

Wiley Series in Probability and Statistics

*Introduction to*

# TIME SERIES ANALYSIS AND FORECASTING

Second Edition



Douglas C. Montgomery  
Cheryl L. Jennings  
Murat Kulahci

**WILEY**



# INTRODUCTION TO TIME SERIES ANALYSIS AND FORECASTING

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.

---

# INTRODUCTION TO TIME SERIES ANALYSIS AND FORECASTING

---

Second Edition

**DOUGLAS C. MONTGOMERY**

Arizona State University  
Tempe, Arizona, USA

**CHERYL L. JENNINGS**

Arizona State University  
Tempe, Arizona, USA

**MURAT KULAHCI**

Technical University of Denmark  
Lyngby, Denmark  
and  
Luleå University of Technology  
Luleå, Sweden

**WILEY**

Copyright © 2016 by John Wiley & Sons, Inc. All rights reserved.

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data applied for.***

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

# CONTENTS

---

<b>PREFACE</b>	<b>xi</b>
<b>1 INTRODUCTION TO FORECASTING</b>	<b>1</b>
1.1 The Nature and Uses of Forecasts / 1	
1.2 Some Examples of Time Series / 6	
1.3 The Forecasting Process / 13	
1.4 Data for Forecasting / 16	
1.4.1 The Data Warehouse / 16	
1.4.2 Data Cleaning / 18	
1.4.3 Imputation / 18	
1.5 Resources for Forecasting / 19	
Exercises / 20	
<b>2 STATISTICS BACKGROUND FOR FORECASTING</b>	<b>25</b>
2.1 Introduction / 25	
2.2 Graphical Displays / 26	
2.2.1 Time Series Plots / 26	
2.2.2 Plotting Smoothed Data / 30	
2.3 Numerical Description of Time Series Data / 33	
2.3.1 Stationary Time Series / 33	

2.3.2	Autocovariance and Autocorrelation Functions / 36
2.3.3	The Variogram / 42
2.4	Use of Data Transformations and Adjustments / 46
2.4.1	Transformations / 46
2.4.2	Trend and Seasonal Adjustments / 48
2.5	General Approach to Time Series Modeling and Forecasting / 61
2.6	Evaluating and Monitoring Forecasting Model Performance / 64
2.6.1	Forecasting Model Evaluation / 64
2.6.2	Choosing Between Competing Models / 74
2.6.3	Monitoring a Forecasting Model / 77
2.7	R Commands for Chapter 2 / 84
	Exercises / 96

### **3 REGRESSION ANALYSIS AND FORECASTING 107**

3.1	Introduction / 107
3.2	Least Squares Estimation in Linear Regression Models / 110
3.3	Statistical Inference in Linear Regression / 119
3.3.1	Test for Significance of Regression / 120
3.3.2	Tests on Individual Regression Coefficients and Groups of Coefficients / 123
3.3.3	Confidence Intervals on Individual Regression Coefficients / 130
3.3.4	Confidence Intervals on the Mean Response / 131
3.4	Prediction of New Observations / 134
3.5	Model Adequacy Checking / 136
3.5.1	Residual Plots / 136
3.5.2	Scaled Residuals and PRESS / 139
3.5.3	Measures of Leverage and Influence / 144
3.6	Variable Selection Methods in Regression / 146
3.7	Generalized and Weighted Least Squares / 152
3.7.1	Generalized Least Squares / 153
3.7.2	Weighted Least Squares / 156
3.7.3	Discounted Least Squares / 161
3.8	Regression Models for General Time Series Data / 177



- 3.8.1 Detecting Autocorrelation: The Durbin–Watson Test / 178
- 3.8.2 Estimating the Parameters in Time Series Regression Models / 184
- 3.9 Econometric Models / 205
- 3.10 R Commands for Chapter 3 / 209
- Exercises / 219

