# applied time series analysis for the social sciences

Richard McCleary
Richard A. Hay, Jr.

with Errol E. Meidinger
and David McDowall

Foreword by Kenneth C. Land

SAGE PUBLICATIONS Beverly Hills London

*For information address:*

SAGE Publications, Inc.
275 South Beverly Drive
Beverly Hills, California 90212

SAGE Publications Ltd.
28 Banner Street
London EC1Y 8QE, England

Printed in the United States of America
**Library of Congress Cataloging in Publication Data**

McCleary, Richard.
Applied time series analysis for the social sciences.

Bibliography: p.
Includes index.
1. Time-series analysis. 2. Prediction theory. 3. Social sciences—Statistical methods. I. Hay, Richard, joint author. II. Title.
HA30.3.M32 300'.1'51955 79-27873
ISBN 0-8039-1205-6
ISBN 0-8039-1206-4 pbk.

FIRST PRINTING

*To Donald T. Campbell*

# CONTENTS

## LIST OF FIGURES

# FOREWORD

Social scientists have been confronted with revolutionary developments in a variety of areas of statistical theory and methodology during the 1970s, including Bayesian statistical inference, exploratory data analysis, log-linear and related models for categorial data, robust statistics, structural equation models in unobservable variables, and time series analysis. Two characteristics of these developments are noteworthy. First, they make the repertoire of available statistical tools considerably more adaptable to the substantive concerns and empirical data limitations encountered in social science research. But, second, this increase in adaptability typically is bought at the expense of increased computational complexity—in the form of iterative estimation algorithms that are manageable only with the assistance of modern electronic computers.

One consequence of these two qualities is that the "cookbook" approach of introductory statistical methods courses (in which statistical recipes are stated in closed form and applied to simplified, often artificial, data sets) no longer is a feasible mode of instruction. Rather, the emphasis must be on the application of sophisticated computer algorithms to real data sets. While the algorithms may remain essentially "black boxes" to the student, the fact that they are computerized facilitates their application to a large and diverse array of empirical data sets. In this way, the student can obtain an intuitive "feel" for the types of problems likely to be encountered in applications.

This new didactic style is exemplified in the present volume by Professors McCleary, Hay, Meidinger, and McDowall. Their subject matter is the synthesis of time series analysis and forecasting methods brought together in 1970 by George E. P. Box and Gwilyn M. Jenkins (*Time Series Analysis: Forecasting and Control,* San Francisco: Holden-Day; 1976, revised edition). While these *A*uto*R*egressive *I*ntegrated *M*oving *A*verage (ARIMA) models (popularly called "Box-Jenkins" models) have been widely applied in engineering, economics, and business for nearly a decade, social science applications outside of economics still are relatively rare. One obvious reason for this is the fact that available textbooks typically assume a mathematical and statistical maturity greater than appropriate for social science audiences outside of economics (which has a long tradition of econometric analyses of time series).

It is precisely this void in the statistical time series textbook literature that McCleary and his coauthors seek to fill. Assuming no training in statistics beyond intermediate statistical methods (at, say, the level of H.M. Blalock, Jr., 1979, *Social Statistics,* revised second edition, New York: McGraw-

Hill), the authors take the reader through a gentle introduction to univariate ARIMA models (emphasizing the Box-Jenkins iterative cycle of model identification, estimation, and diagnosis), impact assessments, and forecasts. This is followed by chapters on multivariate ARIMA models and ARIMA estimation algorithms. The text is noteworthy for its clear, concise exposition punctuated by numerous analyses of real time series. These pertain to a diverse array of noneconomic topics. By including listings of the time series and references to the available ARIMA software, the authors encourage readers to develop "hands on" experience in the didactic mode described above. For the highly motivated student, the authors also include annotated bibliographic references to more advanced literature.

In addition to its value as a didactic aid, this book is a timely and welcome addition to the professional social science literature for two reasons.

First, there is a wide and growing interest in the study of social change via the analysis of historical time series. This has been stimulated, in part, by the "social indicators movement" and the increased availability of time series data on noneconomic social conditions resulting therefrom. (See, for example, Kenneth C. Land and Seymour Spilerman, eds., 1975, *Social Indicator Models,* New York: Russell Sage Foundation, for a statement of some of the analytic problems created by social indicator time series.) Although annual social indicators time series often are too short for the application of ARIMA models and methods, this is less likely to be the case for social indicator time series collected at quarterly, monthly, weekly, or daily intervals. In any case, for sufficiently long series in areas in which there exists little prior theory or research, the ARIMA models described in this book provide powerful analytic tools.

Second, as the authors note, there has been a convergence in recent years between the "statistical time series" and "dynamic structural equation models" literatures (popularly called the "Box-Jenkins" and "econometric" literatures, respectively). For dynamic structural equation modelers, this convergence has taken the form of a greater sensitivity to the stochastic properties of the error or disturbance terms in time series models. Conversely, time series analysts have come to recognize that lag structures need not always be identified *de novo* from the time series to be modeled in areas in which substantial prior theory and/or research exists. The present volume should give nonspecialist readers the background in statistical time series analysis necessary to appreciate more fully the nature of this convergence.

—Kenneth C. Land
University of Illinois at Urbana-Champaign



## ACKNOWLEDGMENTS

Our interest in time series analysis was first stimulated by the pioneering work of Donald T. Campbell. He has remained a constant source of inspiration and support over the years. We are no less indebted to Richard A. Berk who first introduced us to the practical and methodological exigencies of applied social research. He has also generously supported development of the software used in this volume. Berk, Clifford C. Clogg, Bruce Foster, Kenneth C. Land, Michael C. Musheno, and J. Richard Zelonka all read and commented on various drafts of the manuscript; the academic computing centers of Arizona State University and Northwestern University provided computer time and support for the analyses; and our collaborators, Errol E. Meidinger and David McDowall, made invaluable contributions to the organization and content of the volume. Finally, thanks are due to our colleagues at the Center of Criminal Justice, Arizona State University, and the Vogelback Computing Center, Northwestern University, for implicit and explicit support of this project.

—Richard McCleary
Tempe, AZ

—Richard A. Hay, Jr.
Evanston, IL

# 1 Statistical Models for Time Series Analysis

Monographs on time series analysis ordinarily address *either* theoretical issues (e.g., Doob, 1953; Feller, 1971; Anderson, 1975) *or* practical issues (e.g., Nelson, 1973; Glass et al., 1975; Pindyck and Rubinfeld, 1976; Makridakis and Wheelwright, 1978). This volume is of the latter type. We will be concerned largely with the practical or applied aspects of time series analysis and especially with applications of interest to economists, political scientists, psychologists, and sociologists.

Time series analysis can similarly be divided into two methodological areas: harmonic analysis and regression analysis. These two methods are sometimes called *analysis in the frequency domain* and *analysis in the time domain,* respectively. We will not cover harmonic methods at all in this volume. While there are no practical limitations on the use of harmonic methods with social science data (see, e.g., Mayer and Arney, 1974), this type of analysis ordinarily requires a background in the calculus and algebra of complex variables (imaginary numbers). We suspect that most readers will lack this mathematical preparation. Regression approaches to time series analysis, in contrast, have been widely used in the social sciences. All of our readers will have had some training in multiple regression methods. We have consciously addressed our development of time series analysis to this level of understanding.

Readers with broad backgrounds in regression methods may nevertheless find this volume novel. The time series models developed here are not the models ordinarily developed in econometrics texts, but rather are *stochastic* *process* models. The particular class of stochastic process models to be developed are the *A*uto*R*egressive *I*ntegrated *M*oving *A*verage (ARIMA) models of George E. P. Box and Gwilym M. Jenkins (1976; Box and Tiao, 1965, 1975). Although elements of ARIMA modeling can be traced back some 50 years, Box and Jenkins (and George C. Tiao) must be credited with integrating the elements into a comprehensive theory, extending it greatly, and popularizing it. ARIMA modeling has rightly been called the "Box-Jenkins approach to time series analysis."

It will be instructive at this early point to make the tenets of the Box-Jenkins approach explicit. ARIMA models posit a random shock, $a_t$, as the driving force of a time series process, $Y_t$. As an analogy, consider a coffee-brewing machine of the sort widely used in university departments at present. To brew a pot of coffee, 12 cups of cold water are fed into one end of the machine. A few moments later, 12 cups of coffee are delivered from the other end. Of course, we do not always wish to brew 12 cups of coffee. Sometimes we feed only 6 cups of cold water into the machine with the expectation that only 6 cups will be delivered at the other end of the machine. And of course, sometimes we expect to receive 12 cups of coffee from the machine, but for some reason receive only 11.5 cups. We can diagram the coffee brewing process as an input-output process:

$$\text{(Cold Water) } a_t \rightarrow \square \rightarrow Y_t \text{ (Hot Coffee)},$$

Both the $a_t$ input and the $Y_t$ output are measured in cup units. To some extent, there is a prescribed relationship between the size of each input and the size of each output. Inputting 6 cups of cold water, for example, we would be surprised if 12 cups of coffee were delivered at the other end of the machine. On the other hand, we might be equally surprised if *exactly* 6 cups of coffee were delivered. Inside the coffee machine, the unobserved and often mysterious brewing process is at work, sometimes delivering slightly more than 6 cups of coffee and sometimes delivering slightly less.

If we were interested in this mysterious internal process, we could perform an experiment with the aim of unraveling the relationship between $a_t$ and $Y_t$. We would first hire a graduate student to do nothing but brew coffee under scientific conditions. Every 15 minutes, the graduate student would input a precisely measured amount of cold water, and after a few minutes brewing time would receive an amount of hot coffee from the machine to be measured precisely. To ensure experimental control, each input to the machine would be randomly determined. Our graduate student would consult a table of Normal (Gaussian) random numbers to determine how many cups of cold water to feed into the machine for each trial.



After many trials (say, 500 pots of coffee), we could analyze the $Y_t$ output series to draw inferences about the brewing process. We would no doubt discover the following:

(1) The most important determinant of $Y_t$ is $a_t$. Other things being equal, the more cold water input, the more hot coffee output.
(2) To a lesser extent, $Y_t$ may be also be determined by $a_{t-1}$, the previous input. A small percentage of each input, for example, may remain inside the machine to be delivered in the next output. The larger the input on one trial, the larger the residual remaining inside the machine to be delivered in the next output.
(3) To a lesser extent, $Y_t$ may be also be determined by $Y_{t-1}$, the previous output. A particularly large output, for example, as a result of a particularly large input, may somehow reduce or increase the efficiency of the brewing process. This change in efficiency will show up in the next output.
(4) To a much lesser extent, and for the same reasons, $Y_t$ may be determined by inputs and outputs further removed in time, such as $a_{t-2}$ and $Y_{t-2}$.

These are the basic tenets of the Box-Jenkins approach to time series analysis, or ARIMA modeling. In general, we may say that a pot of coffee has its size determined by the few immediately preceding outputs and inputs, that is,

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

Inputs and outputs further removed from the present, $a_{t-3}$ and $Y_{t-3}$, for example, may also play some role in the process but their influence will be so small as to be statistically insignificant. In practice, of course, the number of past inputs and outputs required in the model will be determined empirically, but in almost every case, no more than two prior inputs and outputs will have a statistically significant influence on the present output.

The most important tenet of the ARIMA model is that the present input, $a_t$, will have a greater impact on the present output than any earlier input. This means that the parameter $\theta_1$ must be a fraction. The parameter $\theta_2$, also a fraction, will ordinarily be smaller than $\theta_1$, so in general

$$1 > \Theta_1 > \Theta_2 > \ldots > \Theta_q.$$

This is the most important principle of the ARIMA model. The influence of a past event (or input) on present events diminishes as time passes. This same principle applies to the influence of past outputs:

$$1 > \phi_1 > \phi_2 > \ldots > \phi_p.$$

But if there is a single guiding principle of the Box-Jenkins approach, it is *parsimony*. This principle reflects not only a view of nature but also a view of the relationship between a time series model and nature. In almost all cases, a social science time series process can be modeled as a probabilistic function of a few past inputs (random shocks) and outputs (time series observations). As we develop the algebra of ARIMA modeling in the next chapter, it will become clear that there is a real difference between parsimony and simplicity. While parsimonious, univariate ARIMA models also give a surprisingly sophisticated representation of nature.

The reader who is familiar with the more widely used regression approaches to time series analysis (structural equation or econometric models) should not assume that ARIMA models are substantially different than regression models. While ARIMA models require the novel input-output explanation, the two approaches are in fact identical. The only real difference between ARIMA and regression approaches to time series analysis is a practical one. Whereas regression models can be built on the basis of prior research and/or theory, ARIMA models must be built empirically from the data. Because ARIMA models must be identified from the data to be modeled, relatively long time series are required. *No time series that we analyze in this volume is shorter than 50 observations long*. The reader may use this rule of thumb when deciding whether to analyze time series data from an ARIMA or regression approach. When relatively long time series are available, an empirical ARIMA approach will ordinarily give the best results. But when relatively long series are not available, regression approaches informed by prior research and/or theory will give the best results.

## 1.1 Caveat: The Limits of Time Series Analysis

If our experiences as teachers are typical, there is a danger to learning any sophisticated statistical method. The power of the method may desensitize the student to the more fundamental questions of interpretation. Though absorbing, the statistical problems of analyzing time series data are generally less important than the problems of interpreting the results of an analysis. Lacking an easy interpretation, the time series analysis has failed.

To emphasize this point, we note first that, while the statistical methods of time series analysis are relatively new, the *logic* of time series analysis is not. In his classic investigation into the causes of suicide, for example, Emile Durkheim wrote:

> It is a well known fact that economic crises have an aggravating effect on the suicidal tendency. . . . In Vienna, in 1873 a financial crisis occurred which reached its height in 1874; the number of suicides immediately rose. . . . What



> proves this catastrophe to have been the sole cause of the increase is the special prominence of the increase when the crisis was acute, or during the first four months of 1874 [1951:241].

We would like to believe that Durkheim actually plotted the annual suicide rates of European cities, searched for peaks and valleys in this time series plot, and then compared the peaks and valleys with the economic histories of the cities. In the most general sense, this is a time series analysis.

Aside from an increase in methodological sophistication, nothing has changed. Contemporary economists, political scientists, psychologists, and sociologists share an interest with the early social philosophers in the change of social phenomena over time. At the most pragmatic level, this traditional preoccupation with the temporal ordering of things can be explained as a function of "cause." When causal relationships are an issue, social scientists have traditionally resorted to longitudinal research designs. Time series *analysis* is a statistical *method* for interpreting the results of certain longitudinal research designs. When used appropriately, time series analysis brings a powerful inferential logic to bear on questions of social cause. When used inappropriately, however, the relative weaknesses of time series analysis far outweigh its strengths.

Assuming that time series analysis is the appropriate method for addressing a particular research question, the analyst must pay some attention to defining the time series. On the face of it, *a time series is a set of N time-ordered observations of a process*. Each observation should be an interval level measurement of the process and the time separating successive observations should be constant. Minimal violations of these requirements are acceptable. Monthly time series observations, for example, are sometimes separated by 28 days and sometimes by 30 or 31 days. This minimal departure from the ideal presents no real problems for the analysis, however.

By this definition, a time series is a *discrete* data set. Figure 1.2 shows a plotted time series. The observations of this series, the data that is, are the equally spaced symbols (o) strung out along the time dimension. It is often more realistic (and aesthetically pleasing) to connect the symbols with a broken line as we have done in Figure 1.2. The analyst should nevertheless be conscious that the time series is actually a *discrete* data set.

To be sure, the discrete time series may be a measure of some underlying *continuous* process. Stock market time series, for example, are usually reported as daily closing prices. The price of a stock fluctuates more or less continuously throughout the trading day. *Closing* price is the value of the stock at 3:00 P.M. when the New York Stock Exchange "closes" or ends its business day. We might instead choose to record the daily opening or noon price but the principle is still the same. So long as the continuous process is

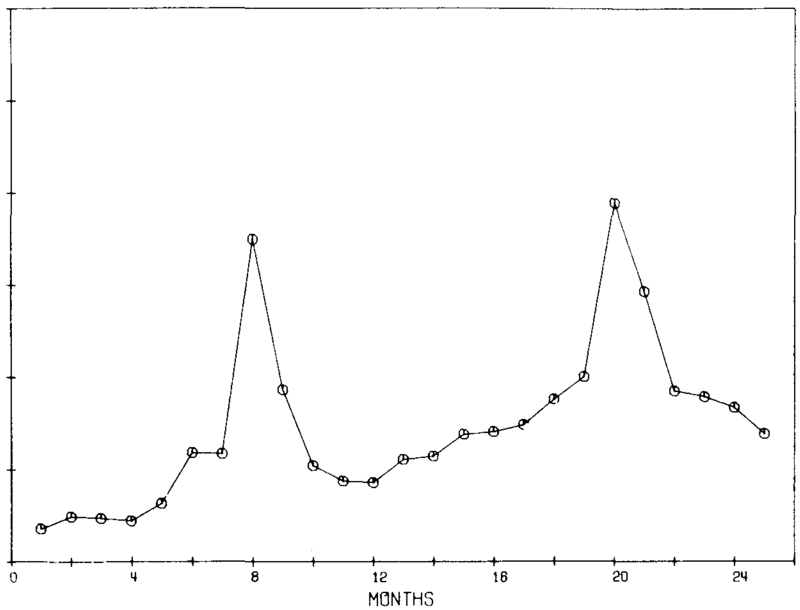



FIGURE 1.2 A Time Series

recorded consistently, a discrete time series is likely to give a good approximation to the continuous process.

Many other time series will approximate processes that are actually discrete. For example, a discrete (but rare) event process can be *aggregated* into a time series. An illustration of this would be a time series of monthly traffic fatalities. Traffic fatalities are discrete, rare events and thus might be best analyzed as Poisson outcomes. If the data are aggregated into monthly totals, however, the resulting time series will ordinarily capture the essence of the rare event process. The average "waiting time" between fatalities in a given month will be roughly proportional to the total number of fatalities in that month.

While there are many other ways in which continuous or discrete social processes can be represented by a time series, the principles of logic are the same. The social scientist is interested in making inferences about the process *underlying* the time series. The time series must thus always give an adequate representation of the true social process. Lacking this quality, inferences based on a time series analysis will be invalid.

But assuming that a time series analysis is the optimal research method, and assuming that an adequate data set is available, some attention must be



paid to the limits of logical inference. For all practical purposes, the time series analyst is interested in "predicting the future" of a social process, that is, in measuring the past "change" in a social process to extrapolate that "change" into the future. Extrapolation cannot be accommodated outside a well-defined axiomatic foundation, however: a theory of the process. Discussing the measurement of economic growth, for example, Simon Kuznets has noted that:

> The difficulties in measuring economic growth, supply of empirical data apart, lie precisely in this point: modern economic growth implies major structural changes and correspondingly large modifications in social and institutional conditions under which the greatly increased product per capita is attained. Yet for purposes of measurement, the changing components of the structure must be reduced to a common denominator; otherwise it would be impossible to compare the product of the economy of the United States with that of China, or the product of an advanced country today with its output a century ago before many of the goods and industries that loom so large today were known [1959:15].

Kuznets's point here is that we cannot compare apples with oranges unless we have some theoretically sound dimension on which they are comparable. To measure change in the output of any process, we must work from a set of axioms that allows a comparison of today's social and economic structures with yesterday's social and economic structures. Gross national product, for example, must be defined in such a way that its meaning is not appreciably time-bound.

In many substantive areas, the interpretation of measured change in a time series presents no significant problems. Even when serious problems arise, however, interpretability can be increased through appropriate changes in operational definitions. There is ordinarily a smaller problem interpreting change over a few months or years, for example, than over a few decades or centuries. Interpretability can thus sometimes be increased by narrowing the time frame of the analysis. Of course, there is a trade-off here: Making the time series more interpretable may also make it more trivial. The limitations of time series analysis in this sense must be clearly understood. Time series analysis is generally more appropriate for gauging incremental change than for gauging structural change. In some cases, meaningful indicators of structural change can be found and, in these cases, the analytical power of time series analysis can be brought to bear. But meaningful indicators of structural change are sometimes not available and, in these cases, any time series analysis will be absurd.

It is also ordinarily easier to understand change in relatively concrete time series than in relatively abstract time series. It would be easier to measure change in "work-related accidental fatalities," for example, than in "unemployment." Our definition of the first indicator has changed little over a long period of time—although it has changed. Our definition of unemployment, on the other hand, changes constantly as a result of changing social norms. What would have been an unthinkably high rate of unemployment during one era may be seen as an absurdly low rate of unemployment during another era. While the absolute meaning of an unemployment rate may not change at all over time, the social meaning of unemployment changes constantly.

We may say the same thing of "crime." Social scientists cannot agree among themselves whether rates of crime and unemployment are higher or lower today than they were 50 years ago. The crux of this disagreement can be eliminated by simply conducting analyses of concretely defined time series. But again, to insist upon concrete, objective indicators may sacrifice substantive importance. We are not arguing that the analyst should ignore a time series problem when the available data are not obviously concrete. But the analyst must always be conscious of and acknowledge the limitations of an analysis.

We have intended this short essay to be a *caveat,* warning the reader that the problems of analyzing time series data are relatively unimportant compared to the problems of interpreting analytic results. Without interpretability, there is nothing. More specifically, the reader has been warned (1) to use time series methods only when those methods are appropriate to the research question; (2) to be sure that a time series gives an adequate representation to the underlying process; and (3) to recognize the limits of each analysis.

## 1.2 An Outline of the Volume

We might have called this section "How To Use Our Book." Here we explain our motives, describing as best as possible what material will be presented in subsequent chapters, in what order this material will be presented, and why. Some readers may become anxious and confused in subsequent chapters, perhaps understanding the material presented, but questioning its relevance to the overall scheme of time series analysis. A careful reading of this outline may reduce anxiety levels, permitting the reader to concentrate more fully on the elements of time series analysis.

ARIMA models and Box-Jenkins time series analysis are not necessarily more difficult a topic than other statistical models and methods used in the social sciences. ARIMA models are nonetheless "different" than most other statistical models and methods in terms of their underlying principles, statistical properties, and applications. One major difference which we have already noted is that ARIMA models are not arbitrarily fit to data, but rather



are built empirically. The analyst selects a particular model for a given time series from the general class of ARIMA models. The decision is based on empirically derived characteristics of the series. To make a wise decision, however, the analyst must be aware of the statistical properties of ARIMA models and must be adept at relating these properties to information about the series.

These two concerns, statistical knowledge or understanding and practical application, have determined the orientation of this volume. In each chapter, we first develop the statistical properties of the time series models. We have attempted to do this in a manner that is accurate and thorough and yet lucid to a reader with only an elementary statistical background. When some esoteric or mathematical point must be made, we use footnotes. For all practical purposes, these footnotes may be disregarded in the first reading of a chapter. After developing statistical properties, we apply the time series models to several time series which, in our opinion, are typical of those encountered in the social sciences. Each of these example series has been carefully selected to illustrate a variety of characteristics. Each of these series is listed in an Appendix to this volume and our intention is that the reader will replicate our analyses.

The volume is divided into six chapters, each covering a distinct topic in time series analysis and each building on material developed in preceding chapters. Chapter 2 is the core of the volume and an understanding of the material presented there is *absolutely* essential. Chapters 3, 4, and 5 extend the basic ARIMA model developed in Chapter 2 to a particular application. These chapters may be read out of order, although we do not recommend it. Chapter 6 deals with several practical issues of parameter estimation which will be of most interest to those readers who are about to start a time series research project.

In Chapter 2, we present the basic concepts of univariate Box-Jenkins time series analysis. The univariate ARIMA model is the baseline "building block" which we use in subsequent chapters for impact assessment, forecasting, and causal modeling. Separate and distinct components of the ARIMA model (the autoregressive, integrated, and moving average components) are developed in sequence and then integrated into the general ARIMA model. Once developed, the general ARIMA model is used in an analysis of four example time series.

Chapter 3 presents a general impact assessment model for the analysis of an "interrupted time series quasi-experiment." This model has been widely used to analyze or assess the effects of planned and unplanned interventions on social systems. The impact assessment model consists of a noise model, as developed in Chapter 2, coupled to an impact model. After developing the algebra of a general impact model, we illustrate its use by analyzing several example time series.

In Chapter 4, we develop the use of ARIMA models for forecasting future values of a time series. Our treatment of this topic is intentionally brief. Most of the books written about applied time series analysis are oriented exclusively to forecasting applications and we would have little original thought to add to this body of literature. Readers still may gain some understanding of ARIMA models and algebra from our treatment of forecasting. After presenting the forecast profiles for several time series, we conclude this chapter with a discussion of the uses (both proper and improper, in our opinion) of forecasting in social science research.

Chapter 5 extends the Box-Jenkins approach to multivariate time series analysis. One or more independent variable time series (inputs) may be used to explain the stochastic behavior of a dependent variable time series (output). Multivariate ARIMA time series models are a rather novel concept in social science research although, in our opinion, they have a great potential. Once again, we present the underlying statistical concepts of multivariate ARIMA models, develop a model-building strategy, and conduct several example analyses.

As in many other areas of quantitative social research, Box-Jenkins time series analysis depends upon sophisticated computer software. ARIMA models are nonlinear, for example, so parameters must be estimated with numerical routines. In Chapter 6, we derive the likelihood function, illustrate the solution procedure, and discuss related topics that may affect the analysis. Several available software packages for the analysis of time series data are reviewed and the use of interactive software is discussed. It may seem unusual to relegate a chapter on "estimation" to the the end of the book. Yet most social science graduate students are able to use multiple regression software packages without actually understanding how parameter estimates are generated inside the computer. The same principle holds for Box-Jenkins parameter estimation. While most readers will have no trouble conducting time series analyses without understanding the mechanics of parameter estimation, the reader who intends to go further in this area must at least understand the general principles of nonlinear estimation. For all readers, Chapter 6 is likely to be insightful. For readers who intend to do major work with time series analysis, Chapter 6 is necessary reading.

## A Note to the Instructor

A graduate seminar in time series can finish the material presented here in 8 to 12 weeks, depending upon many obvious factors. Our seminars have typically included no more than 12 students drawn from several graduate



social science departments. To be admitted to the seminar, students have been required to have a set of time series data and a short research proposal. Grades are based on an article-length research report. The time series data listed in Appendix B, and analyzed in the volume, provide an excellent practicum experience for the seminar. After analyzing these data, however, students must generalize the experience to their own time series data. Of course, the seminar requires access to an appropriate software package, preferably one that is interactive. In Section 6.4, we describe several time series software packages that are available at almost all academic computing centers.

# 2 Univariate ARIMA Models

If this volume has a single most important chapter, it is this one. Here we develop a general ARIMA model-building strategy for a single time series. In subsequent chapters, we develop methods for applying the univariate ARIMA model to problems of social research. An understanding of these applications will require an understanding of the material developed here.

We have divided this chapter into two distinct parts. The first part (Sections 2.1 to 2.9) deals mainly with abstract issues, especially with the statistical properties of univariate ARIMA time series models. This material (and, indeed, the rest of the volume) presumes a knowledge of fundamental statistical concepts and a familiarity with algebra. In general, the reader should have a working knowledge of the material ordinarily presented in a first-year social statistics course: measures of central tendency, variance, covariance, correlation, expected values, the Normal distribution, ordinary least-squares (OLS) regression, and so forth. A short appendix at the end of this chapter summarizes the rules for applying expectation operators to random variables. The reader who is unfamiliar with the concept of expected values may read this appendix before starting the chapter.

We try to present the material in the first half of this chapter in an illustrative manner that is both intuitively plausible and technically correct. Readers are urged not to rely on our algebra, but to work through each demonstration or derivation. Although this may become tedious at times, an understanding of abstract concepts will open the door to an understanding of the general class of ARIMA models. More important, the derivation of model algebra is in its own right an intellectually challenging and satisfying exercise which should not be missed.

In the second part of this chapter (Sections 2.10 to 2.13), we describe the concrete procedures used to build univariate ARIMA models for given sets of time series data. After developing a general model-building strategy, we apply it step by step to the analysis of four time series, each typical of those encountered in social science research. These analyses illustrate how such problems as nonstationarity, seasonality, outliers, and ambiguous identification information can be handled within the general ARIMA model-building strategy.

As an introduction to the first part of this chapter, we now return to the most crucial of all definitions, *a time series is a set of ordered observations:*

$$Y_1, Y_2, Y_3, \ldots, Y_{t-1}, Y_t, Y_{t+1} \ldots.$$

In cross-sectional analyses, the order of observation is not of any great consequence and may even be undefined. If the analyst is measuring the performance of subjects in an experiment, for example, it usually makes no difference which subject was tested first and which subject was tested last. Indeed, it is often the case that all subjects are tested simultaneously, so no order of testing is defined.

In longitudinal analyses, on the other hand, the order of an observation is crucial. If the analyst is measuring the improvement in performance from test to test, for example, the order of an observation is as important as the observation itself. The order of an observation is conventionally denoted by a subscript. The general observation is written as $Y_t$, meaning the $t^{th}$ observation of a time series. This implies that the preceding observation is $Y_{t-1}$ and the subsequent observation is $Y_{t+1}$.

We will make a distinction throughout this chapter between *process* and *realization*. An observed time series is a realization of some underlying stochastic process. In this sense, the relationship between realization and process in time series analysis is analogous to the relationship between *sample* and *population* in cross-sectional analysis.

A related and equally important concept is the *model*. A realization, or time series, is used to build a *model* of the process which generated the series. The procedures used to build this model are broadly referred to as time series *analysis*, which implies that the series is being "picked apart" or decomposed into its components. ARIMA models are built around three process components: the autoregressive, integrated, and moving average components. The first component to be considered is the integrated component which is closely related to the concept of trend.



## 2.1 Trend and Drift

Our discussion of ARIMA algebra begins with the concept of *trend*. When we think of the component parts of a time series process, we tend not to think of ARIMA structures, but rather of more fundamental, commonsensical component parts. Trend is one such component. Even those of us who have no experience with longitudinal analysis have a fundamental understanding of what a trend is. A trend is motion in a specific direction, usually (to simplify matters) upward or downward. We can say thus that the trend in government during this century has been *away* from the state level and *toward* the federal level. Not surprisingly, our notion of trend in terms of ARIMA structures is almost identical to this commonsense notion.

More specifically, we define trend as *any systematic change in the level of a time series*. While this definition lacks the mathematical rigor we might like, it is the best definition we (and mathematicians) can construct. A time series that is trend*less* can ordinarily be represented by the model

$$Y_t = b_0 + N_t,$$

where the parameter $b_0$ is interpreted as the *level* of the time series. $N_t$ is a "noise" component or a stochastic process. The most important type of stochastic process in time series analysis is "white noise," which we represent as $a_t$. Assuming a white noise process, we can rewrite the model for a trendless time series process as

$$Y_t = b_0 + a_t.$$

A white noise process has the statistical property[1]

$$a_t \sim NID(0, \sigma_a^2),$$

that is, white noise consists of a series of random shocks, each distributed Normally and independently about a zero mean with constant variance, $\sigma_a^2$. As the mean of white noise is zero, the expected value of a trendless process is:

$$EY_t = b_0 + Ea_t = b_0.$$

The parameter $b_0$ is the level about which the realized time series fluctuates. As the level of the time series process is constant throughout its course, the time series is trendless with a "flat" appearance.

If a time series process is trendless, the parameter $b_0$ is the arithmetic mean of the series, estimated as

$$\hat{b}_0 = \bar{Y} = 1/N \sum_{i=1}^{N} Y_i,$$

for a time series of N observations.

Unfortunately, most social science time series processes are not well represented by this simple model. A time series process following a linear trend, for example, requires a model with an extra parameter:

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

For this model, the expected value of $Y_t$ is:

$$EY_t = b_0 + b_1 t + Ea_t = b_0 + b_1 t,$$

a regression of $Y_t$ on the time series t ($t = 0, 1, \ldots, N$). The level of the time series process is thus expected to change systematically throughout its course. Starting at $t = 0$,

$$\begin{aligned} EY_0 &= b_0 \\ EY_1 &= b_0 + b_1 \\ EY_2 &= b_0 + 2b_1 \\ &\vdots \\ EY_N &= b_0 + Nb_1. \end{aligned}$$

The parameter $b_0$ is the *intercept* of the model, that is, the expected level of the process when $t = 0$. The parameter $b_1$ is the *slope* of the model, that is, the expected change in level from one observation to the next.

Unfortunately, few social science time series processes appear to be trendless and this presents a problem for the time series analyst. Trend must be removed or modeled. One common (but almost always inappropriate) method of detrending a time series is to use a linear regression model for the trend. With this method, the analyst defines the order of each observation ($t = 1, 2, \ldots, N$) as an independent variable. Then, using the time series itself as a dependent variable, the parameters $b_0$ and $b_1$ are estimated with OLS formulae. For a linear trend, this method yields the detrended series:

$$\hat{y}_t = Y_t - \hat{Y}_t.$$

The detrended series, $\hat{y}_t$, is then analyzed.



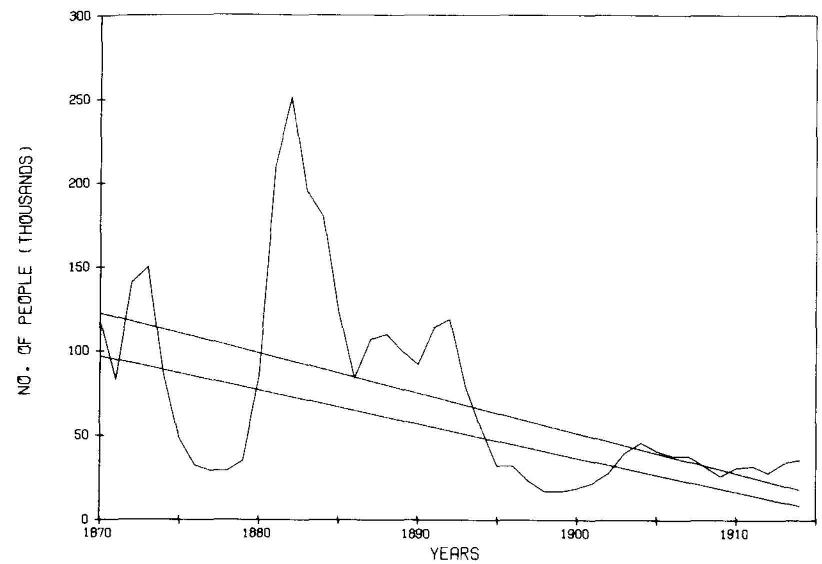



FIGURE 2.1(a) German Immigration to the United States (in Thousands), 1870–1914

*As there are a number of problems with this method, it is not generally recommended.* One major problem with OLS trend models is that the parameters $b_0$ and $b_1$ cannot be estimated with any accuracy. In Figure 2.1(a), for example, we show a time series of total immigration (in thousands) from Germany to the United States during 1870–1914 as reported by Fried (1969). Does this series follow a trend? Reasonable people might agree that total immigration decreased steadily during this period. The OLS trend line for this series is:

$$\hat{Y}_t = 124.78 - 2.37t.$$

But note the abnormal spike in this series starting in 1881. If we ignore this spike, the OLS trend line becomes:

$$\hat{Y}_t = 98,82 - 2.01t.$$

We have superimposed these trend lines over the time series in Figure 2.1(a) to illustrate the estimation problem: *OLS trend estimates are sensitive to outliers.*

The underlying problem is that the OLS linear regression model of trend depends upon the sum of squares function which, for the German immigration time series, is:

$$\sum_{t=1}^{N} (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{N} [b_0 - \hat{b}_0 + (b_1 - \hat{b}_1)t]^2.$$

OLS estimates of $b_0$ and $b_1$ are derived by minimizing this sum of squares function. As the independent variable, t, increases monotonically, however, some of the time series observations are more important than others in the sum of squares function. As a rule, the first and last observations of the time series ($Y_1$ and $Y_N$) usually make the greatest contribution to the sum of squares. In an extreme case, the OLS estimates of $b_0$ and $b_1$ are derived so that the OLS trend line passes through $Y_1$ and $Y_N$ regardless of how well the middle observations ($Y_2, Y_3, \ldots, Y_{N-1}$) are fit.

