.76	556.822	59.90	3.91	-2.18	-0.14
-0.74	521.324	58.69	3.82	-5.90	-0.38
-0.72	489.648	57.71	3.76	-9.21	-0.60
-0.70	461.282	56.90	3.71	-12.11	-0.79
-0.68	435.794	56.23	3.67	-14.65	-0.96
-0.66	412.830	55.69	3.65	-16.86	-1.10
-0.64	392.007	55.25	3.63	-18.78	-1.23
-0.62	373.109	54.88	3.62	-20.43	-1.35
-0.60	355.980	54.59	3.62	-21.83	-1.45
-0.58	340.120	54.36	3.62	-23.01	-1.53
-0.56	325.656	54.18	3.62	-24.03	-1.61
-0.54	312.342	54.04	3.63	-24.87	-1.67
-0.52	300.047	53.94	3.65	-25.55	-1.72
-0.50	288.668	53.87	3.66	-26.09	-1.77
-0.48	278.103	53.82	3.68	-26.52	-1.81
-0.46	268.277	53.80	3.70	-26.83	-1.84
-0.44	259.114	53.80	3.72	-27.05	-1.87
-0.42	250.553	53.82	3.75	-27.18	-1.89
-0.40	242.540	53.85	3.77	-27.23	-1.91
-0.38	235.025	53.90	3.80	-27.21	-1.92
-0.36	227.965	53.95	3.83	-27.14	-1.93
-0.34	221.322	54.01	3.86	-27.00	-1.93
-0.32	215.063	54.08	3.89	-26.82	-1.93
-0.30	209.155	54.15	3.92	-26.60	-1.93
-0.28	203.573	54.22	3.96	-26.33	-1.92
-0.26	198.290	54.30	3.99	-26.04	-1.92
-0.24	193.285	54.37	4.03	-25.72	-1.91
-0.22	188.536	54.44	4.06	-25.37	-1.89
-0.20	184.027	54.52	4.10	-25.00	-1.88
-0.18	179.738	54.59	4.14	-24.62	-1.87
-0.16	175.657	54.65	4.18	-24.23	-1.85
-0.14	171.767	54.71	4.22	-23.82	-1.84
-0.12	168.057	54.77	4.26	-23.42	-1.82
-0.10	164.515	54.82	4.30	-23.00	-1.80
-0.08	161.129	54.87	4.34	-22.59	-1.79
-0.06	157.891	54.91	4.38	-22.19	-1.77
-0.04	154.791	54.95	4.42	-21.78	-1.75
-0.02	151.820	54.98	4.46	-21.39	-1.74
0.0	148.975	55.00	4.51	-21.00	-1.72

SAMPLE PROBLEM FOR TIME SERIES (PART 2)

