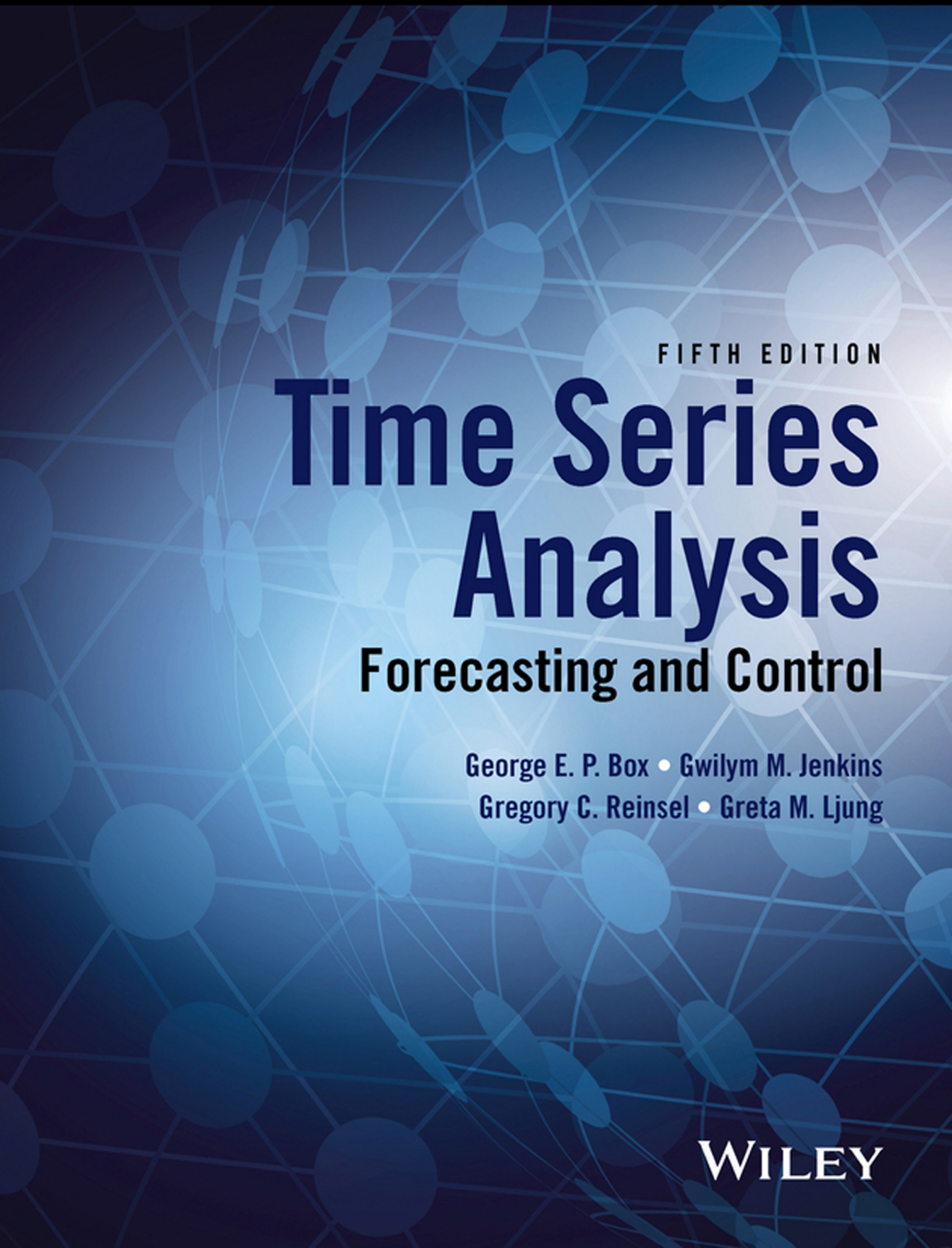
WILEY SERIES IN PROBABILITY AND STATISTICS



FIFTH EDITION

Time Series Analysis

Forecasting and Control

George E. P. Box • Gwilym M. Jenkins
Gregory C. Reinsel • Greta M. Ljung

WILEY

TIME SERIES ANALYSIS

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Garrett M. Fitzmaurice, Geof H. Givens, Harvey Goldstein, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Ruey S. Tsay, Sanford Weisberg*

Editors Emeriti: *J. Stuart Hunter, Iain M. Johnstone, Joseph B. Kadane, Jozef L. Teugels*

A complete list of the titles in this series appears at the end of this volume.

TIME SERIES ANALYSIS

Forecasting and Control

Fifth Edition

GEORGE E. P. BOX
GWILYM M. JENKINS
GREGORY C. REINSEL
GRETA M. LJUNG

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

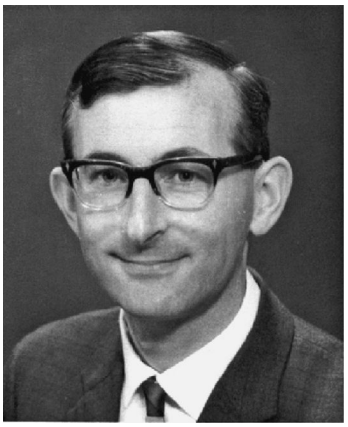
Box, George E. P.
Time series analysis : forecasting and control. – Fifth edition / George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-67502-1 (cloth : alk. paper) 1. Time-series analysis. 2. Prediction theory. 3. Transfer functions. 4. Feedback control systems–Mathematical models. I. Jenkins, Gwilym M. II. Reinsel, Gregory C. III. Ljung, Greta M., 1941- IV. Title.
QA280.B67 2016
519.5'5–dc23
2015015492

Printed in the United States of America

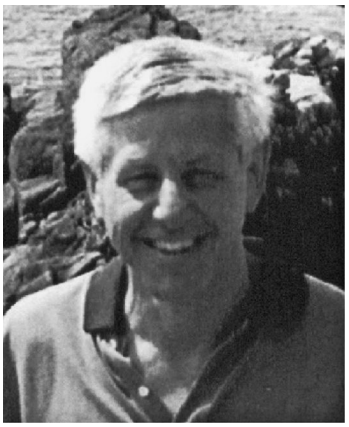
10 9 8 7 6 5 4 3 2 1

To the memory of

George E. P. Box



Gwilym M. Jenkins



Gregory C. Reinsel

CONTENTS

PREFACE TO THE FIFTH EDITION

This book describes statistical models and methods for analyzing discrete time series and presents important applications of the methodology. The models considered include the class of autoregressive integrated moving average (ARIMA) models and various extensions of these models. The properties of the models are examined and statistical methods for model specification, parameter estimation, and model checking are presented. Applications to forecasting nonseasonal as well as seasonal time series are described. Extensions of the methodology to transfer function modeling of dynamic relationships between two or more time series, modeling the effects of intervention events, multivariate time series modeling, and process control are discussed. Topics such as state-space and structural modeling, nonlinear models, long-memory models, and conditionally heteroscedastic models are also covered. The goal has been to provide a text that is practical and of value to both academicians and practitioners.

The first edition of this book appeared in 1970 and around that time there was a great upsurge in research on time series analysis and forecasting. This generated a large influx of new ideas, modifications, and improvements by many authors. For example, several new research directions began to emerge in econometrics around that time, leading to what is now known as time series econometrics. Many of these developments were reflected in the fourth edition of this book and have been further elaborated upon in this new edition.

The main goals of preparing a new edition have been to expand and update earlier material, incorporate new literature, enhance and update numerical illustrations through the use of R, and increase the number of exercises in the book. Some of the chapters in the previous edition have been reorganized. For example, Chapter 14 on multivariate time series analysis has been reorganized and expanded, placing more emphasis on vector autoregressive (VAR) models. The VAR models are by far the most widely used multivariate time series models in applied work. This edition provides an expanded treatment of these models that includes software demonstrations.

Chapter 10 has also been expanded and updated. This chapter covers selected topics in time series analysis that either extend or supplement material discussed in earlier chapters.

This includes unit roots testing, modeling of conditional heteroscedasticity, nonlinear models, and long memory models. A section of unit root testing that appeared in Chapter 7 of the previous edition has been expanded and moved to Section 10.1 in this edition. Section 10.2 deals with autoregressive conditionally heteroscedastic models, such as the ARCH and GARCH models. These models focus on the variability in a time series and are useful for modeling the volatility or variability in economic and financial series, in particular. The treatment of the ARCH and GARCH models has been expanded and several extensions have been added.

Elsewhere in the text, the exposition has been enhanced by revising, modifying, and omitting text as appropriate. Several tables have either been edited or replaced by graphs to make the presentation more effective. The number of exercises has been increased throughout the text and they now appear at the end of each chapter.

A further enhancement to this edition is the use of the statistical software R for model building and forecasting. The R package is available as a free download from the R Project for Statistical Computing at www.r-project.org. A brief description of the software is given in Appendix A1.1 of Chapter 1. Graphs generated using R now appear in many of the chapters along with R code that will help the reader reconstruct the graphs. The software is also used for numerical illustration in many of the examples in the text.

The fourth edition of this book was published by Wiley in 2008. Plans for a new edition began during the fall of 2012. I was deeply honored when George Box asked me to help him with this update. George was my Ph.D. advisor at the University of Wisconsin-Madison and remained a dear friend to me over the years as he did to all his students. Sadly, he was rather ill when the plans for this new edition were finalized towards the end of 2012. He did not have a chance to see the project completed as he passed away in March of 2013. I am deeply grateful for the opportunity to work with him and for the confidence he showed in assigning me this task. The book is dedicated to his memory and to the memory of his distinguished co-authors Gwilym Jenkins and Gregory Reinsel. Their contributions were many and they are all missed.

I also want to express my gratitude to several friends and colleagues in the time series community who have read the manuscript and provided helpful comments and suggestions. These include Ruey Tsay, William Wei, Sung Ahn, and Raja Velu who have read Chapter 14 on multivariate time series analysis, and David Dickey, Johannes Ledolter, Timo Teräsvirta, and Niels Haldrup who have read Chapter 10 on special topics. Their constructive comments and suggestions are much appreciated. Assistance and support from Paul Lindholm in Finland is also gratefully acknowledged. The use of R in this edition includes packages developed for existing books on time series analysis such as Cryer and Chan (2010), Shumway and Stoffer (2011), and Tsay (2014). We commend these authors for making their code and datasets available for public use through the R Project.

Research for the original version of this book was supported by the Air Force Office of Scientific Research and by the British Science Research Council. Research incorporated in the third edition was partially supported by the Alfred P. Sloan Foundation and by the National Aeronautics and Space Administration. Permission to reprint selected tables from *Biometrika Tables for Statisticians*, Vol. 1, edited by E. S. Pearson and H. O. Hartley is also acknowledged. On behalf of my co-authors, I would like to thank George Tiao, David Mayne, David Pierce, Granville Tunnicliffe Wilson, Donald Watts, John Hampton, Elaine Hodkinson, Patricia Blant, Dean Wichern, David Bacon, Paul Newbold, Hiro Kanemasu, Larry Haugh, John MacGregor, Bovas Abraham, Johannes Ledolter, Gina Chen, Raja Velu, Sung Ahn, Michael Wincek, Carole Leigh, Mary Esser, Sandy Reinsel, and

Meg Jenkins, for their help, in many different ways, in preparing the earlier editions. A very special thanks is extended to Claire Box for her long-time help and support.

The guidance and editorial support of Jon Gurstelle and Sari Friedman at Wiley is gratefully acknowledged. We also thank Stephen Quigley for his help in setting up the project, and Katrina Maceda and Shikha Pahuja for their help with the production.

Finally, I want to express my gratitude to my husband Bert Beander for his encouragement and support during the preparation of this revision.

GRETA M. LJUNG

Lexington, MA
May 2015

PREFACE TO THE FOURTH EDITION

It may be of interest to briefly recount how this book came to be written. Gwilym Jenkins and I first became friends in the late 1950s. We were intrigued by an idea that a chemical reactor could be designed that optimized itself automatically and could follow a moving maximum. We both believed that many advances in statistical theory came about as a result of interaction with researchers who were working on real scientific problems. Helping to design and build such a reactor would present an opportunity to further demonstrate this concept.

When Gwilym Jenkins came to visit Madison for a year, we discussed the idea with the famous chemical engineer Olaf Hougen, then in his eighties. He was enthusiastic and suggested that we form a small team in a joint project to build such a system. The National Science Foundation later supported this project. It took 3 years, but suffice it to say, that after many experiments, several setbacks, and some successes the reactor was built and it worked.

As expected, this investigation taught us a lot. In particular, we acquired proficiency in the manipulation of difference equations that were needed to characterize the dynamics of the system. It also gave us a better understanding of nonstationary time series required for realistic modeling of system noise. This was a happy time. We were doing what we most enjoyed doing: interacting with experimenters in the evolution of ideas and the solution of real problems, with real apparatus and real data.

Later there was fallout in other contexts, for example, advances in time series analysis, in forecasting for business and economics, and also developments in statistical process control (SPC) using some notions learned from the engineers.

Originally Gwilym came for a year. After that I spent each summer with him in England at his home in Lancaster. For the rest of the year, we corresponded using small reel-to-reel tape recorders. We wrote a number of technical reports and published some papers but eventually realized we needed a book. The first two editions of this book were written during a period in which Gwilym was, with extraordinary courage, fighting a debilitating illness to which he succumbed sometime after the book had been completed.

Later Gregory Reinsel, who had profound knowledge of the subject, helped to complete the third edition. Also in this fourth edition, produced after his untimely death, the new material is almost entirely his. In addition to a complete revision and updating, this fourth edition resulted in two new chapters: Chapter 10 on nonlinear and long memory models and Chapter 12 on multivariate time series.

This book should be regarded as a tribute to Gwilym and Gregory.

I was especially blessed to work with two such gifted colleagues.

GEORGE E. P. BOX

Madison, Wisconsin
March 2008

PREFACE TO THE THIRD EDITION

This book is concerned with the building of stochastic (statistical) models for time series and their use in important areas of application. This includes the topics of forecasting, model specification, estimation, and checking, transfer function modeling of dynamic relationships, modeling the effects of intervention events, and process control. Coincident with the first publication of *Time Series Analysis: Forecasting and Control,* there was a great upsurge in research in these topics. Thus, while the fundamental principles of the kind of time series analysis presented in that edition have remained the same, there has been a great influx of new ideas, modifications, and improvements provided by many authors.

The earlier editions of this book were written during a period in which Gwilym Jenkins was, with extraordinary courage, fighting a slowly debilitating illness. In the present revision, dedicated to his memory, we have preserved the general structure of the original book while revising, modifying, and omitting text where appropriate. In particular, Chapter 7 on estimation of ARMA models has been considerably modified. In addition, we have introduced entirely new sections on some important topics that have evolved since the first edition. These include presentations on various more recently developed methods for model specification, such as canonical correlation analysis and the use of model selection criteria, results on testing for unit root nonstationarity in ARIMA processes, the state-space representation of ARMA models and its use for likelihood estimation and forecasting, score tests for model checking, structural components, and deterministic components in time series models and their estimation based on regression-time series model methods. A new chapter (12) has been developed on the important topic of *intervention and outlier* analysis, reflecting the substantial interest and research in this topic since the earlier editions.

Over the last few years, the new emphasis on industrial quality improvement has strongly focused attention on the role of control both in process *monitoring* and in process *adjustment.* The control section of this book has, therefore, been completely rewritten to serve as an introduction to these important topics and to provide a better understanding of their relationship.

The objective of this book is to provide practical techniques that will be available to most of the wide audience who could benefit from their use. While we have tried to remove the inadequacies of earlier editions, we have not attempted to produce here a rigorous mathematical treatment of the subject.

We wish to acknowledge our indebtedness to Meg (Margaret) Jenkins and to our wives, Claire and Sandy, for their continuing support and assistance throughout the long period of preparation of this revision.

Research on which the original book was based was supported by the Air Force Office of Scientific Research and by the British Science Research Council. Research incorporated in the third edition was partially supported by the Alfred P. Sloan Foundation and by the National Aeronautics and Space Administration. We are grateful to Professor E. S. Pearson and the Biometrika Trustees for permission to reprint condensed and adapted forms of Tables 1, 8, and 12 of *Biometrika Tables for Statisticians*, Vol. 1, edited by E. S. Pearson and H. O. Hartley, to Dr. Casimer Stralkowski for permission to reproduce and adapt three figures from his doctoral thesis, and to George Tiao, David Mayne, Emanuel Parzen, David Pierce, Granville Wilson, Donald Watts, John Hampton, Elaine Hodkinson, Patricia Blant, Dean Wichern, David Bacon, Paul Newbold, Hiro Kanemasu, Larry Haugh, John MacGregor, Bovas Abraham, Gina Chen, Johannes Ledolter, Greta Ljung, Carole Leigh, Mary Esser, and Meg Jenkins for their help, in many different ways, in preparing the earlier editions.

GEORGE BOX AND GREGORY REINSEL

1

INTRODUCTION

A *time series* is a sequence of observations taken sequentially in time. Many sets of data appear as time series: a monthly sequence of the quantity of goods shipped from a factory, a weekly series of the number of road accidents, daily rainfall amounts, hourly observations made on the yield of a chemical process, and so on. Examples of time series abound in such fields as economics, business, engineering, the natural sciences (especially geophysics and meteorology), and the social sciences. Examples of data of the kind that we will be concerned with are displayed as time series plots in Figures 2.1 and 4.1. An intrinsic feature of a time series is that, typically, adjacent observations are *dependent*. The nature of this dependence among observations of a time series is of considerable practical interest. *Time series analysis* is concerned with techniques for the analysis of this dependence. This requires the development of stochastic and dynamic models for time series data and the use of such models in important areas of application.

In the subsequent chapters of this book, we present methods for building, identifying, fitting, and checking models for time series and dynamic systems. The methods discussed are appropriate for discrete (sampled-data) systems, where observation of the system occurs at equally spaced intervals of time.

We illustrate the use of these time series and dynamic models in five important areas of application:

1. The *forecasting* of future values of a time series from current and past values.
2. The determination of the *transfer function* of a system subject to inertia—the determination of a dynamic input–output model that can show the effect on the output of a system of any given series of inputs.
3. The use of indicator input variables in transfer function models to represent and assess the effects of unusual *intervention* events on the behavior of a time series.

Time Series Analysis: Forecasting and Control, Fifth Edition. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

4. The examination of interrelationships among several related time series variables of interest and determination of appropriate *multivariate* dynamic models to represent these joint relationships among the variables over time.
5. The design of simple *control schemes* by means of which potential deviations of the system output from a desired target may, so far as possible, be compensated by adjustment of the input series values.

1.1 FIVE IMPORTANT PRACTICAL PROBLEMS

1.1.1 Forecasting Time Series

The use at time t of available observations from a time series to forecast its value at some future time $t + l$ can provide a basis for (1) economic and business planning, (2) production planning, (3) inventory and production control, and (4) control and optimization of industrial processes. As originally described by Holt et al. (1963), Brown (1962), and the Imperial Chemical Industries (ICI) monograph on short term forecasting (Coutie, 1964), forecasts are usually needed over a period known as the *lead time*, which varies with each problem. For example, the lead time in the inventory control problem was defined by Harrison (1965) as a period that begins when an order to replenish stock is placed with the factory and lasts until the order is delivered into stock.

We will assume that observations are available at *discrete*, equispaced intervals of time. For example, in a sales forecasting problem, the sales z_t in the current month t and the sales $z_{t-1}, z_{t-2}, z_{t-3}, \ldots$ in previous months might be used to forecast sales for lead times $l = 1, 2, 3, \ldots, 12$ months ahead. Denote by $\hat{z}_t(l)$ the forecast made at *origin* t of the sales z_{t+l} at some future time $t + l$, that is, at *lead time* l. The function $\hat{z}_t(l)$, which provides the forecasts at origin t for all future lead times, based on the available information from the current and previous values $z_t, z_{t-1}, z_{t-2}, z_{t-3}, \ldots$ through time t, will be called the *forecast function* at origin t. Our objective is to obtain a forecast function such that the mean square of the deviations $z_{t+l} - \hat{z}_t(l)$ between the actual and forecasted values is as small as possible *for each lead time* l.

In addition to calculating the best forecasts, it is also necessary to specify their accuracy, so that, for example, the risks associated with decisions based upon the forecasts may be calculated. The accuracy of the forecasts may be expressed by calculating *probability limits* on either side of each forecast. These limits may be calculated for any convenient set of probabilities, for example, 50 and 95%. They are such that the realized value of the time series, when it eventually occurs, will be included within these limits with the stated probability. To illustrate, Figure 1.1 shows the last 20 values of a time series culminating at time t. Also shown are forecasts made from origin t for lead times $l = 1, 2, \ldots, 13$, together with the 50% probability limits.

Methods for obtaining forecasts and estimating probability limits are discussed in detail in Chapter 5. These forecasting methods are developed based on the assumption that the time series z_t follows a *stochastic* model of known form. Consequently, in Chapters 3 and 4 a useful class of such time series models that might be appropriate to represent the behavior of a series z_t, called autoregressive integrated moving average (ARIMA) models, are introduced and many of their properties are studied. Subsequently, in Chapters 6, 7, and 8 the practical matter of how these models may be developed for actual time series data is explored, and the methods are described through the three-stage procedure of tentative

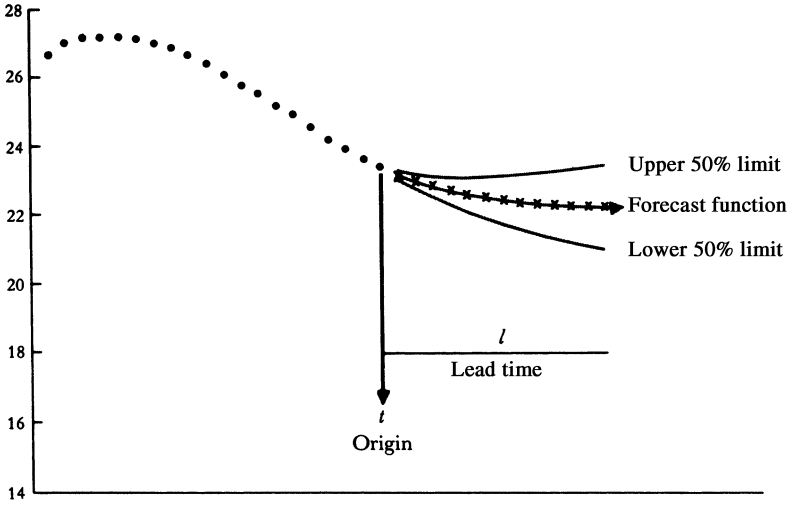


FIGURE 1.1 Values of a time series with forecast function and 50% probability limits.

model identification or specification, estimation of model parameters, and model checking and diagnostics.

1.1.2 Estimation of Transfer Functions

A topic of considerable industrial interest is the study of process dynamics discussed, for example, by Aström and Bohlin (1966, pp. 96–111) and Hutchinson and Shelton (1967). Such a study is made (1) to achieve better control of existing plants and (2) to improve the design of new plants. In particular, several methods have been proposed for estimating the transfer function of plant units from process records consisting of an input time series X_t and an output time series Y_t. Sections of such records are shown in Figure 1.2, where the input X_t is the rate of air supply and the output Y_t is the concentration of carbon dioxide produced in a furnace. The observations were made at 9-second intervals. A hypothetical impulse response function v_j, $j = 0, 1, 2, \ldots$, which determines the *transfer function* for the system through a dynamic linear relationship between input X_t and output Y_t of the form $Y_t = \sum_{j=0}^{\infty} v_j X_{t-j}$, is also shown in the figure as a bar chart. Transfer function models that

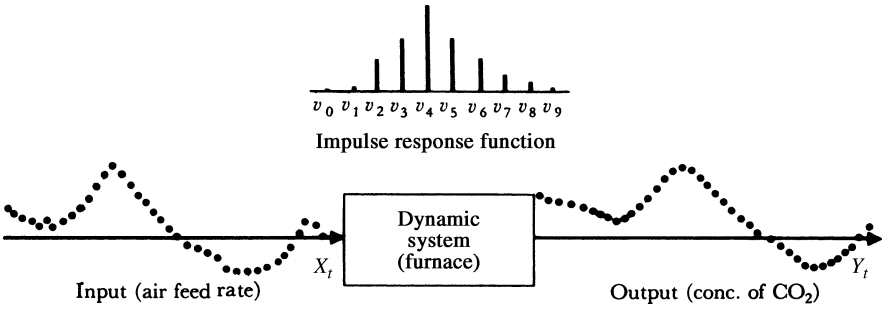


FIGURE 1.2 Input and output time series in relation to a dynamic system.

relate an input process X_t to an output process Y_t are introduced in Chapter 11 and many of their properties are examined.