Intuitively, we would like our estimate of trend to be *dynamic*. In practical terms, this means, first, that each observation of the time series should have more or less the same influence in determining the trend line and, second, that the trend line should fit the start, middle, and end of the series equally well. OLS trend estimates are *static*, not dynamic. An OLS estimate of trend is largely determined by the first and last observations of the series.


FIGURE 2.1(b) Closing Price of IBM Common Stock (Series B from Box and Jenkins)



But there is a more basic problem with OLS detrending methods. Figure 2.1(b) shows what is perhaps the most famous of all time series: "Series B," from Box and Jenkins (1976), 369 consecutive daily closing prices of IBM common stock. Does this series follow a trend? Examining only the first half of this series, reasonable people would see an upward trend. Examining only the second half, the same people would see a downward trend. Brown (1962) demonstrated that this time series could be detrended adequately with a quadratic polynomial of the sort

$$Y_t = b_0 + b_1 t + b_2 t^2.$$

Commenting on this procedure, however, Box and Jenkins note that:

> 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. . . . 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. . . . 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 [1976: 300].

In fact, the IBM series does *not* follow a trend. For want of a better word, we say that this series "drifts," first upward and then downward.

The real difference between trend and drift is that trend is *deterministic* behavior while drift is *stochastic* behavior. Deterministic behavior can be expressed as a systematic or fixed function of time. Stochastic behavior, on the other hand, can be expressed only as the outcome of a process operating through time. Whereas future values of a deterministic process are fixed by the definition of the function, future values of a stochastic process are free to vary in a probabilistic manner. The real problem with regard to time series analysis is that, given an observed realization of finite length, it is extremely difficult to tell whether a progressive change in the level of the series is due to deterministic trend or to stochastic drift. If we model it deterministically, when in fact it is stochastic, then we may make disastrous errors in the prediction of future values or in assessing the magnitude of an exogenous intervention on the series behavior.

We do not mean to imply that trend does not exist. Indeed, many social science time series processes increase or decrease systematically. Demographic time series processes, for example, may trend due to population growth; economic time series processes may trend due to inflation; and time series measures of human performance may trend due to learning. In each case, a few known causal forces underlie the trend. In Chapter 5, we will develop bivariate and multivariate methods that can account for trends of this sort. Our point here is that a time series can drift upward or downward for extremely long periods of time due only to random forces. Unless there is a strong theoretical basis (or empirical evidence) for assuming that a time series process trends deterministically, there are great advantages to be gained by modeling it stochastically.

As Box and Jenkins note, the issue of trend versus drift is really the issue of *fitting* versus *modeling* a time series. OLS detrending methods always require an assumption that change in the realized process is due to the constant, deterministic effects of a few causal forces. If the analyst can make this assumption, then it will be best to include these exogenous forces in the time series model directly rather than attempting to exclude them indirectly through detrending.

The alternative to detrending a time series is to build a dynamic model which accounts for what appears to be (or may actually be) trend. In the next section, we develop difference equation models for drift and trend. A difference equation model accounts for both trend and drift without requiring an a priori distinction between the two. Unlike OLS detrending parameters, the parameters of a difference equation are easily estimated. But most important of all, difference equation models are *dynamic* models. Trend or drift is determined by the entire set of observations, not by the first or last observations.

## 2.2 The Random Walk and Other Integrated Processes

The IBM stock series can be thought of as the result of a random walk. A random walk process is a stochastic process wherein successive random shocks accumulate or *integrate* over time and thus a random walk is called an *integrated* process.

To illustrate the random walk, suppose a gambler bets on the flip of a fair coin. When a coin flip results in heads, the gambler wins one dollar; for tails the gambler loses one dollar. We can then define

$$a_t = +\$1 \text{ if the } t^{th} \text{ coin flip results in a head}$$
$$= -\$1 \text{ if the } t^{th} \text{ coin flip results in a tail.}$$

Because the flip of a fair coin results in heads as often as tails, the series of coin flips is a series of binomial experiments. This means that

$$P(\text{heads}) = P(\text{tails}) = \tfrac{1}{2}$$
$$Ea_t = \tfrac{1}{2}(+\$1) + \tfrac{1}{2}(-\$1) = 0,$$



in the long run, the gambler expects to break even, winning exactly as much money as he or she has lost. Also,

$$Ea_t^2 = (\tfrac{1}{2})(\tfrac{1}{2}) = \tfrac{1}{4}$$
$$Ea_t a_{t+k} = 0.$$

The last equation expresses the notion that successive flips of the coin are expected to have independent outcomes, that is, $a_t$ and $a_{t+k}$ are not related. The outcome of successive coin flips thus approximates white noise. Now define the total amount of money won or lost by the gambler after the $t^{th}$ coin flip as

$$Y_t = \text{total money won (or lost) at the end of t coin flips.}$$

With this definition, we have a random walk process. At the end of the first coin flip,

$$Y_1 = a_1,$$

that is, the total money won (or lost) consists of the money won (or lost) on the first coin flip. At the end of the second coin flip,

$$Y_2 = a_1 + a_2,$$

that is, the total money won (or lost) consists of the money won (or lost) on the first flip plus the money won (or lost) on the second flip. At the end of the third coin flip,

$$Y_3 = a_1 + a_2 + a_3,$$

and at the end of the fourth coin flip,

$$Y_4 = a_1 + a_2 + a_3 + a_4,$$

and at the end of the $t^{th}$ coin flip,

$$Y_t = a_1 + a_2 + \ldots + a_{t-1} + a_t,$$

which is the random walk. $Y_t$ is the sum of the money won (or lost) on t flips of the coin. As each random shock is *expected* to be zero (even though no single random shock can be zero), the sum of t random shocks is also expected to be zero, that is,

$$EY_t = Ea_1 + Ea_2 + \ldots + Ea_{t-1} + Ea_t = 0.$$

If we had to guess the value of the gambler's holding at the end of t flips, we would guess that the gambler had broken even. This guess would be correct in one sense but incorrect (or at least misleading) in another sense. While the expected value of $Y_t$ is zero, the process will almost always drift high above or below its expected value. In fact, it is quite unlikely that the gambler will break even by the $t^{th}$ flip.

This random walk gambling example illustrates a major difference between cross-sectional and longitudinal stochastic processes. If 1000 people each flip a coin simultaneously, we expect to observe 500 heads and 500 tails. It is unlikely that we would observe this exact outcome, of course, but the observed frequencies will ordinarily be quite close to the expected. On the other hand, if one person flips a coin 1000 times, the short-run gain ($Y_t$) is *expected* to be zero also, that is, we would expect 500 heads and 500 tails. However, the observed frequencies are quite likely to be much different than the expected. Heads or tails is likely to be in the lead throughout most of the 1000 flips. If two gamblers each have only a finite amount of cash, say $25, one of the gamblers is likely to win all of the other's money long before the 1000th flip. This phenomenon, "gambler's ruin," is a surprising property of the random walk process.[2]

In Figure 2.2(a), we show the results of a coin flip gamble. This result is typical of a random walk process. The random variate ($Y_t$) makes wide swings away from its expected level. *It drifts,* and if we had only a short

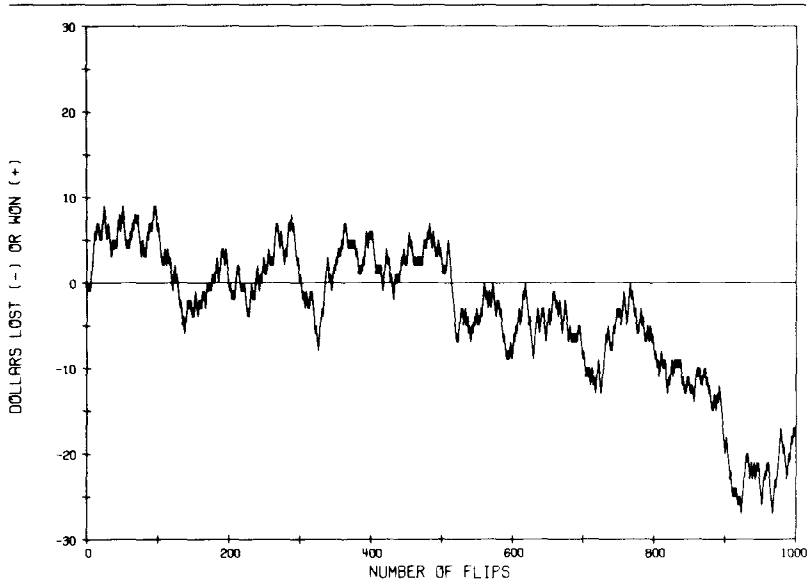



FIGURE 2.2(a) Coin Flip Gamble: A Random Walk



realization of the process, we might conclude that the process follows a trend. The probabilistic forces underlying this drift have a straightforward explanation. If positive and negative shocks were roughly equal in the short run, we would see no drift whatsoever. Relatively long runs of positive or negative shocks can occur, however. In coin flip gambling, the probability that a run of eight heads will be realized is:

$$P(\text{eight consecutive heads}) = (1/2)^8 = .0039063,$$

which is a relatively slight probability. Yet when a run of eight heads does occur, as it inevitably will, the $Y_t$ process drifts far above its expected value of zero. The level of the process stays at this zenith until a run of eight tails drives it back down. Of course, we would expect to wait a long time before realizing a counterbalancing run of eight tails.

Processes that can be thought of as random walks are frequently encountered in the social sciences. A major difference between the random walks we encounter in social science processes, however, and the random walk illustrated by the coin flip gamble is the *size* of each random shock. In coin flip gambling, random shocks are equal in absolute value because a shock must be $\pm\$1$. In the typical social science process, on the other hand, random shocks vary in size as well as in sign. There are large shocks, medium-size shocks, and small shocks in the typical social science random walk.

As an example, consider this simple model for the population of a small town. Let

$Y_t$ = the town population at the end of the $t^{th}$ month

$a_t$ = number of births minus the number of deaths in the $t^{th}$ month.

Then, for some "starting point," $Y_0$, we have the random walk process:

$$\begin{aligned}
Y_0 &= \text{the town population at the end of month zero}\\
Y_1 &= Y_0 + a_1\\
Y_2 &= Y_0 + a_1 + a_2\\
Y_3 &= Y_0 + a_1 + a_2 + a_3\\
&\vdots\\
Y_t &= Y_0 + a_1 + a_2 + \ldots + a_{t-1} + a_t,
\end{aligned}$$

which describes how the population of this small town changes from month to month.

If birth rates and death rates are equal in the long run, then the expected value of each random shock is zero. Actually, the number of births and the number of deaths in a month will seldom be exactly equal, so most random shocks will not be exactly zero. A plausible assumption is that the random shocks will be normally and independently distributed about the mean zero with constant variance, that is,

$$a_t \sim \mathrm{NID}(0,\sigma_a^2).$$

So the random shocks are a white noise process. If this model of the town population is realistic, the expected population after $t$ months is[3]:

$$\begin{aligned} EY_t &= Y_0 + Ea_1 + Ea_2 + \ldots + Ea_{t-1} + Ea_t, \\ &= Y_0 \end{aligned}$$

which was the population at the "starting point." However, we would expect to see the town population drift upward or downward for long periods of time. This feature of the realized time series is only drift, however, not trend.

It might be valuable at this point to explain exactly what a random shock is. In the town population model, random shocks are the thousands of variables which "cause" birth and death. These many factors vary across time and interact in complex and complicated patterns which we call "random." None of these factors alone could explain the birth or death rates which make up an observation of the process. But jointly, the effects of these many factors are aptly described as white noise.

Because a random walk observation is the sum of all past random shocks, there is a rather simple way to model the random walk. The series is simply *differenced*. Differencing a time series amounts to subtracting the first observation from the second, the second from the third, and so forth:

$$\begin{aligned} z_1 &= Y_1 - Y_0 \\ &= (Y_0 + a_1) - Y_0 = a_1 \\ z_2 &= Y_2 - Y_1 \\ &= (Y_0 + a_1 + a_2) - (Y_0 + a_1) = a_2 \\ z_3 &= Y_3 - Y_2 \\ &= (Y_0 + a_1 + a_2 + a_3) - (Y_0 + a_1 + a_2) = a_3 \\ &\vdots \\ z_t &= Y_t - Y_{t-1} \\ &= (Y_0 + a_1 + a_2 + \ldots + a_{t-1} + a_t) \\ &\qquad - (Y_0 + a_1 + a_2 + \ldots + a_{t-2} + a_{t-1}) = a_t. \end{aligned}$$



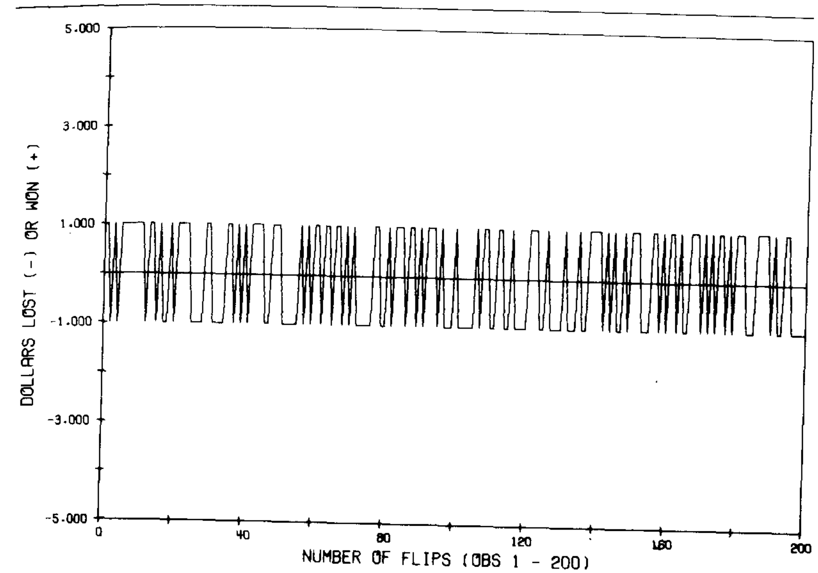



FIGURE 2.2(b) Coin Flip Gamble: A Random Walk Differenced

*In the remainder of the volume, except where explicitly noted, we will represent the differenced $Y_t$ series by $z_t$.*

By differencing the random walk process, we obtain a time series whose observations are the contemporaneous random shocks, $a_1, a_2, \ldots, a_t$. In other words, differencing transforms the random walk into a white noise process. In Figure 2.2(b), we show the coin flip gambling series after it has been differenced. The differenced series no longer drifts away from its expected value, but instead fluctuates about its mean level.

A random walk is an example of an *integrated* process in which "integration" means "addition." If a time series is the realization of an integrated process, it can be modeled by simple differencing. The random walk, for example, is well represented by an ARIMA (0,1,0) model where d = 1 is the number of differences required to make the series stationary.

So far, the random shocks or white noise process has been assumed to have a zero mean. Suppose now that the white noise process has a nonzero mean. In the coin flip gambling example, this implies that heads and tails have different payoffs. In the town population example, this implies that birth and death rates are not equal in the long run. Representing the nonzero level of the white noise process by the constant $\Theta_0$, then,

$$Ea_t = \Theta_0.$$

Successive realizations of the integrated process are now expected to be:

$$
\begin{aligned}
Y_0 &= Y_0 \\
EY_1 &= Y_0 + Ea_1 = Y_0 + \Theta_0 \\
EY_2 &= Y_0 + Ea_1 + Ea_2 = Y_0 + 2\Theta_0 \\
EY_3 &= Y_0 + Ea_1 + Ea_2 + Ea_3 = Y_0 + 3\Theta_0 \\
&\vdots \\
EY_t &= Y_0 + Ea_1 + \ldots + Ea_t = Y_0 + t\Theta_0.
\end{aligned}
$$

So when the white noise process has a nonzero mean, a random walk follows a linear trend. What happens when the $Y_t$ process is differenced?

$$
\begin{aligned}
Ez_1 &= E(Y_1 - Y_0) = Ea_1 = \Theta_0 \\
Ez_2 &= E(Y_2 - Y_1) = Ea_2 = \Theta_0 \\
Ez_3 &= E(Y_3 - Y_2) = Ea_3 = \Theta_0 \\
&\vdots \\
Ez_t &= E(Y_t - Y_{t-1}) = Ea_t = \Theta_0.
\end{aligned}
$$

The differenced series, $z_t$, is expected to equal a nonzero constant, $\Theta_0$. This leads to the difference equation model of linear trend:

$$
\begin{aligned}
Y_t - Y_{t-1} &= \Theta_0 \\
Y_t &= Y_{t-1} + \Theta_0.
\end{aligned}
$$

To illustrate the difference equation model of linear trend, consider the sequence of integers:

$$1, 2, 3, 4, \ldots, t.$$

This sequence has a perfect linear trend which we can represent as

$$Y_t = Y_0 + \Theta_0 t,$$

where $Y_0 = 0$ and $\Theta_0 = 1$. But if we difference this sequence,

$$
\begin{array}{cccc}
2-1, & 3-2, & 4-3, & \ldots, t-(t-1) \\
1, & 1, & 1, & \ldots, 1
\end{array}
$$

the differenced series is equal to the constant, $\Theta_0 = 1$. We may thus write a difference equation model for this sequence of integers as

$$
\begin{aligned}
Y_t - Y_{t-1} &= 1 \\
Y_t &= Y_{t-1} + 1.
\end{aligned}
$$



It might be useful here to draw an analogy between the OLS trend equations which we discussed in Section 2.1 and difference equation models of trend. In the OLS model, we solved the equation

$$Y_t = \hat{b}_0 + \hat{b}_1 t + a_t,$$

using t as an independent variable. The parameter $b_1$ is interpreted as the slope or linear trend of the $Y_t$ process. As noted, it is practically impossible to derive a satisfactory estimate of $b_1$. In the difference equation model, on the other hand, we use the value of $Y_{t-1}$ (rather than t) as an independent variable. The constant of the difference equation, $\Theta_0$, is interpreted in the same way that $b_1$ of the OLS equation is interpreted: as the slope or linear trend of the process. But while the OLS trend equation

$$\hat{Y}_t = \hat{b}_0 + \hat{b}_1 t$$

and the difference equation

$$\hat{Y}_t = Y_{t-1} + \hat{\Theta}_0,$$

describe exactly the same deterministic trend (when $\hat{b}_1 = \hat{\Theta}_0$, that is), there a major *practical* difference between these two equations.[4] There is no satisfactory method of estimating the parameter $b_1$ in the OLS trend equation but the analogous parameter of the difference equation, $\Theta_0$, is easily estimated as

$$\hat{\Theta}_0 = \bar{z} = 1/N \sum_{t=1}^{N} z_t,$$

that is, $\Theta_0$, is estimated as the mean of the differenced series. Beyond this practical point, there are substantive issues which make the difference equation the preferred formulation. We will address these issues at a later point in the chapter.

A time series process that does not require differencing (because it neither drifts nor trends) is said to be *stationary in the homogeneous sense*. A time series process that *does* require differencing (because it drifts or trends) is nonstationary in the homogeneous sense. As noted, a time series that follows a random walk or that follows a linear trend may be best represented by an ARIMA (0,1,0) model written as

$$Y_t = Y_{t-1} + \Theta_0 + a_t.$$

The sense of this ARIMA (0,1,0) model is quite simple. *The best prediction of the current time series observation* ($Y_t$) *comes from the preceding observation* ($Y_{t-1}$) *and a constant*. To distinguish between drift and trend, the analyst may examine the estimated value of $\Theta_0$ with the null hypothesis:

$$H_0\colon \hat{\Theta}_0 = 0.$$

The hypothesis may be tested with a t statistic. If $\hat{\Theta}_0$ is not statistically different from zero, the analyst must conclude that the process is drifting, not trending.

In the general case, an ARIMA (0,d,0) model implies that a time series is white noise after being differenced d times. There are two types of integrated processes which will be well represented by ARIMA (0,d,0) models, $d^{th}$-order random walks and time series processes with $d^{th}$-order polynomial trends. A second-order random walk, for example, would be:

$$Y_t = a_t + 2a_{t-1} + 3a_{t-2} + 4a_{t-3} + \ldots.$$

And differencing this process,

$$\begin{aligned} Y_t - Y_{t-1} &= a_t + 2a_{t-1} + 3a_{t-2} + 4a_{t-3} + \ldots \\ &\quad - a_{t-1} - 2a_{t-2} - 3a_{t-3} - \ldots \\ &= a_t + a_{t-1} + a_{t-2} + a_{t-3} + \ldots. \end{aligned}$$

And differencing this process,

$$\begin{aligned} (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) &= a_t + a_{t-1} + a_{t-2} + a_{t-3} + \ldots \\ &\quad - a_{t-1} - a_{t-2} - a_{t-3} - \ldots \\ &= a_t. \end{aligned}$$

Thus, a second-order random walk will be well represented by an ARIMA (0,2,0) model. To demonstrate how a $d^{th}$-order polynomial trend can be accommodated by an ARIMA (0,2,0) model, consider the sequence of squared integers

$$1, 4, 9, 16, 25, \ldots, t^2$$

If we difference this sequence,

$$\begin{array}{llllll} 4-1, & 9-4, & 16-9, & 25-16, & \ldots, & t^2 - (t-1)^2 \\ 3, & 5, & 7, & 9, & \ldots, & 2t-1 \end{array}$$

and if we difference it again,

$$\begin{array}{lllll} 5-3, & 7-5, & 9-7, & \ldots, & (2t-1) - (2t-3) \\ 2, & 2, & 2, & \ldots, & 2 \end{array}$$



we obtain a sequence of constants. Again, to distinguish $d^{th}$-order drift from $d^{th}$-order trend, the analyst need only test the statistical significance of $\hat{\Theta}_0$. In the social sciences, a time series will ordinarily have to be differenced only once. While ARIMA (0,1,0) processes are quite common, higher order random walks and higher order polynomial trends are rare.

We will close this discussion of integrated processes by applying an input-output analogy to the ARIMA (0,1,0) process. White noise is conceptualized as the "driving force" of a time series process. We may view an ARIMA (0,1,0) model as a "black box" in which the *input* is white noise and the *output* is a time series. For processes that drift or trend, the ARIMA (0,1,0) black box *integrates* random shocks. Once a shock enters the black box, it remains inside, influencing all future outputs. We represent this as

$$a_t \dashrightarrow \boxed{\sum_{i=1}^{\infty} a_{t-i}} \dashrightarrow Y_t.$$

All future outputs of the ARIMA (0,1,0) black box will contain the random shock $a_t$ as well as all prior random shocks back into the infinitely distant past. To unlock this black box, we simply difference the time series.

## 2.3 The Backward Shift Operator

At this point, we introduce the backward shift operator, B, such that

$$B(Y_t) = Y_{t-1}.$$

This expression does not mean "B multiplies $Y_t$," but rather, means that "B operates on $Y_t$ to shift it backward one point in time." B is thus similar to other logical operators such as the derivative or integral operators in calculus or the natural logarithm operator in algebra.

The properties of the backward shift operator are:

$$B^n(Y_t) = Y_{t-n}$$

and

$$B^n B^m(Y_t) = B^{n+m}(Y_t) = Y_{t-n-m}.$$

The operator obeys all the laws of exponents that we routinely use in polynomial algebra. We will make immediate use of the backward shift operator to difference a time series. For this purpose, we use the operator expression

$$(1 - B)Y_t = (1)Y_t - (B)Y_t = Y_t - Y_{t-1} = z_t.$$

Second differencing is accomplished by

$$(1 - B)^2 Y_t = (1 - 2B + B^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2},$$

which is the same result we obtain by differencing a time series and then differencing it again. To demonstrate this identity, we difference the time series, obtaining

$$\begin{aligned} z_1 &= Y_1 - Y_0 \\ z_2 &= Y_2 - Y_1 \\ &\vdots \\ z_t &= Y_t - Y_{t-1}. \end{aligned}$$

We then difference this series, obtaining

$$\begin{aligned} z_1^* &= z_1 - z_0 \\ z_2^* &= z_2 - z_1 = (Y_2 - Y_1) - (Y_1 - Y_0) \\ z_3^* &= z_3 - z_2 = (Y_3 - Y_2) - (Y_2 - Y_1) \\ &\vdots \\ z_t^* &= z_t - z_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}). \end{aligned}$$

We see that the general term $z_t^*$ is:

$$\begin{aligned} z_t^* &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - Y_{t-1} - Y_{t-1} + Y_{t-2} = Y_t - 2Y_{t-1} + Y_{t-2} \\ &= (1 - 2B + B^2)Y_t \\ &= (1 - B)^2 Y_t, \end{aligned}$$

which is the identity we wanted to demonstrate. [redacted] to [redacted].

Another useful property of the backward shift operator is [redacted]. The property of invertibility allows us to move forward and backward in time by applying the operator and its inverse to a time series. We see an analogous relationship between the natural logarithm operator, Ln, and exponentiation. For example,

$$\mathrm{Ln}(K) = x$$

then

$$e^x = K.$$



We will be using the natural logarithm operator at a later point. For now, we introduce it only to illustrate the principle of an inverse operator. The inverse of the backward shift operator, B, is the *forward* shift operator, F. The forward shift operator has the same properties as the backward shift operator, that is,

$$F^n(Y_t) = Y_{t+n}$$

and

$$F^n F^m(Y_t) = F^{n+m}(Y_t) = Y_{t+n+m}.$$

The inverse relationship between B and F is defined by

$$(B)\,(F) = (F)\,(B) = 1,$$

$$(B)\,(F)\,Y_t = (B)Y_{t+1} = Y_t,$$

and

$$B^n F^m Y_t = F^{m-n} Y_t = Y_{t+m-n}.$$

In other words, "F undoes what B has done."

In the remainder of this volume, we will use only the backward shift operator, B. To denote the forward shift operator, F, we will use $B^{-1}$, and thus,

$$B^{-1}B = 1.$$

When the backward shift operator is used to difference a time series, we denote the inverse operation by $(1 - B)^{-1}$, and thus

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

$$(1 - B)^{-1} z_t = Y_t,$$

so

$$(1 - B)^{-1}(1 - B)Y_t = Y_t.$$

In practice, the inverse of the differencing operation implies division,

$$(1 - B)^{-1} z_t = z_t / (1 - B)$$

and division by the backward shift operator is a difficult concept to grasp. The inverse operator can be evaluated with a Taylor series expansion, however, as the infinite series

$$(1 - B)^{-1} = (1 + B + B^2 + \ldots + B^{n-1} + B^n + \ldots).$$

We will not go into the derivation of this identity but instead direct the interested reader to any introductory calculus text. However, we will demonstrate the identity by showing that the differencing operator, $(1 - B)$, and the infinite series are inverses. First, differencing a $Y_t$ process, we obtain

$$(1 - B)Y_t = Y_t - Y_{t-1} = z_t.$$

Then, applying the inverse operator to the $z_t$ process, we obtain

$$\begin{aligned}(1 - B)^{-1}z_t &= z_t(1 + B + B^2 + \ldots B^{n-1} + B^n + \ldots) \\ &= z_t + z_{t-1} + z_{t-2} + \ldots + z_{t-n+1} + z_{t-n} + \ldots \\ &= (Y_t - Y_{t-1}) + (Y_{t-1} - Y_{t-2}) + (Y_{t-2} - Y_{t-3}) + \ldots \\ &\quad + (Y_{t-n+1} - Y_{t-n}) + (Y_{t-n} - Y_{t-n-1}) + \ldots \\ &= Y_t + (Y_{t-1} - Y_{t-1}) + (Y_{t-2} - Y_{t-2}) + \ldots \\ &\quad + (Y_{t-n} - Y_{t-n}) + \ldots \\ &= Y_t,\end{aligned}$$

which demonstrates the identity of $(1 - B)^{-1}$ and the infinite series.

In the following sections of this chapter and in the following chapters, we will routinely use the backward shift operator to describe time series models. The operator greatly simplifies our algebra and, indeed, some of the higher order ARIMA models cannot be written economically without the operator convention.

## 2.4 Variance Stationarity[5]

As noted in Section 2.2, ARIMA models require a time series process to be stationary in the homogeneous sense. While homogeneous sense stationarity is a *necessary* condition of an ARIMA model, however, it is not a *sufficient* condition. Before we can properly represent a time series with an ARIMA model, the time series must also have a stationary *variance*.

Our working definition of homogeneous sense stationarity is based on the process *level*. If a time series process is stationary in the homogeneous sense, then that process has a single constant level throughout its course, that is,

$$EY_t = \Theta_0$$

for all t. By this working definition, if a process drifts or trends, it is *not* stationary in the homogeneous sense. This presents no real problem, however, because when the process is differenced,

$$z_t = (1 - B)^d Y_t$$

and

$$Ez_t = \Theta_0$$

for all t. As the differenced process has a single constant level throughout its course, it is stationary in the homogeneous sense. From this working definition, we might say that a process that is stationary in the homogeneous sense is *stationary in its level*. Finite realizations of such processes appear "flat" or trendless. If we divide the series into two segments of equal length, the first segment will have the same level as the second segment.

A process that is stationary in the homogeneous sense (or one that has been made so by differencing) need not be stationary in its *variance*, however. A process that is stationary in variance will have a single constant variance throughout its course, that is,

$$E(Y_t - \Theta_0)^2 = \sigma_a^2$$

for all t or

$$E(z_t - \Theta_0)^2 = \sigma_a^2$$

for all t. Because the expression for variance is based on the process level, $\Theta_0$, it follows that any process that is stationary in variance is also stationary in level (or has been made so by differencing). The converse is not true, however. A process that is stationary in level need not be stationary in variance.

In Figure 2.4(a), we show a time series of nonfatal disabling mine injuries for the United States during the period 1930–1977. Nonfatal disabling injuries decreased systematically during this period, trending downward. We can guess (correctly) that this series is nonstationary in the homogeneous sense and must be differenced. In Figure 2.4(b), we show the first-differenced time series. Differencing has effectively detrended the series, that is, has made the series stationary in the homogeneous sense. *Note, however, that in both Figures 2.4(a) and Figures 2.4(b), the process variance decreases steadily throughout the length of the series.* Year-to-year fluctuations are much larger in the first half of the series than in the second half. We can guess (correctly) that this series, even after differencing, is not stationary in its variance.

If our experiences are typical, most of the time series that social scientists will be interested in are either stationary or else nonstationary only in the

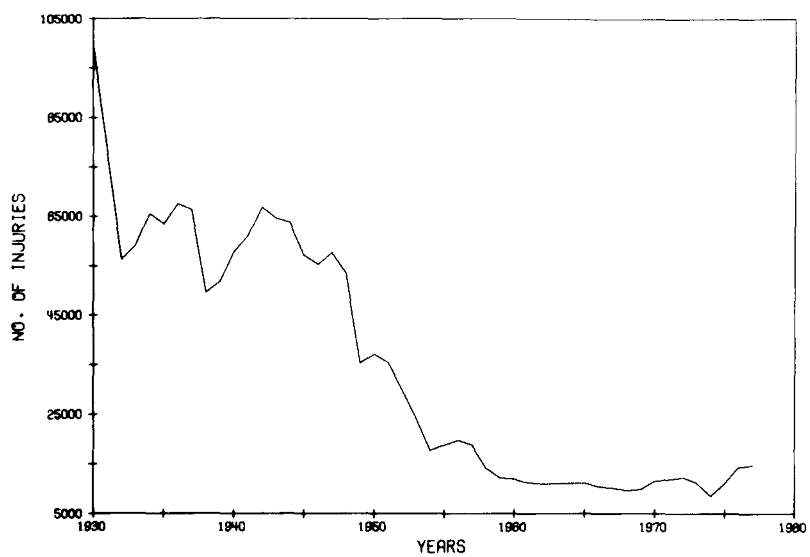



FIGURE 2.4(a) Nonfatal Disabling Mine Injuries, 1930–1978

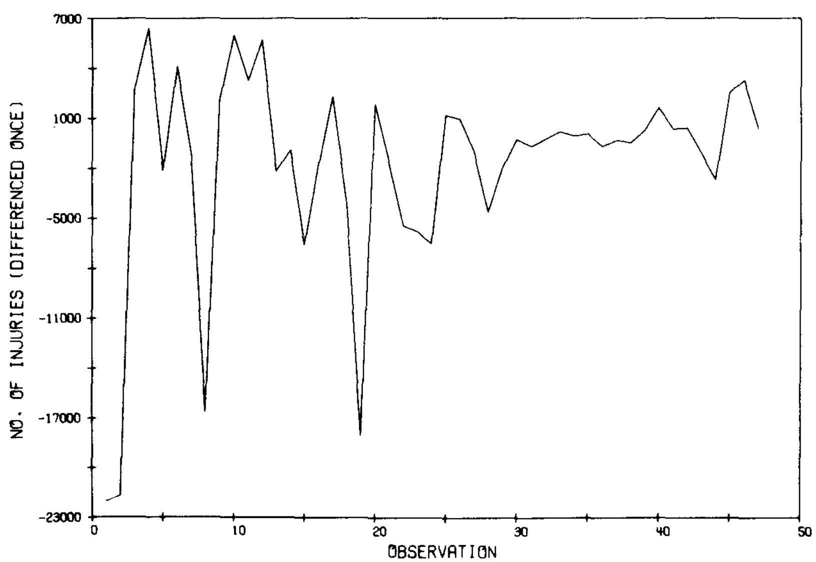



FIGURE 2.4(b) Nonfatal Disabling Mine Injuries, Differenced

homogeneous sense. In practice, then, the analyst will find that most time series can be made stationary in the larger sense (in level and variance) by differencing. Time series of the sort shown in Figure 2.4(a) are not totally



uncommon, however. Many social processes have naturally defined "floors" which constrain the stochastic behavior of the process. As the process approaches its floor, process variance is constrained. In the case of the mine injury series, process variance is roughly proportional to the process level, a relationship which we express as

$$EY_t = \Theta_0 t$$

and

$$\sigma_a^2 \propto \Theta_0 t.$$

What this means simply is that, first, the level of the process decreases from observation to observation by the quantity $\Theta_0$. Second, as the level decreases, process variance decreases proportionally.