## **4 EXPONENTIAL SMOOTHING METHODS 233**

- 4.1 Introduction / 233
- 4.2 First-Order Exponential Smoothing / 239
  - 4.2.1 The Initial Value,  $\tilde{y}_0$  / 241
  - 4.2.2 The Value of  $\lambda$  / 241
- 4.3 Modeling Time Series Data / 245
- 4.4 Second-Order Exponential Smoothing / 247
- 4.5 Higher-Order Exponential Smoothing / 257
- 4.6 Forecasting / 259
  - 4.6.1 Constant Process / 259
  - 4.6.2 Linear Trend Process / 264
  - 4.6.3 Estimation of  $\sigma_e^2$  / 273
  - 4.6.4 Adaptive Updating of the Discount Factor / 274
  - 4.6.5 Model Assessment / 276
- 4.7 Exponential Smoothing for Seasonal Data / 277
  - 4.7.1 Additive Seasonal Model / 277
  - 4.7.2 Multiplicative Seasonal Model / 280
- 4.8 Exponential Smoothing of Biosurveillance Data / 286
- 4.9 Exponential Smoothers and Arima Models / 299
- 4.10 R Commands for Chapter 4 / 300
- Exercises / 311

## **5 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS 327**

- 5.1 Introduction / 327
- 5.2 Linear Models for Stationary Time Series / 328
  - 5.2.1 Stationarity / 329
  - 5.2.2 Stationary Time Series / 329

- 5.3 Finite Order Moving Average Processes / 333
  - 5.3.1 The First-Order Moving Average Process, MA(1) / 334
  - 5.3.2 The Second-Order Moving Average Process, MA(2) / 336
- 5.4 Finite Order Autoregressive Processes / 337
  - 5.4.1 First-Order Autoregressive Process, AR(1) / 338
  - 5.4.2 Second-Order Autoregressive Process, AR(2) / 341
  - 5.4.3 General Autoregressive Process, AR( $p$ ) / 346
  - 5.4.4 Partial Autocorrelation Function, PACF / 348
- 5.5 Mixed Autoregressive–Moving Average Processes / 354
  - 5.5.1 Stationarity of ARMA( $p, q$ ) Process / 355
  - 5.5.2 Invertibility of ARMA( $p, q$ ) Process / 355
  - 5.5.3 ACF and PACF of ARMA( $p, q$ ) Process / 356
- 5.6 Nonstationary Processes / 363
  - 5.6.1 Some Examples of ARIMA( $p, d, q$ ) Processes / 363
- 5.7 Time Series Model Building / 367
  - 5.7.1 Model Identification / 367
  - 5.7.2 Parameter Estimation / 368
  - 5.7.3 Diagnostic Checking / 368
  - 5.7.4 Examples of Building ARIMA Models / 369
- 5.8 Forecasting Arima Processes / 378
- 5.9 Seasonal Processes / 383
- 5.10 Arima Modeling of Biosurveillance Data / 393
- 5.11 Final Comments / 399
- 5.12 R Commands for Chapter 5 / 401
- Exercises / 412

**6 TRANSFER FUNCTIONS AND INTERVENTION MODELS**

- 6.1 Introduction / 427
- 6.2 Transfer Function Models / 428
- 6.3 Transfer Function–Noise Models / 436
- 6.4 Cross-Correlation Function / 436
- 6.5 Model Specification / 438
- 6.6 Forecasting with Transfer Function–Noise Models / 456

- 6.7 Intervention Analysis / 462
- 6.8 R Commands for Chapter 6 / 473
- Exercises / 486

## **7 SURVEY OF OTHER FORECASTING METHODS 493**

- 7.1 Multivariate Time Series Models and Forecasting / 493
  - 7.1.1 Multivariate Stationary Process / 494
  - 7.1.2 Vector ARIMA Models / 494
  - 7.1.3 Vector AR (VAR) Models / 496
- 7.2 State Space Models / 502
- 7.3 Arch and Garch Models / 507
- 7.4 Direct Forecasting of Percentiles / 512
- 7.5 Combining Forecasts to Improve Prediction Performance / 518
- 7.6 Aggregation and Disaggregation of Forecasts / 522
- 7.7 Neural Networks and Forecasting / 526
- 7.8 Spectral Analysis / 529
- 7.9 Bayesian Methods in Forecasting / 535
- 7.10 Some Comments on Practical Implementation and Use of Statistical Forecasting Procedures / 542
- 7.11 R Commands for Chapter 7 / 545
- Exercises / 550