Methods for estimating transfer function models based on deterministic perturbations of the input, such as step, pulse, and sinusoidal changes, have not always been successful. This is because, for perturbations of a magnitude that are relevant and tolerable, the response of the system may be masked by uncontrollable disturbances referred to collectively as *noise*. Statistical methods for estimating transfer function models that make allowance for noise in the system are described in Chapter 12. The estimation of dynamic response is of considerable interest in economics, engineering, biology, and many other fields.

Another important application of transfer function models is in forecasting. If, for example, the dynamic relationship between two time series Y_t and X_t can be determined, past values of *both* series may be used in forecasting Y_t. In some situations, this approach can lead to a considerable reduction in the errors of the forecasts.

1.1.3 Analysis of Effects of Unusual Intervention Events to a System

In some situations, it may be known that certain exceptional external events, *intervention events*, could have affected the time series z_t under study. Examples of such intervention events include the incorporation of new environmental regulations, economic policy changes, strikes, and special promotion campaigns. Under such circumstances, we may use transfer function models, as discussed in Section 1.1.2, to account for the effects of the intervention event on the series z_t, but where the ''input'' series will be in the form of a simple indicator variable taking only the values 1 and 0 to indicate (qualitatively) the presence or absence of the event.

In these cases, the intervention analysis is undertaken to obtain a quantitative measure of the impact of the intervention event on the time series of interest. For example, Box and Tiao (1975) used intervention models to study and quantify the impact of air pollution controls on smog-producing oxidant levels in the Los Angeles area and of economic controls on the consumer price index in the United States. Alternatively, the intervention analysis may be undertaken to adjust for any unusual values in the series z_t that might have resulted as a consequence of the intervention event. This will ensure that the results of the time series analysis of the series, such as the structure of the fitted model, estimates of model parameters, and forecasts of future values, are not seriously distorted by the influence of these unusual values. Models for intervention analysis and their use, together with consideration of the related topic of detection of outlying or unusual values in a time series, are presented in Chapter 13.

1.1.4 Analysis of Multivariate Time Series

For many problems in business, economics, engineering, and physical and environmental sciences, time series data may be available on several related variables of interest. A more informative and effective analysis is often possible by considering individual series as components of a multivariate or vector time series and analyzing the series jointly. For k-related time series variables of interest in a dynamic system, we may denote the series as $z_{1t}, z_{2t}, \ldots, z_{kt}$, and let $\boldsymbol{Z}_t = (z_{1t}, \ldots, z_{kt})'$ denote the $k \times 1$ time series vector at time t.

Methods of *multivariate* time series analysis are used to study the dynamic relationships among the several time series that comprise the vector $\boldsymbol{Z}_t$. This involves the development of statistical models and methods of analysis that adequately describe the interrelationships

among the series. Two main purposes for analyzing and modeling the vector of time series *jointly* are to gain an understanding of the dynamic relationships over time among the series and to improve accuracy of forecasts for individual series by utilizing the additional information available from the related series in the forecasts for each series. Multivariate time series models and methods for analysis and forecasting of multivariate series based on these models are considered in Chapter 14.

1.1.5 Discrete Control Systems

In the past, to the statistician, the words ''process control'' have usually meant the *quality control techniques* developed originally by Shewhart (1931) in the United States (see also Dudding and Jennet, 1942). Later on, the sequential aspects of quality control were emphasized, leading to the introduction of *cumulative sum charts* by Page (1957, 1961) and Barnard (1959) and the *geometric moving average* charts of Roberts (1959). Such basic charts are frequently employed in industries concerned with the manufacture of discrete ''parts'' as one aspect of what is called *statistical process control* (SPC). In particular (see Deming, 1986), they are used for continuous monitoring of a process. That is, they are used to supply a continuous screening mechanism for detecting assignable (or special) causes of variation. Appropriate display of plant data ensures that significant changes are quickly brought to the attention of those responsible for running the process. Knowing the answer to the question ''*when* did a change of this particular kind occur?'' we may be able to answer the question ''*why* did it occur?'' Hence a continuous incentive for process stabilization and improvement can be achieved.

By contrast, in the process and chemical industries, various forms of *feedback and feedforward* adjustment have been used in what we will call *engineering process control* (EPC). Because the adjustments made by engineering process control are usually computed and applied automatically, this type of control is sometimes called *automatic process control* (APC). However, the *manner* in which these adjustments are made is a matter of convenience. This type of control is necessary when there are inherent *disturbances* or *noise* in the system inputs that are impossible or impractical to remove. When we can measure fluctuations in an input variable that can be observed but not changed, it may be possible to make appropriate compensatory changes in some other control variable. This is referred to as *feedforward control*. Alternatively, or in addition, we may be able to use the deviation from target or ''error signal'' of the output characteristic itself to calculate appropriate compensatory changes in the control variable. This is called *feedback control*. Unlike feedforward control, this mode of correction can be employed even when the source of the disturbances is not accurately known or the magnitude of the disturbance is not measured.

In Chapter 15, we draw on the earlier discussions in this book, on time series and transfer function models, to provide insight into the statistical aspects of these control methods and to appreciate better their relationships and different objectives. In particular, we show how some of the ideas of feedback control can be used to design simple charts for *manually adjusting* processes. For example, the upper chart of Figure 1.3 shows hourly measurements of the viscosity of a polymer made over a period of 42 hours. The viscosity is to be controlled about a target value of 90 units. As each viscosity measurement comes to hand, the process operator uses the nomogram shown in the middle of the figure to compute the adjustment to be made in the manipulated variable (gas rate). The lower chart of Figure 1.3 shows the adjustments made in accordance with the nomogram.

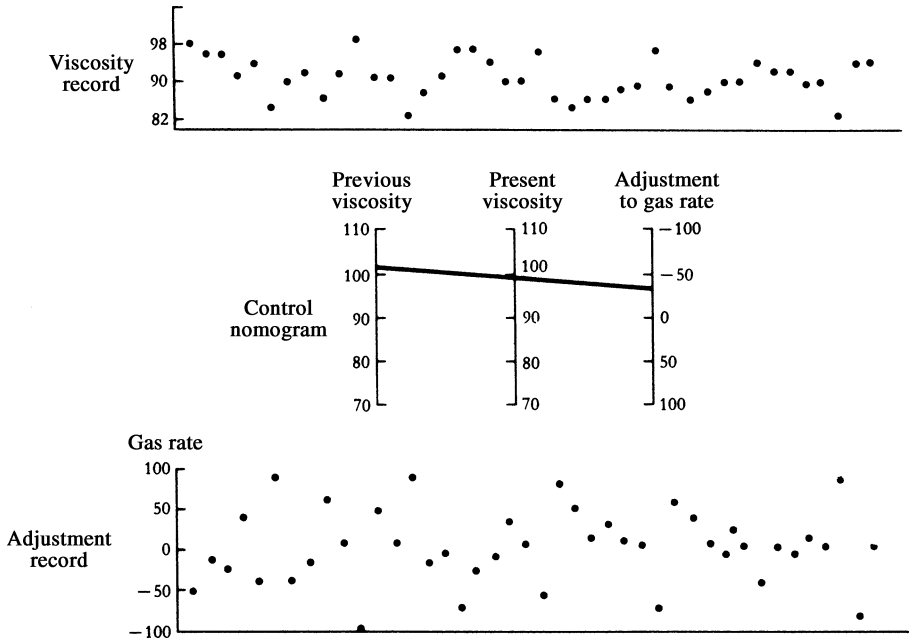


FIGURE 1.3 Control of viscosity. Record of observed viscosity and of adjustments in gas rate made using nomogram.

1.2 STOCHASTIC AND DETERMINISTIC DYNAMIC MATHEMATICAL MODELS

The idea of using a mathematical model to describe the behavior of a physical phenomenon is well established. In particular, it is sometimes possible to derive a model based on physical laws, which enables us to calculate the value of some time-dependent quantity nearly exactly at any instant of time. Thus, we might calculate the trajectory of a missile launched in a known direction with known velocity. If exact calculation were possible, such a model would be entirely *deterministic*.

Probably no phenomenon is totally deterministic, however, because unknown factors can occur such as a variable wind velocity that can throw a missile slightly off course. In many problems, we have to consider a time-dependent phenomenon, such as monthly sales of newsprint, in which there are many unknown factors and for which it is not possible to write a deterministic model that allows exact calculation of the future behavior of the phenomenon. Nevertheless, it may be possible to derive a model that can be used to calculate the *probability* of a future value lying between two specified limits. Such a model is called a probability model or a *stochastic model*. The models for time series that are needed, for example, to achieve optimal forecasting and control, are in fact stochastic models. It is necessary in what follows to distinguish between the probability model or stochastic process, as it is sometimes called, and the actually observed time series. Thus, a time series $z_1, z_2, \ldots, z_N$ of N successive observations is regarded as a sample realization from an infinite population of such time series that could have been generated by the stochastic

process. Very often we will omit the word ''stochastic'' from ''stochastic process'' and talk about the ''process.''

1.2.1 Stationary and Nonstationary Stochastic Models for Forecasting and Control

An important class of stochastic models for describing time series, which has received a great deal of attention, comprises what are called *stationary* models. Stationary models assume that the process remains in *statistical equilibrium* with probabilistic properties that do not change over time, in particular varying about a fixed *constant mean level* and with *constant variance*. However, forecasting has been of particular importance in industry, business, and economics, where many time series are often better represented as nonstationary and, in particular, as having no natural constant mean level over time. It is not surprising, therefore, that many of the economic forecasting methods originally proposed by Holt (1957, 1963), Winters (1960), Brown (1962), and the ICI monographs (Coutie, 1964) that used exponentially weighted moving averages can be shown to be appropriate for a particular type of *nonstationary* process. Although such methods are too narrow to deal efficiently with all time series, the fact that they often give the right kind of forecast function supplies a clue to the *kind of nonstationary* model that might be useful in these problems.

The stochastic model for which the exponentially weighted moving average forecast yields minimum mean square error (Muth, 1960) is a member of a class of *nonstationary* processes called autoregressive integrated moving average processes, which are discussed in Chapter 4. This wider class of processes provides a range of models, stationary and nonstationary, that adequately represent many of the time series met in practice. Our approach to forecasting has been first to derive an adequate stochastic model for the particular time series under study. As shown in Chapter 5, once an appropriate model has been determined for the series, the optimal forecasting procedure follows immediately. These forecasting procedures include the exponentially weighted moving average forecast as a special case.

Some Simple Operators. We employ extensively the *backward shift operator* B, which is defined by $Bz_t = z_{t-1}$; hence $B^m z_t = z_{t-m}$. The inverse operation is performed by the *forward shift operator* $F = B^{-1}$ given by $Fz_t = z_{t+1}$; hence $F^m z_t = z_{t+m}$. Another important operator is the *backward difference operator*, ∇, defined by $\nabla z_t = z_t - z_{t-1}$. This can be written in terms of B, since

$$\nabla z_t = z_t - z_{t-1} = (1 - B)z_t$$

Linear Filter Model. The stochastic models we employ are based on an idea originally due to Yule (1927) that an observable time series z_t in which successive values are highly dependent can frequently be regarded as generated from a series of *independent* ''shocks'' a_t. These shocks are *random* drawings from a fixed distribution, usually assumed normal and having mean zero and variance σ_a^2. Such a sequence of independent random variables $a_t, a_{t-1}, a_{t-2}, \ldots$ is called *a white noise* process.

The white noise process a_t is supposed transformed to the process z_t by what is called a *linear filter*, as shown in Figure 1.4. The linear filtering operation simply takes a weighted

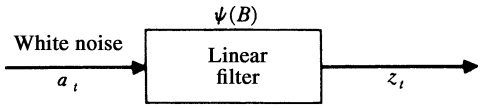


FIGURE 1.4 Representation of a time series as the output from a linear filter.

sum of previous random shocks a_t, so that

$$\begin{aligned} z_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= \mu + \psi(B) a_t \end{aligned} \tag{1.2.1}$$

In general, μ is a parameter that determines the ''level'' of the process, and

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$$

is the linear operator that transforms a_t into z_t and is called the *transfer function* of the filter. The model representation (1.2.1) can allow for a flexible range of patterns of dependence among values of the process $\{z_t\}$ expressed in terms of the independent (unobservable) random shocks a_t.

The sequence $\psi_1, \psi_2, \ldots$ formed by the weights may, theoretically, be finite or infinite. If this sequence is finite, or infinite and *absolutely summable* in the sense that $\sum_{j=0}^{\infty} |\psi_j| < \infty$, the filter is said to be *stable* and the process z_t is stationary. The parameter μ is then the mean about which the process varies. Otherwise, z_t is nonstationary and μ has no specific meaning except as a reference point for the level of the process.

Autoregressive Models. A stochastic model that can be extremely useful in the representation of certain practically occurring series is the *autoregressive* model. In this model, the current value of the process is expressed as a finite, linear aggregate of *previous values of the process* and a random shock a_t. Let us denote the values of a process at equally spaced times $t,\ t-1,\ t-2,\ \ldots$ by $z_t,\ z_{t-1},\ z_{t-2}, \ldots$. Also let $\tilde{z}_t = z_t - \mu$ be the series of deviations from μ. Then

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{1.2.2}$$

is called an *autoregressive* (AR) *process of order p*. The reason for this name is that a linear model

$$\tilde{z} = \phi_1 \tilde{x}_1 + \phi_2 \tilde{x}_2 + \cdots + \phi_p \tilde{x}_p + a$$

relating a ''dependent'' variable z to a set of ''independent'' variables $x_1, x_2, \ldots, x_p$, plus a random error term a, is referred to as a *regression* model, and z is said to be ''regressed'' on $x_1, x_2, \ldots, x_p$. In (1.2.2) the variable z is regressed on previous values of itself; hence the model is *auto*regressive. If we define an *autoregressive operator* of order p in terms of the backward shift operator B by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

the autoregressive model (1.2.2) may be written economically as

$$\phi(B)\tilde{z}_t = a_t$$

The model contains $p+2$ unknown parameters $\mu, \phi_1, \phi_2, \ldots, \phi_p, \sigma_a^2$, which in practice have to be estimated from the data. The additional parameter σ_a^2 is the variance of the white noise process a_t.

It is not difficult to see that the autoregressive model is a special case of the linear filter model of (1.2.1). For example, we can eliminate $\tilde{z}_{t-1}$ from the right-hand side of (1.2.2) by substituting

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \phi_2 \tilde{z}_{t-3} + \cdots + \phi_p \tilde{z}_{t-p-1} + a_{t-1}$$

Similarly, we can substitute for $\tilde{z}_{t-2}$, and so on, to yield eventually an infinite series in the a's. Consider, specifically, the simple first-order ($p = 1$) AR process, $\tilde{z}_t = \phi \tilde{z}_{t-1} + a_t$. After m successive substitutions of $\tilde{z}_{t-j} = \phi \tilde{z}_{t-j-1} + a_{t-j}$, $j = 1, \ldots, m$ in the right-hand side we obtain

$$\tilde{z}_t = \phi^{m+1} \tilde{z}_{t-m-1} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots + \phi^m a_{t-m}$$

In the limit as $m \to \infty$ this leads to the convergent infinite series representation $\tilde{z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j}$ with $\psi_j = \phi^j$, $j \geq 1$, provided that $|\phi| < 1$. Symbolically, in the general AR case we have that

$$\phi(B)\tilde{z}_t = a_t$$

is equivalent to

$$\tilde{z}_t = \phi^{-1}(B) a_t = \psi(B) a_t$$

with $\psi(B) = \phi^{-1}(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

Autoregressive processes can be stationary or nonstationary. For the process to be stationary, the ϕ's must be such that the weights $\psi_1, \psi_2, \ldots$ in $\psi(B) = \phi^{-1}(B)$ form a convergent series. The necessary requirement for stationarity is that the autoregressive operator, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, considered as a polynomial in B of degree p, must have all roots of $\phi(B) = 0$ greater than 1 in absolute value; that is, all roots must lie outside the unit circle. For the first-order AR process $\tilde{z}_t = \phi \tilde{z}_{t-1} + a_t$ this condition reduces to the requirement that $|\phi| < 1$, as the argument above has already indicated.

Moving Average Models. The autoregressive model (1.2.2) expresses the deviation $\tilde{z}_t$ of the process as a *finite* weighted sum of p previous deviations $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots, \tilde{z}_{t-p}$ of the process, plus a random shock a_t. Equivalently, as we have just seen, it expresses $\tilde{z}_t$ as an *infinite* weighted sum of the a's.

Another kind of model, of great practical importance in the representation of observed time series, is the finite *moving average* process. Here we take $\tilde{z}_t$, linearly dependent on a *finite* number q of previous a's. Thus,

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{1.2.3}$$

is called a *moving average* (MA) *process of order q*. The name ''moving average'' is somewhat misleading because the weights $1, -\theta_1, -\theta_2, \ldots, -\theta_q$, which multiply the a's, need not total unity nor need they be positive. However, this nomenclature is in common use, and therefore we employ it.

If we define a *moving average operator* of order q by

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

the moving average model may be written economically as

$$\tilde{z}_t = \theta(B)a_t$$

It contains $q + 2$ unknown parameters μ, $\theta_1, \ldots, \theta_q$, σ_a^2, which in practice have to be estimated from the data.

Mixed Autoregressive–Moving Average Models. To achieve greater flexibility in fitting of actual time series, it is sometimes advantageous to include both autoregressive and moving average terms in the model. This leads to the mixed *autoregressive–moving average* (ARMA) model:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{1.2.4}$$

or

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

The model employs $p + q + 2$ unknown parameters μ, $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$, σ_a^2, that are estimated from the data. This model may also be written in the form of the linear filter (1.2.1) as $\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t = \psi(B)a_t$, with $\psi(B) = \phi^{-1}(B)\theta(B)$. In practice, it is frequently true that an adequate representation of actually occurring stationary time series can be obtained with autoregressive, moving average, or mixed models, in which p and q are not greater than 2 and often less than 2. We discuss the classes of autoregressive, moving average, and mixed models in much greater detail in Chapters 3 and 4.

Nonstationary Models. Many series actually encountered in industry or business (e.g., stock prices and sales figures) exhibit nonstationary behavior and in particular do not vary about a fixed mean. Such series may nevertheless exhibit homogeneous behavior over time of a kind. In particular, although the general level about which fluctuations are occurring may be different at different times, the broad behavior of the series, when differences in level are allowed for, may be similar over time. We show in Chapter 4 and later chapters that such behavior may often be represented by a model in terms of a generalized autoregressive operator $\varphi(B)$, in which one or more of the zeros of the polynomial $\varphi(B)$ [i.e., one or more of the roots of the equation $\varphi(B) = 0$] lie on the unit circle. In particular, if there are d unit roots and all other roots lie outside the unit circle, the operator $\varphi(B)$ can be written

$$\varphi(B) = \phi(B)(1 - B)^d$$

where $\phi(B)$ is a stationary autoregressive operator. Thus, a model that can represent homogeneous nonstationary behavior is of the form

$$\varphi(B)z_t = \phi(B)(1 - B)^d z_t = \theta(B)a_t$$

that is,

$$\phi(B)w_t = \theta(B)a_t \tag{1.2.5}$$

where

$$w_t = (1 - B)^d z_t = \nabla^d z_t \tag{1.2.6}$$

Thus, homogeneous nonstationary behavior can sometimes be represented by a model that calls for the dth difference of the process to be stationary. In practice, d is usually 0, 1, or at most 2, with $d = 0$ corresponding to stationary behavior.

The process defined by (1.2.5) and (1.2.6) provides a powerful model for describing stationary and nonstationary time series and is called an *autoregressive integrated moving average process*, of order (p, d, q), or ARIMA(p, d, q) process. The process is defined by

$$w_t = \phi_1 w_{t-1} + \cdots \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{1.2.7}$$

with $w_t = \nabla^d z_t$. Note that if we replace w_t, by $z_t - \mu$, when $d = 0$, the model (1.2.7) includes the stationary mixed model (1.2.4), as a special case, and also the pure autoregressive model (1.2.2) and the pure moving average model (1.2.3).

The reason for inclusion of the word ''integrated'' (which should perhaps more appropriately be ''summed'') in the ARIMA title is as follows. The relationship, which is the inverse to (1.2.6), is $z_t = S^d w_t$, where $S = \nabla^{-1} = (1 - B)^{-1} = 1 + B + B^2 + \cdots$ is the *summation operator* defined by

$$Sw_t = \sum_{j=0}^{\infty} w_{t-j} = w_t + w_{t-1} + w_{t-2} + \cdots$$

Thus, the general ARIMA process may be generated by summing or ''integrating'' the stationary ARMA process $w_t d$ times. In Chapter 9, we describe how a special form of the model (1.2.7) can be employed to represent seasonal time series. The chapter also includes a discussion of regression models where the errors are autocorrelated and follow an ARMA process.

Chapter 10 includes material that may be considered more specialized and that either supplements or extends the material presented in the earlier chapters. The chapter begins with a discussion of unit root testing that may be used as a supplementary tool to determine if a time series is nonstationary and can be made stationary through differencing. This is followed by a discussion of conditionally heteroscedastic models such as the ARCH and GARCH models. These models assume that the conditional variance of an observation given its past vary over time and are useful for modeling time varying volatility in economic and financial time series, in particular. In Chapter 10, we also discuss nonlinear time series models and fractionally integrated long-memory processes that allow for certain more general features in a time series than are possible using the linear ARIMA models.

1.2.2 Transfer Function Models

An important type of dynamic relationship between a continuous input and a continuous output, for which many physical examples can be found, is that in which the deviations of input X and output Y, from appropriate mean values, are related by a *linear* differential equation. In a similar way, for discrete data, in Chapter 11 we represent the transfer relationship between an output Y and an input X, each measured at equispaced times, by

the difference equation

$$(1 + \xi_1\nabla + \cdots + \xi_r\nabla^r)Y_t = (\eta_0 + \eta_1\nabla + \cdots + \eta_s\nabla^s)X_{t-b} \tag{1.2.8}$$

in which the differential operator $D = d/dt$ is replaced by the difference operator $\nabla = 1 - B$. An expression of the form (1.2.8), containing only a few parameters ($r \leq 2, s \leq 2$), may often be used as an approximation to a dynamic relationship whose true nature is more complex.

The linear model (1.2.8) may be written equivalently in terms of past values of the input and output by substituting $B = 1 - \nabla$ in (1.2.8), that is,

$$\begin{aligned}(1 - \delta_1 B - \cdots - \delta_r B^r)Y_t &= (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)X_{t-b} \\ &= (\omega_0 B^b - \omega_1 B^{b+1} - \cdots - \omega_s B^{b+s})X_t\end{aligned} \tag{1.2.9}$$

or

$$\delta(B)Y_t = \omega(B)B^b X_t = \Omega(B)X_t$$

Alternatively, we can say that the output Y_t and the input X_t are linked by a linear filter

$$\begin{aligned}Y_t &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots \\ &= v(B)X_t\end{aligned} \tag{1.2.10}$$

for which the transfer function

$$v(B) = v_0 + v_1 B + v_2 B^2 + \cdots \tag{1.2.11}$$

can be expressed as a ratio of two polynomial operators,

$$v(B) = \frac{\Omega(B)}{\delta(B)} = \delta^{-1}(B)\Omega(B)$$

The linear filter (1.2.10) is said to be *stable* if the series (1.2.11) converges for $|B| \leq 1$, equivalently, if the coefficients $\{v_j\}$ are absolutely summable, $\sum_{j=0}^{\infty} |v_j| < \infty$. The sequence of weights $v_0, v_1, v_2, \ldots$, which appear in the transfer function (1.2.11), is called the *impulse response function*. We note that for the model (1.2.9), the first b weights $v_0, v_1, \ldots, v_{b-1}$, are zero. A hypothetical impulse response function for the system of Figure 1.2 is shown in the center of that diagram.