Fortunately, there is a rather simple transformation which may be applied to such processes to make them stationary in the larger sense. In Figure 2.4(c), we show a time series of natural logarithms of the nonfatal disabling injuries. The log-transformed series still follows a downward trend, that is, is still nonstationary in the homogeneous sense. However, log-transformation has made the year-to-year fluctuations roughly the same in both halves of the series. In Figure 2.4(d), we show the first-differenced log-

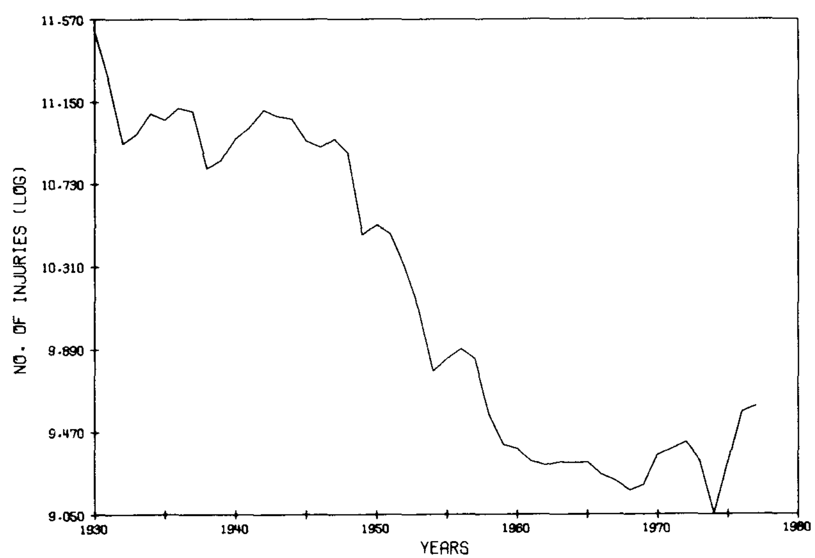



FIGURE 2.4(c) Nonfatal Disabling Mine Injuries, Logged

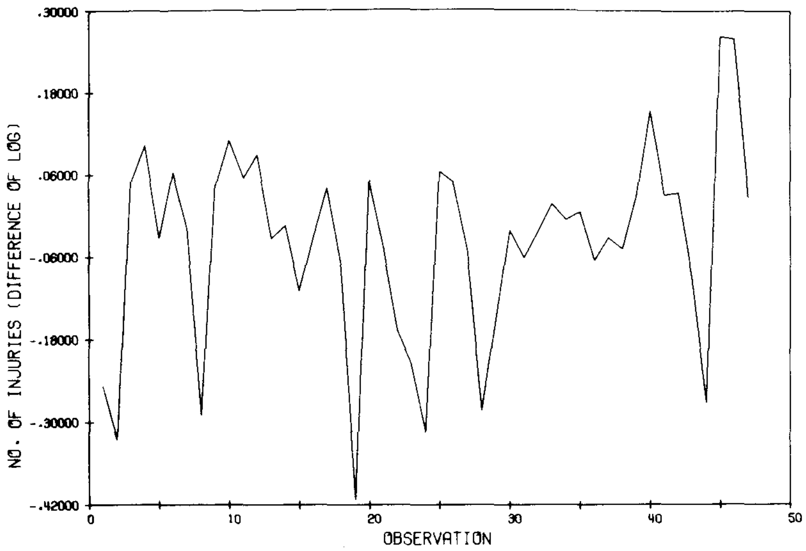



FIGURE 2.4(d) Nonfatal Disabling Mine Injuries, Logged and Differenced

transformed series. After differencing, the series has a single level and single variance throughout its course. We may guess (correctly) that the first-differenced log-transformation series is stationary in the larger sense.

In general, whenever the variance of a time series decreases (or increases) as the level of the series decreases (or increases), the series can be made stationary in the larger sense (in both level and variance) by log-transformation and differencing. Log-transformation is effective in such cases because the absolute value of successive random shocks changes systematically. Thus,

$$a_t^2 = Ka_{t-1}^2$$

for some constant of proportionality K. From this we see that

$$\mathrm{Ln}(a_t^2) = \mathrm{Ln}(a_{t-1}^2) + \mathrm{Ln}(K)$$
$$\mathrm{Ln}(a_t^2) - \mathrm{Ln}(a_{t-1}^2) = \mathrm{Ln}(K).$$

Log-transformation and differencing results in a constant variance. Readers who are familiar with variance stabilizing transformations in the context of regression models (e.g., Draper and Smith, 1966: 131–134) will recognize similarities in the approach taken here.



We note in conclusion that the analyst must always be able to assume that a time series process is stationary in both its level *and* variance. A series that is nonstationary in variance must be transformed prior to analysis. If the analyst ignores this problem (and if the assumption of larger sense stationarity is not satisfied), the analysis may lead to incorrect inferences about the time series process. These assumptions are similar in kind to the assumptions typically required of a regression analysis. Homoskedasticity of disturbance terms must be assumed to derive least-squares estimates of regression model parameters, for example, and ARIMA models are not different in this respect.

## 2.5 Autoregressive Processes of Order p: ARIMA (p,0,0) Models

Autoregressive processes of order p may be modeled by using p lagged observations of the series to predict the current observation, that is,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + a_t,$$

which, using the backward shift operator, may be rewritten as

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) y_t = a_t.$$

As a convention, we will use $y_t$ to denote a *deviate* time series, that is,

$$y_t = Y_t - \Theta_0$$

for the stationary process, $Y_t$. Autoregressive processes are always stationary processes and ARIMA (p,0,0) models are always stationary models. Should a process be nonstationary, however, it may happen that its difference will be well represented by an ARIMA (p,0,0) model. Thus, for the nonstationary process, $Y_t$,

$$z_t = (1 - B)^d Y_t$$
$$y_t = z_t - \Theta_0$$
$$= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + a_t.$$

The process is well represented by an ARIMA (p,d,0) model.

There is no essential difference in the algebras of ARIMA (p,0,0) and ARIMA (p,d,0) models except, of course, that ARIMA (p,d,0) models imply a nonstationary process. To simplify our discussion, then, we will deal only with ARIMA (p,0,0) models and only with deviate time series.

When a process is nonstationary, of course, our argument will apply to the differenced process.

The most commonly encountered autoregressive processes in the social sciences are first-order autoregressive processes. First-order autoregression is well represented by an ARIMA (1,0,0) model:

$$y_t = \phi_1 y_{t-1} + a_t$$
$$(1 - \phi_1 B) y_t = a_t.$$

It may be instructive to think of the ARIMA (1,0,0) model as an OLS regression model in which the current time series observation is regressed on the preceding time series observation. Unlike the OLS regression model, however, the parameter $\phi_1$ must be constrained to the interval

$$-1 < \phi_1 < +1.$$

The purpose of these constraints will be soon made clear.

Tracking a random shock through time, the ARIMA (1,0,0) model exhibits a distinctive pattern of stochastic behavior called *autoregression*. Let

$$y_0 = a_0,$$

then,

$$\begin{aligned}
y_1 &= \phi_1 y_0 + a_1 \\
&= \phi_1 a_0 + a_1 \\
y_2 &= \phi_1 y_1 + a_2 \\
&= \phi_1(\phi_1 a_0 + a_1) + a_2 \\
&= \phi_1^2 a_0 + \phi_1 a_1 + a_2 \\
y_3 &= \phi_1 y_2 + a_3 \\
&= \phi_1(\phi_1^2 a_0 + \phi_1 a_1 + a_2) + a_3 \\
&= \phi_1^3 a_0 + \phi_1^2 a_1 + \phi_1 a_2 + a_3 \\
&\vdots \\
y_t &= \phi_1^t a_0 + \phi_1^{t-1} a_1 + \ldots + \phi_1 a_{t-1} + a_t.
\end{aligned}$$



Recalling that $\phi_1$ is a fraction, successive powers of $\phi_1$ converge to zero. If $\phi_1 = .5$, for example, then

$$\begin{aligned}
\phi_1^2 &= .25 \\
\phi_1^3 &= .125 \\
\phi_1^4 &= .0625 \\
&\vdots \\
\phi_1^t &\simeq 0
\end{aligned}$$

for t large. So while the initial random shock stays in the process indefinitely, its impact diminishes exponentially. After one observation, the impact of $a_0$ is only a fraction of its initial impact. By time t, the impact of $a_0$ is so small that we may think of it as zero. Returning to the black box input-output analogy used in Section 2.2 to describe integrated processes, we see that a portion of the random shock "leaks" out of the autoregressive black box as time passes:

$$a_t \dashrightarrow \boxed{\sum_{i=1}^{\infty} \phi_1^i a_{t-i}} \dashrightarrow y_t$$
$$\text{"leakage"} \dashrightarrow \sum_{i=1}^{\infty} (1 - \phi_1^i)\, a_{t-i}$$

The autoregressive black box accumulates random shocks in the same manner as the integrated black box accumulates random shocks. There is a significant difference between these two black boxes, however. Random shocks leak out of the autoregressive black box over time. For an initial random shock, $a_0$, the portion remaining in the black box at successive points in time is:

| Time | Portion Remaining | Portion Lost Through Leakage |
|---|---|---|
| $t = 0$ | $a_0$ | . . . |
| $t = 1$ | $\phi_1 a_0$ | $(1 - \phi_1)a_0$ |
| $t = 2$ | $\phi_1^2 a_0$ | $(1 - \phi_1^2)a_0$ |
| $\vdots$ | | |
| $t = t$ | $\phi_1^t a_0 \approx 0$ | $(1 - \phi_1^t)a_0 \approx a_0.$ |

After t moments, the portion of $a_0$ remaining in the black box is so small that we may think of it as zero. The portion of $a_0$ lost through leakage is so large that we may think of it as 100% of $a_0$.

To be perfectly correct in this analogy, the $y_t$ process must then be passed through another black box:

$$y_t \dashrightarrow \boxed{+\Theta_0} \dashrightarrow Y_t,$$

which adds a constant to each observation, thus transforming the $y_t$ process into the $Y_t$ process. Of course, for an ARIMA (p,d,0) model, a nonstationary process is implied and the sequence of black boxes is:

$$y_t \dashrightarrow \boxed{+\Theta_0} \dashrightarrow z_t,$$

and then,

$$z_t \dashrightarrow \boxed{\sum_{i=1}^{\infty} z_{t-i}} \dashrightarrow Y_t.$$

Thus, while we have confined our discussion to ARIMA (p,0,0) models and to deviate time series, it is clear that any argument can be generalized to ARIMA (p,d,0) models and to the raw $Y_t$ process by simply defining an appropriate black box filter.

Autoregression refers to a stochastic behavior in which a random shock has an exponentially diminishing impact over time. While it may not be obvious, this aspect of autoregression is determined by the bounds placed on the parameter $\phi_1$:

$$-1 < \phi_1 < +1.$$

These are called the *bounds of stationarity for autoregressive parameters*. If an ARIMA (1,0,0) model is written as

$$(1 - \phi_1 B)y_t = a_t,$$

the implications of these bounds became apparent. If $\phi_1 = 1$,

$$(1 - B)y_t = a_t,$$

the ARIMA (1,0,0) model becomes an ARIMA (0,1,0) model which reflects a nonstationary process. Nonstationary processes result from an integration of random shocks over time:

$$Y_t = Y_0 + a_1 + a_2 + \ldots + a_{t-1} + a_t.$$



The impact of shocks from the distant past do not diminish over time and this is not consistent with autoregressive behavior.

This implication may be demonstrated again by introducing an important property of the ARIMA (1,0,0) model: The ARIMA (1,0,0) model can be expressed identically as *the infinite sum of exponentially weighted past random shocks*. To deduce this important property, write the ARIMA (1,0,0) model as

$$y_t = \phi_1 y_{t-1} + a_t$$

and

$$y_{t-1} = \phi_1 y_{t-2} + a_{t-1}.$$

Substituting for $y_{t-1}$,

$$\begin{aligned} y_t &= \phi_1(\phi_1 y_{t-2} + a_{t-1}) + a_t \\ &= \phi_1^2 y_{t-2} + \phi_1 a_{t-1} + a_t. \end{aligned}$$

Similarly,

$$y_{t-2} = \phi_1 y_{t-3} + a_{t-2}$$

which we may substitute into the expression for $y_t$ to obtain

$$\begin{aligned} y_t &= \phi_1^2(\phi_1 y_{t-3} + a_{t-2}) + \phi_1 a_{t-1} + a_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 a_{t-2} + \phi_1 a_{t-1} + a_t. \end{aligned}$$

Continuing the substitution process back into the infinite past,

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

which demonstrates that an ARIMA (1,0,0) process is identical to the infinite sum of exponentially weighted past shocks.

This is a rather important point. In Section 2.3, we demonstrated that the inverse difference operator was identical with the infinite series

$$\begin{aligned} (1 - B)^{-1} &= 1 + B + B^2 + \ldots + B^k + \ldots \\ &= \sum_{k=0}^{\infty} B^k. \end{aligned}$$

Writing the ARIMA (1,0,0) model as an infinite sum of exponentially weighted past random shocks, it is clear that the inverse first-order autoregressive factor is identical with the infinite series

$$(1 - \phi_1 B)^{-1} = 1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^k B^k + \ldots$$
$$= \sum_{k=0}^{\infty} \phi_1^k B^k.$$

This identity gives a "solution" of the ARIMA (1,0,0) model as

$$(1 + \phi_1 B)y_t = a_t$$
$$y_t = (1 - \phi_1 B)^{-1} a_t$$
$$= (1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^k B^k + \ldots)a_t$$
$$= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots + \phi_1^k a_{t-k} + \ldots.$$

Moreover, as the infinite series formulation of the inverse autoregressive operator converges to zero, it may be truncated after a few terms without appreciably changing an evaluation of the solution. For example, if $\phi_1 = .5$,

$$y_t = a_t + .5a_{t-1} + .25a_{t-2} + .0625a_{t-3} + .03125a_{t-4} + \ldots + (.5)^k a_{t-k} + \ldots.$$

After a few terms, the value of $\phi_1^k$ is approximately zero. All successive terms of the series could be ignored.

While the infinite series identity is important in and of itself, it gives a crucial insight into the nature of process stationarity. Whenever $\phi_1 = 1$, an ARIMA (0,1,0) model is implied but when $\phi_1 > 1$, *a growth process is implied*. Suppose, for example, that $\phi_1 = 1.5$. The ARIMA (1,0,0) process is then

$$(1 - 1.5B)y_t = a_t$$
$$y_t = (1 + 1.5B)^{-1} a_t$$
$$= (1 + 1.5B + 2.25B^2 + \ldots + (1.5)^k B^k + \ldots)a_t$$
$$= a_t + 1.5a_{t-1} + 2.25a_{t-2} + \ldots (1.5)^k a_{t-k} + \ldots.$$

Past random shocks have larger weights and, thus, are more important to the $y_t$ process than the current random shock. As time passes, the random shock becomes more important.

As an example of such a process, consider the hypothetical situation in



which a savings account earns 5% monthly compounded interest. If the monthly deposit to the account is essentially random, that is,

$a_t$ = deposit in the $t^{th}$ month

and

$y_t$ = savings and accrued interest in the $t^{th}$ month,

then the level of the savings account is given by

$$y_t = 1.05y_{t-1} + a_t.$$

In other words, the level of the savings account in this month is equal to 105% of the previous month's level ($y_{t-1}$) plus the current month's deposit ($a_t$). The current level of the account can be expressed identically in terms of past deposits only.

$$y_t = a_t + 1.05a_{t-1} + 1.1025a_{t-2} + 1.157625a_{t-3} + \ldots + (1.05)^k a_{t-k} + \ldots.$$

When expressed in this form, it is apparent that *the most important determinant of the current level of the savings account is the first deposit*. This is contrary to the principles of autoregression, and to avoid this situation, the value of $\phi_1$ must be constrained to the bounds of stationarity for autoregressive parameters.

The principles of autoregression can be generalized to higher order processes. An ARIMA (2,0,0) model, for example, is:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t$$

or

$$(1 - \phi_1 B - \phi_2 B^2)y_t = a_t.$$

For p = 2, the current series observation is equal to a portion of the two preceding observations. In the general case, an ARIMA (p,0,0) model reflects a type of stochastic behavior in which the present observation is equal to portions of the p preceding observations. In practice, however, we have found that most autoregressive social science processes are well represented by ARIMA (1,0,0) models. ARIMA (2,0,0) processes are less common and higher order ARIMA (p,0,0) processes are quite rare. Indeed, in the next section, we will describe a class of moving average ARIMA models which will parsimoniously represent higher order autoregressive processes.

To demonstrate the principles of autoregression for an ARIMA (2,0,0) process, we begin with the expression for $y_t$ and then substitute backward into time:

$$\begin{aligned}
y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t \\
&= \phi_1(\phi_1 y_{t-2} + \phi_2 y_{t-3} + a_{t-1}) \\
&\quad + \phi_2(\phi_1 y_{t-3} + \phi_2 y_{t-4} + a_{t-2}) \\
&= a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} \\
&\quad + \phi_1^2 y_{t-2} + 2\phi_1\phi_2 y_{t-3} + \phi_2^2 y_{t-4} \\
&= a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} \\
&\quad + \phi_1^2(\phi_1 y_{t-3} + \phi_2 y_{t-4} + a_{t-2}) \\
&\quad + 2\phi_1\phi_2(\phi_1 y_{t-4} + \phi_2 y_{t-5} + a_{t-3}) \\
&\quad + \phi_2^2(\phi_1 y_{t-5} + \phi_2 y_{t-6} + a_{t-4}) \\
&= a_t + \phi_1 a_{t-1} + (\phi_2 + \phi_1^2)a_{t-2} + 2\phi_1\phi_2 a_{t-3} + \phi_2^2 a_{t-4} \\
&\quad + \phi_1^3 y_{t-3} + 3\phi_1^2\phi_2 y_{t-4} + 3\phi_1\phi_2^2 y_{t-5} + \phi_2^3 y_{t-6}
\end{aligned}$$

and so forth and so on. Clearly, with enough arithmetic, we can write the ARIMA (2,0,0) process as the infinite sum of exponentially weighted past shocks. Unlike the ARIMA (1,0,0) process, however, the infinite series is not a simple function of $\phi_1$ and $\phi_2$.

Of course, like the ARIMA (1,0,0) process, the parameters $\phi_1$ and $\phi_2$ must be constrained to the bounds of stationarity for autoregressive parameters. On commonsense grounds, one might think that the bounds of stationarity for an ARIMA (2,0,0) model should be:

$$-1 < \phi_1, \phi_2 < +1.$$

But there are a number of $\phi_1\phi_2$ interaction terms in the infinite series, so these simple bounds will not ensure stationarity. While we will not do so here, it can be demonstrated that the bounds of stationarity for an ARIMA (2,0,0) model are:[6]

$$\begin{aligned}
-1 &< \phi_2 < +1 \\
\phi_1 + \phi_2 &< +1 \\
\phi_2 - \phi_1 &< +1.
\end{aligned}$$



So long as the values of $\phi_1$ and $\phi_2$ satisfy these bounds, a stationary process is implied. To demonstrate this, let

$$\phi_1 + \phi_2 = 1$$

or

$$\phi_2 = 1 - \phi_1,$$

which violates the bounds of stationarity. The ARIMA (2,0,0) process then becomes:

$$\begin{aligned}
(1 - \phi_1 B - \phi_2 B^2)y_t &= a_t \\
(1 - \phi_1 B - (1 - \phi_1)B^2)y_t &= a_t \\
(1 - \phi_1 B + \phi_1 B^2 - B^2)y_t &= a_t \\
y_t - y_{t-2} &= \phi_1(y_{t-1} - y_{t-2}) + a_t,
\end{aligned}$$

which is a difference equation representation of a nonstationary process.

ARIMA (p,0,0) models describe a type of stochastic behavior in which the current observation is a weighted sum of p preceding observations. In general, ARIMA (p,0,0) processes can be written as an infinite sum of exponentially weighted past shocks. As noted, most autoregressive social processes can be well represented by ARIMA (1,0,0) models. ARIMA (2,0,0) processes are less common. Higher order ARIMA (p,0,0) processes are extremely rare and, in any event, are more parsimoniously represented by the ARIMA (0,0,q) models which we will now introduce.

## 2.6 Moving Average Processes of Order q: ARIMA (0,0,q) Models

A white noise process is conveniently thought of as the "driving force" of all ARIMA (p,d,q) models. We have shown that integrated processes are realized as the sum of all past shocks and, thus, are well represented by ARIMA (0,d,0) models. Autoregressive processes are realized as an exponentially weighted sum of all past shocks and, thus, are well represented by ARIMA (p,0,0) models. The unifying factor between integrated and autoregressive processes is the persistence of a random shock. Each shock persists indefinitely, although for autoregressive processes, the impact of a shock diminishes rapidly. Moving average processes, in contrast, are characterized by a *finite* persistence. A random shock enters the system and then persists for no more than q observations before it vanishes entirely.

Moving average processes are well represented by ARIMA (0,0,q) models. An ARIMA (0,0,1) model is written as

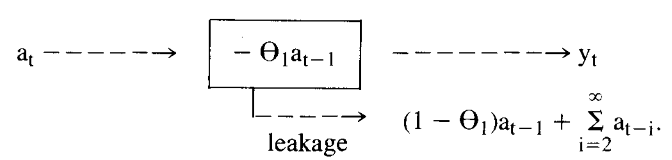
$$y_t = a_t - \Theta_1 a_{t-1}$$

or

$$y_t = (1 - \Theta_1 B)a_t.$$

As are autoregressive processes, moving average processes are stationary. Again, there is no essential difference between ARIMA (0,0,q) and ARIMA (0,d,q) models except, of course, that ARIMA (0,d,q) models reflect a nonstationary process. To simplify our discussion, we will deal only with stationary processes and ARIMA (0,0,q) models, noting that when a process is nonstationary, our argument applies to the differenced process.

The general principle of moving average processes is that a random shock persists for exactly q observations and then is gone. Using the black box input-output analogy again, we can think of a first-order moving average process as




Here the random shock $a_t$ enters the black box, is joined with a portion of the preceding random shock, $a_{t-1}$, and leaves the black box as the time series observation $y_t$. A portion of the preceding random shock, $a_{t-1}$, has already leaked out of the system along with all prior random shocks back into the distant past.

This basic principle of the moving average process can be generalized to higher order processes. For example, an ARIMA (0,0,2) process is written as

$$y_t = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2}$$
$$y_t = (1 - \Theta_1 B - \Theta_2 B^2)a_t.$$

Here the current time series observation consists of the current shock, $a_t$, and portions of the two preceding shocks, $a_{t-1}$ and $a_{t-2}$. A shock persists for only two observations and then is gone. In the general case, an ARIMA (0,0,q) process is written as

$$y_t = (1 - \Theta_1 B - \Theta_2 B^2 - \ldots - \Theta_{q-1} B^{q-1} - \Theta_q B^q)a_t.$$



And in the general case, a shock persists for exactly q observations and then is gone. In practice, we have found that most moving average social processes are well represented by ARIMA (0,0,1) models. ARIMA (0,0,2) models are less common and higher order moving average models are extremely rare.

In Section 2.5, we noted that ARIMA (p,0,0) processes in which $p > 2$ will be more parsimoniously represented by lower order moving average models. This hints at an instructive relationship between ARIMA (p,0,0) and ARIMA (0,0,q) models. Writing an ARIMA (0,0,1) process at two points in time,

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

and

$$y_{t-1} = a_{t-1} - \Theta_1 a_{t-2}$$

thus

$$a_{t-1} = y_{t-1} + \Theta_1 a_{t-2}.$$

If we substitute the expression for $a_{t-1}$ into the expression for $y_t$,

$$\begin{aligned} y_t &= a_t - \Theta_1(y_{t-1} + \Theta_1 a_{t-2}) \\ &= a_t - \Theta_1 y_{t-1} - \Theta_1^2 a_{t-2}. \end{aligned}$$

Similarly, for $a_{t-2}$,

$$\begin{aligned} y_t &= a_t - \Theta_1 y_{t-1} - \Theta_1^2(y_{t-2} + \Theta_1 a_{t-3}) \\ &= a_t - \Theta_1 y_{t-1} - \Theta_1^2 y_{t-2} - \Theta_1^3 a_{t-3}. \end{aligned}$$

And continuing this substitution back into time,

$$y_t = a_t - \sum_{i=1}^{\infty} \Theta_1^i y_{t-i}.$$

So an ARIMA (0,0,1) process can be expressed identically as the infinite sum of exponentially weighted past observations of the process. While we will not do so here, it can be demonstrated that any ARIMA (0,0,q) process can be expressed as an infinite series of exponentially weighted past observations.

Given this relationship, it is clear that we must constrain the values of moving average parameters. These constraints are identical to the constraints placed on autoregressive parameters. For an ARIMA (0,0,1) model,

$$-1 < \Theta_1 < +1$$

and for an ARIMA (0,0,2) model,

$$-1 < \Theta_2 < +1$$
$$\Theta_1 + \Theta_2 < +1$$
$$\Theta_2 - \Theta_1 < +1.$$

These are called the *bounds of invertibility for moving average parameters*. While the name is different, the bounds of invertibility play much the same role as the bounds of stationarity for autoregressive parameters. Suppose, for example, that $\Theta_1 = 1.5$ for some ARIMA (0,0,1) process. Then, writing this process as the infinite series,

$$y_t = a_t - 1.5y_{t-1} - 2.25y_{t-2} - 3.37y_{t-3} - \ldots$$

the weights associated with observations in the distant past become greater and greater.

In practice, when $\phi_p$ or $\Theta_q$ parameters exceed the bounds of stationarity or the bounds of invertability, the analyst may assume either that the series is nonstationary and must be differenced or that it was differenced too many times. Even if the parameters do not exceed their bounds, however, large values of $\phi_1$ or $\Theta_1$ may indicate that the model selected for the time series is inappropriate. We will return to these issues at a later point.

## 2.7 The General ARIMA (p,d,q) Model

Our development so far has treated p and q, the ARIMA autoregressive and moving average structural parameters, separately. Our development of autoregressive processes, for example, considered only ARIMA (p,0,0) models, models in which $q = 0$. Likewise, our development of moving average processes considered only ARIMA (0,0,q) models, models in which $p = 0$. This development reflects our experience with social science time series. If our experiences are typical, only a few social science time series in a thousand will have both p and $q \neq 0$.

While ARIMA (p,0,q) models are not logically *impossible*, the relationships between ARIMA (p,0,0) and ARIMA (0,0,q) models which we have discussed place some logical limits on ARIMA (p,0,q) models. These limitations are ordinarily referred to as the limitations of *parameter redundancy*. To illustrate parameter redundancy, write the ARIMA (1,0,1) process as

$$y_t = \phi_1 y_{t-1} + a_t - \Theta_1 a_{t-1}$$

or

$$(1 - \phi_1 B)y_t = (1 - \Theta_1 B)a_t$$



and solving for $y_t$

$$y_t = (1 - \phi_1 B)^{-1}(1 - \Theta_1 B)a_t.$$

In Section 2.4, we demonstrated that the inverse autoregressive operator, $(1 - \phi_1 B)^{-1}$, is identical to an infinite series:

$$(1 - \phi_1 B)^{-1} = 1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots$$

so $$y_t = (1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots)(1 - \Theta_1 B)a_t.$$

We can use this infinite series form of the ARIMA (1,0,1) model to examine the problem of parameter redundancy under certain conditions.

When $\phi_1 = \Theta_1$, both parameters are completely redundant. Substituting $\phi_1$ for $\Theta_1$ in the infinite series form of the ARIMA (1,0,1) process,

$$\begin{aligned} y_t &= (1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots)(1 - \phi_1 B)a_t \\ &= [(1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots) \\ &\qquad - \phi_1 B(1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots)]a_t \\ &= (1 - \phi_1 B + \phi_1 B - \phi_1^2 B^2 + \phi_1^2 B^2 - \ldots - \phi_1^n B^n + \phi_1^n B^n - \ldots)a_t \\ &= a_t. \end{aligned}$$

So when $\phi_1 = \Theta_1$, an ARIMA (1,0,1) model reduces to an ARIMA (0,0,0) model.

Now suppose that $\phi_1 \neq \Theta_1$ but that the parameters are *close* to each other in value. If $\Theta_1 = \phi_1 + c$, where c is a very small number, the infinite series form of the ARIMA (1,0,1) model is:

$$\begin{aligned} y_t &= (1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots)(1 - \phi_1 B - cB)a_t \\ &= [(1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots) \\ &\qquad - \phi_1 B(1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots) \\ &\qquad - cB(1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots)]a_t \\ &= (1 - cB - c\phi_1 B^2 - c\phi_1^2 B^3 - \ldots - c\phi_1^n B^{n+1} + \ldots)a_t. \end{aligned}$$

Now when both c and $\phi_1$ are small, $c\phi_1$ will be approximately zero and the ARIMA (1,0,1) model reduces approximately to an ARIMA (0,0,1) model:

$$y_t = (1 - \Theta_1 B)a_t,$$

where $c = \Theta_1$. And if $c\phi_1$ is not approximately zero, then the ARIMA

(1,0,1) model reduces approximately to an ARIMA (0,0,2) model:

$$yt = (1 - \Theta_1 B - \Theta_2 B^2)a_t$$

where $c\phi_1 = \Theta_2$. And of course, whenever $c \approx \phi_1$, the ARIMA (1,0,1) model reduces approximately to an ARIMA (1,0,0) model. In practice, the analyst must always remain skeptical of ARIMA (p,0,q) models. ARIMA (0,0,0), ARIMA (p,0,0), and ARIMA (0,0,q) models should always be ruled out before ARIMA (p,0,q) models are entertained.

## 2.8 The Autocorrelation Function

We have discussed the general ARIMA model so far without paying any attention to the task of model building. For a given time series, that is, how can the analyst select an appropriate ARIMA (p,d,q) model? *Identification*, the procedure whereby the values of p, d, and q are determined for a given time series, relies on a statistic called the autocorrelation function (ACF). For a time series process, $Y_t$, the ACF is defined as

$$ACF(k) = COV(Y_t\, Y_{t+k}) / VAR(Y_t).$$

Given a realization of the $Y_t$ process, a finite time series of N observations, the ACF is estimated from the formula

$$ACF(k) = \frac{\sum_{t=1}^{N-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{N} (Y_t - \bar{Y})^2} \left[\frac{N}{N-k}\right].$$

The ACF(k) is a measure of correlation between $Y_t$ and $Y_{t+k}$. For a given lag k, however, variance (the denominator of the formula) is estimated over all N observations while covariance (the numerator of the formula) is estimated over only N−k pairs of observations. Strictly speaking, then, ACF(k) is not the familiar Pearson product-moment correlation coefficient between time series observation k units apart.

To illustrate the formula for estimating ACF(k), lag the $Y_t$ series forward in time.

| | | | | | |
|---|---|---|---|---|---|
| lag-0 | $Y_1$ | $Y_2$ | $Y_3 \ldots Y_N$ | | |
| lag-1 | | $Y_1$ | $Y_2 \ldots Y_{N-1}$ | $Y_N$ | |
| lag-2 | | | $Y_1 \ldots Y_{N-2}$ | $Y_{N-1}$ | $Y_N$ |



and so forth. ACF(1) is the correlation coefficient estimated between the time series (lag-0) and its first lag (lag-1); ACF(2) is the correlation coefficient estimated between the time series (lag-0) and its second lag (lag-2); and, in general, ACF(k) is the correlation coefficient estimated between the time series (lag-0) and its $k^{th}$ lag (lag-k).

There are three points to be noted about the ACF. First, by definition, ACF(0) = 1; a time series is always perfectly correlated with itself. Second, also by definition, ACF(k) = ACF(−k); in other words, the ACF(k) is the same whether the series is lagged forward or backward. Because the ACF is symmetrical about lag-0, only the positive half of the ACF need be examined. Third, each time the series is lagged, one pair of observations is lost from the estimate of ACF(k). ACF(1) is estimated from N−1 pairs of observations; ACF(2) is estimated from N−2 pairs of observations; and so forth. As the value of k increases, confidence in the estimate of ACF(k) diminishes.

In theory, each time series process has a unique ACF. The $Y_t$ process is fully determined by its ACF, and if two processes have the same ACF, they are identical. In Section 2.11, we will describe a procedure whereby the ACF estimated from a realization of the process (from a finite time series, that is) is used to *identify* the ARIMA structure of the underlying process. In this section, we will derive the theoretical or expected ACFs for a variety of ARIMA (p,d,q) processes.