THETA	ERR VAR	LEVEL	T STAT	LEV CHG	T STAT
0.02	146.239	55.02	4.55	-20.61	-1.71
0.04	143.614	55.03	4.60	-20.27	-1.69
0.06	141.041	55.03	4.64	-19.92	-1.68
0.08	138.666	55.03	4.69	-19.60	-1.67
0.10	136.322	55.02	4.74	-19.29	-1.66
0.12	134.085	55.00	4.79	-19.00	-1.65
0.14	131.921	54.98	4.83	-18.74	-1.65
0.16	129.835	54.95	4.89	-18.49	-1.64
0.18	127.823	54.92	4.94	-18.27	-1.64
0.20	125.903	54.88	4.99	-18.08	-1.64
0.22	124.004	54.83	5.05	-17.91	-1.65
0.24	122.201	54.78	5.10	-17.77	-1.66
0.26	120.453	54.72	5.16	-17.65	-1.67
0.28	118.764	54.66	5.22	-17.56	-1.68
0.30	117.130	54.59	5.29	-17.50	-1.69
0.32	115.551	54.52	5.35	-17.46	-1.71
0.34	114.022	54.44	5.42	-17.45	-1.74
0.36	112.542	54.37	5.49	-17.46	-1.76
0.38	111.109	54.28	5.57	-17.50	-1.80
0.40	109.722	54.20	5.65	-17.57	-1.83
0.42	108.378	54.11	5.73	-17.66	-1.87
0.44	107.076	54.03	5.81	-17.77	-1.91
0.46	105.816	53.94	5.91	-17.91	-1.96
0.48	104.597	53.85	6.00	-18.06	-2.01
0.50	103.418	53.77	6.11	-18.24	-2.07
0.52	102.280	53.69	6.21	-18.43	-2.13
0.54	101.182	53.61	6.33	-18.65	-2.20
0.56	100.127	53.55	6.46	-18.87	-2.28
0.58	99.117	53.49	6.60	-19.12	-2.36
0.60	98.155	53.44	6.74	-19.37	-2.44
0.62	97.246	53.42	6.90	-19.64	-2.54
0.64	96.346	53.41	7.08	-19.92	-2.64
0.66	95.613	53.42	7.27	-20.21	-2.75
0.68	94.908	53.46	7.48	-20.51	-2.87
0.70	94.296	53.54	7.72	-20.82	-3.00
0.72	93.797	53.66	7.98	-21.14	-3.15
0.74	93.436	53.83	8.28	-21.48	-3.30
0.76	93.250	54.06	8.61	-21.83	-3.48
0.78	93.288	54.37	8.99	-22.22	-3.68
0.80	93.617	54.76	9.43	-22.61	-3.90
0.82	94.236	55.26	9.94	-23.04	-4.15
0.84	95.592	55.90	10.54	-23.62	-4.43
0.86	97.605	56.71	11.25	-24.24	-4.81
0.88	100.723	57.74	12.11	-24.98	-5.24
0.90	105.505	59.08	13.19	-25.87	-5.78
0.92	112.855	60.84	14.61	-26.92	-6.47
0.94	124.113	63.22	16.63	-28.13	-7.40
0.96	140.447	66.39	19.43	-29.34	-8.78
0.98	159.024	69.90	26.59	-30.30	-11.05

MINIMUM ERROR VARIANCE OF 93.2503 FOUND AT 1 VALUE(S) OF PHI AND/OR THETA BELOW

THETA = 0.76

APPENDIX C

TIME01 PROGRAM

UNIVERSITY OF ALBERTA

Division of Educational Research Services

Computer Program Documentation

TITLE: TIME SERIES ANALYSIS (PART 1) (GLASS)*
MACHINE: IBM 360/67
LANGUAGE: FORTRAN IV
PROGRAM TYPE: Complete
SUBPROGRAMS: (XDER:SUB) TITLE
LIMITS: Total of pre-intervention plus post intervention points must be less than 300.
TIME: Time depends upon options selected.
PROGRAMMER: C.P. Bower, W.L. Padia, G.V. Glass and modified by T.C. Montgomerie
DOCUMENTED: T.C. Montgomerie

Description:

TIME01 and TIME02 are a pair of programs required for carrying out a time series analysis. Given a series of measures prior to an intervention point, and a series of measures after the intervention point, the programs allow an analysis to be made of the effects of the intervention. TIME01 permits the identification of the appropriate model to be used as well as its parameters. The parameters from TIME01 become input to TIME02 for a 'least squares normal theory analysis'.

References

Bower, C.P. Padia, W.L. & Glass, G.V., TMS: Two Fortran IV Programs for analysis of Time-Series Experiments. Laboratory of Educational Research, University of Colorado, Boulder, Col., October 1974.

NOTE: In the above reference CORREL should be read as TIME01, and TSX should be read as TIME02 in order that the reader may relate the programs described in the manual to those named in the DERS library.

---

*This program was obtained from the Laboratory of Educational Research at the University of Colorado. It was originally written for the CDC 6400 by C.P. Bower, W. Padia, and G.V. Glass.

