1-1-1985

The assessment of intervention effects in time series processes.

John B. White
University of Massachusetts Amherst

THE ASSESSMENT OF INTERVENTION EFFECTS IN TIME SERIES PROCESSES

A Dissertation Presented

By

JOHN B. WHITE

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1985

Psychology

THE ASSESSMENT OF INTERVENTION EFFECTS
IN TIME SERIES PROCESSES

A Dissertation Presented

By

JOHN B. WHITE

Approved as to style and content by:

Hariharan Swaminathan, Chairperson

Icek Aizen, Member

Michael Sutherland, Member

Arnold D. Well, Member

Seymour M. Berger, Department Head
Psychology

ACKNOWLEDGMENTS

I would like to express my appreciation to each of the members of my dissertation committee. Serving as a committee member is only part of each individual's contribution to my experience as a graduate student.

I am extremely grateful to have had the opportunity to study with H. Swaminathan. As we have worked together, my professional and personal respect for him has continued to grow. His ability to convey complex theoretical issues with clarity has been instrumental in the development of my interest in applied statistics. I look forward to a continuation of our research activities.

Icek Aizen has been an excellent role model as a social scientist. His ability to quickly identify critical research issues has always been impressive and instructive. Mike Sutherland's enthusiasm and interest in the practical issues involved in the application of statistical methods has been refreshing. Arnie Well has always been a pleasure to interact with, and his interest in my academic studies has often provided me with encouragement.

Mary Beth Regan has always provided sound advice, reassurance, and emotional support that helped to alleviate much of the stress involved in completing this dissertation.

ABSTRACT

The Assessment of Intervention Effects
in Time Series Processes
May 1985

John B. White, B.S., Kenyon College
M.A., University of Massachusetts
Ph.D., University of Massachusetts
Directed by: Professor H. Swaminathan

This dissertation presents four Monte Carlo experiments that provide information with respect to the small sample properties of several estimators utilized in time series analysis. Studies one through three investigate procedures that are used in the model identification stage of ARIMA(p,d,q) time series analysis, while study four examines the small sample properties of Box and Tiao's (1965,1975) test statistic for the presence of an intervention effect in an ongoing time series process. All of the studies manipulate two factors; the nature of the autocorrelation structure and the length of the time series realization.

On the basis of the research presented in this

dissertation, it is recommended that time series realizations consist of at least 90 observations. The length of the time series realization plays a critical role in determining the quality of the estimates that are obtained when applying the procedures examined in this investigation. Almost all of the estimation problems that have been investigated - bias, the magnitude of standard errors, the accuracy of estimated standard errors, inflation of Type I error rates, and lack of power - are much less severe for more lengthy time series realizations.

It is also important for researchers to be aware of the severity of the estimation problems that are encountered when the autocorrelation among data points is extremely large. For almost all of the conditions examined in the present research, extreme serial dependence magnifies the problems that are observed in estimation procedures. In the model identification stage of time series analysis, both the bias in the autocorrelation estimator and the over-estimation of the standard error of autocorrelation coefficients becomes more severe as the serial dependence becomes more severe. Furthermore, problems with the estimation of the intervention component become more severe as serial dependence increases; the inflation of the Type I error rate becomes greater and there is a large reduction in the statistical power to detect an intervention effect.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv

ABSTRACT v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

I. INTRODUCTION 1

Introduction 1
Applications of the Interrupted Time Series Design 4

II. TIME SERIES MODELS 9

Introduction 9
Autocorrelation and Autocovariance 10
Stationarity 12
Modeling Time Series Data 26
Autoregressive models 26
Moving average models 33
The duality of MA and AR processes . . 37
ARMA and ARIMA models 39
Model Building 43
Interrupted Time Series Experiments 56
Statement of the Problem 67

III. METHODOLOGY 71

Introduction 71
Study One 72
Study Two 74
Study Three 76
Study Four

IV. RESULTS AND DISCUSSION 88

Study One 88
Study Two 111
Study Three 135
Study Four 140

V. SUMMARY, LIMITATIONS, AND CONCLUSIONS 151

Summary 151
Limitations 154
Conclustions 159

REFERENCES 163

LIST OF TABLES

1. First Order Autoregressive Processes: Mean Estimated Autocorrelation and Empirical Standard Error 89
2. First Order Autoregressive Processes: Mean Estimated Partial Autocorrelation and Empirical Standard Error 98
3. Second Order Autoregressive Processes: Mean Estimated Autocorrelation and Empirical Standard Error 100
4. Second Order Autoregressive Processes: Mean Estimated Partial Autocorrelation and Empirical Standard Error 108
5. First and Second Order Moving Average Processes: Mean Estimated Autocorrelation and Empirical Standard Error 109
6. First and Second Order Moving Average Processes: Mean Estimated Partial Autocorrelation and Empirical Standard Error 110
7. First Order Autoregressive Processes: Mean Estimated Standard Error and Empirical Standard Error 113
8. Second Order Autoregressive Processes: Mean Estimated Standard Error and Empirical Standard Error 117
9. First and Second Order Moving Average Processes: Mean Estimated Standard Error and Empirical Standard Error 122
10. First and Second Order Autoregressive Processes: Percentage of Null Hypotheses Rejected for Three Lags of Estimated Autocorrelations and Partial Autocorrelations 125
11. First and Second Order Moving Average Processes: Percentage of Null Hypotheses Rejected for Three Lags of Estimated Autocorrelations and Partial Autocorrelations 134
12. Unbiased Estimates of Autoregressive Processes: Percentage of Autocorrelation Estimates > 1.0 . . 137
13. First and Second Order Autoregressive Processes: Empirical Standard Error of Unbiased and Biased Autocorrelation Estimates 139
14. Significance Test of Intervention Parameter: Percentage of Null Hypotheses Rejected for Four Magnitudes of Intervention Effects . . 142
15. Mean Estimated Standard Error and Empirical Standard Error of Intervention Effect Estimates 148

LIST OF FIGURES

1. AR1 Processes - Lag 1 Autocorrelations 91
2. AR1 Processes - Lag 2 Autocorrelations 92
3. AR1 Processes - Lag 3 Autocorrelations 93
4. Standard Error of Estimated Autocorrelations - Lag 1 95
5. Standard Error of Estimated Autocorrelations - Lag 2 96
6. Standard Error of Estimated Autocorrelations - Lag 3 97
7. AR2 Processes - Lag 1 Autocorrelations 101
8. AR2 Processes - Lag 2 Autocorrelations 102
9. AR2 Processes - Lag 3 Autocorrelations 103
10. Standard Error of Estimated Autocorrelations - Lag 1 (AR2 Processes) 105
11. Standard Error of Estimated Autocorrelations - Lag 2 (AR2 Processes) 106
12. Standard Error of Estimated Autocorrelations - Lag 3 (AR2 Processes) 107
13. Empirical versus Approximate Estimates: AR1 Processes - Lag 1 Standard Error Estimates . 114
14. Empirical versus Approximate Estimates: AR1 Processes - Lag 2 Standard Error Estimates . 115
15. Empirical versus Approximate Estimates: AR1 Processes - Lag 3 Standard Error Estimates . 116
16. Empirical versus Approximate Estimates: AR2 Processes - Lag 1 Standard Error Estimates . 119
17. Empirical versus Approximate Estimates: AR2 Processes - Lag 2 Standard Error Estimates . 120
18. Empirical versus Approximate Estimates: AR2 Processes - Lag 3 Standard Error Estimates . 121
19. Power of Autocorrelation Test - Lag 1 (AR1) . . . 126
20. Power of Autocorrelation Test - Lag 2 (AR1) . . . 127
21. Power of Autocorrelation Test - Lag 3 (AR1) . . . 128
22. Power of Autocorrelation Test - Lag 1 (AR2) . . . 129
23. Power of Autocorrelation Test - Lag 2 (AR2) . . . 130
24. Power of Autocorrelation Test - Lag 3 (AR2) . . . 131
25. Power of Partial Autocorrelation Test - Lag 2 . . 133
26. Power of Intervention Test - Intervention Effect = .5 144
27. Power of Intervention Test - Intervention Effect = .8 145
28. Power of Intervention Test - Intervention Effect = 1.1 146
29. Intervention Analysis - Empirical and Estimated Standard Errors 150

CHAPTER I

INTRODUCTION

Introduction

In recent years, social scientists have more frequently been using data analysis procedures that are based on statistical models of time series processes. In particular, the interrupted time series design has been widely utilized as a research paradigm for social scientists affiliated with a wide spectrum of disciplines. This quasi-experimental design (Campbell, 1963; Campbell and Stanley, 1963, 1966), offers several potential benefits as an alternative to traditional experimental paradigms. The design was originally introduced by Campbell as a technique for assessing the impact of an intervention (e.g., a change in social policy) on some ongoing social process. This dissertation will discuss the appropriate procedures for the analysis of data from interrupted time series experiments, and point out potential difficulties that may be encountered in applying these procedures to "real life" data sets.

In the quasi-experimental time series design, a series of measures are 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 times series with the post-intervention time series. The research design is conceptually simple, and researchers may be tempted to use an ordinary t-test to compare the mean of measures collected before the intervention with the mean of those collected after the intervention. The results of such a procedure are extremely unreliable, however, due to the non-independence of the measures that are assessed over time. More sophisticated statistical models, which take the interdependence of measures into account, are necessary 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.

There are several sources of information about interrupted time series designs that are intended for an audience of social scientists. The most comprehensive discussion of the interrupted time series design is provided by Glass, Willson, and Gottman in Design and Analysis of Time Series Experiments (1975). This book thoroughly considers basic quasi-experimental designs, and in particular the statistical procedures that are commonly used in the analysis of data collected from interrupted time series experiments. A brief introduction to the most prevalent data analysis procedures is presented by McDowall, McCleary, Meidinger and Hay (1980). Gottman (1981) provides a brief discussion of some alternative data analysis procedures, which he suggests may often be preferable to more widely used techniques. Finally, McCleary, Hay, Meidinger and McDowall (1980) discuss the interrupted time series design, as well as other time series procedures that are utilized by social scientists.

Applications of the Interrupted Time Series Design

The interrupted time series design is appealing to social scientists for a number of reasons. Many topics of study would be virtually impossible to investigate within the structure of traditional experimental designs. Time series quasi-experiments, on the other hand, often allow researchers to meaningfully interpret data that are collected in the absence of the rigorous control over variables that is necessary in traditional experiments. Time series experiments also permit hypothesis testing of treatment effects in studies involving only a single subject or unit of observation. Finally, and perhaps most importantly, the interrupted time series experiment provides information concerning the nature of the intervention effect over a period of time. This advantage may be of particular importance in many areas of social science research. The impact of an intervention on human behavior is likely to be extremely complex, and probably not consistent over time. It is often of interest to evaluate the immediacy, duration, and pattern over time of the intervention effect by examining the post-intervention data.

One type of research that has greatly benefited from the interrupted time series design involves testing post hoc hypotheses using archival data. The researcher can generate causal hypotheses concerning the effect of historical events on some variable of interest, and test these hypotheses using the standard interrupted time series data analysis procedures. Some examples of this type of quasi-experiment include the impact of new traffic laws (Campbell and Ross, 1968; Glass, 1968), the effect of air pollution control laws (Box and Tiao, 1975), and the impact of gun control laws (Deutsch and Alt, 1977; Hay and McCleary, 1979; Zimring, 1975). Extreme caution must be exercised in drawing causal inferences on the basis of archival data, however. Interventions are likely to be accompanied by other events that may also influence the variable that is being studied, and thus, viable alternative explanations for an intervention effect will generally be present. Whenever possible, replications of the study under different conditions and/or planned experiments should be conducted to lend greater credence to the veracity of the causal inference.

A second situation in which the interrupted time series design may be useful to social scientists occurs when the feasibility of comparison groups is questionable. There are situations in which it is very difficult to expose one group of subjects to a treatment, while simultaneously observing a second comparable group of subjects. This type of situation often arises in educational or societal settings, where entire populations are affected by an intervention. Comparisons with a separate population from a different school or geographical region may be meaningless, since there may be substantial underlying discrepancies between the populations. In other situations, it is sometimes unethical to withhold a beneficial intervention from a sample of people in order to scientifically examine the effect of a treatment. Under the circumstances discussed above, the impact of an intervention is best evaluated using an interrupted times series paradigm.

A similar circumstance involves the desirability of single subject experimental designs. For a variety of reasons, researchers often prefer to investigate treatment effects using a single experimental unit. One area of research that relies heavily on the use of single subject designs is the field of behavioral psychology. Experiments generally involve an operant conditioning procedure that is administered to a single person or animal. Time series designs allow the researcher to evaluate the impact of the conditioning procedure on an individual unit.

Finally, the interrupted time series design provides longitudinal information about the impact of the intervention. Conventional experimental designs generally assess the impact of a treatment at a single time point after the intervention has occurred. Glass et. al. (1975) suggest that the most valuable asset of time series experiments is the capability of examining an intervention impact over a period of time.

> The most important advantage of the time series design is not that it offers an alternative when a traditional, randomized, comparative experimental design is not feasible, but it offers a unique perspective on the evaluation of intervention (or "treatment") effects. Simultaneous comparative designs in the Fisherian tradition may blind the experimenter to important observations when such designs become a thoughtless habit of mind. The Fisherian design which has so captured the attention of social and behavioral scientists was originally developed for use in evaluating agricultural field trials. The methodology was appropriate to comparing two or more agricultural methods with respect to their relative yields. The yields were crops which were harvested when they were ripe; it was irrelevant in this application whether the crops grew slowly or rapidly or whether they rotted six months after harvest. For social systems, there are no planting and harvest times... The value of an intervention is properly judged not by whether the effect is observable at the fall harvest, but by whether the effect occurs immediately or is delayed, whether it increases or decays, whether it is only temporarily or constantly superior to the effects of alternative interventions. The time series design provides a methodology appropriate to the complexity of the effects of interventions into social organizations or with human beings (pp. 4-5).

It is clear that the interrupted time series design is a useful tool for social scientists interested in a wide variety of research areas. The statistical procedures that have evolved to analyze time series data have made possible the investigation of new topic areas, and have provided a unique perspective for the study of traditional fields of research. It is important that those who conduct research that may benefit from these techniques thoroughly understand both the research opportunities that are afforded by the availability of these procedures, and the limitations and drawbacks of these methods. As with any statistical procedure, those who wisely apply the method to research problems will benefit greatly from the information generated, while those who are less prudent in their applications will often be misled to erroneous conclusions.

The discussion in the following chapters deals with the basic underlying statistical procedures that are necessary to model time series processes and to test for the effect of interventions. More importantly, some of the potential limitations and problems encountered in the application of time series analysis will be discussed. The extent to which some of these potential problems may adversely affect statistical inferences is investigated empirically via computer simulations.

CHAPTER II

TIME SERIES MODELS

Introduction

The analysis of time series data requires different considerations than are generally encountered in more traditional data analysis procedures. The distinguishing aspect of the structure of time series data is the non-independence of observations. Most statistical models are based on the premise that observations are independent, or uncorrelated, with other observations. This basic assumption is seldom fulfilled for data that is collected on the same experimental unit across time, however. Instead, observations are likely to be related to other observations collected in close temporal proximity and relatively independent from more distant observations.

The most common method for resolving this problem of serial dependence is to empirically model the autocorrelation of the measures, and then test for the presence of an intervention effect while controlling for the autocorrelation. The most widely used time series model is the Autoregressive Integrated Moving Average (ARIMA) model, which was developed primarily by Box and Jenkins

(1970). The problem of model identification is general to all time series analyses based on ARIMA models (e.g. forecasting and the use of "lead indicators"). Thus, the statistical modeling of serial dependence in quasi-experimental time series data is a preliminary step for hypothesis testing of an intervention effect. The model identification process is of critical importance in the analysis of interrupted time series experiments, and the actual test of the basic hypothesis is relatively straightforward if the time series model is properly identified. Unfortunately, difficulties are often encountered when attempting to model the dependency of "real life" data sets.

Autocorrelation and Autocovariance

The autocorrelation of a time series process is defined as the correlation between all pairs of observations that are separated by a fixed number of points in the time series. Suppose that an individual's overall mood is measured on a daily basis over some period of time. The estimated correlations between the subject's mood on day 1 vs. day 2, day 2 vs. day 3 ... through day t vs. day t+1 can be computed as an ordinary Pearson product moment correlation coefficient. This first-order autocorrelation coefficient is an indication of how well an individual's overall daily mood can be predicted on the basis of the subject's mood on the previous day. Similarly, the second-order autocorrelation coefficient (or autocorrelation at a lag of 2) can be computed by correlating observations on day t vs. day t+2.

The estimate of the first-order autocorrelation is calculated as:

$$r_1 = \{ \sum_{t=1}^{n-1} (X_t - \overline{X})(X_{t+1} - \overline{X})\} / \{ \sum_{t=1}^{n} (X_t - \overline{X})^2\}$$

where X_t is the observation on day t, X_{t+1} represents the observation on day t+1, and $\overline{X}$ is the mean of all observations. This estimator is based on the assumption of stationarity, which is discussed below. The formula for the second order autocorrelation coefficient is of the same form, with X_{t+1} simply replaced by X_{t+2}.

The estimate of the autocovariance at lag 1 of a time series is defined as

$$\text{cov}(X_t, X_{t+1}) = \{ \sum_{t=1}^{n-1} (X_t - \overline{X})(X_{t+1} - \overline{X})\} / n \quad .$$

As with the estimate for the second-order autocorrelation, the autocovariance at lag 2 is estimated by substituting X_{t+2} for X_{t+1} in the formula presented above.

Stationarity

When estimating a statistical model of a time series process that may have generated an observed data set, it is necessary to make certain assumptions about the nature of the underlying process. A set of assumptions that is fundamental to most time series models is often referred to as the conditions of stationarity. These conditions are based on the assumption that specific characteristics of the underlying time series process remain stable over time. It is easy to see, on an intuitive level, why certain aspects of a time series process must remain stable over time if a statistical model of the process is to be fitted. A finite set of parameters must be estimated to determine the model that is believed to have generated an observed realization of the time series process. If crucial aspects of the process were not consistent across time, it would be impossible to apply the same model to different portions of the time series. For example, forecasting future points in time would not be possible if an observed trend was not consistent across the time period.

The first condition of stationarity is that the mean and variance of a time series process do not change with historical time. In other words, a stationary time series process will oscillate around a constant level with uniform variability over time. The second condition of stationarity is that the autocovariance of a time series process is independent of historical time. Thus the covariance of two time points is completely determined by the relative lag of the time points, irrespective of the portion of the time series that is under consideration. Therefore, the accuracy with which time point 2 can be predicted from time point 1 is equal to the predictability of point 48 from point 47 (or any other pair of adjacent time points).

Time series data must conform to the conditions of stationarity before the researcher can properly model the time series process. In actual practice, however, data sets often do not conform to these requirements. Instead, a realization of a time series process is likely to exhibit one or more of the following characteristics: 1) a change in the level of the series over time, 2) periodicity, 3) nonconstant variance, or 4) a shift in the autocovariance structure of the time series. The two most common forms of nonstationarity are the presence of "trend" in the series and a tendency for the series to display indications of periodicity. Methods of analyzing data sets that exhibit these types of nonstationarity have been developed, and generally appear to provide an adequate means for modeling time series processes with these tendencies.

Nonstationarity involving a shift in the variance of the observations or a change in the autocovariance structure of the time series are less common and may present greater difficulties for the researcher.

Despite the fundamental nature of the concept of stationarity, there appears to be no precise method for determining the stationarity of a time series process, and thus, the researcher must exercise caution when inspecting data for indications of stationarity. 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. The specification of these models and the transformation of the data can greatly affect the conclusions drawn on the basis of statistical analyses, so once again, caution is demanded of the researcher.

Several characteristics of the observed time series should be considered when examining data for stationarity. Some of the tools that may be useful may be useful in deciding whether the data are stationary include: a) a plot of the time series data, b) a correlogram of the data, c) tables of means, variances, and autocorrelations for different segments of the observed time series, and d) the spectral density function of the time series data. A careful consideration of this information concerning the underlying structure of the data set is necessary to assess the validity of assuming that the conditions of stationarity are fulfilled. In addition, these methods are useful in determining which type of transformation or model may be effective in removing nonstationary aspects of the data set.

The first step in studying stationarity is to visually inspect a plot of the time series realization. The plot of the observed data points over time will often reveal evidence that is pertinent to the stationarity conditions. It is often possible to visually detect changes in the level of the time series, or in the variability of the data points around that level. Periodic trends in the data (e.g., a seasonal component) may also become evident when examining the plot of time series data. In short, the careful examination of the data points plotted over time may alert the researcher to possible violations of the assumption of stationarity.

The correlogram is simply a plot of the autocorrelation coefficients at each lag as a function of the lag. In general, the autocorrelations of a stationary time series process will approach zero after a relatively small number of lags. It should be emphasized that not all stationary time series processes conform to this pattern of autocorrelations; however, most stationary time series data that is encountered in practical applications will exhibit this tendency. The interdependence of data generated by a stationary time series process can usually be explained in terms of a small number of lags, and thus, the autocorrelation at relatively large lags is essentially zero. In contrast, nonstationary data (especially data in which the level is not stationary) will generally result in autocorrelations that approach zero very slowly as the number of lags increases. It is easy to see why this would be the case for data that follows a linear trend over time, such as a gradual increase in the level of the time series. In such instances, data points will be correlated to some extent with observations that are separated by several points in time. In other words, data points that are relatively distant will be somewhat useful in predicting the location of future observations.

The correlogram is also helpful in detecting cyclic components of a time series process. Periodicity in the time series may be considered to be either a deterministic or nondeterministic process. A deterministic process implies that future time points are completely determined by past observations, whereas a nondeterministic (or stochastic) process indicates that observations are only partially determined by previous occurrences. The implication of a deterministic periodic component is that the predictive accuracy of previous points in time does not diminish as the distance between observations increases. In contrast, the predictive accuracy of a nondeterministic periodic component will attenuate as the amount of time between data points increases. The correlogram provides indications of both deterministic and nondeterministic cycles in time series processes. For example, cyclic components that are based on seasonal variation corresponding to measurements obtained on a monthly basis will often be indicated by a large autocorrelation coefficient at lag 12. This type of periodicity indicates that observations from the corresponding month of the previous year are useful in predicting the current observation. If the cycle is of a deterministic nature, the autocorrelation coefficient at lag 24 and at lag 36 will be of the same magnitude as the autocorrelation at lag 12. A nondeterministic cycle, on the other hand, will display autocorrelations that tend to decrease with each cycle.

It is also useful to divide the data set into several segments and construct tables of the mean, variance, and autocorrelations within each segment. Under the conditions of stationarity, these values will remain relatively constant across segments. If the data are nonstationary with respect to one of these characteristics, however, a disparity in the values across segments of the time series may be apparent.

Finally, the spectral density function is very useful in identifying cyclic components in time series data. The present paper will not consider spectral density models, which belong to the class of frequency domain time series models. At an intuitive level, however, spectral decomposition involves modeling the time series by forming a summation of sine waves. This decomposition allows one to identify the frequencies of underlying periodic components of the time series process. The existence of periodicity in the data must then be taken into account when using time domain approaches (e.g. ARIMA models) to model the process.

One procedure that is sometimes utilized in the analysis of nonstationarity time series data 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 process. This approach assumes that most nonstationarity consists of two components; 1) trend, which is usually linear although in some cases it may be necessary to remove polynomial trends, and 2) deterministic cycles. These two components may be modeled via ordinary least squares fitting procedures. Gottman (1981) advocates the use of this technique as outlined below.

A linear trend for a time series is modeled as

$$Y_t = b_0 + b_1 t + a_t$$

where b_0 is the estimated level of the time series at time 0, b_1 is the estimated slope of the linear trend in the data, and a_t is the residual (representing a stationary time series process if the nonstationarity of the data is adequately modeled using only a linear trend component). Least squares fitting of polynomial trend follows directly from the linear model presented above.

Assuming that the time series has been "detrended" as described above, deterministic cyclical components may then be removed by fitting the model

$$Y_t = A(\sin 2\pi ft) + B(\cos 2\pi ft)$$

using ordinary least squares procedures to estimate A and B, while assuming a frequency (defined as the reciprocal of the length of the period) of f. Here Y_t is used to represent the residual of the time series after removing the linear trend component. It should be emphasized that this method of obtaining "de-sined" data is not appropriate for data sets exhibiting stochastic periodicity.

An alternative method for analyzing data sets that are nonstationary with respect to level involves a transformation of the data referred to as "differencing". As implied by its label, "differencing" is performed by calculating differences between pairs of observed values separated by a fixed number of time points. The simplest and most common form of differencing is called "first-order differencing". In this case, all observations are subtracted from the observation that immediately precedes it. Thus, the first differencing of a time series is defined as

$$\nabla X_t = X_t - X_{t-1}$$

where ∇ is used to indicate that the time series has been differenced.

It is easy to see the effect of first order differencing on a time series whose level increases monotonically, as shown below.

t =	1	2	3	4	5	6	7
X_t =	2.5	3.7	4.4	6.5	7.6	8.4	9.6
∇X_t =	---	1.2	.7	1.1	1.1	.8	1.3

The original data set shows a clear linear trend, and thus is nonstationary in the homogeneous sense (meaning with respect to level). The first differenced series, however, is relatively stable and fluctuates around a level of 1. The resulting time series can be thought of as a series of estimates representing the rate of change in the original data set.

First order differencing will also dramatically change the pattern of data sets exhibiting trends that are not simply increasing or decreasing over time. First differencing transforms data solely on the basis of adjacent time points, and thus, even temporary trends in the data will be removed by the differencing operation. For example, a time series realization that follows an upward linear trend during the first portion of the series, a stable level throughout the middle section of the series, and a decreasing linear trend during the last section of the time series will be transformed into a series with a relatively constant level.

Time series data that follows a quadratic trend can also be transformed to stationarity by differencing the observations. In this case it is necessary to perform "second-order differencing", which is simply differencing the first differenced series. A time series realization that has been second differenced is represented by $\nabla^2 X_t$. Similarly, data sets displaying n^{th} polynomial trends may be transformed to stationarity by n^{th} order differencing. It should be emphasized, however, that in practical applications of time series analysis it is rarely necessary to difference beyond the second order.

There is presently some controversy concerning which of the two procedures for analyzing nonstationary time series data generally possesses more desirable properties. Differencing is the most widely advocated approach for analyzing time series data that is nonstationary in the level. However, some researchers maintain that modeling nonstationary components of a time series process and subsequently removing these components from the data set is generally preferable.

Gottman (1981) contends that the modeling of nonstationary components is generally preferable to differencing, and should almost always be attempted before resorting to the differencing procedure. His argument in favor of modeling nonstationarity is based on two premises; 1) differencing radically transforms the data set, and 2) modeling the nonstationarity often suggests a meaningful interpretation of the nature of the nonstationarity. Gottman points out that differencing a white noise series (i.e. $Y_t = b_0 + a_t$, where $a_t \sim NID(0, \sigma_a^2)$ and b_0 represents the level of the series) actually introduces dependency in the data set. He argues that the potential misuse of such a powerful transformation may outweigh the benefits of the procedure. Gottman also believes that modeling of nonstationary components permits a more readily interpretable analysis of the time series data. The researcher can easily describe underlying trends in the data set in terms of an ordinary least squares regression (OLS) equation. Because of these advantages, he recommends that the researcher attempt to model nonstationarity before relying on the alternative of the differencing procedure.

Horne, Yang, and Ware (1982) also advocate the method of removing deterministic trends and seasonal components of time series data, and subsequently, modeling the residuals of the time series data using autoregressive and moving average parameters (i.e. ARMA models). Their basis for preferring this method of time series analysis is similar to Gottman's point concerning the interpretability of the analysis; it is argued that the underlying trends in a data set are of great interest to many research issues, and thus, the OLS regression modeling of trend provides a more meaningful insight into the time series process in comparison to differencing.

Many other authors contend that the procedure of removing trend via OLS regression analysis is generally inappropriate. According to McCleary et. al., "One common (but almost always inappropriate) method of detrending a time series is to use a linear regression model for the trend." They go on to point out several potential problems with the procedure. 1) OLS estimates of slope and intercept are sensitive to outliers, and thus, the existence of a small number of extreme data points can dramatically alter the estimates of the coefficients. 2) The point in time (t) of each observation serves as the independent variable in this procedure. This variable (t) increases monotonically, however, rather than being normally distributed. As a result, observations that are close to the beginning or the end of the series tend to have greater impact on the sum of squares function than observations that are close to the middle of the time series. In essence, the extreme values of t have an effect similar to outliers in OLS analyses, with the additional problem that the absence of numerous observations close to the mean value of t tends to further exaggerate the importance of these points in minimizing the residual sum of squares. 3) The distinction between "deterministic trend" and "stochastic drift" is emphasized by McCleary et. al. The modeling of trend via OLS regression analysis assumes an underlying deterministic process that will continue in the future as a fixed function of time. Differencing, on the other hand, assumes an underlying stochastic process that is free to vary in a probabilistic manner.

Box and Jenkins (1976) elaborate on the last point raised by McCleary et. al.; the assumption of a fixed, deterministic trend is often unjustified on the basis of a finite sample from a time series process.

> One of the deficiencies in the analysis of time series in the past has been the confusion between fitting a series and forecasting it. For example, suppose that a time series has shown a tendency to increase over a particular period and also follow a seasonal pattern. A common method of analysis is to decompose the series arbitrarily into three components; a "trend," a "seasonal component," and a "random component." The trend might be fitted by a polynomial and the seasonal component by a Fourier series. A forecast was then made by projecting these fitted functions.
>
> Such methods can give extremely misleading results..... Now, it is true that short lengths of Series B do look as if they might be fitted by quadratic curves. This simply reflects the fact that a sum of random deviates can sometimes have this appearance. However, there is no basis for the use of a quadratic forecast function, which produces very poor forecasts. Of course, genuine systematic effects which can be explained physically should be taken into account by the inclusion of a suitable deterministic component in the model (P.301).