First, an ARIMA (0,0,0) or white noise process written as

$$Y_t = a_t + \Theta_0$$

is expected to have a uniformly zero ACF. This follows from the definition of white noise. For all k,

$$COV(a_t\, a_{t+k}) = 0.$$

Second, an ARIMA (0,1,0) process written as

$$(1 - B)Y_t = a_t + \Theta_0$$

is expected to have an ACF that is positive and dies out slowly from lag to lag, that is,

$$ACF(1) \simeq ACF(2) \simeq \ldots \simeq ACF(k).$$

A trending process, for example, has the expected value

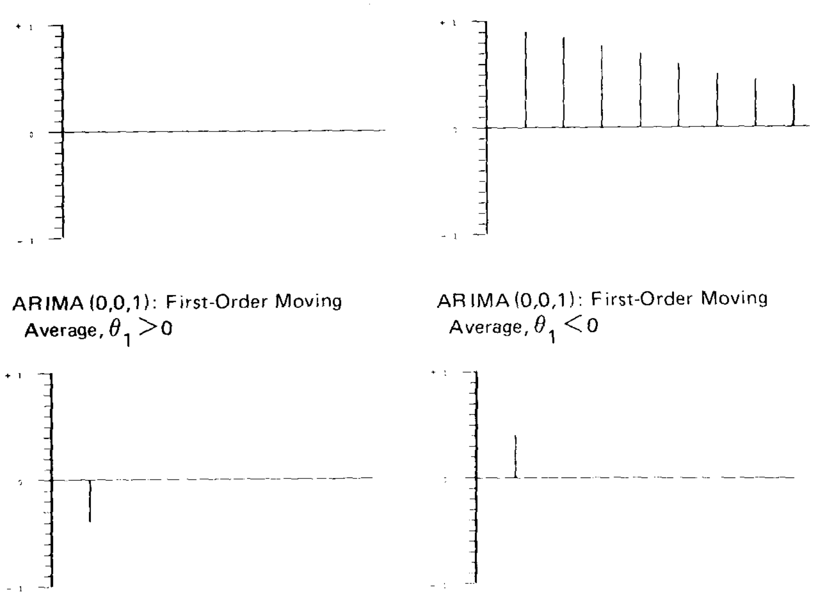
$$EY_t = \Theta_0 t.$$

A realization of this process will have a mean, $\overline{Y}$, approximately equal to some middle observation of the series. When $\overline{Y}$ is subtracted from each observation of the series, the resulting deviate series has the expected form

$$\ldots, -2\Theta_0, -1\Theta_0, 0, +1\Theta_0, +2\Theta_0, \ldots.$$

The first half of the deviate series will be negative numbers and the second half will be positive numbers. Thus, for a series of N observations, the estimate of ACF(1) will be based on N−2 pairs of observations with the same sign; the estimate of ACF(2) will be based on N−3 pairs of observations with the same sign; and so forth.

Figure 2.8(a) shows the expected ACFs for several ARIMA processes. The ACF of a nonstationary process is expected to have a relatively high positive value for ACF(1) and successive lags of the ACF are expected to die out slowly to zero. In particular, ACF(k) is expected to be approximately





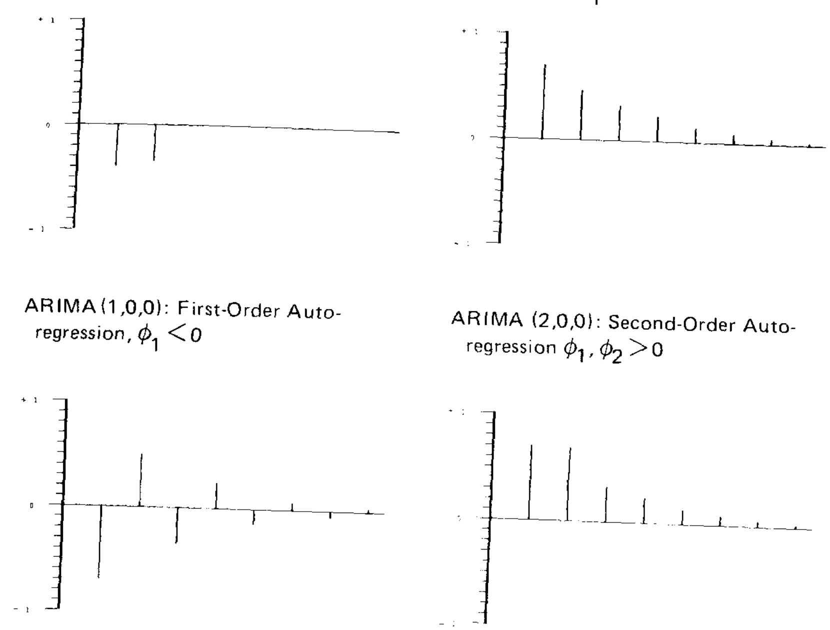




FIGURE 2.8(a) Expected ACFs for Several ARIMA Processes

equal to ACF(k+1) for all lags. In general, the ACF of any nonstationary process is expected to have the form indicated in Figure 2.8(a).

Third, an ARIMA (0,0,1) process written as

$$y_t = (1 - \Theta_1 B)a_t$$

is expected to have a nonzero ACF(1). All other lags of the ACF are expected to be zero. This is easily demonstrated by noting that COV $(y_t\, y_{t+1})$ is:

$$\begin{aligned} E(y_t y_{t+1}) &= E[(a_t - \Theta_1 a_{t-1})(a_{t+1} - \Theta_1 a_t)] \\ &= E(a_t a_{t+1} - \Theta_1 a_t^2 - \Theta_1 a_{t-1} a_{t+1} + \Theta_1^2 a_{t-1} a_t) \\ &= Ea_t a_{t+1} - \Theta_1 Ea_t^2 - \Theta_1 Ea_{t-1} a_{t+1} + \Theta_1^2 Ea_{t-1} a_t. \end{aligned}$$

As the property of white noise is that each random shock is independent of every other random shock, $Ea_ta_{t+k} = 0$ and all terms except one are zero:

$$E(y_ty_{t+1}) = -\Theta_1 Ea_t^2 = -\Theta_1\sigma_a^2.$$

To obtain the expected value of ACF (1), this result must be divided by the $y_t$ process variance. VAR ($y_t$) is:

$$\begin{aligned} E(y_t^2) &= E[(a_t - \Theta_1 a_{t-1})^2] \\ &= E(a_t^2 - 2\Theta_1 a_t a_{t-1} + \Theta_1^2 a_{t-1}^2) \\ &= Ea_t^2 - 2\Theta_1 Ea_ta_{t-1} + \Theta_1^2 Ea_{t-1}^2 \\ &= \sigma_a^2(1+\Theta_1^2). \end{aligned}$$

From these two results,

$$E[ACF(1)] = \frac{-\Theta_1\sigma_a^2}{\sigma_a^2(1+\Theta_1^2)} = \frac{-\Theta_1}{1+\Theta_1^2}.$$

Through the same procedure,

$$\begin{aligned} E(y_ty_{t+2}) &= E[(a_t - \Theta_1 a_{t-1})(a_{t+2} - \Theta_1 a_{t+1})] \\ &= E(a_ta_{t+2} - \Theta_1 a_ta_{t+1} - \Theta_1 a_{t-1}a_{t+2} + \Theta_1^2 a_{t-1}a_{t+1}) \\ &= Ea_ta_{t+2} - \Theta_1 Ea_ta_{t+1} - \Theta_1 Ea_{t-1}a_{t+2} + \Theta_1^2 Ea_{t-1}a_{t+1} = 0. \end{aligned}$$

From this,

$$E[ACF(2)] = \frac{0}{\sigma_a^2(1+\Theta_1^2)} = 0.$$

Through this same procedure, it can be demonstrated that ACF (3), ACF (4), . . . , ACF (k) are all expected to be zero.

It is important to note that an ARIMA (0,d,1) process is nonstationary and is thus expected to have an ACF typical of all nonstationary processes. If the process is differenced, however,

$$z_t = (1-B)^d Y_t = a_t - \Theta_1 a_{t-1}.$$

An ACF for the $z_t$ process will be that expected of an ARIMA (0,0,1) process.



Fourth, an ARIMA (0,0,2) process written as

$$y_t = (1 - \Theta_1 B - \Theta_2 B^2)a_t$$

is expected to have nonzero values for ACF (1) and ACF (2). The values of ACF (3) and all successive ACF (k) are expected to be zero. To demonstrate this, note that COV ($y_t\, y_{t+1}$) is:

$$\begin{aligned} E(y_ty_{t+1}) &= E[(a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2})(a_{t+1} - \Theta_1 a_t - \Theta_2 a_{t-1})] \\ &= E(a_ta_{t+1} - \Theta_1 a_t^2 - \Theta_2 a_ta_{t-1} - \Theta_1 a_{t-1}a_{t+1} + \Theta_1^2 a_{t-1}a_t \\ &\quad + \Theta_1\Theta_2 a_{t-1}^2 - \Theta_2 a_{t-2}a_{t+1} + \Theta_2\Theta_1 a_{t-2}a_t \\ &\quad + \Theta_2^2 a_{t-2}a_{t-1}) \\ &= -\Theta_1 Ea_t^2 + \Theta_1\Theta_2 Ea_{t-1}^2 \\ &= \sigma_a^2\Theta_1(\Theta_2 - 1). \end{aligned}$$

The process variance is:

$$\begin{aligned} E(y_t^2) &= E[(a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2})^2] \\ &= E(a_t^2 - 2\Theta_1 a_ta_{t-1} - \Theta_2 a_ta_{t-2} + \Theta_1^2 a_{t-1}^2 + 2\Theta_1\Theta_2 a_{t-1}a_{t-2} \\ &\quad - \Theta_2 a_ta_{t-2} + \Theta_2^2 a_{t-2}^2) \\ &= Ea_t^2 + \Theta_1^2 Ea_{t-1}^2 + \Theta_2^2 Ea_{t-2}^2 \\ &= \sigma_a^2(1 + \Theta_1^2 + \Theta_2^2). \end{aligned}$$

From these two results,

$$E[ACF(1)] = \frac{\sigma_a^2\Theta_1(\Theta_2 - 1)}{\sigma_a^2(1+\Theta_1^2+\Theta_2^2} = \frac{\Theta_1(\Theta_2-1)}{1+\Theta_1^2+\Theta_2^2}.$$

For the second lag of the ACF,

$$\begin{aligned} E(y_t\, y_{t+2}) &= E(a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2})(a_{t+2} - \Theta_1 a_{t+1} - \Theta_2 a_t) \\ &= -\Theta_2 Ea_t^2 \\ &= -\Theta_2\sigma_a^s. \end{aligned}$$

This gives the result

$$E[ACF(2)] = \frac{-\Theta_2\sigma_a^2}{\sigma_a^2(1+\Theta_1^2+\Theta_2^2} = \frac{-\Theta_2}{1+\Theta_1^2+\Theta_2^2}.$$

For the third lag of the ACF,

$$E(y_t y_{t+3}) = E[(a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2})(a_{t+3} - \Theta_1 a_{t+2} - \Theta_2 a_{t+1})]$$
$$= 0$$

so

$$E[ACF(3)] = 0.$$

The values of ACF (4), ACF (5), . . . , ACF (k) are all expected to be zero for the same reason.

Continuing this procedure, it can be demonstrated that an ARIMA (0,0,q) process is expected to have nonzero values for ACF (1), . . . , ACF (q) while ACF (q+1) and all successive lags are expected to be zero.

Fifth, an ARIMA (1,0,0) process written as

$$(1 - \phi_1 B) y_t = a_t$$

is expected to have an ACF that decays exponentially beginning with the first lag. To demonstrate this, note that COV ($y_t\, y_{t+1}$) is:

$$E(y_t y_{t+1}) = E[(y_t)(y_{t+1})] = E[(y_t)(\phi_1 y_t + a_{t+1})]$$
$$= E(\phi_1 y_t^2 + y_t a_{t+1}).$$

As $y_t$ and $a_{t+1}$ are independent, $Ey_t a_{t+1} = 0$, so

$$E(y_t y_{t+1}) = \phi_1 E y_t^2 = \phi_1 \sigma_y^2.$$

The variance of the process, of course, is $\sigma_y^2$. Dividing covariance by variance,

$$E[ACF(1)] = \frac{\phi_1 \sigma_y^2}{\sigma_y^2} = \phi_1.$$

For lag-2 of the ACF,

$$E(y_t y_{t+2}) = E[(y_t)(\phi_1 y_{t+1} + a_{t+2})]$$
$$= E[(y_t)(\phi_1(\phi_1 y_t + a_{t+1}) + a_{t+2})]$$
$$= E(\phi_1^2 y_t^2 + \phi_1 y_t a_{t+1} + y_t a_{t+2})$$
$$= \phi_1^2 E y_t^2 = \phi_1^2 \sigma_y^2.$$

Dividing covariance by variance again,

$$E[ACF(2)] = \phi_1^2 \sigma_y^2 / \sigma_y^2 = \phi_1^2.$$



By the same procedure,

$$E[ACF(3)] = \phi_1^3$$
$$\vdots$$
$$E[ACF(k)] = \phi_1^k.$$

This gives the ACF of an ARIMA (1,0,0) process a distinctive pattern of exponential decay from lag to lag. If $\phi_1 = .5$, for example, then,

$$ACF(1) = .5$$
$$ACF(2) = (.5)^2 = .25$$
$$ACF(3) = (.5)^3 = .125$$
$$\vdots$$
$$ACF(k) = (.5)^k \simeq 0.$$

Should $\phi_1$ be negative, however, successive lags of the ACF are alternately negative and positive. So if $\phi_1 = -.5$,

$$ACF(1) = -.5$$
$$ACF(2) = (-.5)^2 = .25$$
$$ACF(3) = (-.5)^3 = -.125$$
$$\vdots$$
$$ACF(k) = (-.5)^k \simeq 0.$$

For both positive and negative values of $\phi_1$, the expected ACF (k) grows smaller and smaller from lag to lag until, after three or four lags, ACF (k) is approximately zero.

Sixth, an ARIMA (2,0,0) process written as

$$(1 - \phi_1 B - \phi_2 B^2) y_t = a_t$$

is also expected to have an ACF that decays exponentially beginning with the first lag. For higher order ARIMA (p,0,0) processes, COV ($y_t\, y_{t+k}$) is difficult to derive. It can be demonstrated, nevertheless, that the expected ACF for an ARIMA (2,0,0) process is given by[7]:

$$ACF(k) = \phi_1\, ACF(k-1) + \phi_2\, ACF(k-2).$$

Recalling that ACF (0) = 1 and that ACF (−k) = ACF (k), the ACF of an ARIMA (2,0,0) process is expected to be:

$$\mathrm{ACF}\,(1) = \phi_1\,\mathrm{ACF}\,(0) + \phi_2\,\mathrm{ACF}\,(-1) = \phi_1 + \phi_2\,\mathrm{ACF}\,(1)$$
$$= \frac{\phi_1}{1-\phi_2}$$
$$\mathrm{ACF}\,(2) = \phi_1\mathrm{ACF}\,(1) + \phi_2\,\mathrm{ACF}\,(0)$$
$$= \frac{\phi_1^2}{1-\phi_2} + \phi_2$$
$$\mathrm{ACF}\,(3) = \phi_1\,\mathrm{ACF}\,(2) + \phi_2\,\mathrm{ACF}\,(1)$$
$$= \frac{\phi_1(\phi_2+\phi_1^2)}{1-\phi_2} + \phi_1\phi_2$$
$$\mathrm{ACF}\,(4) = \phi_1\,\mathrm{ACF}\,(3) + \phi_2\,\mathrm{ACF}\,(2)$$
$$= \frac{\phi_1^2(2\phi_2+\phi_1^2)}{1-\phi_2} + \phi_2\,(\phi_1^2+\phi_2)$$
$$\vdots$$
$$\mathrm{ACF}\,(k) = \phi_1\,\mathrm{ACF}\,(k-1) + \phi_2\,\mathrm{ACF}\,(k-2)\,.$$

To illustrate this expected ACF, let $\phi_1 = \phi_2 = .4$, then,

ACF (1) = .677
ACF (2) = .677
ACF (3) = .533
ACF (4) = .479
ACF (5) = .343
ACF (6) = .329

and so forth. Alternatively, if $\phi_1 = .4$ and $\phi_2 = -.4$,

ACF (1) = .286
ACF (2) = −.286
ACF (3) = −.229
ACF (4) = .023

and so forth. The pattern of decay in the expected ACF is always determined by the values of $\phi_1$ and $\phi_2$. In the general case, an ARIMA (p,0,0) process is expected to have an ACF that decays from lag to lag with the rate of decay determined by the values of $\phi_1, \phi_2, \ldots, \phi_p$.



Examining the expected ACFs in Figure 2.8(a), a method for *identifying* the ARIMA structure of a time series process can be seen. Given a realization of the process, a finite time series, the ACF can be estimated and used to infer the process structure. If the estimated ACF is zero for all lags, the analyst can infer that the time series was generated by an ARIMA (0,0,0) process. If the estimated ACF (1) is large and positive, say ACF (1) ≥ .7, and if the ACF dies out slowly from lag to lag, the analyst can infer that the process is nonstationary; the series must be differenced. If the estimated ACF (1) is nonzero but ACF (2) and all successive lags are zero, the analyst can infer that the time series was generated by an ARIMA (0,0,1) process. Finally, if the estimated ACF dies out exponentially from lag to lag, the analyst can infer that the time series was generated by an ARIMA (1,0,0) process.

But in practice, *identification* may not always be a simple task. The ACFs shown in Figure 2.8(a) are *expected* ACFs which presume either a knowledge of the process itself or else an infinitely long realization of the process. In practice, the process itself is always unknown and only a finite realization of the process (an N-observation time series) is available. To be sure, the estimated ACFs of white noise and nonstationary processes are so distinctive that the analyst cannot mistake them. Similarly, the estimated ACFs of higher order ARIMA (0,0,q) and ARIMA (p,0,0) processes are so different that the analyst will not ordinarily mistake one for the other. The estimated ACFs of ARIMA (0,0,q) and ARIMA (1,0,0) processes are quite similar, however, and in practice, it is nearly impossible to distinguish between these two processes on the basis of an estimated ACF alone.

Fortunately, another *identification* statistic, the partial autocorrelation function (PACF), can be used to distinguish a higher order ARIMA (0,0,q) process from an ARIMA (p,0,0) process. The PACF has an interpretation not unlike that of any other measure of partial correlation. The lag-k PACF, PACF (k), is a measure of correlation between time series observations k units apart *after the correlation at intermediate lags has been controlled or "partialed out."*

Unlike the ACF, the PACF cannot be estimated from a simple, straightforward formula: PACF (k) is estimated from a solution of the Yule-Walker equation system (See Box and Jenkins, 1976:64).[8] While we will not do so here, it can be demonstrated that the solution gives the values of

$$\mathrm{PACF}\,(1) = \mathrm{ACF}\,(1)$$
$$\mathrm{PACF}\,(2) = \frac{\mathrm{ACF}\,(2) - [\mathrm{ACF}\,(1)]^2}{1 - [\mathrm{ACF}\,(1)]^2}$$
$$\mathrm{ACF}\,(3) + \mathrm{ACF}\,(1)\,[\mathrm{ACF}\,(2)]^2 + [\mathrm{ACF}\,(1)]^3$$

$$\text{PACF}(3) = \frac{-2\text{ACF}(1)\,\text{ACF}(2) - [\text{ACF}(1)]^2\,\text{ACF}(3)}{1 + 2[\text{ACF}(1)]^2\,\text{ACF}(2) - [\text{ACF}(2)]^2 - [\text{ACF}(1)]^2}$$

and so forth. Expressing the expected PACF in this form, the role of the PACF as a measure of partial correlation is made explicit. The PACF in fact is a "partial" (or "partialed") ACF. Expressing the PACF in this form also makes explicit the tedious arithmetic involved in its estimation. Without the proper software, the estimated PACF would be of little use to the time series analyst.

As the expected PACF is a function of the expected ACF, and as the expected ACFs of several ARIMA processes have already been derived, the expected PACFs are as given.

First, an ARIMA (1,0,0) process whose ACF is expected to be

$$\text{ACF}(k) = \phi_1^k$$

is expected to have a nonzero PACF (1) while PACF (2) and all successive lags are expected to be zero:

$$\text{PACF}(1) = \phi_1$$

$$\text{PACF}(2) = \frac{\phi_1^2 - \phi_1^2}{1 - \phi_1^2} = 0$$

$$\text{PACF}(3) = \frac{\phi_1^3 + \phi_1\phi_1^4 + \phi_1^3 - 2\phi_1\phi_1^2 - \phi_1^2\phi_1^3}{1 + 2\phi_1^2\phi_1^2 - \phi_1^4 - 2\phi_1^2}$$

$$= \frac{2\phi_1^3 + \phi_1^5 - 2\phi_1^3 - \phi_1^5}{1 + \phi_1^4 - 2\phi_1^2} = 0.$$

Successive lags of PACF (k) are also expected to be zero.

Second, an ARIMA (2,0,0) process whose ACF is expected to be

$$\text{ACF}(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\text{ACF}(2) = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

$$\text{ACF}(3) = \frac{\phi_1(\phi_2 + \phi_1^2)}{1 - \phi_2} + \phi_1\phi_2$$



is expected to have nonzero values of PACF (1) and PACF (2) while PACF (3) and all successive lags are expected to be zero. Substituting the expected ACF (k) in the formula (too tedious a procedure to be presented here),

$$\text{PACF}(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\text{PACF}(2) = \frac{\phi_2(\phi_2 - 1)^2 - \phi_1\phi_2}{(1 - \phi_2)^2 - \phi_1^2}$$

$$\text{PACF}(3) = 0.$$

Successive lags are all expected to be zero. In the general case, an ARIMA (p,0,0) process is expected to have nonzero values for PACF (1), . . . , PACF (p) while PACF (p+1) and all successive lags are expected to be zero.

Third, an ARIMA (0,0,1) process whose ACF is expected to be

$$\text{ACF}(1) = \frac{-\Theta_1}{1 + \Theta_1^2}$$

$$\text{ACF}(2) = \ldots = \text{ACF}(k) = 0$$

has a decaying PACF, that is, all PACF (k) are expected to be nonzero:

$$\text{PACF}(1) = \frac{-\Theta_1}{1 + \Theta_1^2}$$

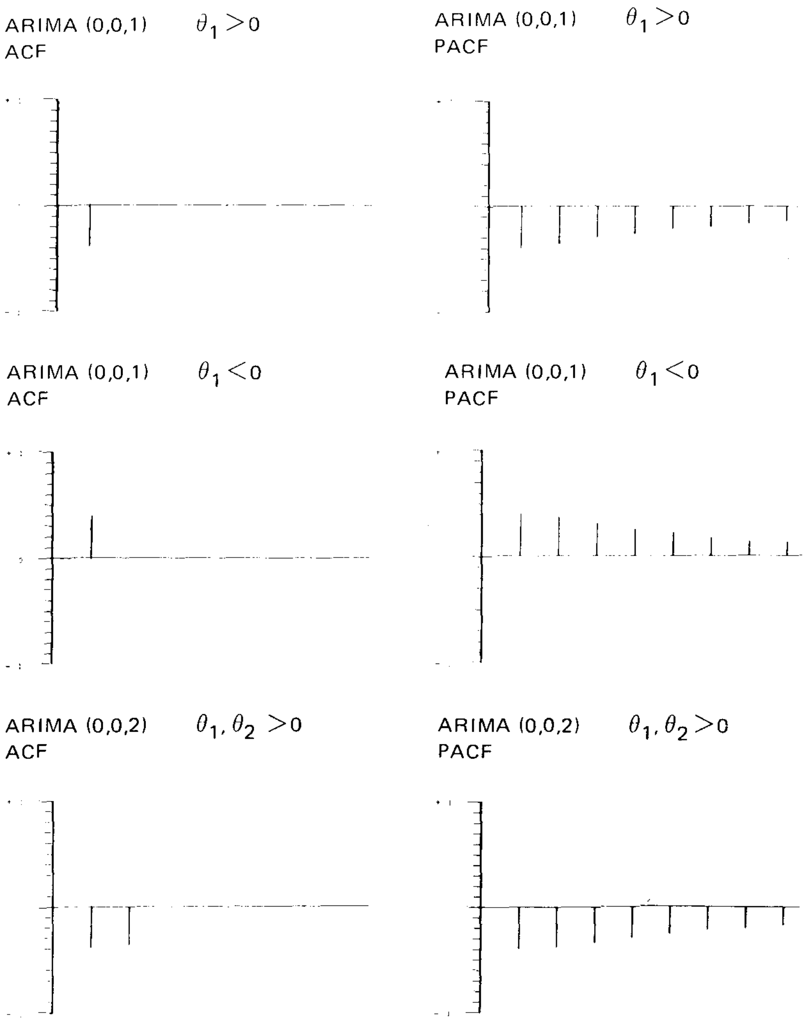
$$\text{PACF}(2) = \frac{-\Theta_1^2}{1 + \Theta_1^2 + \Theta_1^4}$$

$$\text{PACF}(3) = \frac{-\Theta_1^3}{1 + \Theta_1^2 + \Theta_1^4 + \Theta_1^6}.$$

Successive lags of the expected PACF grow smaller and smaller in absolute value. If $\Theta_1 = .7$, for example,

$$\text{PACF}(1) = -.469$$
$$\text{PACF}(2) = -.283$$
$$\text{PACF}(3) = -.186$$

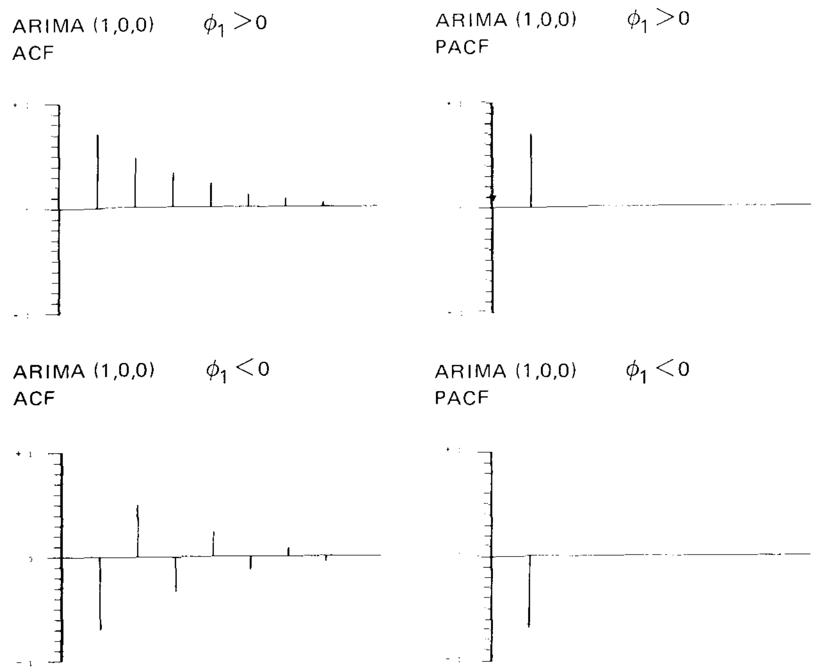
and so forth. In the general case, the PACF of an ARIMA (0,0,q) process is expected to decay in this same manner but at a rate determined by the values of $\Theta_1, \ldots, \Theta_q$.




*(Figure 2.8(b) continued on p. 79)*

Figure 2.8(b) shows the expected ACFs and PACFs for several ARIMA (p,0,0) and ARIMA (0,0,q) processes. Autoregressive processes are characterized by decaying ACFs and spiking PACFs. An ARIMA (p,0,0) process is expected to have exactly p nonzero spikes in the first p lags of its PACF. All successive lags of the PACF are expected to be zero. Moving average






*(Figure 2.8(b) continued on p. 80)*

processes are characterized by spiking ACFs and decaying PACFs. An ARIMA (0,0,q) process is expected to have exactly q nonzero spikes in the first q lags of its ACF. All successive lags of the ACF are expected to be zero. Finally, an ARIMA (p,0,q) process is expected to have both decaying ACF and PACF.

There is one more theoretical issue to be covered before the practical issues of model identification can be considered. In Section 2.11, we will describe the procedures required to identify an appropriate ARIMA (p,d,q) model from estimated ACFs and PACFs. In Section 2.12, we will present four example analyses which illustrate in detail the model-building procedure.

## 2.9 Seasonality

If it were not for seasonality, time series analysis might become a rather simple, pleasant task. Most social science time series would be well repre-

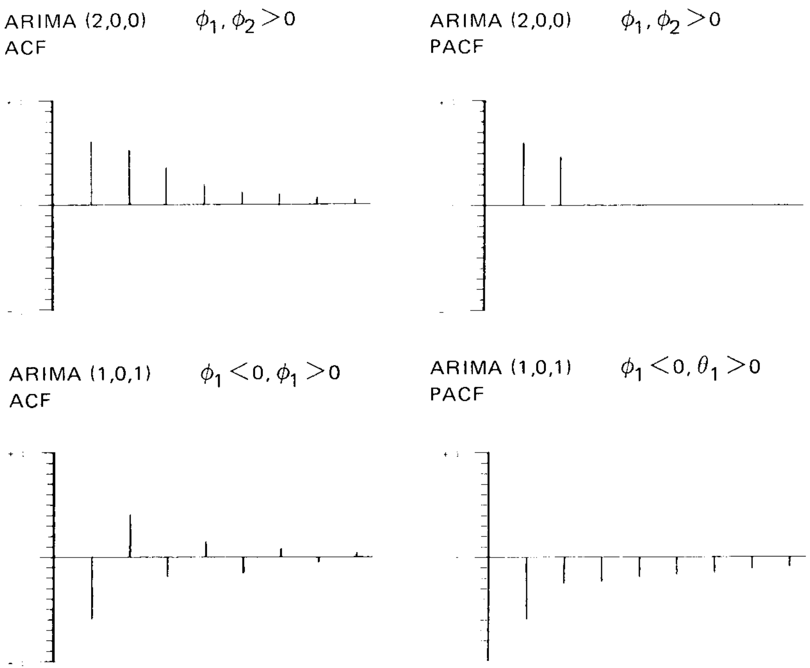



FIGURE 2.8(b) Expected ACFs and PACFs for Several ARIMA Processes

sented by lower order models such as ARIMA (1,0,0), ARIMA (0,0,1), and ARIMA (0,1,1) which are easily identified in most cases. But it is an unfortunte fact that almost all monthly or quarterly social science time series have strong seasonal components. This seasonality complicates the task of time series analysis generally and ARIMA modeling specifically.

We define seasonality as *any cyclical or periodic fluctuation in a time series that recurs or repeats itself at the same phase of the cycle or period.* Retail sales indicators, for example, normally peak in December when families shop for Christmas presents. If we knew nothing else about a retail sales indicator, then, we could guess that the series would reach a highpoint in each December.



To the time series analyst, seasonality is process variance which must be removed or controlled. One method of controlling for seasonal variance is to *deseasonalize* the time series prior to analysis. Johnston (1972: 186) describes a multiple regression deseasonalization method wherein dummy variables are used to estimate the seasonal variance of each month or quarter of the cycle. This estimated variance is then subtracted from the series. Makridakis and Wheelwright (1978: Chapter 16) describe similar (though more complicated) deseasonalization methods used by the U.S. Labor and Commerce Departments. All of these deseasonalization methods require that seasonal variation be "adjusted" or subtracted from the series prior to analysis. In *Design and Analysis of Time Series Experiments,* Glass et al. (1975) recommended deseasonalization, noting that it was the only practical means of handling seasonality. This recommendation was based on the limitations of time series software, however. Since that time, the state of the art in time series software has advanced to the point at which deseasonalization cannot generally be recommended. *Our comments in Section 2.1 about detrending a time series apply as well to deseasonalizing a time series: No adequate deseasonalization methods are available.*

The absolute "best" method of handling seasonality is to build a causal model of seasonal forces. On this point, we cite Nerlove:

> In one sense, the whole problem of seasonal adjustment of economic time series is a spurious one. Seasonal variations have causes (for example, variations in the weather), and insofar as these causes are measurable they should be used to explain changes that are normally regarded as seasonal . . . Ideally, one should formulate a complete econometric model in which the causes of seasonality are incorporated directly in the equations. . . . On the practical side the problems include the lack of availability of many relevant series, the non-measurability of key items, and the lack of appropriate statistical methodology . . . In addition, the precise structure of the model will very much affect the analysis of seasonal effects . . . On the conceptual side, the problem is basically one of continuing structural change, which is essentially the sort of thing which causes seasonality to show up [1964: 263].

In Chapter 5, where we develop bivariate ARIMA models, it will be clear that two time series

$$Y_t = N_t + S_t$$
$$X_t = N'_t + S'_t$$

may share the same set of seasonal causors. That is, for the seasonal components $S_t$ and $S'_t$, there is some relationship

$$W_t \begin{array}{l} \nearrow S_t \\ \searrow S_t'. \end{array}$$

So in the bivariate model,

$$Y_t = f(X_{t-n}) + f(N_t, N_t')$$

the two seasonal components, $S_t$ and $S_t'$, will cancel each other out.

In the univariate situation, however, which is our only concern in this chapter, seasonality must be accounted for in the univariate ARIMA model. The seasonal ARIMA model, in contrast to deseasonalization methods, controls seasonal variance by incorporating seasonal correlations into the model. To illustrate this general principle, we may write out an N-month time series as

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ | $Y_6$ | $Y_7$ | $Y_8$ | $Y_9$ | $Y_{10}$ | $Y_{11}$ | $Y_{12}$ |
| $Y_{13}$ | $Y_{14}$ | $Y_{15}$ | $Y_{16}$ | $Y_{17}$ | $Y_{18}$ | $Y_{19}$ | $Y_{20}$ | $Y_{21}$ | $Y_{22}$ | $Y_{23}$ | $Y_{24}$ |
| $\vdots$ | | | | | | | | | | | $\vdots$ |
| $Y_{N-23}\ldots$ | | | | | | | | | | | $\ldots Y_{N-12}$ |
| $Y_{N-11}\ldots$ | | | | | | | | | | | $\ldots Y_N$ |

Now an ARIMA (1,0,0) model uses the prior observation to predict the current observation, that is,

$$y_t = \phi_1 y_{t-1} + a_t.$$

The ARIMA (1,0,0) model defines a relationship between $Y_1$ and $Y_2$, between $Y_2$ and $Y_3$, and so forth. Similarly, an ARIMA (0,0,1) model

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

defines a relationship between $Y_1$ and $a_2$, $Y_2$ and $a_3$, and so forth. If a time series is seasonal, however, it makes good sense to suspect that there will be similar relationships between $Y_1$, or January of the first year, and $Y_{13}$, January of the second year; between $Y_2$, February of the first year, and $Y_{14}$, February of the second year, and so forth.

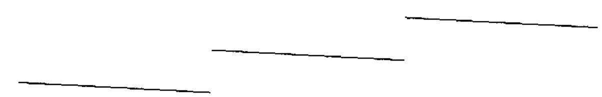


## Seasonal Nonstationarity

A process may drift or trend in annual steps or increments. Agricultural production series tend to exhibit nonstationarity of this sort due, we assume, to the prominence of crop seasons in the process. To account for seasonal drift or trend, difference the series seasonally. That is, for monthly data, subtract $Y_1$ from $Y_{13}$, $Y_2$ from $Y_{14}$, and so forth. We represent this procedure by the difference operator

$$(1 - B^{12})\,Y_t = \Theta_0$$
$$\text{or} \quad Y_t = Y_{t-12} + \Theta_0.$$

Processes that are seasonally nonstationary drift or trend in annual steps such as



rather than as observation-to-observation steps.