Models with Superimposed Noise. We have seen that the problem of estimating an appropriate model, linking an output Y_t and an input X_t, is equivalent to estimating the transfer function $v(B) = \delta^{-1}(B)\Omega(B)$, for example, specifying the parametric form of the transfer function $v(B)$ and estimating its parameters. However, this problem is complicated in practice by the presence of noise N_t, which we assume corrupts the true relationship between input and output according to

$$Y_t = v(B)X_t + N_t$$

where N_t and X_t are independent processes. Suppose, as indicated by Figure 1.5, that the noise N_t can be described by a stationary or nonstationary stochastic model of the form

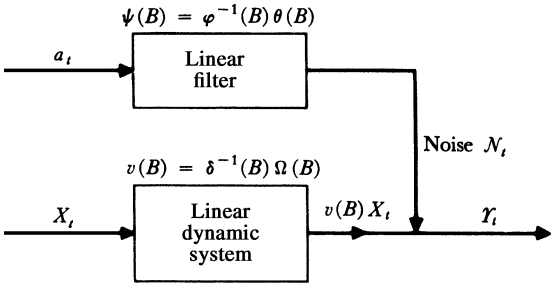


FIGURE 1.5 Transfer function model for dynamic system with superimposed noise model.

(1.2.5) or (1.2.7), that is,

$$N_t = \psi(B)a_t = \varphi^{-1}(B)\theta(B)a_t$$

Then the observed relationship between output and input will be

$$\begin{aligned} Y_t &= v(B)X_t + \psi(B)a_t \\ &= \delta^{-1}(B)\Omega(B)X_t + \varphi^{-1}(B)\theta(B)a_t \end{aligned} \tag{1.2.12}$$

In practice, it is necessary to estimate the transfer function

$$\psi(B) = \varphi^{-1}(B)\theta(B)$$

of the linear filter describing the noise, in addition to the transfer function $v(B) = \delta^{-1}(B)\Omega(B)$, which describes the dynamic relationship between the input and the output. Methods for doing this are discussed in Chapter 12.

1.2.3 Models for Discrete Control Systems

As stated in Section 1.1.5, control is an attempt to compensate for disturbances that infect a system. Some of these disturbances are measurable; others are not measurable and only manifest themselves as unexplained deviations from the target of the characteristic to be controlled. To illustrate the general principles involved, consider the special case where unmeasured disturbances affect the output Y_t of a system, and suppose that feedback control is employed to bring the output as close as possible to the desired target value by adjustments applied to an input variable X_t. This is illustrated in Figure 1.6. Suppose that N_t represents the effect at the output of various unidentified disturbances within the system, which in the absence of control could cause the output to drift away from the desired target value or *set point* T. Then, despite adjustments that have been made to the process, an error

$$\begin{aligned} \varepsilon_t &= Y_t - T \\ &= v(B)X_t + N_t - T \end{aligned}$$

will occur between the output and its target value T. The object is to choose a control equation so that the errors ε have the smallest possible mean square. The control equation expresses the adjustment $x_t = X_t - X_{t-1}$ to be taken at time t, as a function of the present deviation ε_t, previous deviations $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$, and previous adjustments $x_{t-1}, x_{t-2}, \ldots$. The mechanism (human, electrical, pneumatic, or electronic) that carries out the control action called for by the control equation is called the *controller*.

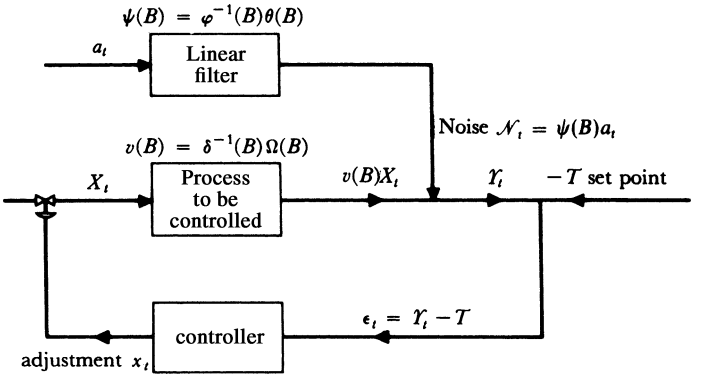


FIGURE 1.6 Feedback control scheme to compensate an unmeasured disturbance N_t.

One procedure for designing a controller is equivalent to forecasting the deviation from target which would occur *if no control were applied*, and then calculating the adjustment that would be necessary to cancel out this deviation. It follows that the forecasting and control problems are closely linked. In particular, if a minimum mean square error forecast is used, the controller will produce minimum mean square error control. To forecast the deviation from target that could occur if no control were applied, it is necessary to build a model

$$N_t = \psi(B)a_t = \varphi^{-1}(B)\theta(B)a_t$$

for the disturbance. Calculation of the adjustment x_t that needs to be applied to the input at time t to cancel out a predicted change at the output requires the building of a dynamic model with transfer function

$$v(B) = \delta^{-1}(B)\Omega(B)$$

which links the input with output. The resulting adjustment x_t will consist, in general, of a linear aggregate of previous adjustments and current and previous control errors. Thus the control equation will be of the form

$$x_t = \zeta_1 x_{t-1} + \zeta_2 x_{t-2} + \cdots + \chi_0 \varepsilon_t + \chi_1 \varepsilon_{t-1} + \chi_2 \varepsilon_{t-2} + \cdots \tag{1.2.13}$$

where $\zeta_1, \zeta_2, \ldots, \chi_0, \chi_1, \chi_2, \ldots$ are constants.

It turns out that, in practice, minimum mean square error control sometimes results in unacceptably large adjustments x_t to the input variable. Consequently, modified control schemes are employed that restrict the amount of variation in the adjustments. Some of these issues are discussed in Chapter 15.

1.3 BASIC IDEAS IN MODEL BUILDING

1.3.1 Parsimony

We have seen that the mathematical models we need to employ contain certain constants or parameters whose values must be estimated from the data. It is important, in practice, that

we employ the *smallest possible* number of parameters for adequate representations. The central role played by this principle of *parsimony* (Tukey, 1961) in the use of parameters will become clearer as we proceed. As a preliminary illustration, we consider the following simple example.

Suppose we fitted a dynamic model (1.2.9) of the form

$$Y_t = (\omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_s B^s)X_t \tag{1.3.1}$$

when dealing with a system that was adequately represented by

$$(1 - \delta B)Y_t = \omega_0 X_t \tag{1.3.2}$$

The model (1.3.2) contains only two parameters, δ and ω_0, but for s sufficiently large, it could be represented approximately by the model (1.3.1), through

$$Y_t = (1 - \delta B)^{-1}\omega_0 X_t = \omega_0(1 + \delta B + \delta^2 B^2 + \cdots)X_t$$

with $|\delta| < 1$. Because of experimental error, we could easily fail to recognize the relationship between the coefficients in the fitted equation. Thus, we might needlessly fit a relationship like (1.3.1), containing $s + 1$ parameters, where the much simpler form (1.3.2), containing only two, would have been adequate. This could, for example, lead to unnecessarily poor estimation of the output Y_t for given values of the input $X_t, X_{t-1}, \ldots$.

Our objective, then, must be to obtain adequate but parsimonious models. Forecasting and control procedures could be seriously deficient if these models were either inadequate or unnecessarily prodigal in the use of parameters. Care and effort is needed in selecting the model. The process of selection is necessarily iterative; that is, it is a process of evolution, adaptation, or trial and error and is outlined briefly below.

1.3.2 Iterative Stages in the Selection of a Model

If the physical mechanism of a phenomenon were completely understood, it would be possible theoretically to write down a mathematical expression that described it exactly. This would result in a *mechanistic* or *theoretical* model. In most instances the complete knowledge or large experimental resources needed to produce a mechanistic model are not available, and we must resort to an empirical model. Of course, the exact mechanistic model and the exclusively empirical model represent extremes. Models actually employed usually lie somewhere in between. In particular, we may use incomplete theoretical knowledge to indicate a suitable class of mathematical functions, which will then be fitted empirically (e.g., Box and Hunter, 1965); that is, the number of terms needed in the model and the numerical values of the parameters are estimated from experimental data. This is the approach that we adopt in this book. As we have indicated previously, the stochastic and dynamic models we describe can be justified, at least partially, on theoretical grounds as having the right general properties.

It is normally supposed that successive values of the time series under consideration or of the input–output data are available for analysis. If possible, at least 50 and preferably 100 observations or more should be used. In those cases where a past history of 50 or more observations is not available, one proceeds by using experience and past information to derive a preliminary model. This model may be updated from time to time as more data become available.

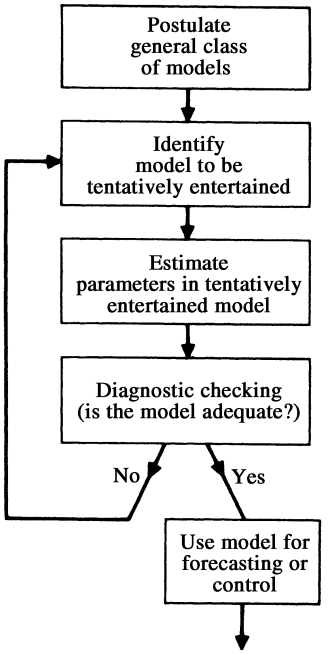


FIGURE 1.7 Stages in the iterative approach to model building.

In fitting dynamic models, a theoretical analysis can sometimes tell us not only the appropriate form for the model, but may also provide us with good estimates of the numerical values of its parameters. These values can then be checked later by analysis of data.

Figure 1.7 summarizes the iterative approach to model building for forecasting and control, which is employed in this book.

1. From the interaction of theory and practice, a *useful class of models* for the purposes at hand is considered.
2. Because this class is too extensive to be conveniently fitted directly to data, rough methods for *identifying* subclasses of these models are developed. Such methods of model identification employ data and knowledge of the system to suggest an appropriate parsimonious subclass of models that may be tentatively entertained. In addition, the identification process can be used to yield rough preliminary estimates of the parameters in the model.
3. The tentatively entertained model is *fitted* to data and its parameters *estimated*. The rough estimates obtained during the identification stage can now be used as starting values in more refined iterative methods for estimating the parameters, such as the nonlinear least squares and maximum likelihood methods.
4. *Diagnostic checks* are applied with the goal of uncovering possible lack of fit and diagnosing the cause. If no lack of fit is indicated, the model is ready to use. If any inadequacy is found, the iterative cycle of identification, estimation, and diagnostic checking is repeated until a suitable representation is found.

Identification, estimation, and diagnostic checking are discussed for univariate time series models in Chapters 6, 7, 8, and 9, for transfer function models in Chapter 12, for intervention models in Chapter 13, and for multivariate time series models in Chapter 14.

The model building procedures will be illustrated using actual time series with numerical calculations performed using the R software and other tools. A brief description of the R software is included in Appendix A1.1 along with references for further study. Exercises at the end of the chapters also make use of the software.

APPENDIX A1.1 USE OF THE R SOFTWARE

The R software for statistical computing and graphics is a common choice for data analysis and development of new statistical methods. R is available as Free Software under the terms of the Free Software Foundations's GNU General Public License in source code form. It compiles and runs on all common operating systems including Windows, MacOS X, and Linux. The main website for the R project is http://www.r-project.org.

The R environment consists of a base system, which is developed and maintained by the R Core Team, and a large set of user contributed packages. The base system provides the source code that implements the basic functionality of R. It also provides a set of standard packages that include commonly used probability distributions, graphical tools, classic datasets from the literature, and a set of statistical methods that include regression analysis and time series analysis. In addition to these base packages, there are now thousands of contributed packages developed by researchers around the world. Packages useful for time series modeling and forecasting include the stats package that is part of the base distribution and several contributed packages that are available for download. These include the TSA package by K-S Chan and Brian Ripley, the astsa package by David Stoffer, the Rmetrics packages fGarch and fUnitRoots for financial time series analysis by Diethelm Wuertz and associates, and the MTS package for multivariate time series analysis by Ruey Tsay. We use many functions from these packages in this book. We also use datasets available for download from the R datasets package, and the TSA and astsa packages.

Both the base system and the contributed packages are distributed through a network of servers called the Comprehensive R Archive Network (CRAN) that can be accessed from the official R website. Contributed packages that are not part of the base distribution can be installed directly from the R prompt ''>'' using the command install.package(). Under the Windows system, the installation can also be done from a drop-down list. The command will prompt the user to select a *CRAN Mirror*, after which a list of packages available for installation appears. To use a specific package, it also needs to be loaded into the system at the start of each session. For example, the TSA package can be loaded using the commands library(TSA) or require(TSA). The command data() will list all datasets available in the loaded packages. The command data(airquality) will load the dataset airquality from the R datasets package into memory. Data stored in a text file can be read into R using the command is read.table. For a .csv file, the command is read.csv. To get help on specific functions, e.g. the arima function which fits an ARIMA model to a time series, type help(arima) or ?arima.

R is object-oriented software and allows the user to create many objects. For example, the command ts() will create a time series object. This has advantages for plotting the time series and for certain other applications. However, it is not necessary to create a time series

object for many of the applications discussed in this book. The structure of the data in R can be examined using commands such as class(), str(), and summary().

The data used for illustration in this book, as well as in some of the exercises, include a set of time series listed in Part Five of the book. These series are also available at http://pages.cs.wisc.edu/ reinsel/bjr-data/index.html. At least three of the series are also included in the R datasets package and can be accessed using the data() command described above. Some of the exercises require the use of R and it will be assumed that the reader is already familiar with the basics of R, which can be obtained by working through relevant chapters of texts such as Crawley (2007) and Adler (2010). Comprehensive documentation in the form of manuals, contributed documents, online help pages, and FAQ sheets is also available on the R website. Since R builds on the S language, a useful reference book is also Venables and Ripley (2002).

EXERCISES

1.1. The dataset airquality in the R datasets package includes information on daily air quality measurements in New York, May to September 1973. The variables included are mean ozone levels at Roosevelt Island, solar radiation at Central Park, average wind speed at LaGuardia Airport, and maximum daily temperature at LaGuardia Airport; see help(airquality) for details.

(a) Load the dataset into R.

(b) Investigate the structure of the dataset.

(c) Plot each of the four series mentioned above using the plot() command in R; see help(plot) for details and examples.

(d) Comment on the behavior of the four series. Do you see any issues that may require special attention in developing a time series model for each of the four series.

1.2. Monthly totals of international airline passengers (in thousands of passengers), January 1949–December 1960, are available as Series G in Part Five of this book. The data are also available as series AirPassengers in the R datasets package.

(a) Load the dataset into R and examine the structure of the data.

(b) Plot the data using R and describe the behavior of the series.

(c) Perform a log transformation of the data and plot the resulting series. Compare the behavior of the original and log-transformed series. Do you see an advantage in using a log transformation for modeling purposes?

1.3. Download a time series of your choosing from the Internet. Note that financial and economic time series are available from sources such as Google Finance and the Federal Reserve Economic Data (FRED) of Federal Reserve Bank in St. Louis, Missouri, while climate data is available from from NOAA's National Climatic Data Center (NCDC).

(a) Store the data in a text file or a .csv file and read the data into R.

(b) Examine the properties of your series using plots or other appropriate tools.

(c) Does your time series appear to be stationary? If not, would differencing and/or some other transformation make the series stationary?

PART ONE

STOCHASTIC MODELS AND THEIR FORECASTING

In the first part of this book, which includes Chapters 2, 3, 4, and 5, a valuable class of stochastic models is described and its use in forecasting discussed.

A model that describes the probability structure of a sequence of observations is called a *stochastic process*. A time series of N successive observations $\mathbf{z}' = (z_1, z_2, \ldots, z_N)$ is regarded as a sample realization, from an infinite population of such samples, which could have been generated by the process. A major objective of statistical investigation is to infer properties of the population from those of the sample. For example, to make a forecast is to infer the probability distribution of a *future observation* from the population, given a sample $\mathbf{z}$ of past values. To do this, we need ways of describing stochastic processes and time series, and we also need classes of stochastic models that are capable of describing practically occurring situations. An important class of stochastic processes discussed in Chapter 2 is the *stationary* processes. They are assumed to be in a specific form of statistical equilibrium, and in particular, vary over time in a stable manner about a fixed mean. Useful devices for describing the behavior of stationary processes are the *autocorrelation function* and the *spectrum*.

Particular stationary stochastic processes of value in modeling time series are the autoregressive (AR), moving average (MA), and mixed autoregressive–moving average (ARMA) processes. The properties of these processes, in particular their autocorrelation structures, are described in Chapter 3.

Because many practically occurring time series (e.g., stock prices and sales figures) have nonstationary characteristics, the stationary models introduced in Chapter 3 are developed further in Chapter 4 to give a useful class of nonstationary processes called autoregressive integrated moving-average (ARIMA) models. The use of all these models in forecasting time series is discussed in Chapter 5 and is illustrated with examples.

Time Series Analysis: Forecasting and Control, Fifth Edition. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

2

AUTOCORRELATION FUNCTION AND SPECTRUM OF STATIONARY PROCESSES

A central feature in the development of time series models is an assumption of some form of *statistical equilibrium*. A particularly useful assumption of this kind (but an unduly restrictive one, as we will see later) is that of *stationarity*. Usually, a stationary time series can be usefully described by its mean, variance, and *autocorrelation function* or equivalently by its mean, variance, and *spectral density function*. In this chapter, we consider the properties of these functions and, in particular, the properties of the autocorrelation function, which will be used extensively in developing models for actual time series.

2.1 AUTOCORRELATION PROPERTIES OF STATIONARY MODELS

2.1.1 Time Series and Stochastic Processes

Time Series. A time series is a set of observations generated sequentially over time. If the set is continuous, the time series is said to be *continuous*. If the set is discrete, the time series is said to be *discrete*. Thus, the observations from a discrete time series made at times $\tau_1, \tau_2, \ldots, \tau_t, \ldots, \tau_N$ may be denoted by $z(\tau_1), z(\tau_2), \ldots, z(\tau_t), \ldots, z(\tau_N)$. In this book, we consider only discrete time series where observations are made at a fixed interval h. When we have N *successive* values of such a series available for analysis, we write $z_1, z_2, \ldots, z_t, \ldots, z_N$ to denote observations made at equidistant time intervals $\tau_0 + h, \tau_0 + 2h, \ldots, \tau_0 + th, \ldots, \tau_0 + Nh$. For many purposes, the values of τ_0 and h are unimportant, but if the observation times need to be defined exactly, these two values can be specified. If we adopt τ_0 as the origin and h as the unit of time, we can regard z_t as the observation *at time t*.

Time Series Analysis: Forecasting and Control, Fifth Edition. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

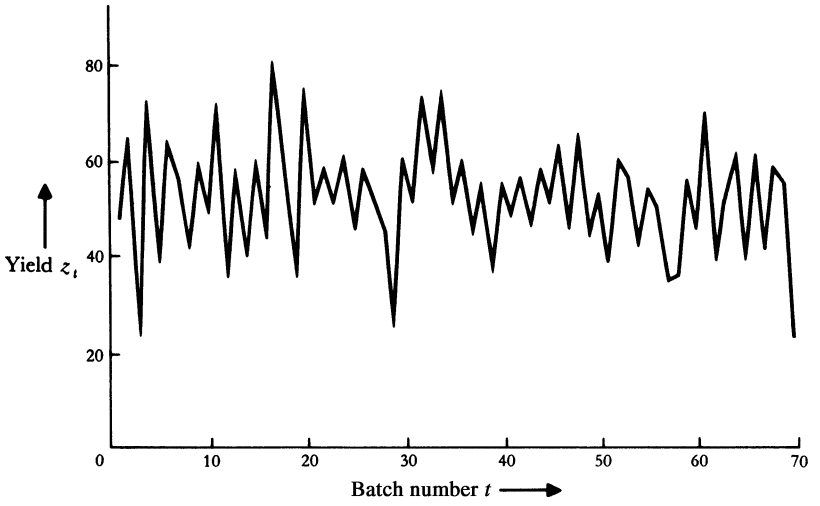


FIGURE 2.1 Yields of 70 consecutive batches from a chemical process.

Discrete time series may arise in two ways:

1. By *sampling* a continuous time series: For example, in the situation shown in Figure 1.2, where the continuous input and output from a gas furnace was sampled at intervals of 9 seconds.
2. By *accumulating* a variable over a period of time: Examples are rainfall, which is usually accumulated over a period such as a day or a month, and the yield from a batch process, which is accumulated over the batch time. For example, Figure 2.1 shows a time series consisting of the yields from 70 consecutive batches of a chemical process. The series shown here is included as Series F in Part Five of this book.

Deterministic and Statistical Time Series. If future values of a time series are exactly determined by some mathematical function such as

$$z_t = \cos(2\pi f t)$$

the time series is said to be *deterministic*. If future values can be described only in terms of a probability distribution, the time series is said to be nondeterministic or simply a *statistical time series*. The batch data of Figure 2.1 provide an example of a statistical time series. Thus, although there is a well-defined high–low pattern in the series, it is impossible to forecast the exact yield for the next batch. It is with such statistical time series that we are concerned in this book.

Stochastic Processes. A statistical phenomenon that evolves in time according to probabilistic laws is called a *stochastic process*. We will often refer to it simply as a *process*, omitting the word ''stochastic.'' The time series to be analyzed may then be thought of as a particular *realization*, produced by the underlying probability mechanism, of the system under study. In other words, *in analyzing a time series we regard it as a realization of a stochastic process.*

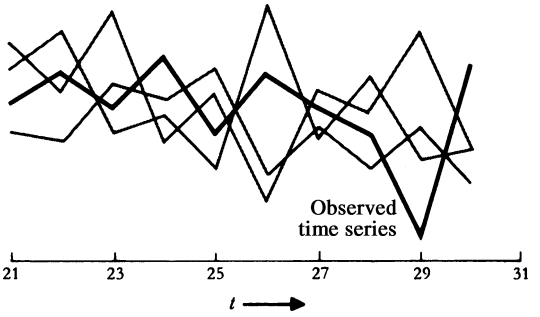


FIGURE 2.2 Observed time series (thick line), with other time series representing realizations of the same stochastic process.

For example, to analyze the batch data in Figure 2.1, we can imagine other sets of observations (other realizations of the underlying stochastic process), which might have been generated by the same chemical system, in the same $N = 70$ batches. Thus, Figure 2.2 shows the yields from batches $t = 21$ to $t = 30$ (thick line), together with other time series that *might* have been obtained from the population of time series defined by the underlying stochastic process. It follows that we can regard the observation z_t at a given time t, say $t = 25$, as a realization of a random variable z_t with probability density function $p(z_t)$. Similarly, the observations at any two times, say $t_1 = 25$ and $t_2 = 27$, may be regarded as realizations of two random variables z_{t_1} and z_{t_2} with joint probability density function $p(z_{t_1}, z_{t_2})$. For illustration Figure 2.3 shows contours of constant density for such a joint distribution, together with the marginal distribution at time t_1. In general, the observations making up an equispaced time series can be described by an N-dimensional random variable $(z_1, z_2, \ldots, z_N)$ with probability distribution $p(z_1, z_2, \ldots, z_N)$.

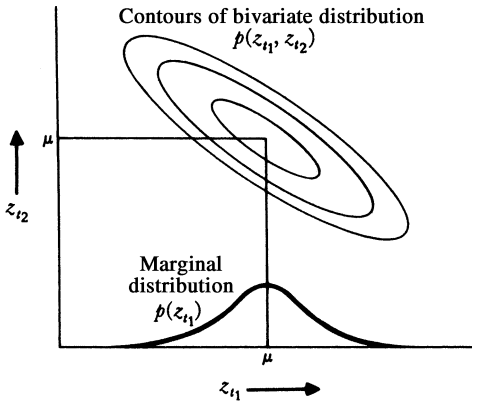


FIGURE 2.3 Contours of constant density of a bivariate probability distribution describing a stochastic process at two times t_1, t_2, together with the marginal distribution at time t_1.

2.1.2 Stationary Stochastic Processes

A very special class of stochastic processes, called *stationary processes*, is based on the assumption that the process is in a particular state of *statistical equilibrium*. A stochastic process is said to be *strictly stationary* if its properties are unaffected by a change of time origin, that is, if the joint probability distribution associated with m observations $z_{t_1}, z_{t_2}, \ldots, z_{t_m}$, made at *any* set of times $t_1, t_2, \ldots, t_m$, is the same as that associated with m observations $z_{t_1+k}, z_{t_2+k}, \ldots, z_{t_m+k}$, made at times $t_1 + k, t_2 + k, \ldots, t_m + k$. Thus, for a discrete process to be strictly stationary, the joint distribution of any set of observations must be unaffected by shifting all the times of observation forward or backward by any integer amount k.