## **APPENDIX A STATISTICAL TABLES 561**

## **APPENDIX B DATA SETS FOR EXERCISES 581**

## **APPENDIX C INTRODUCTION TO R 627**

## **BIBLIOGRAPHY 631**

## **INDEX 639**



# PREFACE

---

Analyzing time-oriented data and forecasting future values of a time series are among the most important problems that analysts face in many fields, ranging from finance and economics to managing production operations, to the analysis of political and social policy sessions, to investigating the impact of humans and the policy decisions that they make on the environment. Consequently, there is a large group of people in a variety of fields, including finance, economics, science, engineering, statistics, and public policy who need to understand some basic concepts of time series analysis and forecasting. Unfortunately, most basic statistics and operations management books give little if any attention to time-oriented data and little guidance on forecasting. There are some very good high level books on time series analysis. These books are mostly written for technical specialists who are taking a doctoral-level course or doing research in the field. They tend to be very theoretical and often focus on a few specific topics or techniques. We have written this book to fill the gap between these two extremes.

We have made a number of changes in this revision of the book. New material has been added on data preparation for forecasting, including dealing with outliers and missing values, use of the variogram and sections on the spectrum, and an introduction to Bayesian methods in forecasting. We have added many new exercises and examples, including new data sets in Appendix B, and edited many sections of the text to improve the clarity of the presentation.

Like the first edition, this book is intended for practitioners who make real-world forecasts. We have attempted to keep the mathematical level modest to encourage a variety of users for the book. Our focus is on short- to medium-term forecasting where statistical methods are useful. Since many organizations can improve their effectiveness and business results by making better short- to medium-term forecasts, this book should be useful to a wide variety of professionals. The book can also be used as a textbook for an applied forecasting and time series analysis course at the advanced undergraduate or first-year graduate level. Students in this course could come from engineering, business, statistics, operations research, mathematics, computer science, and any area of application where making forecasts is important. Readers need a background in basic statistics (previous exposure to linear regression would be helpful, but not essential), and some knowledge of matrix algebra, although matrices appear mostly in the chapter on regression, and if one is interested mainly in the results, the details involving matrix manipulation can be skipped. Integrals and derivatives appear in a few places in the book, but no detailed working knowledge of calculus is required.

Successful time series analysis and forecasting requires that the analyst interact with computer software. The techniques and algorithms are just not suitable to manual calculations. We have chosen to demonstrate the techniques presented using three packages: Minitab<sup>®</sup>, JMP<sup>®</sup>, and R, and occasionally SAS<sup>®</sup>. We have selected these packages because they are widely used in practice and because they have generally good capability for analyzing time series data and generating forecasts. Because R is increasingly popular in statistics courses, we have included a section in each chapter showing the R code necessary for working some of the examples in the chapter. We have also added a brief appendix on the use of R. The basic principles that underlie most of our presentation are not specific to any particular software package. Readers can use any software that they like or have available that has basic statistical forecasting capability. While the text examples do utilize these particular software packages and illustrate some of their features and capability, these features or similar ones are found in many other software packages.

There are three basic approaches to generating forecasts: regression-based methods, heuristic smoothing methods, and general time series models. Because all three of these basic approaches are useful, we give an introduction to all of them. Chapter 1 introduces the basic forecasting problem, defines terminology, and illustrates many of the common features of time series data. Chapter 2 contains many of the basic statistical tools used in analyzing time series data. Topics include plots, numerical