Preparation of Header Cards for TIME01

Card Seq.	See Notes*	Card Type	Cols.	Description
1		Title Card	1-80	Any description of the problem
2		Parameter	1- 5	Number of pre-intervention points
			6-10	Number of post-intervention
	1		15	Blank for correlogram output of raw data 1 gives correlogram for raw & log transformed data 2 gives correlogram for log transformed data only (If 1 or 2, card output is made)
			16-20	Constant to be added to observation points prior to log transformation. (Use F format); else leave blank.
	2		21-25	Length of seasonal cycle, (I format); else leave blank.
3		Format	1-80	F format for data
Next		Data cards		As described in Card Type 3. Data points must be less than 1000 in absolute value.
Next		End of job		Blank card if no further jobs to be run; else this should be the Title card of the next job.

USER NOTES*

1) If option 1 or 2 is given be sure to specify the number of cards expected on the $SIGNON card.

2) The original reference notes the following: "For non-seasonal data, the lag k auto-correlations are computed for differences of order zero through four; for seasonal (cyclic) data, the lag k autocorrelations are computed for seasonal differences of order one and two and within each seasonal difference for order zero, one, and two. In both the seasonal and non-seasonal case, k lag auto correlations are calculated for each difference where k is half the number of differenced or non-differenced data points."

Manipulation of Data Prior to Processing

No DATRAN subroutine is available in these programs.

Data Output for TIME01

The following data is output by the program:

1. Summary of parameter card and format card(s).
2. The pre-intervention observations.
3. The post-intervention observations.
4. The autocorrelations and partial autocorrelations for the pre-intervention observations.
5. The autocorrelations and partial autocorrelations for the post-intervention observations.
6. The log-transformed pre-intervention observations.
7. The log-transformed post-intervention observations.
8. The autocorrelations and partial autocorrelations for the log-transformed pre-intervention observations.
9. The autocorrelations and partial auto correlations for the log-transformed post-intervention observations.
10. Card output of the log-transformed pre-intervention and post-intervention observations. If the log-transformation of input observations is selected, the log-transformed data will be punched on cards in the format (10F8.4).

(Note - the output of items 2-10 above are optional, depending on the specifications of the parameter card.)

Example of Input Data for TIME01

```
$SIGNON CSID                                     (SIGNON CARD)
PSWORD                                           (PASSWORD CARD)
$RUN XDEP:PROGRAM PAR=TIME01                     (RUN CARD)
SAMPLE PROBLEM FOR TIME SERIES (PART 1)          (TITLE CARD)
   60   60                                       (PARAMETER CARD)
(10F3.1)                                         (FORMAT CARD)
55 56 48 46 56 46 59 60 53 58                    (DATA CARDS)
73 69 72 59 72 69 68 69 79 77
53 63 80 65 78 64 72 77 82 77
35 79 73 73 77 76 83 73 78 91
70 88 98 85 77 63 91 94 72 83
88 79 84 78 75 75 86 79 75 97
66 73 62 27 52 47 65 59 77 47
51 47 49 54 58 56 50 54 45 66
39 51 39 27 39 37 43 41 27 29
27 26 29 31 29 38 37 26 31 45
38 33 33 25 24 29 37 35 32 31
28 40 31 37 34 43 38 33 28 35
                                                 (BLANK CARD)
$SIGNOFF                                         (SIGNOFF CARD)
```

Example of Output for TIME01

See pages 5 & 6.

Error Messages for TIME01

The error messages for TIME01 are self explanatory.