In summary, there are legitimate arguments supporting the use of either method for removing nonstationarity from time series data. The arguments presented by McCleary et al. and by Box and Jenkins appear to convincingly rule out a simple procedure of routinely attempting to model nonstationary components of the time series data using OLS methods. Rather, the modeling procedure may be considered to be appropriate only when it can safely be assumed that the nonstationarity is of a deterministic nature. The practical consequences of choosing one procedure in favor of the other are not readily apparent.

Modeling Time Series Data

Autoregressive Models

Autoregressive time series models are an extension of the more common regression models used in a wide variety of applications. Autoregressive models simply predict observations in a time series from a previous set of observations in the series. For example, a time series realization may be modeled adequately by predicting each observation from the two observations that immediately precede the observation. In this case, the autoregressive model would be specified as

$$X_t = a_1X_{t-1} + a_2X_{t-2} + e_t$$

where a_1 and a_2 are the regression coefficients that minimize the square error, $\sum_{t=1}^{n} e_t^2$; and where t varies from three to n, the number of observations in the time series.

The variance of e_t is represented as σ_e^2, e_t is assumed to have a mean of 0, and e_t is assumed to be uncorrelated with $X_{t'}$, where $t' \neq t$.

If a time series realization is generated by a "second-order" autoregressive process, then the values of X_t will be predicted accurately by the previous two observations of the time series, X_{t-1} and X_{t-2}. It should be apparent that many time series processes that are studied in the social sciences are well described by autoregressive models. It is logical to expect that individual observations of an on-going social process will be accurately predicted by previous observations that are in close temporal proximity.

The observations of time series processes are usually represented as deviations from the mean observation, since algebraic manipulations of time series processes represented in this manner are greatly simplified. The theoretical conclusions that are arrived at by using this alternative representation are generally unaltered, and therefore, X_t will be used to represent $(X_t - \overline{X})$ throughout the remainder of this dissertation unless otherwise noted. This representation of time series processes simply "centers" the series around a mean observation of 0.

The first-order autoregressive process (AR(1)) is represented as

$$X_t = a_1 X_{t-1} + e_t .$$

It is important to note that the absolute value of a_1 must be less than one if a first-order autoregressive process is to be stationary. It will be later shown that the variance of an AR(1) process is

$$\sigma_x^2 = \sigma_e^2 \{ 1 + a_1 + (a_1)^2 + (a_1)^3 \ldots \} .$$

If the absolute value of a_1 is greater than or equal to one, the value of σ_x^2 will increase without bound. This type of time series process is referred to as an explosive series, and it is impossible to model such a series using time series models which are based on the assumption of stationarity. In addition, an autoregressive model in which a_1 is greater than one runs counter to the type of dependency that is generally assumed to be present in time series data. It will become apparent that if a_1 is greater than one, a given observation will be more strongly related to those observations that are temporarily distant in comparison to those that are close in temporal proximity. The autocovariance function of a first-order autoregressive process is calculated by multiplying the autoregressive equation by X_{t-k} and taking expected values.

$$X_{t-k}X_t = a_1 X_{t-k}X_{t-1} + e_t X_{t-k}$$
$$\mathrm{cov}(X_{t-k},X_t) = a_1 \mathrm{cov}(X_{t-k},X_{t-1}) + \mathrm{cov}(e_t,X_{t-k})$$
$$\gamma_k = a_1 \gamma_{k-1}$$

Here, γ_k is used to represent the autocovariance at lag k. The covariance between e_t and X_{t-k} is equal to zero, since e_t is independent of all observations other than X_t. The autocorrelation function is derived by dividing the autocovariance function by the variance of the observations ($\gamma_0 = \sigma_x^2$).

$$\rho_k = a_1 \rho_{k-1}$$

Thus, the autocorrelation at lag 1 (k=1) for a first order autoregressive process is

$$\rho_1 = a_1$$

since ρ_0 is equal to one by definition. The autocorrelation of observations at lag 2 (k=2) is

$$\rho_2 = a_1 \rho_1 = a_1^2 \quad (\text{since } \rho_1 = a_1) .$$

In general,the autocorrelation function of a first-order autoregressive process is defined as

$$\rho_k = a_1^{\,k} \quad .$$

A set of linear equations, referred to as the Yule-Walker equations, are used to express the parameters of autoregressive models in terms of autocorrelations and variances. The derivation of the equation is straightforward. A p^{th} order autoregressive process (AR(p)) can be represented as

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \ldots a_pX_{t-p} + e_t \ .$$

The autocovariance of a time series process is defined as

$$\gamma_k = E(X_t, X_{t-k})$$

where, as usual, X_t represents the deviation of observation t from the mean of the observations. Substituting for X_t results in

$$\begin{aligned}\gamma_k &= E\{(a_1X_{t-1} + a_2X_{t-2} + \ldots a_pX_{t-p} + e_t)(X_{t-k})\} \\ &= a_1E(X_{t-1}X_{t-k}) + a_2E(X_{t-2}X_{t-k}) + \ldots a_pE(X_{t-p}X_{t-k}) \\ &\quad + E(e_tX_{t-k}) \ .\end{aligned}$$

The definition of a stationary time series process demands that the autocovariance is a function only of the lag between observations, regardless of the values of t. As a result,

$$E(X_{t-1}X_{t-k}) = E(X_{t+s-1}X_{t+s-k})$$

for all values of s. It follows that

$$\gamma_k = a_1 \gamma_{k-1} + a_2 \gamma_{k-2} + \dots a_p \gamma_{k-p} + E(e_t X_{t-k})$$

since

$$(t+s-1) - (t+s-k) = k-1, \ (t+s-2) - (t+s-k) = k-2, \text{ etc.}$$

$E(e_t X_{t-k})$ is equal to zero for all $k > 0$, since e_t is by definition independent from all $X_{t'}$, where $t' \neq t$. Finally, the autocorrelation at lag k is determined by dividing by γ_0, the variance of the observations in the series

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2} + \dots a_p \rho_{k-p} \quad (\text{for } k > 0)$$

If k is set equal to zero, the autocovariance at lag 0 is defined as the variance of the observations

$$\gamma_0 = \sigma_x^2 = a_1 \gamma_{-1} + a_2 \gamma_{-2} + \ldots a_p \gamma_{-p} + E(e_t X_t)$$

The $\gamma_{(-s)} = \gamma_s$ by the definition of a stationary time series process, and thus

$$\gamma_0 = \sigma_x^2 = a_1 \gamma_1 + a_2 \gamma_2 + \ldots a_p \gamma_p + E(e_t X_t) \quad .$$

The expected value of $(e_t X_t)$ is σ_e^2, since the only portion of X_t which is correlated with e_t is the contribution of e_t. Therefore,

$$\gamma_0 = \sigma_x^2 = a_1 \gamma_1 + a_2 \gamma_2 + \ldots a_p \gamma_p + \sigma_e^2$$

Dividing by γ_0 results in

$$1 = \{a_1 \rho_1 + a_2 \rho_2 + \ldots a_p \rho_p + \sigma_e^2\} / \sigma_x^2$$

or

$$\sigma_e^2 = \sigma_x^2(1 - a_1 \rho_1 - a_2 \rho_2 - \ldots a_p \rho_p) \quad .$$

The Yule-Walker equations are extremely important in the estimation of unknown parameters of an autoregressive model. Given that the order of the autoregressive process is known, and the quantities $\hat{\sigma}_x^2$, $\hat{\rho}_1$, $\hat{\rho}_2, \ldots \hat{\rho}_p$ can be estimated from the data set, estimates of a_1, a_2, ... a_p and σ_e^2 can be computed using the Yule-Walker equations. The linearity of the Yule-Walker equations provide closed form solutions for the estimation of autoregressive parameters in AR models. In contrast, solutions for moving average and ARIMA models involve non-linear equations, and thus iterative procedures are necessary to estimate the parameters of interest. The relative simplicity of parameter estimation in autoregressive models has led some researchers (e.g. Gottman) to recommend their use almost exclusively. The consideration of only AR models will often result in a rather large number of autoregressive parameters, however. As a result other authors maintain that ARIMA models are preferable, since they provide more parsimonious models.

Moving Average Models

The discussion of autoregressive time series models described the basic characteristic of a stationary autoregressive process as a dependency between observations that decays exponentially as the numbers of time points between observations increases. It will be shown that a class of models that are referred to as moving average models will be useful in describing time series processes that are characterized by a different type of dependency.

Moving average processes exhibit a dependency between observations that are separated by a finite number of time points. Observations that are 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 using previous error terms (e_t) that are usually referred to as random shocks. The general principle of the model is that an observation X_t is a function of the current random shock e_t, and a portion of a fixed number of previous random shocks. The model is represented as

$$X_t = e_t - b_1 e_{t-1} - b_2 e_{t-2} - \cdots b_q e_{t-q}$$

which can also be written as

$$X_t = e_t - \sum_{s=1}^{q} b_s e_{t-s} \quad ,$$

Where e_t is a white noise series with variance σ^2_e, and X_t is used to represent $(X_t - \overline{X})$.

It can also be seen that the term moving average is actually a misnomer, since the values of b_q are usually not equal. Moving average models are actually defined as a weighted summation of random shocks.

It is instructive to initially consider the properties of a first-order moving average process, which is represented as

$$X_t = e_t - b_1 e_{t-1} .$$

The autocovariance function of any time series process is defined as

$$\gamma_k = E(X_t X_{t-k})$$

Substituting for X_t and X_{t-k}, and then taking expected values results in

$$\begin{aligned}\gamma_k &= E\{(e_t - b_1 e_{t-1})(e_{t-k} - b_1 e_{t-k-1})\} \\ &= E(e_t e_{t-k} - b_1 e_t e_{t-k-1} - b_1 e_{t-1} e_{t-k} + b_1^2 e_{t-1} e_{t-k-1}) \\ &= E(e_t e_{t-k}) - b_1 E(e_t e_{t-k-1}) - b_1 E(e_{t-1} e_{t-k}) + \\ &\qquad b_1^2 E(e_{t-1} e_{t-k-1}) .\end{aligned}$$

The variance of the series is

$$\begin{aligned}\gamma_0 &= \sigma^2_x \\ &= E(e_t^2) - b_1 E(e_t e_{t-1}) - bE(e_{t-1} e_t) + b^2 E(e_{t-1}) .\end{aligned}$$

Since the process e_t is a white noise process, with constant variance

$$\sigma^2_x = \sigma^2_e + b^2\sigma^2_e = (1 + b^2)\sigma^2_e .$$

The autocovariance of a first-order moving average model at lag 1 is

$$\begin{aligned}\gamma_1 &= E(e_t e_{t-1}) - b_1 E(e_t e_{t-2}^2) - b_1 E(e_{t-1}) + \\ &\qquad b^2 E(e_{t-1} e_{t-2}) \\ &= -b_1 E(e_{t-1}^2) \\ &= -b_1 \sigma^2_e\end{aligned}$$

If k is greater than one, $\gamma_k = 0$, since all e_t are assumed to be independent from $e_{t'}$ where $t' \neq t$. Thus, the autocovariance function of a MA(1) model truncates after a single lag. In comparison, the autocorrelation of an AR(1) model was shown to decrease exponentially ($\rho_1 = a$; $\rho_2 = a^2$; $\rho_p = a^p$).

Extending the derivation to a MA(q) model, it can be shown that the autocovariance function truncates to zero after q lags. The general form of the autocovariance function is

$$\gamma_k = -b_k\sigma^2_e + \sum_{s=k+1}^{q} b_s b_{s-k}\sigma^2_e \qquad .$$

The Duality of MA and AR Processes

It can be shown that a stationary AR(1) time series process can be represented as an infinite order MA process. Similarly, under certain conditions, a MA(1) process can be expressed as an infinite order autoregressive process. The practical importance of this duality is related to the flexibility that it provides in modeling time series data. For example, the researcher can adequately model a MA(1) process with an autoregressive model, AR(p), in which p is relatively large.

The equivalence of an AR(1) model and an infinite order moving average model will be considered first. Given our knowledge of the autocorrelation functions of MA and AR models, one might intuitively expect an AR(1) process to share similarities with a relatively high order MA process. The autocorrelation function of an AR(1) process decays exponentially over time, while that of a moving average process truncates after lag q of a MA(q) process. Logically, the only potential for modeling the dependency of an AR(1) process with a MA(q) model would be to specify a large value of q.

The first-order autoregressive model is specified as

$$X_t = a_1 X_{t-1} + e_t \quad .$$

The model can then be rewritten by substituting for X_{t-1},

$$X_t = a_1(a_1X_{t-2} + e_{t-1}) + e_t$$
$$= a_1^2X_{t-2} + a_1e_{t-1} + e_t \quad .$$

Substituting successively for each observation represented on the right side of the equation results in

$$X_t = a_1^2(a_1X_{t-3} + e_{t-2}) + (a_1e_{t-1} + e_t)$$
$$= a_1^3X_{t-3} + a_1^2e_{t-2} + a_1e_{t-1} + e_t$$
$$= e_t + a_1e_{t-1} + a_1^2e_{t-2} + a_1^3e_{t-3} + \ldots$$
$$= \sum_{i=0}^{\infty} a_1^i \, e_{t-i}$$

which is a MA(∞) model. It should be apparent that this relationship is only reasonable if $|a_1| < 1$, so that $a_1^k \to 0$ as $k \to \infty$, and thus, the series converges to a finite limit.

The first-order moving average model

$$X_t = e_t - be_{t-1}$$

also implies

$$X_{t-1} = e_{t-1} - be_{t-2} \quad .$$

Solving for e_{t-1} and substituting in the first equation results in

$$X_t = e_t - b(X_{t-1} + be_{t-2})$$
$$= -bX_{t-1} + e_t - b^2 e_{t-2} .$$

Rewriting the MA(1) model as

$$X_{t-2} = e_{t-2} - be_{t-3}$$

suggests a substitution for e_{t-2} in the previous equation

$$X_t = -bX_{t-1} + e_t - b^2(X_{t-2} + be_{t-3})$$
$$= -bX_{t-1} - b^2 X_{t-2} + e_t - b^3 e_{t-3} .$$

The repeated substitution results in

$$X_t = -bX_{t-1} - b^2 X_{t-2} - b^3 X_{t-3} - b^4 X_{t-4} \cdots$$

Which is an AR(∞) process. It should also be pointed out that this relationship will hold up only if $|b| < 1$, which is referred to as the invertibility condition of a MA(1) process.

ARMA and ARIMA models

Time series processes are sometimes best modeled by including both autoregressive and moving average parameters in the model. Although many authors believe that it is rarely useful to include both AR and MA parameters in a single model, it is sometimes convenient to represent time series processes using a single unified model. The ARIMA class of models incorporates the differencing operation into a model that also includes AR and MA parameters. ARIMA is an acronym for autoregressive integrated moving average process. The number of parameters of each type are specified by using the notation ARIMA (p,d,q), where p indicates the number of autoregressive parameters, d represents the degree of differencing that is required to obtain stationarity, and q indicates the number of moving average parameters in the model. In a comment on the general utility of models with both AR and MA parameters, McCleary et. al. (1980) state "if our experiences are typical, only a few social science time series in a thousand will have both p and $q \neq 0$".

At this point it is helpful to introduce some additional notation that is often used in represented "mixed" autoregressive moving average time series models. The backward shift operator, which is typically represented as B, acts as an operator which shifts the time series backward one point in time. Thus, the notation $B(X_t)$ is used to represent X_{t-1}. Superscripts are also used, which follow the general laws of exponents that are routinely used in algebraic manipulations. The notation $B^n(X_t)$ represents X_{t-n}, and $B^nB^m(X_t) = B^{n+m}(X_t) = X_{t-n-m}$.

The representation of the differencing operator is simplified by using the backward shift notation. First differencing can be expressed as

$$\nabla X_t = X_t - X_{t-1} = X_t - B(X_t) = (1-B)X_t \ .$$

Similarly, second differencing can be written as

$$\begin{aligned}\nabla^2 X_t &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} \\ &= (1 - 2B + B^2)X_t \\ &= (1 - B)^2 X_t \ .\end{aligned}$$

The backward shift operator also has the property of invertibility, so that $B^{-1}B = 1$. This is a useful property since B^{-1} can be used to represent a forward shift operator.

The backward shift operator is used to represent an AR(p) process as follows:

$$\begin{aligned}X_t &= a_1X_{t-1} + a_2X_{t-2} + \dots a_pX_{t-p} + e_t \\ e_t &= (X_t - a_1X_{t-1} - a_2X_{t-2} - \dots a_pX_{t-p})\end{aligned}$$

$$e_t = (1 - a_1B - a_2B^2 - \dots a_pB^p)X_t \quad .$$

Similarly, a MA(q) process is represented as

$$X_t = e_t - b_1e_{t-1} - b_2e_{t-2} - \dots b_qe_{t-q}$$
$$X_t = (1 - b_1B - b_2B^2 - \dots b_qB^q)e_t \; .$$

An ARIMA (1,0,1) model, which would represent a stationary time series process with one autoregressive parameter, and one moving average parameter can be written as

$$X_t = aX_{t-1} + e_t - be_{t-1}$$
$$X_t - aX_{t-1} = e_t - be_{t-1}$$
$$(1 - aB)Xt = (1 - bB)e_t \quad ,$$

or as

$$X_t = \frac{(1 - bB)}{(1 - aB)} e_t \quad .$$

Once again, the variance of the time series will converge to a finite limit only if the stationarity and the invertibility conditions are satisfied. In general, the roots of the polynomials of an ARIMA model will be complex numbers that can be represented as

$$S_k = U_k + iV_k$$

where U_k and V_k are real numbers, and $i = (-1)^{1/2}$. The modulus of the complex number s_k is represented as $|s_k|$, and defined as

$$|s_k| = (U_k{}^2 + V_k{}^2)^{1/2} \quad .$$

An AR or MA process is said to be stationary or invertible if all of the roots (s_k) have a modulus greater than one.

Model Building

Box and Jenkins (1976) have proposed a model building strategy which is the most widely employed procedure in the analysis of time series data. This strategy consists of an iterative procedure that is divided into three stages; identification, estimation, and diagnostic checking. The identification process involves selecting an ARIMA model that may parsimoniously describe the data set. Next, the parameters of the potential model that is chosen are estimated. Finally, the adequacy of the model is assessed in the diagnostic checking stage. This procedure will generally be repeated several times until, in the judgment of the researcher, the "best" model is determined. As with any model fitting procedure, there is ultimately a trade off between parsimony and further improvement in the fit of the model. Adding more parameters will generally improve the fit of the model, however, this more complicated model may simply be explaining idiosyncrasies of a single data set, and drawing inferences on the basis of a less parsimonious model may be difficult.

Model identification is perhaps the most important key to the analysis of time series data. This is also the portion of the model building strategy that requires the most subjective judgment on the part of the researcher. There is no precise objective method for determining the best values of p, d, and q in an ARIMA(p,d,q) model. Instead, the data set must be carefully examined for clues that provide suggestions of potential models. The information that is most useful in the identification of models is the autocorrelation function and the partial autocorrelation function (which will be discussed in the next paragraph). This is the information that most directly describes the dependency of time series observations, and thus, this information is essential in the proper identification of models that are designed to explain time dependency. In addition, it is it is often useful to present this information in graphical form to facilitate the model identification process.

The partial autocorrelation function is extremely useful in model identification. The actual autocorrelation function of a time series process is never known, and thus, a finite realization of the time series process must be used to estimate the true autocorrelation function. The imprecision of these estimates can result in considerable confusion as to the choice of the potential model. The interpretation of the partial autocorrelation function is very similar to the usual partial correlation coefficient, except the correlations of intermediate observations are "partialled out" instead of the intercorrelation of a third variable. Thus, the partial autocorrelation coefficient (k) represents the correlation between observations that are separated by a lag of k, after the autocorrelation of intermediate lags has been controlled for.

For purposes of clarification, it is useful to consider the partial autocorrelation function of an AR(1) time series process. It has already been shown that the autocorrelation function of an AR(1) process is $\rho_k = (a_1)^k$. The purpose of the partial autocorrelation function is to determine if the dependency between observations is adequately explained by the first-order autoregressive process, or alternatively, if the observations demonstrate a dependency even after the first-order autoregressive process is taken into account. Representing the time points

X_t, X_{t+1}, and X_{t+2} as 1, 2, and 3 in the standard equation for a partial correlation coefficient,

$$\rho_{13\cdot 2} = \frac{\rho_{13} - \rho_{12}\,\rho_{13}}{(1 - \rho_{12}^2)^{1/2}\,(1 - \rho_{32}^2)^{1/2}} \quad ,$$

we can determine the autocorrelation between observations X_t and X_{t+2}, after controlling for the intermediate observation X_{t+1}. By using the previously derived properties of the autocorrelation function of an AR(1) process ($\rho_{13} = \rho_2 = a_1{}^2$ and $\rho_{12} = \rho_{32} = a_1$), it can seen that the numerator of the partial autocorrelation is $a_1{}^2 - (a_1)(a_1) = 0$. Therefore the partial autocorrelation function of an AR(1) process at lag 2 is equal to zero. Similarly, the partial autocorrelation of all lags greater than 2 are equal to zero.

In contrast to the partial autocorrelation function of an AR(p) process which truncates after lag p, it can be shown that the partial autocorrelation function of a MA(q) process decays gradually. The general properties of the autocorrelation function and the partial autocorrelation function are summarized below.

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

It should be obvious at this point that estimates of the autocorrelation function and the partial autocorrelation function are extremely important in the identification of time series models. The characteristics of the autocorrelation function are virtually the only means of distinguishing one time series process from another. If the true autocorrelation function was known, the identification of time series models would be a relatively routine and precise procedure. In practice, however, the autocorrelation (and partial autocorrelation) functions are estimated on the basis of a finite set of observations.

The construction of confidence intervals around the estimates of autocorrelations and partial autocorrelations is useful in the identification of ARIMA models. The confidence intervals provide assistance for the researcher who is trying to determine which of the apparent dependencies in the data set are of a magnitude that is large enough to warrant consideration in the model building procedure. Bartlett (1946) has shown that the standard error of an autocorrelation coefficient at lag k may be estimated from the formula

$$SE(r_k) = \{(1/N)(1 + 2 \sum_{i=1}^{k-1} \rho_i^2)\}^{1/2}$$

where r_k is the estimate of the autocorrelation function ρ_k, N is the number of observations in the time series, and ρ_i are the true autocorrelations for all lags less than k. In practice, the estimated autocorrelations are substituted for ρ_i .

Quenouille (1947) has shown that the partial autocorrelations of a realization of an AR(p) process are distributed with variance 1/n for all partial autocorrelation at a lag of p+1 or greater. Thus, the standard error of the partial autocorrelation at lag k is estimated as

$$SE\{PACF(k)\} = 1/\sqrt{n} \quad .$$

Approximate 95% confidence intervals for the autocorrelations and partial autocorrelations can be formed around the zero value using the values ± 2 SE. If autocorrelations or partial autocorrelations fall within this interval, they are generally considered to be not significantly different form zero. Computer programs that plot the autocorrelation and partial autocorrelation functions with the appropriate confidence intervals are of great assistance in the identification of time series models. The width of the confidence bands are directly related to the number of observations in the time series process, and thus, certainty in the identification of the time series process is increased as the number of time points in the realization becomes greater.

The information provided by the estimated autocorrelation and partial autocorrelation functions is sufficient to specify a tentative ARIMA (p,d,q) model. It should be remembered that the researcher will usually examine the fit of several models, and therefore, the preliminary identification of a model does not imply a commitment to the model. The preliminary identification is based on an informal consideration of the general characteristics of the estimated autocorrelation and partial autocorrelation functions. The researcher simply examines these estimates for similarities to the known properties of various time series processes.

The first consideration is always the stationarity of the time series process. This issue was thoroughly considered in the discussion of stationarity and the analysis of nonstationary data. A tentative identification of the differencing parameter (d) is determined at this point. It should be noted that over-differencing is one of the most common errors in the use of ARIMA models. If the specified value of d is too large, as often happens when the researcher attempts to remove all evidence of nonstationarity in the data, dependencies among the observations are introduced. It is then necessary to remove these dependencies with AR and MA parameters, which usually results in a cumbersome ARIMA model with a large number of parameters. Furthermore, most authors contend that time series processes which require a value of d greater than 2 are extremely rare. Box and Jenkins (1976) state that "In practice d is usually 0, 1, or at most 2 (p.11)." McCleary et. al. (1980) also express the opinion that applications of time series analyses almost never require differencing beyond the second order. It is wise to avoid the problem of over-differencing by applying the difference operator only when the estimated autocorrelation function unambiguously demonstrates that the time series process is nonstationary.

The next step in the model building process is the estimation of the AR and MA parameters (i.e. $a_1, a_2, \ldots a_p, b_1, b_2, \ldots b_p$). It was shown previously that the Yule-Walker equations provide a closed form solution for the estimation of autoregressive models. Unfortunately, the general ARIMA parameters can not be estimated by using analytical solutions. Instead, numerical solutions are are required in the estimation procedure. There are several alternative estimation procedures, which are usually based on an iterative algorithm that is used to numerically derive maximum likelihood estimates of the ARIMA parameters.

Box and Jenkins suggest using a grid search procedure to minimize the residual sum of squares ($\sum_{i=1}^{n} e_i^2$) of an ARIMA (p,d,q) model. The procedure is conceptually simple, but the amount of computation involved makes the procedure inefficient even on today's high speed computers. The grid search procedure simply involves repeated substitution for the parameters $(a_1,a_2,...a_p,b_1,b_2,...b_p)$ to obtain estimates of $\sum_{i=1}^{n} e_i^2$, with the set of parameters which minimize the sum of squares function providing the best estimate of the parameters. It is easy to see that this procedure becomes impractical with even a relatively small number of parameters to estimate. Considerably more complex estimation procedures that converge to a minimum more rapidly are used in computer software designed to analyze time series data. Marquardt's (1963) algorithm, or minor variations of this algorithm, are generally applied in time series software.

After the parameters of an ARIMA (p,d,q) model have been estimated, the model must be evaluated in the diagnostic checking stage of the model building procedure. The analysis of the residuals of a time series model forms the basis of the diagnosis stage, in which the adequacy of the tentative model is evaluated. As usual, the estimated autocorrelation function is the principle source of information available to the analyst of time series data. McCleary et. al. (1980) outline some guidelines that can be used to evaluate the adequacy of an ARIMA model. First, there should be no dependency between the estimated autocorrelations at the first or second lag. A large autocorrelation would obviously suggest that the ARIMA model is not adequately accounting for the observed dependency of the data points.

Box and Jenkins (1976,p.290) point out that the approximate standard errors of the estimated autocorrelations of the residuals ($1/\sqrt{n}$) tend to be inflated at low lags. As a result, they suggest that the researcher consider the confidence intervals of the autocorrelations at low lags to be an upper bound of the true confidence intervals. Thus, discrepancies from the expected autocorrelation of zero at lags 1 or 2 should, for diagnostic purposes, be considered to be significant if they approach a value of $\pm$ two standard errors.

The second check of the residuals is designed to evaluate whether the residuals are distributed as white noise. This diagnostic check considers an entire set of autocorrelations simultaneously to evaluate whether the entire set of estimated autocorrelations are significantly different from 0. Box and Pierce (1970) suggest testing the independence of the residuals by the statistic Q, defined as

$$Q = n \sum_{j=1}^{k} r_j^2 \quad ,$$

which is distributed approximately as chi-square with k-p-q degrees of freedom. Here, n represents N-d (the number of observations used to estimate the autocorrelation function), k is the number of lags that the estimated ACF is calculated for, and r_j is the estimate of the autocorrelation at lag j. It is suggested that a minimum of 20 lags (k=20) should always be used to compute the Q statistic. McCleary et. al. claim that a large value of k (for example 50) will tend to lack power in rejecting the null hypothesis of independent observations, whereas a value of k less than 20 will tend to be over-sensitive to indications of serial dependency and lead to rejections of the null hypothesis even when the residuals are distributed as white noise. As a result, they recommend setting the value of k between 20 and 30 when using the Q statistic to detect serial dependency of the residuals.

It is also recommended that plots of the residuals and plots of observed values versus predicted values should be carefully examined by the researcher. The visual inspection of these plots is very useful in assessing the adequacy of the fit. In addition, these plots are useful in the detection of outliers and may provide indications of nonstationarity.

Finally, another useful procedure in the diagnosis stage of modeling is referred to as over-fitting. The basic concept of over-fitting is the attempt to find a better fitting model by adding parameters to the tentative ARIMA model that is being evaluated. These additional parameters should be selected on the basis of knowledge concerning possible sources of dependency that may not have adequately been modeled. The analysis of residuals that has just been discussed is the most logical source of information for the selection of the new parameters in the over-fitted model.

It is useful to compute a statistic that describes the amount of variance accounted for by each of the alternative models. A statistic analogous to the percentage of variance accounted for by a regression analysis can be defined as

$$R^2 = 1 - \sum_{t=1}^{n} \{\hat{e}_t^{\,2} / X_t^{\,2}\} \quad ,$$

where $X_t = (X_t - \overline{X})$, and R^2 indicates the amount of variance accounted for by the AR and MA parameters in the model. Another related measure of the goodness-of-fit of the model is the residual mean square statistic, defined as

$$RMS = \{1/n\}\{\sqrt{\Sigma \hat{e}_t^2}\} \quad .$$

Models that have a smaller residual mean square are better fitting than those models with a larger value.

The lack of precision in the model building procedures of time series analysis are very apparent in the use of the Q statistic, the estimation of the standard errors of the autocorrelation function, and the evaluation of the goodness-of-fit of a model. There is a large degree of subjective judgment involved in the interpretation of information that is used in analyzing time series data, as many of the properties of the estimators that are used in ARIMA modeling are not precisely defined at this time. Nevertheless, those who are familiar with time series analysis believe that a well informed researcher can use the available information to make judgments that will provide useful time series models. In addition, there is considerable interest in further developing and refining the procedures of time series analysis. Both theoretical and empirical research that is being conducted should lead to significant advances in the field.