Mean and Variance of a Stationary Process. When $m = 1$, the stationarity assumption implies that the probability distribution $p(z_t)$ is the same for all times t and may be written as $p(z)$. Hence, the stochastic process has a constant mean

$$\mu = E[z_t] = \int_{-\infty}^{\infty} z p(z) dz \tag{2.1.1}$$

which defines the level about which it fluctuates, and a constant variance

$$\sigma_z^2 = E[(z_t - \mu)^2] = \int_{-\infty}^{\infty} (z - \mu)^2 p(z) dz \tag{2.1.2}$$

which measures its *spread* about this level. Since the probability distribution $p(z)$ is the same for all times t, its shape can be inferred by forming the histogram of the observations $z_1, z_2, \ldots, z_N$, making up the observed time series. In addition, the mean μ of the stochastic process can be estimated by the sample mean

$$\bar{z} = \frac{1}{N} \sum_{t=1}^{N} z_t \tag{2.1.3}$$

of the time series, and the variance σ_z^2 of the stochastic process can be estimated by the sample variance

$$\hat{\sigma}_z^2 = \frac{1}{N} \sum_{t=1}^{N} (z_t - \bar{z})^2 \tag{2.1.4}$$

of the time series.

Autocovariance and Autocorrelation Coefficients. The stationarity assumption also implies that the joint probability distribution $p(z_{t_1}, z_{t_2})$ is the same for all times t_1, t_2, which are a constant interval apart. In particular, it follows that the covariance between values z_t and z_{t+k}, separated by k intervals of time, or by *lag* k, must be the same for all t under the stationarity assumption. This covariance is called the *autocovariance* at lag k and is defined by

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \tag{2.1.5}$$

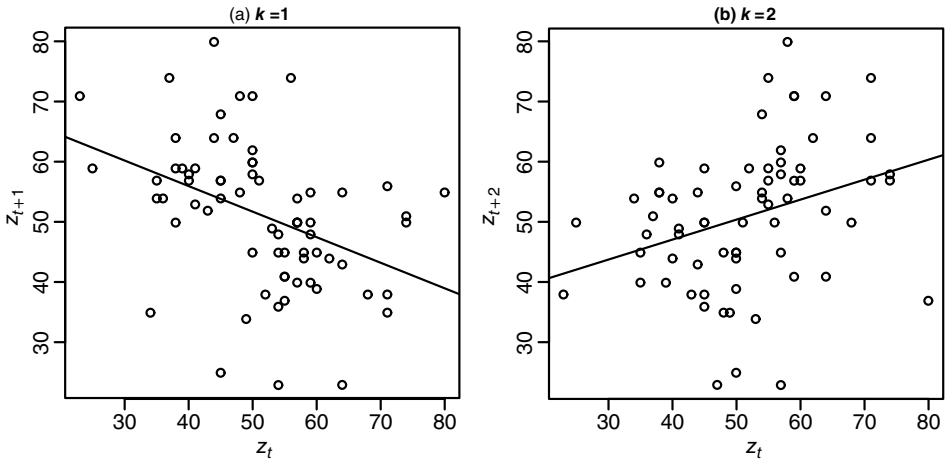


FIGURE 2.4 Scatter diagrams at lags (a) $k = 1$ and (b) $k = 2$ for the batch data of Figure 2.1.

Similarly, the *autocorrelation* at lag k is

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2]}} = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2}$$

since, for a stationary process, the variance $\sigma_z^2 = \gamma_0$ is the same at time $t + k$ as at time t. Thus, the autocorrelation at lag k, that is, the correlation between z_t and z_{t+k}, is

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{2.1.6}$$

which implies, in particular, that $\rho_0 = 1$.

It also follows for a stationary process that the nature of the joint probability distribution $p(z_t, z_{t+k})$ of values separated by k intervals of time can be inferred by plotting a scatter diagram using pairs of values (z_t, z_{t+k}) of the time series, separated by a constant interval or lag k. For the batch data displayed in Figure 2.1, Figure 2.4(a) shows a scatter diagram for lag $k = 1$, obtained by plotting z_{t+1} versus z_t, while Figure 2.4(b) shows a scatter diagram for lag $k = 2$, obtained by plotting z_{t+2} versus z_t. We see that neighboring values of the time series are correlated. The correlation between z_t and z_{t+1} appears to be negative and the correlation between z_t and z_{t+2} positive. Figure 2.4 was generated in R as follows:

```
> Yield = read.table("SeriesF.txt",header=TRUE)
> y1=Yield[2:70]
> x1=Yield[1:69]
> y2=Yield[3:70]
> x2=Yield[1:68]
> win.graph(width=5,height=2.7,pointsize=5)
> par(mfrow=c(1,2))  % Places two graphs side-by-side
> plot(y=y1,x=x1,ylab=expression(z[t+1]),xlab=expression(z[t]),
```

```
        main="(a): k=1",type='p')
> abline(lsfit(x1,y1))
> plot(y=y2,x=x2,ylab=expression(z[t+2]),xlab=expression(z[t]),
        main="(b): k=2",type='p')
> abline(lsfit(x2,y2))
```

2.1.3 Positive Definiteness and the Autocovariance Matrix

The covariance matrix associated with a stationary process for observations $(z_1, z_2, \ldots, z_n)$ made at n successive times is

$$\boldsymbol{\Gamma}_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix}$$
$$= \sigma_z^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} = \sigma_z^2 \mathbf{P}_n \tag{2.1.7}$$

A covariance matrix $\boldsymbol{\Gamma}_n$ of this form, which is symmetric with constant elements on any diagonal, is called an *autocovariance matrix*, and the corresponding correlation matrix $\mathbf{P}_n$ is called an *autocorrelation matrix*. Now, consider any linear function of the random variables $z_t, z_{t-1}, \ldots, z_{t-n+1}$:

$$L_t = l_1 z_t + l_2 z_{t-1} + \cdots + l_n z_{t-n+1} \tag{2.1.8}$$

Since $\text{cov}[z_i, z_j] = \gamma_{|j-i|}$ for a stationary process, the variance of L_t is

$$\text{var}[L_t] = \sum_{i=1}^{n} \sum_{j=1}^{n} l_i l_j \gamma_{|j-i|}$$

which is necessarily greater than zero if the l's are not all zero. It follows that both the autocovariance matrix and the autocorrelation matrix are positive definite for any stationary process. Correspondingly, it is seen that both the autocovariance function $\{\gamma_k\}$ and the autocorrelation function $\{\rho_k\}$, viewed as functions of the lag k, are positive-definite functions in the sense that $\sum_{i=1}^{n} \sum_{j=1}^{n} l_i l_j \gamma_{|j-i|} > 0$ for every positive integer n and all constants $l_1, \ldots, l_n$.

Conditions Satisfied by the Autocorrelations of a Stationary Process. The positive definiteness of the autocorrelation matrix (2.1.7) implies that its determinant and all principal minors are greater than zero. In particular, for $n = 2$,

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0$$

so that

$$1 - \rho_1^2 > 0$$

and hence

$$-1 < \rho_1 < 1$$

Similarly, for $n = 3$, we must have

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0 \qquad \begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} > 0$$

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} > 0$$

which implies that

$$\begin{aligned} -1 &< \rho_1 < 1 \\ -1 &< \rho_2 < 1 \\ -1 &< \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1 \end{aligned}$$

and so on. Since $\mathbf{P}_n$ must be positive definite for *all* values of n, the autocorrelations of a stationary process must satisfy a very large number of conditions. As will be shown in Section 2.2.3, all of these conditions can be brought together in the definition of the spectrum.

Stationarity of Linear Functions. It follows from the definition of stationarity that the process L_t, obtained by performing the linear operation (2.1.8) on a stationary process z_t for fixed n and fixed coefficients $l_1, \ldots, l_n$, is also stationary. The autocovariance of the process L_t, at a general lag $k \geq 0$, is given by

$$\text{cov}[L_t, L_{t-k}] = \sum_{i=1}^{n} \sum_{j=1}^{n} l_i l_j \text{cov}[z_{t+1-i}, z_{t+1-k-j}] = \sum_{i=1}^{n} \sum_{j=1}^{n} l_i l_j \gamma_{|k+j-i|}$$

In particular, the first difference $\nabla z_t = z_t - z_{t-1}$ and higher differences $\nabla^d z_t$ are stationary. This result is of particular importance to the discussion of nonstationary time series presented in Chapter 4.

The result also extends to infinite linear operations or infinite linear (time-invariant) filters applied to a stationary process $\{z_t\}$, under a condition of absolute summability. That is, if $\{z_t\}$ is a stationary process and $\{y_t\}$ is defined by the infinite linear (time-invariant) filter

$$y_t = \psi_0 z_t + \psi_1 z_{t-1} + \psi_2 z_{t-2} + \cdots = \sum_{i=0}^{\infty} \psi_i z_{t-i} \tag{2.1.9}$$

with fixed coefficients $\{\psi_i\}$ such that $\sum_{i=0}^{\infty} |\psi_i| < \infty$, then $\{y_t\}$ is also stationary. The absolute summability condition, $\sum_{i=0}^{\infty} |\psi_i| < \infty$, guarantees that the variables y_t in (2.1.9) are well-defined finite random variables (with probability one) and represent the limit of the sequence $\sum_{i=0}^{n} \psi_i z_{t-i}$ as $n \to \infty$. The variance of y_t in (2.1.9) (taking $E[z_t] = 0$ for convenience) is

$$\text{var}[y_t] = E[y_t^2] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_{|j-i|}$$

This variance is finite since $|\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_{|j-i|}| \le \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |\psi_i||\psi_j||\gamma_{|j-i|}| \le \gamma_0 \{\sum_{i=0}^{\infty} |\psi_i|\}^2 < \infty$. The autocovariance of y_t at any lag $k \ge 0$ is then

$$\text{cov}[y_t, y_{t-k}] = \lim_{n \to \infty} \sum_{i=0}^{n} \sum_{j=0}^{n} \psi_i \psi_j \gamma_{|k+j-i|} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_{|k+j-i|} \tag{2.1.10}$$

which converges by the dominated convergence result.

Gaussian Processes. If the probability distribution of observations associated with *any* set of times is a multivariate normal distribution, the process is called a *normal* or *Gaussian* process. Since the multivariate normal distribution is fully characterized by its moments of first and second order, the existence of a fixed mean μ and an autocovariance matrix $\mathbf{\Gamma}_n$ of the form (2.1.7) for all n would be sufficient to ensure the stationarity of a Gaussian process.

Weak Stationarity. We have seen that for a process to be strictly stationary, the whole probability structure must depend only on time differences. A less restrictive requirement, called *weak stationarity* of order f, is that the moments up to some order f depend only on time differences. For example, the existence of a fixed mean μ and an autocovariance matrix $\mathbf{\Gamma}_n$ of the form (2.1.7) is sufficient to ensure stationarity up to second order. That is, a process $\{z_t\}$ is *weakly stationary* (of order 2), or second-order stationary, if the mean $E[z_t] = \mu$ is a fixed constant for all t and the autocovariances $\text{cov}[z_t, z_{t+k}] = \gamma_k$ depend only on the time difference or time lag k for all t. Thus, second-order stationarity and an assumption of normality are sufficient to produce strict stationarity.

White Noise Process. The most fundamental example of a stationary process is a sequence of *independent and identically distributed* random variables, denoted as $a_1, \ldots, a_t, \ldots$, which we also assume to have mean zero and variance σ_a^2. This process is strictly stationary and is referred to as a *white noise* process. Because independence implies that the a_t are uncorrelated, its autocovariance function is simply

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \ne 0 \end{cases}$$

If one concentrates only on the second-order properties, then a sequence of random variables a_t, which are *uncorrelated*, have mean zero, and common variance σ_a^2 has the same autocovariance function γ_k as above, and is weakly (second-order) stationary. Such a process may also be referred to as a white noise process (in the weak sense), when the focus

is only on the second-order properties. Although the white noise process has very basic properties, this process plays an important role in the building of processes with much more interesting and more complicated properties through linear filtering operations as in (2.1.8) and (2.1.9).

2.1.4 Autocovariance and Autocorrelation Functions

It was seen in Section 2.1.2 that the autocovariance coefficient γ_k, at lag k, measures the covariance between two values z_t and z_{t+k} a distance k apart. The plot of γ_k versus lag k is called the *autocovariance function* $\{\gamma_k\}$ of the stochastic process. Similarly, the plot of the autocorrelation coefficient ρ_k as a function of the lag k is called the *autocorrelation function* $\{\rho_k\}$ of the process. Note that the autocorrelation function is dimensionless, that is, independent of the scale of measurement of the time series. Since $\gamma_k = \rho_k \sigma_z^2$, knowledge of the autocorrelation function $\{\rho_k\}$ and the variance σ_z^2 is equivalent to knowledge of the autocovariance function $\{\gamma_k\}$.

The autocorrelation function, shown in Figure 2.5 as a plot of the diagonals of the autocorrelation matrix, reveals how the correlation between any two values of the series changes as their separation changes. Since $\rho_k = \rho_{-k}$, the autocorrelation function is

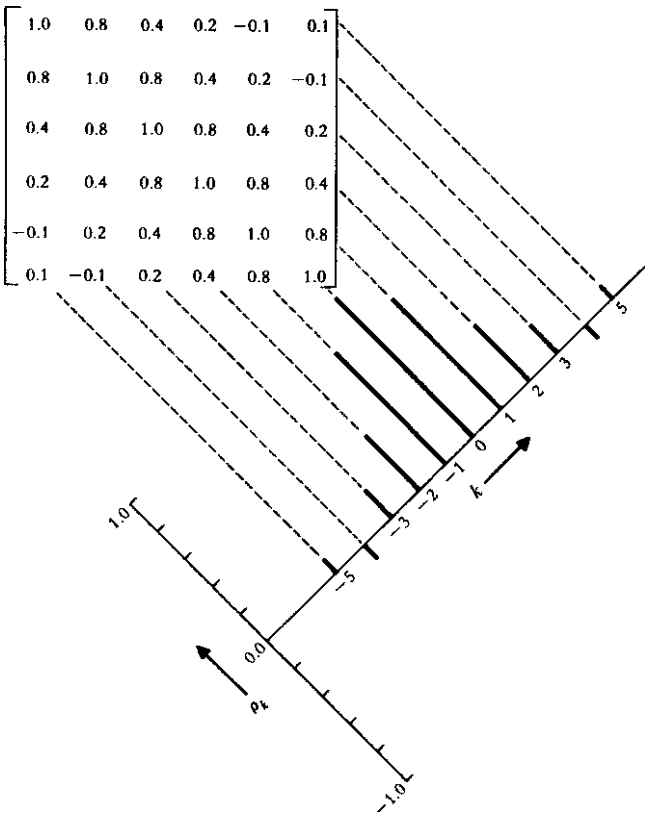


FIGURE 2.5 Autocorrelation matrix and corresponding autocorrelation function of a stationary process.

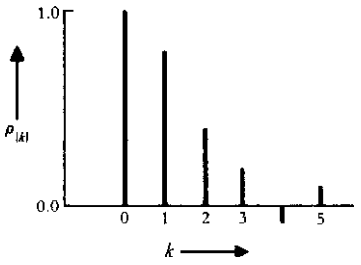


FIGURE 2.6 Positive half of the autocorrelation function of Figure 2.5.

necessarily symmetric about zero, and in practice it is only necessary to plot the positive half of this function. Figure 2.6 shows the positive half of the autocorrelation function given in Figure 2.5. When we speak of the autocorrelation function, we typically mean the positive half. In the past, the autocorrelation function has sometimes been called the *correlogram*.

From what has previously been shown, a *normal* stationary process z_t is completely characterized by its mean μ and its autocovariance function $\{\gamma_k\}$, or equivalently by its mean μ, variance σ_z^2, and autocorrelation function $\{\rho_k\}$.

2.1.5 Estimation of Autocovariance and Autocorrelation Functions

Up to now, we have only considered the theoretical autocorrelation function that describes a stochastic process. In practice, we have a finite time series $z_1, z_2, \ldots, z_N$ of N observations, from which we can only obtain *estimates* of the mean μ and the autocorrelations. The mean $\mu = E[z_t]$ is estimated as in (2.1.3) by the sample mean $\bar{z} = \sum_{t=1}^{N} z_t/N$. It is easy to see that $E[\bar{z}] = \mu$, so that $\bar{z}$ is an unbiased estimator of μ. As a measure of precision of $\bar{z}$ as an estimator of μ, we find that

$$\operatorname{var}[\bar{z}] = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\gamma_{t-s} = \frac{\gamma_0}{N}\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right]$$

A ''large-sample'' approximation for this variance is given by

$$\operatorname{var}[\bar{z}] = \left(\frac{\gamma_0}{N}\right)\left(1 + 2\sum_{k=1}^{\infty}\rho_k\right)$$

in the sense that $N\operatorname{var}[\bar{z}] \to \gamma_0\left(1 + 2\sum_{k=1}^{\infty}\rho_k\right)$ as $N \to \infty$, assuming that $\sum_{k=-\infty}^{\infty}|\rho_k| < \infty$. Notice that the first factor in $\operatorname{var}[\bar{z}]$, γ_0/N, is the familiar expression for the variance of $\bar{z}$ obtained from independent random samples of size N, but the presence of autocorrelation among the z_t values can substantially affect the precision of $\bar{z}$. For example, in the case where a stationary process has autocorrelations $\rho_k = \phi^{|k|}$, $|\phi| < 1$, the large-sample approximation for the variance of $\bar{z}$ becomes $\operatorname{var}[\bar{z}] = (\gamma_0/N)[(1+\phi)/(1-\phi)]$, and the second factor can obviously differ substantially from 1.

A number of estimates of the autocorrelation function have been suggested in the literature, and their properties are discussed by Jenkins and Watts (1968), among others. It

TABLE 2.1 Estimated Autocorrelation Function of Batch Data

k	r_k	k	r_k	k	r_k
1	−0.39	6	−0.05	11	0.11
2	0.30	7	0.04	12	−0.07
3	−0.17	8	−0.04	13	0.15
4	0.07	9	0.00	14	0.04
5	−0.10	10	0.01	15	−0.01

is concluded that the most satisfactory estimate of the kth lag autocorrelation ρ_k is

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0} \tag{2.1.11}$$

where

$$c_k = \hat{\gamma}_k = \frac{1}{N}\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z}) \qquad k = 0, 1, 2, \ldots, K \tag{2.1.12}$$

is the estimate of the autocovariance γ_k and $\bar{z}$ is the sample mean of the time series. The values r_k in (2.1.11) may be called the *sample* autocorrelation function. To obtain a useful estimate of the autocorrelation function in practice, we would typically need at least 50 observations, and the estimated autocorrelations r_k would be calculated for $k = 0, 1, \ldots, K$, where K was not larger than, say, $N/4$.

The estimated autocorrelation function r_k of the batch data in Figure 2.1 is given in Table 2.1 and plotted in Figure 2.7. The autocorrelation function is characterized by correlations that alternate in sign and tend to damp out with increasing lag. Autocorrelation functions of this kind are not uncommon in production data and can arise because of ''carryover'' effects. In this particular example, a high-yielding batch tended to produce tarry residues, which were not entirely removed from the vessel and adversely affected the yield of the next batch.

Figure 2.7 and the autocorrelations shown in Table 2.1 were generated in R as follows:

```
> Yield = read.table("SeriesF.txt",header=TRUE)
> ACF = acf(Yield,15)
> ACF    % retrieves the values shown in Table 2.1
```

2.1.6 Standard Errors of Autocorrelation Estimates

To identify a model for a time series, using methods to be described in Chapter 6, it is useful to have a rough check on whether ρ_k is effectively zero beyond a certain lag. For this purpose, we can use the following expression for the approximate variance of the estimated autocorrelation coefficient of a stationary normal process given by Bartlett (1946):

$$\text{var}[r_k] \simeq \frac{1}{N}\sum_{v=-\infty}^{\infty}(\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2) \tag{2.1.13}$$

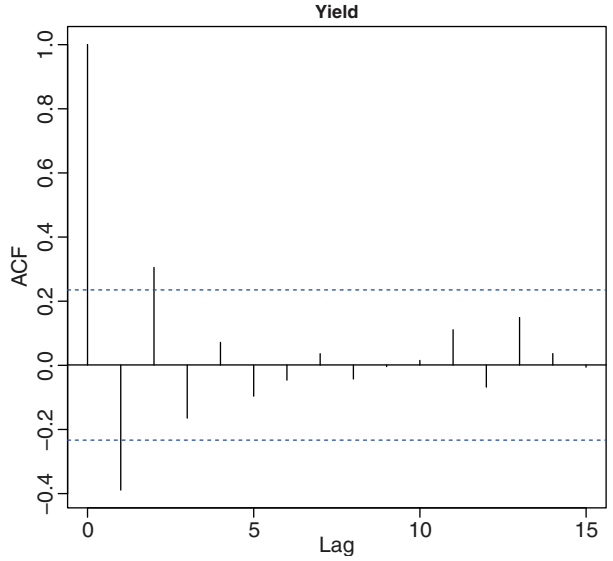


FIGURE 2.7 Estimated autocorrelation function of batch data.

For example, if $\rho_k = \phi^{|k|}$ $(-1 < \phi < 1)$, that is, the autocorrelation function damps out exponentially, (2.1.13) gives

$$\text{var}[r_k] \simeq \frac{1}{N}\left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}\right] \tag{2.1.14}$$

and in particular

$$\text{var}[r_1] \simeq \frac{1}{N}(1-\phi^2)$$

For any process for which all the autocorrelations ρ_v are zero for $v > q$, all terms except the first appearing in the right-hand side of (2.1.13) are zero when $k > q$. Thus, for the variance of the estimated autocorrelation r_k, at lags k greater than some value q beyond which the theoretical autocorrelation function may be deemed to have ''*died out*'', Bartlett's approximation gives

$$\text{var}[r_k] \simeq \frac{1}{N}\left(1 + 2\sum_{v=1}^{q}\rho_v^2\right) \qquad k > q \tag{2.1.15}$$

To use this result in practice, the estimated autocorrelations r_k $(k = 1, 2, \ldots, q)$ are substituted for the theoretical autocorrelations ρ_k, and when this is done, we refer to the square root of (2.1.15) as the *large-lag standard error*. On the assumption that the ρ_k are all zero beyond some lag $k = q$, the large-lag standard error approximates the standard deviation of r_k for suitably large lags $(k > q)$. We will show in Chapter 3 that the moving average (MA) process in (1.2.3) has a correlation structure such that the approximation (2.1.15) applies to this process.

Similar expressions for the approximate covariance between the estimated autocorrelations r_k and r_{k+s} at two different lags k and $k + s$ were also given by Bartlett (1946). In

particular, the large-lag approximation reduces to

$$\text{cov}[r_k, r_{k+s}] \simeq \frac{1}{N} \sum_{v=-q}^{q} \rho_v \rho_{v+s} \qquad k > q \tag{2.1.16}$$

This result shows that care is required in the interpretation of individual autocorrelations because large covariances can exist between neighboring values. This effect can sometimes distort the visual appearance of the sample autocorrelation function, which may fail to damp out according to expectation.

A case of particular interest occurs for $q = 0$, that is, when the ρ_k are taken to be zero for all lags (other than lag 0), and hence the series is completely random or white noise. Then, the standard errors from (2.1.15) for estimated autocorrelations r_k take the simple form

$$\text{se}[r_k] \simeq \frac{1}{\sqrt{N}} \qquad k > 0$$

In addition, in this case the result in (2.1.16) indicates that estimated autocorrelations r_k and r_{k+s} at two different lags are not correlated, and since the r_k are also known to be approximately normally distributed for large N, a collection of estimated autocorrelations for different lags will tend to be independently and normally distributed with mean 0 and variance $1/N$.

Two standard error limits determined under the assumption that the series is completely random are included for the autocorrelation function of the batch data in Figure 2.7. Since N equals 70 in this case, the two standard errors limits are around ± 0.24. The magnitude of the estimated autocorrelation coefficients are clearly inconsistent with the assumption that the series is white noise.

Example. For further illustration, assume that the following estimated autocorrelations were obtained from a time series of length $N = 200$ observations, generated from a stochastic process for which it was *known* that $\rho_1 = -0.4$ and $\rho_k = 0$ for $k \geq 2$:

k	r_k	k	r_k
1	−0.38	6	0.00
2	−0.08	7	0.00
3	0.11	8	0.00
4	−0.08	9	0.07
5	0.02	10	−0.08

On the assumption that the series is completely random, that is, white noise, we have $q = 0$. Then, for *all lags*, (2.1.15) yields

$$\text{var}[r_k] \simeq \frac{1}{N} = \frac{1}{200} = 0.005$$

The corresponding standard error is $0.07 = (0.005)^{1/2}$. Since the value of -0.38 for r_1 is over five times this standard error, it can be concluded that ρ_1 is nonzero. Moreover, the estimated autocorrelations for lags greater than 1 are all small. Therefore, it might be reasonable to ask next whether the series was compatible with a hypothesis (whose relevance will be discussed later) whereby ρ_1 was nonzero, but $\rho_k = 0$ $(k \geq 2)$. Using

(2.1.15) with $q = 1$ and substituting r_1 for ρ_1, the estimated large-lag variance under this assumption is

$$\text{var}[r_k] \simeq \frac{1}{200}[1 + 2(-0.38)^2] = 0.0064 \qquad k > 1$$

yielding a standard error of 0.08. Since the estimated autocorrelations for lags greater than 1 are small compared with this standard error, there is no reason to doubt the adequacy of the model $\rho_1 \neq 0$, $\rho_k = 0$ $(k \geq 2)$.