Technical Notes

The computational formulae used in TIME01 and TIME02 can be found in Appendix A of Bower, Padia, & Glass.

```
INPUT PARAMETERS

    N1 = 60,  N2 = 60

    CORRELOGRAM OPTION =  0   (RAW DATA)                                  SAMPLE PROFILES FOR TIME SERIES (PART.1)

    LENGTH OF SEASONAL CYCLE =     0

INPUT DATA FORMAT      (10F3.1)


PRE-INTERVENTION OBSERVATIONS      N1= 60
   55.0000    56.0000    49.0000    46.0000    56.0000    86.0000    59.0000    60.0000    53.0000    58.0000
   73.0000    63.0000    73.0000    51.0000    72.0000    69.0000    68.0000    69.0000    79.0000    77.0000
   53.0000    63.0000    82.0000    65.0000    78.0000    64.0000    72.0000    77.0000    82.0000    77.0000
   35.0000    74.0000    71.0000    73.0000    77.0000    76.0000    83.0000    73.0000    74.0000    91.0000
   70.0000    88.0000    88.0000    85.0000    77.0000    63.0000    91.0000    94.0000    72.0000    83.0000
   88.0000    78.0000    84.0000    78.0000    75.0000    75.0000    86.0000    79.0000    75.0000    87.0000


POST-INTERVENTION OBSERVATIONS      N2= 60
   66.0000    73.0000    62.0000    27.0000    52.0000    47.0000    65.0000    59.0000    77.0000    47.0000
   51.0000    47.0000    49.0000    58.0000    58.0000    56.0000    50.0000    54.0000    45.0000    66.0000
   39.0000    51.0000    39.0000    27.0000    39.0000    37.0000    43.0000    41.0000    27.0000    29.0000
   27.0000    26.0000    29.0000    31.0000    28.0000    38.0000    37.0000    26.0000    31.0000    45.0000
   38.0000    33.0000    33.0000    25.0000    28.0000    29.0000    37.0000    35.0000    32.0000    31.0000
   28.0000    40.0000    31.0000    37.0000    34.0000    43.0000    38.0000    33.0000    28.0000    35.0000


AUTOCORRELATIONS FOR PRE-INTERVENTION DATA


THE DATA IS OF DIFFERENCE ORDER  0  FOR LAGS 1 TO 30
MEAN=    71.6500      VARIANCE=    161.0260
 R=     0.4275       0.3064       0.4712       0.4040       0.4276       0.2388       0.3470       0.2675
     0.1661       0.3236       0.1085       0.1547       0.1497       0.0905       0.1802      -0.0001
     0.1203       0.0930       0.0060       0.0394       0.0084      -0.0177      -0.0157       0.0118
     0.0646      -0.1779      -0.0515       0.0125      -0.1090      -0.0787
STANDARD ERROR
        0.1291       0.1509       0.1643       0.1874       0.2014       0.2160       0.2204       0.2293
     0.2345       0.2368       0.2437       0.2445       0.2461       0.2476       0.2492       0.2503
     0.2503       0.2514       0.2520       0.2520       0.2521       0.2521       0.2521       0.2521
     0.2521       0.2524       0.2545       0.2547       0.2547       0.2554
CHI-SQUARE STATISTIC =    43.73
     DEGREES OF FREEDOM ARE 30 -1(IF MEAN INCLUDED, 0 OTHERWISE) - P - Q - P(SEASONAL) - Q(SEASONAL)


THE DATA IS OF DIFFERENCE ORDER  1  FOR LAGS 1 TO 29
MEAN=      0.5424      VARIANCE=    170.5103
 R=    -0.4695      -0.1355       0.1238      -0.0525       0.1773      -0.2574       0.1636       0.0162
    -0.2460       0.3720      -0.2308       0.0181       0.0186      -0.0626       0.2237      -0.2977
     0.1087       0.0567      -0.0089      -0.0119       0.0288       0.0081      -0.0628       0.0062
     0.2463      -0.3297       0.0677       0.1656      -0.0815
STANDARD ERROR
        0.1302       0.1553       0.1582       0.1599       0.1602       0.1635       0.1702       0.1728
     0.1729       0.1737       0.1914       0.1960       0.1961       0.1961       0.1964       0.2006
     0.2090       0.2099       0.2100       0.2106       0.2106       0.2107       0.2107       0.2110
     0.2110       0.2158       0.2238       0.2241       0.2262
CHI-SQUARE STATISTIC =    30.27
     DEGREES OF FREEDOM ARE 29 -1(IF MEAN INCLUDED, 0 OTHERWISE) - P - Q - P(SEASONAL) - Q(SEASONAL)
```