## Seasonal Autoregression

The current observation of a process may depend to some extent upon the corresponding observation from the preceding cycle or period. We express this relationship for monthly data as

$$y_t = \phi_{12} y_{t-12} + a_t$$
$$\text{or} \quad (1 - \phi_{12}B^{12})\,y_t = a_t.$$

And, of course, the current observation may depend to some extent upon the corresponding observations from the two preceding cycles or periods:

$$(1 - \phi_{12}B^{12} - \phi_{24}B^{24})y_t = a_t.$$

Of course, our comments on higher order autoregression apply here as well. For seasonal ARIMA structures, too, first-order structures are more common than second-order structures. Higher order structures are almost never encountered.

## Seasonal Moving Averages

Finally, the current observation of a time series process may depend to some extent upon the random shocks from a year or two years earlier:

$$y_t = (1 - \Theta_{12}B^{12})a_t$$

or

$$y_t = (1 - \Theta_{12}B^{12} - \Theta_{24}B^{24})a_t.$$

And, of course, the seasonal ARIMA structure may be any combination of integrated, autoregressive, and moving average components.

The general ARIMA seasonal model is denoted by ARIMA (p,d,q) $(P,D,Q)_S$ where P, D, and Q are analogous to p, d, and q. The structural parameter S indicates the length of the naturally occurring period or cycle. Thus, for monthly data, S = 12. For quarterly data S = 4, and for weekly data S = 52.

Most time series with seasonal ARIMA behavior will exhibit regular ARIMA behavior as well. It might seem commonsensical to incorporate regular and seasonal structures additively into the ARIMA model. For example, a time series with both regular and seasonal autoregressive structures incorporated additively would be written as

$$(1 - \phi_1 B - \phi_{12}B^{12})y_t = a_t$$

or

$$y_t = \phi_1 y_{t-1} + \phi_{12}y_{t-12} + a_t.$$

However, regular and seasonal ARIMA structures are ordinarily incorporated *multiplicatively*. A time series with both regular and seasonal autoregressive structures would be written as

$$(1 - \phi_{12}B^{12})(1 - \phi_1 B)y_t = a_t.$$

The two autoregressive terms in this expression are called *factors*. The difference between additive and multiplicative seasonal models is made explicit by expanding the two-factor model:

$$(1 - \phi_{12}B^{12})(1 - \phi_1 B)y_t = a_t$$
$$(1 - \phi_1 B - \phi_{12}B^{12} + \phi_1\phi_{12}B^{13})y_t = a_t$$
$$y_t = \phi_1 y_{t-1} + \phi_{12}y_{t-12} - \phi_1\phi_{12}y_{t-13} + a_t.$$

The multiplicative model has a cross-product term, $\phi_1\phi_{12}B^{13}$, which an additive model lacks. Clearly, when both $\phi_1$ and $\phi_{12}$ are small, their product, $\phi_1\phi_{12}$, is approximately zero and, as a result, there will be little difference between the additive and multiplicative models. When $\phi_1$ and $\phi_{12}$ are not small, however, the two-factor multiplicative model reflects a much different process than the additive model. The two-factor model uses one extra piece of information ($y_{t-13}$) to predict the current observation, so it will ordinarily give a better fit to seasonal data than the additive model.



The model we have just demonstrated is an ARIMA (1,0,0) $(1,0,0)_{12}$ model. Other common seasonal models are (1) the ARIMA (0,0,1) $(0,0,1)_{12}$

$$\begin{aligned} y_t &= (1 - \Theta_1 B)(1 - \Theta_{12}B^{12})a_t \\ &= (1 - \Theta_1 B - \Theta_{12}B^{12} + \Theta_1\Theta_{12}B^{13})a_t \\ &= a_t - \Theta_1 a_{t-1} - \Theta_{12}a_{t-12} + \Theta_1\Theta_{12}a_{t-13} \end{aligned}$$

(2) the ARIMA (0,1,1) $(0,0,1)_{12}$

$$\begin{aligned} (1 - B)Y_t &= \Theta_0 + (1 - \Theta_1 B)(1 - \Theta_{12}B^{12})a_t \\ &= \Theta_0 + (1 - \Theta_1 B - \Theta_{12}B^{12} + \Theta_1\Theta_{12}B^{13})a_t \\ &= \Theta_0 + a_t - \Theta_1 a_{t-1} - \Theta_{12}a_{t-12} + \Theta_1\Theta_{12}a_{t-13} \end{aligned}$$

and (3) the ARIMA (0,1,1) $(0,1,1)_{12}$

$$\begin{aligned} (1 - B)(1 - B^{12})Y_t &= \Theta_0 + (1 - \Theta_1 B)(1 - \Theta_{12}B^{12})a_t \\ &= \Theta_0 + (1 - \Theta_1 B - \Theta_{12}B^{12} + \Theta_1\Theta_{12}B^{13})a_t \\ &= \Theta_0 + a_t - \Theta_1 a_{t-1} - \Theta_{12}a_{t-12} + \Theta_1\Theta_{12}a_{t-13}. \end{aligned}$$

In the ARIMA (0,1,1) $(0,1,1)_{12}$ model, we describe a time series process that drifts or trends both regularly *and* seasonally. The order of differencing in practice (that is, regular differencing and then seasonal differencing or vice versa) does not matter because

$$\begin{aligned} (1 - B)(1 - B^{12}) &= (1 - B^{12})(1 - B) \\ &= (1 - B - B^{12} + B^{13}). \end{aligned}$$

So we difference the series regularly,

$$z_t = Y_t - Y_{t-1},$$

and then difference it seasonally,

$$z_t^* = z_t - z_{t-12},$$

or vice versa without changing the result. And, of course, we could accomplish both differences simultaneously with the operation

$$z_t^* = (1 - B)(1 - B^{12})\,Y_t$$
$$z_t^* = (1 - B - B^{12} + B^{13})\,Y_t$$
$$z_t^* = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}.$$

Our comments about regular autoregression and moving averages are general to the seasonal cases as well. In particular, the ARIMA (p,d,q) $(P,D,Q)_S$ model must have parameters constrained to the bounds of stationarity which, for autoregressive models, are[9]

$$-1 < \phi_1, \phi_s < +1$$

and for the ARIMA (0,d,1) $(0,D,1)_S$ model, the bounds of invertibility are

$$-1 < \Theta_1, \Theta_S < +1.$$

Also, ARIMA (p,0,0) $(P,0,0)_S$ models may be written as an infinite series of exponentially weighted past shocks. Similarly, ARIMA (0,d,q) $(0,D,Q)_S$ models may be written as an infinite series of exponentially weighted past observations. The reader may demonstrate these truths by substitution. We will not do so here because it involves too much arithmetic. Nevertheless, it can be demonstrated that the seasonal autoregressive factor has as its inverse the infinite series

$$(1 - \phi_S B^S)^{-1} = 1 + \phi_S B^S + \phi_S^2 B^{2S} + \ldots + \phi_S^n B^{nS} + \ldots.$$

So to "solve" an ARIMA (1,0,0) $(1,0,0)_{12}$,

$$(1 - \phi_1 B)(1 - \phi_{12} B^{12}) y_t = a_t$$
$$y_t = (1 - \phi_1 B)^{-1}(1 - \phi_{12} B^{12})^{-1} a_t$$
$$= (1 + \phi_1 B + \phi_1^2 B^2 + \ldots + \phi_1^n B^n + \ldots)$$
$$(1 + \phi_{12} B^{12} + \phi_{12}^2 B^{24} + \ldots + \phi_{12}^n B^{12n} + \ldots) a_t.$$

As both infinite series converge, their product converges.

There should be some transfer of understanding here. The principles demonstrated for ARIMA (p,d,q) models generalize one-to-one to ARIMA $(P,D,Q)_S$ models. ARIMA(p,d,q)$(P,D,Q)_S$ models imply a stochastic behavior determined by the polynomial multiplication of ARIMA (p,d,q) and ARIMA $(P,D,Q)_S$ models. In the general case, the reader who understands the behavior of ARIMA (p,d,q) models and the rules of polynomial multiplication can deduce the behavior of ARIMA (p,d,q) $(P,D,Q)_S$ models. We will turn our attention to a more practical matter: identifying an appropriate seasonal model for a time series.



## 2.10 Identifying a Seasonal Model

As one might suspect, an ARIMA $(P,D,Q)_S$ model can be identified for a time series through an inspection of the *seasonal* lags of the ACF and PACF. For monthly data, the seasonal lags are -12, -24, and -36. For quarterly data, the seasonal lags are -4, -8, and -12. And in general, for the cycle S, the seasonal lags of the ACF and PACF are lags-S, -2S, and -3S. In Figure 2.10(a), we show the expected ACFs and PACFs for several ARIMA $(P,D,Q)_{12}$ processes. The patterns of spiking and decay for these ARIMA $(P,D,Q)_{12}$ processes are identical with the patterns expected of the analogous ARIMA (p,d,q) processes—except that the spiking and decay occur at seasonal lags.

First, seasonal nonstationarity is indicated by large and nearly equal values of the ACF at seasonal lags, that is,

$$\text{ACF}(S) \approx \text{ACF}(2S) \approx \ldots \approx \text{ACF}(kS).$$

Seasonal differencing will make an ARIMA $(0,D,0)_S$ process stationary.

Second, ARIMA $(0,0,Q)_S$ processes are expected to have Q spikes at the first Q seasonal lags of the ACF. All successive lags are expected to be zero. The PACF of an ARIMA $(0,0,Q)_S$ process is expected to decay from seasonal lag to seasonal lag. The rate of decay is determined by the values of $\Theta_S, \Theta_{2S}, \ldots, \Theta_{QS}$.

Third, ARIMA (P,0,0) processes are expected to have a decaying ACF, the rate of decay determined by the values of $\phi_S, \phi_{2S}, \ldots, \phi_{PS}$. The PACF of an ARIMA (P,0,0) process is expected to have P spikes at the first P seasonal lags. All successive lags of the PACF are expected to be zero.

Of course, ARIMA $(P,D,Q)_S$ processes are rarely encountered in the social sciences. Most social science processes, if seasonal at all, are best represented by ARIMA (p,d,q) $(P,D,Q)_S$ models. The identification of an ARIMA (p,d,q) $(P,D,Q)_S$ model is complicated by interaction terms. Whereas an ARIMA (p,d,q) model can be identified on the basis of the first few lags of the ACF and PACF, and whereas an ARIMA $(P,D,Q)_S$ model can be identified on the basis of the first few seasonal lags of the ACF and PACF, an ARIMA (p,d,q) $(P,D,Q)_S$ model must be identified on the basis of the entire ACF and PACF.

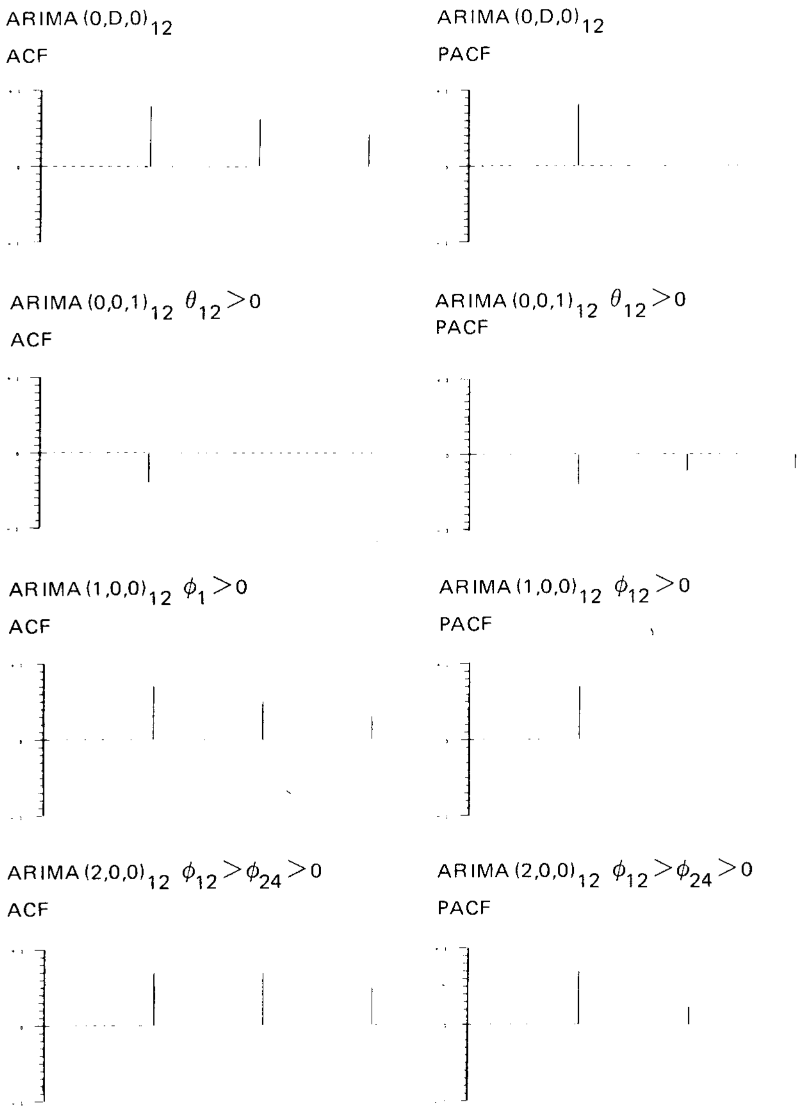



FIGURE 2.10(a) Expected ACFs for Several ARIMA $(P,D,Q)_{12}$ Processes



If it were not for the rather simple nature of social science time series processes, the identification of an ARIMA (p,d,q) $(P,D,Q)_S$ model would be an infinitely complicated task. There are a number of features which simplify the task, however.

For example, the analyst usually knows the value of S. For monthly data, S = 12. For quarterly data, S = 4, and so on. In other substantive areas, time series analysts may not know the value of S and this complicates model identification. Because we know the length of the seasonal cycle, we need only examine a few specific lags of the ACF and PACF to identify a model.

Similarly, social science processes typically have small integer values (almost always 0 or 1, sometimes 2) of p, q, P, and Q. If the ARIMA (p,d,q) $(P,D,Q)_S$ structural parameters took larger integer values, as seems to be the case in other substantive areas, the analyst would be forced to assess the statistical significance of many dozen lags of the ACF and PACF. This in turn would require much longer time series (say 300 observations or more) than are ordinarily available to the social scientist.

Finally, as a general rule, the regular and seasonal factors of an ARIMA (p,d,q) $(P,D,Q)_S$ model will be of the same type, that is, *either* autoregressive *or* moving average. If the regular factor is autoregressive, the analyst can usually rule out seasonal factors that are not autoregressive. As our discussion of parameter redundancy in section 2.7 suggests, ARIMA (p,0,0) $(0,0,Q)_S$ and ARIMA (0,0,q) $(P,0,0)_S$ models will often reduce to simpler ARIMA (p,0,0) $(P,0,0)_S$ and ARIMA (0,0,q) $(0,0,Q)_S$ models.

In Figure 2.10(b), we show the expected ACFs and PACFs of the most commonly encountered ARIMA (p,d,q) $(P,D,Q)_{12}$ processes. As shown, ARIMA (0,d,0) $(0,D,0)_{12}$ processes are characterized by persistently high values of the ACF at the regular *and* seasonal lags. The ACF shown indicates that the time series must be differenced both regularly and seasonally. Many social science time series processes are nonstationary only in the regular factor *or* the seasonal factor and, thus, should be differenced only regularly *or* seasonally. The ACFs of regularly, seasonally, and joint regularly/seasonally nonstationary processes are so distinctive that, in practice, the analyst will seldom mistake the type or number of differences required to make a process stationary.

An ARIMA (0,0,1) $(0,0,1)_{12}$ process is expected to have spikes at lags-11, -12, and -13 of the ACF. Higher lags of the ACF are expected to be zero. The PACF is expected to decay at both regular and seasonal lags but the key to identification is clearly the ACF. In the general case, an ARIMA (0,0,q) $(0,0,Q)_S$ process has a distinctive ACF with

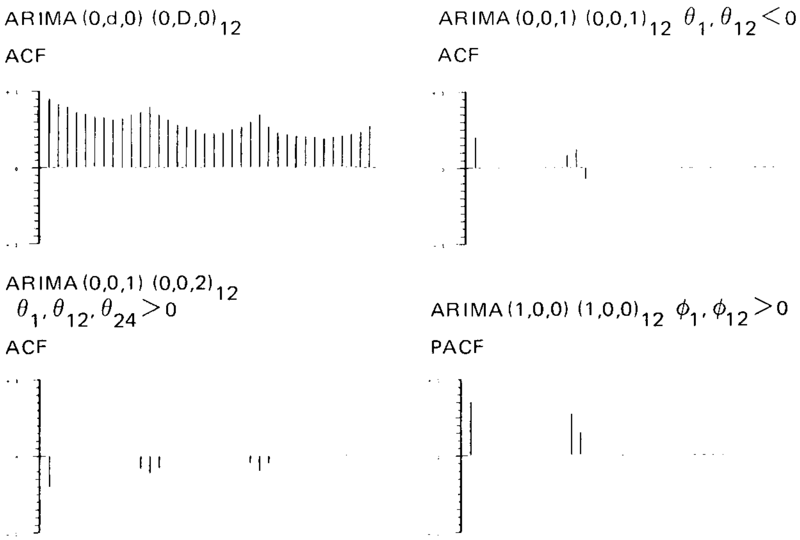



FIGURE 2.10(b) Expected ACF or PACF for Several ARIMA (p,d,q) $(P,D,Q)_{12}$ Processes

ACF (1), . . . , ACF (q) expected to be *non*zero
ACF (q+1) , . . . , ACF (S−q) expected to be zero
ACF (S−q+1) , . . . , ACF (S+q) expected to be *non*zero
ACF (S+q+1) , . . . , ACF (2S−q) expected to be zero
ACF (2S−q+1) , . . . , ACF (2S+q) expected to be *non*zero

and so forth. For Q>2, the process repeats itself. An ARIMA (0,0,1) $(0,0,2)_{12}$ process then is expected to have nonzero spikes at lags-11, -12, and -13 of the ACF and at lags-23, -24, and -25 of the ACF.

An ARIMA (1,0,0) $(1,0,0)_{12}$ process is expected to have an ACF which decays exponentially from seasonal lag to seasonal lag. The key to identification, however, is the PACF which will have a spike at lag-1 and at lags-12 and -13. In the general case, the PACF of an ARIMA (p,0,0) $(P,0,0)_{12}$ process is expected to have spikes at the first p lags, at lags S, . . . , SP, and at lags PS+p. For an ARIMA (2,0,0) $(2,0,0)_{12}$ process, then, we expect to see spikes at lags-1 and -2, at lags-12 and -13, and at lags-24 and -25 of the PACF. All other lags of the PACF are expected to be zero.



In theory, the identification of an ARIMA (p,d,q) $(P,D,Q)_S$ model for a time series reduces to a set of logical steps. First, use the estimated ACF to determine whether the series is stationary. If not, difference it appropriately. Second, use the ACF and PACF to determine the integer values of p and/or q. Third, use the ACF to determine the value of Q or the PACF to determine the value of P. Having thus identified an ARIMA (p,d,q) $(P,D,Q)_S$ model for the time series, we are ready to build the model.

## 2.11 Model Building

ARIMA modeling has been called an "art" by many authors. This seems to imply that one must be an artist (either by virtue of innate talent or lengthy training) to successfully create a model. We disagree. We prefer to think of ARIMA modeling as a *craft,* similar perhaps to carpentry. As a craft activity, basic ARIMA techniques are accessible to everyone after only a relatively short apprenticeship (usually an intensive workshop or a half-semester of coursework). And after acquiring a firm grasp of the essentials, one can develop journeyman skills by working with different types of data and more challenging substantive applications.

An ARIMA model is custom-built to fit a particular time series. Like a carpenter, the time series analyst uses tools (the statistical models we have developed), materials (the data), and plans (a model-building strategy) to create a model. In this section, we develop a detailed model-building strategy which, when followed by the analyst and supplemented by skills, will usually produce a sturdy, craftlike model. Our model-building strategy is essentially the one recommended by Box and Jenkins (1976) with the addition of several procedures we have found valuable in the analysis of social science data.

Our model-building strategy is generally conservative. We prefer to see a simple, robust model rather than a flimsy, flashy one. If a model does not fit the data well, we expect the craftsman to acknowledge that fact (rather than trying to force a fit by bending the model parts out of shape). Having built the model, the craftsman will critically evaluate its quality and, if found wanting, will make appropriate adjustments or will report its shortcomings.

ARIMA models are built through an iterative identification/estimation/diagnosis strategy which we have outlined as a flow chart in Figure 2.11(a). Before starting, the analyst should inspect a plot of the raw time series. Particular attention should be paid to sources of nonstationarity which may be visible in the series plot. While nonstationarity due only to systematic trend or drift is easily detected in an inspection of the series ACF, nonstationarity associated with other causes (variance nonstationarity, for exam-

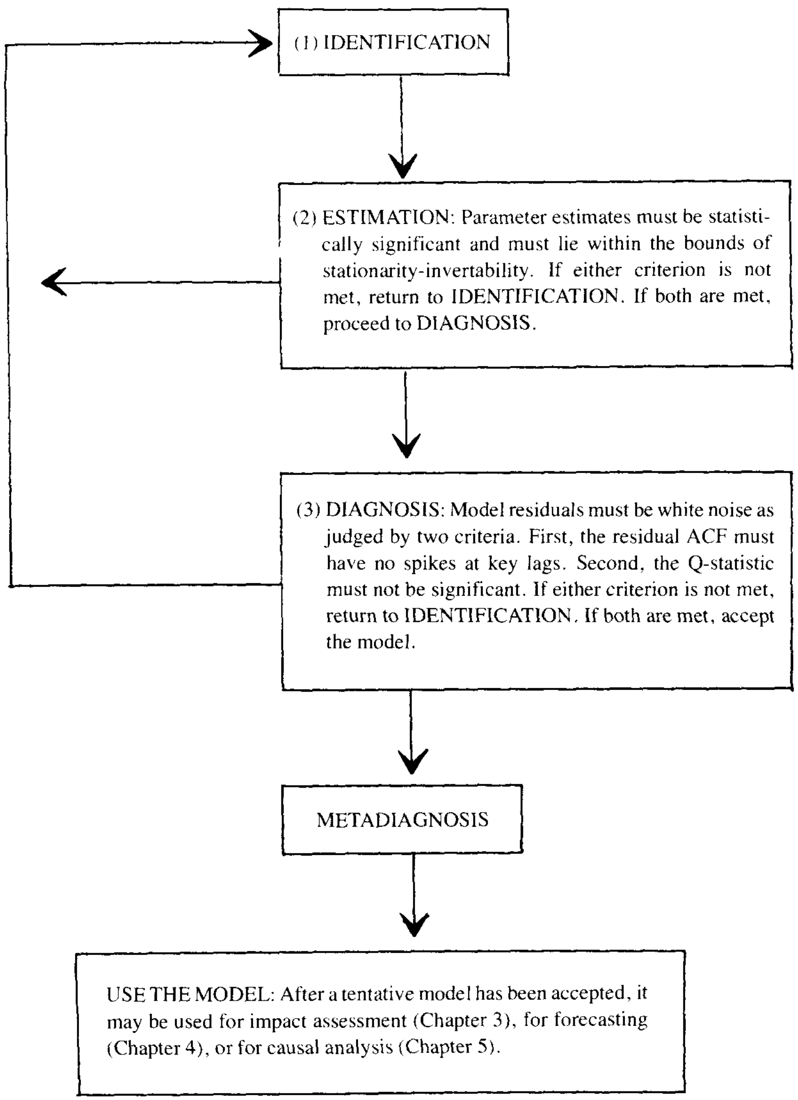



FIGURE 2.11(a) The ARIMA Model-Building Strategy

ple) can usually be detected only through an inspection of the time series plot.



The iterative identification/estimation/diagnosis strategy begins with the *identification* of a tentative ARIMA model for the series. Patterns of autocorrelation observed in the data are compared with the patterns expected of various ARIMA models. If nonstationarity is indicated—by an ACF which fails to "die out"; see Figure 2.8(a)—it will be necessary to difference and/or transform the series prior to identifying a tentative model.

Next, the parameters of the tentative model are *estimated*. All parameter estimates must lie within the bounds of stationarity-invertibility and must be statistically significant. If the parameter estimates do not satisfy these criteria, a new model must be identified and its parameters estimated.

After a tentative model has been identified and its parameters satisfactorily estimated, it must be *diagnosed*. To pass diagnosis, the residuals of the tentative model must be white noise. If this criterion is not satisfied, the tentative model is inadequate and must be rejected; the model-building procedure begins anew. Another model is identified, its parameters estimated, and its residuals diagnosed. The iterative identification/estimation/diagnosis procedure continues until an adequate model has been created for the time series.

The strategy we recommend is a model-*building* strategy. It leads to a model that is statistically adequate and yet parsimonious for a given time series. Alternative strategies might be better described as model-*fitting* strategies. The analyst might begin, for example, by fitting a general ARIMA (p,d,q) $(P,D,Q)_S$ model to the time series and then deleting unnecessary terms from the model. This alternative strategy will generally lead to a model that is statistically adequate (that "fits" the data, that is) but that is not necessarily parsimonious. A model that is adequate but not parsimonious is arbitrary and may lead to a set of confused and confusing inferences.

Our treatment of ARIMA modeling so far has been in the abstract. We have been concerned primarily with "expected" statistics and "general" cases. We will now consider the more practical aspects of modeling, starting with a more detailed description of the general model-building strategy.

## Identification

*Identification* is the key to model building. An ARIMA model must have some empirical basis. That is, put simply, there should be some reason for selecting one tentative model over another. The empirical basis will ordinarily be the patterns of autocorrelation found in the ACFs and PACFs estimated from the time series. If two competing models are both adequate, the model that best fits the ACF and PACF is the "better" of the two models.

In practice, estimated ACFs and PACFs will not be identical to the expected ACFs and PACFs shown in Figures 2.8(a), 2.8(b), 2.10(a), and 2.10(b). An ARIMA (1,0,0) process, for example, is expected to show "perfect" exponential decay in the ACF and to have a single spike at PACF(1). These expected patterns can be counted on only when the process realization (the time series, that is) is infinitely long, however. If a time series is not infinitely long, the estimated ACF and PACF will not "perfectly" match the expected ACF and PACF of the underlying process.

Ambiguity in identification sometimes amounts to differences of opinion or interpretation. One analyst may see two spikes in the estimated ACF whereas some other analyst may see only one spike. The first analyst will then conclude that an ARIMA (0,0,2) model adequately represents the series

$$y_t = (1 - \Theta_1 B - \Theta_2 B^2)a_t$$

while the second analyst will conclude that an ARIMA (0,0,1) model adequately represents the series

$$y_t = (1 - \Theta_1 B)a_t.$$

When we discuss parameter estimation, it will be apparent that differences of opinion such as this can be decided absolutely. For the time being, however, we note that ambiguity in estimated ACFs and PACFs can be lessened somewhat by placing confidence bands around the ACFs and PACFs. For the ACF, standard errors (SE) of the ACF(k) are estimated from the formula

$$SE\,[ACF\,(k)] = \sqrt{1/N(1 + 2 \sum_{i=1}^{k} [ACF\,(i)]^2}.$$

For the PACF, standard errors of the PACF (k) are estimated from the formula

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

In the example analyses of Section 2.12, we will plot confidence intervals around the ACF and PACF at $\pm$ 2 SE. Values of ACF (k) and PACF (k) which lie inside this interval will be considered not significantly different than zero.

The difference between the pattern of autocorrelation in the theoretical ACF of an infinitely long series and the pattern generated by finite samples is



vividly illustrated in Figures 2.11(b) and 2.11(c). Figure 2.11(b) presents lags 1 to 10 of the theoretical ACF of the ARIMA (1,0,0) model

$$y_t = .5y_{t-1} + a_t.$$

Using this same model, we have generated six realizations of the process, each 100 observations long. Each realization was generated with NID (0,1) random shocks. Figure 2.11(c) shows the first 10 lags of the ACF for these realizations.

*Note that, even though these six realizations were generated by the same ARIMA (1,0,0) process, there is a wide range in their estimated ACFs.* The distinctive exponential decay pattern of an ARIMA (1,0,0) ACF is clearly seen in some of the estimated ACFs but is largely obscured in others. The estimated PACFs for these six realizations (not presented) clearly indicate an ARIMA (1,0,0) process, however, with statistically significant estimates of PACF(1) and all other lags not statistically significant. It is also important to note that, as the length of realization increases, the pattern of decay in the estimated ACF converges to the pattern expected of an ARIMA (1,0,0) process. If these six realizations were 200 observations long, identification of an ARIMA (1,0,0) model from the ACF would be more certain.

These simulated identifications illustrate the value of using all available data. The analyst should also use both the ACF and the PACF to identify a tentative model, rather than relying solely on the ACF and its SE. In general, a conservative approach to identification is urged. The more parsimonious

```
-1   -.8  -.6  -.4  -.2   0.   .2   .4   .6   .8  +1
+----+----+----+----+----+----+----+----+----+----+
                      (    IXXXX)XXXXXXXX
                     (     IXXXXX*
                     (     IXXX   )
                     (     IXX    )
                     (     IX     )
                     (     I      )
                     (     I      )
                     (     I      )
                     (     I      )
                     (     I      )
```

FIGURE 2.11(b) Expected ACF for the Process $Y_t = .5y_{t-1} + a_t$; to aid comparison with figure 2.11(c), standard errors have been calculated with T=100 observations

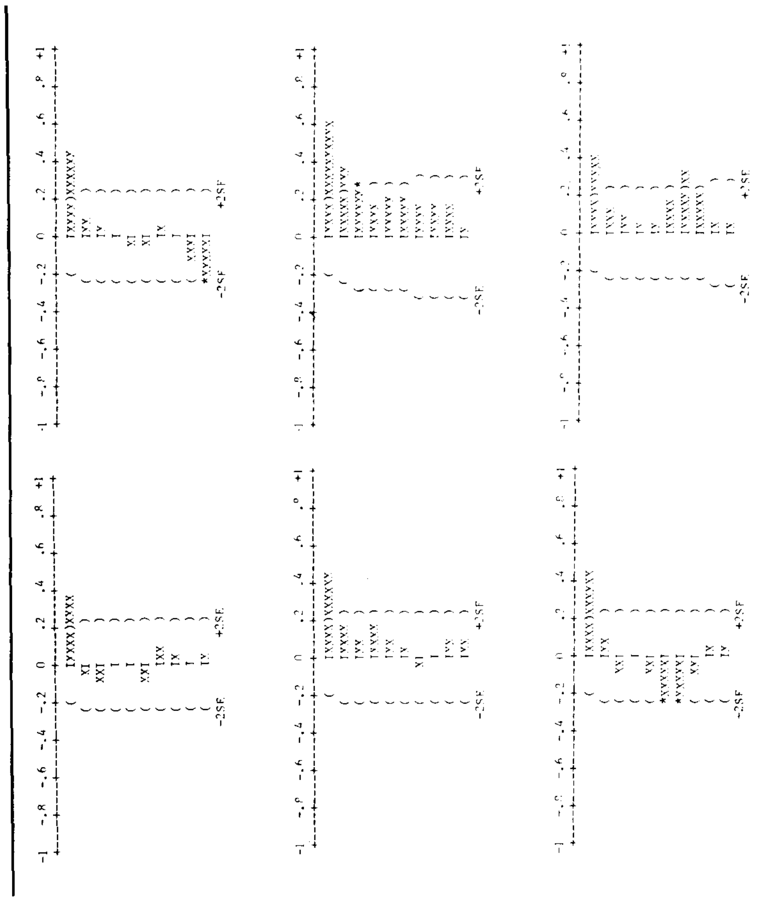


FIGURE 2.11(c) Estimated ACFs for the Process

models (p,P,q, and Q = 0 or 1) should be ruled out before more complicated models are entertained. Should a more parsimonious model prove inadequate, the inadequacy will become apparent at a later stage of the model-building strategy. An overly complicated model, on the other hand, may pass through all subsequent stages of the strategy without notice.

The conservative approach is especially urged when differencing is an issue. The impatient analyst may difference a time series generated by a stationary process and, as a result, may select an overly complicated, cum-



bersome model. The costs of *over*differencing a time series are easily demonstrated. A relatively simple ARIMA (0,0,1) process, for example,

$$y_t = (1 - \Theta_1 B)a_t,$$

when differenced becomes

$$\begin{aligned} y_t - y_{t-1} &= (1 - \Theta_1 B)a_t - (1 - \Theta_1 B)a_{t-1} \\ &= a_t - \Theta_1 a_{t-1} - a_{t-1} + \Theta_1 a_{t-2} \\ &= a_t - (\Theta_1 + 1)a_{t-1} + \Theta_1 a_{t-2} \end{aligned}$$

an ARIMA (2,0,0) process. Moreover, as the parameters of this process are likely to exceed the bounds of invertibility, the analyst may end up with a high-order ARIMA (p,0,0) model for the time series. Whenever the realization of a stationary process is differenced (when the time series is overdifferenced, that is), autoregressive or moving average structures must be incorporated in the model to remove the effects of differencing. This is a classic example of parameter redundancy. To avoid this problem, a time series should not be differenced unless its ACF clearly indicates a nonstationary process. And of course, a time series should never be differenced before its ACF is examined.

## Estimation

*Estimation* follows identification. Having tentatively identified an ARIMA (p,d,q) $(P,D,Q)_S$ model for the time series, the $\phi_p$, $\phi_P$, $\Theta_q$, and $\Theta_Q$ parameters of the tentative model must be estimated. *An ARIMA (p,d,q) $(P,D,Q)_S$ model is generally nonlinear in its parameters and this means that standard OLS regression software such as SPSS cannot be used for estimation.* All university computing centers will have either a general nonlinear regression program or a software package designed especially for ARIMA estimation.[10] The analyst will input the time series and an ARIMA (p,d,q) $(P,D,Q)_S$ model and will receive parameter estimates as output. The analyst will have two major concerns here.

*First, the estimated autoregressive and moving average parameters should be statistically significant.* Any parameter whose estimated value is not significantly different than zero should be dropped from the tentative model. In our discussion of identification, we cited a hypothetical situation in which ARIMA (0,0,1) and ARIMA (0,0,2) models were posited for the same time series, noting that this type of disagreement could be resolved absolutely. In fact, if the estimate of $\Theta_2$ for the ARIMA (0,0,2) model is not statistically significant, the $\Theta_2$ parameter should be dropped from the model.

*Second, the estimated autoregressive and moving average parameters must lie within the bounds of stationarity and invertibility.* If the estimated parameters of the tentative model do not satisfy the stationarity-invertibility conditions, then the tentative model must be rejected. If a stationary time series has been differenced incorrectly, or if a nonstationary time series has not be differenced, autoregressive and/or moving average parameter estimates will invariably exceed the bounds of stationarity or invertibility. Whatever the cause of the problem, however, the tentative model must be rejected if the estimated parameters do not satisfy the stationarity-invertibility conditions.