Interrupted Time Series Experiments

The modeling of time series processes is usually only a preliminary step in the analysis of time series data. The researcher is generally interested in modeling a time series process for one of three purposes; 1) to forecast future points in time, 2) to draw causal inferences concerning the interrelationship of separate time series processes, or 3) to assess an intervention effect. The identification of an appropriate ARIMA model is usually the most complex and difficult aspect of the analysis of time series data, and the applications of ARIMA models for the purposes stated above are relatively straightforward.

Forecasting is widely used by social scientists in the fields of economics and political science. Most other areas of study are less interested in simply predicting future time points. Instead, they tend to focus on theoretical issues of causation; therefore, time series analyses that assess the relationship between different ongoing processes or evaluate the impact of an intervention offer greater potential utility to most social scientists. Of these two potential applications of time series models, only impact assessment has been widely used by social scientists. This is probably a reflection of current computer software capabilities rather than a response to the inherent usefulness of procedures involving multiple time series processes. In recent years, software programs that are capable of assessing univariate intervention effects have been developed and widely distributed. Unfortunately, programs that are designed for the analysis of multivariate time series processes are not generally available at the present time.

The most widely used method for the analysis of the interrupted time series experiment is an approach developed by Box and Tiao (1965) and discussed by Glass, Willson, and Gottman (1975). These methods involve the simultaneous estimation of the intervention component and the parameters of an ARIMA model using nonlinear estimation procedures. Gottman (1981,p.365) discusses another procedure that he proposes as a "simple, yet uninvestigated alternative". This procedure 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 method of analysis is extremely simple in comparison to the more widely used procedures, since all of the parameters of the model can be estimated with closed form solutions.

A third procedure for assessing the effect of interventions has been proposed by Box and Tiao (1975) and is recommended in the writings of McCleary et. al. (1980). This is the only procedure to be considered in the present discussion. The original Box and Tiao (1965) method is a special case of this general model, and thus, will not be described. The Gottman procedure is not often utilized, and will therefore also be excluded from this discussion.

The Box and Tiao (1975) approach to modeling intervention effects provides a straightforward conceptualization of the intervention effect, as well as providing a more manageable technique for incorporating complex intervention effects into the time series model. As mentioned previously, one type of intervention effect has been evaluated almost exclusively; that of an abrupt, constant change in level, where a constant value, δ , is added to each post-intervention observation. The recent development of this alternative method provides greater flexibility in the modeling of intervention effects, and thus, may lead to the evaluation of a wide variety of intervention effects. Of course, the distribution of the necessary computer software is a prerequisite to the widespread use of these techniques. As with any type of parameter estimation in the general class of ARIMA models, closed form solutions for the estimations do not exist, which results in tremendous computational difficulties.

The time series process with an intervention component can be represented succinctly as

$$X_t = f(I_t) + N_t$$

where N_t represents the "noise component" of the time series process, and $f(I_t)$ denotes a function of the variable I_t, which represents the intervention component. Thus, the time series process is simply assumed to be the outcome of two components; 1) the stochastic process of an ARIMA(p,d,q) process, represented by N_t, and 2) the deterministic effect of an intervention component ($f(I_t)$).

The simplest type of intervention effect is that of an abrupt, constant shift in level. In this case, the intervention component can be represented as

$$f(I_t) = w_0 I_t \quad ,$$

where $I_t = 0$ prior to the intervention and $I_t = 1$ after the intervention. The full impact assessment model can be written as

$$X_t = w_0 I_t + N_t \quad ,$$

and the intervention effect itself can be represented as

$$X_t - N_t = w_0 I_t \quad ,$$

or,

$$X_t^* = w_0 I_t \quad ,$$

where X_t^* represents the time series after removing the stochastic process modeled as ARIMA(p,d,q).

The general procedure of evaluating the effect of an intervention can be described as follows. An ARIMA(p,d,q) model is identified according to the general model building strategies that were previously explained. This entire procedure should be carried out separately for both the pre-intervention and post-intervention data, since the intervention effect will sometimes change the nature of the ARIMA process. In the event that the ARIMA process is altered by the intervention, it is not entirely clear how the researcher should proceed. There is an underlying assumption that the stochastic process of the time series data is equivalent before and after intervention. All of the impact assessment models that have been discussed rely on this implicit assumption, and thus one approach to the problem is to stop the analysis at this point with the conclusion that the intervention altered the nature of the time series process. Other authors have suggested explicitly modeling the two distinct time series processes (Stoline, Huitema, and Mitchell, 1980), or relying on the pre-intervention data to identify the ARIMA model (McCleary et. al., 1980) and then fit the entire model including the intervention component.

Assuming that the noise component of the pre- and post-intervention time series models are similar, and that the most appropriate ARIMA(p,d,q) model has been identified, the intervention component of the model is added and all of the parameters of the full model are estimated. The intervention effect is then evaluated using the estimate for the parameter w_0.

Other types of intervention effects can be incorporated into this model of the interrupted time series process by modifying the function $f(I_t)$. The general form of the modified function is

$$f(I_t) = \{w_0/(1 - \delta_1 B)\} I_t$$

where the parameter δ_1 is in the interval -1 to +1, and B indicates the backward shift operator. It will be seen that the parameter δ_1 estimates the rate at which the intervention effect approaches the asymptote, or eventual change in the level, of X_t. Thus, models in which δ_1 is not equal to zero imply a gradual change in level as a result of some intervention until an asymptotic level is reached. The series then remains relatively stable throughout the post-intervention phase of the time series realization.

Once again, the time series process after removing the stochastic noise component can be represented as

$$X_t^* = X_t - N_t$$
$$= \{w_0/(1 - \delta_1 B)\} I_t \quad .$$

Therefore,

$$(1 - \delta_1 B) X_t^* = w_0 I_t$$
$$X_t^* = \delta_1 X_{t-1}^* + w_0 I_t \quad .$$

The level of the time series X_t^* prior to the intervention is equal to zero, as $I_t = 0$;

$$X_t^* = \delta_1 X_{t-1}^* + w_o I_t$$
$$= \delta_1(0) + w_0(0) \quad .$$

The first post-intervention time point occurs at $t=n_1+1$, and the value of I_{n1+1} is equal to 1. Therefore,

$$X_{n1+1}^{*} = \delta_1 X_{n1}^{*} + w_0 I_{n1+1}$$
$$= \delta_1(0) + w_0(1)$$
$$= w_0 \quad .$$

At the second observation after the intervention,

$$X_{n1+2}^{*} = \delta_1 X_{n1+1}^{*} + w_0 I_{n1+2}$$
$$= \delta_1(w_0) + w_0(1)$$
$$= \delta_1 w_0 + w_0 \quad .$$

It can be seen that the recursive substitution for each subsequent post-intervention time point will result in the general representation of X_{n1+s}^{*} as

$$X_{n1+s}^{*} = \delta_1 X_{n1+s-1} + w_0 I_{n1+s}$$
$$= \delta_1(\delta_1^{s-1} w_0 +\ldots \delta_1 w_0 + w_0) + w_0$$
$$= \sum_{s=0}^{n-n1} \delta_1^{s} w_0 \quad .$$

Since the absolute value of δ_1 is always less than 1, the additional change in level at each post-intervention time point becomes smaller as time passes until an asymptotic level is reached.

As the value of δ_1 approaches 1, the rate of change tends to continue at a relatively constant amount, rather than decreasing as an asymptotic level is reached. In the most extreme case, where $\delta_1 = 1$, the level continues to increase at a continuous rate instead of eventually reaching a constant level. This intervention effect can be represented as

$$X_t^* = \{w_0 / (1 - B)\} I_t$$

This model indicates that the level of the series prior to intervention is 0, since $I_t = 0$. After intervention, however, the value of w_0 is added to each successive observation. This implies a change in level as follows:

$$X_{n1+1}^* = w_0$$
$$X_{n1+2}^* = 2w_0$$
$$X_{n1+3}^* = 3w_0$$
$$\vdots$$
$$X_{n1+s}^* = Sw_0$$

It is obvious that this type of intervention effect is virtually impossible in practice. As time passes, the additional impact of an intervention will inevitably decay.

At the opposite extreme, an intervention whose full impact is realized at the first post-intervention time point can be modeled by setting $\delta_1 = 0$. In this case the intervention component reduces to $w_0 I_t$, which is the model that has been implicit in almost all interrupted time series experiments. An abrupt, temporary intervention effect can also be modeled by slightly altering the representation of the intervention component. In this case, the complete model becomes

$$X_t = \{[w_0 / (1 - \delta_1 B)] (1 - B)\} I_t + N_t$$

or,

$$X_t^* = \{ [w_0 / (1 - \delta_1 B)] (1 - B) \} I_t$$

$$X_t^* = \delta_1 X_{t-1}^* + w(1 - B) I_t$$

where $I_t = 0$ before the intervention is introduced, and $I_t = 1$ after the onset of the intervention $(t = n_1 + 1)$. Assuming this model,

$$(1 - B) I_{n1+1} = I_{n1+1} - I_{n1} = 1 - 0 = 1$$

thus,

$$X_{n1+1}^{*} = \delta_1 (0) + w_0(1)$$
$$= w_0 \quad .$$

At the second post-intervention observation,

$$(1 - B)I_{n1+2} = I_{n1+2} - I_{n1+1}$$
$$= 1 - 1$$
$$= 0 \quad .$$

and therefore,

$$X_{n1+2}^{*} = \delta_1(w_0) + w_0(0)$$
$$= \delta_1 w_0 \quad .$$

It follows that at time point $t = n_1+s$, the intervention effect can be represented as

$$X_{n1+s} = \delta_1^{s-1} w_0$$

It is readily apparent that this form of the intervention component represents an abrupt impact with a magnitude of w_0 at the point of intervention. The effect of the intervention then decays at a rate determined by δ_1.

Three different types of intervention effects have been illustrated using the interrupted time series model which was first proposed by Box and Tiao (1975) and later advocated by McCleary et. al.(1980). First, the modeling of an abrupt, permanent impact was discussed. Next, a model that implies a gradual realization of a permanent intervention effect was described. And finally, an intervention model that assumes an immediate effect that gradually diminishes was described. Although these three forms of the intervention component probably represent the most commonly encountered intervention effects, they are by no means the only types on intervention effects that can be modeled using this approach. In fact, virtually any type of intervention effect can be evaluated by adding a second (or even third) intervention component to the model, or by creating more complex intervention components with more than one rate parameter (represented by δ_n).

Statement of the Problem

The most common methods for the analysis of interrupted time series experiments have been discussed in this chapter. As with any type of time series analysis, the primary focus of the procedures involves modeling the dependency among data points that is characteristic of most time series processes. The stationarity conditions are a prerequisite for modeling the autocorrelation of the time series process, and thus, careful attention must also be devoted to the ramifications of this assumption. The central importance of these issues is reflected in the proportion of this chapter that is concerned with the conditions of stationarity and the modeling of serial dependency.

Only those time series modeling procedures that are closely related to the general class of ARIMA models have been examined. Even within this limited class of models, only the essential aspects of ARIMA models were considered. The lengthy volume of Box and Jenkins' (1976) attests to the complexity and number of issues that can be involved in ARIMA models of time series processes. In addition to the general class of ARIMA models, many other approaches to the analysis of time series processes have been developed. The most widely used of these procedures is spectral analysis, which was only briefly mentioned in this chapter. There has also been a recent proliferation of other alternative procedures, many of which are still in the process of being developed and refined.

An approach recently suggested by Box and Tiao (1975) and advocated by McCleary et. al. (1980), is perhaps the most promising available method for the analysis of the interrupted time series experiment. This procedure offers the flexibility of modeling almost any type of intervention effect, and thus has the potential of providing researchers with more explicit information about the nature of the intervention effect.

There are numerous issues concerning interrupted time series data analysis that are unanswered at the present time. First, all of the statistical methods that have been discussed are based on asymptotic theory. There has been a conspicuous absence of research devoted to the investigation of the small sample properties of these procedures, and thus, it is impossible to determine the circumstances under which the application of these procedures is appropriate. The most fundamental question that the researcher is faced with concerns the number of time points that are necessary to accurately estimate the effect of an intervention. Although there are some vague guidelines available, an accurate answer in the context of many situations is not presently available.

The general problem of a lack of information concerning the small sample properties of these estimation procedures has several implications for the evaluation of intervention effects. The ARIMA(p,d,q) 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 p, d, and q; but at present, it is not clear what length is generally sufficient. Furthermore, it is not known how severely the misidentification of p, d, and q will distort the test of an intervention effect. It is possible that ARIMA(p,d,q) models that are similar, but not identical, to the true ARIMA(p,d,q) process will result in a negligible effect on the test of intervention effects.

Other issues that may be of interest to researchers involve the consequences of violating the underlying assumptions of time series models. In practical applications, the theoretical assumptions of any statistical procedure are never completely fulfilled. The investigation of the robustness of time series procedures is an important line of research that has not yet been pursued.

In summary, there are a wide variety of practical issues that are extremely important to researchers who utilize the interrupted time series experimental design. Although a variety of theoretical approaches to the analysis of interrupted time series data have been developed, there has been virtually no research concerning the problems of applying these procedures to "real-life" data sets. To address these issues, it is necessary to empirically examine the small sample properties of these procedures under a variety of conditions by conducting Monte Carlo experiments.

CHAPTER III

METHODOLOGY

Introduction

The research methodology will be presented as a series of four interrelated studies. The first three Monte Carlo experiments examine the small sample properties of the autocorrelation and partial autocorrelation functions, as they are commonly utilized in the identification of ARIMA(p,d,q) time series models. The importance 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. The fourth study investigates the small sample properties of the test statistic for the analysis of intervention effect prescribed by Box and Tiao (1965,1975).

For each specific condition in the four studies, 1000 time series realizations were randomly generated according to a given ARIMA(p,d,q) process using the IMSL subroutine FTGEN. The data generating program allows the user to specify the population autoregressive. moving average, and white noise parameters of the time series process, as well as the length of the time series realization. In study four, an intervention effect of an abrupt permanent change in level was imposed within a general FORTRAN program by simply adding a constant to all of the post-intervention time points.

Each of the four studies varies either two or three factors in all possible combinations. Therefore, the design of the simulations can be thought of as being analogous to a completely crossed, factorial experimental design. The two factors that were manipulated in all four studies were 1) the values of autoregressive or moving parameters, and 2) the number of the time points in the data sets. Study four adds a third factor to the experimental design.

Study One

The primary purpose of study one is the investigation of the sampling variability of the autocorrelation and partial autocorrelation functions under a variety of conditions. The bias of the estimates will also be considered, although the magnitude of bias can be theoretically derived given the population parameters of the ARIMA process and the length of the time series realization.

For each condition examined, 1000 data sets were generated according to the parameters specified in the condition. The discrepancy between the mean of the 1000 parameter estimates and the true parameters was used to assess the degree of bias in various conditions. The sampling variability of the estimates was measured by computing the standard deviation of the 1000 parameter estimates. This measure can be considered to be an empirical estimate of the standard error of the autocorrelation and partial autocorrelation coefficients.

The first factor that was manipulated in study one was the nature of the autocorrelation structure. Four different ARIMA(p,d,q) processes were considered, with three different parameter values examined for each of the four models. The three AR(1) processes were generated with autoregressive coefficients of .3, .6, and .9. The three AR(2) processes had autoregressive parameters of (.4, .3), (.5, .3), and (.6, .3). The MA(1) coefficients were -.3, -.6, and -.9, while the MA(2) parameters were (-.4, -.3), (-.5, -.3), and (-.6, -.3).

The second characteristic that was varied is the number of time points in the data sets. Each replication consisted of either 30, 60, 90, or 120 time points. The lengths of the realizations were intended to cover a wide range; from the minimum number of time points for which time series analysis might be considered (30), to a number that would be considered to be adequate by most researchers (120).

The estimates of the autocorrelation and partial autocorrelation functions were computed with the IMSL subroutine FTAUTO. The estimators are those that are most commonly employed, and are described in detail in Chapter Two. Procedures for estimating the mean and standard deviation of the estimates were programmed in FORTRAN.

Study Two

Study two investigates the accuracy of approximate standard errors of the estimated autocorrelation and partial autocorrelation functions. In addition, the directly related issue of the adequacy of the procedure for constructing approximate confidence intervals around the estimates of the coefficients is considered. The specific research topics investigated in study two are:

- o the discrepancy between the estimated and empirical standard errors of the autocorrelation functions;
- o the Type I error rate and power of the statistical test for a non-zero autocorrelation coefficient; and,
- o the Type I error rate and power of the statistical test for a non-zero partial autocorrelation coefficient.

As discussed in Chapter Two, the confidence intervals are based on the formulae for approximate standard errors;

$$SE(r_k) = [(1/N)(1 + 2\sum_{i=1}^{k-1} \rho_i^2]^{1/2} \quad \text{, and}$$

$$SE[PACF(k)] = 1/\sqrt{N} \; .$$

It is important to remember that these expressions are based on the assumption that ρ_k is known and that all of the autocorrelations at lag k or greater are equal to zero. Consequently, the accuracy of these approximate standard errors for finite sample sizes is questionable.

Study two was executed as a subroutine of the FORTRAN program used in study one. Thus, two factors were manipulated in a manner identical to study one. As described in study one, a total of 12 different ARIMA processes were examined, with data sets of length 30, 60, 90, and 120 time points.

Several results of interest were computed. The mean (over the 1000 replications) of the estimated standard errors of the autocorrelation coefficients were computed for the first five lags. The mean values of the estimated standard errors are then compared with the empirical standard errors which were computed in study one. In addition, test statistics were constructed to investigate the Type I error rate and power of testing the null hypotheses H_o: $\rho_k = 0$ and H_o:PAC(k) = 0.

Study Three

Study three examines the usefulness of Quenouille's (1956) unbiased estimator of the autocorrelation function. As previously discussed, the bias in the usual estimator of the autocorrelation function is very large for relatively short time series realizations (i.e. less than 100 observations). Quenouille proposed a jackknife procedure to correct for bias using the formula below

$$R_k = 2r - 1/2\{r_1 + r_2\},$$

where R_k is the unbiased estimate of autocorrelation at lag k, r represents the ordinary autocorrelation estimate (at lag k) for the entire series, and r_1 and r_2 are the usual autocorrelation estimates (at lag k) for the first and second halves of the series, respectively.

There are two potential problems with the application of this estimation procedure. First, unlike the ordinary estimation procedure, there is no assurance that the value of R will be within the theoretical bounds of $-1 \leq R \leq 1$. Thus, estimates that are theoretically impossible may occur. Secondly, the standard error of the unbiased estimator may be considerably larger than that of the biased estimator, and consequently, it may be a less desirable estimator despite its attribute of unbiasedness. The purpose of study three can be summarized as follows:

- o to determine the percentage of unbiased estimates that are outside the theoretical boundaries of the parameter; and,
- o to compare the sampling variability of the unbiased estimates with those of the ordinary biased estimates.

As with studies one and two, the two factors manipulated were the nature of the ARIMA process and the number of time points in the realization.

The IMSL subroutine FTAUTO was used to estimate the ordinary biased autocorrelation function of a) the entire series, b) the first half of each realization, and c) the second half of each realization. Based on the results generated by FTAUTO, FORTRAN statements were used to calculate the unbiased estimates of the autocorrelation function. The percentage of estimates that exceeded the absolute value of 1.0 were then tabulated. Furthermore, the standard deviation of the estimates over 1000 replications in each condition were calculated as an empirical measure of standard error.

Study Four

This study is designed to investigate the small sample properties of the test statistic of intervention effect proposed by Box and Tiao (1965,1975). Several important properties of the test statistic are examined including:

- the distribution of the test statistic;
- the Type I error rate of the test statistic;
- the statistical power of the test statistic; and,
- the accuracy of the estimates of standard error.

The form of 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 several considerations. Most importantly, an abrupt permanent change in level is the most common form of impact assessment in social science research applications. Furthermore, it is important to gain a thorough understanding of the sampling properties of the most straightforward intervention model before attempting to study more complicated intervention components involving additional parameters.

The scope of the present study is also limited to the first order autoregressive process. The AR(1) process is considered because it is probably the most prevalent model employed in social science research. Moreover, it is important to begin the systematic investigation of the intervention analysis procedure with the more basic time series processes. Problems which are encountered in the test statistic of relatively simple time series processes are likely to be common to other processes, and may very well create even more serious difficulties in ARIMA models of greater complexity.

The present investigation employs estimation procedures based on the exact likelihood function. Most computing algorithms exploit the relationship between maximum likelihood estimation and least squares procedures (e.g., BMDP and IMSL). However, according to Harvey and Phillips (1979), many authors have recently stressed the importance of computing ARIMA estimates using the exact likelihood function.

The logarithm of the exact likelihood function of a first order autoregressive process without an intervention component (Fuller, 1976,p.328) is

$$\log L\ (x;\mu,\rho,\sigma^2) = -(n/2)\log 2\pi - (n/2)\log\sigma^2 + (1/2)\log(1-\rho^2)$$

$$-(1/2\sigma^2)\{(X_1 - \mu)^2(1 - \rho^2) + \sum_{t=2}^{n}[(X_t - \mu) - \rho(X_{t-1}-\mu)]^2\}$$

The likelihood function can be modified to incorporate the intervention component as follows:

$$\begin{aligned}\log L\ (x;\mu_1,\mu_2,\rho,\sigma^2) = & -(n/2)\log 2\pi - (n/2)\log\sigma^2 + \\ & + (1/2)\log(1 - \rho^2) \\ & - (1/2\sigma^2)\{(X_1 - \mu_1)^2(1 - \rho^2) \\ & + \sum_{t=2}^{r}[(X_t - \mu_1) - \rho(X_{t-1} - \mu_1)]^2 \\ & + [(X_{r+1} - \mu_2) - \rho(X_r - \mu_1)]^2 \\ & + \sum_{t=r+2}^{n}[(X_t - \mu_2) - \rho(X_{t-1} - \mu_2]^2\}\end{aligned}$$

X_r is used to represent the midpoint of the time series realization. The point at which the likelihood function is maximum provides the maximum likelihood estimates of μ_1, μ_2, ρ, and σ^2.

Several procedures based directly on the joint maximization of the likelihood function are available for maximum likelihood estimation of the parameters of the intervention model. The IMSL subroutine ZXMIN, which employs a quasi-Newton method to find the minimum of a function of variables, was used to minimize -(log L).

Starting values for each replication were determined on the basis of closed form expressions for each of the parameters, based on the assumption that estimates of the other parameters are population parameters. Thus the starting values that were provided were

$$\mu_1 = \{\sum_{i=1}^{r} X_i\}/r$$

$$\mu_2 = \{\sum_{i=r+1}^{n} X_i\}/(n-r)$$

$$\rho_1 = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}$$

$$\sigma^2 = \text{var}\ (1 - \rho^2)$$

Unfortunately, a failure to meet the convergence criterion ($|t_{i+1} - t_i| < .001$) was a problem often encountered with the joint maximization procedure. It was not uncommon for 30-40% of the replications to fail to converge in certain conditions.

Several computational methods designed to increase the percentage of replications that successfully converged were attempted. The most promising procedure was to compute estimates using a stage-wise estimation procedure, and then use these estimates as starting values for the simultaneous estimation procedure. The stage-wise procedure involves fixing μ_1, μ_2, and σ^2 in the likelihood function, and maximizing the function with ρ as the only variable. Then, closed form solutions for the three parameters that were previously fixed (μ_1, μ_2, σ^2) are computed by treating the estimate of ρ as a known parameter.

First, an estimate of σ^2 is obtained by setting

$$\frac{\log L}{\partial\sigma^2} = 0$$

which results in the closed form expression

$$\sigma^2 = (1/n)(1 - \rho^2)(X_1 - \mu_1)^2 + \sum_{t=2}^{r} \{(X_t - \mu_1) - \rho(X_{t-1} - \mu_1)^2\}$$
$$+ \{(X_{r+1} - \mu_2) - \rho(X_r - \mu_1)^2\}$$
$$+ \sum_{t=r+2}^{n} \{(X_t - \mu_2) - \rho(X_{t-1} - \mu_2)^2\}$$

Previous estimates of ρ, μ_1, and μ_2 are substituted and the equation is solved.

Next, setting

$$\frac{\partial \log L}{\partial\mu_1} = 0 \quad \text{and} \quad \frac{\partial \log L}{\partial\mu_2} = 0$$

yields the system of linear equations

$$[1+(r-1)(1-\rho)^2]\mu_1 - \rho\mu_2 = (1-\rho^2)X_1 + (1-\rho)\sum_{t=2}^{r}\{(X_t - \rho X_{t-1}) - \rho(X_{r+1} - \rho X_r)\}$$

$$\rho\mu_1 - [(r-1)(\rho-1)^2+1]\mu_2 = \rho X_r - X_{r+1} - (1-\rho)\sum_{t=r+2}^{n}(X_t - \rho X_{t-1})$$

By substituting the previous estimate of ρ as a known quantity, the simultaneous solution of the two equations provides estimates of μ_1 and μ_2. The entire stage-wise estimation procedure is then repeated until the convergence criterion ($|t_{i+1} - t_i| < .001$) is reached. Finally, the parameter estimates of the stage-wise procedure were used as starting values for the simultaneous estimation of all four parameters.

This procedure produced a much higher percentage of successful convergence (over 95% for most conditions), but created another problem. The amount of computing time required to go through the two separate iteration procedures for the 1000 replications in each condition made the procedure cumbersome to execute. It was decided that the results of the stage-wise procedure should be compared with the full simultaneous estimation procedure. If the comparison of the two procedures yielded nearly identical results, the stage-wise estimation procedure would be used for the investigation of the test of intervention effect.

The comparison of the two procedures was conducted by generating 500 replications of several conditions (varying ρ , σ^2, n, and the magnitude of the intervention), and estimating the parameters using both the stage-wise procedure described above and the procedure for estimating all four parameters simultaneously. Only data sets that successfully converged for both estimation procedures were included in the comparison of the procedures. Descriptive statistics were then computed on the two sets of estimated parameters and examined for similarities and differences. The two procedures were nearly identical with respect to mean parameter estimates, standard deviations of the estimates, and other descriptions of the distribution of the estimates such as skewness and kurtosis.

A second method of comparison was the computation of Pearson product-moment coefficients of the corresponding pairs of estimates computed according to the two alternative procedures. In all conditions the correlation coefficients between the estimates of the two procedures was greater than .95, and in most instances the coefficients approached 1.0. Consequently, it was concluded that the stage-wise estimation procedure would be utilized because of its advantage in computational efficiency.

The other quantity that must be estimated to construct the test statistic is the standard error of the estimate $\mu_1 - \mu_2$. The appropriate estimate of the standard error is derived as follows. The information matrix (given below) of the parameters to be estimated is obtained. The negative of the inverse of this matrix yields the variances of the maximum likelihood estimators of the parameters. The information matrix for the current problem is defined as:

$$\varepsilon \begin{bmatrix} \frac{\partial^2 \log L}{\partial \mu_1^2} & \frac{\partial^2 \log L}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 \log L}{\partial \rho \partial \mu_1} & \frac{\partial^2 \log L}{\partial \sigma^2 \partial \mu_1} \\ \frac{\partial^2 \log L}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 \log L}{\partial \mu_2^2} & \frac{\partial^2 \log L}{\partial \rho \partial \mu_2} & \frac{\partial^2 \log L}{\partial \sigma^2 \partial \mu_2} \\ \frac{\partial^2 \log L}{\partial \rho \partial \mu_1} & \frac{\partial^2 \log L}{\partial \rho \partial \mu_2} & \frac{\partial^2 \log L}{\partial \rho^2} & \frac{\partial^2 \log L}{\partial \sigma^2 \partial \rho} \\ \frac{\partial^2 \log L}{\partial \sigma^2 \partial \mu_1} & \frac{\partial^2 \log L}{\partial \sigma^2 \partial \mu_2} & \frac{\partial^2 \log L}{\partial \sigma^2 \partial \rho} & \frac{\partial^2 \log L}{\partial (\sigma^2)^2} \end{bmatrix}$$

It can be shown that this matrix simplifies to

$$- \begin{bmatrix} \frac{1+(r-1)(1-\rho)^2}{\sigma^2} & -\rho/\sigma^2 & 0 & 0 \\ -\rho/\sigma^2 & \frac{1+(r-1)(1-\rho)^2}{\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{1}{1-\rho^2}\{\frac{1+\rho^2}{1-\rho^2} + (n-2)\} & \rho/\sigma^2(1-\rho^2) \\ 0 & 0 & \rho/\sigma^2(1-\rho^2) & n/(2\sigma^4) \end{bmatrix}$$

where r represents 1/2 the length of the time series.

Because μ_1 and μ_2 are the only parameters of interest in the present context and the matrix is of the form of a diagonal block, the 2 x 2 block can be inverted independently. The negative of the inverse of the block provides in the variance-covariance matrix of μ_1 and μ_2 is

$$\{1 + (r-1)(1-\rho)^2\}^2 \begin{bmatrix} 1 + (r-1)(1-\rho)^2 & \rho \\ \rho & 1 + (r-1)(1-\rho)^2 \end{bmatrix}$$

It follows that the standard error of $(\mu_1 - \mu_2)$ is

$$\sqrt{\{(2\sigma^2)(1-\rho)[1+(r-1)(1-\rho)]\} \,/\, \{\,[1+(r-1)(1-\rho)^2]^2 - \rho\,\}}$$

It is assumed that the test statistic is distributed as t with degrees of freedom equal to n-2, where n is equal to the length of the realization.

The design of the Monte Carlo experiment can be thought of as a completely crossed three factor design. The factors that were systematically manipulated were 1) the magnitude of the autoregressive parameter (.3, .6, or .9), 2) the length of the time series realization (60, 90, 120, or 150 time points) and 3) the magnitude of the intervention (0, 0.5, 0.8, or 1.1). The variance of the time series processes were held constant at 1.0 by adjusting the value of the white noise parameter. The variance is a function of both the white noise variance of the series and the autocorrelation of the series,

$$VAR(X_t) = \sigma^2 / (1 - \rho^2)$$

Thus, fixing the variance of the process at 1.0 implies the following white noise parameters: (ρ =.3,σ^2 =.91); (ρ =.6, σ^2 =.64); (ρ =.9, σ^2=.19).