Remark. The limits shown in Figure 2.7, which assume that the series is white noise, are generated by default in R. Alternative limits, consistent with the assumptions underlying (2.1.15), can be obtained by adding the argument ci.type="ma" to the acf() command.

2.2 SPECTRAL PROPERTIES OF STATIONARY MODELS

2.2.1 Periodogram of a Time Series

Another way of analyzing a time series is based on the assumption that it is made up of sine and cosine waves with different frequencies. A device that uses this idea, introduced by Schuster (1898), is the *periodogram*. The periodogram was originally used to detect and estimate the amplitude of a sine component, of known frequency, buried in noise. We will use it later to provide a check on the randomness of a series (usually, a series of residuals after fitting a particular model), where we consider the possibility that periodic components of unknown frequency may remain in the series.

To illustrate the calculation of the periodogram, suppose that the number of observations $N = 2q + 1$ is odd. We consider fitting the Fourier series model

$$z_t = \alpha_0 + \sum_{i=1}^{q}(\alpha_i c_{it} + \beta_i s_{it}) + e_t \tag{2.2.1}$$

where $c_{it} = \cos(2\pi f_i t)$, $s_{it} = \sin(2\pi f_i t)$, and $f_i = i/N$, which is the ith harmonic of the fundamental frequency $1/N$ associated with the ith sine wave component in (2.2.1) with frequency f_i and period $1/f_i = N/i$. The least squares estimates of the coefficients α_0 and (α_i, β_i) will be

$$a_0 = \bar{z} \tag{2.2.2}$$

$$a_i = \frac{2}{N}\sum_{t=1}^{N} z_t c_{it} \tag{2.2.3}$$

$$i = 1, 2, \ldots, q$$

$$b_i = \frac{2}{N}\sum_{t=1}^{N} z_t s_{it} \tag{2.2.4}$$

since $\sum_{t=1}^{N} c_{it}^2 = \sum_{t=1}^{N} s_{it}^2 = N/2$, and all terms in (2.2.1) are mutually orthogonal over $t = 1, \ldots, N$. The periodogram then consists of the $q = (N-1)/2$ values

$$I(f_i) = \frac{N}{2}(a_i^2 + b_i^2) \qquad i = 1, 2, \ldots, q \tag{2.2.5}$$

where $I(f_i)$ is called the *intensity* at frequency f_i. When N is even, we set $N = 2q$ and (2.2.2)–(2.2.5) apply for $i = 1, 2, \ldots, (q-1)$, but

$$a_q = \frac{1}{N}\sum_{t=1}^{N}(-1)^t z_t$$
$$b_q = 0$$

and

$$I(f_q) = I(0.5) = N a_q^2$$

Note that the highest frequency is 0.5 cycle per time interval because the smallest period is two intervals.

2.2.2 Analysis of Variance

In an analysis of variance table associated with the fitted regression (2.2.1), when N is odd, we can isolate $(N-1)/2$ pairs of degrees of freedom, after eliminating the mean. These are associated with the pairs of coefficients $(a_1, b_1), (a_2, b_2), \ldots, (a_q, b_q)$, and hence with the frequencies $1/N, 2/N, \ldots, q/N$. The periodogram $I(f_i) = (N/2)(a_i^2 + b_i^2)$ is seen to be simply the ''sum of squares'' associated with the pair of coefficients (a_i, b_i) and hence with the frequency $f_i = i/N$ or period $p_i = N/i$. Thus,

$$\sum_{t=1}^{n}(z_t - \bar{z})^2 = \sum_{i=1}^{q} I(f_i) \tag{2.2.6}$$

When N is even, there are $(N-2)/2$ pairs of degrees of freedom and a further single degree of freedom associated with the coefficient a_q.

If the series were truly random, containing no systematic sinusoidal component, that is,

$$z_t = \alpha_0 + e_t$$

with α_0 the fixed mean, and the e's independent and normal, with mean zero and variance σ^2, each component $I(f_i)$ would have expectation $2\sigma^2$ and would be distributed[1] as $\sigma^2\chi^2(2)$, independently of all other components. By contrast, if the series contained a systematic sine component having frequency f_i, amplitude A, and phase angle F, so that

$$z_t = \alpha_0 + \alpha\cos(2\pi f_i t) + \beta\sin(2\pi f_i t) + e_t$$

with $A\sin F = \alpha$ and $A\cos F = \beta$, the sum of squares $I(f_i)$ would tend to be inflated since its expected value would be $2\sigma^2 + N(\alpha^2 + \beta^2)/2 = 2\sigma^2 + NA^2/2$.

In practice, it is unlikely that the frequency f of an unknown systematic sine component would exactly match any of the frequencies f_i for which intensities have been calculated. In this case the periodogram would show an increase in the intensities in the immediate vicinity of f.

[1]It is to be understood that $\chi^2(m)$ refers to a random variable having a chi-square distribution with m degrees of freedom, defined explicitly, for example, in Appendix A7.1.

TABLE 2.2 Mean Monthly Temperatures for Central England in 1964

t	z_t	c_{1t}	t	z_t	c_{1t}
1	3.4	0.87	7	16.1	−0.87
2	4.5	0.50	8	15.5	−0.50
3	4.3	0.00	9	14.1	0.00
4	8.7	−0.50	10	8.9	0.50
5	13.3	−0.87	11	7.4	0.87
6	13.8	−1.00	12	3.6	1.00

Example. A large number of observations would generally be used in calculation of the periodogram. However, to illustrate the details of the calculation, we use the set of 12 mean monthly temperatures (in degrees Celsius) for central England during 1964, given in Table 2.2. The table gives $c_{it} = \cos(2\pi t/12)$, which is required in the calculation of a_1, obtained from

$$\begin{aligned} a_1 &= \frac{1}{6}[(3.4)(0.87) + \cdots + (3.6)(1.00)] \\ &= -5.30 \end{aligned}$$

The values of the a_i, b_i, $i = 1, 2, \ldots, 6$, are given in Table 2.3 and yield the analysis of variance of Table 2.4. As would be expected, the major component of these temperature data has a period of 12 months, that is, a frequency of 1/12 cycle per month.

2.2.3 Spectrum and Spectral Density Function

For completeness, we add here a brief discussion of the spectrum and spectral density function. The use of these important tools is described more fully by Jenkins and Watts (1968), Bloomfield (2000), and Shumway and Stoffer (2011, Chapter 4), among others. We do not apply them to the analysis of time series in this book, and this section can be omitted on first reading.

Sample Spectrum. The definition of the periodogram in (2.2.5) assumes that the frequencies $f_i = i/N$ are harmonics of the fundamental frequency $1/N$. By way of introduction to the spectrum, we relax this assumption and allow the frequency f to vary continuously

TABLE 2.3 Amplitudes of Sines and Cosines at Different Harmonics for Temperature Data

i	a_i	b_i
1	−5.30	−3.82
2	0.05	0.17
3	0.10	0.50
4	0.52	−0.52
5	0.09	−0.58
6	−0.30	

TABLE 2.4 Analysis of Variance Table for Temperature Data

	Frequency				
i	f_i	Period	Periodogram $I(f_i)$	Degrees of Freedom	Mean Square
1	1/12	12	254.96	2	127.48
2	1/6	6	0.19	2	0.10
3	1/4	4	1.56	2	0.78
4	1/3	3	3.22	2	1.61
5	5/12	12/5	2.09	2	1.05
6	1/2	2	1.08	1	1.08
			263.10	11	23.92

in the range of 0–0.5 cycle. The definition (2.2.5) of the periodogram may be modified to

$$I(f) = \frac{N}{2}(a_f^2 + b_f^2) \qquad 0 \le f \le \frac{1}{2} \tag{2.2.7}$$

and $I(f)$ is then referred to as the *sample spectrum* (Jenkins and Watts, 1968). Like the periodogram, it can be used to detect and estimate the amplitude of a sinusoidal component of unknown frequency f buried in noise and is, indeed, a more appropriate tool for this purpose if it is known that the frequency f is not harmonically related to the length of the series. Moreover, it provides a starting point for the theory of spectral analysis, using a result given in Appendix A2.1. This result shows that the sample spectrum $I(f)$ and the estimate c_k of the autocovariance function are linked by the important relation

$$I(f) = 2\left[c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi f k)\right] \qquad 0 \le f \le \frac{1}{2} \tag{2.2.8}$$

That is, the sample spectrum is the Fourier cosine transform of the estimate of the autocovariance function.

Spectrum. The periodogram and sample spectrum are appropriate tools for analyzing time series made up of mixtures of sine and cosine waves, at *fixed* frequencies buried in noise. However, stationary time series of the kind described in Section 2.1 are characterized by random changes of frequency, amplitude, and phase. For this type of series, the sample spectrum $I(f)$ fluctuates wildly and is not capable of any meaningful interpretation.

However, suppose that the sample spectrum was calculated for a time series of N observations, which is a realization of a stationary normal process. As already mentioned, such a process would not have any cosine or sine deterministic components, but we could formally carry through the Fourier analysis and obtain values of (a_f, b_f) for any given frequency f. If repeated realizations of N observations were taken from the stochastic process, we could build up a population of values for a_f, b_f, and $I(f)$. Thus, we could calculate the mean value of $I(f)$ in repeated realizations of size N, namely,

$$E[I(f)] = 2\left[E[c_0] + 2\sum_{k=1}^{N-1} E[c_k]\cos(2\pi f k)\right] \tag{2.2.9}$$

For large N, it may be shown (e.g., Jenkins and Watts, 1968) that the mean value of the estimate c_k of the autocovariance coefficient in repeated realizations tends to the theoretical autocovariance γ_k, that is,

$$\lim_{N\to\infty} E[c_k] = \gamma_k$$

On taking the limit of (2.2.9) as N tends to infinity, the *power spectrum* $p(f)$ is defined by

$$p(f) = \lim_{N\to\infty} E[I(f)] = 2\left[\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \cos(2\pi f k)\right] \qquad 0 \le f \le \frac{1}{2} \tag{2.2.10}$$

We note that since

$$\begin{aligned} |p(f)| &\le 2\left[|\gamma_0| + 2\sum_{k=1}^{\infty}|\gamma_k||\cos(2\pi f k)|\right] \\ &\le 2\left(|\gamma_0| + 2\sum_{k=1}^{\infty}|\gamma_k|\right) \end{aligned} \tag{2.2.11}$$

a sufficient condition for the spectrum to converge is that γ_k damps out rapidly enough for the series (2.2.11) to converge. *Since the power spectrum is the Fourier cosine transform of the autocovariance function*, knowledge of the autocovariance function is mathematically equivalent to knowledge of the spectrum, and vice versa. From now on, we refer to the power spectrum as simply the spectrum.

On integrating (2.2.10) between the limits 0 and $\frac{1}{2}$, the variance of the process z_t is

$$\gamma_0 = \sigma_z^2 = \int_0^{1/2} p(f)\, df \tag{2.2.12}$$

Hence, in the same way that the periodogram $I(f)$ shows how the variance (2.2.6) of a series, consisting of mixtures of sines and cosines, is distributed between the various distinct harmonic frequencies, the spectrum $p(f)$ shows how the variance of a stochastic process is distributed between a continuous range of frequencies. One can interpret $p(f)\,df$ as measuring approximately the variance of the process in the frequency range of f to $f + df$. In addition, from the definition in (2.2.10), the spectral representation for the autocovariance function $\{\gamma_k\}$ can be obtained as

$$\gamma_k = \int_0^{1/2} \cos(2\pi f k) p(f)\, df$$

which together with (2.2.10) directly exhibits the one-to-one correspondence between the power spectrum and the autocovariance function of a process. Conversely, since the γ_k form a positive-definite sequence, provided the series (2.2.11) converges, it follows from Herglotz's theorem (see, e.g., Loève, 1977) that a unique function $p(f)$ exists such that γ_k have the spectral representation $\gamma_k = \frac{1}{2}\int_{-1/2}^{1/2} e^{i2\pi f k} p(f)\, df$. Consequently, the power spectrum $p(f)$ of a stationary process, for which (2.2.11) converges, can be defined as this unique function, which is guaranteed to exist and must have the form of the right-hand side of (2.2.10) by the spectral representation.

The fundamental property of the spectrum that $p(f) \geq 0$ for all $0 \leq f \leq \frac{1}{2}$ follows from $I(f) \geq 0$ and the definition in (2.2.10). In fact, a function $p(f)$ defined on $0 \leq f \leq \frac{1}{2}$ can be the spectrum of a stationary process if and only if it satisfies $p(f) \geq 0$ for $0 \leq f \leq \frac{1}{2}$ and $\int_0^{1/2} p(f)\, df \leq \infty$. Conversely, a sequence $\{\gamma_k\}_{k=0}^{\infty}$ can be the autocovariance function of a stationary process if and only if $\{\gamma_k\}$ is a nonnegative-definite sequence, and this is equivalent to the condition that $p(f) \geq 0$, $0 \leq f \leq \frac{1}{2}$, with $p(f)$ defined by (2.2.10).

Spectral Density Function. It is sometimes more convenient to base the definition (2.2.10) of the spectrum on the autocorrelations ρ_k rather than on the autocovariances γ_k. The resulting function

$$\begin{aligned} g(f) &= \frac{p(f)}{\sigma_z^2} \\ &= 2\left[1 + 2\sum_{k=1}^{\infty} \rho_k \cos(2\pi f k)\right] \qquad 0 \leq f \leq \frac{1}{2} \end{aligned} \tag{2.2.13}$$

is called the *spectral density function*. Using (2.2.12), it is seen that the spectral density function has the property

$$\int_0^{1/2} g(f)\, df = 1$$

Since $g(f)$ is also positive, it has the same properties as an ordinary probability density function. This analogy extends to the estimation properties of these two functions, as we discuss next.

Estimation of the Spectrum. One would expect that an estimate of the spectrum could be obtained from (2.2.10), by replacing the theoretical autocovariances γ_k with their estimates c_k. Because of (2.2.8), this corresponds to taking the sample spectrum as an estimate of $p(f)$. However, it can be shown (e.g., Jenkins and Watts, 1968) that the sample spectrum of a stationary time series fluctuates violently about the theoretical spectrum. An intuitive explanation of this fact is that the sample spectrum corresponds to using an interval, in the frequency domain, whose width is too small. This is analogous to using too small a group interval for the histogram when estimating an ordinary probability distribution. By using a modified or *smoothed* estimate

$$\hat{p}(f) = 2\left[c_0 + 2\sum_{k=1}^{N-1} \lambda_k c_k \cos(2\pi f k)\right] \tag{2.2.14}$$

where the λ_k are suitably chosen weights called a *lag window*, it is possible to increase the *bandwidth* of the estimate and to obtain a smoother estimate of the spectrum. The weights λ_k in (2.2.14) are typically chosen so that they die out to zero for lags $k > M$, where M is known as the truncation point and $M < N$ is moderately small in relation to series length N. As an alternative computational form, one can also obtain an estimate of the spectrum smoother than the sample spectrum $I(f)$ by forming a weighted average of a number of periodogram values $I(f_{i+j})$ in a small neighborhood of frequencies around a

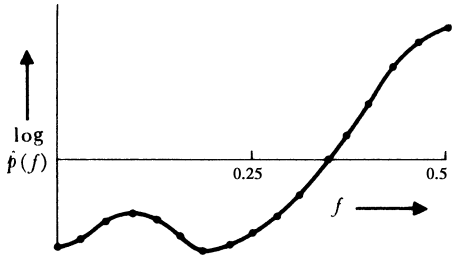


FIGURE 2.8 Estimated power spectrum of batch data.

given frequency f_i. Specifically, a smoothed periodogram estimator of $p(f_i)$ takes the form

$$\hat{p}(f_i) = \sum_{j=-m}^{m} W(f_j) I\left(f_i + \frac{j}{N}\right)$$

where $\sum_{j=-m}^{m} W(f_j) = 1$, the symmetric weighting function $W(f_i)$ is referred to as the *spectral window*, and m is chosen to be much smaller than $N/2$.

Figure 2.8 shows an estimate of the spectrum of the batch data. It is seen that most of the variance of the series is concentrated at high frequencies. This is due to the rapid oscillations in the original series, shown in Figure 2.1.

Remark. The command spectrum() can be used to estimate the power spectrum in R. To use this command, a smoothing window must be specified; see help(spectrum) and the references therein for details. The following command will generate a graph roughly similar to Figure 2.8:

```
spectrum(Yield,spans=c(7,7),taper=0)
```

As an alternative, the R program spec.ar() fits an autoregressive model of order p to the series and computes the spectral density of the fitted model. The lag order p is selected using a model selection criterion such as the AIC to be discussed in Chapter 6.

2.2.4 Simple Examples of Autocorrelation and Spectral Density Functions

For illustration, we now show equivalent representations of two simple stationary stochastic processes based on:

1. Their theoretical models
2. Their theoretical autocorrelation functions
3. Their theoretical spectra

Consider the two processes

$$z_t = 10 + a_t + a_{t-1} \qquad z_t = 10 + a_t - a_{t-1}$$

where $a_t, a_{t-1}, \ldots$ are a sequence of uncorrelated normal random variables with mean zero and variance σ_a^2, that is, Gaussian white noise. From the result in Section 2.1.3 on

stationarity of linear functions, it is clear that the two processes above are stationary. Using the definition (2.1.5),

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)]$$

where $E[z_t] = E[z_{t+k}] = \mu = 10$, and the autocovariances of these two stochastic processes are obtained from

$$\begin{aligned}\gamma_k &= \text{cov}[a_t + a_{t-1}, a_{t+k} + a_{t+k-1}] \\ &= \text{cov}[a_t, a_{t+k}] + \text{cov}[a_t, a_{t+k-1}] + \text{cov}[a_{t-1}, a_{t+k}] + \text{cov}[a_{t-1}, a_{t+k-1}]\end{aligned}$$

and $\gamma_k = \text{cov}[a_t - a_{t-1}, a_{t+k} - a_{t+k-1}]$, respectively. Hence, the autocovariances are

$$\gamma_k = \begin{cases} 2\sigma_a^2 & k = 0 \\ \sigma_a^2 & k = 1 \\ 0 & k \geq 2 \end{cases} \qquad \gamma_k = \begin{cases} 2\sigma_a^2 & k = 0 \\ -\sigma_a^2 & k = 1 \\ 0 & k \geq 2 \end{cases}$$

Thus, the theoretical autocorrelation functions are

$$\rho_k = \begin{cases} 0.5 & k = 1 \\ 0.0 & k \geq 2 \end{cases} \qquad \rho_k = \begin{cases} -0.5 & k = 1 \\ 0.0 & k \geq 2 \end{cases}$$

and using (2.2.13), the theoretical spectral density functions are

$$g(f) = 2[1 + \cos(2\pi f)] \qquad g(f) = 2[1 - \cos(2\pi f)]$$

The autocorrelation functions and spectral density functions are plotted in Figure 2.9 together with a sample time series from each process.

1. It should be noted that for these two stationary processes, knowledge of either the autocorrelation function or the spectral density function, with the mean and variance of the process, is equivalent to knowledge of the model (given the normality assumption).
2. It will be seen that the autocorrelation function reflects one aspect of the behavior of the series. The comparatively smooth nature of the first series is accounted for by the positive association between successive values. The alternating tendency of the second series, in which positive deviations usually follow negative ones, is accounted for by the negative association between successive values.
3. The spectral density throws light on a different but equivalent aspect. The predominance of low frequencies in the first series and high frequencies in the second is shown by the spectra.

Remark. The two models considered in Figure 2.9 are special cases of the moving average model defined in (1.2.3). Specifically, the models are first-order moving average, or MA(1), models with parameters $\theta = -1$ and $\theta = +1$, respectively. As such, they are also special cases of the more general autoregressive integrated moving average (ARIMA) model defined in (1.2.7), where the order now is (0, 0, 1). Figure 2.9 was generated in R by taking advantage of special functions for simulating ARIMA processes and for computing the

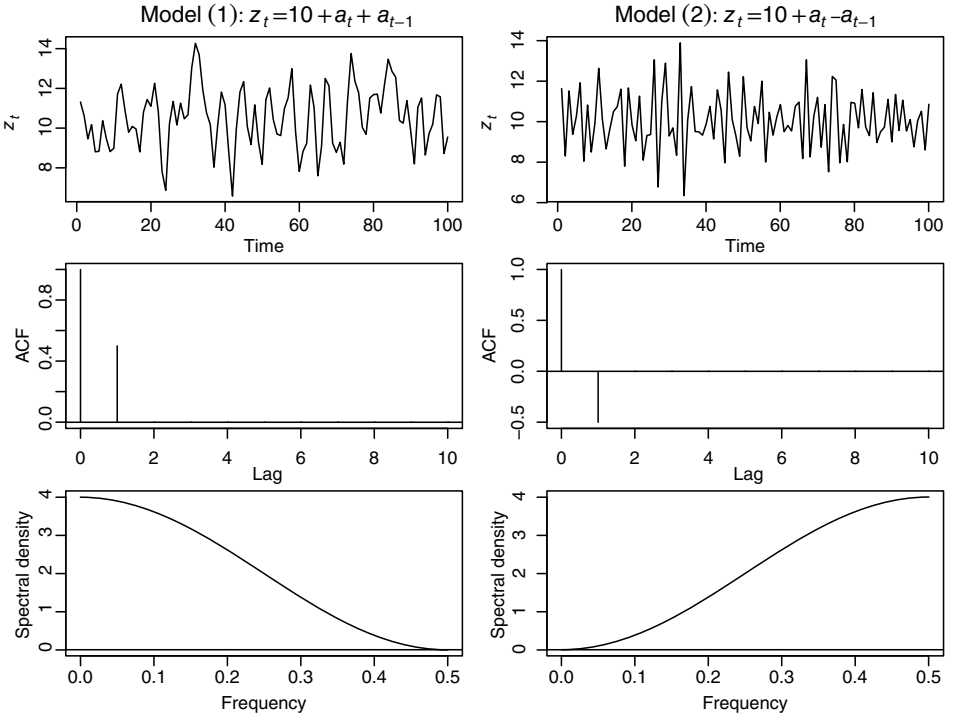


FIGURE 2.9 Two simple stochastic models with their corresponding theoretical autocorrelation functions and spectral density functions.

theoretical autocorrelation function and power spectrum for these processes. The function arima.sim() simulates a time series from a specified model, while ARMAacf() computes its theoretical autocorrelation. Both functions are available in the stats library of R. The TSA library includes a function ARMAspec() that computes and plots the theoretical spectrum of an autoregressive–moving average (ARMA) process. The commands used to generate Figure 2.9 are given below. Note, however, that the MA parameters are entered as +1 and −1, since R uses a definition that has positive signs of the MA parameters in (1.2.3).

```
> library(TSA)
> set.seed(12345)
> par(mfrow=c(3,2)) % Specifies panels in three rows and two columns
> plot(10+arima.sim(list(order=c(0,0,1), ma = +1.0), n=100),ylab =
  expression(z[t]),main=(expression(Model~(1):z[t] == 10+a[t]+a[t-1])))
> plot(10+arima.sim(list(order=c(0,0,1), ma = -1.0), n=100),ylab =
  expression(z[t]),main=(expression(Model~(2):z[t] == 10+a[t]-a[t-1])))
> plot(ARMAacf(ar=0,ma=1.0,10),type="h",x=(0:10),xlab="lag",ylab="ACF")
> abline(h=0)
> plot(ARMAacf(ar=0,ma=-1.0,10),type="h",x=(0:10),xlab="lag",ylab="ACF")
> abline(h=0)
> ARMAspec(model=list(ma=1.0),freq=seq(0,0.5,0.001),plot=TRUE)
> ARMAspec(model=list(ma=-1.0),freq=seq(0,0.5,0.001),plot=TRUE)
```

2.2.5 Advantages and Disadvantages of the Autocorrelation and Spectral Density Functions

Because the autocorrelation function and the spectrum are transforms of each other, they are mathematically equivalent, and therefore any discussion of their advantages and disadvantages turns not on mathematical questions but on the representational value. Because, as we have seen, each sheds light on a different aspect of the data, they should be regarded not as rivals but as allies. Each contributes something to an understanding of the stochastic process in question.