PARTIAL AUTOCORRELATIONS

```
DIFFERENCE IS ORDER    0
  STANDARD ERROR=          0.1336
  PHI(1,1)=        0.4275
  PHI(2,2)=        0.2492
  PHI(3,3)=        0.2502
  PHI(4,4)=        0.2412
  PHI(5,5)=        0.4165
  PHI(6,6)=        0.1949

DIFFERENCE IS ORDER    1
  STANDARD ERROR=          0.1336
  PHI(1,1)=       -0.4695
  PHI(2,2)=       -0.4565
  PHI(3,3)=       -0.3309
  PHI(4,4)=        0.0085
  PHI(5,5)=        0.1606
  PHI(6,6)=       -0.2491


 AUTOCORRELATIONS FOR POST-INTERVENTION DATA

THE DATA IS OF DIFFERENCE ORDER  0  FOR LAGS 1 TO 30
MEAN=     40.9933       VARIANCE=     166.5479
                                                                   0.4551          0.5583          0.5150          0.4239
                                                 0.4519
 R=      0.6152        0.5385         0.3702                     0.2786          0.1806          0.1219          0.0142
     0.3185        0.2377         0.2790         0.2180        -0.0707         -0.1742         -0.2222         -0.1918
     0.0403        0.0169         0.0218         0.0121        -0.2497         -0.2680
    -0.1366       -0.1685        -0.1790        -0.2278
STANDARD ERROR
                                                                   0.2243          0.2392          0.2600          0.2765
         0.1291        0.1711         0.1978         0.2086       0.3029          0.3070          0.3088          0.3096
     0.2871        0.2930         0.2962         0.3003         0.3097          0.3100          0.3116          0.3143
     0.3096        0.3097         0.3097         0.3097         0.3230          0.3262
     0.3162        0.3172         0.3187         0.3203
CHI-SQUARE STATISTIC =   80.76
       DEGREES OF FREEDOM ARE 30 -1(IF MEAN INCLUDED, 0 OTHERWISE) - P - Q - P(SEASONAL) - Q(SEASONAL)

THE DATA IS OF DIFFERENCE ORDER  1  FOR LAGS 1 TO 29
MEAN=     -0.5254       VARIANCE=     118.8577
                                                                  -0.1334          0.1888          0.0819         -0.0393
                                                  0.0260
 R=      -0.4734        0.1719        -0.2265                     0.1967         -0.0740          0.0696         -0.1709
     0.0601       -0.1763         0.1366        -0.1624        -0.0274         -0.0426         -0.0600         -0.0773
     0.0621       -0.0116        -0.0526         0.2066        -0.0100
     0.1207       -0.0514         0.0579         0.0067
STANDARD ERROR
                                                  0.1652          0.1653          0.1671          0.1692          0.1699
         0.1302        0.1567         0.1598                     0.1778          0.1814          0.1819          0.1824
     0.1700        0.1704         0.1734         0.1752         0.1896          0.1896          0.1898          0.1901
     0.1851        0.1854         0.1855         0.1857         0.1925
     0.1906        0.1919         0.1922         0.1925
CHI-SQUARE STATISTIC =    17.78
       DEGREES OF FREEDOM ARE 29 -1(IF MEAN INCLUDED, 0 OTHERWISE) - P - Q - P(SEASONAL) - Q(SEASONAL)

                                   PARTIAL AUTOCORRELATIONS

DIFFERENCE IS ORDER    0
  STANDARD ERROR=        0.1336
  PHI(1,1)=        0.6152
  PHI(2,2)=        0.2575
  PHI(3,3)=       -0.0277
  PHI(4,4)=        0.1857
  PHI(5,5)=        0.5128
  PHI(6,6)=        0.5256

DIFFERENCE IS ORDER    1
  STANDARD ERROR=        0.1336
  PHI(1,1)=       -0.4734
  PHI(2,2)=       -0.3674
  PHI(3,3)=       -0.1468
  PHI(4,4)=        0.0116
  PHI(5,5)=       -0.1524
  PHI(6,6)=        0.1508


 [illegible] NORMALLY TERMINATED WITH A BLANK CARD ON AUG 12, 1975

13:28.50 1.942 [illegible]=0
```