As with studies one through four, 1000 time series realizations were generated by IMSL subroutine FTGEN for each condition under consideration. After generating the each data set, a constant (.5, .8, or 1.1) was added to each of the post-intervention data points of conditions in which an intervention effect was present. Those conditions in which the model did not include an intervention effect (i.e., H_o: $\mu_1 - \mu_2 = 0$ is true) were left unaltered.

Estimates of the intervention component ($\mu_1 - \mu_2$) and its standard error were obtained as described above. SPSS routines were employed to test for the normality of the distribution of the test statistic (Kolmogorov-Smirnov one-sample test), to generate descriptive statistics of all of the estimated parameters, and to calculate the percentage of statistically significant rejections of the null hypothesis.

CHAPTER IV

RESULTS AND DISCUSSION

Study One

Monte Carlo methods were used to examine the sampling properties of the autocorrelation and partial autocorrelation functions under a variety of conditions. 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, and therefore, the primary interest in study one is the empirical estimation of the standard errors of the autocorrelation and partial autocorrelation coefficients. Information concerning both the standard error and bias of the estimates is essential for determining the length of time series realization that is necessary to ensure a reasonably high likelihood of an appropriately identified ARIMA(p,d,q) model.

The first model to be considered is the first order autoregressive model. The mean (over 1000 replications) estimates of the autocorrelation function for lags 1 through 3, and the standard deviation of these estimates, are presented in Table 1. The true parameter value of each autocorrelation coefficient is also presented for the

Table 1

First Order Autoregressive Processes

Mean Estimated Autocorrelation (AC) and
Empirical Standard Error (SE)

AR1	AR2	TP	LAG1 AC	LAG1 SE	LAG2 AC	LAG2 SE	LAG3 AC	LAG3 SE
.9	–	ρ	.900	–	.810	–	.729	–
		120	.859	.051	.736	.090	.628	.120
		90	.844	.062	.709	.107	.590	.141
		60	.809	.087	.650	.146	.519	.182
		30	.724	.133	.508	.197	.336	.227
.6	–	ρ	.600	–	.360	–	.216	–
		120	.574	.078	.326	.105	.177	.118
		90	.566	.088	.313	.118	.163	.129
		60	.538	.114	.276	.154	.135	.161
		30	.488	.156	.211	.192	.054	.200
.3	–	ρ	.300	–	.090	–	.027	–
		120	.285	.091	.077	.095	.013	.096
		90	.280	.103	.070	.109	.005	.107
		60	.260	.126	.048	.139	–.002	.132
		30	.232	.169	.025	.173	–.044	.170

purpose of comparison.

The mean estimate of the autocorrelation function reflects the considerable magnitude of bias in the estimator when applied to finite sample sizes. The bias is a function of both the true autocorrelation parameter (ρ_k) and the number of observations in the series (n). The bias is always downward, which results in an underestimation the autocorrelation. The degree of underestimation increases as ρ becomes larger and as n becomes smaller. Furthermore, the magnitude of the bias increases as the lag of the autocorrelation coefficient increases.

The results summarized in Table 1 are also presented graphically in a series of figures. The first graph illustrates the downward bias of the autocorrelation estimates at lag 1. As the number of time points in the realization increases, the mean of the 1000 estimated autocorrelations approaches the true values of .9, .6, and .3. The same pattern of results are apparent for lags 2 and 3, which are displayed in Figures 2 and 3. The bias becomes even more severe as the lag increases, especially for the condition with the greatest serial correlation (AR1 = .9).

The standard deviations of the 1000 parameter estimates provide an empirical measure of the standard error, which is extremely important given the absence of a theoretically

Figure 1

AR1 PROCESSES – LAG 1 AUTOCORRELATIONS

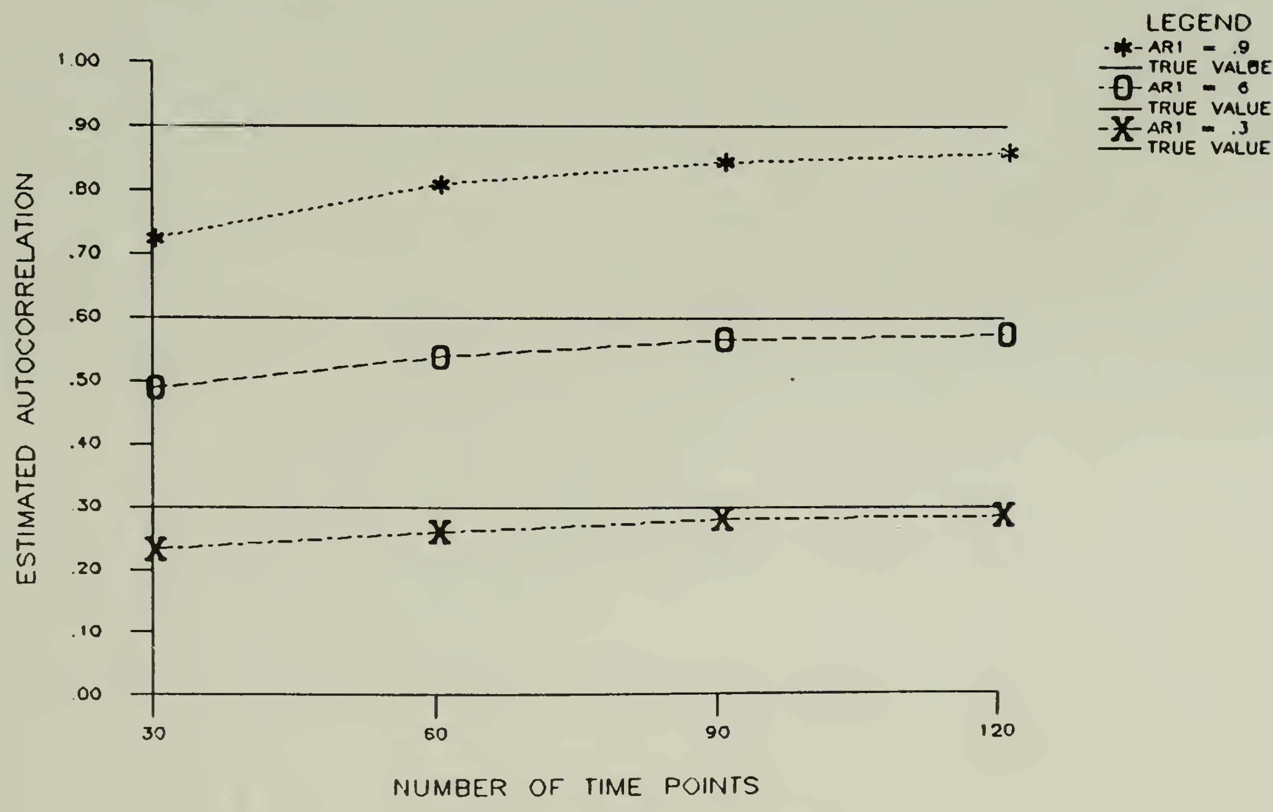


Figure 2

AR1 PROCESSES – LAG 2 AUTOCORRELATIONS

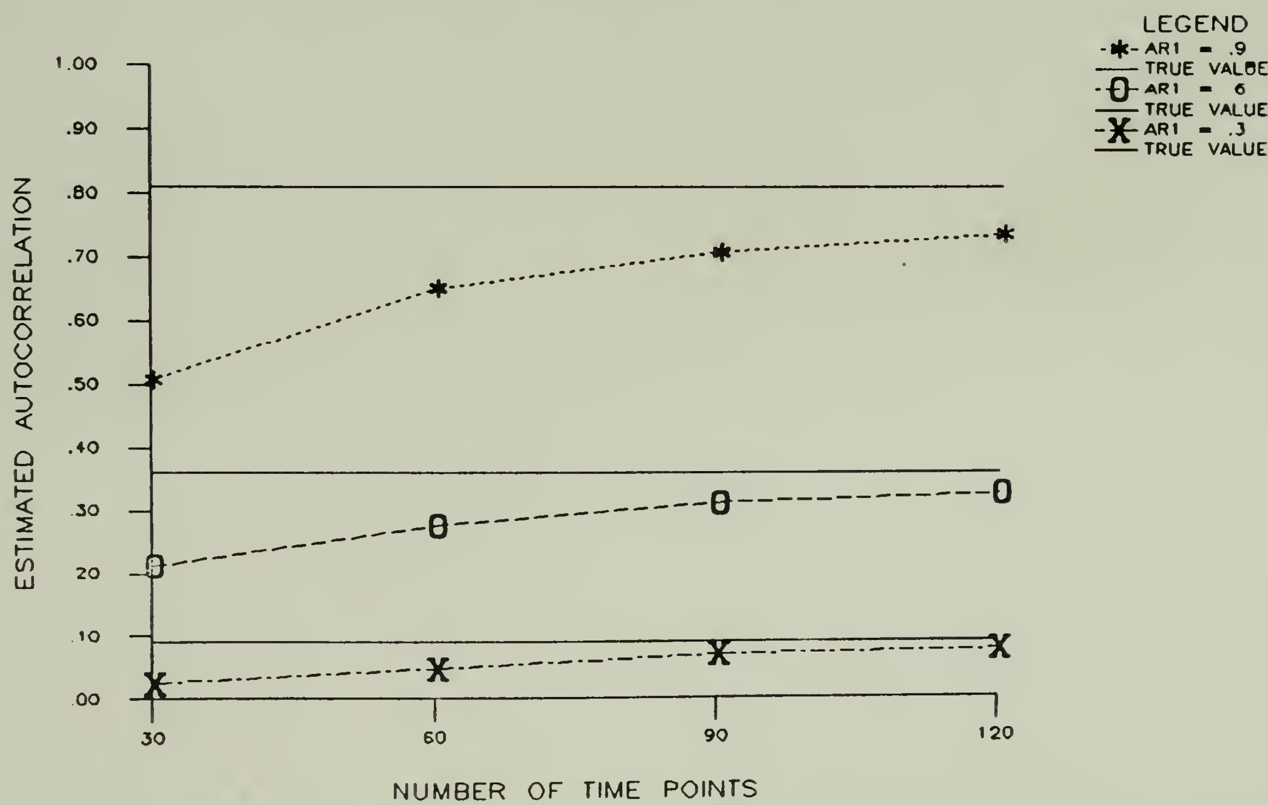
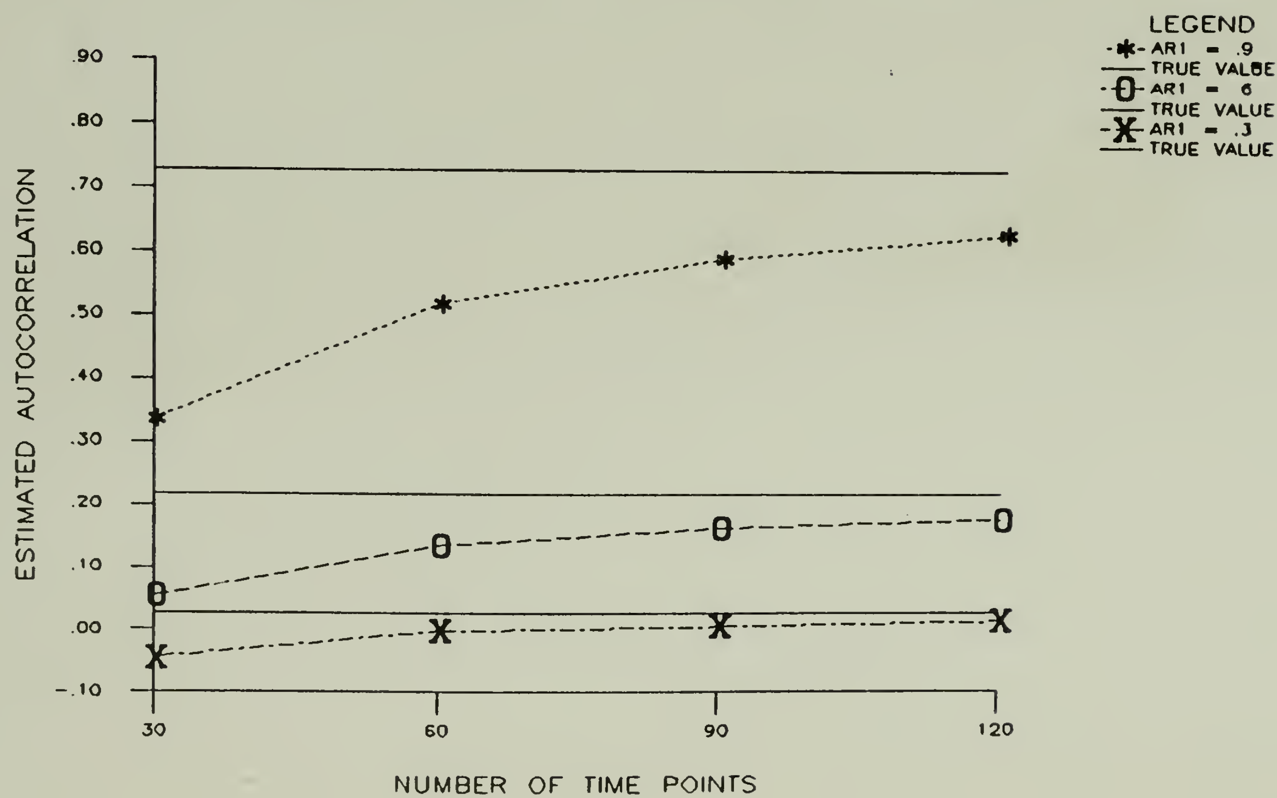


Figure 3

AR1 PROCESSES – LAG 3 AUTOCORRELATIONS



derived expression for an exact estimate of the standard error. The standard error at lag 1 is a function of ρ and n; the standard error becomes smaller as ρ and n become larger. In addition, autocorrelation estimates are more variable at larger lags. At lag 3, the relative magnitude of the standard error for the three values of AR(1) reverses, with the estimates of AR(1) = .9 demonstrating the most variablility. This result can be explained in terms of the formula for the approximate standard error of an autocorrelation coefficient, which is presented in Chapter II. 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, autocorrelation estimates will be more variable. Figures 4 through 6 show the estimated standard errors of the autocorrelation coefficients at lags 1, 2, and 3. In each case, the estimated standard error decreases rapidly as the number of time points becomes larger.

The small sample properties of the partial autocorrelation function at lags 2, 3, and 4 of the AR(1) process are presented in Table 2. The true value of the parameter for all conditions is, of course, zero. In all cases, the bias in the estimator results in a mean estimate

Figure 4

STANDARD ERROR OF ESTIMATED AUTOCORRELATIONS – LAG 1

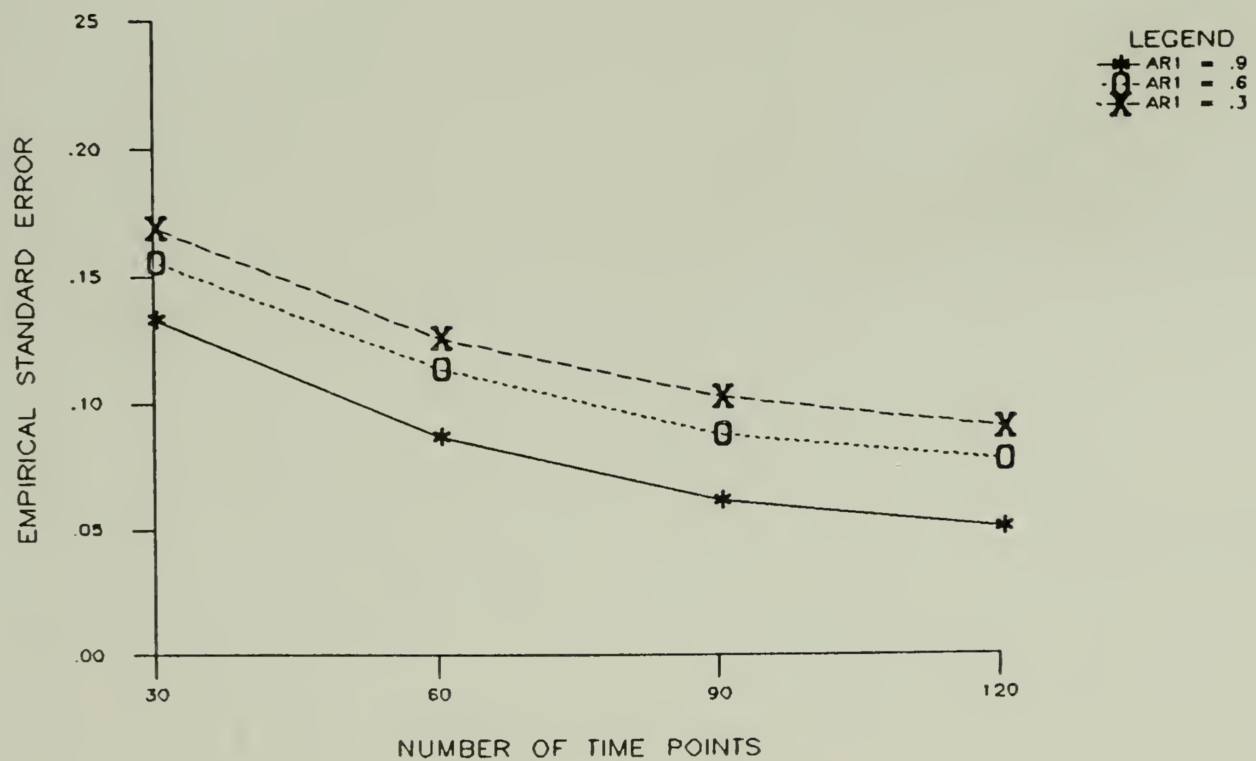


Figure 5

STANDARD ERROR OF ESTIMATED AUTOCORRELATIONS - LAG 2

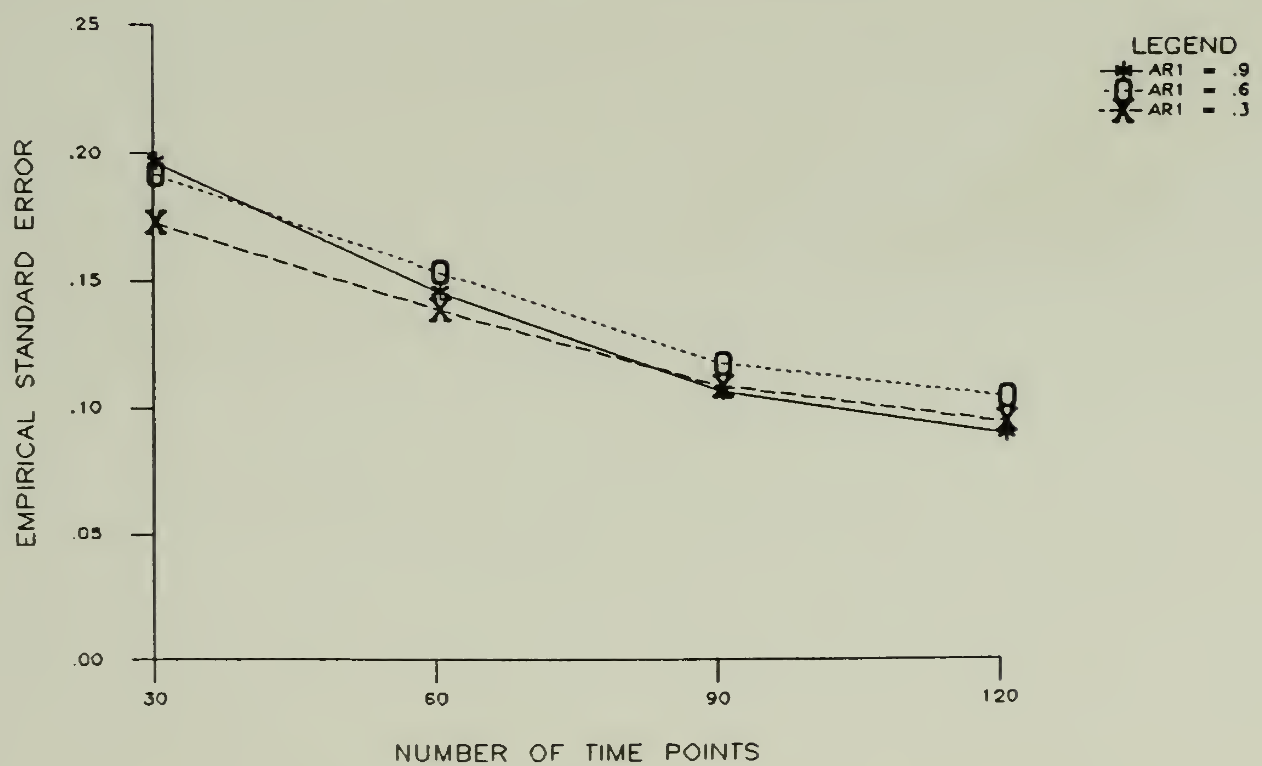


Figure 6

STANDARD ERROR OF ESTIMATED AUTOCORRELATIONS - LAG 3

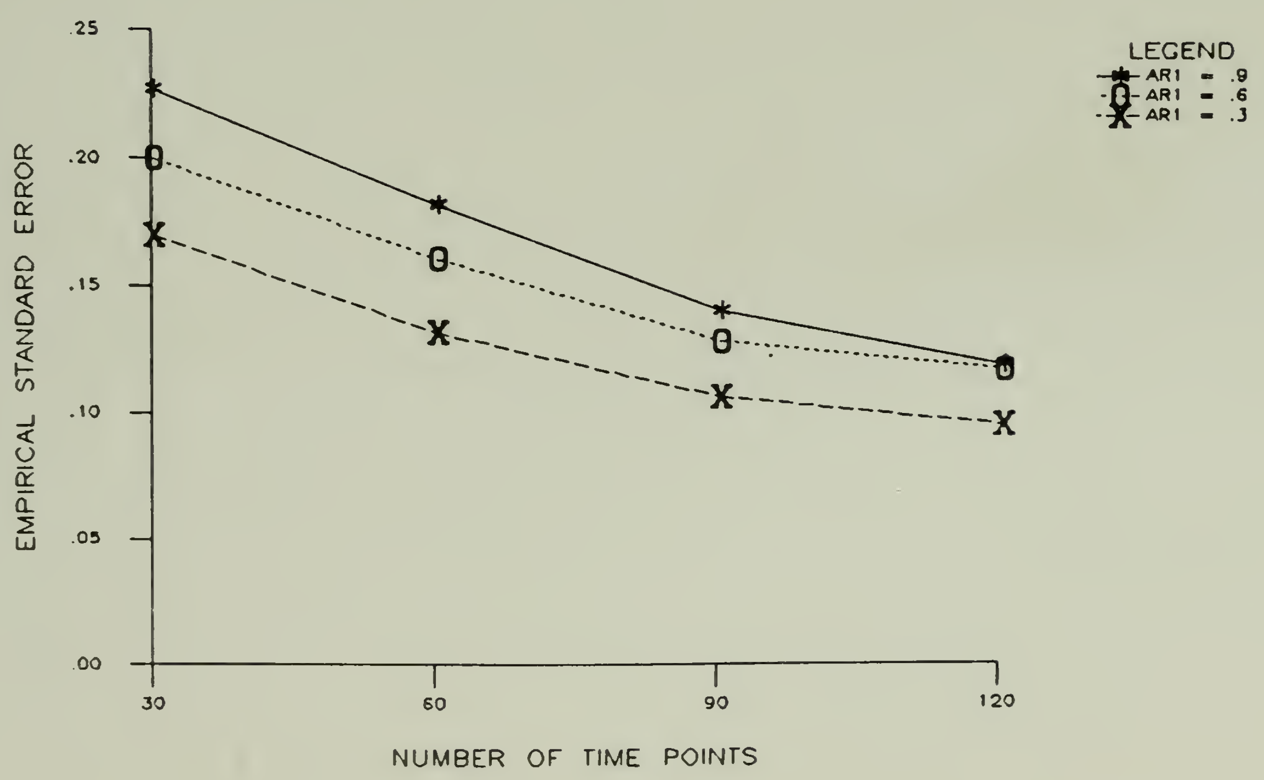


Table 2

First Order Autoregressive Processes

Mean Estimated Partial Autocorrelation (PAC) and Empirical Standard Error (SE)

AR1	AR2	TP	LAG2 PAC	LAG2 SE	LAG3 PAC	LAG3 SE	LAG4 PAC	LAG4 SE
.9	-	TRUE	.000	-	.000	-	.000	-
		120	-.018	.089	-.014	.083	-.022	.086
		90	-.030	.102	-.020	.098	-.025	.094
		60	-.038	.120	-.018	.121	-.043	.111
		30	-.081	.155	-.054	.150	-.069	.145
.6	-	TRUE	.000	-	.000	-	.000	-
		120	-.015	.092	-.010	.087	-.022	.088
		90	-.023	.104	-.014	.100	-.021	.100
		60	-.037	.127	-.012	.125	-.038	.120
		30	-.071	.169	-.049	.162	-.067	.153
.3	-	TRUE	.000	-	.000	-	.000	-
		120	-.014	.089	-.008	.089	-.021	.089
		90	-.021	.103	-.013	.099	-.020	.102
		60	-.039	.129	-.012	.124	-.037	.120
		30	-.064	.169	-.046	.164	-.069	.159

less than zero. The extent of the bias is negligible for all conditions in which the length of the series is greater than or equal to 60 time points. The underestimation of the partial autocorrelation coefficient is considerably larger for realizations of length 30, however. It can also be seen that the variablility of the partial autocorrelation estimates as the length of the realization becomes greater.

The next process considered is the ARIMA(2,0,0) model. One thousand time series realizations following the AR(2) model were generated based on three sets of autoregressive parameters; 1) AR1=.6, AR2=.3, 2) AR1=.5, AR2=.3, and 3) AR1=.4, AR2=.3. The mean of the 1000 estimates of the autocorrelation function for the first three lags and the standard deviation of these estimates are presented in Table 3.