The obtaining of sample estimates of the autocorrelation function and of the spectrum are nonstructural approaches, analogous to the representation of an empirical distribution function by a histogram. They are both ways of letting data from stationary series ''speak for themselves'' and provide a first step in the analysis of time series, just as a histogram can provide a first step in the distributional analysis of data, pointing the way to some parametric model on which subsequent analysis will be based.

Parametric time series models such as those of Section 2.2.4, are not necessarily associated with a simple autocorrelation function or a simple spectrum. Working with either of these nonstructural methods, we may be involved in the estimation of many lag correlations and many spectral ordinates, even when a parametric model containing only one or two parameters could represent the data. Each correlation and each spectral ordinate is a parameter to be estimated, so that these nonstructural approaches might be very prodigal with parameters, when the approach via the model could be parsimonious. On the other hand, initially, we probably do not know what type of model may be appropriate, and initial use of one or the other of these nonstructural approaches is necessary to *identify* the type of model that is needed (in the same way that plotting a histogram helps to indicate which family of distributions may be appropriate). The choice between the spectrum and the autocorrelation function as a tool in model building depends upon the nature of the models that turn out to be practically useful. The models that we have found useful, which we consider in later chapters of this book, are simply described in terms of the autocorrelation function, and it is this tool that we will employ for model specification.

APPENDIX A2.1 LINK BETWEEN THE SAMPLE SPECTRUM AND AUTOCOVARIANCE FUNCTION ESTIMATE

Here, we derive the result (2.2.8):

$$I(f) = 2\left[c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi f k)\right] \qquad 0 \le f \le \frac{1}{2}$$

which links the sample spectrum $I(f)$ and the estimate c_k of the autocovariance function. Suppose that the least square estimates a_f and b_f of the cosine and sine components, at frequency f, in a series are combined according to $d_f = a_f - ib_f$, where $i = -\sqrt{-1}$; then

$$\begin{aligned} I(f) &= \frac{N}{2}(a_f - ib_f)(a_f + ib_f) \\ &= \frac{N}{2} d_f d_f^* \end{aligned} \tag{A2.1.1}$$

where d_f^* is the complex conjugate of d_f. Then, using (2.2.3) and (2.2.4), we obtain

$$\begin{aligned} d(f) &= \frac{2}{N}\sum_{t=1}^{N} z_t[\cos(2\pi ft) - i\sin(2\pi ft)] \\ &= \frac{2}{N}\sum_{t=1}^{N} z_t e^{-i2\pi ft} \\ &= \frac{2}{N}\sum_{t=1}^{N}(z_t - \bar{z})e^{-i2\pi ft} \end{aligned} \tag{A2.1.2}$$

Substituting (A2.1.2) in (A2.1.1) yields

$$I(f) = \frac{2}{N}\sum_{t=1}^{N}\sum_{t'=1}^{N}(z_t - \bar{z})(z_{t'} - \bar{z})e^{-i2\pi f(t-t')} \tag{A2.1.3}$$

Since

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z})$$

the transformation $k = t - t'$ transforms (A2.1.3) into the following required result:

$$\begin{aligned} I(f) &= 2\sum_{k=-N+1}^{N-1} c_k e^{-i2\pi fk} \\ &= 2\left[c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi fk)\right] \qquad 0 \le f \le \frac{1}{2} \end{aligned}$$

EXERCISES

2.1. The following are temperature measurements z_t made every minute on a chemical reactor:

200, 202, 208, 204, 204, 207, 207, 204, 202, 199, 201, 198, 200,
202, 203, 205, 207, 211, 204, 206, 203, 203, 201, 198, 200, 206,
207, 206, 200, 203, 203, 200, 200, 195, 202, 204.207, 206, 200

(a) Plot the time series.
(b) Plot z_{t+1} versus z_t.
(c) Plot z_{t+2} versus z_t.
After inspecting the graphs, do you think that the series is autocorrelated?

2.2. State whether or not a stationary stochastic process can have the following values of autocorrelations:

(a) $\rho_1 = 0.80,\ \rho_2 = 0.55,\ \rho_k = 0,$ for $k > 2$

(b) $\rho_1 = 0.80,\ \rho_2 = 0.28,\ \rho_k = 0,$ for $k > 2$

2.3. Two stationary stochastic processes z_{1t} and z_{2t} have the following autocovariance functions:

$$z_{1t}: \quad \gamma_0 = 0.5,\ \gamma_1 = 0.2,\ \gamma_j = 0 \quad (j \geq 2)$$
$$z_{2t}: \quad \gamma_0 = 2.30,\ \gamma_1 = -1.43,\ \gamma_2 = 0.30,\ \gamma_j = 0 \quad (j \geq 3)$$

Calculate the autocovariance function of the process $z_{3t} = z_{1t} + 2z_{2t}$ and verify that it is a valid stationary process.

2.4. Calculate $c_0, c_1, c_2, c_3, r_1, r_2, r_3$ for the series given in Exercise 2.1. Make a graph of $r_k, k = 0, 1, 2, 3$.

2.5. On the assumption that $\rho_j = 0$ for $j > 2$, obtain the following:

(a) Approximate standard errors for r_1, r_2, and $r_j, j > 2$.

(b) The approximate correlation between r_4 and r_5.

2.6. The annual sales of mink furs by a North American company during 1911–1950 are included as Series N in Part Five of this book. The series is also available at http://pages.stat.wisc.edu/ reinsel/bjr-data/.

(a) Plot the time series using R. Calculate and plot the sample autocorrelation function of the series.

(b) Repeat the analysis in part (a) for the logarithm of the series. Do you see an advantage in using the log transformation in this case?

2.7. Repeat the calculations in Exercise 2.6 for the annual sunspot series given as Series E in Part Five of this book. Use a square root transformation of the data in part (b) in Exercise 2.6. (*Note:* This series is also available for a slightly longer time period as series sunspot.year in the datasets package of R).

2.8. Calculate and plot the theoretical autocorrelation function and the spectral density function for the AR(1) process $z_t = 0.95z_{t-1} + a_t$. (*Hint:* See the R code provided for Figure 2.9). Based on the results, how would you expect a time series generated from this model to fluctuate relative to its mean?

2.9. Calculate and plot the theoretical autocorrelation function and the spectral density function for the AR(2) process $z_t + 0.35z_{t-1} - 0.20z_{t-2} = a_t$.

2.10. Simulate a time series of length $N = 300$ from the AR(2) model specified in Exercise 2.9 and plot the resulting series.

(a) Estimate and plot the autocorrelation function for the simulated series. Compare the results with the theoretical autocorrelation function derived in Exercise 2.9.

(b) Repeat the calculations performed above for a series of length $N = 70$ generated from the same process and compare the results with those for $N = 200$.

(c) Do the estimated autocorrelation functions derived above show any similarity to autocorrelation function of the chemical yield series shown in Figure 2.7. If so, what would you conclude?

2.11. Using the data of Exercise 2.1, calculate the periodogram for periods 36, 18, 12, 9, 36/5, and 6 and construct an analysis of variance table showing the mean squares associated with these periods and the residual mean square.

2.12. A *circular* stationary stochastic process with period N is defined by $z_t = z_{t+N}$.

(a) Show that (see, e.g., Brockwell and Davis, 1991; Fuller, 1996; Jenkins and Watts, 1968) when $N = 2n$, the latent roots of the $N \times N$ autocorrelation matrix of z_t are

$$\lambda_k = 1 + 2\sum_{i=1}^{n-1} \rho_i \cos\left(\frac{\pi i k}{n}\right) + \rho_n \cos(\pi k)$$

$k = 1, 2, \ldots, N$ and the latent vectors corresponding to λ_k, λ_{N-k} (with $\lambda_k = \lambda_{N-k}$) are

$$\ell'_k = \left(\cos\left(\frac{\pi k}{n}\right), \cos\left(\frac{2\pi k}{n}\right), \ldots, \cos(2\pi k)\right)$$
$$\ell'_{N-k} = \left(\sin\left(\frac{\pi k}{n}\right), \sin\left(\frac{2\pi k}{n}\right), \ldots, \sin(2\pi k)\right)$$

(b) Verify that as N tends to infinity, with k/N fixed, λ_k tends to $g(k/N)/2$, where $g(f)$ is the spectral density function, showing that in the limit the latent roots of the autocorrelation matrix trace out the spectral curve.

3

LINEAR STATIONARY MODELS

In this chapter, we describe a general linear stochastic model that assumes that the time series is generated by a linear aggregation of random shocks. For practical representation, it is desirable to employ models that use parameters parsimoniously. Parsimony may often be achieved by representation of the linear process in terms of a small number of autoregressive–moving average (ARMA) terms. The properties of the resulting ARMA models are discussed in preparation for their use in model building in subsequent chapters.

3.1 GENERAL LINEAR PROCESS

3.1.1 Two Equivalent Forms for the Linear Process

In Section 1.2.1, we discussed the representation of a stochastic process as the output from a linear filter, whose input is white noise a_t, that is,

$$\begin{aligned} \tilde{z}_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \end{aligned} \tag{3.1.1}$$

where $\tilde{z}_t = z_t - \mu$ is the deviation of the process from some origin, or from its mean, if the process is stationary. The *general linear process* (3.1.1) allows us to represent $\tilde{z}_t$ as a weighted sum of present and past values of the ''white noise'' process a_t. Important early references on the development of linear stochastic models include Yule (1927), Walker (1931), Slutsky (1937), Wold (1938), Kendall (1945), Bartlett (1946), Quenouille (1952, 1957), Doob (1953), Grenander and Rosenblatt (1957), Hannan (1960), Robinson (1967),

Time Series Analysis: Forecasting and Control, Fifth Edition. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

among others. The usefulness of these models is well-documented in subsequent literature. The white noise process a_t may be regarded as a *series of shocks* that drive the system. It consists of a sequence of uncorrelated random variables with mean zero and constant variance, that is,

$$E[a_t] = 0 \qquad \text{var}[a_t] = \sigma_a^2$$

Since the random variables a_t are assumed uncorrelated, it follows that their autocovariance function is

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{3.1.2}$$

Thus, the autocorrelation function of white noise has a particularly simple form

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{3.1.3}$$

A fundamental result in the development of stationary processes is that of Wold (1938), who established that any zero-mean purely nondeterministic stationary process $\tilde{z}_t$ possesses a linear representation as in (3.1.1) with $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. The a_t are uncorrelated with common variance σ_a^2 but *need not be independent*. We will reserve the term *linear processes* for processes $\tilde{z}_t$ of the form of (3.1.1) in which the a_t are independent random variables.

For $\tilde{z}_t$ defined by (3.1.1) to represent a valid stationary process, it is necessary for the coefficients ψ_j to be *absolutely summable*, that is, for $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Under suitable conditions (see Koopmans, 1974, p. 254), $\tilde{z}_t$ is also a weighted sum of past $\tilde{z}_t$'s and an added shock a_t, that is,

$$\begin{aligned} \tilde{z}_t &= \pi_1 \tilde{z}_{t-1} + \pi_2 \tilde{z}_{t-2} + \cdots + a_t \\ &= \sum_{j=1}^{\infty} \pi_j \tilde{z}_{t-j} + a_t \end{aligned} \tag{3.1.4}$$

In this alternative form, the current deviation $\tilde{z}_t$ from the level μ may be thought of as being ''regressed'' on past deviations $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots$ of the process.

Relationships between the ψ Weights and the π Weights. The relationships between the ψ weights and the π weights may be obtained by using the previously defined *backward shift operator* B, such that

$$Bz_t = z_{t-1} \quad \text{and hence} \quad B^j z_t = z_{t-j}$$

Later, we will also need to use the forward shift operator $F = B^{-1}$, such that

$$Fz_t = z_{t+1} \quad \text{and} \quad F^j z_t = z_{t+j}$$

As an example of the use of the operator B, consider the following model

$$\tilde{z}_t = a_t - \theta a_{t-1} = (1 - \theta B) a_t$$

in which $\psi_1 = -\theta$, $\psi_j = 0$ for $j > 1$. Expressing a_t in terms of the $\tilde{z}_t$'s, we obtain

$$(1 - \theta B)^{-1}\tilde{z}_t = a_t$$

Hence, for $|\theta| < 1$,

$$(1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \cdots)\tilde{z}_t = a_t$$

and the deviation $\tilde{z}_t$ expressed in terms of previous deviations, as in (3.1.4), is

$$\tilde{z}_t = -\theta\tilde{z}_{t-1} - \theta^2\tilde{z}_{t-2} - \theta^3\tilde{z}_{t-3} - \cdots + a_t$$

so that for this model, $\pi_j = -\theta^j$.

Using the backshift operator B, the model (3.1.1) can be written as

$$\tilde{z}_t = \left(1 + \sum_{j-1}^{\infty} \psi_j B^j\right) a_t$$

or

$$\tilde{z}_t = \psi(B)a_t \tag{3.1.5}$$

where

$$\psi(B) = 1 + \sum_{j=1}^{\infty} \psi_j B^j = \sum_{j=0}^{\infty} \psi_j B^j$$

with $\psi_0 = 1$. As mentioned in Section 1.2.1, $\psi(B)$ is called the *transfer function* of the linear filter relating $\tilde{z}_t$ to a_t. It can be regarded as the *generating function* of the ψ weights, with B now treated simply as a variable whose jth power is the coefficient of ψ_j.

Similarly, (3.1.4) may be written as

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right)\tilde{z}_t = a_t$$

or

$$\pi(B)\tilde{z}_t = a_t \tag{3.1.6}$$

Thus,

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

is the generating function of the π weights. After operating on both sides of this expression by $\psi(B)$, we obtain

$$\psi(B)\pi(B)\tilde{z}_t = \psi(B)a_t = \tilde{z}_t$$

Hence, $\psi(B)\pi(B) = 1$, so that

$$\pi(B) = \psi^{-1}(B) \tag{3.1.7}$$

This relationship may be used to derive the π weights, knowing the ψ weights, and vice versa.

3.1.2 Autocovariance Generating Function of a Linear Process

A basic data analysis tool for identifying models in Chapter 6 will be the autocorrelation function. Therefore, it is important to know the autocorrelation function of a linear process. It is shown in Appendix A3.1 that the autocovariance function of the linear process (3.1.1) is given by

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \tag{3.1.8}$$

In particular, by setting $k = 0$, we find that its variance is

$$\gamma_0 = \sigma_z^2 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 \tag{3.1.9}$$

It follows that the stationarity condition of absolute summability of the coefficients ψ_j, $\sum_{j=0}^{\infty} |\psi_j| < \infty$, implies that the series on the right of this equation converges, and hence guarantees that the process will have a finite variance.

Another way of obtaining the autocovariances of a linear process is via the *autocovariance generating function*

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k \tag{3.1.10}$$

where γ_0, the variance of the process, is the coefficient of $B^0 = 1$, while γ_k, the autocovariance of lag k, is the coefficient of both B^j and $B^{-j} = F^j$. It is shown in Appendix A3.1 that

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 \psi(B)\psi(F) \tag{3.1.11}$$

For example, suppose that $\tilde{z}_t = a_t - \theta a_{t-1} = (1 - \theta B)a_t$ so that $\psi(B) = (1 - \theta B)$. Then,

$$\begin{aligned} \gamma(B) &= \sigma_a^2 (1 - \theta B)(1 - \theta B^{-1}) \\ &= \sigma_a^2 [-\theta B^{-1} + (1 + \theta^2) - \theta B] \end{aligned}$$

Comparing with (3.1.10), the autocovariances are

$$\begin{aligned} \gamma_0 &= (1 + \theta^2)\sigma_a^2 \\ \gamma_1 &= -\theta \sigma_a^2 \\ \gamma_k &= 0 \qquad k \geq 2 \end{aligned}$$

In the development that follows, when treated as a variable in a generating function, B will be able to take on complex values. In particular, it will often be necessary to consider the different cases when $|B| < 1$, $|B| = 1$, or $|B| > 1$, that is, when the complex number B lies inside, on, or outside the unit circle.

3.1.3 Stationarity and Invertibility Conditions for a Linear Process

Stationarity. The convergence of the series (3.1.9) ensures that the process has a finite variance. Also, we have seen in Section 2.1.3 that the autocovariances and autocorrelations must satisfy a set of conditions to ensure stationarity. For a linear process (3.1.1), these conditions are guaranteed by the single condition that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. This condition can also be embodied in the condition that the series $\psi(B)$, which is the generating function of the ψ weights, must converge for $|B| \leq 1$, that is, on or within the unit circle. This result is discussed in Appendix A3.1.

Spectrum of a Linear Stationary Process. It is shown in Appendix A3.1 that if we substitute $B = e^{-i2\pi f}$, where $i = \sqrt{-1}$, in the autocovariance generating function (3.1.11), we obtain one half of the power spectrum. Thus, the spectrum of a linear process is

$$\begin{aligned} p(f) &= 2\sigma_a^2 \psi(e^{-i2\pi f})\psi(e^{i2\pi f}) \\ &= 2\sigma_a^2 |\psi(e^{-i2\pi f})|^2 \qquad 0 \leq f \leq \frac{1}{2} \end{aligned} \tag{3.1.12}$$

In fact, this is the well-known expression (e.g., Jenkins and Watts, 1968) that relates the spectrum $p(f)$ of the output from a linear system to the uniform spectrum $2\sigma_a^2$ of a white noise input by multiplying it with the squared gain $G^2(f) = |\psi(e^{-i2\pi f})|^2$ of the system.

Invertibility. We have seen that the ψ weights of a linear process must satisfy the condition that $\psi(B)$ converges on or within the unit circle if the process is to be stationary. We now consider a similar restriction applied to the π weights to ensure what is called *invertibility*. This invertibility condition is independent of the stationarity condition and is also applicable to the nonstationary linear models, which we introduce in Chapter 4.

To illustrate the basic idea of invertibility, consider again the special case

$$\tilde{z}_t = (1 - \theta B)a_t \tag{3.1.13}$$

Expressing the a_t's in terms of the present and past $\tilde{z}_t$'s, this model becomes

$$a_t = (1 - \theta B)^{-1}\tilde{z}_t = (1 + \theta B + \theta^2 B^2 + \cdots + \theta^k B^k)(1 - \theta^{k+1} B^{k+1})^{-1}\tilde{z}_t$$

that is,

$$\tilde{z}_t = -\theta\tilde{z}_{t-1} - \theta^2\tilde{z}_{t-2} - \cdots - \theta^k\tilde{z}_{t-k} + a_t - \theta^{k+1}a_{t-k-1} \tag{3.1.14}$$

If $|\theta| < 1$, on letting k tend to infinity, we obtain the infinite series

$$\tilde{z}_t = -\theta\tilde{z}_{t-1} - \theta^2\tilde{z}_{t-2} - \cdots + a_t \tag{3.1.15}$$

and the π weights of the model in the form of (3.1.4) are $\pi_j = -\theta^j$. Whatever the value of θ, $\tilde{z}_t = (1 - \theta B)a_t$ defines a perfectly proper stationary process. However, if $|\theta| \geq 1$, the current deviation $\tilde{z}_t$ in (3.1.14) depends on $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots, \tilde{z}_{t-k}$, with weights that increase as k increases. We avoid this situation by requiring that $|\theta| < 1$. We then say that the series is *invertible.* We see that this condition is equivalent to $\sum_{j=0}^{\infty} |\theta|^j \equiv \sum_{j=0}^{\infty} |\pi_j| < \infty$, so

that the series

$$\pi(B) = (1 - \theta B)^{-1} = \sum_{j=0}^{\infty} \theta^j B^j$$

converges for all $|B| \leq 1$, that is, on or within the unit circle. The invertibility requirement is needed to associate present events with *past* values in a sensible manner.

The general linear process (3.1.1) is invertible and can be written in the form

$$\pi(B)\tilde{z}_t = a_t$$

if the weights π_j are absolutely summable, that is, if $\sum_{j=0}^{\infty} |\pi_j| < \infty$, which implies that the series $\pi(B)$ converges on or within the unit circle.

Thus, to summarize, a linear process (3.1.1) is *stationary* if $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and is *invertible* if $\sum_{j=0}^{\infty} |\pi_j| < \infty$, where $\pi(B) = \psi^{-1}(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$.

3.1.4 Autoregressive and Moving Average Processes

The representations (3.1.1) and (3.1.4) of the general linear process would not be very useful in practice if they contained an infinite number of parameters ψ_j and π_j. We now describe a way to introduce parsimony and arrive at models that are representationally useful for practical applications.

Autoregressive Processes. Consider first the special case of (3.1.4) in which only the first p of the weights are nonzero. The model may be written as

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{3.1.16}$$

where we now use the symbols $\phi_1, \phi_2, \ldots, \phi_p$ for the *finite* set of weight parameters. The resulting process is called an *autoregressive* process of order p, or more succinctly, an AR(p) process. In particular, the AR(1) and AR(2) models

$$\begin{aligned} \tilde{z}_t &= \phi_1 \tilde{z}_{t-1} + a_t \\ &= \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \end{aligned}$$

are of considerable practical importance.

The AR(p) model can be written in the equivalent form

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\tilde{z}_t = a_t$$

or

$$\phi(B)\tilde{z}_t = a_t \tag{3.1.17}$$

This implies that

$$\tilde{z}_t = \frac{1}{\phi(B)} a_t = \phi^{-1}(B) a_t \equiv \psi(B) a_t$$

Hence, the autoregressive process can be thought of as the output $\tilde{z}_t$ from a linear filter with transfer function $\phi^{-1}(B) = \psi(B)$ when the input is white noise a_t.

Moving Average Processes. Next consider the special case of (3.1.1), when only the first q of the ψ weights are nonzero. The process may be written as

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{3.1.18}$$

where we now use the symbols $-\theta_1, -\theta_2, \ldots, -\theta_q$ for the *finite* set of weight parameters. This process is called a *moving average* process[1] of order q, which we often abbreviate as MA(q). The special cases of MA(1) and MA(2) models

$$\begin{aligned}\tilde{z}_t &= a_t - \theta_1 a_{t-1} \\ &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}\end{aligned}$$

are again particularly important in practice.

Using the backshift operator $Ba_t = a_{t-1}$, the MA(q) model can be written in the equivalent form as

$$\tilde{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t$$

or more succinctly as

$$\tilde{z}_t = \theta(B)a_t \tag{3.1.19}$$

Hence, the moving average process can be thought of as the output $\tilde{z}_t$ from a linear filter with transfer function $\theta(B)$ when the input is white noise a_t.

Mixed Autoregressive–Moving Average Processes. As discussed in Section 3.1.1, the *finite* moving average process

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 B)a_t \qquad |\theta_1| < 1$$

can also be written as an *infinite* autoregressive process

$$\tilde{z}_t = -\theta_1 \tilde{z}_{t-1} - \theta_1^2 \tilde{z}_{t-2} - \cdots + a_t$$

However, if the process really was MA(1), we would not obtain a parsimonious representation using an autoregressive model. Conversely, an AR(1) process could not be parsimoniously represented using a moving average model. In practice, to obtain parsimonious parameterization, it is often useful to include both autoregressive and moving average terms in the model. The resulting model

$$\tilde{z}_1 = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

or

$$\phi(B)\tilde{z}_t = \theta(B)a_t \tag{3.1.20}$$

is called the *mixed autoregressive–moving average* process of order (p, q), which we abbreviate as ARMA(p, q). For example, the ARMA(1, 1) process is

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t - \theta_1 a_{t-1}$$

[1] As we remarked in Chapter 1, the term ''moving average'' is somewhat misleading since the weights do not sum to unity. However, this nomenclature is now well established and we will use it here.

Now writing

$$\begin{aligned}\tilde{z}_t &= \phi^{-1}(B)\theta(B)a_t \\ &= \frac{\theta(B)}{\phi(B)}a_t = \frac{1-\theta_1 B-\cdots-\theta_q B^q}{1-\phi_1 B-\cdots-\phi_p B^p}a_t\end{aligned}$$

we see that the mixed ARMA process can be thought of as the output $\tilde{z}_t$ from a linear filter, whose transfer function is the ratio of two polynomial operators $\theta(B)$ and $\phi(B)$, when the input is white noise a_t. Furthermore, since $\tilde{z}_t = z_t - \mu$, where $\mu = E[z_t]$ is the mean of the process in the stationary case, the general ARMA(p, q) process can also be written in terms of the original process z_t as

$$\phi(B)z_t = \theta_0 + \theta(B)a_t \tag{3.1.21}$$

where the constant term θ_0 is

$$\theta_0 = (1-\phi_1-\phi_2-\cdots-\phi_p)\mu \tag{3.1.22}$$

In the next sections, we discuss some important characteristics of autoregressive, moving average, and mixed models. In particular, we study their variances, autocorrelation functions, spectra, and the stationarity and invertibility conditions that must be imposed on their parameters.