summaries of time series data including the autocovariance and autocorrelation functions, transformations, differencing, and decomposing a time series into trend and seasonal components. We also introduce metrics for evaluating forecast errors and methods for evaluating and tracking forecasting performance over time. Chapter 3 discusses regression analysis and its use in forecasting. We discuss both crosssection and time series regression data, least squares and maximum likelihood model fitting, model adequacy checking, prediction intervals, and weighted and generalized least squares. The first part of this chapter covers many of the topics typically seen in an introductory treatment of regression, either in a stand-alone course or as part of another applied statistics course. It should be a reasonable review for many readers. Chapter 4 presents exponential smoothing techniques, both for time series with polynomial components and for seasonal data. We discuss and illustrate methods for selecting the smoothing constant(s), forecasting, and constructing prediction intervals. The explicit time series modeling approach to forecasting that we have chosen to emphasize is the autoregressive integrated moving average (ARIMA) model approach. Chapter 5 introduces ARIMA models and illustrates how to identify and fit these models for both nonseasonal and seasonal time series. Forecasting and prediction interval construction are also discussed and illustrated. Chapter 6 extends this discussion into transfer function models and intervention modeling and analysis. Chapter 7 surveys several other useful topics from time series analysis and forecasting, including multivariate time series problems, ARCH and GARCH models, and combinations of forecasts. We also give some practical advice for using statistical approaches to forecasting and provide some information about realistic expectations. The last two chapters of the book are somewhat higher in level than the first five.

Each chapter has a set of exercises. Some of these exercises involve analyzing the data sets given in Appendix B. These data sets represent an interesting cross section of real time series data, typical of those encountered in practical forecasting problems. Most of these data sets are used in exercises in two or more chapters, an indication that there are usually several approaches to analyzing, modeling, and forecasting a time series. There are other good sources of data for practicing the techniques given in this book. Some of the ones that we have found very interesting and useful include the U.S. Department of Labor—Bureau of Labor Statistics (<http://www.bls.gov/data/home.htm>), the U.S. Department of Agriculture—National Agricultural Statistics Service, Quick Stats Agricultural Statistics Data ([http://www.nass.usda.gov/Data\\_and\\_Statistics/Quick\\_Stats/index.asp](http://www.nass.usda.gov/Data_and_Statistics/Quick_Stats/index.asp)), the U.S. Census Bureau (<http://www.census.gov>), and the U.S.

Department of the Treasury (<http://www.treas.gov/offices/domestic-finance/debt-management/interest-rate/>). The time series data library created by Rob Hyndman at Monash University (<http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/index.htm>) and the time series data library at the Mathematics Department of the University of York (<http://www.york.ac.uk/depts/math/data/ts/>) also contain many excellent data sets. Some of these sources provide links to other data. Data sets and other materials related to this book can be found at <ftp://ftp.wiley.com/public/scitechmed/timeseries>.

We would like to thank the many individuals who provided feedback and suggestions for improvement to the first edition. We found these suggestions most helpful. We are indebted to Clifford Long who generously provided the R codes he used with his students when he taught from the book. We found his codes very helpful in putting the end-of-chapter R code sections together. We also have placed a premium in the book on bridging the gap between theory and practice. We have not emphasized proofs or technical details and have tried to give intuitive explanations of the material whenever possible. The result is a book that can be used with a wide variety of audiences, with different interests and technical backgrounds, whose common interests are understanding how to analyze time-oriented data and constructing good short-term statistically based forecasts.

We express our appreciation to the individuals and organizations who have given their permission to use copyrighted material. These materials are noted in the text. Portions of the output contained in this book are printed with permission of Minitab Inc. All material remains the exclusive property and copyright of Minitab Inc. All rights reserved.

DOUGLAS C. MONTGOMERY  
CHERYL L. JENNINGS  
MURAT KULAHCI



# CHAPTER 1

---

## INTRODUCTION TO FORECASTING

---

It is difficult to make predictions, especially about the future  
NEILS BOHR, *Danish physicist*