## Diagnosis

Having identified a tentative ARIMA (p,d,q) $(P,D,Q)_S$ model, and having satisfactorily estimated its parameters, the model must be *diagnosed*. A statistically *adequate* model is defined as one whose residuals ($\hat{a}_t$) satisfy two diagnostic criteria.

*First, the residuals of the tentative model must be independent at a first and second lag,* that is,

$$E(\hat{a}_t \hat{a}_{t+1}) = E(\hat{a}_t \hat{a}_{t+2}) = 0.$$

To evaluate the statistical adequacy of the tentative model by this criterion, we estimate an ACF for the residuals. If the tentative model is statistically adequate by this criterion, then

$$ACF(1) = ACF(2) = 0.$$

The ACF for the model residuals must have no statistically significant values at the first two lags, that is, ACF (1) and ACF (2) must lie within the confidence intervals we have plotted around the ACF. For seasonal data, we will also require

$$ACF(S) = ACF(2S) = 0,$$

that is, the model residuals must be free of autocorrelation at the seasonal lags. Box and Jenkins (1976: 291) note that the standard error of the ACF may significantly underestimate the true standard error of the residual ACF depending on model form and parameter values. Therefore, in residual diagnosis, the ACF standard error should be considered an upper limit. If a low-order lag autocorrelation is slightly less than two standard errors in magnitude, then the prudent analyst may wish to consider it significant for diagnostic purposes.



If the tentative model is not statistically adequate by this criterion, the analyst may often remedy the situation by increasing the values of p, P, q, or Q. In the hypothetical situation of ARIMA (0,0,1) and ARIMA (0,0,2) models posited for the same time series, we see that the disagreement could be settled at the diagnosis stage also. If the analyst posits an ARIMA, (0,0,1) model but finds statistically significant autocorrelation at lag-2 of the residual ACF, then the ARIMA (0,0,1) model must be rejected. Similarly, if the analyst posits an ARIMA (p,d,q) model for a time series but finds statistically significant autocorrelation at seasonal lags of the residual ACF, then the ARIMA (p,d,q) model must be rejected in favor of an ARIMA (p,d,q) $(P,D,Q)_S$ model.

*The second criterion of statistical adequacy is that the residuals of the tentative model must be distributed as white noise.* The ACF of a white noise process is expected to be uniformly zero. For 20 or 30 lags of an ACF, however, given a .05 level of statistical significance, we anticipate that there will be two or three significant spikes by chance alone. To test whether the *entire* residual ACF is different from what would be expected of a white noise process, the analyst may use a Q statistic given by the formula[11]

$$Q = N \sum_{i=1}^{k} [ACF(i)]^2$$

with df = k − p − q − P − Q. The Q statistic is distributed approximately chi-square with the degrees of freedom as indicated. A null hypothesis that the model residuals are white noise is:

$$H_0\colon ACF(1) = \ldots = ACF(k) = 0.$$

That is, the null hypothesis states that the residual ACF is not different than a white noise ACF. If the Q statistic for the residual ACF is significant, the null hypothesis must be rejected; the model residuals are not white noise, so the tentative model is not statistically adequate and must be rejected. If the Q statistic is not statistically significant, the null hypothesis is not rejected; the model residuals are not significantly different from white noise and the tentative model is accepted.

The Q statistic is sensitive to the value of k, that is, to the number of lags in the residual ACF. For a relatively long ACF, say 50 lags, the Q statistic is likely to understate the serial correlation in the model residuals. Even if the residuals are not white noise, the Q statistic for a relatively long ACF is not likely to be statistically significant. The problem here is that, using an ACF of 20 lags, the null hypothesis might be rejected. For the same set of residuals, using an ACF of 40 lags, the null hypothesis might not be rejected. To solve this problem, one should use an ACF that is neither "too long" nor "too short." In the example analyses in Section 2.12, we usually use ACFs of 25 lags. This length allows us to examine the seasonal lags of monthly data (lags 12 and 24 of the ACF and PACF) while still allowing a fair test for the Q statistic. In our experience, 25 lags is usually neither "too long" nor "too short" for most data but, of course, this length is somewhat arbitrary. Granger and Newbold (1977) do, however, recommend that a minimum of 20 lags always be used in the calculation of Q.

We note finally that each of the two criteria of statistical adequacy is *necessary* but not *sufficient* grounds for accepting a tentative model. Obviously, a set of model residuals can meet one criterion but not the other. To be accepted, however, the tentative model must be statistically adequate; and to be statistically adequate, the tentative model must meet both of these diagnostic criteria.

A variety of other residual checks may prove useful in diagnosing the estimated model. Box and Jenkins (1976) suggest methods for checking for Normality and for investigating possible seasonal dependencies in the residuals. The latter are often of doubtful utility given the limited length of many social science time series. We strongly recommend inspecting a plot of the residual series and a plot of the predicted values versus the observed values. Both of these are invaluable for assessing the fit and adequacy of the model, particularly with regard to potential sources of variance nonstationarity such as outliers and variance proportional to the series level.

## Metadiagnosis

After the tentative model has been identified, its parameters estimated, and its residuals diagnosed, the analyst may accept the tentative model. However, the prudent analyst may wish to consider a set of factors over and above those implied by the identification/estimation/diagnosis strategy. We call this procedure *meta*diagnosis.

Having accepted the tentative model, the analyst can be sure only that the model is statistically adequate and parsimonious. These are *relative* qualities of the model which say very little about certain absolute concerns. In building an ARIMA model for a time series, the analyst plans to *use* the model for some purpose. We do not discuss the uses of ARIMA models in this chapter. But generally, the analyst plans to use the ARIMA model for impact assessment (which we discuss in Chapter 3), for forecasting (which we cover in Chapter 4), or for multivariate causal analyses (which we cover in Chapter 5). If the ARIMA model is used for any or all of these purposes, the analyst



should consider the absolute qualities of the model in light of its projected use.

First, in absolute terms, how good is the ARIMA model? There are a number of criteria which could be used for this assessment. However, the most reasonable criterion would seem to be the $R^2$ statistic computed as

$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}} = 1 - \sum_{t=1}^{N} \frac{\hat{a}_t^2}{Y_t^2}.$$

The $R^2$ statistic has the same interpretation here as in cross-sectional multiple regression analysis. It is the percent of variance in the time series that is "explained" by the model. The reader who is not familiar with time series methods may be surprised to learn that time series models routinely have $R^2$ statistics higher than .9. Of course, the greatest portion of this explained variance is due to the parameter $\Theta_0$. A more realistic $R^2$ statistic, then, will be:

$$R^2 = 1 - \sum_{t=1}^{N} \frac{\hat{a}_t^2}{y_t^2}.$$

where $y_t = Y_t - \Theta_0$ for a stationary process and $y_t = z_t - \Theta_0$ for a nonstationary process. By subtracting the parameter $\Theta_0$ from each time series observation, the analyst obtains an $R^2$ statistic which measures the percent variance explained only by the autoregressive and/or moving average parameters of the model. Naturally, the analyst will require a relatively high $R^2$ statistic for the model.

A statistic related to the $R^2$ is the residual mean square (RMS) statistic computed from the formula

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

for a set of N residuals. Like the $R^2$, the RMS statistic gives a "goodness-of-fit" measure for the model. Unlike the $R^2$, however, the RMS statistic is not standardized. By tradition, the RMS statistic is more widely used in time series analysis than the $R^2$ and we will follow that tradition so far as possible.

Metadiagnosis ordinarily begins with *over*modeling. If the iterative identification/estimation/diagnosis procedure has lead to an ARIMA (0,1,1) model for a time series, the analyst should try to fit an ARIMA (0,1,2) model. If the ARIMA (0,1,1) model has been judged statistically adequate in diagnosis, the estimated $\Theta_2$ parameter of the ARIMA (0,1,2) model should be statistically insignificant. In general, overmodeling amounts to increasing the values of p, P, q, and Q for the ARIMA (p,d,q) $(P,D,Q)_S$ model and, in general, if the accepted model has been judged statistically adequate, the estimated parameters of the higher order model should be statistically insignificant.

Another dimension of overmodeling might be called *under*modeling. In Sections 2.5 and 2.6, we demonstrated that an ARIMA (1,0,0) model could be expressed as an infinite order ARIMA (0,0,q) model, that is,

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + a_t \\ &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots + \phi_1^n a_{t-n} + \ldots, \end{aligned}$$

and that an ARIMA (0,0,1) model could be expressed as an infinite order ARIMA (p,0,0) model, that is,

$$\begin{aligned} y_t &= a_t - \Theta_1 a_{t-1} \\ &= a_t - \Theta_1 y_{t-1} - \Theta_1^2 y_{t-2} - \ldots - \Theta_1^n y_{t-n} - \ldots, \end{aligned}$$

and due to the conditions of stationarity-invertibility, the infinite series both converge absolutely to zero. More to the point, after two or three terms, for small values of $\phi_1$ and $\Theta_1$, the values of $\phi_1^n$ and $\Theta_1^n$ are approximately zero. Given this, it is possible that an ARIMA (0,0,2) model might be approximately identical with an ARIMA(1,0,0) model. In practice, then, if the accepted model is ARIMA(0,0,2),

$$y_t = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2}$$

and if $\Theta_2 \simeq \Theta_1^2$, the analyst can almost always find a statistically adequate ARIMA (1,0,0) model for the series

$$y_t = \phi_1 y_{t-1} + a_t$$

where $\phi_1 \simeq \Theta_1$ Now which of these two alternative models is the "better?"

In this case, the competition might be decided on the basis of parsimony alone. The ARIMA (1,0,0) model has only one parameter and this is "better" than the ARIMA (0,0,2) model. The analyst must remember, however, that the ARIMA (1,0,0) and ARIMA (0,0,2) models are only *approximately* identical. In fact, if the underlying process is ARIMA (0,0,2), an ARIMA (0,0,2) model will be the "better" under all circumstances. Parsimony not withstanding, the RMS statistics of the two models can be used as a measure



of quality. If the ARIMA (1,0,0) model has a lower (or even approximately equal) RMS statistic, it must be deemed the "better" model because it is the more parsimonious of the two. But if the ARIMA (0,0,2) model has a significantly lower RMS statistic, the rule of parsimony may be waived.

The rough equivalence of autoregressive and moving average models becomes more of a problem when seasonal factors are considered. An ARIMA (0,0,1) $(0,0,2)_{12}$ model, for example,

$$y_t = (1 - \Theta_1 B)(1 - \Theta_{12}B^{12} - \Theta_{24}B^{24})a_t$$

will be approximately identical with an ARIMA (0,0,1) $(1,0,0)_{12}$ model

$$(1 - \phi_{12}B^{12})y_t = (1 - \Theta_1 B)a_t$$

whenever $\Theta_{24} \simeq \Theta_{12}^2$. The reader may demonstrate this simple fact by "solving" the ARIMA (0,0,1) $(1,0,0)_{12}$ model. Whenever P or Q is greater than one,then, the prudent analyst will undermodel the series, comparing alternatives.

Metadiagnosis is perhaps the most critical stage of the model-building strategy. In metadiagnosis, the analyst plays the role of devil's advocate, criticizing and arguing as best as possible that a "better" model can be found. Failing, the analyst should be convinced that the "best" possible model has been built for the time series. We will return to this topic in the next few chapters, discussing metadiagnostic techniques which pertain to the specific applications of ARIMA modeling: impact analysis, forecasting, and multivariate analysis.

## 2.12 Example Analyses

We suspect that, after digesting the material preceding this, the reader still has a number of unanswered questions. In this section, we hope to answer many of these questions by presenting a few in-depth example analyses. The series we will analyze here are listed in an appendix to this volume and the reader is invited to replicate our analyses, checking (or challenging) our results. If our experiences in teaching time series analysis are typical, many questions the reader may have can be answered only through personal experience. The surest way to learn time series analysis is through informed practice.

We have selected these series for analysis because each illustrates one or more practical problems that the analyst is likely to encounter. Collectively, we have analyzed hundreds of social science time series in the last few years and the series we analyze here are "typical." We include stationary series and nonstationary series and one time series that is nonstationary in both level and variance. One of the series is "not a time series" in the strictest sense and another is distorted by a deviant value or outlier.

The ACFs and PACFs for these analyses are the printed output of SCRUNCH (Hay, 1979), an interactive software package for Box-Jenkins time series analysis. The output of SCRUNCH is similar to most Box-Jenkins time series packages. Parentheses about the ACFs and PACFs indicate confidence intervals of $\pm 2$ standard errors. Hence, any estimated ACF(k) or PACF(k) within the parentheses are not significantly different than zero. The values of each ACF (k) or PACF (k) along with their respective standard errors are listed alongside the correlogram plot.

### 2.12.1 Sutter County Workforce

The time series plotted in Figure 2.12.1(a) are monthly workforce statistics (the total number of people employed in the workforce) for Sutter County, California. The first observation of this series is January, 1946. The 252nd observation is December 1966. From an eyeball inspection of the plotted series, it seems obvious that this series is nonstationary and seasonal. In fact, as the series level appears to increase in annual steps, we suspect that this series may be seasonally nonstationary. We will reserve judgment on this issue until after we have inspected the ACFs and PACFs, however. *And of course, we will follow the model-building strategy we have outlined in Figure 2.11(a).*

*Identification.* The ACF of the raw time series shown in Figure 2.12.1(b) indicates nonstationarity as we had suspected. There is no evidence of decay and the high-order lags remain significant. The series must be differenced. The ACF of the regularly differenced series shows seasonal nonstationarity as well. The basis for this identification is seen at lags $-12$ and $-24$ of the ACF. Both lags are large and nearly equal. After differencing this series both regularly and seasonally, the ACF and PACF suggests an ARIMA $(0,1,1)(0,1,1)_{12}$ model. We arrive at this identification by noting the ACF spikes at lag$-1$ and lag$-12$ while the PACF exhibits rough patterns of decay beginning at lag$-1$ and lag$-12$. We write this tentative model as

$$(1-B)(1-B^{12})Y_t = \Theta_0 + (1-\Theta_1 B)(1-\Theta_{12}B^{12})a_t$$

or

$$Y_t = \frac{\Theta_0 + (1-\Theta_1 B)(1-\Theta_{12}B^{12})}{(1-B)(1-B^{12})}a_t.$$



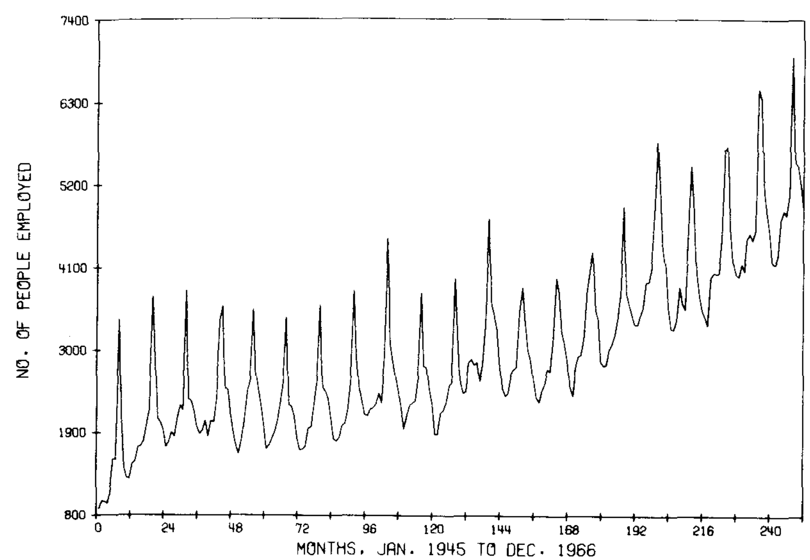



FIGURE 2.12.1(a) Sutter County Workforce

The model has three parameters, $\Theta_0$, $\Theta_1$, and $\Theta_{12}$, which must now be estimated.

*Estimation.* Parameter estimates for the tentative model are[12]:

$$\hat{\Theta}_0 = .52 \text{ with t statistic} = .22$$
$$\hat{\Theta}_1 = .60 \text{ with t statistic} = 11.38$$
$$\hat{\Theta}_{12} = .68 \text{ with t statistic} = 13.33.$$

First, we note that both $\hat{\Theta}_1$ and $\hat{\Theta}_{12}$ lie within the bounds of invertibility for moving average parameters, so both parameters are acceptable by that criterion. Using a .05 level of significance, however, we require a t statistic of $\pm 1.96$ and, at this level, the estimated value of $\Theta_0$ is not statistically different than zero. What this means is that the upward motion of this time series is not significantly different than *drift* and, as a result, $\Theta_0$ is dropped from our tentative model. Both $\hat{\Theta}_1$ and $\hat{\Theta}_{12}$ are statistically significant, so our tentative model is:

$$(1-B)(1-B^{12})Y_t = (1-.60B)(1-.68B^{12})a_t.$$

*(text continued on p. 110)*

```
SERIES.. EMPLOY    (NOBS= 252)  SUTTER COUNTY EMPLOYMENT, 1/45 TO 12/66
NO. OF VALID OBSERVATIONS  =    252.

AUTOCORRELATIONS OF LAGS 1 - 30.
Q( 30, 252) =  2277.8      SIG =     0.000

LAG   CORR    SE   -1  -.8  -.6  -.4  -.2   0   .2   .4   .6   .8   +1
                   +----+----+----+----+----+----+----+----+----+----+
  1   .874  .063                         (  IXX)XXXXXXXXXXXXXXXXXXX
  2   .757  .100                       (    IXXXX)XXXXXXXXXXXXXX
  3   .664  .121                      (     IXXXXX)XXXXXXXXXXX
  4   .585  .134                     (      IXXXXXX)XXXXXXXX
  5   .514  .144                     (      IXXXXXX)XXXXXX
  6   .484  .151                    (       IXXXXXXX)XXXX
  7   .490  .157                    (       IXXXXXXX)XXXX
  8   .534  .163                    (       IXXXXXXX)XXXXX
  9   .580  .170                   (        IXXXXXXXX)XXXXX
 10   .633  .178                   (        IXXXXXXXX)XXXXXXX
 11   .722  .186                   (        IXXXXXXXX)XXXXXXXXX
 12   .792  .197                  (         IXXXXXXXXX)XXXXXXXXXX
 13   .700  .209                  (         IXXXXXXXXX)XXXXXXX
 14   .595  .219                 (          IXXXXXXXXXX)XXXX
 15   .512  .225                 (          IXXXXXXXXXX)XX
 16   .432  .229                 (          IXXXXXXXXXX*
 17   .368  .233                (           IXXXXXXXXX  )
 18   .341  .235                (           IXXXXXXXXX  )
 19   .347  .237                (           IXXXXXXXXX  )
 20   .392  .239                (           IXXXXXXXXXX )
 21   .444  .241                (           IXXXXXXXXXXX)
 22   .499  .245                (           IXXXXXXXXXXX*
 23   .578  .249                (           IXXXXXXXXXXX)XX
 24   .653  .254               (            IXXXXXXXXXXXX)XXX
 25   .578  .261               (            IXXXXXXXXXXXX)X
 26   .486  .266               (            IXXXXXXXXXXXX)
 27   .405  .269               (            IXXXXXXXXXX  )
 28   .333  .272              (             IXXXXXXXX     )
 29   .278  .273              (             IXXXXXXX      )
 30   .249  .274              (             IXXXXXX       )
                            -2SE                         +2SE

PARTIAL AUTOCORRELATIONS OF LAGS 1 - 30.

LAG   CORR    SE   -1  -.8  -.6  -.4  -.2   0   .2   .4   .6   .8   +1
                   +----+----+----+----+----+----+----+----+----+----+
  1   .874  .063                         (  IXX)XXXXXXXXXXXXXXXXXXX
  2  -.027  .063                         ( XI  )
  3   .038  .063                         (  IX )
  4   .007  .063                         (  I  )
  5  -.008  .063                         (  I  )
  6   .142  .063                         (  IXX)X
  7   .149  .063                         (  IXX)X
  8   .217  .063                         (  IXX)XX
  9   .114  .063                         (  IXX*
 10   .165  .063                         (  IXX)X
 11   .336  .063                         (  IXX)XXXXX
 12   .224  .063                         (  IXX)XXX
 13  -.512  .063               XXXXXXXXXX(XXI  )
 14  -.163  .063                        X(XXI  )
 15  -.005  .063                         (  I  )
 16  -.047  .063                         ( XI  )
 17   .024  .063                         (  IX )
 18   .003  .063                         (  I  )
 19  -.004  .063                         (  I  )
 20   .062  .063                         (  IXX)
 21   .106  .063                         (  IXX*
 22   .052  .063                         (  IX )
 23  -.009  .063                         (  I  )
 24   .200  .063                         (  IXX)XX
 25  -.182  .063                       XX(XXI  )
 26  -.072  .063                         (XXI  )
 27  -.018  .063                         (  I  )
 28   .012  .063                         (  I  )
 29   .022  .063                         (  IX )
 30  -.050  .063                         ( XI  )
                                       -2SE  +2SE
```

FIGURE 2.12.1(b) ACF and PACF for the Raw Series



```
SERIES.. EMPLOY    (NOBS= 252)  SUTTER COUNTY EMPLOYMENT, 1/45 TO 12/66
DIFFERENCED  1 TIME(S) OF ORDER  1.
NO. OF VALID OBSERVATIONS  =    251.

AUTOCORRELATIONS OF LAGS 1 - 30.
Q( 30, 251) =  296.02      SIG =     0.000

LAG   CORR    SE   -1  -.8  -.6  -.4  -.2   0   .2   .4   .6   .8   +1
                   +----+----+----+----+----+----+----+----+----+----+
  1  -.046  .063                         ( XI  )
  2  -.116  .063                         *XXI  )
  3  -.062  .064                         (XXI  )
  4  -.066  .064                         (XXI  )
  5  -.124  .065                         *XXI  )
  6  -.151  .066                        X(XXI  )
  7  -.116  .067                         *XXI  )
  8  -.041  .068                         ( XI  )
  9  -.043  .068                         ( XI  )
 10  -.157  .068                        X(XXI  )
 11   .080  .069                         (  IXX)
 12   .713  .070                         (  IXX)XXXXXXXXXXXXXXX
 13   .045  .094                       (    IX   )
 14  -.096  .095                       (  XXI    )
 15  -.051  .095                       (   XI    )
 16  -.074  .095                       (  XXI    )
 17  -.101  .095                       ( XXXI    )
 18  -.144  .096                       (XXXXI    )
 19  -.127  .097                       ( XXXI    )
 20  -.049  .097                       (   XI    )
 21  -.032  .097                       (   XI    )
 22  -.111  .097                       ( XXXI    )
 23   .021  .098                       (    IX   )
 24   .651  .098                       (    IXXXX)XXXXXXXXXXX
 25   .074  .114                      (     IXX    )
 26  -.057  .114                      (    XI     )
 27  -.072  .114                      (   XXI     )
 28  -.069  .114                      (   XXI     )
 29  -.077  .114                      (   XXI     )
 30  -.171  .115                      ( XXXXI     )
                                   -2SE         +2SE

PARTIAL AUTOCORRELATIONS OF LAGS 1 - 30.

LAG   CORR    SE   -1  -.8  -.6  -.4  -.2   0   .2   .4   .6   .8   +1
                   +----+----+----+----+----+----+----+----+----+----+
  1  -.048  .063                         ( XI  )
  2  -.119  .063                         *XXI  )
  3  -.075  .063                         (XXI  )
  4  -.090  .063                         (XXI  )
  5  -.156  .063                        X(XXI  )
  6  -.208  .063                       XX(XXI  )
  7  -.221  .063                      XXX(XXI  )
  8  -.199  .063                       XX(XXI  )
  9  -.249  .063                      XXX(XXI  )
 10  -.472  .063                XXXXXXXXX(XXI  )
 11  -.496  .063                XXXXXXXXX(XXI  )
 12   .406  .063                         (  IXX)XXXXXXX
 13   .183  .063                         (  IXX)XX
 14   .028  .063                         (  IX )
 15   .019  .063                         (  I  )
 16   .008  .063                         (  I  )
 17   .060  .063                         (  IX )
 18   .060  .063                         (  IX )
 19  -.013  .063                         (  I  )
 20  -.093  .063                         (XXI  )
 21  -.055  .063                         ( XI  )
 22   .055  .063                         (  IX )
 23  -.257  .063                      XXX(XXI  )
 24   .066  .063                         (  IXX)
 25   .033  .063                         (  IX )
 26   .029  .063                         (  IX )
 27  -.035  .063                         ( XI  )
 28  -.026  .063                         ( XI  )
 29   .020  .063                         (  IX )
 30  -.075  .063                         (XXI  )
                                       -2SE  +2SE
```

FIGURE 2.12.1(c) ACF and PACF for the Regularly Differenced Series

```
SERIES.. EMPLOY    (NOBS= 252)  SUTTER COUNTY EMPLOYMENT, 1/45 TO 12/66
DIFFERENCED  1 TIME(S) OF ORDER  1.
DIFFERENCED  1 TIME(S) OF ORDER 12.
NO. OF VALID OBSERVATIONS  =    239.

AUTOCORRELATIONS OF LAGS 1 - 30.
Q( 30, 239) =  151.93       SIG =      .000

 LAG   CORR    SE   -1   -.8  -.6  -.4  -.2  0    .2   .4   .6   .8   +1
                    +----+----+----+----+----+----+----+----+----+----+
   1  -.430  .065                 XXXXXXXX(XXI    )
   2   .065  .076                        (   IXX )
   3  -.076  .076                        ( XXI   )
   4   .002  .076                        (   I   )
   5  -.056  .076                        (  XI   )
   6   .014  .076                        (   I   )
   7   .049  .076                        (   IX  )
   8   .039  .077                        (   IX  )
   9   .040  .077                        (   IX  )
  10  -.144  .077                        *XXXI   )
  11   .306  .078                        (    IXXX)XXXX
  12  -.439  .083                 XXXXXXX(XXXI    )
  13   .120  .092                       (    IXXX )
  14  -.037  .093                       (   XI    )
  15   .066  .093                       (    IXX  )
  16  -.045  .093                       (   XI    )
  17   .025  .093                       (    IX   )
  18   .058  .093                       (    IX   )
  19  -.105  .093                       ( XXXI    )
  20   .030  .094                       (    IX   )
  21  -.021  .094                       (   XI    )
  22   .088  .094                       (    IXX  )
  23  -.150  .094                       (XXXXI    )
  24   .049  .095                       (    IX   )
  25   .001  .095                       (    I    )
  26   .131  .095                       (    IXXX )
  27  -.132  .096                       ( XXXI    )
  28   .030  .097                       (    IX   )
  29   .060  .097                       (    IXX  )
  30  -.123  .097                       ( XXXI    )
                                      -2SE      +2SE

PARTIAL AUTOCORRELATIONS OF LAGS 1 - 30.

 LAG   CORR    SE   -1   -.8  -.6  -.4  -.2  0    .2   .4   .6   .8   +1
                    +----+----+----+----+----+----+----+----+----+----+
   1  -.430  .065                  XXXXXXXX(XXI  )
   2  -.147  .065                         X(XXI  )
   3  -.135  .065                          *XXI  )
   4  -.103  .065                          *XXI  )
   5  -.136  .065                          *XXI  )
   6  -.104  .065                          *XXI  )
   7  -.007  .065                          (  I  )
   8   .061  .065                          (  IXX)
   9   .115  .065                          (  IXX*
  10  -.081  .065                          (XXI  )
  11   .300  .065                          (  IXX)XXXXX
  12  -.230  .065                       XXX(XXI  )
  13  -.191  .065                        XX(XXI  )
  14  -.134  .065                          *XXI  )
  15  -.090  .065                          (XXI  )
  16  -.119  .065                          *XXI  )
  17  -.129  .065                          *XXI  )
  18   .008  .065                          (  I  )
  19  -.103  .065                          *XXI  )
  20  -.011  .065                          (  I  )
  21   .082  .065                          (  IXX)
  22   .029  .065                          (  IX )
  23   .062  .065                          (  IXX)
  24  -.162  .065                         X(XXI  )
  25  -.141  .065                         X(XXI  )
  26   .040  .065                          (  IX )
  27  -.093  .065                          (XXI  )
  28  -.129  .065                          *XXI  )
  29  -.041  .065                          ( XI  )
  30  -.074  .065                          (XXI  )
                                         -2SE  +2SE
```

FIGURE 2.12.1(d) ACF and PACF for the Regularly *and* Seasonally Differenced Series



```
SERIES.. RESIDUAL (NOBS= 239)   SUTTER COUNTY EMPLOYMENT RESIDUALS
NO. OF VALID OBSERVATIONS  =    239.

AUTOCORRELATIONS OF LAGS 1 - 30.
Q( 28, 239) =  28.304       SIG =      .448

 LAG   CORR    SE   -1   -.8  -.6  -.4  -.2  0    .2   .4   .6   .8   +1
                    +----+----+----+----+----+----+----+----+----+----+
   1   .050  .065                          (  IX )
   2  -.016  .065                          (  I  )
   3  -.112  .065                          *XXI  )
   4  -.054  .066                          ( XI  )
   5  -.024  .066                          ( XI  )
   6   .077  .066                          (  IXX)
   7   .080  .066                          (  IXX)
   8   .087  .067                          (  IXX)
   9  -.025  .067                          ( XI  )
  10  -.127  .067                          *XXI  )
  11   .037  .068                          (  IX )
  12  -.009  .068                          (  I  )
  13  -.034  .068                          ( XI  )
  14  -.031  .068                          ( XI  )
  15  -.065  .068                          (XXI  )
  16  -.061  .069                          (XXI  )
  17   .025  .069                          (  IX )
  18   .034  .069                          (  IX )
  19  -.033  .069                          ( XI  )
  20   .008  .069                          (  I  )
  21  -.020  .069                          ( XI  )
  22  -.034  .069                          ( XI  )
  23  -.113  .069                          *XXI  )
  24   .017  .070                          (  I  )
  25   .123  .070                          (  IXX*
  26   .098  .071                         (   IXX )
  27  -.072  .071                         ( XXI   )
  28  -.024  .072                         (  XI   )
  29   .037  .072                         (   IX  )
  30  -.043  .072                         (  XI   )
                                        -2SE    +2SE

PARTIAL AUTOCORRELATIONS OF LAGS 1 - 30.

 LAG   CORR    SE   -1   -.8  -.6  -.4  -.2  0    .2   .4   .6   .8   +1
                    +----+----+----+----+----+----+----+----+----+----+
   1   .050  .065                          (  IX )
   2  -.018  .065                          (  I  )
   3  -.111  .065                          *XXI  )
   4  -.044  .065                          ( XI  )
   5  -.023  .065                          ( XI  )
   6   .067  .065                          (  IXX)
   7   .064  .065                          (  IXX)
   8   .078  .065                          (  IXX)
   9  -.017  .065                          (  I  )
  10  -.106  .065                          *XXI  )
  11   .074  .065                          (  IXX)
  12  -.016  .065                          (  I  )
  13  -.065  .065                          (XXI  )
  14  -.044  .065                          ( XI  )
  15  -.078  .065                          (XXI  )
  16  -.056  .065                          ( XI  )
  17   .029  .065                          (  IX )
  18   .029  .065                          (  IX )
  19  -.062  .065                          (XXI  )
  20   .013  .065                          (  I  )
  21   .021  .065                          (  IX )
  22  -.025  .065                          ( XI  )
  23  -.113  .065                          *XXI  )
  24   .021  .065                          (  IX )
  25   .104  .065                          (  IXX*
  26   .060  .065                          (  IXX)
  27  -.074  .065                          (XXI  )
  28  -.007  .065                          (  I  )
  29   .064  .065                          (  IXX)
  30  -.032  .065                          ( XI  )
                                         -2SE  +2SE
```

FIGURE 2.12.1(e) *Diagnosis:* ACF and PACF for the Model Residuals

If the residuals from this tentative model fail to meet our diagnostic criteria, we will return to the identification stage.

*Diagnosis*. The ACF and PACF—Figure 2.12.1(d)—for the residuals of our tentative model show no spikes at lag-1 or at the seasonal lags. The Q statistic for the ACF is not statistically significant. With 28 degrees of freedom, the value of Q = 28.3 is associated with a .448 level of significance. The residuals of the tentative model meet both diagnostic criteria, so the model is accepted.

This time series, as well as the others used as examples, are listed in Appendix A of this volume. We suggest that the reader reanalyze the series as an exercise. The Sutter County Workforce series is a textbook example of an ARIMA (0,1,1) $(0,1,1)_{12}$ process. The ACFs and PACFs give clear and unambiguous evidence for this model.

In December 1955, the 120th observation of this series, a major flood forced the evacuation of Sutter County. In the next chapter, we will use the ARIMA model we have identified here to assess the impact of this flood on the level of the Sutter County Workforce series.

### 2.12.2 Boston Armed Robberies

The monthly number of reported armed robberies in Boston, Massachusetts, is plotted in Figure 2.12.2(a). The first observation is January 1966 and the 118th observation is October 1975. Deutsch and Alt (1977: Deutsch, 1979) analyzed this series along with a large number of other Uniform Crime Report time series. In our opinion, the model proposed by Deutsch and Alt does not adequately represent the series. By following the iterative model-building strategy presented in Section 2.11, we will contrast the inadequacies of their model with the empirical characteristics of the data.

It should be noted that there are a number of ways to construct an ARIMA model other than the empirically based procedure we have outlined. We refer to these other methods as *arbitrary methods* and emphasize that they usually will not result in a parsimonious *and* statistically adequate model. For example, an analyst might have identified the same ARIMA model for a number of series all belonging to the same substantive class, such as crime series. The analyst might then be tempted to assume that all other similar substantive time series (e.g., *all* crime series) could best be fit by the same ARIMA model. Furthermore, the analyst may attempt to infer that identification of the same ARIMA model for a number of series provides evidence that the same social process was generating all of the series.