3.2 AUTOREGRESSIVE PROCESSES

3.2.1 Stationarity Conditions for Autoregressive Processes

The parameters $\phi_1, \phi_2, \ldots, \phi_p$ of an AR(p) process

$$\tilde{z}_t = \phi_1\tilde{z}_{t-1} + \cdots + \phi_p\tilde{z}_{t-p} + a_t$$

or

$$(1-\phi_1 B-\cdots-\phi_p B^p)\tilde{z}_t = \phi(B)\tilde{z}_t = a_t$$

must satisfy certain conditions for the process to be stationary. For illustration, the AR(1) process

$$(1-\phi_1 B)\tilde{z}_t = a_t$$

may be written as

$$\tilde{z}_t = (1-\phi_1 B)^{-1}a_t = \sum_{j=0}^{\infty}\phi_1^j a_{t-j}$$

provided that the infinite series on the right converges in an appropriate sense. Hence,

$$\psi(B) = (1-\phi_1 B)^{-1} = \sum_{j=0}^{\infty}\phi_1^j B^j \tag{3.2.1}$$

We have seen in Section 3.1.3 that for stationarity, $\psi(B)$ must converge for $|B| \leq 1$, or equivalently that $\sum_{j=0}^{\infty} |\phi_1|^j < \infty$. This implies that the parameter ϕ_1 of an AR(1) process must satisfy the condition $|\phi_1| < 1$ to ensure stationarity. Since the root of $1 - \phi_1 B = 0$ is $B = \phi_1^{-1}$, this condition is equivalent to saying that the root of $1 - \phi_1 B = 0$ must lie *outside* the unit circle.

The general AR(p) process $\phi(B)\tilde{z}_t = a_t$ can be written as

$$\tilde{z}_t = \phi^{-1}(B)a_t \equiv \psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

provided that the right-side expression is convergent. Using the factorization

$$\phi(B) = (1 - G_1 B)(1 - G_2 B) \cdots (1 - G_p B)$$

where $G_1^{-1}, \ldots, G_p^{-1}$ are the roots of $\phi(B) = 0$, and expanding $\theta^{-1}(B)$ in partial fractions yields

$$\tilde{z}_t = \phi^{-1}(B)a_t = \sum_{i=1}^{p} \frac{K_i}{1 - G_i B} a_t$$

Hence, if $\psi(B) = \phi^{-1}(B)$ is to be a convergent series for $|B| \leq 1$, that is, if the weights $\psi_j = \sum_{i=1}^{p} K_i G_i^j$ are to be absolutely summable so that the AR(p) process is stationary, we must have $|G_i| < 1$, for $i = 1, \ldots, p$. Equivalently, the roots of the $\phi(B) = 0$ must lie *outside* the unit circle. The roots of the equation $\phi(B) = 0$ may be referred to as the zeros of the polynomial $\phi(B)$. Thus, for stationarity, the zeros of $\phi(B)$ must lie *outside* the unit circle. A similar argument may be applied when the zeros of $\phi(B)$ are not all distinct. The equation $\phi(B) = 0$ is called the *characteristic equation* for the process.

Note also that the roots of $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = 0$ are the reciprocals to the roots of the polynomial equation in m,

$$m^p - \phi_1 m^{p-1} - \cdots - \phi_p = 0$$

Hence, the stationarity condition that all roots of $\phi(B) = 0$ must lie outside the unit circle, that is, be greater than 1 in absolute value, is equivalent to the condition that all roots of $m^p - \phi_1 m^{p-1} - \cdots - \phi_p = 0$ must lie *inside* the unit circle, that is, be less than 1 in absolute value.

Since the series $\pi(B) = \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is finite, no restrictions are required on the parameters of an autoregressive process to ensure invertibility.

ψ Weights. Since $\psi(B) = 1/\phi(B)$ so that $\phi(B)\psi(B) = 1$, it readily follows that the weights ψ_j for the AR(p) process satisfy the difference equation

$$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \cdots + \phi_p \psi_{j-p} \qquad j > 0$$

with $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$, from which the weights ψ_j can easily be computed recursively in terms of the ϕ_i. In fact, as seen from the principles of linear difference equations as discussed in Appendix A4.1, the fact that the weights ψ_j satisfy the difference equation discussed earlier implies that they have an explicit representation in the form of $\psi_j = \sum_{i=1}^{p} K_i G_i^j$ for the case of distinct roots.

3.2.2 Autocorrelation Function and Spectrum of Autoregressive Processes

Autocorrelation Function. An important recurrence relation for the autocorrelation function of a stationary autoregressive process is found by multiplying throughout in

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + a_t$$

by $\tilde{z}_{t-k}$, for $k \geq 0$, to obtain

$$\tilde{z}_{t-k}\tilde{z}_t = \phi_1 \tilde{z}_{t-k}\tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-k}\tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-k}\tilde{z}_{t-p} + \tilde{z}_{t-k}a_t \tag{3.2.2}$$

Now, on taking expected values, we obtain the difference equation

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} \qquad k > 0 \tag{3.2.3}$$

Note that the expectation $E[\tilde{z}_{t-k}a_t]$ is zero for $k > 0$, since $\tilde{z}_{t-k}$ can only involve the shocks a_j up to time $t-k$, which are uncorrelated with a_t. On dividing throughout in (3.2.3) by γ_0, we see that the autocorrelation function satisfies the same form of difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \qquad k > 0 \tag{3.2.4}$$

Note that this is analogous to the difference equation satisfied by the process $\tilde{z}_t$ itself, but without the random shock input a_t.

Now suppose that this equation is written as

$$\phi(B)\rho_k = 0$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and B now operates on k and not t. Then, writing

$$\phi(B) = \prod_{i=1}^{p}(1 - G_i B)$$

the general solution for ρ_k in (3.2.4) (see, e.g., Appendix A4.1) is

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k \tag{3.2.5}$$

where $G_1^{-1}, G_2^{-1}, \ldots, G_p^{-1}$ are the roots of the *characteristic equation*

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = 0$$

or equivalently, $G_1, G_2, \ldots, G_p$ are the roots of $m^p - \phi_1 m^{p-1} - \cdots - \phi_p = 0$.

For stationarity, we require that $|G_i| < 1$. Thus, two situations can arise in practice if we assume that the roots G_i are distinct.

1. A root G_i is real, in which case a term $A_i G_i^k$ in (3.2.5) decays to zero geometrically as k increases. We often refer to this as a damped exponential.
2. A pair of roots G_i and G_j are complex conjugates, in which case they contribute a term

   $$D^k \sin(2\pi f k + F)$$

   to the autocorrelation function (3.2.5), which follows a damped sine wave, with damping factor $D = |G_i| = |G_j|$ and frequency f such that $2\pi f = \cos^{-1}[|\mathrm{Re}(G_i)|/D]$.

In general, the autocorrelation function of a stationary autoregressive process will consist of a mixture of damped exponentials and damped sine waves.

Autoregressive Parameters in Terms of the Autocorrelations: Yule–Walker Equations. If we substitute $k = 1, 2, \ldots, p$ in (3.2.4), we obtain a set of linear equations for $\phi_1, \phi_2, \ldots, \phi_p$ in terms of $\rho_1, \rho_2, \ldots, \rho_p$, that is,

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\ &\vdots \quad \vdots \quad \cdots \quad \vdots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p \end{aligned} \tag{3.2.6}$$

These are the well-known *Yule–Walker* equations (Yule, 1927; Walker, 1931). We obtain *Yule--Walker estimates* of the parameters by replacing the theoretical autocorrelations ρ_k by the estimated autocorrelations r_k. Note that, if we write

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad \mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}$$

the solution of (3.2.6) for the parameters $\boldsymbol{\phi}$ in terms of the autocorrelations may be written as

$$\boldsymbol{\phi} = \mathbf{P}_p^{-1}\boldsymbol{\rho}_p \tag{3.2.7}$$

Variance. When $k = 0$, the contribution from the term $E[\tilde{z}_{t-k}a_t]$, on taking expectations in (3.2.2), is $E[a_t^2] = \sigma_a^2$, since the only part of $\tilde{z}_t$ that will be correlated with a_t is the most recent shock, a_t. Hence, when $k = 0$,

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \cdots + \phi_p\gamma_{-p} + \sigma_a^2$$

On substituting $\gamma_{-k} = \gamma_k$ and writing $\gamma_k = \gamma_0\rho_k$, the variance $\gamma_0 = \sigma_z^2$ may be written as

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p} \tag{3.2.8}$$

Spectrum. For the AR(p) process, $\psi(B) = \phi^{-1}(B)$ and

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Therefore, using (3.1.12), the spectrum of an autoregressive process is

$$p(f) = \frac{2\sigma_a^2}{|1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f} - \cdots - \phi_p e^{-i2p\pi f}|^2} \qquad 0 \le f \le \tfrac{1}{2} \tag{3.2.9}$$

We now discuss two particularly important autoregressive processes, those of first and second order.

3.2.3 The First-Order Autoregressive Process

The first-order autoregressive process is

$$\begin{aligned}\tilde{z}_t &= \phi_1 \tilde{z}_{t-1} + a_t \\ &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \end{aligned} \tag{3.2.10}$$

where it has been shown in Section 3.2.1 that ϕ_1 must satisfy the condition $-1 < \phi_1 < 1$ for the process to be stationary.

Autocorrelation Function. Using (3.2.4), the autocorrelation function satisfies the first-order difference equation

$$\rho_k = \phi_1 \rho_{k-1} \qquad k > 0 \tag{3.2.11}$$

which, with $\rho_0 = 1$, has the solution

$$\rho_k = \phi_1^k \qquad k \geq 0 \tag{3.2.12}$$

Since $-1 < \phi < 1$, the autocorrelation function decays exponentially to zero when ϕ_1 is positive but decays exponentially to zero and oscillates in sign when ϕ_1 is negative. In particular, we note that

$$\rho_1 = \phi_1 \tag{3.2.13}$$

Variance. Using (3.2.8), the variance of the process is

$$\begin{aligned}\sigma_z^2 &= \frac{\sigma_a^2}{1 - \rho_1 \phi_1} \\ &= \frac{\sigma_a^2}{1 - \phi_1^2} \end{aligned} \tag{3.2.14}$$

on substituting $\rho_1 = \phi_1$

Spectrum. Finally, using (3.2.9), the spectrum is

$$\begin{aligned} p(f) &= \frac{2\sigma_a^2}{|1 - \phi_1 e^{-i2\pi f}|^2} \\ &= \frac{2\sigma_a^2}{1 + \phi_1^2 - 2\phi_1 \cos(2\pi f)} \qquad 0 \leq f \leq \tfrac{1}{2} \end{aligned} \tag{3.2.15}$$

Example. Figure 3.1 shows realizations from two AR(1) processes with $\phi_1 = 0.8$ and $\phi_1 = -0.8$, and the corresponding theoretical autocorrelation functions and spectra. Thus, when the parameter has the large positive value $\phi_1 = 0.8$, neighboring values in the series are similar and the series exhibits marked trends. This is reflected in the autocorrelation function, which slowly decays to zero, and in the spectrum, which is dominated by low frequencies. When the parameter has the large negative value $\phi_1 = -0.8$, the series tends to oscillate rapidly, and this is reflected in the autocorrelation function, which alternates

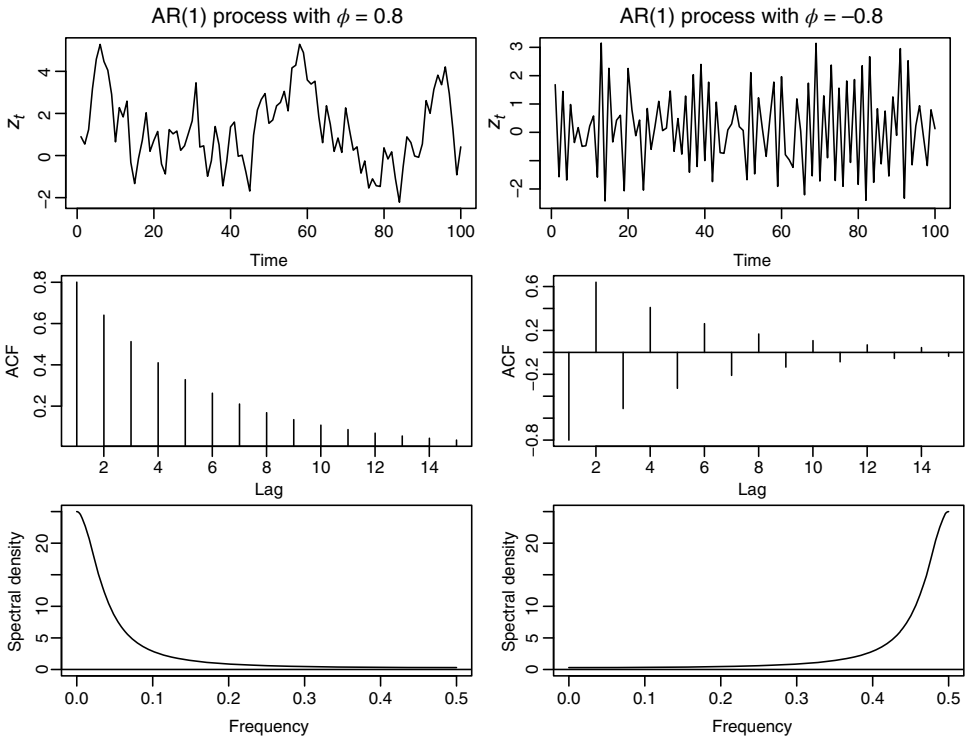


FIGURE 3.1 Realizations from two first-order autoregressive processes and their corresponding theoretical autocorrelation functions and spectral density functions.

in sign as it decays to zero, and in the spectrum, which is dominated by high frequencies. Figure 3.1 was generated in R and can be reproduced as follows:

```
>library(TSA)
>set.seed(12345)
>par(mfrow=c(3,2))
>plot(arima.sim(list(order=c(1,0,0),ar = 0.8), n=100),ylab=
 expression(z[t]),main=expression("AR(1) process with "*phi*"=0.8"))
>plot(arima.sim(list(order=c(1,0,0),ar = -0.8), n=100), ylab=
 expression(z[t]),main=expression("AR(1) process with "*phi*"=-0.8"))
>plot(ARMAacf(ar=0.8,ma=0,15)[-1],type="h",ylab="ACF",xlab="lag")
>abline(h=0)
>plot(ARMAacf(ar=-0.8,ma=0,15)[-1],type="h",ylab="ACF",xlab="lag")
>abline(h=0)
>ARMAspec(model=list(ar=0.8),freq=seq(0,0.5,0.001),plot=TRUE)
>ARMAspec(model=list(ar=-0.8),freq=seq(0,0.5,0.001),plot=TRUE)
```

3.2.4 Second-Order Autoregressive Process

Stationarity Condition. The second-order autoregressive process can be written as

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \tag{3.2.16}$$

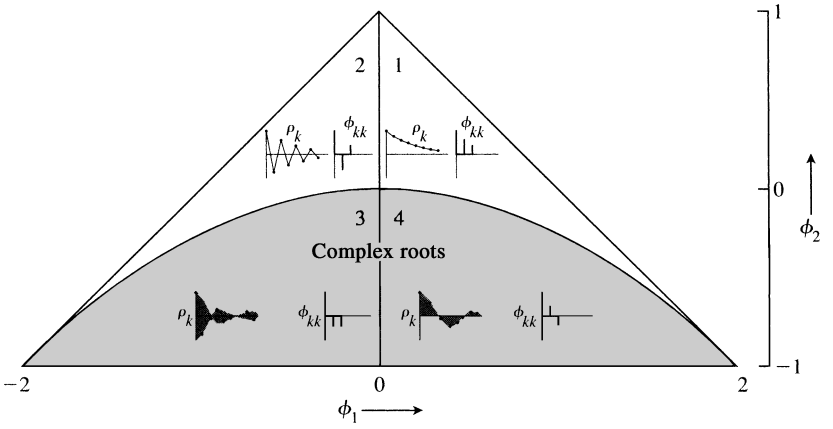


FIGURE 3.2 Typical autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various stationary AR(2) models (*Source:* Stralkowski, 1968).

For stationarity, the roots of

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0 \tag{3.2.17}$$

must lie outside the unit circle, which implies that the parameters ϕ_1 and ϕ_2 must lie in the triangular region

$$\begin{aligned} \phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 < \phi_2 &< 1 \end{aligned} \tag{3.2.18}$$

as shown in Figure 3.2.

Autocorrelation Function. Using (3.2.4), the autocorrelation function satisfies the second-order difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \qquad k > 0 \tag{3.2.19}$$

with starting values $\rho_0 = 1$ and $\rho_1 = \phi_1/(1 - \phi_2)$. From (3.2.5), the general solution to this difference equation is

$$\begin{aligned} \rho_k &= A_1 G_1^k + A_2 G_2^k \\ &= \frac{G_1(1 - G_2^2)G_1^k - G_2(1 - G_1^2)G_2^k}{(G_1 - G_2)(1 + G_1 G_2)} \end{aligned} \tag{3.2.20}$$

where G_1^{-1} and G_2^{-1} are the roots of the characteristic equation $\phi(B) = 0$. When the roots are real, the autocorrelation function consists of a mixture of damped exponentials. This occurs when $\phi_1^2 + 4\phi_2 \geq 0$ and corresponds to regions 1 and 2, which lie above the parabolic boundary in Figure 3.2. Specifically, in region 1, the autocorrelation function remains positive as it damps out, corresponding to a positive dominant root in (3.2.20). In region 2, the autocorrelation function alternates in sign as it damps out, corresponding to a negative dominant root.

If the roots G_1 and G_2 are complex ($\phi_1^2 + 4\phi_2 < 0$), a second-order autoregressive process displays *pseudoperiodic behavior*. This behavior is reflected in the autocorrelation function, for on substituting $G_1 = De^{i2\pi f_0}$ and $G_2 = De^{-i2\pi f_0}$ ($0 < f_0 < \frac{1}{2}$) in (3.2.20), we obtain

$$\rho_k = \frac{D^k \sin(2\pi f_0 k + F)}{\sin F} \tag{3.2.21}$$

We refer to this as a *damped sine wave* with damping factor D, *frequency* f_0, and *phase* F. These factors are related to the process parameters as follows:

$$D = |G_i| = \sqrt{-\phi_2} \tag{3.2.22}$$

where the positive square root is taken,

$$\cos(2\pi f_0) = \frac{\mathrm{Re}(G_i)}{D} = \frac{\phi_1}{2\sqrt{-\phi_2}} \tag{3.2.23}$$

$$\tan F = \frac{1 + D^2}{1 - D^2} \tan(2\pi f_0) \tag{3.2.24}$$

Again referring to Figure 3.2, the autocorrelation function is a damped sine wave in regions 3 and 4, the phase angle F being less than 90° in region 4 and lying between 90° and 180° in region 3. This means that the autocorrelation function starts with a positive value throughout region 4 but always switches sign from lag 0 to lag 1 in region 3.

Yule–Walker Equations. For the AR(2) model, the Yule–Walker equations become

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \end{aligned} \tag{3.2.25}$$

which, when solved for ϕ_1 and ϕ_2, give

$$\begin{aligned} \phi_1 &= \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \\ \phi_2 &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \end{aligned} \tag{3.2.26}$$

These equations can also be solved to express ρ_1 and ρ_2 in terms of ϕ_1 and ϕ_2 to give

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1 - \phi_2} \\ \rho_2 &= \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \end{aligned} \tag{3.2.27}$$

which provide the starting values for the recursions in (3.2.19). Expressions (3.2.20) and (3.2.21) are useful for explaining the different patterns for ρ_k that may arise in practice. However, for computing the autocorrelations of an AR(2) process, it is simplest to make direct use of the recursions implied by (3.2.19).

Using the stationarity condition (3.2.18) and the expressions for ρ_1 and ρ_2 in (3.2.27), it can be seen that the admissible values of ρ_1 and ρ_2, for a stationary AR(2) process, must

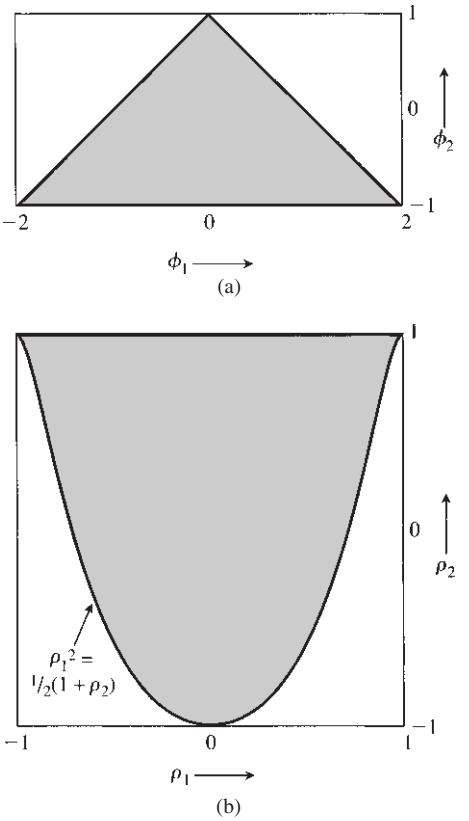


FIGURE 3.3 Admissible regions for (a) ϕ_1, ϕ_2 and (b) ρ_1, ρ_2, for a stationary AR(2) process.

lie in the region

$$\begin{aligned} -1 < \rho_1 < 1 \\ -1 < \rho_2 < 1 \\ \rho_1^2 < \tfrac{1}{2}(\rho_2 + 1) \end{aligned}$$

The admissible region for the parameters ϕ_1 and ϕ_2 is shown in Figure 3.3(a), while Figure 3.3(b) shows the corresponding admissible region for ρ_1 and ρ_2.

Variance. From (3.2.8), the variance of the AR(2) process is

$$\begin{aligned} \sigma_z^2 &= \frac{\sigma_a^2}{1 - \rho_1\phi_1 - \rho_2\phi_2} \\ &= \frac{1 - \phi_2}{1 + \phi_2} \frac{\sigma_a^2}{(1 - \phi_2)^2 - \phi_1^2} \end{aligned} \tag{3.2.28}$$

Spectrum. From (3.2.9), the spectrum is

$$p(f) = \frac{2\sigma_a^2}{|1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f}|^2}$$
$$= \frac{2\sigma_a^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos(2\pi f) - 2\phi_2 \cos(4\pi f)} \qquad 0 \le f \le \tfrac{1}{2} \tag{3.2.29}$$

The spectrum also reflects the pseudoperiodic behavior that the series exhibits when the roots of the characteristic equation are complex. For illustration, Figure 3.4(a) shows 70 values of a series generated from the AR(2) model

$$\tilde{z}_t = 0.75\tilde{z}_{t-1} - 0.50\tilde{z}_{t-2} + a_t$$

Figure 3.4(b) shows the corresponding theoretical autocorrelation function. The roots of the characteristic equation

$$1 - 0.75B + 0.5B^2 = 0$$

are complex, so that the pseudoperiodic behavior observed in the series is to be expected. We clearly see this behavior reflected in the theoretical autocorrelation function of Figure 3.4(b), the average apparent period being about 6. The damping factor D and frequency

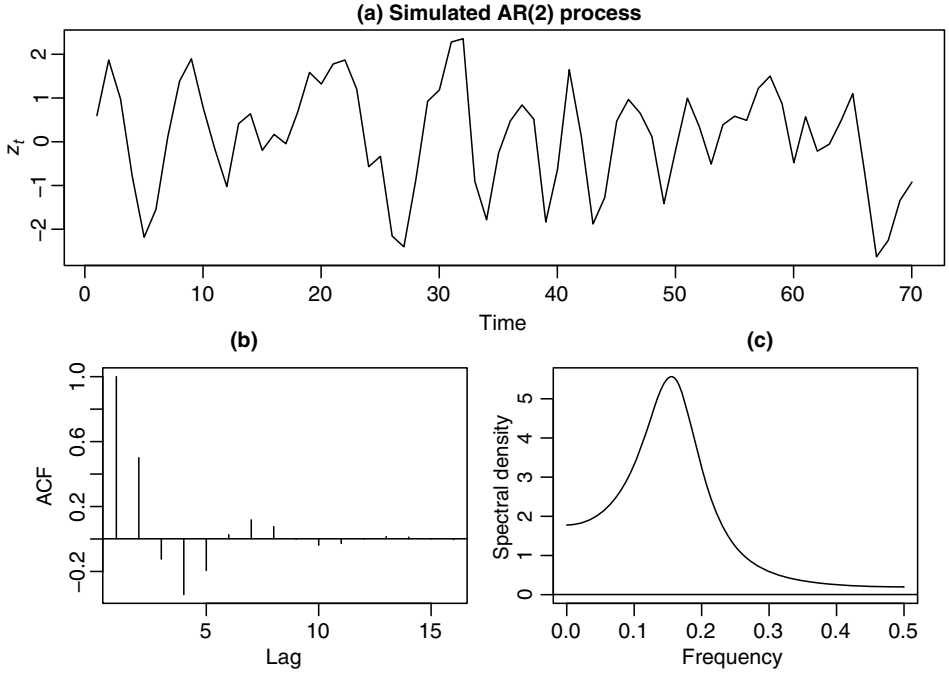


FIGURE 3.4 (a) Time series generated from a second-order autoregressive process $\tilde{z}_t = 0.75\,\tilde{z}_{t-1} - 0.50\,\tilde{z}_{t-2} + a_t$, along with (b) the theoretical autocorrelation function, and (c) the spectral density function for the same process.

f_0, from (3.2.22) and (3.2.23), are

$$D = \sqrt{0.50} = 0.71 \qquad f_0 = \frac{\cos^{-1}(0.5303)}{2\pi} = \frac{1}{6.2}$$

Thus, the fundamental period of the autocorrelation function is 6.2. In addition, the theoretical spectral density function in Figure 3.4(c) shows that a large proportion of the variance of the series is accounted for by frequencies in the neighborhood of f_0.