The general pattern of results is similar to that of the AR(1) process. However, the downward bias of the estimates is greater and the standard error of the estimates tends to be slightly larger. Thus, the troublesome attributes of the estimated autocorrelation function are magnified for the second order autoregressive process. The bias of the estimated autocorrelation function of the AR(2) process is illustrated in Figures 7, 8, and 9, which plot the mean estimate over 1000

Table 3

Second Order Autoregressive Processes

Mean Estimated Autocorrelation (AC) and Empirical Standard Error (SE)

AR1	AR2	TP	LAG1 AC	LAG1 SE	LAG2 AC	LAG2 SE	LAG3 AC	LAG3 SE
.6	.3	ρ	.857	-	.814	-	.746	-
		120	.791	.083	.727	.098	.628	.126
		90	.768	.099	.695	.116	.586	.149
		60	.718	.132	.630	.153	.509	.185
		30	.601	.192	.482	.197	.318	.222
.5	.3	ρ	.714	-	.657	-	.543	-
		120	.661	.098	.592	.104	.459	.129
		90	.643	.112	.569	.118	.428	.145
		60	.598	.146	.516	.154	.373	.180
		30	.502	.202	.403	.193	.228	.214
.4	.3	ρ	.571	-	.529	-	.383	-
		120	.528	.110	.480	.103	.322	.124
		90	.513	.123	.462	.115	.299	.136
		60	.474	.156	.418	.149	.259	.168
		30	.395	.210	.333	.185	.147	.202

Figure 7

AR2 PROCESSES – LAG 1 AUTOCORRELATIONS

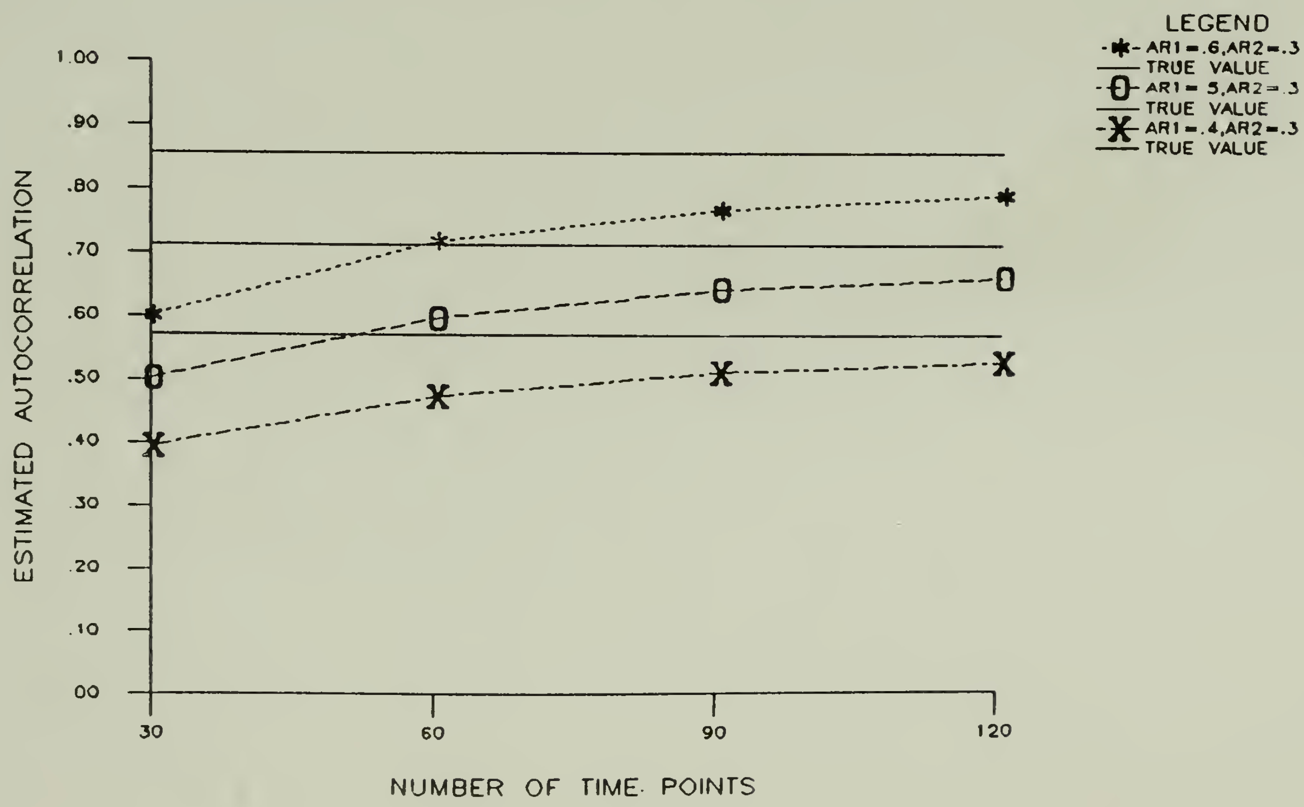


Figure 8

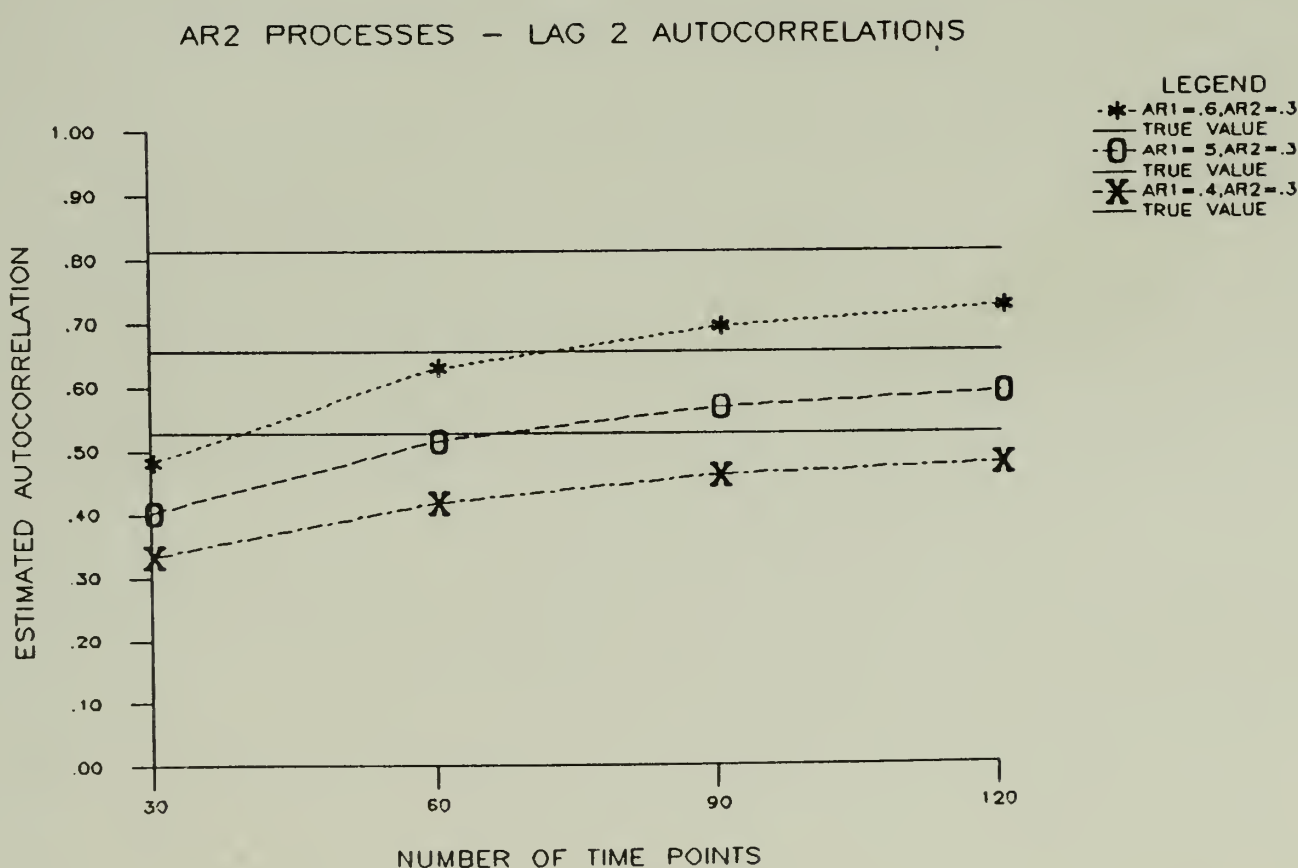
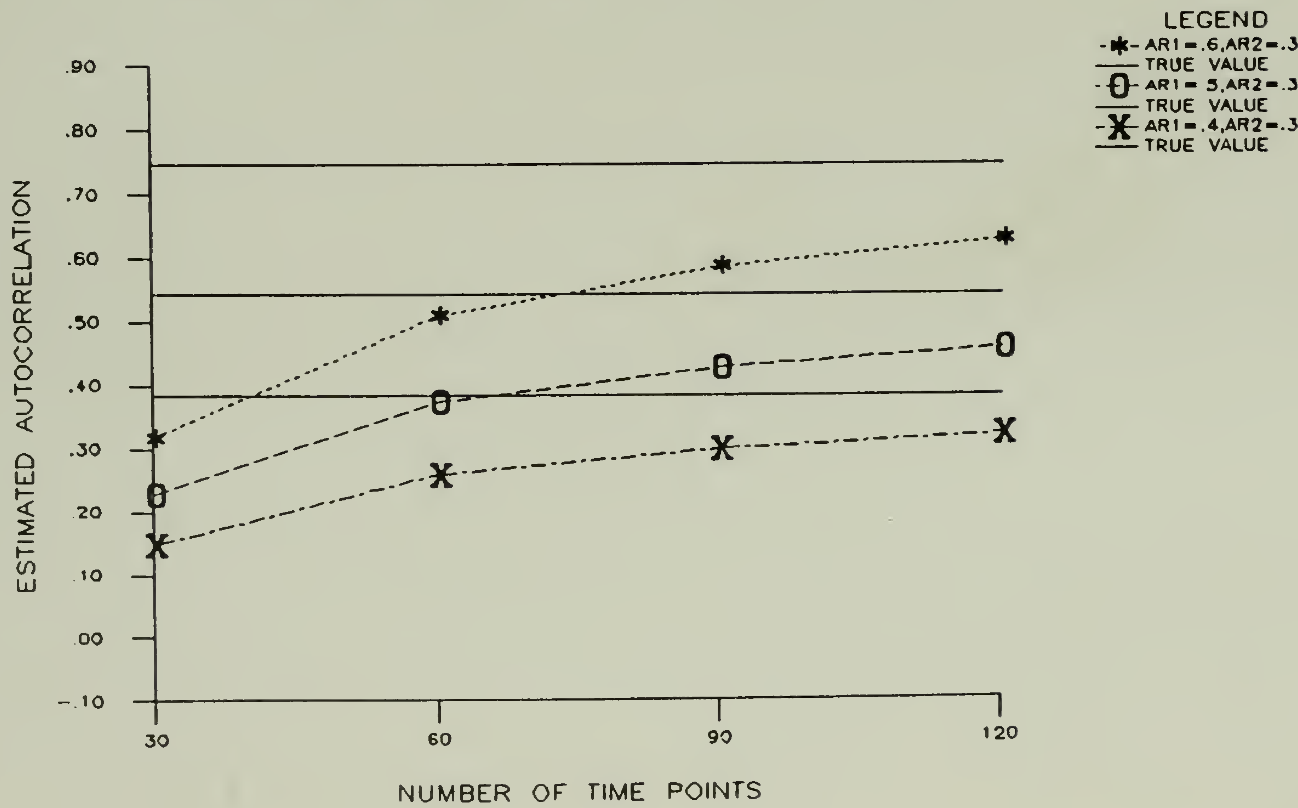

AR2 PROCESSES - LAG 2 AUTOCORRELATIONS
ESTIMATED AUTOCORRELATION
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
.00
30
60
90
120
NUMBER OF TIME POINTS
LEGEND
AR1=.6,AR2=.3
TRUE VALUE
AR1=.5,AR2=.3
TRUE VALUE
AR1=.4,AR2=.3
TRUE VALUE

Figure 9

AR2 PROCESSES – LAG 3 AUTOCORRELATIONS



replications for lag 1, lag 2, and lag 3, respectively. The empirical standard error of the estimates at each lag are presented graphically in Figures 10, 11, and 12.

The next table shows the mean and standard deviation over 1000 replications of the estimated partial autocorrelation function. Once again, the bias of the estimates is considerable for relatively short time series realizations and the standard error of the estimates is much larger than would be desirable.

Finally, the procedure followed for the AR(1) and AR(2) processes was repeated for first and second order moving average processes. One thousand time series realizations were generated for each of the conditions under consideration. The mean and standard deviation of the 1000 estimates of the autocorrelation and partial autocorrelations are presented in tables 5 and 6. The results are similar to those for the autoregressive processes, and will not be discussed further at this time.

In summary, study one examined the autocorrelation and partial autocorrelation functions of four different time series processes; AR(1), AR(2), MA(1), and MA(2). For each of the models, the length of the time series realization and the nature of the serial dependence of the observations was varied. The results of the study suggest that the proper identification of an ARIMA(p,d,q) process may be

Figure 10

STANDARD ERROR OF ESTIMATED AUTOCORRELATIONS – LAG 1

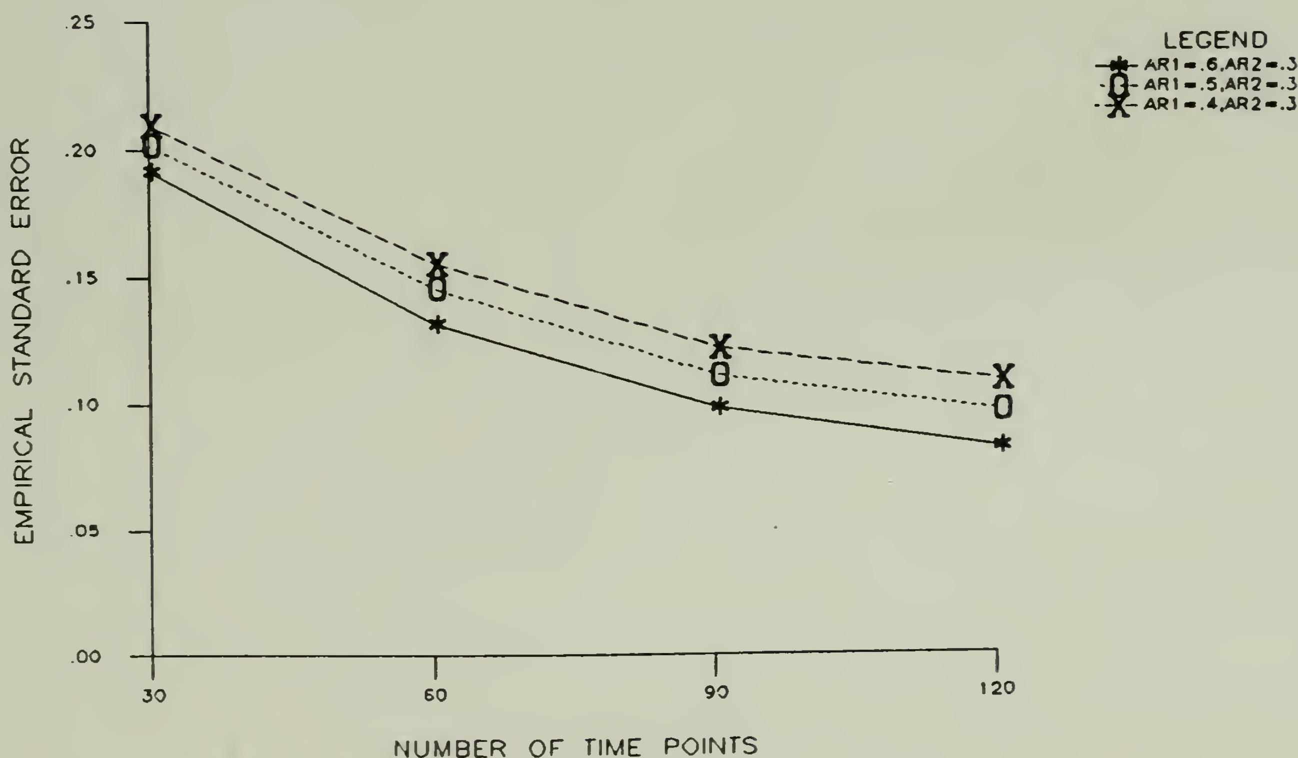


Figure 11

STANDARD ERROR OF ESTIMATED AUTOCORRELATIONS – LAG 2

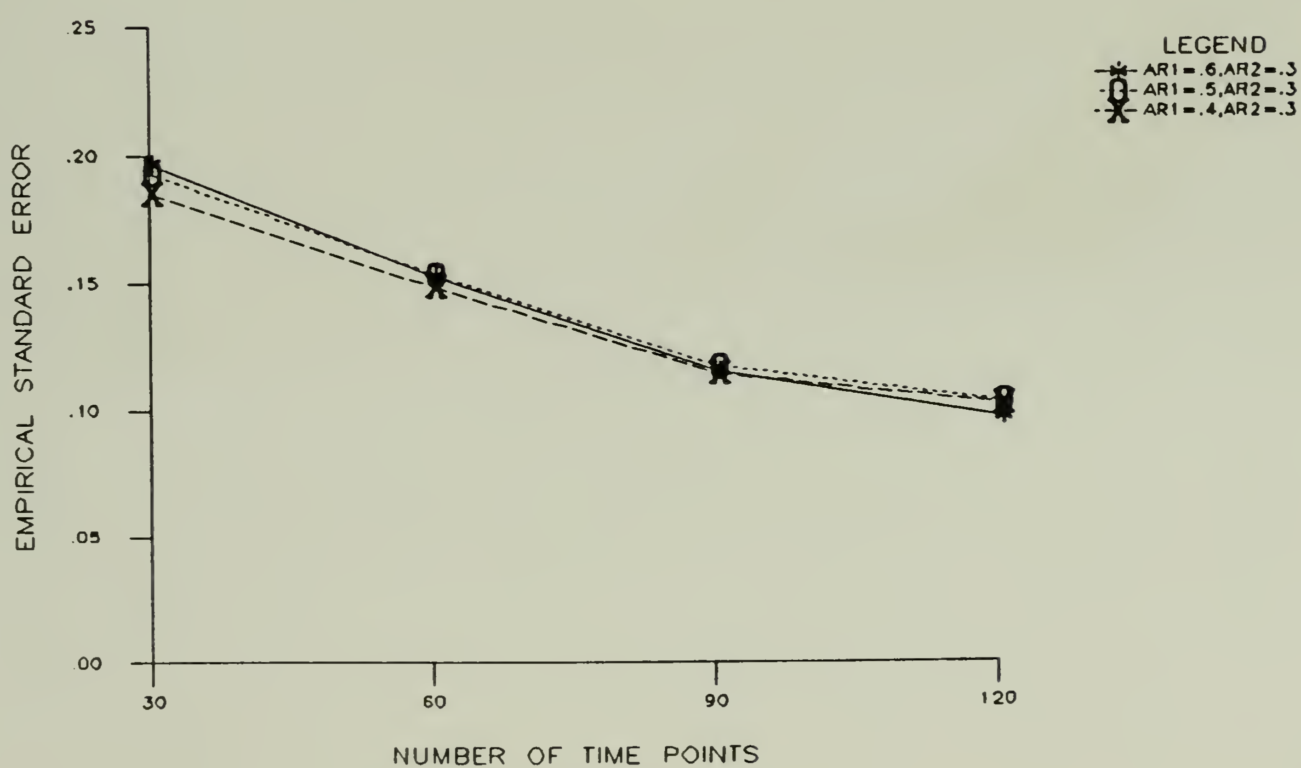


Figure 12

STANDARD ERROR OF ESTIMATED AUTOCORRELATIONS – LAG 3

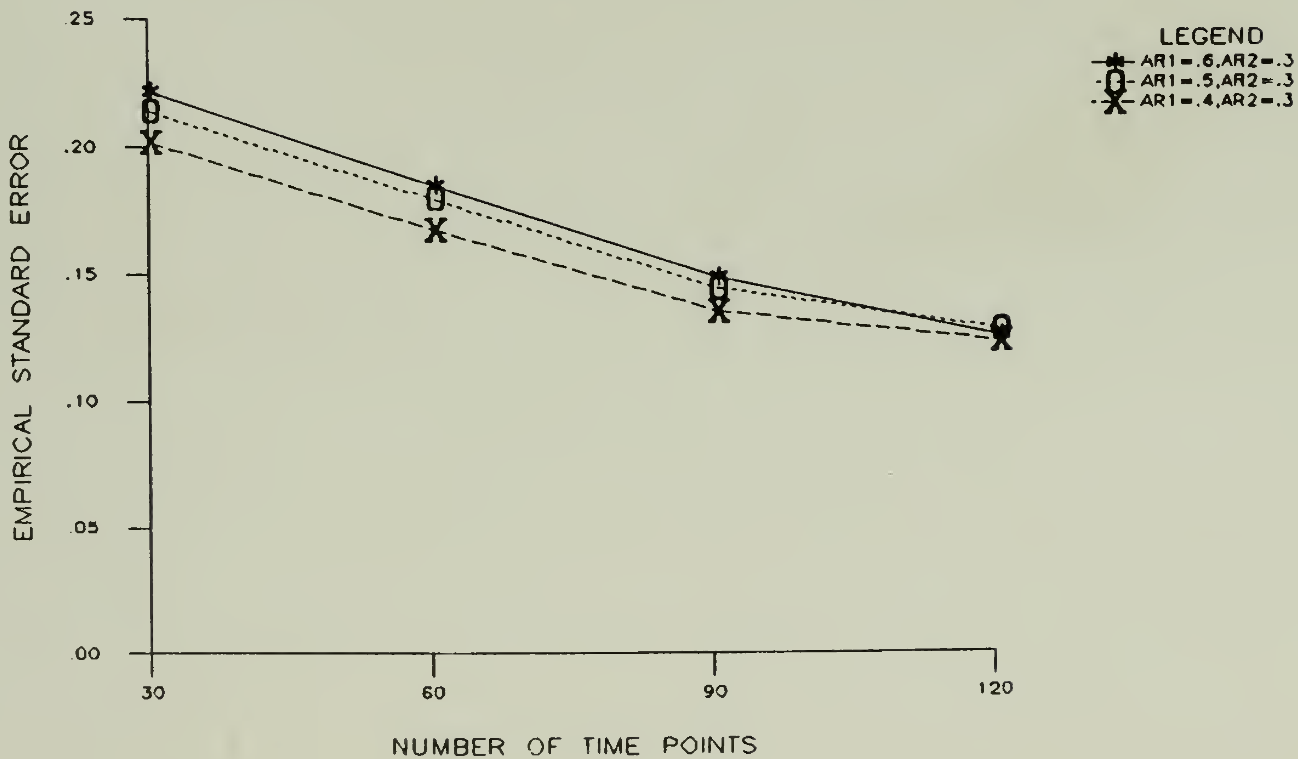


Table 4

Second Order Autoregressive Processes

Mean Estimated Partial Autocorrelation (PAC) and Empirical Standard Error (SE)

AR1	AR2	TP	LAG2 PAC	LAG2 SE	LAG3 PAC	LAG3 SE	LAG4 PAC	LAG4 SE
.6	.3	TRUE	.300	-	.000	-	.000	-
		120	.254	.089	-.017	.084	-.024	.087
		90	.237	.101	-.027	.097	-.027	.097
		60	.209	.122	-.023	.123	-.042	.113
		30	.136	.161	-.063	.151	-.074	.148
.5	.3	TRUE	.300	-	.000	-	.000	-
		120	.263	.090	-.013	.086	-.023	.088
		90	.248	.102	-.022	.099	-.025	.100
		60	.221	.124	-.018	.125	-.039	.117
		30	.154	.166	-.058	.155	-.074	.151
.4	.3	TRUE	.300	-	.000	-	.000	-
		120	.267	.089	-.011	.088	-.022	.088
		90	.254	.101	-.019	.099	-.024	.101
		60	.225	.126	-.015	.126	-.037	.119
		30	.164	.169	-.055	.158	-.073	.153

Table 5

First and Second Order Moving Average Processes

Mean Estimated Autocorrelation (AC) and Empirical Standard Error (SE)

MA1	MA2	TP	LAG1 AC	LAG1 SE	LAG2 AC	LAG2 SE	LAG3 AC	LAG3 SE
-.9	-		.497	-	.000	-	.000	-
		120	.479	.065	-.022	.109	-.020	.109
		90	.475	.075	-.031	.125	-.030	.123
		60	.465	.094	-.039	.152	-.031	.153
		30	.421	.132	-.084	.196	-.056	.187
-.6	-		.441	-	.000	-	.000	-
		120	.422	.069	-.020	.105	-.019	.104
		90	.419	.080	-.028	.120	-.029	.119
		60	.408	.099	-.035	.146	-.029	.148
		30	.364	.139	-.076	.188	-.053	.181
-.3	-		.275	-	.000	-	.000	-
		120	.257	.081	-.015	.096	-.015	.095
		90	.255	.094	-.021	.109	-.024	.110
		60	.246	.116	-.026	.133	-.024	.136
		30	.202	.159	-.056	.175	-.044	.169
-.6	-.3		.538	-	.207	-	.000	-
		120	.514	.068	.180	.099	-.029	.111
		90	.513	.080	.179	.113	-.032	.124
		60	.495	.097	.162	.140	-.042	.154
		30	.449	.148	.102	.187	-.087	.190
-.5	-.3		.485	-	.224	-	.000	-
		120	.460	.077	.199	.095	-.028	.108
		90	.459	.090	.198	.108	-.032	.120
		60	.440	.109	.181	.134	-.041	.149
		30	.395	.164	.123	.180	-.083	.184
-.4	-.3		.416	-	.240	-	.000	-
		120	.389	.087	.217	.090	-.027	.104
		90	.389	.102	.216	.104	-.031	.117
		60	.370	.123	.200	.128	-.040	.143
		30	.327	.182	.144	.173	-.078	.178

Table 6

First and Second Order Moving Average Processes

Mean Esimated Partial Autocorrelation (PAC)
and Empirical Standard Error (SE)

MA1	MA2	TP	LAG2 PAC	LAG2 SE	LAG3 PAC	LAG3 SE	LAG4 PAC	LAG4 SE
-.9	-	TRUE	-.328	-	.243	-	-.191	-
		120	-.333	.074	.216	.084	-.195	.081
		90	-.339	.086	.209	.091	-.194	.097
		60	-.338	.107	.201	.115	-.196	.105
		30	-.344	.141	.159	.144	-.205	.153
-.6	-	TRUE	-.242	-	.141	-	-.083	-
		120	-.248	.080	.119	.088	-.097	.085
		90	-.255	.091	.112	.097	-.097	.102
		60	-.256	.113	.107	.122	-.103	.114
		30	-.266	.151	.074	.153	-.126	.160
-.3	-	TRUE	-.082	-	.025	-	-.007	-
		120	-.094	.088	.012	.089	-.028	.089
		90	-.102	.100	.004	.102	-.031	.103
		60	-.108	.123	.004	.128	-.038	.121
		30	-.130	.168	-.014	.161	-.071	.161
-.6	-.3	TRUE	-.116	-				
		120	-.122	.093	-.099	.091	.066	.089
		90	-.125	.109	-.102	.103	.061	.098
		60	-.125	.136	-.099	.125	.040	.118
		30	-.160	.173	-.102	.166	-.006	.159
-.5	-.3	TRUE	-.015	-				
		120	-.025	.096	-.142	.088	.050	.089
		90	-.029	.113	-.144	.098	.046	.100
		60	-.034	.140	-.138	.121	.026	.119
		30	-.078	.178	-.137	.161	-.016	.160
-.4	-.3	TRUE	-.081	-				
		120	.067	.094	-.158	.086	.018	.091
		90	.062	.111	-.161	.095	.015	.101
		60	.053	.138	-.153	.118	-.003	.121
		30	.001	.178	-.151	.158	-.040	.162

difficult for relatively short realizations. Given the combination of severe bias and large standard errors of the estimated autocorrelation and partial autocorrelation functions, the accuracy of ARIMA models identified on the basis of fewer than 90 observations is questionable.

Study Two

A second tool used in the model identification stage of time series analysis is investigated in study two. The construction of approximate confidence intervals around the estimated autocorrelation and partial autocorrelation functions is often prescribed as a useful tool in the identification of ARIMA models. The first result of interest is the comparison of the estimated standard error of the autocorrelation function (which will be referred to as the estimated standard error) with the empirical standard error assessed as the standard deviation of the estimates over 1000 replications of the procedure (which will be referred to as the empirical standard error) . For each realization in the conditions under consideration, the estimated standard error of the autocorrelation function at lags 1 through 3 was computed using the formula previously discussed in the methodology section. The mean over 1000 replications was then computed to serve as a measure of central tendency of the distribution of the estimates. Table 7 presents the mean estimated standard errors for lags 1 through 3 with the empirical measure of the standard deviation of the estimated autocorrelations over 1000 replications.

The results for the three AR(1) processes presented in Table 7 show that the mean estimated standard error is always greater than the empirical standard error for this set of conditions. It is also apparent that the discrepancy between the estimated and empirical standard errors is largest for autoregressive processes with greater values of ρ . In addition, the discrepancy is related to the length of the time series realization. Figures 13, 14, and 15 graphically illustrate the results of Table 7 at lags 1, 2, and 3 respectively. It should be noted that the estimated standard errors at lag 1 do not vary (since the estimation procedure assumes that $\rho_1 = 0$), and thus the three conditions are represented by a single line.

The parallel results for three AR(2) processes are presented in Table 8. The results for lags 2 and 3 are similar to those of the AR(1) processes, with the estimated standard errors consistently larger than the empirical standard errors. In contrast to the AR(1) processes and the AR(2) results at lags 2 and 3, the results at lag 1 show that the estimated standard errors are consistently

Table 7

First Order Autoregressive Processes

Mean Estimated Standard Error (EST)
and Empirical Standard Error (EMP)

AR1	AR2	TP	LAG1 EST	LAG1 EMP	LAG2 EST	LAG2 EMP	LAG3 EST	LAG3 EMP
.9	-	120	.091	.051	.144	.090	.173	.120
		90	.105	.062	.164	.107	.196	.141
		60	.129	.087	.197	.146	.230	.182
		30	.183	.133	.263	.197	.296	.227
.6	-	120	.091	.078	.118	.105	.126	.118
		90	.105	.088	.135	.118	.144	.129
		60	.129	.114	.163	.154	.172	.161
		30	.183	.156	.224	.192	.235	.200
.3	-	120	.091	.091	.099	.095	.100	.096
		90	.105	.103	.114	.109	.116	.107
		60	.129	.126	.139	.139	.142	.132
		30	.183	.169	.197	.173	.202	.170

Figure 13

Empirical versus Approximate Estimates:

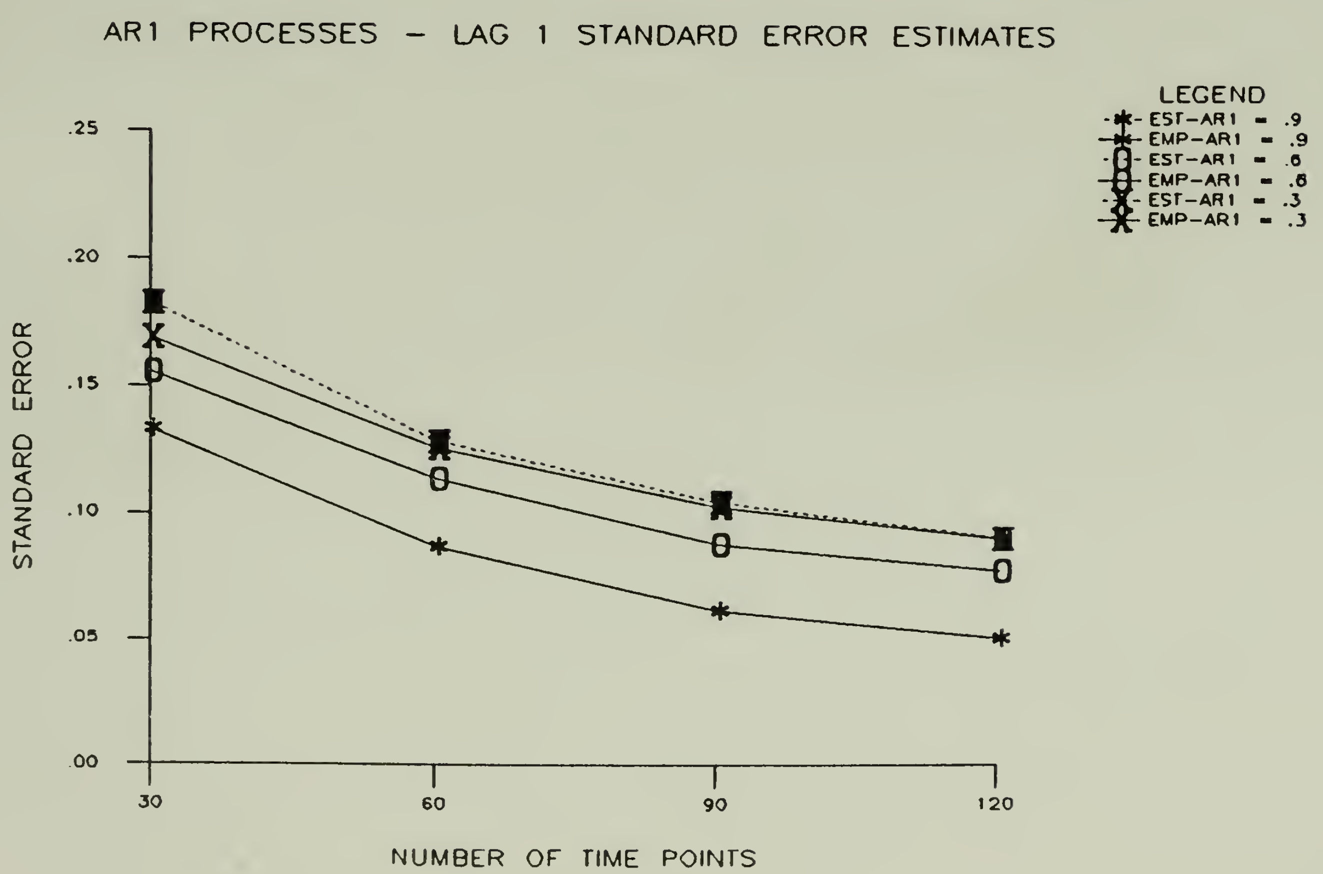


Figure 14

Empirical versus Approximate Estimates:

AR1 PROCESSES – LAG 2 STANDARD ERROR ESTIMATES

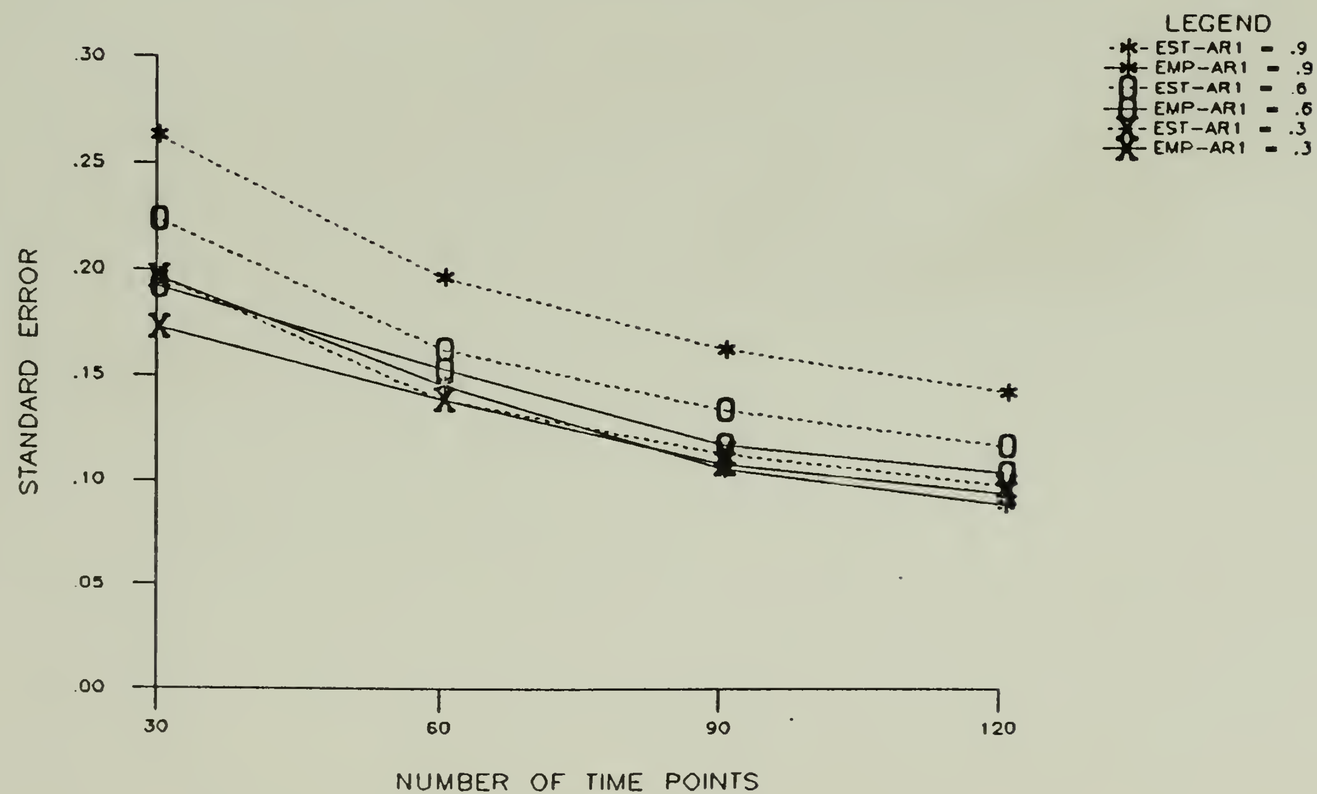


Figure 15

Empirical versus Approximate Estimates:

AR1 PROCESSES – LAG 3 STANDARD ERROR ESTIMATES

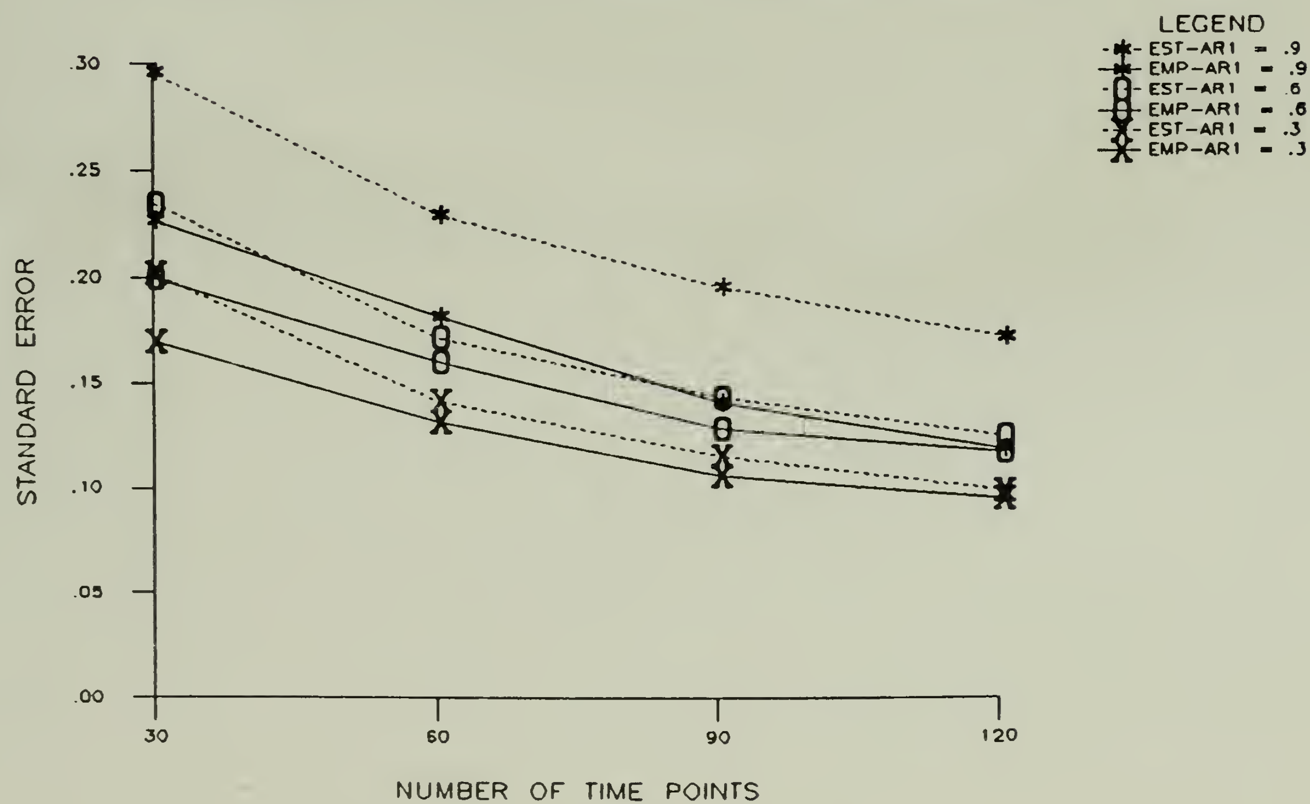


Table 8

Second Order Autoregressive Processes

Mean Estimated Standard Error (EST)
and Empirical Standard Error (EMP)

AR1	AR2	TP	LAG1 EST	LAG1 EMP	LAG2 EST	LAG2 EMP	LAG3 EST	LAG3 EMP
.6	.3	120	.091	.083	.137	.098	.166	.126
		90	.105	.099	.156	.116	.188	.149
		60	.129	.132	.185	.153	.218	.185
		30	.183	.192	.243	.197	.276	.222
.5	.3	120	.091	.098	.125	.104	.147	.129
		90	.105	.112	.143	.118	.167	.145
		60	.129	.146	.171	.154	.196	.180
		30	.183	.202	.228	.193	.254	.214
.4	.3	120	.091	.110	.115	.103	.131	.124
		90	.105	.123	.131	.115	.149	.136
		60	.129	.156	.157	.149	.176	.168
		30	.183	.210	.215	.185	.235	.202

smaller than the empirical standard errors. These results are graphically displayed in Figures 16, 17, and 18.

Table 9 presents the results of comparable simulations for moving average processes. An examination of the results shows that the estimated standard errors at those lags where the true autocorrelation coefficient is equal to 0 (lags 2 and 3 for MA(1) processes and lag 3 for MA(2) processes) are close to the empirical estimates of the standard error for relatively lengthy realizations ($n \geq 60$). On the other hand, the results for MA(1) processes at lag 1 and MA(2) processes at lags 1 and 2 demonstrate a tendency to over-estimate the standard error of the autocorrelation coefficients.

To briefly summarize, the results indicate that the procedure for estimating the approximate standard errors of the autocorrelation coefficients often over-estimates the actual magnitude of the standard error. The over-estimation is greatest for situations in which the autocorrelation coefficient is considerably different from zero. Those conditions for which the autocorrelation parameter is close to or equal to zero provide reasonably accurate estimates of the standard error of the parameter.

The results presented above have obvious implications for the use of approximate confidence intervals in the identification of time series processes. It is necessary to

Figure 16

Empirical versus Approximate Estimates:

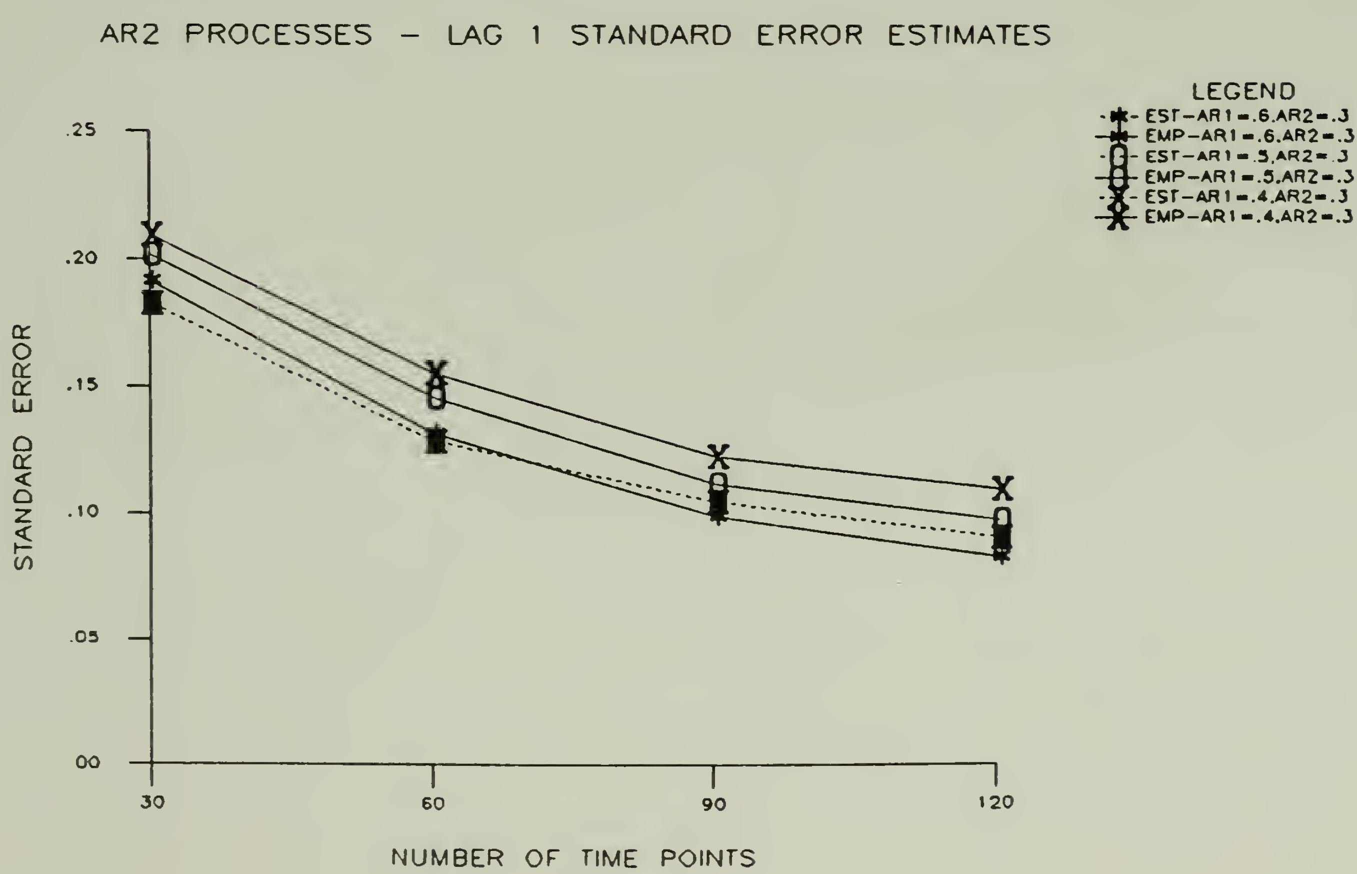


Figure 17

Empirical versus Approximate Estimates:

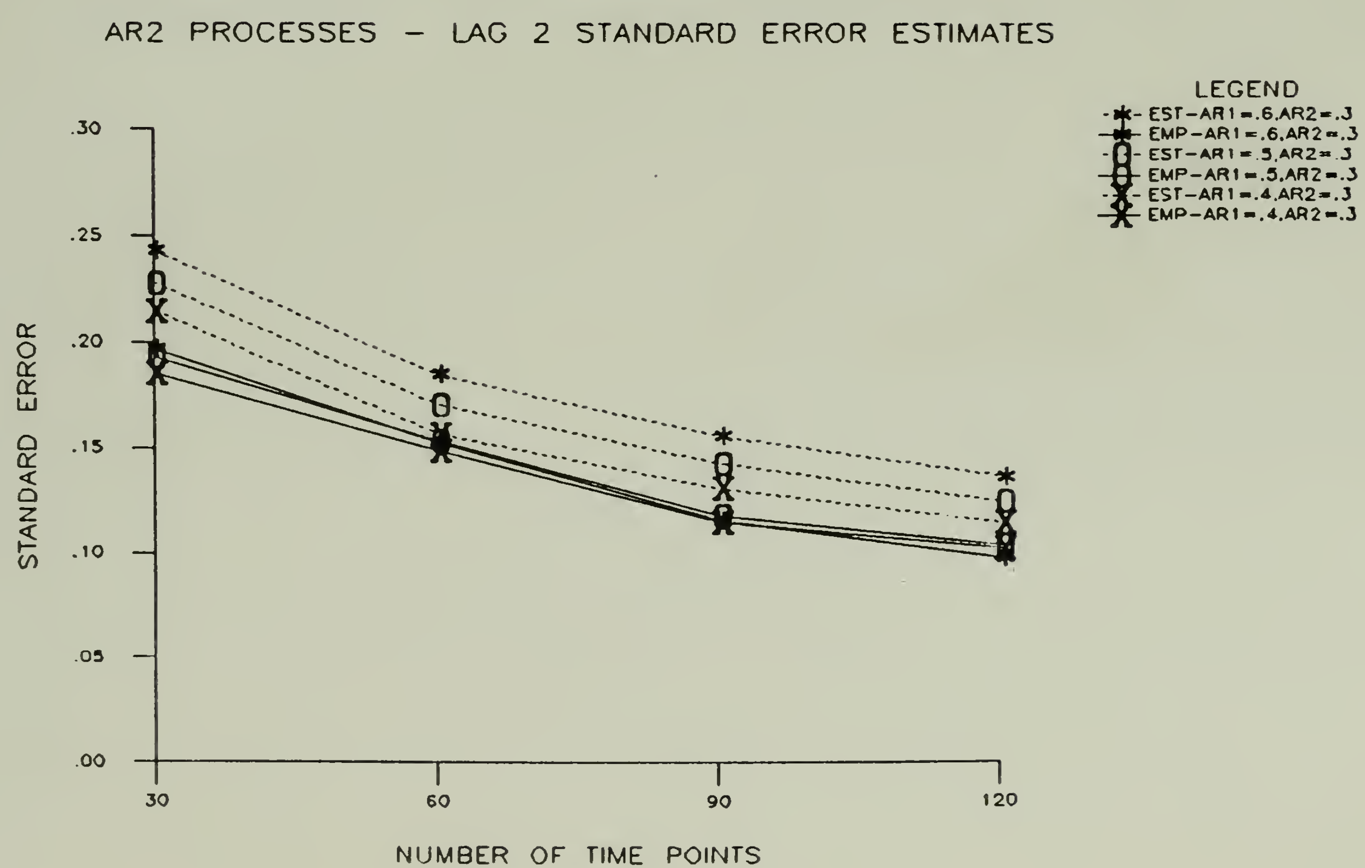


Figure 18

Empirical versus Approximate Estimates:

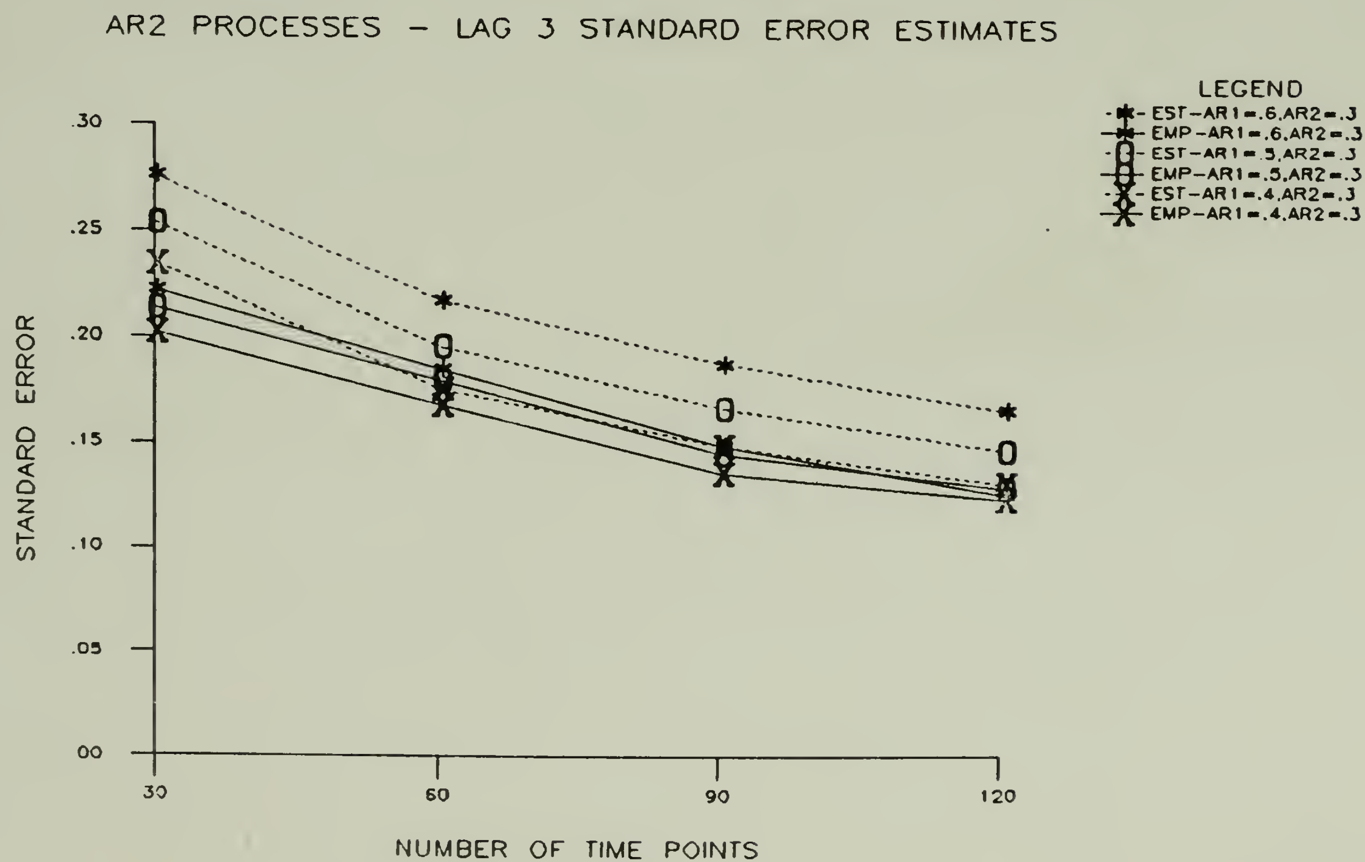


Table 9

First and Second Order Moving Average Processes

Mean Estimated Standard Error (EST)
and Empirical Standard Error (EMP)

MA1	MA2	TP	EST	EMP	EST	EMP	EST	EMP
-.9	-	120	.091	.065	.111	.109	.111	.108
		90	.105	.075	.127	.125	.129	.123
		60	.129	.094	.155	.152	.158	.153
		30	.183	.132	.215	.196	.222	.187
-.6	-	120	.091	.069	.107	.105	.108	.104
		90	.105	.080	.123	.120	.124	.119
		60	.129	.099	.150	.146	.152	.148
		30	.183	.139	.208	.188	.214	.181
-.3	-	120	.091	.081	.098	.096	.098	.095
		90	.105	.094	.113	.109	.114	.110
		60	.129	.116	.138	.133	.140	.136
		30	.183	.159	.194	.175	.200	.169
-.6	-.3	120	.091	.068	.113	.099	.116	.111
		90	.105	.080	.131	.113	.134	.124
		60	.129	.097	.158	.140	.163	.154
		30	.183	.148	.219	.187	.225	.190
-.5	-.3	120	.091	.087	.105	.090	.109	.104
		90	.105	.102	.121	.104	.126	.117
		60	.129	.123	.147	.128	.153	.144
		30	.183	.182	.206	.173	.213	.178
-.4	-.3	120	.091	.087	.105	.090	.109	.104
		90	.105	.102	.121	.104	.126	.117
		60	.129	.123	.147	.128	.153	.144
		30	.183	.182	.206	.173	.213	.178

assume that the parameter is equal to zero in order to construct confidence intervals around the hypothetical autocorrelation parameter of zero. If the parameter is indeed zero, the approximate confidence intervals are relatively accurate. In contrast, non-zero autocorrelation parameters lead to standard error estimates which deviate from the true standard error; in most circumstances, the estimated standard errors are larger than the true values. It follows that confidence bands around parameters equal to zero will result in a true type one error rate, while the confidence intervals for non-zero parameters will often be too wide and result in a large Type II error rate.

The next part of study two systematically examines the Type I and Type II error rates of the test of significance of the autocorrelation parameters. The test statistic is based on the 95% confidence interval around the zero value of the autocorrelation coefficient. For each replication, estimated autocorrelations were substituted for the population parameters that are assumed to be known in the estimation procedure. The percentage of replications for which the estimate was outside of the 95% confidence interval were calculated, resulting in a Type I error rate for conditions in which the true autocorrelation is equal to zero, and a power estimate (1 - Type II error rate) for those conditions with a non-zero population parameter.

Type I error rates and power estimates were also assessed for the partial autocorrelation coefficients using the same procedure with the standard error estimate of the partial autocorrelation coefficient.

The results of the Monte Carlo simulations for AR(1) and AR(2) processes are presented in Table 10. The population autocorrelation function is greater than zero for lags 1, 2, and 3, as shown in the rows labeled "ρ ". Thus, the rejection of the null hypothesis H_0: $\rho = 0$ is a measure of the power of the procedure to detect a significant autocorrelation parameter. It can be seen that the power of this test varies considerably with the length of the time series realization. The percentage of replications for which the test was able to detect a non-zero autocorrelation parameter is also obviously related to the magnitude of the parameter. As the lag increases, the population parameter of an AR(1) or AR(2) process becomes smaller and more difficult to detect. These results are also presented graphically in Figures 19 through 24.

The percentage of null hypotheses rejected for the test of the partial autocorrelation coefficients are presented in the last three columns of Table 10. The population parameters are zero for all cases other than the partial autocorrelation coefficient at lag 2 of the AR(2)

Table 10

First and Second Order Autoregressive Processes

Percentage of Null Hypotheses Rejected (p<.05) for Three Lags of Estimated Autocorrelations (AC) and Partial Autocorrelations (PC)

AR1	AR2	TP	AC1	AC2	AC3	PC2	PC3	PC4
.9	-	TRUE	.900	.810	.729	.000	.000	.000
		120	1.000	1.000	.994	.045	.032	.043
		90	1.000	.999	.943	.044	.037	.031
		60	1.000	.966	.715	.043	.039	.034
		30	.980	.543	.027	.039	.032	.028
.6	-	TRUE	.600	.360	.216	.000	.000	.000
		120	1.000	.839	.268	.049	.042	.052
		90	1.000	.679	.156	.062	.042	.033
		60	.987	.401	.068	.048	.039	.050
		30	.805	.082	.008	.044	.036	.031
.3	-	TRUE	.300	.090	.027	.000	.000	.000
		120	.875	.106	.042	.041	.049	.049
		90	.766	.082	.032	.057	.043	.045
		60	.537	.050	.031	.066	.046	.051
		30	.234	.021	.014	.040	.034	.035
.6	.3	TRUE	.857	.814	.746	.300	.000	.000
		120	1.000	1.000	.994	.807	.039	.045
		90	1.000	.999	.950	.621	.045	.040
		60	.997	.966	.736	.381	.051	.032
		30	.895	.572	.036	.088	.034	.026
.5	.3	TRUE	.714	.657	.543	.300	.000	.000
		120	1.000	.999	.941	.834	.038	.045
		90	1.000	.994	.804	.670	.050	.041
		60	.981	.910	.518	.422	.048	.039
		30	.773	.421	.019	.106	.032	.030
.4	.3	TRUE	.571	.529	.383	.300	.000	.000
		120	.995	.990	.747	.840	.042	.049
		90	.989	.969	.531	.691	.052	.043
		60	.898	.809	.284	.435	.049	.035
		30	.613	.304	.010	.122	.031	.034