These are fallacies, of course. A univariate ARIMA model is a stochastic or probabilistic description of the outcome of a process operating through time. It provides no information about the inputs generating that process. As



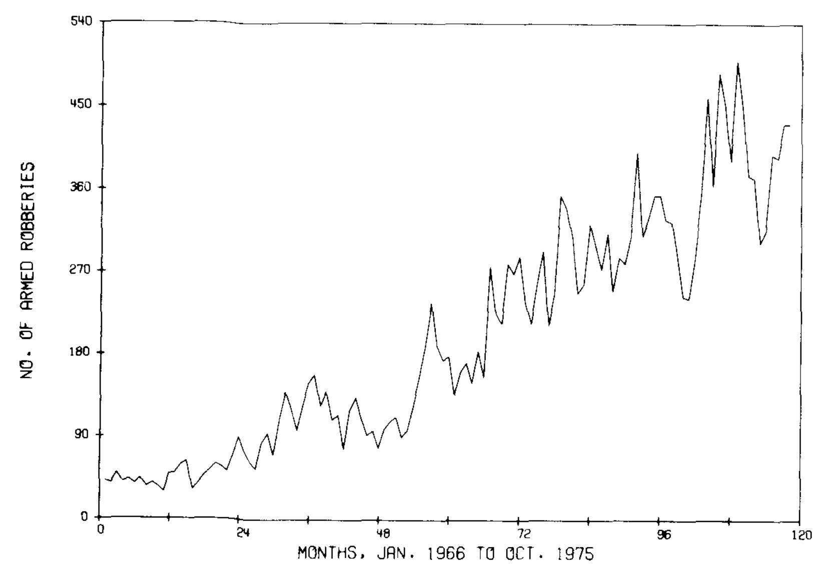



FIGURE 2.12.2(a) Boston Armed Robberies

Hibbs (1977) succinctly points out, "[Univariate] Box-Tiao or Box-Jenkins models are essentially models of ignorance that are not based in theory and, in this sense, are devoid of explanatory power." As in other areas of the social sciences, inference of a causal relationship in time series analysis can only be made through assessment of covariation betweeen one or more explanatory variables and a dependent variable—a crime rate in this case. We develop the methodology for this type of analysis in Chapter 5.

A careful reading of Deutsch and Alt does not clearly reveal their model selection procedure. They do not report identification statistics such as the ACF and PACF and this makes it difficult to assess the adequacy of their models. We recommend that such statistics be routinely reported in time series research so that the social science community may make informed appraisals of the quality of ARIMA models. Although we cannot second guess the Deutsch-Alt model selection procedure, their results are not incongruous with the arbitrary procedure described above. We will now contrast these results with those produced by use of the model-building procedure presented in Section 2.11. The reader is referred to Hay and McCleary (1979) for a more detailed discussion of these issues.

*(text continued on p. 115)*

```
SERIES.. BAR        (NOBS= 118)  BOSTON / MONTHLY ARMED ROBBERY
NO. OF VALID OBSERVATIONS  =    118.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 25, 118) =  1221.4       SIG =     0.000

 LAG   CORR    SE  -1    -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
   1   .928  .092                        (    IXXXX)XXXXXXXXXXXXXXXXXX
   2   .881  .152                      (      IXXXXXXX)XXXXXXXXXXXXXX
   3   .853  .190                    (        IXXXXXXXXX)XXXXXXXXXXX
   4   .808  .220                   (         IXXXXXXXXXX)XXXXXXXXX
   5   .797  .244                  (          IXXXXXXXXXXX)XXXXXXXX
   6   .774  .265                 (           IXXXXXXXXXXXX)XXXXXX
   7   .749  .284                (            IXXXXXXXXXXXXX)XXXXX
   8   .744  .300               (             IXXXXXXXXXXXXXX)XXXX
   9   .712  .315              (              IXXXXXXXXXXXXXXX)XX
  10   .681  .329              (              IXXXXXXXXXXXXXXX)X
  11   .684  .340             (               IXXXXXXXXXXXXXXXX*
  12   .668  .352            (                IXXXXXXXXXXXXXXXXX)
  13   .628  .362            (                IXXXXXXXXXXXXXXXX )
  14   .599  .372           (                 IXXXXXXXXXXXXXXX   )
  15   .550  .380           (                 IXXXXXXXXXXXXXX    )
  16   .519  .386           (                 IXXXXXXXXXXXXX     )
  17   .505  .392          (                  IXXXXXXXXXXXXX      )
  18   .483  .398          (                  IXXXXXXXXXXXX       )
  19   .478  .403          (                  IXXXXXXXXXXXX       )
  20   .465  .407          (                  IXXXXXXXXXXXX       )
  21   .434  .412         (                   IXXXXXXXXXXX         )
  22   .428  .416         (                   IXXXXXXXXXXX         )
  23   .412  .420         (                   IXXXXXXXXXX          )
  24   .394  .423         (                   IXXXXXXXXXX          )
  25   .377  .426         (                   IXXXXXXXXX           )
                        -2SE                                     +2SE
```

```
PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

 LAG   CORR    SE  -1    -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
   1   .928  .092                        (    IXXXX)XXXXXXXXXXXXXXXXXX
   2   .147  .092                        (    IXXXX)
   3   .145  .092                        (    IXXXX)
   4  -.092  .092                        (  XXI    )
   5   .217  .092                        (    IXXXX*
   6  -.034  .092                        (   XI    )
   7   .032  .092                        (    IX   )
   8   .092  .092                        (    IXX  )
   9  -.116  .092                        ( XXXI    )
  10  -.032  .092                        (   XI    )
  11   .188  .092                        (    IXXXX*
  12  -.007  .092                        (    I    )
  13  -.245  .092                       X(XXXXI    )
  14  -.025  .092                        (   XI    )
  15  -.124  .092                        ( XXXI    )
  16   .031  .092                        (    IX   )
  17   .065  .092                        (    IXX  )
  18   .065  .092                        (    IXX  )
  19  -.005  .092                        (    I    )
  20  -.017  .092                        (    I    )
  21  -.002  .092                        (    I    )
  22   .100  .092                        (    IXXX )
  23  -.060  .092                        (   XI    )
  24   .010  .092                        (    I    )
  25  -.078  .092                        (  XXI    )
                                       -2SE      +2SE
```

FIGURE 2.12.2(b) ACF and PACF for the Raw Series



```
SERIES.. BAR        (NOBS= 118)  BOSTON / MONTHLY ARMED ROBBERY
DIFFERENCED  1 TIME(S) OF ORDER  1.
NO. OF VALID OBSERVATIONS  =   117.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 25, 117) =   63.074       SIG =     .000

 LAG   CORR    SE  -1    -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
   1  -.259  .092                        X(XXXXI    )
   2  -.137  .098                         ( XXXI    )
   3   .156  .100                         (    IXXXX)
   4  -.238  .102                        X(XXXXI    )
   5   .113  .107                         (    IXXX )
   6  -.055  .108                         (   XI    )
   7  -.215  .108                        *XXXXI     )
   8   .177  .112                        (     IXXXX )
   9  -.092  .114                        (   XXI     )
  10  -.205  .115                        (XXXXXI     )
  11   .113  .118                        (     IXXX  )
  12   .190  .119                        (     IXXXXX)
  13   .040  .121                        (     IX     )
  14   .076  .121                        (     IXX    )
  15  -.082  .122                        (   XXI      )
  16  -.082  .122                        (   XXI      )
  17   .135  .123                        (     IXXX   )
  18  -.163  .124                        ( XXXXI      )
  19   .023  .126                        (     IX     )
  20   .114  .126                        (     IXXX   )
  21  -.251  .127                        *XXXXXI      )
  22   .079  .131                       (      IXX     )
  23   .024  .131                       (      IX      )
  24   .000  .131                       (      I       )
  25   .174  .131                       (      IXXXX   )
                                      -2SE         +2SE
```

```
PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

 LAG   CORR    SE  -1    -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
   1  -.259  .092                        X(XXXXI    )
   2  -.219  .092                         *XXXXI    )
   3   .062  .092                         (    IXX  )
   4  -.229  .092                        X(XXXXI    )
   5   .026  .092                         (    IX   )
   6  -.125  .092                         ( XXXI    )
   7  -.235  .092                        X(XXXXI    )
   8  -.042  .092                         (   XI    )
   9  -.131  .092                         ( XXXI    )
  10  -.330  .092                      XXX(XXXXI    )
  11  -.253  .092                        X(XXXXI    )
  12   .098  .092                         (    IXX  )
  13   .058  .092                         (    IX   )
  14   .083  .092                         (    IXX  )
  15  -.001  .092                         (    I    )
  16  -.130  .092                         ( XXXI    )
  17  -.011  .092                         (    I    )
  18  -.094  .092                         (  XXI    )
  19   .024  .092                         (    IX   )
  20   .051  .092                         (    IX   )
  21  -.132  .092                         ( XXXI    )
  22   .008  .092                         (    I    )
  23   .036  .092                         (    IX   )
  24   .078  .092                         (    IXX  )
  25   .020  .092                         (    I    )
                                        -2SE      +2SE
```

FIGURE 2.12.2(c) ACF and PACF for the Regularly Differenced Series

```
SERIES.. RESIDUAL (NOBS= 105)  BOSTON ARMED ROBBERY RESIDUALS
NO. OF VALID OBSERVATIONS  =    105.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 23, 105) =  41.871      SIG =     .009

 LAG   CORR    SE  -1   -.8  -.6  -.4  -.2   0   .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
   1   .010  .098                       (    I    )
   2  -.033  .098                       (   XI    )
   3   .158  .098                       (    IXXXX)
   4  -.215  .100                      *XXXXI     )
   5   .009  .104                       (    I    )
   6  -.135  .104                       ( XXXI    )
   7  -.307  .106                    XXX(XXXXI    )
   8   .094  .114                      (     IXX    )
   9  -.088  .115                      (   XXI      )
  10  -.193  .116                      (XXXXXI      )
  11   .070  .119                      (     IXX    )
  12   .025  .119                      (     IX     )
  13   .114  .119                      (     IXXX   )
  14   .190  .120                      (     IXXXXX)
  15  -.043  .123                      (    XI      )
  16  -.100  .123                      (   XXI      )
  17   .082  .124                      (     IXX    )
  18  -.105  .124                      (  XXXI      )
  19  -.018  .125                      (     I      )
  20   .046  .125                      (     IX     )
  21  -.137  .125                      (  XXXI      )
  22   .122  .127                      (     IXXX   )
  23   .006  .128                      (     I      )
  24  -.137  .128                      (  XXXI      )
  25   .146  .129                      (     IXXXX )
                                     -2SE        +2SE


PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

 LAG   CORR    SE  -1   -.8  -.6  -.4  -.2   0   .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
   1   .010  .098                       (    I    )
   2  -.033  .098                       (   XI    )
   3   .159  .098                       (    IXXXX)
   4  -.226  .098                     X(XXXXI     )
   5   .037  .098                       (    IX    )
   6  -.194  .098                      *XXXXI     )
   7  -.241  .098                     X(XXXXI     )
   8   .047  .098                       (    IX    )
   9  -.083  .098                       (  XXI     )
  10  -.194  .098                      *XXXXI     )
  11  -.072  .098                       (  XXI     )
  12   .029  .098                       (    IX    )
  13   .042  .098                       (    IX    )
  14   .081  .098                       (    IXX   )
  15  -.038  .098                       (   XI     )
  16  -.244  .098                     X(XXXXI     )
  17  -.002  .098                       (    I     )
  18  -.054  .098                       (   XI     )
  19   .024  .098                       (    IX    )
  20   .016  .098                       (    I     )
  21  -.095  .098                       (  XXI     )
  22   .046  .098                       (    IX    )
  23  -.029  .098                       (   XI     )
  24  -.037  .098                       (   XI     )
  25   .013  .098                       (    I     )
                                     -2SE       +2SE
```

FIGURE 2.12.2(d) *Diagnosis:* ACF and PACF for the Model Residuals



*Identification.* The ACFs and PACFs for this series are shown in Figures 2.12.2(b), 2.12.2(c), and 2.12.2(d). Deutsch and Alt picked an ARIMA $(0,1,1)\,(0,1,1)_{12}$ model for this series. This is the same model we identified for the Sutter County Workforce time series, so the reader can compare those ACFs and PACFs with these.

The ACF for the raw armed robbery time series indicates nonstationarity, so the series must be differenced. The ACF and PACF of the differenced series do not indicate *seasonal* nonstationarity, however. The ACF and PACF of the regularly and seasonally differenced series do not unambiguously suggest the ARIMA $(0,1,1)\,(0,1,1)_{12}$ model used by Deutsch and Alt. To be perfectly frank, these ACFs and PACFs baffle us. We see no clear patterns of spiking and/or decay which would lead us to accept any tentative model.

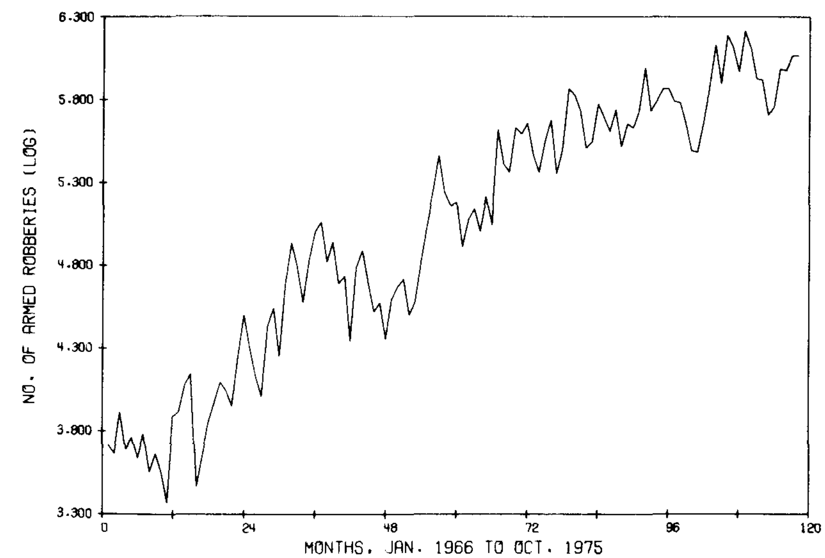



FIGURE 2.12.2(e) Boston Armed Robberies (Logged)

*Estimation*. For the ARIMA (0,1,1) $(0,1,1)_{12}$ model proposed by Deutsch and Alt, we obtain parameter estimates of

$$\hat{\Theta}_0 = .31 \text{ with t statistic } = .47$$
$$\hat{\Theta}_1 = .52 \text{ with t statistic } = 6.08$$
$$\hat{\Theta}_{12} = .73 \text{ with t statistic } = 8.29.$$

As $\hat{\Theta}_0$ is not statistically significant, we drop it from the tentative model. $\hat{\Theta}_1$ and $\hat{\Theta}_{12}$ are both statistically significant and both lie within the bounds of invertibility for moving average parameters. The tentative model is, then,

$$Y_t = \frac{(1 - .52B)(1 - .73B^{12})}{(1 - B)(1 - B^{12})} a_t.$$

*Diagnosis*. The ACF and PACF for the residuals of this model do not inspire confidence. There are statistically significant spikes at lags-4, -7, and -10 of the ACF and -16 of the PACF as well as a handful of "marginally significant" spikes at other lags. The Q statistic for this ACF is also quite large. With 23 degrees of freedom, Q = 41.87, a value of Q associated with a .009 significance level. As the Q statistic is significant, we must conclude that these residuals are not white noise. The tentative model fails our diagnostic criteria and we thus reject it.

*Identification*. Because we have rejected the tentative model, we return to the identification stage. The ACFs and PACFs of the raw and differenced series are not much help here. As noted, we see no evidence for a parsimonious ARIMA model in these statistics. One alternative at this point would be to include extra moving average parameters in the model, that is, to increase the values of q and/or Q. We see no evidence of higher order moving averages in these ACFs and PACFs, however; and, moreover, the model is already rather complicated and cumbersome.

Another alternative is to explore a transformation of the time series. In Section 2.4, we discussed variance stationarity, noting that a time series process must be made stationary in variance as well as in level. Examining the plotted armed robbery time series in Figure 2.12.2(a), we can see that the series variance is roughly proportional to the the series level. In the first half of the series, when the process is at its lowest level, month-to-month fluctuations are small. In the second half of the series, when the process is at its highest level, month-to-month fluctuations are relatively large. In Figure

*(text continued on p. 120)*



```
SERIES.. VAR3      (NOBS= 118)  VAR3      = LOG (BAR       )
NO. OF VALID OBSERVATIONS  =    118.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 25, 118) =  1249.4      SIG =    0.000

LAG    CORR    SE  -1   -.8  -.6  -.4  -.2   0   .2   .4   .6   .8  +1
                   +----+----+----+----+----+----+----+----+----+----+
  1   .946  .092                       (    IXXXX)XXXXXXXXXXXXXXXXXXX
  2   .906  .154                    (       IXXXXXXX)XXXXXXXXXXXXXX
  3   .883  .194                  (         IXXXXXXXXX)XXXXXXXXXXX
  4   .855  .225                 (          IXXXXXXXXXX)XXXXXXXXX
  5   .836  .251               (            IXXXXXXXXXXXX)XXXXXXXX
  6   .808  .274              (             IXXXXXXXXXXXXX)XXXXXX
  7   .782  .293             (              IXXXXXXXXXXXXXX)XXXXX
  8   .761  .311            (               IXXXXXXXXXXXXXXX)XXX
  9   .726  .326            (               IXXXXXXXXXXXXXXX)XX
 10   .695  .339           (                IXXXXXXXXXXXXXXXXX*
 11   .672  .351          (                 IXXXXXXXXXXXXXXXXXX)
 12   .650  .362          (                 IXXXXXXXXXXXXXXXXX )
 13   .616  .372         (                  IXXXXXXXXXXXXXXX   )
 14   .591  .380         (                  IXXXXXXXXXXXXXXX   )
 15   .560  .388         (                  IXXXXXXXXXXXXXX    )
 16   .528  .395        (                   IXXXXXXXXXXXXX      )
 17   .501  .401        (                   IXXXXXXXXXXXXX      )
 18   .479  .406        (                   IXXXXXXXXXXXX       )
 19   .463  .411       (                    IXXXXXXXXXXXX        )
 20   .442  .415       (                    IXXXXXXXXXXX         )
 21   .417  .419       (                    IXXXXXXXXXX          )
 22   .397  .423       (                    IXXXXXXXXXX          )
 23   .377  .426       (                    IXXXXXXXXX           )
 24   .363  .429       (                    IXXXXXXXXX           )
 25   .341  .431      (                     IXXXXXXXXX            )
                   -2SE                                          +2SE


PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

LAG    CORR    SE  -1   -.8  -.6  -.4  -.2   0   .2   .4   .6   .8  +1
                   +----+----+----+----+----+----+----+----+----+----+
  1   .946  .092                       (    IXXXX)XXXXXXXXXXXXXXXXXXX
  2   .107  .092                       (    IXXX )
  3   .153  .092                       (    IXXXX)
  4  -.011  .092                       (    I    )
  5   .099  .092                       (    IXX  )
  6  -.079  .092                       (  XXI    )
  7   .014  .092                       (    I    )
  8   .004  .092                       (    I    )
  9  -.111  .092                       ( XXXI    )
 10  -.021  .092                       (   XI    )
 11   .033  .092                       (    IX   )
 12   .023  .092                       (    IX   )
 13  -.124  .092                       ( XXXI    )
 14   .065  .092                       (    IXX  )
 15  -.094  .092                       (  XXI    )
 16  -.014  .092                       (    I    )
 17  -.011  .092                       (    I    )
 18   .074  .092                       (    IXX  )
 19   .032  .092                       (    IX   )
 20  -.012  .092                       (    I    )
 21  -.020  .092                       (   XI    )
 22   .009  .092                       (    I    )
 23  -.006  .092                       (    I    )
 24   .050  .092                       (    IX   )
 25  -.064  .092                       (  XXI    )
                                     -2SE      +2SE
```

FIGURE 2.12.2(f) ACF and PACF for the Raw (Logged) Series

```
SERIES.. VAR3        (NOBS= 118)  VAR3        = LOG (BAR        )
DIFFERENCED  1 TIME(S) OF ORDER  1.
NO. OF VALID OBSERVATIONS   =   117.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 25, 117) =  28.966        SIG =      .265

 LAG   CORR    SE  -1   -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
  1  -.229  .092                       X(XXXXI    )
  2  -.162  .097                         (XXXXI    )
  3  -.004  .099                         (    I    )
  4  -.078  .099                         (  XXI    )
  5   .102  .100                         (    IXXX )
  6  -.040  .101                         (   XI    )
  7  -.134  .101                         ( XXXI    )
  8   .162  .103                         (    IXXXX)
  9  -.112  .105                         ( XXXI    )
 10  -.120  .106                         ( XXXI    )
 11   .083  .107                         (    IXX  )
 12   .144  .107                         (    IXXXX)
 13  -.029  .109                         (   XI    )
 14   .054  .109                         (    IX   )
 15  -.136  .109                         ( XXXI    )
 16   .023  .111                        (     IX     )
 17   .008  .111                        (     I      )
 18  -.064  .111                        (   XXI      )
 19   .051  .111                        (     IX     )
 20   .025  .111                        (     IX     )
 21  -.116  .111                        (  XXXI      )
 22   .055  .112                        (     IX     )
 23  -.027  .113                        (    XI      )
 24   .060  .113                        (     IX     )
 25   .032  .113                        (     IX     )
                                      -2SE        +2SE
```

```
PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

 LAG   CORR    SE  -1   -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
  1  -.229  .092                     X(XXXXI    )
  2  -.227  .092                     X(XXXXI    )
  3  -.113  .092                      ( XXXI    )
  4  -.166  .092                      (XXXXI    )
  5   .012  .092                      (    I    )
  6  -.063  .092                      (  XXI    )
  7  -.169  .092                      (XXXXI    )
  8   .062  .092                      (    IXX  )
  9  -.122  .092                      ( XXXI    )
 10  -.203  .092                      *XXXXI    )
 11  -.063  .092                      (  XXI    )
 12   .114  .092                      (    IXXX )
 13  -.023  .092                      (   XI    )
 14   .100  .092                      (    IXXX )
 15  -.044  .092                      (   XI    )
 16  -.037  .092                      (   XI    )
 17  -.070  .092                      (  XXI    )
 18  -.061  .092                      (  XXI    )
 19  -.027  .092                      (   XI    )
 20  -.021  .092                      (   XI    )
 21  -.086  .092                      (  XXI    )
 22  -.017  .092                      (    I    )
 23  -.052  .092                      (   XI    )
 24  -.018  .092                      (    I    )
 25  -.028  .092                      (   XI    )
                                    -2SE      +2SE
```

FIGURE 2.12.2(g) ACF and PACF for the Regularly Differenced (Logged) Series



```
SERIES.. RESDLOG  (NOBS= 117)  BOSTON ARMED ROBBERY (LOG) RESIDUALS
NO. OF VALID OBSERVATIONS  =   117.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 23, 117) =  16.315       SIG =      .841

 LAG   CORR    SE  -1   -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
  1   .041  .092                         (    IX   )
  2  -.152  .093                         (XXXXI    )
  3  -.053  .095                         (   XI    )
  4  -.109  .095                         ( XXXI    )
  5   .077  .096                         (    IXX  )
  6  -.046  .097                         (   XI    )
  7  -.149  .097                         (XXXXI    )
  8   .094  .099                         (    IXX  )
  9  -.099  .099                         (  XXI    )
 10  -.099  .100                         (  XXI    )
 11   .097  .101                         (    IXX  )
 12  -.012  .102                         (    I    )
 13   .021  .102                         (    IX   )
 14   .040  .102                         (    IX   )
 15  -.098  .102                         (  XXI    )
 16  -.001  .103                         (    I    )
 17  -.031  .103                         (   XI    )
 18  -.040  .103                         (   XI    )
 19   .043  .103                         (    IX   )
 20  -.029  .103                         (   XI    )
 21  -.066  .103                         (  XXI    )
 22   .032  .104                         (    IX   )
 23  -.017  .104                         (    I    )
 24   .060  .104                         (    IXX  )
 25   .073  .104                         (    IXX  )
                                       -2SE      +2SE
```

```
PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

 LAG   CORR    SE  -1   -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                    +----+----+----+----+----+----+----+----+----+----+
  1   .041  .092                         (    IX   )
  2  -.154  .092                         (XXXXI    )
  3  -.040  .092                         (   XI    )
  4  -.132  .092                         ( XXXI    )
  5   .076  .092                         (    IXX  )
  6  -.098  .092                         (  XXI    )
  7  -.134  .092                         ( XXXI    )
  8   .080  .092                         (    IXX  )
  9  -.154  .092                         (XXXXI    )
 10  -.102  .092                         ( XXXI    )
 11   .053  .092                         (    IX   )
 12  -.036  .092                         (   XI    )
 13  -.029  .092                         (   XI    )
 14   .026  .092                         (    IX   )
 15  -.075  .092                         (  XXI    )
 16  -.062  .092                         (  XXI    )
 17  -.053  .092                         (   XI    )
 18  -.034  .092                         (   XI    )
 19  -.045  .092                         (   XI    )
 20  -.036  .092                         (   XI    )
 21  -.079  .092                         (  XXI    )
 22  -.028  .092                         (   XI    )
 23  -.043  .092                         (   XI    )
 24   .016  .092                         (    I    )
 25   .016  .092                         (    I    )
                                       -2SE      +2SE
```

FIGURE 2.12.2(h) *Diagnosis:* ACF and PACF for the Model Residuals (Logged)

2.12.2(e), we show the log-transformed series. In the natural log metric, month-to-month fluctuations are more or less the same size throughout the length of the series. We can now try to identify a simple ARIMA model for the log series.

The ACF and PACF for the raw log-transformed series—Figure 2.12.2(f)—indicate nonstationarity. The ACF and PACF for the first-differenced log series—(Figure 2.12.2(g)—indicate that one regular difference is sufficient. The absence of spikes at the seasonal lags indicates that seasonal differencing is not required. There is a rough pattern of decay beginning at lag-1 of the PACF and a single spike at lag-1 of the ACF which suggest a ARIMA (0,1,1) model for this series. Spiking and decay at lag-12 of the ACF and PACF lie within the confidence bands and, thus, are not statistically significant. Nevertheless, we see seasonal variation in the plotted series. To account for this seasonality, we propose an ARIMA (0,1,1) $(0,0,1)_{12}$ model. If $\hat{\Theta}_{12}$ proves not to be statistically significant, we can drop it from our model at the estimation stage.

*Estimation*. Our tentative model is ARIMA (0,1,1) $(0,0,1)_{12}$ for the log-transformed series. We write this as

$$\mathrm{Ln}(Y_t) = \frac{\Theta_0 + (1 - \Theta_1 B)(1 - \Theta_{12} B^{12})}{1 - B} a_t.$$

The parameters of this model are estimated as

$$\hat{\Theta}_0 = .0195 \text{ with t statistic} = 1.57$$
$$\hat{\Theta}_1 = .4321 \text{ with t statistic} = 4.99$$
$$\hat{\Theta}_{12} = .1884 \text{ with t statistic} = -1.97.$$

As the estimate of $\Theta_0$ is not statistically significant, this parameter is dropped from the tentative model. The estimates of $\Theta_1$ and $\Theta_{12}$ are both statistically significant (though $\hat{\Theta}_{12}$, just barely) and both lie within the bounds of invertibility for moving average parameters.

*Diagnosis*. The ACF and PACF for the model residuals—Figure 2.12.2(h)—have no spikes at lag-1 or at the seasonal lags. With 23 degrees of freedom, $Q = 16.315$, a value of the Q statistic associated with a .84 significance level. As the residuals from our tentative model appear to be white noise, we can accept the model.

Our decision to include a $\Theta_{12}$ parameter in the tentative model despite the lack of evidence for seasonality in the ACFs and PACFs might be criticized.



Our decision was based primarily on the seasonal appearance of the time series. In this case, however, had we started with a ARIMA (0,1,1) model, metadiagnosis would have lead to the ARIMA (0,1,1) $(0,0,1)_{12}$ model. Metadiagnosis, particularly overfitting the tentative model, is crucial when a seasonal effect is possible. While including a seasonal component in the ARIMA model when no seasonality is present is an error, failing to include a seasonal component when one is present is a more dangerous error.

The value of inspecting model residuals as a metadiagnostic step is apparent in this example analysis. It would be desirable to have a simple statistical test for variance nonstationarity. Unfortunately, as Granger and Newbold note, "No completely satisfactory techniques are available for testing whether or not a series contains a trend in mean and/or variance. A number of sensible procedures can be suggested, but a decision based on the plot of the data is likely to be a reasonable one." (1977: 37).

To test for a variance nonstationary process in this time series, Hay and McCleary (1979) devided the series into equal interval segments and noted that the mean and standard deviation of each segment showed a nearly monotonic increase over time. Two tests for variance homogeneity, Cochran's C and the Bartlett-Box F tests, also were applied to the segments. Yet none of these statistical tests is as compelling as the visual evidence. In Figure 2.12.2(i), we show a plot of the estimated residuals of the ARIMA (0,1,1) $(0,1,1)_{12}$ model for this series. The variance of the residuals increases as a function of time, clearly indicating variance nonstationarity.

Our point is this example is that, unless an ARIMA model is built through the iterative identification/estimation/diagnosis strategy that we recommend, an arbitrary ARIMA model may result. In the next chapter, we will return to this time series to demonstrate how serious this error can be. In April 1975, the Massachusetts legislature passed a strict gun control law which (presumably) would have an impact on the level of this time series. On the basis of an arbitrary and statistically inadequate ARIMA model, Deutsch and Alt concluded that the law had an abrupt and profound impact on this series. We will demonstrate in the next chapter that the evidence does not support this conclusion.

### 2.12.3 Swedish Harvest Index

When is a time series not a time series? In Figure 2.12.3(a), we show the annual Swedish Harvest Index for the period 1749–1850 as reported by Thomas (1940). These data come close to being a "time series that is not a time series." In each year, the Swedish grain harvest was rated on a nine-

*(text continued on p. 124)*

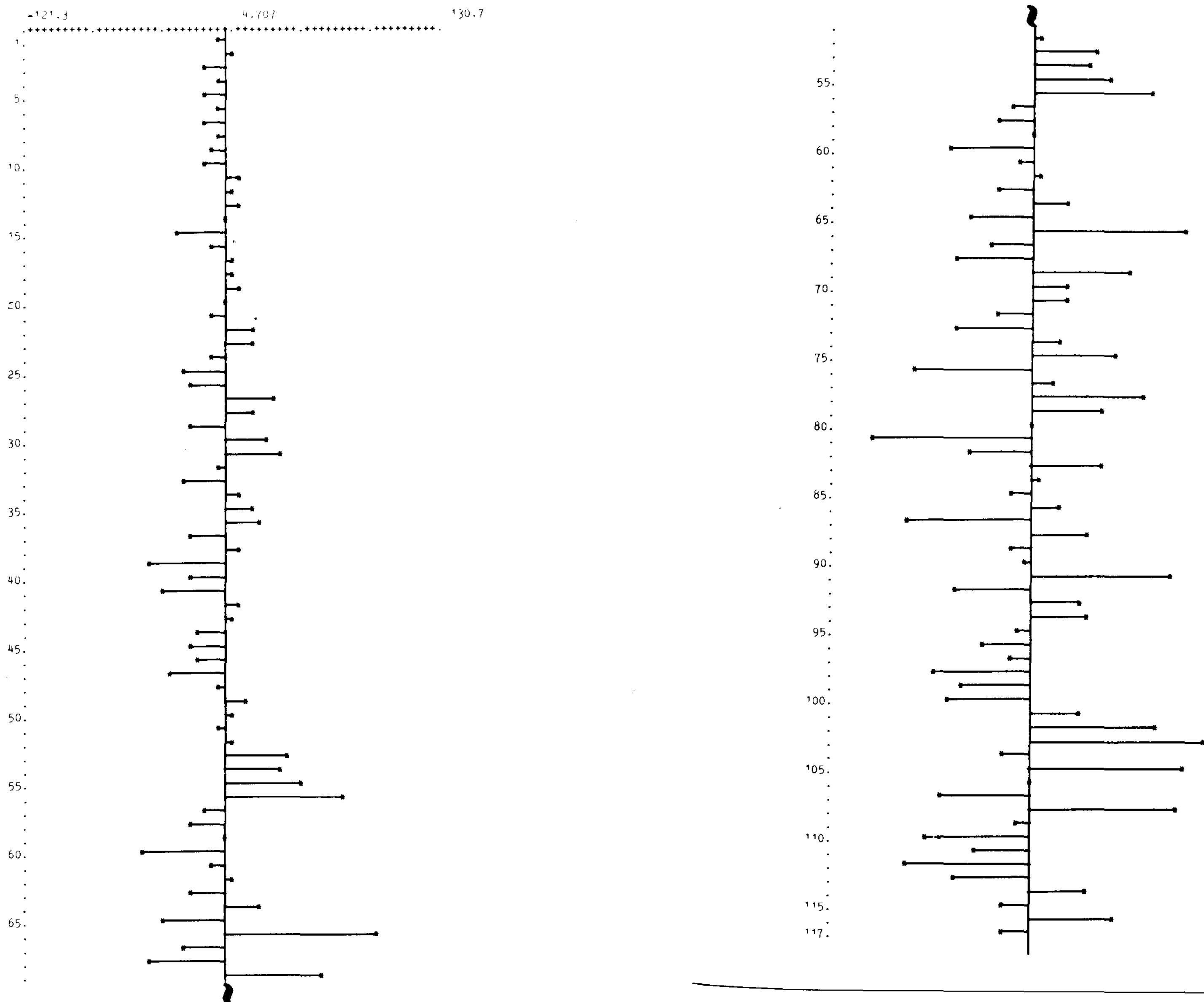



FIGURE 2.12.2(i) The Residuals of the Unlogged Model Indicate that the Generating Process is Nonstationary in Variance

point scale with a total crop failure scored as zero and a superabundant crop scored as nine. The problem with this time series is its level of measurement. While we require a time series process to be measured at the interval level, this one is measured at the ordinal level. In only 13 of the 102 years does the index take on a noninteger value. Because this series does not have a real interval level variance (and covariance), we are not optimistic about building a good ARIMA model for it.

*Identification*. The ACF of the raw time series indicates that the process is stationary. A single spike at lag-1 of the ACF and rough decay beginning at lag-1 of the PACF suggest an ARIMA (0,0,1) model. We write this tentative model as

$$Y_t = \Theta_0 + (1 - \Theta_1 B)a_t,$$

where the parameter $\Theta_0$ is interpreted as the level of the stationary series.

