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

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

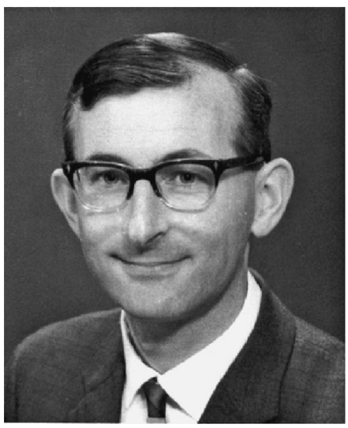
2015015492

Printed in the United States of America

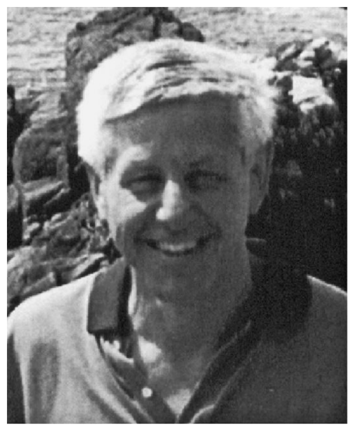
10 9 8 7 6 5 4 3 2 1

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

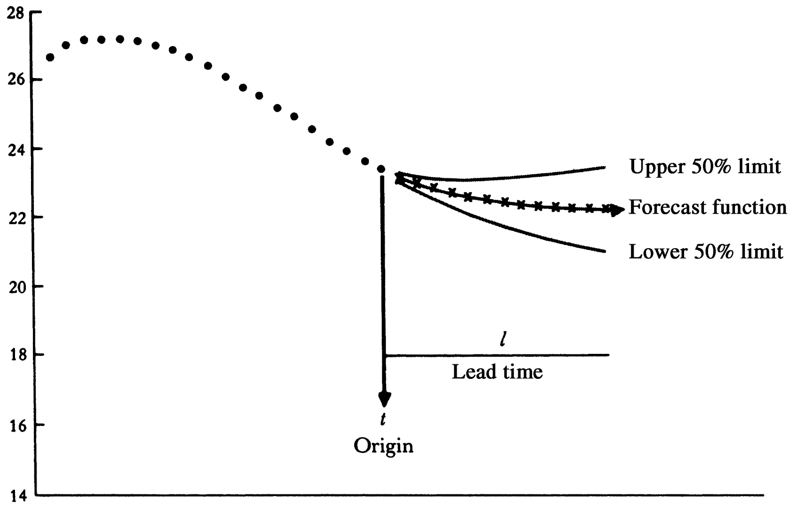


FIGURE 1.1 Values of a time series with forecast function and 50% probability limits.

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

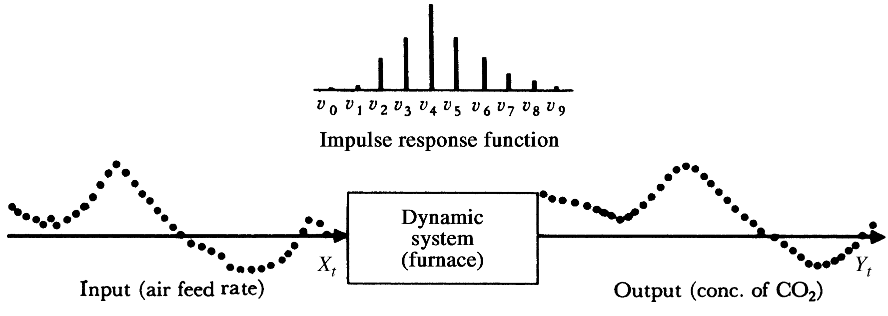


FIGURE 1.2 Input and output time series in relation to a dynamic system.

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.

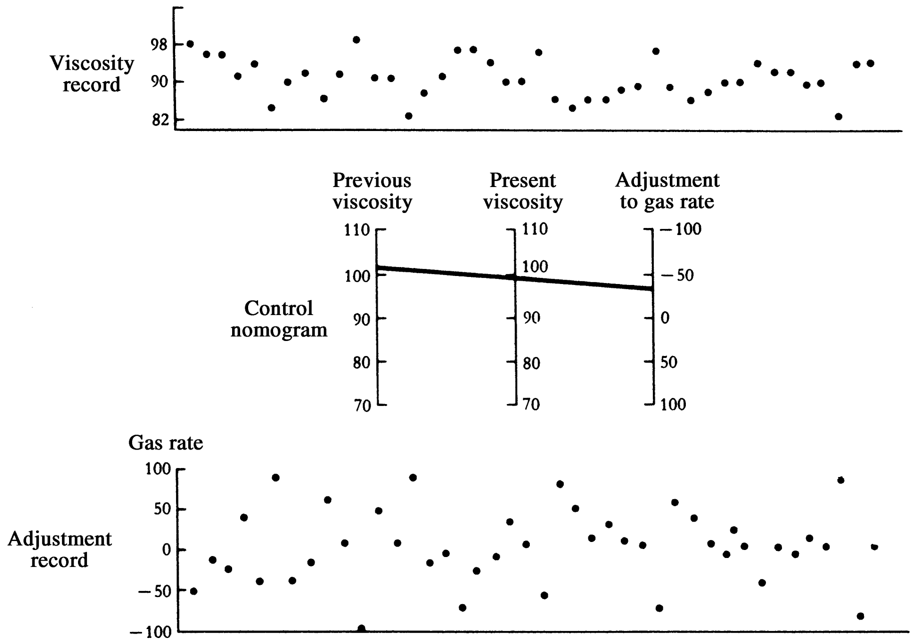


FIGURE 1.3 Control of viscosity. Record of observed viscosity and of adjustments in gas rate made using nomogram.

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

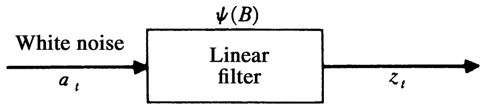


FIGURE 1.4 Representation of a time series as the output from a linear filter.

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 μ, $\phi_1, \phi_2, \ldots, \phi_p, \sigma_a^2$, 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.