Figure 19

POWER OF AUTOCORRELATION TEST – LAG 1

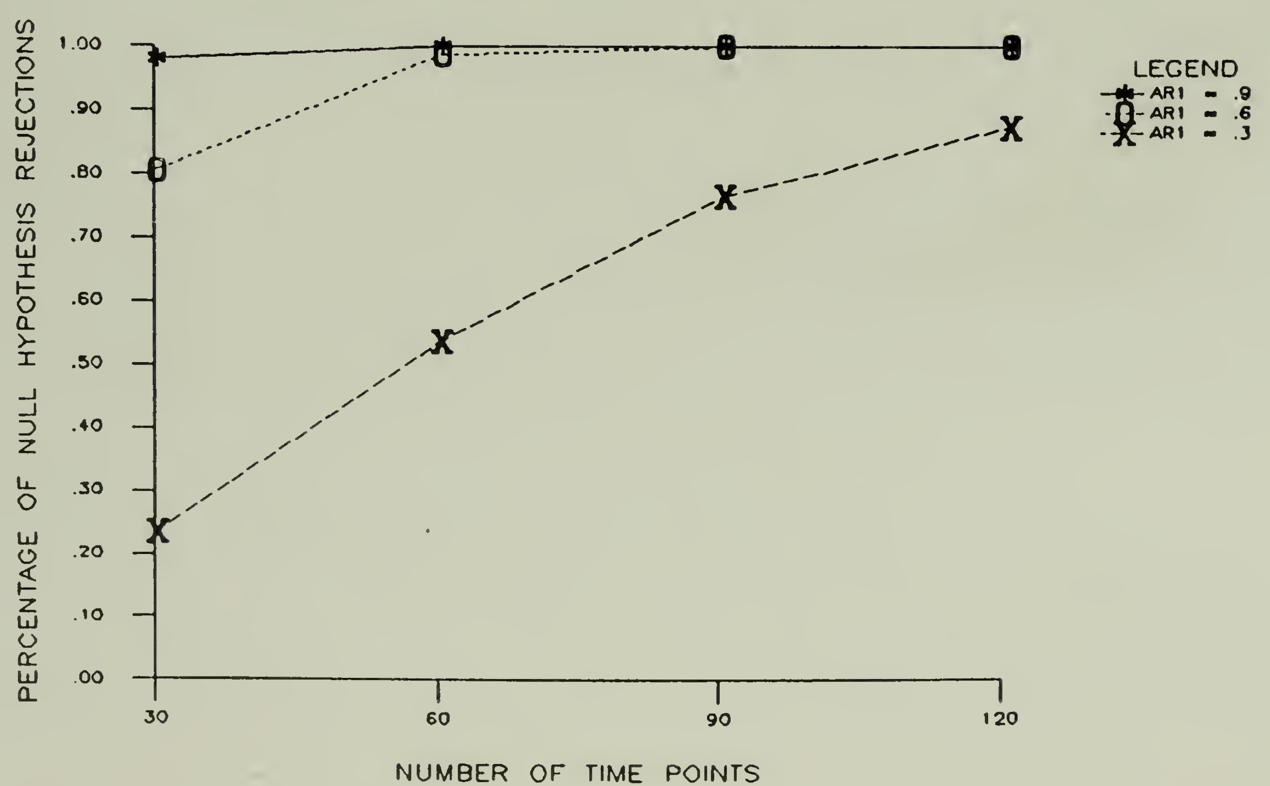


Figure 20

POWER OF AUTOCORRELATION TEST – LAG 2

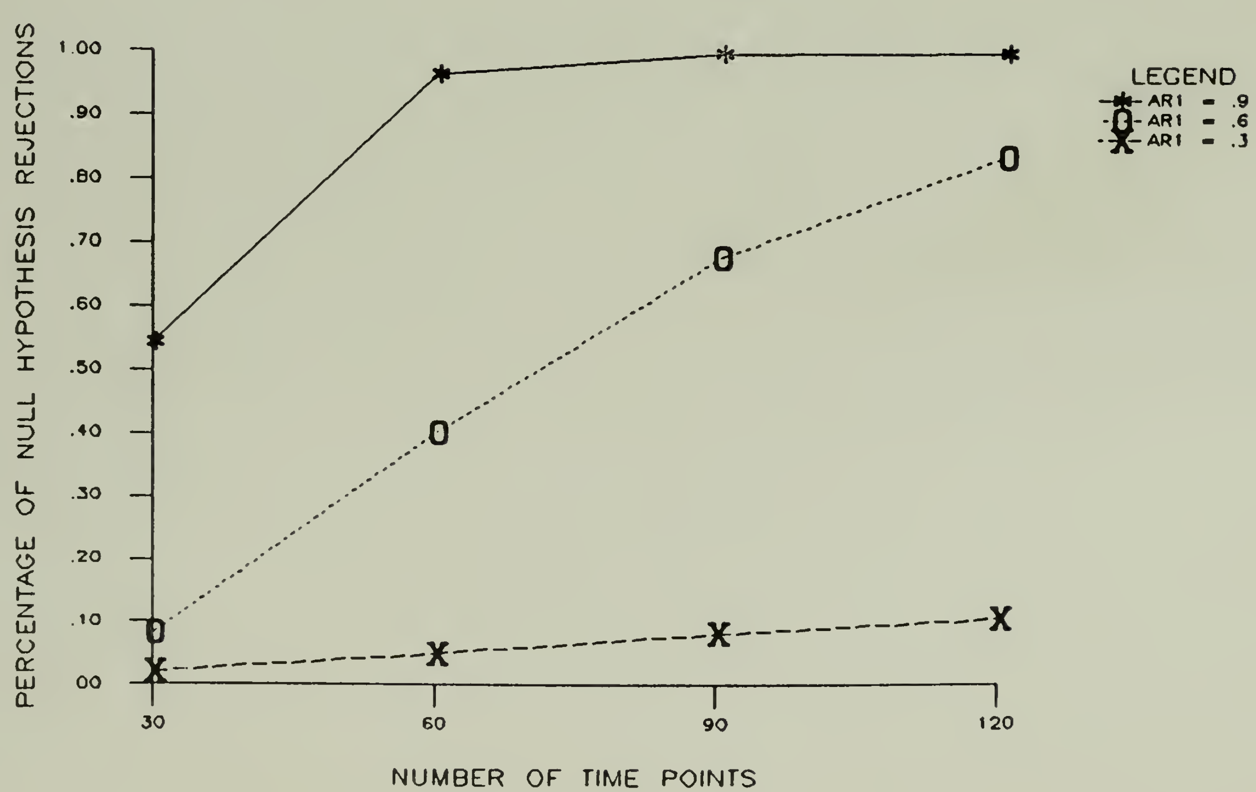


Figure 21

POWER OF AUTOCORRELATION TEST — LAG 3

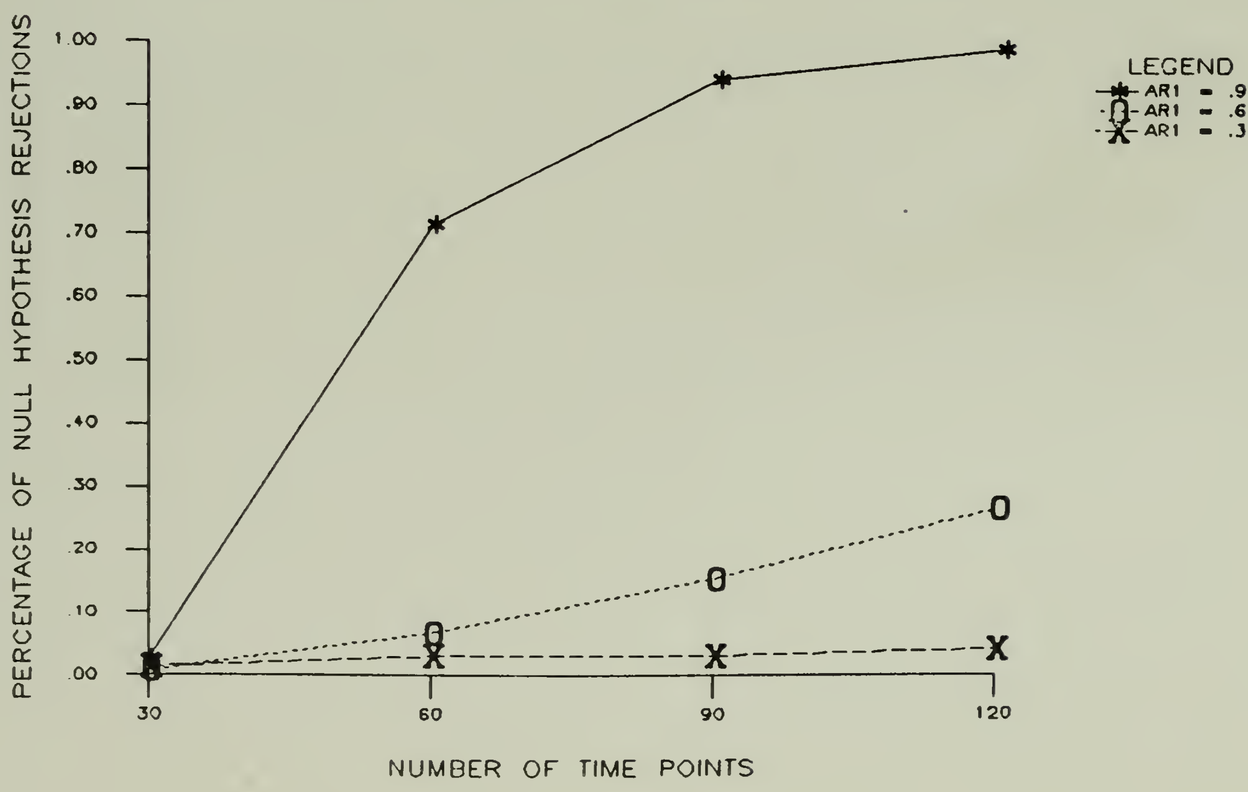


Figure 22

POWER OF AUTOCORRELATION TEST — LAG 1

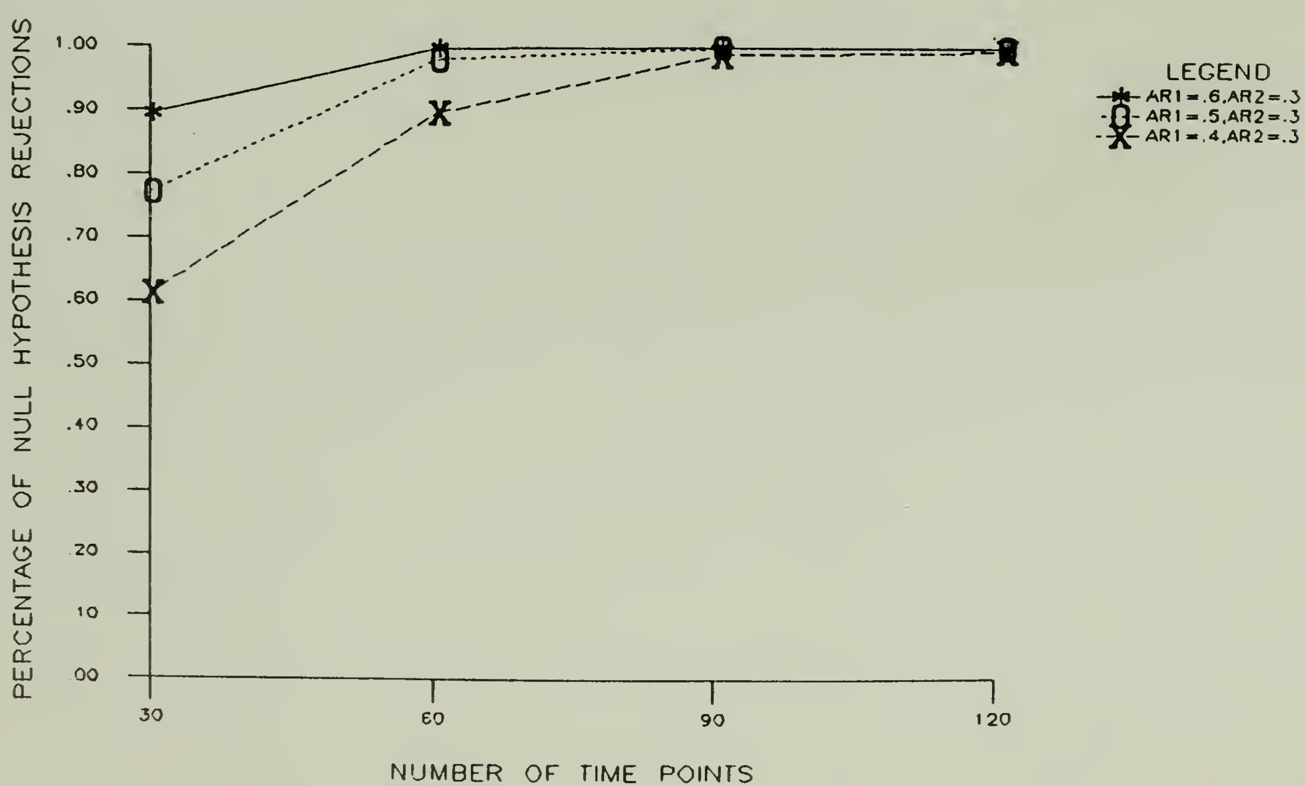


Figure 23

POWER OF AUTOCORRELATION TEST — LAG 2

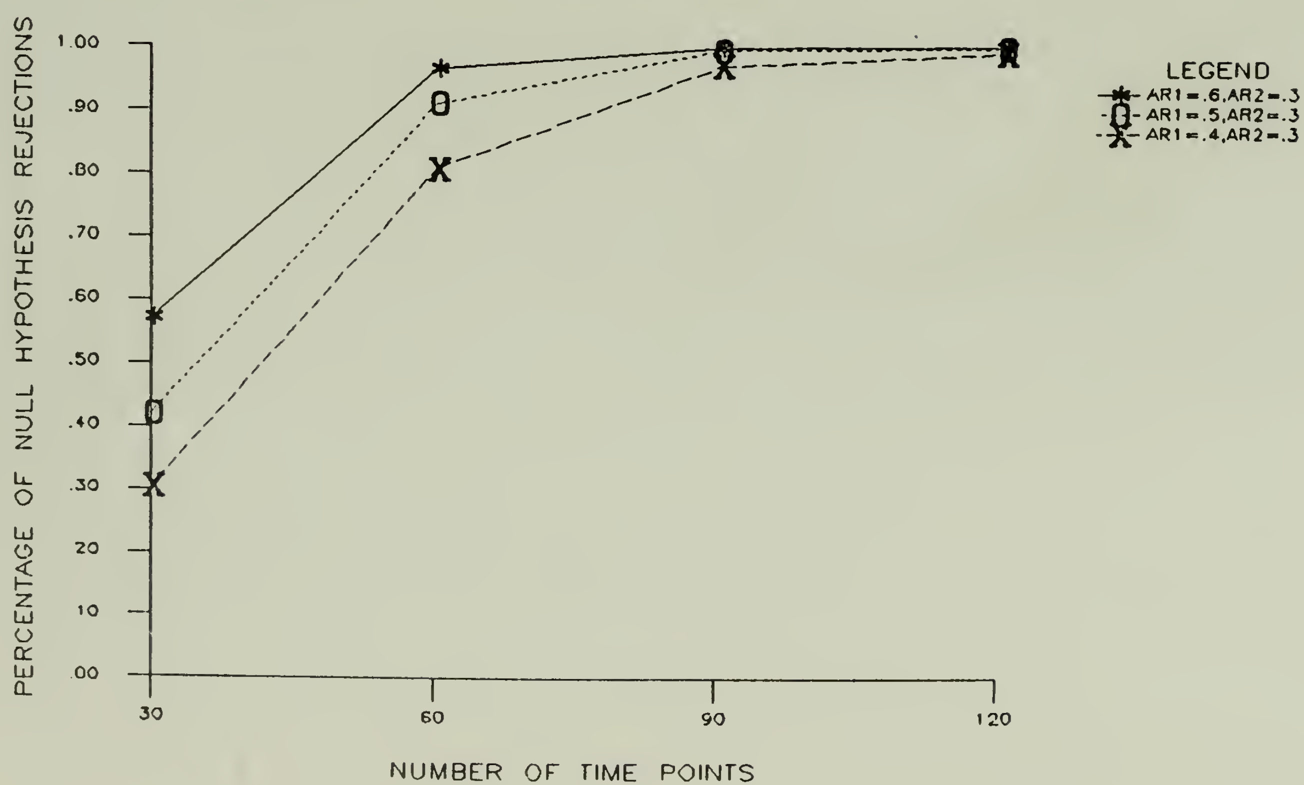
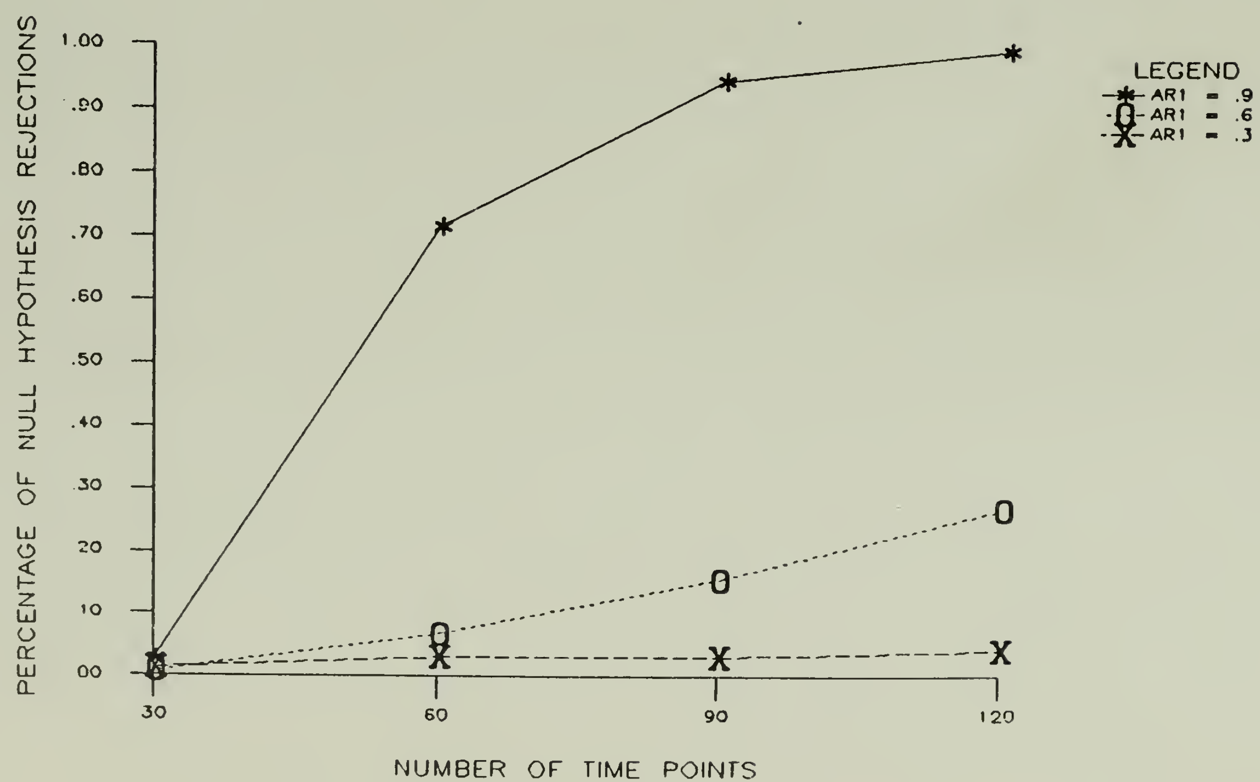


Figure 24

POWER OF AUTOCORRELATION TEST – LAG 3



processes. It can be seen that the type I error rates in these conditions are very close to the nominal value of p=.05, with a slight tendency towards a deflated type I error rate.

The power of the statistical test of the partial autocorrelation coefficient is examined for the AR(2) processes. The length of the realization has a tremendous effect on the power to reject the null hypothesis that the lag 2 partial autocorrelation coefficient is different from zero. This relationship is displayed in Figure 25.

Table 11 presents the corresponding results for moving average processes. In this situation, the population autocorrelation coefficient at lags greater than the number of moving average parameters is equal to zero. It is apparent that the empirical Type I error rates in these conditions are relatively close to the nominal p=.05 level. The power to detect true differences from zero at lag 1 is very good for time series realizations of length 60 or greater. For the MA(2) processes however, the ability to detect non-zero autocorrelations at lag 2 is poor, even for realizations of 120 time points.

The practical implications of insufficient power should be pointed out at this time. For first order autoregressive processes, the theoretical autocorrelation function is characterized by an exponential rate of decay.

Figure 25

POWER OF PARTIAL AUTOCORRELATION TEST – LAG 2

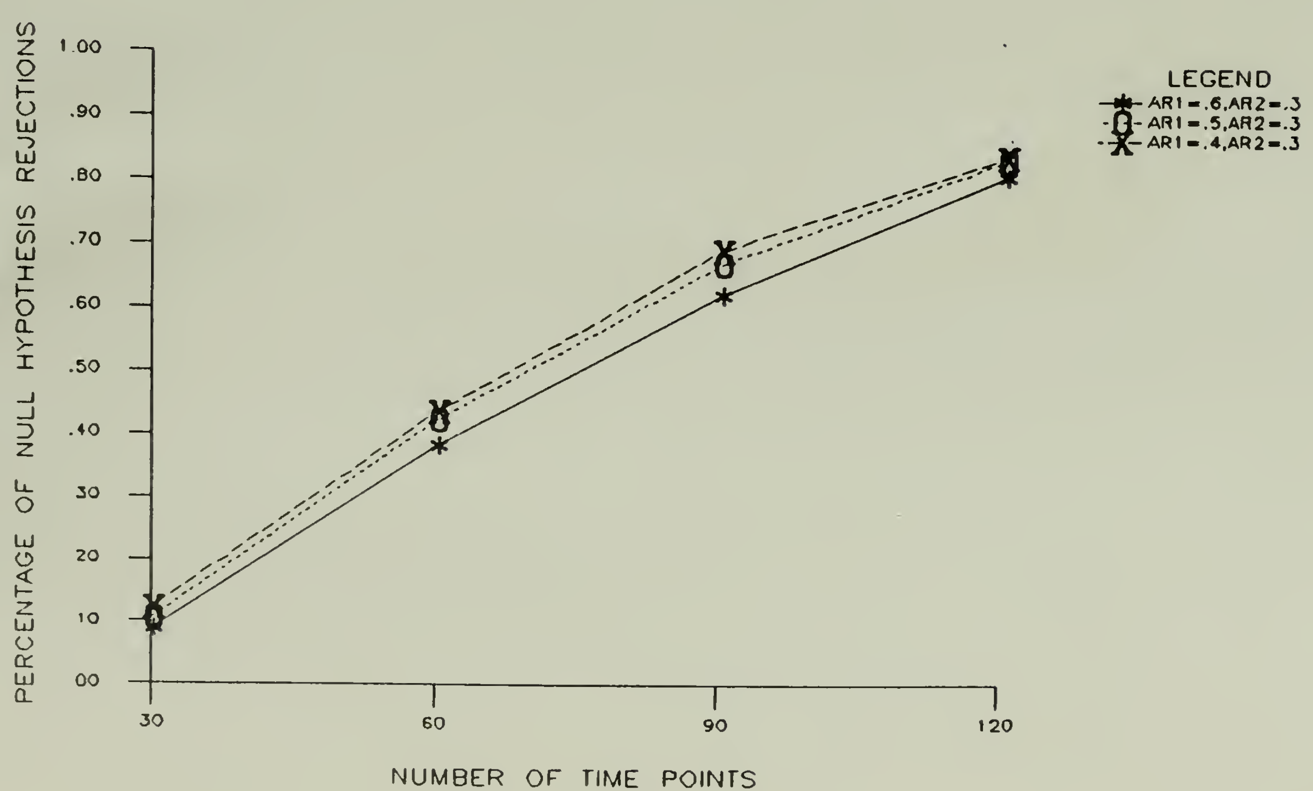


Table 11

First and Second Order Moving Average Processes

Percentage of Null Hypotheses rejected (p<.05) for Three Lags of Autocorrelation Estimates (AC) and Partial Autocorrelation Estimates (PC)

MA1	MA2	TP	AC1	AC2	AC3	PC2	PC3	PC4
-.3	-	120	.831	.045	.045	.170	.043	.046
		90	.697	.051	.044	.156	.044	.053
		60	.487	.044	.036	.123	.048	.049
		30	.165	.038	.018	.084	.026	.037
-.6	-	120	1.000	.057	.044	.815	.257	.172
		90	.992	.053	.040	.700	.164	.138
		60	.934	.052	.035	.511	.114	.099
		30	.552	.054	.019	.287	.033	.072
-.9	-	120	1.000	.059	.047	.982	.665	.585
		90	.998	.054	.038	.937	.521	.466
		60	.984	.058	.035	.794	.345	.299
		30	.716	.064	.018	.473	.091	.163
-.4	-.3	120	.988	.563	.060	.111	.416	.059
		90	.952	.429	.042	.104	.318	.043
		60	.831	.237	.055	.078	.220	.037
		30	.473	.059	.040	.043	.088	.026
-.5	-.3	120	.999	.450	.062	.071	.342	.078
		90	.991	.330	.038	.069	.280	.057
		60	.946	.174	.058	.080	.178	.040
		30	.628	.038	.041	.056	.079	.027
-.6	-.3	120	1.000	.336	.058	.277	.190	.104
		90	.999	.236	.037	.241	.159	.071
		60	.990	.118	.061	.172	.115	.045
		30	.750	.028	.040	.122	.055	.020

The failure to consider autocorrelation coefficients at lags 2 and beyond as different from zero may result in a misidentification of an AR(1) process as a MA(1) process. The apparent form of the autocorrelation function would be a truncation after lag 1 rather than an exponential decay. One of the keys to identifying an AR(2) process is recognizing the partial autocorrelation at lag 2 as being different from zero, and thus a lack of power in testing this coefficient may result in the misidentification of AR(2) processes. In short, the researcher must make accurate judgments about form of the autocorrelation and partial autocorrelation functions in order to properly identify a time series process. A lack of power in testing whether the autocorrelation and partial autocorrelation parameters are different from zero can be a serious impediment to proper model identification.

Study Three

This study considers the practical implications of applying the unbiased estimator of the autocorrelation function proposed by Quenouille (1956). This procedure for estimating the autocorrelation function is designed to remove the severe bias of the ordinary estimator of the autocorrelation function that occurs when time series realizations are relatively short in length. Although the estimator is unbiased, two other considerations are necessary in evaluating the relative desirability of the estimator. First, unlike the usual estimator of the autocorrelation function, estimates may be obtained that are theoretically impossible. Secondly, it is very important to consider the magnitude of standard error of the estimator.

The first result of this set of Monte Carlo simulations addresses the issue of estimates outside the theoretical bounds of the parameter. It is of interest to examine the frequency of such estimates under a variety of conditions. The results presented below (Table 12) indicate that estimated autocorrelation coefficients greater than 1.0 are quite common for autoregressive processes with a high degree of serial correlation. As would be expected, shorter realizations are more likely to produce such estimates than are longer realizations. A set of simulations conducted for moving average processes showed that the incidence of estimates greater than 1.0 was very small, since the expected values of the autocorrelation coefficients are smaller than those of the autoregressive processes considered.

Next, the empirical standard errors of the unbiased estimator are compared with those of the usual biased

Table 12

Unbiased Estimates of Autoregressive Processes

Percentage of Autocorrelation Estimates > 1.0

AR1	AR2	TP	AC1	AC2	AC3	AC4	AC5
.9	-	120	.075	.061	.054	.047	.041
		90	.125	.107	.086	.071	.062
		60	.199	.173	.143	.121	.106
		30	.327	.252	.189	.130	.062
.6	-	120	.000	.000	.000	.000	.000
		90	.000	.000	.000	.000	.000
		60	.002	.000	.000	.000	.000
		30	.020	.011	.007	.005	.000
.3	-	120	.000	.000	.000	.000	.000
		90	.000	.000	.000	.000	.000
		60	.000	.000	.000	.000	.000
		30	.001	.001	.000	.000	.000
.6	.3	120	.079	.073	.072	.063	.061
		90	.137	.133	.119	.104	.092
		60	.192	.177	.152	.127	.110
		30	.274	.222	.168	.110	.050
.5	.3	120	.004	.003	.003	.003	.002
		90	.017	.019	.016	.010	.006
		60	.057	.051	.039	.035	.028
		30	.144	.103	.077	.038	.014
.4	.3	120	.000	.000	.000	.000	.000
		90	.000	.000	.000	.000	.000
		60	.016	.016	.013	.007	.005
		30	.060	.048	.026	.011	.004

estimator. Table 13 presents the standard deviation of the estimated autocorrelation coefficients for Quenouille's unbiased estimate next to the usual biased estimates. In all cases the empirical standard error of the biased estimator is smaller than that of the unbiased estimator. The magnitude of the difference in standard error is relatively small for time series realizations of length 120, especially for conditions with less serial dependency. As the series becomes shorter and ρ increases in magnitude, however, the biased estimator exhibits considerably less variability.

The results of study three suggest that the unbiased estimation procedure proposed by Quenouille is of limited usefulness. The conditions under which bias is troublesome are the same circumstances that result in problems with the unbiased estimation procedure; relatively short time series realizations and relatively high degrees of serial dependence. In conclusion, it appears that utilizing the unbiased estimation procedure avoids one problem at the expense of a large increase in the magnitude of the standard error of the estimates.

Table 13

First and Second Order Autoregressive Processes

Empirical Standard Error of Unbiased (UNBS) and Biased (BIAS) Autocorrelation Estimates

AR1	AR2	TP	LAG1 UNBS	LAG1 BIAS	LAG2 UNBS	LAG2 BIAS	LAG3 UNBS	LAG3 BIAS
.9	-	120	.068	.051	.121	.090	.166	.120
		90	.085	.062	.151	.107	.206	.141
		60	.126	.087	.217	.146	.280	.182
		30	.216	.133	.334	.197	.398	.227
.6	-	120	.085	.078	.121	.105	.139	.118
		90	.099	.088	.141	.118	.161	.129
		60	.136	.114	.197	.154	.218	.161
		30	.212	.156	.283	.192	.311	.200
.3	-	120	.097	.091	.105	.095	.107	.096
		90	.111	.103	.122	.109	.124	.107
		60	.142	.126	.165	.139	.163	.132
		30	.210	.169	.233	.173	.239	.170
.6	.3	120	.109	.083	.135	.098	.178	.126
		90	.133	.099	.167	.116	.221	.149
		60	.186	.132	.229	.153	.287	.185
		30	.296	.192	.330	.197	.389	.222
.5	.3	120	.117	.098	.130	.104	.167	.129
		90	.138	.112	.157	.118	.200	.145
		60	.194	.146	.220	.154	.269	.180
		30	.297	.202	.310	.193	.363	.214
.4	.3	120	.124	.110	.123	.103	.153	.124
		90	.144	.123	.145	.115	.180	.136
		60	.197	.156	.203	.149	.241	.168
		30	.295	.210	.282	.185	.330	.202

Study Four

The final set of Monte Carlo simulations investigate the small sample properties of the intervention analysis procedure developed by Box and Tiao (1965, 1975). The type of intervention process examined is an immediate, permanent change in level of an ARIMA(1,0,0) stationary time series process. Three characteristics of the time series process were systematically manipulated; 1) the magnitude of the autoregressive parameter (.3, .6, or .9), 2) the number of time points in the time series process (60, 90, 120, or 150), and 3) the magnitude of the intervention parameter (0, .5, .8, or 1.1). Also, the value of the white noise parameter changed with the value of the autoregressive parameter in order to generate time series realizations with a constant variance of one (if AR(1) = .3, σ^2 = .91; if AR(1) = .6, σ^2 = .64; if AR(1) = .9, σ^2 = .19).

Maximum likelihood estimates of the intervention parameter were obtained for 1000 replications of each condition. Confidence intervals (95% and 99% C.I.) were constructed around the estimates based on the t distribution and asymptotic approximation of the standard error of the estimator. The percentage of replications that resulted in the rejection of the null hypothesis, $H_o : \mu_1 - \mu_2 = 0$ (where μ_1 and μ_2 represent the pre- and post- intervention means of the time series process), were computed for each condition. This measure provides an estimate of the empirical Type I error rate 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 results of this set of simulations are presented in Table 14. 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. The inflation of Type I error is somewhat smaller for time series realizations of greater length. However, even data sets of 150 time points exhibit error rates are considerably inflated. Furthermore, the inflation becomes more severe as the value of the autoregressive parameter increases.

The power of the test of the intervention effect will be considered next. As mentioned previously, the variance of the generated data sets was fixed at 1.0 in order to facilitate the comparison of conditions in which the autoregressive parameter differed. Table 14 shows the percentage of rejections of H_o: $\mu_1 - \mu_2 = 0$ for situations in which the intervention component is equal to .5, .8, and 1.1. Obviously, the power of the test statistic increases as the magnitude of the intervention parameter becomes larger.

Table 14

Significance Test of Intervention Parameter

Percentage of Null Hypotheses Rejected for
Four Magnitudes of Intervention Effects (I)

		p < .05				p < .01			
AR1	TP	I=0	I=.5	I=.8	I=1.1	I=0	I=.5	I=.8	I=1.1
.9	150	.097	.345	.642	.855	.034	.194	.442	.717
	120	.122	.344	.628	.837	.037	.188	.413	.685
	90	.129	.351	.598	.811	.054	.190	.405	.655
	60	.137	.327	.561	.781	.065	.190	.356	.606
.6	150	.077	.433	.777	.945	.018	.241	.566	.866
	120	.086	.383	.700	.914	.027	.198	.493	.776
	90	.087	.327	.620	.849	.027	.166	.412	.683
	60	.102	.269	.470	.730	.026	.133	.283	.510
.3	150	.067	.654	.954	1.000	.015	.444	.865	.996
	120	.065	.573	.919	.996	.019	.352	.773	.969
	90	.067	.465	.829	.974	.022	.247	.648	.913
	60	.075	.337	.660	.908	.024	.173	.418	.736

The impact of the autoregressive parameter on the power of the test statistic is of more interest. It can be seen that the power of the test statistic is greatly diminished as the autoregressive parameter becomes larger. Figures 26, 27, and 28 illustrate the importance of serial correlation in determining the probability of correctly rejecting the null hypothesis at the $p < .05$ level.