*Estimation*. Parameter estimates for the tentative model are:

$$\hat{\Theta}_0 = \quad 5.21 \text{ with t statistic} = 15.19$$

$$\hat{\Theta}_{11} = -\ .39 \text{ with t statistic} = -4.20.$$

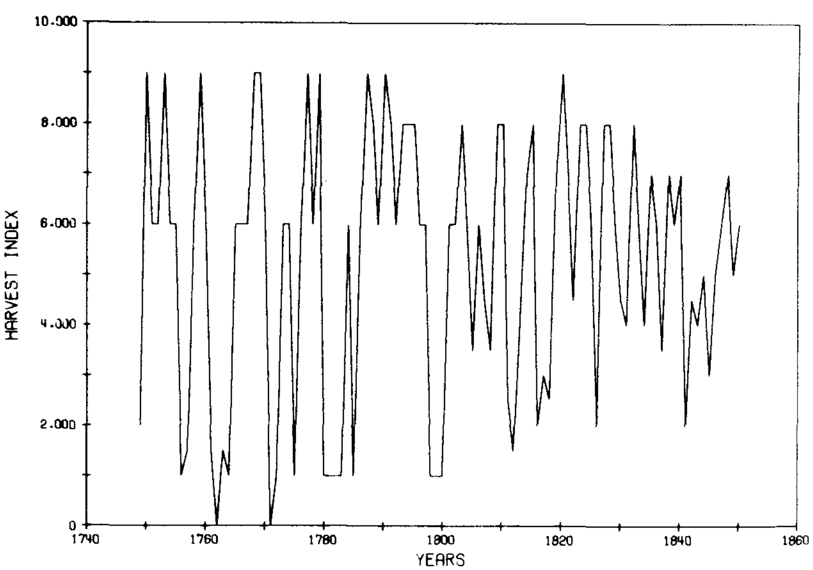



FIGURE 2.12.3(a) Swedish Harvest Index



```
SERIES.. HARVEST  (NOBS= 102)  SWEDISH HARVEST INDEX 1749-1860
NO. OF VALID OBSERVATIONS  =   102.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 25, 102) =  36.048      SIG =      .071

  LAG   CORR    SE -1   -.8  -.6  -.4  -.2   0   .2   .4   .6   .8  +1
                   +----+----+----+----+----+----+----+----+----+----+
    1   .340  .099                     (    IXXXX)XXX
    2  -.044  .110                     (   XI    )
    3  -.139  .110                    (  XXXI     )
    4  -.097  .112                    (   XXI     )
    5  -.065  .113                    (   XXI     )
    6  -.161  .113                    ( XXXXI     )
    7  -.095  .115                    (   XXI     )
    8   .029  .116                    (     IX    )
    9   .026  .116                    (     IX    )
   10   .018  .116                    (     I     )
   11  -.062  .116                    (   XXI     )
   12  -.170  .116                    ( XXXXI     )
   13  -.050  .119                    (    XI     )
   14   .042  .119                    (     IX    )
   15   .015  .119                    (     I     )
   16  -.062  .119                    (   XXI     )
   17   .040  .120                    (     IX    )
   18   .199  .120                    (     IXXXXX)
   19   .171  .123                    (     IXXXX )
   20  -.007  .125                    (     I     )
   21  -.167  .125                    ( XXXXI     )
   22  -.155  .127                    ( XXXXI     )
   23  -.047  .129                    (    XI     )
   24   .008  .129                    (     I     )
   25   .023  .129                    (     IX    )
                                    -2SE        +2SE


PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

  LAG   CORR    SE -1   -.8  -.6  -.4  -.2   0   .2   .4   .6   .8  +1
                   +----+----+----+----+----+----+----+----+----+----+
    1   .340  .099                     (    IXXXX)XXX
    2  -.181  .099                     *XXXXI    )
    3  -.070  .099                     (  XXI    )
    4  -.030  .099                     (   XI    )
    5  -.050  .099                     (   XI    )
    6  -.169  .099                     (XXXXI    )
    7   .000  .099                     (    I    )
    8   .030  .099                     (    IX   )
    9  -.057  .099                     (   XI    )
   10   .003  .099                     (    I    )
   11  -.089  .099                     (  XXI    )
   12  -.175  .099                     (XXXXI    )
   13   .046  .099                     (    IX   )
   14   .011  .099                     (    I    )
   15  -.075  .099                     (  XXI    )
   16  -.083  .099                     (  XXI    )
   17   .093  .099                     (    IXX  )
   18   .102  .099                     (    IXXX )
   19   .041  .099                     (    IX   )
   20  -.037  .099                     (   XI    )
   21  -.133  .099                     ( XXXI    )
   22  -.067  .099                     (  XXI    )
   23   .019  .099                     (    I    )
   24   .002  .099                     (    I    )
   25   .014  .099                     (    I    )
                                     -2SE      +2SE
```

FIGURE 2.12.3(b) ACF and PACF for the Raw Series

```
SERIES.. RESIDUAL (NOBS= 102)  SWEDISH HARVEST INDEX RESIDUALS
NO. OF VALID OBSERVATIONS  =   102.

AUTOCORRELATIONS OF LAGS 1 - 25.
Q( 23, 102) =  15.058      SIG =      .893

  LAG   CORR    SE -1   -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                     +----+----+----+----+----+----+----+----+----+----+
    1   .010  .099                       (    I    )
    2  -.009  .099                       (    I    )
    3  -.117  .099                       ( XXXI    )
    4  -.058  .100                       (   XI    )
    5   .001  .101                       (    I    )
    6  -.141  .101                       (XXXXI    )
    7  -.065  .103                       (  XXI    )
    8   .050  .103                       (    IX   )
    9   .002  .103                       (    I    )
   10   .020  .103                       (    I    )
   11  -.013  .103                       (    I    )
   12  -.164  .103                       (XXXXI    )
   13  -.005  .106                       (    I    )
   14   .033  .106                       (    IX   )
   15   .030  .106                       (    IX   )
   16  -.076  .106                       (  XXI    )
   17   .013  .107                       (    I    )
   18   .154  .107                       (    IXXXX)
   19   .123  .109                       (    IXXX )
   20  -.005  .110                      (     I     )
   21  -.131  .110                      (  XXXI     )
   22  -.107  .112                      (  XXXI     )
   23  -.015  .113                      (     I     )
   24   .017  .113                      (     I     )
   25  -.007  .113                      (     I     )
                                      -2SE        +2SE


PARTIAL AUTOCORRELATIONS OF LAGS 1 - 25.

  LAG   CORR    SE -1   -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                     +----+----+----+----+----+----+----+----+----+----+
    1   .010  .099                       (    I    )
    2  -.009  .099                       (    I    )
    3  -.117  .099                       ( XXXI    )
    4  -.057  .099                       (   XI    )
    5  -.000  .099                       (    I    )
    6  -.159  .099                       (XXXXI    )
    7  -.081  .099                       (  XXI    )
    8   .044  .099                       (    IX   )
    9  -.039  .099                       (   XI    )
   10  -.017  .099                       (    I    )
   11  -.010  .099                       (    I    )
   12  -.196  .099                       *XXXXI    )
   13  -.035  .099                       (   XI    )
   14   .036  .099                       (    IX   )
   15  -.024  .099                       (   XI    )
   16  -.118  .099                       ( XXXI    )
   17   .017  .099                       (    I    )
   18   .108  .099                       (    IXXX )
   19   .077  .099                       (    IXX  )
   20   .016  .099                       (    I    )
   21  -.104  .099                       ( XXXI    )
   22  -.113  .099                       ( XXXI    )
   23  -.012  .099                       (    I    )
   24   .017  .099                       (    I    )
   25  -.014  .099                       (    I    )
                                       -2SE      +2SE
```

FIGURE 2.12.3(c) *Diagnosis:* ACF and PACF for the Model Residuals



Both parameter estimates are statistically significant. The estimate of $\Theta_1$ is well within the bounds of invertibility for moving average parameters, so the tentative model poses no problems at this stage.

*Diagnosis*. The residual ACF has no spikes at early lags. Moreover, the Q statistic for this ACF is not significant. As these residuals satisfy our diagnostic criteria, we infer that they are white noise and accept the model.

The model for the Swedish Harvest Index time series is:

$$Y_t = 5.21 + a_t + .39a_{t-1}.$$

Due to the form of these data, however, we should not immediately conclude that this model is of the highest quality. While the model is clearly adequate in the sense that its residuals are white noise, the model may not necessarily be of much use to us. In general, the utility of a model depends upon its predictive ability. How well does it fit the data? To answer this question, we note that the $R^2$ statistic is:

$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}} = 1 - \frac{604.47}{3489.25}$$

$$= .83$$

We interpret the $R^2$ statistic to mean that 83% of the variance in the time series is explained by the model. This is a respectable figure. However, the largest portion of the explained variance is due to the mean. We see that the total sum of squares can be divided into three parts:

| | |
|---|---|
| residual sum of squares | 604.47 |
| sum of squares due to $\Theta_0$ | 2758.89 |
| sum of squares due to $\Theta_1$ | 125.89 |
| total sum of squares | 3489.25 |

This partition of the total sum of squares is analogous to an analysis of variance partition. We see that the moving average accounts for less than 4% of the variance in the time series. Though this percentage is statistically significant and not at all trivial, it is small compared to the variance explained by the model mean.

Our point in this example analysis is that, sometimes, data that appear to be a time series do not really constitute a time series. This is nearly true of the Harvest Index time series. When data do not constitute a time series, of course, an ARIMA model fit to the data might lead to incorrect conclusions. We will return to this time series in Chapter 5.

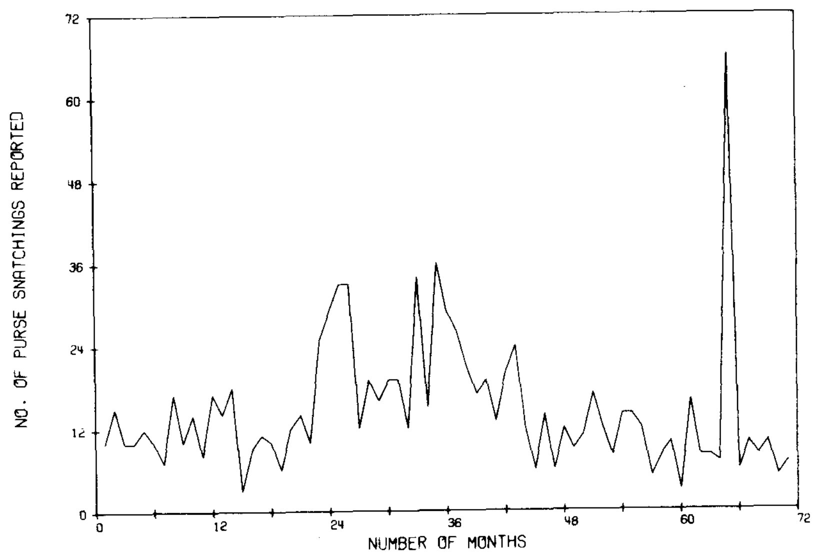



FIGURE 2.12.4(a) Hyde Park Purse Snatchings

## 2.12.4 Hyde Park Purse Snatchings: Outliers

In Figure 2.12.4(a), we show a time series of purse snatchings reported 13 times per year (every 28 days) in the Hyde Park neighborhood of Chicago from January 1969 to September 1973. These data were collected by Reed (1978) for an evaluation of Operation Whistlestop, a community crime prevention program. Our attention is immediately drawn to the 65th observation of the series which is approximately five times larger than adjacent observations. This is an outlier. In an analysis of these data, Reed arrived at an ambiguous conclusion and suspected that the problem was due to this outlier. Since there was no apparent explanation for this single extreme observation, Reed concluded that it was due to an error in recording the data. Returning to the primary data collection sheets, Reed was able to arrive at a more reasonable, correct number for the 65th observation.

Outliers are a somewhat obscure topic in the literature of time series statistical analysis. While it appears that extreme values may have a distorting effect on the identification and estimation of ARIMA models (we will demonstrate the size of this distortion shortly), there is little consensus on



```
SERIES.. HPPS        (NOBS=  71)  HYDE PARK PURSE SNATCHINGS (W/OUTLIER)
NO. OF VALID OBSERVATIONS  =     71.

AUTOCORRELATIONS OF LAGS 1 - . .
Q( 15,   71) =  12.359       S:  :        .652

 LAG   CORR    SE   -1    -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                     +----+----+----+----+----+----+----+----+----+----+
  1    .166  .119                        (      IXXXX )
  2    .230  .122                        (      IXXXXX*
  3    .111  .128                        (      IXXX  )
  4    .150  .129                        (      IXXXX )
  5   -.017  .132                       (       I       )
  6   -.004  .132                       (       I       )
  7    .089  .132                       (       IXX     )
  8    .028  .133                       (       IX      )
  9    .082  .133                       (       IXX     )
 10    .136  .133                       (       IXXX    )
 11    .112  .135                       (       IXXX    )
 12    .075  .137                       (       IXX     )
 13    .060  .137                       (       IX      )
 14   -.001  .138                       (       I       )
 15   -.053  .138                       (      XI       )
                                       -2SE           +2SE

PARTIAL AUTOCORRELATIONS OF LAGS 1 - 15.

 LAG   CORR    SE   -1    -.8  -.6  -.4  -.2   0    .2   .4   .6   .8  +1
                     +----+----+----+----+----+----+----+----+----+----+
  1    .166  .119                        (      IXXXX )
  2    .209  .119                        (      IXXXXX)
  3    .049  .119                        (      IX    )
  4    .088  .119                        (      IXX   )
  5   -.086  .119                        (    XXI     )
  6   -.049  .119                        (     XI     )
  7    .108  .119                        (      IXXX  )
  8    .012  .119                        (      I     )
  9    .064  .119                        (      IXX   )
 10    .114  .119                        (      IXXX  )
 11    .029  .119                        (      IX    )
 12    .008  .119                        (      I     )
 13   -.003  .119                        (      I     )
 14   -.068  .119                        (    XXI     )
 15   -.065  .119                        (    XXI     )
                                       -2SE        +2SE
```

FIGURE 2.12.4(b) ACF and PACF for the Raw Series (Outlier Included)

just what characteristics define an outlier. How (relatively) large must an observation be before it is called an outlier?

It is important to note here that outliers are *sample* phenomena, not *population* phenomena. In cross-sectional analyses, for example, it would be unusual to find a seven-foot-tall person in a sample of five people. This person would be considered an outlier. If the sample size was increased, however, there would come a point at which a seven-foot-tall person would no longer be considered an outlier. This principle holds for longitudinal analyses as well. The deviant 65th observation of the Hyde Park time series

```
SERIES.. HPPS2     (NOBS=  71)  HYDE PARK PURSE SNATCHINGS (WO/OUTLIER)
NO. OF VALID OBSERVATIONS  =    71.

AUTOCORRELATIONS OF LAGS 1 - ...
Q( 15,   71) =   81.002      SIG             .000

 LAG    CORR     SE   -1    -.8   -.6   -.4   -.2    0     .2    .4    .6    .8   +1
                       +----+----+----+----+----+----+----+----+----+----+
  1     .493   .119                         (       IXXXXX)XXXXXX
  2     .534   .145                        (        IXXXXXX)XXXXXX
  3     .363   .170                      (          IXXXXXXXX*
  4     .294   .181                      (          IXXXXXXX )
  5     .261   .167                      (          IXXXXXXX )
  6     .163   .192                     (           IXXXX       )
  7     .243   .194                     (           IXXXXXX     )
  8     .183   .199                     (           IXXXXX      )
  9     .179   .201                     (           IXXXX       )
 10     .243   .203                     (           IXXXXXX     )
 11     .204   .207                     (           IXXXXX      )
 12     .227   .210                    (            IXXXXXX      )
 13     .147   .214                    (            IXXXX        )
 14    -.022   .215                    (           XI            )
 15    -.023   .215                    (           XI            )
                                     -2SE                      +2SE

PARTIAL AUTOCORRELATIONS OF LAGS 1 - 15.

 LAG    CORR     SE   -1    -.8   -.6   -.4   -.2    0     .2    .4    .6    .8   +1
                       +----+----+----+----+----+----+----+----+----+----+
  1     .493   .119                         (       IXXXXX)XXXXXX
  2     .385   .119                         (       IXXXXX)XXXX
  3     .018   .119                         (       I     )
  4    -.038   .119                         (      XI     )
  5     .053   .119                         (       IX    )
  6    -.048   .119                         (      XI     )
  7     .141   .119                         (       IXXXX )
  8     .044   .119                         (       IX    )
  9    -.026   .119                         (      XI     )
 10     .136   .119                         (       IXXX  )
 11     .036   .119                         (       IX    )
 12     .001   .119                         (       I     )
 13    -.056   .119                         (      XI     )
 14    -.296   .119                        X(XXXXXI       )
 15    -.067   .119                         (     XXI     )
                                          -2SE         +2SE
```

FIGURE 2.12.4(c) ACF and PACF for the Raw Series (Outlier Excluded)

is considered an outlier in a time series of 71 observations. If the length of this series was increased, however, there would come a point at which the 65th observation would no longer be considered an outlier. The question, of course, is how long must the series be before an outlier can be accommodated by an ARIMA model?

We will answer this question indirectly. Although detection of the outlier is obvious in this example, it will be of some value to analyze this series both with and without the deviant 65th observation. By comparing the two analyses, we can make a statement about the effect that this outlier has on the identification and estimation of an ARIMA model.



*Identification.* Figure 2.12.4(b) shows the ACF and PACF for the series including the deviant 65th observation. There is no identifiable pattern in either the ACF or PACF, so we must conclude that this series is white noise. This judgment is confirmed by the Q statistic ($Q = 12.36$ with 15 degrees of freedom) which is significant only at the .65 level. In Figure 2.12.4(c), however, we show the ACF and the PACF for the series excluding the deviant 65th observation. An ARIMA (2,0,0) process is now indicated. The ACF decays and the PACF has spikes at the first two lags. The presence of only one outlier in a series of 71 observations has severely distorted the ACF, leading to an incorrect identification of an ARIMA (0,0,0) model for this series.

A single outlier can have such a profound effect because of the nature of the *estimated* ACF. In estimating an ACF (k), deviated time series observations are weighted by their absolute distance to the series mean. Outlying observations thus exert a profound effect on the estimated ACF (k). We can see this effect more clearly by comparing the components of the lag-1 ACF estimated with and without the outlier:

| | Covariance | Variance | ACF (1) |
|---|---|---|---|
| Estimated with the correct observation | 27.47 | 55.71 | .49 |
| Estimated with the outlier included | 15.49 | 93.29 | .17 |

Examining these figures, we note that removing the outlier results in a 44% *in*crease in covariance and a 67% *de*crease in variance. Since the estimated value of ACF (1) is the ratio of covariance to variance, it is not surprising that a decrease in numerator and an increase in denominator seriously underestimates ACF (1). This will generally be true: A single outlier in a short time series will result in a biased underestimation of low lags of the ACF. Because this biased estimate of ACF (1) is not statistically significant, we incorrectly identified an ARIMA (0,0,0) model rather than the more appropriate ARIMA (2,0,0) model.

*Estimation.* For the ARIMA (2,0,0) model, our parameter estimates are:

$$\hat{\phi}_1 = .31 \text{ with t statistic} = 2.67$$
$$\hat{\phi}_2 = .40 \text{ with t statistic} = 3.45.$$

Both estimates are statistically significant and otherwise acceptable.

It is difficult in practice to untangle the distortions in model estimation introduced by outliers. In this time series, for example, the inappropriate ARIMA (0,0,0) model (with the outlier included) has RMS = 94.62. Fitting the more appropriate ARIMA (2,0,0) model to the series (with the outlier included) results in RMS = 92.82; and both autoregressive parameter estimates are statistically insignificant at a .05 level. The ARIMA (2,0,0) model fit to the series with the outlier excluded results in RMS = 37.75. Much of this difference in the RMS statistics is due only to the outlier.

*Diagnosis*. The residuals from the ARIMA (2,0,0) model without the outlier appear to be white noise. There are no statistically significant spikes at low lags of the ACF, and Q = 17.8 with 22 degrees of freedom is statistically significant only at the .71 level. We accept this tentative model, then, for the time series with the corrected value for the 65th observation.

We will have more to say about this time series and about outliers generally in the next chapter. Since univariate ARIMA models are always built with some purpose in mind (such as impact assessment, forecasting, or multivariate analyses), outliers can confound an analysis in more complicated ways than those we have discussed. In impact assessment, for example, the location of outliers relative to the point of intervention may directly bias the estimate of impact. This issue will be discussed in detail in the next chapter.

## 2.13 Conclusion

We conclude this chapter with an overview of ARIMA models and modeling. It is often instructive to see an ARIMA model as a series of linear filters as shown in Figure 2.13(a). In general, a linear filter expresses an output time series as a function of an input white noise process. For any ARIMA model, we start with a random shock input drawn from a Normal (Gaussian) distribution with a zero mean and constant variance. As indicated in this figure (and as explained in Sections 2.2, 2.5, and 2.6), a moving average filter, an autoregressive filter, and/or a nonstationary summation filter is applied to the input to produce an output. Autocorrelation and stochastic behavior in the output time series is determined by the $\phi$ and $\Theta$ parameter values of the filter.

ARIMA model-building procedures, of course, can be represented as inverse filtering. We start with an observed time series and, through the empirical model-building strategy, determine the likely filters and parameter values which will produce an output series of white noise. This process is illustrated in Figure 2.13(b). Note that in both of these diagrams, the input to



$a_t$ ----→ [Moving Average Filter] → [Autoregressive Filter] → [Integration Filter] ----→ $Y_t$

FIGURE 2.13(a) An Input-Output Representation of the ARIMA Model

$Y_t$ ----→ [Differencing Filter] → [Autoregressive Filter] → [Moving Average Filter] ----→ $a_t$

FIGURE 2.13(b) An Input-Output Representation of the ARIMA Model-Building Strategy

both the autoregressive and moving average filters is a stationary series. Application of the model to nonstationary time series is achieved through use of the summation and differencing filters. As we explained in Section 2.2, each of these filters is the inverse of the other. In Figure 2.13(a), the summation filter sums (or integrates) a stationary input series to produce a nonstationary output series. In Figure 2.13(b), the differencing filter differences a nonstationary input to produce a stationary output (which is then passed through autoregressive and/or moving average filters).

An attractive statistical property of ARIMA models is that they can be "run" in both directions. Given an observed time series, we "run the model forward" as in Figure 2.13(b) and reduce the $Y_t$ input to white noise. This requires that statistically adequate parameters be estimated for an appropriate and parsimonious filter structure. Having determined these parameters, we then "run the model backward" as in Figure 2.13(a) to produce predicted values of the $Y_t$ series.

At this point, we suggest that the reader pause to review the material developed in this chapter. We first discussed the statistical properties of ARIMA models: trend, drift, integrated processes, stationarity, variance, autoregressive processes, moving average processes, the expected ACF and PACF, and seasonality. Algebraic operators and formulae were presented to facilitate expression and manipulation of the models. We then developed an iterative model-building strategy wherein the analyst applies these concepts to the problem of constructing a model for an observed time series. Finally, the model-building strategy was applied to the analysis of four time series typical of those encountered in social science research.

We suspect that the reader may feel somewhat frustrated at this point. So far, we have expended a great deal of effort explaining how to build a time series model without explaining how to *use* the model. A univariate ARIMA model in and of itself is admittedly of little interest or utility. This necessary but unfortunate state of affairs will now be remedied. In Chapter 3, we will combine univariate ARIMA models with intervention components to build a variety of models for social impact assessment. In Chapter 4, we will employ univariate ARIMA models to forecast future values of a time series. And in Chapter 5, we will use several univariate ARIMA models to create multivariate time series models of social phenomena.

## For Further Reading

There are several treatments of univariate ARIMA modeling which develop this same material at various levels of sophistication. McCain and McCleary (1979), for example, require only an introductory course in statistics while Box and Jenkins (1976: Chapters 6–8) require a solid mathematical background. Intermediate level treatments are given by Granger and Newbold (1977: Chapter 3), Pindyck and Rubinfeld (1976: Chapters 14–15), and Nelson (1973: Chapter 5).

## NOTES TO CHAPTER 2

1. The symbol "~" means "is distributed as." The meanings of other symbols and conventions are given in a glossary appendix.

2. See Feller (1968: Chapter III) for an illuminating discussion of drift in the random walk. The reader who wishes to study stochastic processes generally is directed to this work and to Feller (1971). While Feller's development lacks nothing in mathematical rigor, he is clearly the most understandable and readable authority on stochastic processes.

3. We use this example only to illustrate the random walk process. In Chapter 5, we build a population growth model, but of course that model is much more complicated than this one. The reader who plans to do a time series analysis of population statistics would do well to first read Keyfitz (1977: Chapter 1).

4. The two equations

$$Y_t = Y_0 + \Theta_0 t$$

and

$$Y_t - Y_{t-1} = \Theta_0$$

are related in a rather straightforward manner. The first equation is the unique solution of the second. We require no background in difference equations for this volume. However, the



interested reader is directed to Goldberg (1958) for an introduction to difference equations written especially for social scientists.

5. Our discussion of stationarity is necessarily a conceptual discussion. A process that is stationary in the *widest* sense is one in which both the process variance, $\text{VAR}(y_t)$, and the process covariance, $\text{COV}(y_t y_{t+k})$, are independent of t. Such a process is fully described by its variance and covariance. For a more precise definition of widest sense stationarity, see Dhrymes (1974: 385) or Malinvaud (1970: 418–419). Our discussion of transformations is similarly conceptual. See Box and Cox (1964) for a general discussion of transformations.

6. The bounds of stationarity for an ARIMA (p,0,0) process are determined by the roots of the characteristic equation

$$1 + \phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p = 0.$$

If the process is nonstationary, then all roots must be greater than unity in absolute value. Thus, for an ARIMA (1,0,0) model

$$1 + \phi_1 B = 0$$
$$\phi_1 B = -1$$
$$B = -(1/\phi_1).$$

This root will be greater than unity in absolute value only when $\phi_1$ is less than unity in absolute value. Similarly, for an ARIMA (2,0,0) model

$$1 + \phi_1 B + \phi_2 B^2 = 0.$$

The roots of this characteristic equation are given by the formula

$$B = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 - 4\phi_2}}{2\phi_2}$$

These roots will be greater than unity in absolute value only when the bounds of stationarity are satisfied.

7. In the general case, an ARIMA (p,0,0) process is:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + a_t.$$

Multiplying the process by $y_{t-k}$ gives us

$$y_t y_{t-k} = \phi_1 y_{t-1} y_{t-k} + \phi_2 y_{t-2} y_{t-k} + \ldots + \phi_p y_{t-p} y_{t-k} + a_t y_{t-k}.$$

Then taking the expectation of this process and dividing by $\sigma_y^2$, we obtain the expected value of ACF (k):

$$\text{ACF}(k) = \phi_1 \text{ACF}(k-1) + \phi_2 \text{ACF}(k-2) + \ldots + \phi_p \text{ACF}(k-p).$$

The reader may use this general expression to derive the expected ACF of higher order ARIMA (p,0,0) processes.

8. The YuleWalker equation system is:

$$\begin{bmatrix} \text{ACF}(0) & \text{ACF}(1) & \ldots & \text{ACF}(k-1) \\ \text{ACF}(1) & \text{ACF}(2) & \ldots & \text{ACF}(k-2) \\ \cdot & \cdot & \ddots & \\ \cdot & \cdot & \ldots & \\ \cdot & \cdot & \ldots & \\ \text{ACF}(k-2) & \text{ACF}(k-3) & \ldots & \text{ACF}(1) \\ \text{ACF}(k-1) & \text{ACF}(k-2) & \ldots & \text{ACF}(0) \end{bmatrix} \times \begin{bmatrix} \text{PACF}(1) \\ \text{PACF}(2) \\ \cdot \\ \cdot \\ \cdot \\ \text{PACF}(k-1) \\ \text{PACF}(k) \end{bmatrix} = \begin{bmatrix} \text{ACF}(1) \\ \text{ACF}(2) \\ \cdot \\ \cdot \\ \cdot \\ \text{ACF}(k-1) \\ \text{ACF}(k) \end{bmatrix}$$

Cramer's Rule can be applied to this k-equation system to obtain solutions for the k unknown values of the PACF. Of course, this assumes that the true values of ACF (1), ACF (2), . . . , ACF (k) are available. Box and Jenkins (1976: 82–84) present a recursive method for calculating PACF (k) which is attributed to Durbin (1960). In practice, a time series computer program routinely estimates the PACF. Our concern here is largely with the *interpretation* of an empirical PACF in the context of model building.

9. Cf. Note 6 above. If

$$(1 - \phi_1 B)(1 - \phi_S B^S) = 0,$$

then either $(1 - \phi_1 B) = 0$

or $(1 - \phi_S B^S) = 0.$

This implies that the two factors will have identical bounds of stationarity-invertibility.

10. Nonlinear estimation is covered as a separate topic in Chapter 6. We have two motives for relegating this topic to the last chapter of the volume. First, many readers will have no interest in the topic. Indeed, the analyst can perform and competently interpret the results of an ARIMA analysis without ever knowing the mechanical details of parameter estimation. Second, nonlinear estimation requires a slightly higher degree of mathematical sophistication of the reader. A superficial knowledge of calculus is assumed, for example. The topic of nonlinear estimation (and Chapter 6) in this sense is not consistent with the more general topics of time series analysis (and the other chapters). In any event, the details of estimation will be better understood after the reader has absorbed the basics of ARIMA modeling.

11. Ljung and Box (1976) have proposed a modification of the Q statistic which increases its value slightly and thus makes it a slightly more conservative test of the hypothesis that the ACF is not different from white noise. We have used the original formula for the Q statistic (Box and Jenkins, 1976: Chapter 8.2) throughout this volume.

12. Model parameter estimates presented in this volume were obtained without backcasting initialization. See Chapter 4, Note 2 and Chapter 6, Notes 1 and 2 for a description of this method and the underlying issues.

## Appendix to Chapter 2: Expected Values

The reader who has a working knowledge of calculus is directed to Feller (1966: Chapter 9) for a rigorous but readable discussion of expected values. The reader who lacks this background is directed instead to any introductory statistics text (e.g., Hays, 1973: 871) for an introduction to these concepts. This appendix will deal only with the algebra of expectations required for an understanding of Chapters 2, 3, 4, and 5.

If a random variable, x, is *discrete*, it takes on only a finite set of values. This is ordinarily written as

$$x = \{x_1, \ldots, x_n\}.$$

If each element of this set has an associated probability, $p(x_i)$, then the *expected value* of x is defined as

$$E(x) = \sum_{i=1}^{n} x_i \cdot p(x_i).$$



For example, if x is the number observed on the roll of a die, then x takes on only six values:

$$x = \{1, 2, 3, 4, 5, 6\}.$$

Each of these six values is equiprobable, so the expected value of x is:

$$E(x) = \sum_{i=1}^{6} x_i (1/6) = 3.5.$$

On the other hand, if the random variable is *continuous*, it may take on any value in the real line. This is ordinarily written as

$$-\infty < x < +\infty.$$

If a probability density function, f(x), is defined for the real line, then the expected value of x is:

$$E(x) = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx.$$

A random shock, may take on any real value, so the expected value of a random shock is:

$$E(a)_t = \int_{-\infty}^{+\infty} a_t \cdot f(a_t) \cdot da_t = 0,$$

where $f(a_t)$ is the Normal probability density function.

The proof of any expected value theorem is done by summing (in the case of a discrete random variable) or integrating (in the case of a continuous random variable) the product of a random variable and its probability function. A random variable of particular interest to the time series analyst is white noise: the random shock. Each shock is distributed Normally and independently with zero-mean and constant variance, that is:

$$a_t \sim NID(0, \sigma_a^2).$$

The implication of this distribution is that

$$E(a_t) = 0$$
$$E(a_t^2) = \sigma_a^2$$
$$E(a_t a_{t+n}) = 0.$$

Each of these expected values can be derived by integrating the product of the term and its density function as indicated. We will take these expected values as givens.

The expectation operator, E, is applied to a random variable, or to a combination of random variables, to derive the expected value. The expectation operator is a *linear* operator, so the procedure of applying the operator follows the common rules of linear algebra. These rules consist of the five listed below.

*First, the operator is applied to a function only after all other operations have been performed.* For example, to take the expected value of the term

$$(a_t - \Theta_1 a_{t-1})^2,$$

the expectation operator is applied

$$E\,[(a_t - \Theta_1 a_{t-1})^2].$$

However, all operations indicated inside the brackets ([ ]) must be performed before the expected value is taken, that is,

$$E\,[(a_t - \Theta_1 a_{t-1})^2] = E\,[a_t^2 - 2\Theta_1^2 a_t a_{t-1} + a_{t-1}^2].$$

Taking the expected value before the bracketed operations have been performed will ordinarily not give the expected value of the function.

*Second, the expected value of a constant is the constant,* for example,

$$E\,(\Theta_0) = \Theta_0.$$

*Third, the expected value of the product of a constant and a random variable is the product of the constant and the expected value of the random variable,* for example,

$$E\,(\Theta_1 a_{t-1}) = \Theta_1\, E\,(a_{t-1})\,.$$

*Fourth, the expected value of the sum of two random variables is the sum of the expected values,* for example,

$$E\,(a_t - \Theta_1 a_{t-1}) = E\,(a_t) - \Theta_1 E\,(a_{t-1}).$$

This rule generalizes to any linear combination of random variables.

*Fifth, the expected value of the product of two independent random variables is the product of their expected values,* for example,

$$E\,(a_t a_{t-1}) = E\,(a_t)\, E\,(a_{t-1})$$

The key word here is *independence*. Random shocks are independent by definition, so the expected value of random shock products is equal to the product of their expected values. If two random variables are *not* independent, however, the expected value of their product is not generally the product of their expected values. For example, successive realizations of an ARIMA (0,0,1) process

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

$$\text{and } y_{t+1} = a_{t+1} - \Theta_1 a_t$$

have the same zero expected values:

$$E\,(y_t) = E\,(y_{t+1}) = 0$$

but because these two random variables are not independent, the expected value of their product is *not* the product of their expected values:

$$E\,(y_t y_{t+1}) \neq E\,(y_t) E(y_{t+1}).$$

Instead, the expected value of their product is:

$$\begin{aligned} E\,(y_t\, y_{t+1}) &= E\,[(a_t - \Theta_1 a_{t-1})\,(a_{t+1} - \Theta_1 a_t)] \\ &= E\,[a_t a_{t+1} - \Theta_1 a_t^2 - \Theta_1 a_{t-1} a_{t+1} + \Theta_1^2 a_{t-1} a_t] \\ &= E\,(a_t a_{t+1}) - \Theta_1 E\,(a_t^2) - \Theta_1 E\,(a_{t-1} a_{t+1}) \\ &\qquad + \Theta_1^2 E\,(a_{t-1} a_t) \\ &= 0 - \Theta_1 \sigma_a^2 - 0 + 0 = -\,\Theta_1 \sigma_a^2. \end{aligned}$$

Our use of the expectation operator in Chapters 2, 3, 4, and 5 employs these five rules. Although a particular demonstration may appear formidable, the algebraic manipulations are all straightforward. The reader is urged to learn the rules for applying expectation operators, as developed in this appendix, and to replicate each derivation.