Figure 3.4 was generated in R using the following commands:

```
> library(TSA)
> ar.acf=ARMAacf(model=list(ar=c(0.75,-0.5)))
> ar.spec=ARMAspec(model=list(ar=c(0.75,-0.5),freq=seq(0,0.5,0.0005)))
> layout(matrix(c(1,1,2,3),2,2,byrow=TRUE))
> plot(arima.sim(list(order=c(2,0,0),ar=c(0.75,-0.5)), n=70), ylab=
      expression(z[t]),xlab="Time",main=("Simulated AR(2) process"))
> plot(ar.acf, main="b")
> plot(ar.spec, main="c")
```

3.2.5 Partial Autocorrelation Function

In practice, we typically do not know the order of the autoregressive process initially, and the order has to be specified from the data. The problem is analogous to deciding on the number of independent variables to be included in a multiple regression. The partial autocorrelation function is a tool that exploits the fact that, whereas an AR(p) process has an autocorrelation function that is infinite in extent, the partial autocorrelations are zero beyond lag p.

The partial autocorrelations can be described in terms of p nonzero *functions* of the autocorrelations. Denote by ϕ_{kj} the jth coefficient in an autoregressive representation of order k, so that ϕ_{kk} is the last coefficient. From (3.2.4), the ϕ_{kj} satisfy the set of equations

$$\rho_j = \phi_{k1}\rho_{j-1} + \cdots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \qquad j = 1, 2, \ldots, k \tag{3.2.30}$$

leading to the Yule–Walker equations (3.2.6), which may be written as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \tag{3.2.31}$$

or

$$\mathbf{P}_k \boldsymbol{\phi}_k = \boldsymbol{\rho}_k \tag{3.2.32}$$

Solving these equations for $k = 1, 2, 3, \ldots$, successively, we obtain

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \tag{3.2.33}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

In general, for ϕ_{kk}, the determinant in the numerator has the same elements as that in the denominator, but with the last column replaced by $\boldsymbol{\rho}_k$. The quantity ϕ_{kk}, regarded as a function of the lag k, is called the *partial autocorrelation* function.

For an AR(p) process, the partial autocorrelations ϕ_{kk} will be nonzero for $k \leq p$ and *zero for* $k > p$. In other words, the partial autocorrelation function of the AR(p) process has a *cutoff* after lag p. For the AR(2) process, partial autocorrelation functions ϕ_{kk} are shown in each of the four regions of Figure 3.2. As a numerical example, the partial autocorrelations of the AR(2) process $\tilde{z}_t = 0.75\tilde{z}_{t-1} - 0.50\tilde{z}_{t-2} + a_t$ considered in Figure 3.4 are $\phi_{11} = \rho_1 = 0.5$, $\phi_{22} = (\rho_2 - \rho_1^2)/(1 - \rho_1^2) = -0.5 \equiv \phi_2$, and $\phi_{kk} = 0$, for all $k > 2$.

The quantity ϕ_{kk} is called the *partial autocorrelation* of the process $\{z_t\}$ at lag k, since it equals the partial correlation between the variables z_t and z_{t-k} adjusted for the intermediate variables $z_{t-1}, z_{t-2}, \ldots, z_{t-k+1}$ (or the correlation between z_t and z_{t-k} not accounted for by $z_{t-1}, z_{t-2}, \ldots, z_{t-k+1}$). Now, it is easy to establish from least squares theory that the values $\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk}$, which are the solutions to (3.2.31), are the regression coefficients in the linear regression of z_t on $z_{t-1}, \ldots, z_{t-k}$, that is, they are the values of coefficients $b_1, \ldots, b_k$, which minimize $E[(z_t - b_0 - \sum_{i=1}^{k} b_i z_{t-i})^2]$. Hence, assuming for convenience that the process $\{z_t\}$ has mean zero, the best linear predictor, in the mean squared error sense, of z_t based on $z_{t-1}, z_{t-2}, \ldots, z_{t-k+1}$ is

$$\hat{z}_t = \phi_{k-1,1} z_{t-1} + \phi_{k-1,2} z_{t-2} + \cdots + \phi_{k-1,k-1} z_{t-k+1}$$

whether the process is an AR or not. Similarly, the best linear predictor of z_{t-k} based on the (future) values $z_{t-1}, z_{t-2}, \ldots, z_{t-k+1}$ is

$$\hat{z}_{t-k} = \phi_{k-1,1} z_{t-k+1} + \phi_{k-1,2} z_{t-k+2} + \cdots + \phi_{k-1,k-1} z_{t-1}$$

Then, the lag k partial autocorrelation of $\{z_t\}$, ϕ_{kk}, can be defined as the correlation between the residuals from these two regressions on $z_{t-1}, \ldots, z_{t-k+1}$, that is,

$$\phi_{kk} = \text{corr}[z_t - \hat{z}_t, z_{t-k} - \hat{z}_{t-k}] \tag{3.2.34}$$

TABLE 3.1 Estimated Partial Autocorrelation Function for the Chemical Yield Data in Figure 2.1

k	$\hat{\phi}_{kk}$	k	$\hat{\phi}_{kk}$	k	$\hat{\phi}_{kk}$
1	−0.39	6	−0.12	11	0.14
2	0.18	7	0.02	12	−0.01
3	0.00	8	0.00	13	0.09
4	−0.04	9	− 0.06	14	0.17
5	−0.07	10	0.00	15	0.00

As examples, we find that $\phi_{11} = \text{corr}[z_t, z_{t-1}] = \rho_1$, while

$$\begin{aligned}\phi_{22} &= \text{corr}[z_t - \rho_1 z_{t-1}, z_{t-2} - \rho_1 z_{t-1}] \\ &= \frac{\gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0}{[(\gamma_0 + \rho_1^2\gamma_0 - 2\rho_1\gamma_1)^2]^{1/2}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}\end{aligned}$$

which agrees with the results in (3.2.33) derived from the Yule–Walker equations. Higher order partial autocorrelations ϕ_{kk} defined through (3.2.34) can similarly be shown to be the solution to the appropriate set of Yule–Walker equations.

3.2.6 Estimation of the Partial Autocorrelation Function

The partial autocorrelations may be estimated by fitting successively autoregressive models of orders $1, 2, 3, \ldots$ by least squares and picking out the estimates $\hat{\phi}_{11}, \hat{\phi}_{22}, \hat{\phi}_{33}, \ldots$ of the last coefficient fitted at each stage. Alternatively, if the values of the parameters are not too close to the nonstationary boundaries, approximate Yule–Walker estimates of the successive autoregressive models may be employed. The estimated partial autocorrelations can then be obtained by substituting estimates r_j for the theoretical autocorrelations in (3.2.30), to yield

$$r_j = \hat{\phi}_{k1} r_{j-1} + \hat{\phi}_{k2} r_{j-2} + \cdots + \hat{\phi}_{k(k-1)} r_{j-k+1} + \hat{\phi}_{kk} r_{j-k} \qquad j = 1, 2, \ldots, k \qquad (3.2.35)$$

and solving the resultant equations for $k = 1, 2, \ldots$. This can be done using a simple recursive method due to Levinson (1947) and Durbin (1960), which we describe in Appendix A3.2. However, these estimates obtained from (3.2.35) become very sensitive to rounding errors and should not be used if the values of the parameters are close to the nonstationary boundaries.

3.2.7 Standard Errors of Partial Autocorrelation Estimates

It was shown by Quenouille (1949) that on the hypothesis that the process is autoregressive of order p, the estimated partial autocorrelations of order $p + 1$, and higher, are approximately independently and normally distributed with zero mean. Also, if n is the number of observations used in fitting,

$$\text{var}[\hat{\phi}_{kk}] \simeq \frac{1}{n} \qquad k \geq p + 1$$

Thus, the standard error (SE) of the estimated partial autocorrelation $\hat{\phi}_{kk}$ is

$$\text{SE}[\hat{\phi}_{kk}] = \hat{\sigma}[\hat{\phi}_{kk}] \simeq \frac{1}{\sqrt{n}} \qquad k \geq p+1 \tag{3.2.36}$$

3.2.8 Calculations in R

The estimation of the partial autocorrelation function is conveniently performed in R. For example, the command pacf(Yield) in the stats package gives the estimated partial autocorrelations shown in Table 3.1 for the chemical yield data plotted in Figure 2.1. An alternative is to use the command acf2() in the R package astsa. This command has the advantage that it produces plots of the autocorrelation and partial autocorrelation functions in a single graph. This allows easy comparison of the two functions, which will be useful for specifying a model for the time series. Figure 3.5 shows a graph of the 15 first autocorrelations and partial autocorrelations for the chemical yield data produced using this routine. The patterns of the two functions resemble those of an AR(1) process with a negative value of ϕ_1, or possibly an AR(2) process with a dominant negative root (see region 2 of Figure 3.2). Also shown in Figure 3.5 by dashed lines are the two SE limits calculated on the assumption that the process is white noise. Since $\hat{\phi}_{22}$ is the second biggest partial autocorrelation, the possibility that the process is AR(2) should be kept in mind.

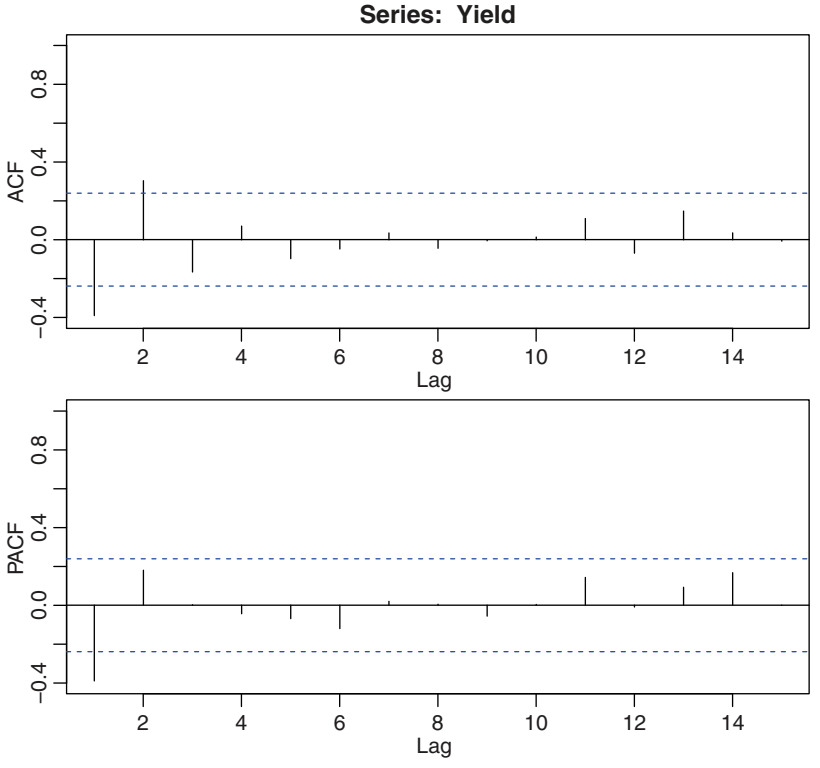


FIGURE 3.5 Estimated autocorrelation and partial autocorrelation functions for the chemical yield data in Series F.

The use of the autocorrelation and partial autocorrelation functions for model specification will be discussed more fully in Chapter 6. Figure 3.5 was generated using the following R commands:

```
> library(astsa)
> seriesF=read.table("SeriesF.txt,header=TRUE)
> Yield=ts(seriesF)
> acf2(Yield,15)
```

3.3 MOVING AVERAGE PROCESSES

3.3.1 Invertibility Conditions for Moving Average Processes

We now derive the conditions that the parameters $\theta_1, \theta_2, \ldots, \theta_q$ must satisfy to ensure the invertibility of the MA(q) process:

$$\begin{aligned}\tilde{z}_t &= a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \\ &= (1 - \theta_1 B - \cdots - \theta_q B^q) a_t \\ &= \theta(B) a_t \end{aligned} \tag{3.3.1}$$

We have already seen in Section 3.1.3 that the first-order moving average process

$$\tilde{z}_t = (1 - \theta_1 B) a_t$$

is invertible if $|\theta_1| < 1$; that is,

$$\pi(B) = (1 - \theta_1 B)^{-1} = \sum_{j=0}^{\infty} \theta_1^j B^j$$

converges on or within the unit circle. However, this is equivalent to saying that the root, $B = \theta_1^{-1}$ of $(1 - \theta_1 B) = 0$, lies *outside* the unit circle.

The invertibility condition for higher order MA processes may be obtained by writing $\tilde{z}_t = \theta(B) a_t$ as

$$a_t = \theta^{-1}(B) \tilde{z}_t$$

Hence, if

$$\theta(B) = \prod_{i=1}^{q} (1 - H_i B)$$

where $H_1^{-1}, \ldots, H_q^{-1}$ are the roots of $\theta(B) = 0$, then, on expanding in partial fractions, we obtain

$$\pi(B) = \theta^{-1}(B) = \sum_{i=1}^{q} \left(\frac{M_i}{1 - H_i B} \right)$$

which converges, or equivalently, the weights $\pi_j = -\sum_{i=1}^{q} M_i H_i^j$ are absolutely summable, if $|H_i| < 1$, for $i = 1, 2, \ldots, q$. It follows that the invertibility condition for

an MA(q) process is that the roots H_i^{-1} of the characteristic equation

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = 0 \tag{3.3.2}$$

lie *outside* the unit circle. From the relation $\theta(B)\pi(B) = 1$, it follows that the weights π_j satisfy the difference equation

$$\pi_j = \theta_1 \pi_{j-1} + \theta_2 \pi_{j-2} + \cdots + \theta_q \pi_{j-q} \qquad j > 0$$

with the convention that $\pi_0 = -1$ and $\pi_j = 0$ for $j < 0$, from which the weights π_j can easily be computed recursively in terms of the θ_i.

Note that since the series

$$\psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

is finite, no restrictions are needed on the parameters of the moving average process to ensure stationarity.

3.3.2 Autocorrelation Function and Spectrum of Moving Average Processes

Autocorrelation Function. The autocovariance function of an MA(q) process is

$$\begin{aligned}\gamma_k &= E[(a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \cdots - \theta_q a_{t-k-q})] \\ &= -\theta_k E[a_{t-k}^2] + \theta_1\theta_{k+1} E[a_{t-k-1}^2] + \cdots + \theta_{q-k}\theta_q E[a_{t-q}^2]\end{aligned}$$

since the a_t are uncorrelated, and $\gamma_k = 0$ for $k > q$. Hence, the variance of the process is

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_a^2 \tag{3.3.3}$$

and

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q)\sigma_a^2 & k = 1, 2, \ldots, q \\ 0 & k > q \end{cases}$$

Thus, the autocorrelation function is

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & k = 1, 2, \ldots, q \\ 0 & k > q \end{cases} \tag{3.3.4}$$

We see that the autocorrelation function of an MA(q) process is zero, beyond the order q of the process. In other words, the autocorrelation function of a moving average process has a *cutoff* after lag q.

Moving Average Parameters in Terms of Autocorrelations. If $\rho_1, \rho_2, \ldots, \rho_q$ are known, the q equations (3.3.4) may be solved for the parameters $\theta_1, \theta_2, \ldots, \theta_q$. However, unlike the Yule–Walker equations (3.2.6) for an autoregressive process, the equations (3.3.4) are nonlinear. Hence, except in the simple case where $q = 1$, which is discussed shortly, these equations have to be solved iteratively. Estimates of the moving average parameters may be obtained by substituting estimates r_k for ρ_k and solving the resulting equations. However, unlike the autoregressive estimates obtained from the Yule–Walker equations,

the resulting moving average estimates may not have high statistical efficiency. Nevertheless, they can provide useful rough estimates at the model identification stage discussed in Chapter 6. Furthermore, they provide useful starting values for an iterative parameter estimation procedure, discussed in Chapter 7, which converges to the efficient maximum likelihood estimates.

Spectrum. For the MA(q) process,

$$\psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

Therefore, using (3.1.12), the spectrum of an MA(q) process is

$$p(f) = 2\sigma_a^2 |1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f} - \cdots - \theta_q e^{-i2q\pi f}|^2 \qquad 0 \le f \le \frac{1}{2} \tag{3.3.5}$$

We now discuss in greater detail the moving average processes of first and second order, which are of considerable practical importance.

3.3.3 First-Order Moving Average Process

We have already introduced the MA(1) process

$$\begin{aligned} \tilde{z}_t &= a_t - \theta_1 a_{t-1} \\ &= (1 - \theta_1 B) a_t \end{aligned}$$

and we have seen that θ_1 must lie in the range $-1 < \theta_1 < 1$ for the process to be invertible. The process is, of course, stationary for all values of θ_1.

Autocorrelation Function. It is easy to see that the variance of this process equals

$$\gamma_0 = (1 + \theta_1^2)\sigma_a^2$$

The autocorrelation function is

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases} \tag{3.3.6}$$

from which it is noted that ρ_1 must satisfy $|\rho_1| = |\theta_1|/(1 + \theta_1^2) \le \frac{1}{2}$. Also, for $k = 1$, we find that

$$\rho_1 \theta_1^2 + \theta_1 + \rho_1 = 0 \tag{3.3.7}$$

with roots for θ_1 equal to $\theta_1 = (-1 \pm \sqrt{1 - 4\rho_1^2})/(2\rho_1)$. Since the product of the roots is unity, we see that if θ_1 is a solution, so is θ_1^{-1}. Furthermore, if θ_1 satisfies the invertibility condition $|\theta_1| < 1$, the other root θ_1^{-1} will be greater than unity and will not satisfy the condition. For example, if $\rho_1 = -0.4$, the two solutions are $\theta_1 = 0.5$ and $\theta_1 = 2.0$. However, only the solution $\theta_1 = 0.5$ corresponds to an invertible model.

Spectrum. Using (3.3.5), the spectrum of the MA(1) process is

$$\begin{aligned} p(f) &= 2\sigma_a^2 |1 - \theta_1 e^{-i2\pi f}|^2 \\ &= 2\sigma_a^2 [1 + \theta_1^2 - 2\theta_1 \cos(2\pi f)] \qquad 0 \le f \le \frac{1}{2} \end{aligned} \tag{3.3.8}$$

In general, when θ_1 is negative, ρ_1 is positive, and the spectrum is dominated by low frequencies. Conversely, when θ_1 is positive, ρ_1 is negative, and the spectrum is dominated by high frequencies.

Partial Autocorrelation Function. Using (3.2.31) with $\rho_1 = -\theta_1/(1 + \theta_1^2)$ and $\rho_k = 0$, for $k > 1$, we obtain after some algebraic manipulation

$$\phi_{kk} = \frac{-\theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}$$

Thus, $|\phi_{kk}| < |\theta_1|^k$, and the partial autocorrelation function is dominated by a damped exponential. If ρ_1 is positive, so that θ_1 is negative, the partial autocorrelations alternate in sign. If, however, ρ_1 is negative, so that θ_1 is positive, the partial autocorrelations are negative. From (3.1.15), it has been seen that the weights π_j for the MA(1) process are $\pi_j = -\theta_1^j$, and hence since these are coefficients in the infinite autoregressive form of the process, it makes sense that the partial autocorrelation function ϕ_{kk} for the MA(1) essentially mimics the exponential decay feature of the weights π_j.

We now note a duality between the AR(1) and the MA(1) processes. Thus, whereas the autocorrelation function of an MA(1) process has a cutoff after lag 1, the autocorrelation function of an AR(1) process tails off exponentially. Conversely, whereas the partial autocorrelation function of an MA(1) process tails off and is dominated by a damped exponential, the partial autocorrelation function of an AR(1) process has a cutoff after lag 1. It turns out that a corresponding approximate duality of this kind occurs in general in the autocorrelation and partial autocorrelation functions between AR and MA processes.

3.3.4 Second-Order Moving Average Process

Invertibility Conditions. The second-order moving average process is defined by

$$\begin{aligned} \tilde{z}_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) a_t \end{aligned}$$

and is stationary for all values of θ_1 and θ_2. However, it is invertible only if the roots of the characteristic equation

$$1 - \theta_1 B - \theta_2 B^2 = 0 \tag{3.3.9}$$

lie outside the unit circle, that is,

$$\begin{aligned} \theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 < \theta_2 &< 1 \end{aligned} \tag{3.3.10}$$

These are parallel to conditions (3.2.18) required for the *stationarity* of an AR(2) process.

Autocorrelation Function. Using (3.3.3), the variance of the process is

$$\gamma_0 = \sigma_a^2(1 + \theta_1^2 + \theta_2^2)$$

and using (3.3.4), the autocorrelation function is

$$\begin{aligned} \rho_1 &= \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2 &= \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \\ \rho_k &= 0 \qquad k > 2 \end{aligned} \tag{3.3.11}$$

Thus, the autocorrelation function has a cutoff after lag 2.

It follows from (3.3.10) and (3.3.11) that the first two autocorrelations of an invertible MA(2) process must lie within the area bounded by segments of the curves

$$\begin{aligned} \rho_2 + \rho_1 &= -0.5 \\ \rho_2 - \rho_1 &= -0.5 \\ \rho_1^2 &= 4\rho_2(1 - 2\rho_2) \end{aligned} \tag{3.3.12}$$

The invertibility region (3.3.10) for the parameters is shown in Figure 3.6(a) and the corresponding admissible region (3.3.12) for the autocorrelations in Figure 3.6(b). The latter shows whether a given pair of autocorrelations ρ_1 and ρ_2 is consistent with the assumption that the model is an MA(2) process. If they are consistent, the values of the parameters θ_1 and θ_2 can be obtained by solving the nonlinear equations (3.3.11). To facilitate this calculation, Chart C in the Collection of Tables and Charts in Part Five has been prepared so that the values of θ_1 and θ_2 can be read off directly, given ρ_1 and ρ_2.

Spectrum. Using (3.3.5), the spectrum of the MA(2) process is

$$\begin{aligned} p(f) &= 2\sigma_a^2|1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f}|^2 \\ &= 2\sigma_a^2[1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1-\theta_2)\cos(2\pi f) - 2\theta_2\cos(4\pi f)] \\ & \qquad\qquad 0 < f < \frac{1}{2} \end{aligned} \tag{3.3.13}$$

and is the reciprocal of the spectrum (3.2.29) of a second-order autoregressive process, apart from the constant $2\sigma_a^2$.

Partial Autocorrelation Function. The exact expression for the partial autocorrelation function of an MA(2) process is complicated, but it is dominated by the sum of two exponentials if the roots of the characteristic equation $1 - \theta_1 B - \theta_2 B^2 = 0$ are real, and by a damped sine wave if the roots are complex. Thus, it behaves like the autocorrelation function of an AR(2) process. The autocorrelation functions and partial autocorrelation functions for various values of the parameters within the invertible region are shown in Figure 3.7. Comparison of Figure 3.7 with Figure 3.2, which shows the corresponding