The length of the time series realization is extremely important in determining the power of the test statistic. Sixty time points is apparently not a sufficient length to assure reasonable certainty of rejecting the null hypothesis when an intervention effect is in fact present. It can be seen that there is very little power for a time series realization of sixty time points and a moderately large autoregressive parameter of .6. Even an intervention effect of 1.1, which is greater than the variance of the time series, only results in a rejection of the null hypothesis in 73% of the replications at the $p < .05$ level.

The test statistic for time series realizations of greater length demonstrates a larger percentage of rejections of the null hypothesis, however, it appears that a lack of power may be a problem in many of the longer time series conditions as well. In particular, as the magnitude of the autoregressive parameter increases, the ability to detect an intervention effect diminishes. For the case of

Figure 26

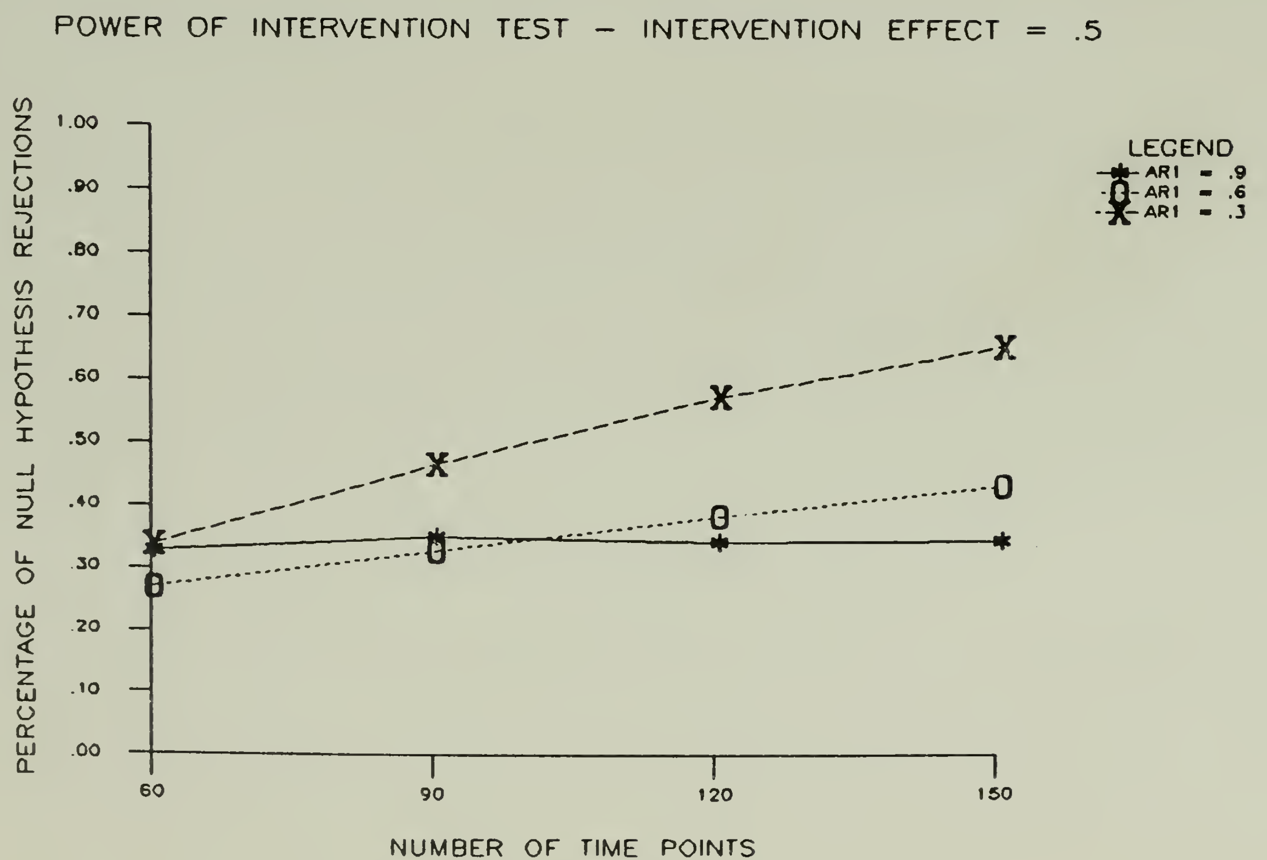

POWER OF INTERVENTION TEST – INTERVENTION EFFECT = .5
PERCENTAGE OF NULL HYPOTHESIS REJECTIONS
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
.00
60
90
120
150
NUMBER OF TIME POINTS
LEGEND
AR1 = .9
AR1 = .6
AR1 = .3

Figure 27

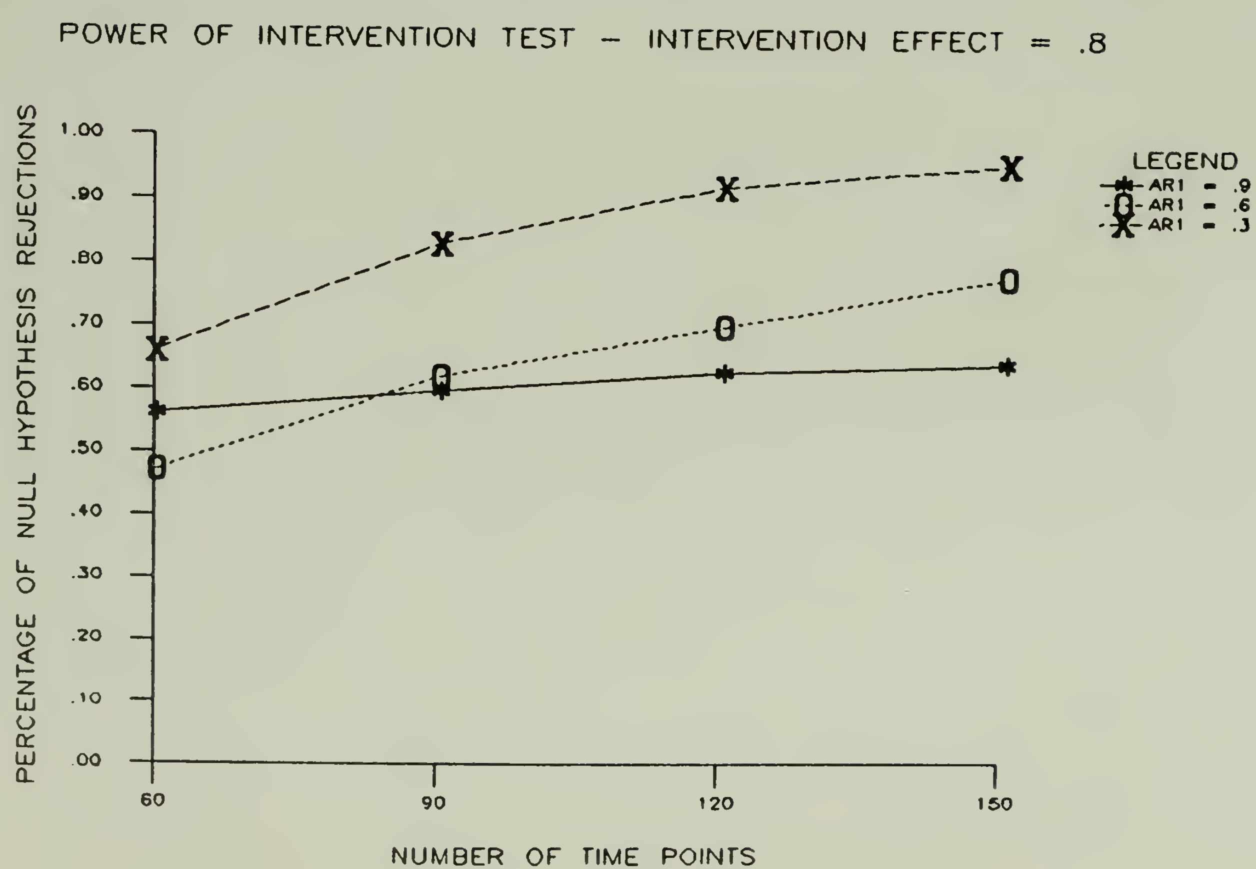
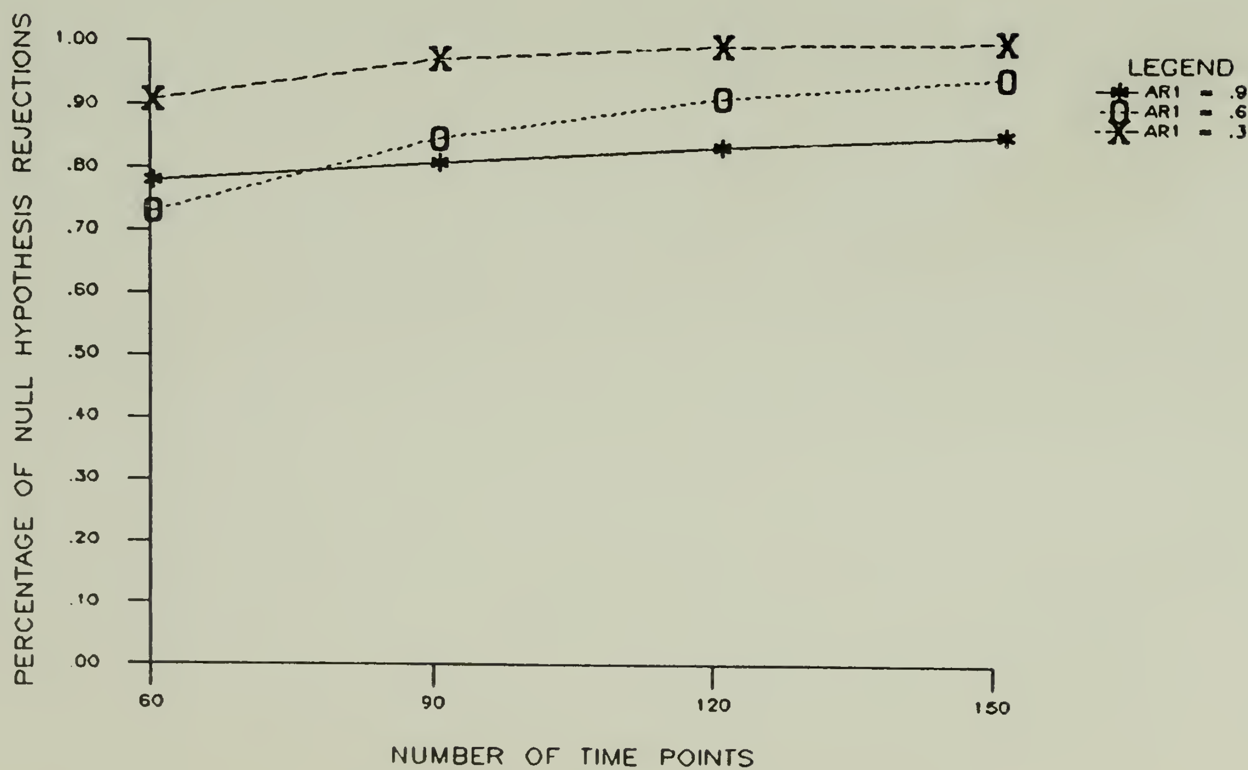

POWER OF INTERVENTION TEST – INTERVENTION EFFECT = .8
PERCENTAGE OF NULL HYPOTHESIS REJECTIONS
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
.00
60
90
120
150
NUMBER OF TIME POINTS
LEGEND
AR1 = .9
AR1 = .6
AR1 = .3

Figure 28

POWER OF INTERVENTION TEST – INTERVENTION EFFECT = 1.1



AR(1) = .9 and 150 time points, an intervention effect of 1.1 is necessary to achieve reasonable power.

Study four also considers the sampling distribution of the test statistic. It is assumed that the test statistic is distributed normally for large sample sizes and as a t distribution for smaller sample sizes. The Kolmogorov-Smirnov test of normality was applied to the distributions of the test statistics in each condition with an intervention parameter equal to zero. The null hypothesis that the sampling distribution was normal could not be rejected for any of the conditions studied. However, the probability levels of the Kolmogorov-Smirnov test statistic were less than .20 for three conditions; AR(1)=.9 with the number of time points equal to 120, 90, and 60. Thus it appears that the distribution of the test statistic may deviate somewhat from normality as the autoregressive parameter of the time series process becomes very large.

The final set of results presented in study four concerns the accuracy of the estimated standard error of the intervention parameter. The results presented in Table 15 are a comparison of the mean estimated standard error of the intervention parameter over 1000 replications and the empirical measure of the standard error obtained by computing the standard deviation of the 1000 estimates of the intervention component. As would be expected, the

Table 15

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

		I=0		I=.5		I=.8		I=1.1	
		EST	EMP	EST	EMP	EST	EMP	EST	EMP
.9	150	.341	.393	.342	.392	.342	.388	.341	.393
	120	.348	.416	.347	.416	.348	.421	.348	.419
	90	.351	.443	.351	.441	.351	.445	.351	.440
	60	.353	.463	.354	.469	.354	.469	.353	.472
.6	150	.291	.309	.291	.307	.291	.308	.291	.307
	120	.318	.344	.318	.342	.318	.344	.318	.344
	90	.350	.388	.349	.382	.349	.390	.349	.390
	60	.400	.452	.399	.452	.400	.454	.400	.455
.3	150	.213	.219	.213	.219	.213	.217	.213	.220
	120	.237	.247	.237	.247	.237	.246	.237	.244
	90	.269	.280	.269	.278	.268	.276	.268	.273
	60	.328	.351	.326	.346	.323	.337	.322	.337

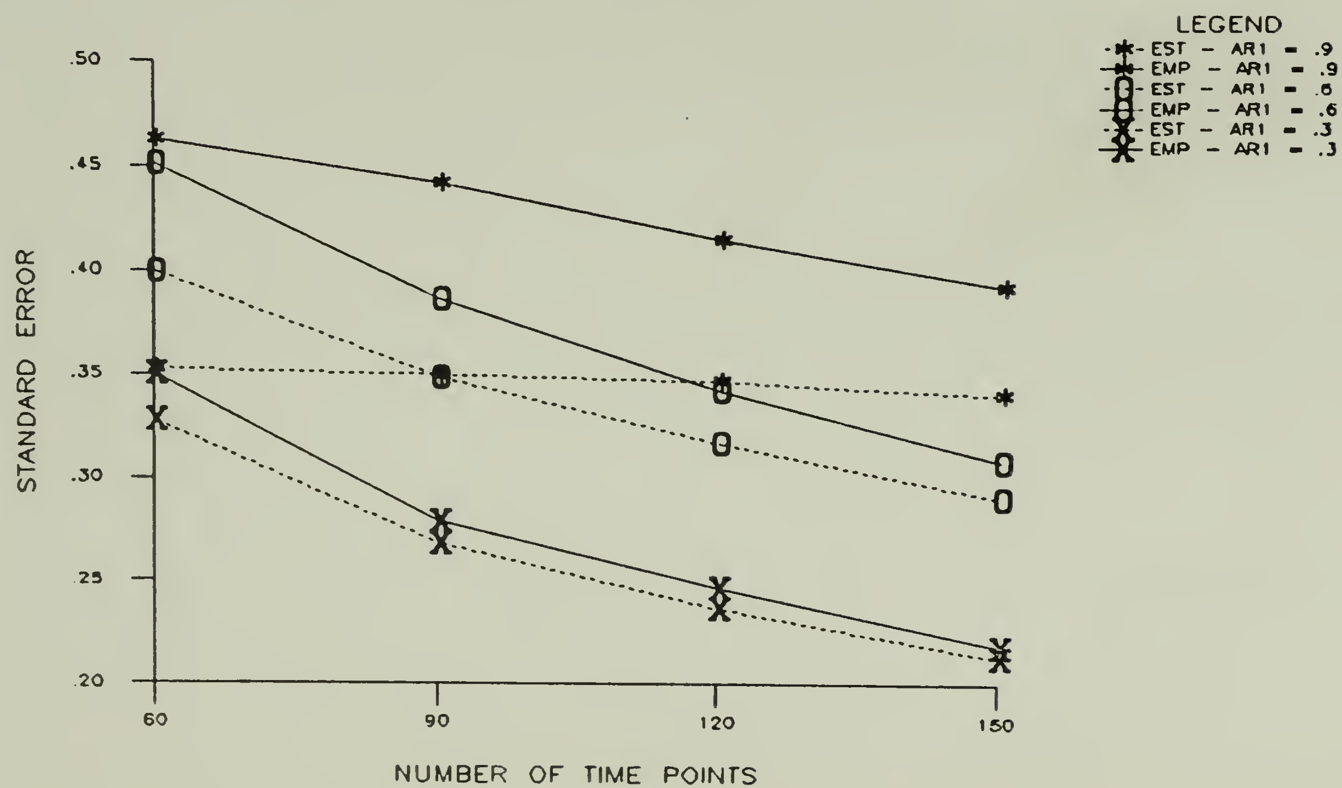
intervention component does not influence the magnitude of either the estimated or empirical standard error, and thus only the values for conditions with the intervention component equal to zero need to be considered.

It can be seen immediately that the mean estimated standard error is consistently smaller than the empirically obtained standard error. This discrepancy between the estimated and empirical standard errors becomes greater as the AR(1) parameter becomes larger, and as the length of the realization becomes shorter. This finding is also presented graphically in Figure 29. The underestimation of the standard error of the intervention component is the most reasonable explanation for the consistent inflation of Type I error rate reported above.

In conclusion, the results of study four suggest two problems in the application of the Box and Tiao (1965,1975) method of testing for the presence of an intervention effect. The empirical Type I error rate of the estimator is consistently greater than the nominal error rate, and the power of the test statistic is often insufficient. These undesirable properties are most severe in those conditions with the largest autoregressive parameter. Both of these problems tend to be alleviated as the length of the time series realizations approaches the maximum length investigated, 150 time points.

Figure 29

INTERVENTION ANALYSIS–ESTIMATED AND EMPIRICAL STANDARD ERRORS



CHAPTER V

SUMMARY, LIMITATIONS, AND CONCLUSIONS

Summary

The four Monte Carlo studies conducted in this research project provide important information with respect to the small sample properties of several estimators utilized in time series analysis. Studies one through three investigate procedures that are used in the model identification stage of ARIMA(p,d,q) time series analysis, while study four examines the small sample properties of Box and Tiao's (1965,1975) test statistic for the presence of an intervention effect in an ongoing time series process. All of the studies manipulate two factors; the nature of the autocorrelation structure and the length of the time series realization.

Study one examines the properties of the usual 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 time series analysis. The results of study one emphasize the importance of measuring the time series process over a sufficient number of time points. As demonstrated in this set of simulations, the problems of bias and variability of the estimates are attenuated as the number of observations in the time series becomes greater.

The second study investigates the estimators of the approximated standard error of the autocorrelation and partial autocorrelation coefficients. The estimators are commonly used to detect non-zero coefficients by constructing confidence intervals around zero for each of the parameters in the autocorrelation or partial autocorrelation function. The expression used to estimate the standard error of the autocorrelation coefficient is based on two assumptions that are likely to be violated. First, it is assumed that the population parameters of the autocorrelation function at every lag is known. Furthermore, it is assumed that the parameter for which the standard error is being estimated is equal to zero, and that the autocorrelation at all lags greater than that of the parameter being tested are equal to zero.

The simulations conducted in study two emphasize several problems in utilizing the estimated standard errors. The study compares the estimated values of standard error with empirical estimates of the standard error. These results indicate a wide discrepancy for many of the conditions investigated, with the estimated standard error generally exceeding the empirical standard error.

Study two also examines the statistical tests of significance of the autocorrelation and partial autocorrelation coefficients. These tests of significance are based on the estimated standard errors, and thus, the results of this part of the study are closely related to those described above. Study two demonstrates that while the empirical Type I error rates are reasonably accurate, the power of the two test statistics tend to be insufficient for the less lengthy realizations that were considered. This is not surprising given the finding that the standard error of the autocorrelation coefficient is over-estimated when the population parameter is different from zero.

Study three investigates the unbiased estimator of the autocorrelation function proposed by Quenouille (1956). It was found that while the estimator is unbiased, two other problems arise that may be more troublesome. First, the occurrence of parameter estimates outside of the theoretical bounds of the parameter were not uncommon. Perhaps more importantly, the estimates exhibited much greater variability in comparison to those of the usual biased estimator. These undesirable properties were most severe in conditions with larger values of ρ , and in those with shorter time series realizations.

Study four examines the statistical test of an abrupt permanent change in level of a stationary time series process, as proposed by Box and Tiao (1965, 1975). Based on the results of the Kolmogorov-Smirnov test of normality, the sampling distribution of the test statistic does not significantly deviate from normality. The results of study four also indicate that the estimated standard error of the intervention component was consistently smaller than the standard deviation of the 1000 estimates of the intervention effect. In addition, the Type I error rate of the test statistic was inflated for all conditions considered, with the inflation increasing as the length of the time series realization becomes shorter. This result can most likely be attributed to the underestimation of the standard error noted above. Finally, the power of the test statistic is less than desirable for many conditions that were studied.

Limitations

The findings of the present research project are limited in several respects, and many areas of future investigation are warranted. The conditions selected for investigation were limited to:

- the AR(1), AR(2), MA(1), and MA(2) processes in studies one, two and three;
- the AR(1) process for study four, which investigated the statistical test of an intervention effect; o an intervention component of an abrupt permanent change in level; o time series realizations of length 30, 60, 90, and 120 in studies one, two, and three;
- and, realizations of length 60, 90, 120, and 150 in study four.

Numerous other conditions could have been selected for investigation, and may have led to slightly different results.

An important area or research that was not investigated involves the robustness of the statistical test of the intervention component. All of the simulations in the present research are based on data generated according to the AR(1) process, which is the model that was assumed by the statistical test. It is also important to examine the consequences of violating the assumption, by testing for an intervention when the ARIMA(p,d,q) model is misidentified.

The present research attempted to investigate the issue of robustness, but a severe problem in obtaining convergence was encountered during the estimation procedure. Time series realizations were generated according to a MA(1) process and then tested using the maximum likelihood function based on the AR(1) process. Estimates were unobtainable for virtually all of the realizations that were generated.

A second type of model misspecification was attempted by generating AR(2) realizations and testing for an intervention effect using the likelihood function of the AR(1) process. Once again, reaching convergence was a problem, but in this instance, the difficulties were somewhat less severe. The percentage of replications for which convergence was attained was roughly 50% for 90 time points, and 20% for 150 time points. For those data sets that successfully converged, the Type I error rate was severely inflated. For the limited number of conditions considered, the empirical Type I error rate was roughly 20% when tested at the nominal $p < .05$ level. Obviously, the validity of the results are questionable given the number of data sets that were eliminated due to convergence problems. Nevertheless, the results of this informal investigation do suggest that model misidentification may be a critical problem in the application of statistical tests of intervention.

Another type of model misspecification that warrants future research is the instance of a change in the time series process at the point of intervention. This is likely to be a common problem for the researcher, since the intervention component may simultaneously affect both the level of the time series process and the nature of the interdependence among the observations. This study did not attempt to investigate this form of model misspecification, and there is apparently existing research that addresses the issue.

Another type of limitation of the research involves the estimation procedures employed in study four, which examines the properties of the intervention component. As described in the methodology chapter, efforts to estimate the four parameters of the model simultaneously by maximizing the full likelihood function were not successful. As a result, a stage-wise estimation procedure was utilized to obtain the maximum likelihood estimates. The extensive comparison of the stage-wise and simultaneous estimation procedures provide reasonable assurance that results based on a simultaneous estimation procedure would not differ from those of the present research. Nevertheless, the possibility that the results would not be precisely replicated using a simultaneous estimation procedure cannot be entirely eliminated.

A related issue concerns the convergence failure of roughly 2 to 5% of the data sets that were generated.

Although the realizations that failed to converge did not appear to differ systematically from those that were successfully analyzed, there is a possibility of biased results due to a systematic difference in those realizations that were eliminated.

In summary, the present research is limited with respect to two general aspects. First, the results of the research are based on a set of specific conditions that were selected for investigation, and these conditions will not necessarily generalize to other conditions that may be of interest. Secondly, the estimation procedure that was employed was not the most desirable procedure that is available. The practical considerations involved in conducting 1000 replications for each condition necessitated using a stage-wise estimation procedure rather than the full maximum likelihood estimation of all parameters simultaneously. The researcher who is analyzing a limited set of time series realizations is more able to employ the full maximum likelihood estimation procedure, and, in the event of non-convergence, modify the starting values to obtain solutions in most circumstances.

Conclusions

In addition to providing specific information with respect to the small sample properties of several estimators utilized in time series analysis, two general conclusions can be drawn by considering the entire set of four Monte Carlo simulations. The most important conclusion concerns the length of time series realization that is necessary to obtain meaningful results on the basis of the statistical procedures discussed in this paper. In addition, researchers should be wary of the problems that are likely to be encountered when analyzing time series data with extreme serial dependence.

It is extremely difficult to provide definitive guidelines for the application of interrupted time analysis. 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 made more troublesome by the large standard error of estimated autocorrelation coefficients that are based on a small number of time points. The extensive variability 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, the researcher will often be forced to determine the number of observations to be measured in the absence of knowledge concerning the autocorrelation structure of the process.

Additional problems in the application of time series analysis to actual data sets involve the assumptions of stationarity and the proper identification of an ARIMA (p,d,q) model. The results presented in this dissertation are based on simulated data sets that are generated according to known stationary AR and MA processes. In actual practice, data sets are not likely to fit an identified ARIMA (p,d,q) process as closely as the simulated data sets of the Monte Carlo experiments. Furthermore, time series data encountered in practice is not likely to precisely conform to the assumption of stationarity. The results of the present research are obtained under optimal conditions, and thus, the idiosyncrasies of actual data sets may magnify the undesirable properties that have been discussed.

Based on the research presented in this dissertation, it is recommended that time series realizations consist of at least 90 observations. The length of the time series realization plays a critical role in determining the quality of the estimates that are obtained when applying the procedures that have been described. Almost all of the estimation problems that have been investigated - bias, the magnitude of standard errors, the accuracy of estimated standard errors, inflation of Type I error rates, and lack of power - are much less severe for more lengthy time series realizations. Thus, researchers should make every effort possible to obtain lengthy data sets, and recognize that conclusions drawn on the basis of shorter time series realizations may be misleading.

The second general conclusion suggested by the results of these experiments involves the degree of serial dependence in time series processes. 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 extremely high. For almost all of the conditions examined in the present research, extreme serial dependence increases the problems that are observed in the estimation procedures. In the model identification stage of time series analysis, both the bias in the autocorrelation estimator and the over-estimation of the standard error of autocorrelation coefficients becomes more severe as the serial dependence becomes more severe. Furthermore, problems with the estimation of the intervention component become more severe as ρ increases; inflation of the Type I error rate becomes greater and there is a decrease in the statistical power to detect an intervention effect.

In conclusion, although methods for the statistical analysis of time series processes are a valuable tool for the researcher, it is important for those applying these procedures to be aware of the inherent limitations and potential problems with the procedures. It is hoped that the research presented here will prove useful to those interested in the application of ARIMA models.

REFERENCES

Bartlett, M.S.(1946) On the theoretical specification and sampling properties of autocorrelated time series. Journal of the Royal Statistical Society, B8, 27-41.

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

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

Box, G.E.P., & Tiao, G.C.(1965) A change in level of a nonstationary time series. Biometrika, 52, 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, 70-92.

Campbell, D.T.(1963) From description to experimentation: Interpreting trends as quasi-experiments. In Harris, C.W. (Ed), Problems in measuring change. Madison, WI: University of Wisconsin Press.

Campbell, D.T., & Ross, H.L.(1968) The Connecticut crackdown on speeding: Time series data in quasi-experimental analysis. Law and Society Review, 3, 33-53.

Campbell, D.T., & Stanley, J.C.(1963) Experimental and quasi-experimental designs for research on teaching. In Gage, N.L. (Ed.), Handbook of research on teaching. Chicago: Rand McNally.

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

Deutsch, S.J. & Alt, F.B.(1977) The effect of Massachusetts' gun control on gun-related crimes in the city of Boston. Evaluation Quarterly, 1, 543-568.

Fuller, W. A.(1976) Introduction to Statistical Time Series. New York: John Wiley & sons.

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

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

Gottman, J.M.(1981) Time-series analysis: A comprehensive intro- duction for social scientists. New York: Cambridge University Press.

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

Hay, R.A.,& McCleary, R.(1976) Box-Tiao time series models for impact assessment: A comment on the recent work of Deutsch and Alt. Evaluation Quarterly, 3, 277-314.

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

Kendall, M. G.(1976) Time Series. London: Charles Griffin & Company, Ltd.

Mann, H.B., & Wald, A.(1943) On the statistical treatment of linear stochastic difference equations. Econometrica, 11, 173-220.

Marquardt, D.W.(1963) An algorithm for least squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, 2, 431-441.

McCleary,R., Hay,R.A., Meidinger,E.E., & McDowall, D.(1980) Applied time series for the social sciences. Beverly Hills: Sage Publications.

McDowell,D., McCleary,R., Meidinger,E.E. & Hay,R.A.(1980) Interrupted time series analysis. Beverly Hills: Sage University Press.

Quenouille, M.H.(1947) A large-sample test for the goodness of fit of autoregressive schemes. Journal of the Royal Statistical Society, 110, 123-129.

Quenouille, M.H.(1956) Notes on bias in estimation. Biometrika, 36, 353.

Zimring, F.E.(1975) Firearms and federal law: The gun control act of 1968. Journal of Legal Studies, 4, 133-198.