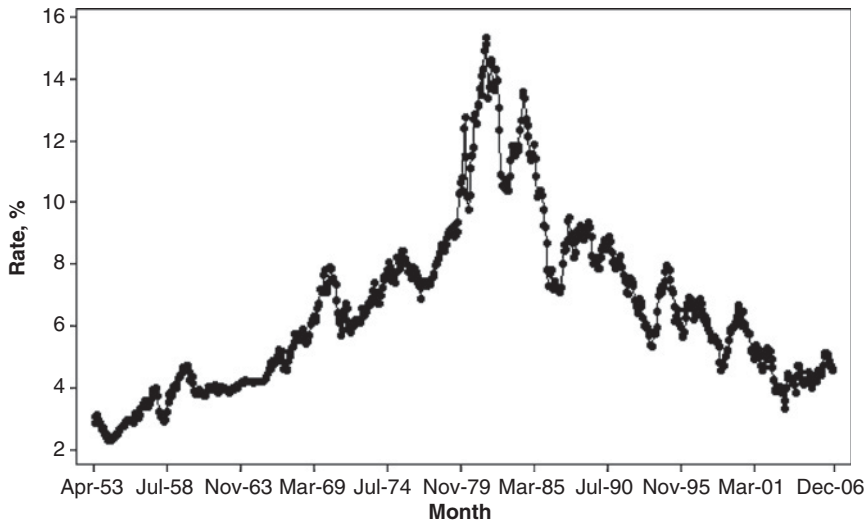
### 1.1 THE NATURE AND USES OF FORECASTS

A **forecast** is a prediction of some future event or events. As suggested by Neils Bohr, making good predictions is not always easy. Famously “bad” forecasts include the following from the book *Bad Predictions*:

- “The population is constant in size and will remain so right up to the end of mankind.” *L’Encyclopedie*, 1756.
- “1930 will be a splendid employment year.” U.S. Department of Labor, *New Year’s Forecast* in 1929, just before the market crash on October 29.
- “Computers are multiplying at a rapid rate. By the turn of the century there will be 220,000 in the U.S.” *Wall Street Journal*, 1966.

Forecasting is an important problem that spans many fields including business and industry, government, economics, environmental sciences, medicine, social science, politics, and finance. Forecasting problems are often classified as short-term, medium-term, and long-term. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, and months) into the future. Medium-term forecasts extend from 1 to 2 years into the future, and long-term forecasting problems can extend beyond that by many years. Short- and medium-term forecasts are required for activities that range from operations management to budgeting and selecting new research and development projects. Long-term forecasts impact issues such as strategic planning. Short- and medium-term forecasting is typically based on identifying, modeling, and extrapolating the patterns found in historical data. Because these historical data usually exhibit inertia and do not change dramatically very quickly, statistical methods are very useful for short- and medium-term forecasting. This book is about the use of these statistical methods.

Most forecasting problems involve the use of time series data. A **time series** is a time-oriented or chronological sequence of observations on a variable of interest. For example, Figure 1.1 shows the market yield on US Treasury Securities at 10-year constant maturity from April 1953 through December 2006 (data in Appendix B, Table B.1). This graph is called a **time**



**FIGURE 1.1** Time series plot of the market yield on US Treasury Securities at 10-year constant maturity. *Source:* US Treasury.

**series plot.** The rate variable is collected at equally spaced time periods, as is typical in most time series and forecasting applications. Many business applications of forecasting utilize daily, weekly, monthly, quarterly, or annual data, but any reporting interval may be used. Furthermore, the data may be instantaneous, such as the viscosity of a chemical product at the point in time where it is measured; it may be cumulative, such as the total sales of a product during the month; or it may be a statistic that in some way reflects the activity of the variable during the time period, such as the daily closing price of a specific stock on the New York Stock Exchange.

The reason that forecasting is so important is that prediction of future events is a critical input into many types of planning and decision-making processes, with application to areas such as the following:

1. *Operations Management.* Business organizations routinely use forecasts of product sales or demand for services in order to schedule production, control inventories, manage the supply chain, determine staffing requirements, and plan capacity. Forecasts may also be used to determine the mix of products or services to be offered and the locations at which products are to be produced.
2. *Marketing.* Forecasting is important in many marketing decisions. Forecasts of sales response to advertising expenditures, new promotions, or changes in pricing policies enable businesses to evaluate their effectiveness, determine whether goals are being met, and make adjustments.
3. *Finance and Risk Management.* Investors in financial assets are interested in forecasting the returns from their investments. These assets include but are not limited to stocks, bonds, and commodities; other investment decisions can be made relative to forecasts of interest rates, options, and currency exchange rates. Financial risk management requires forecasts of the volatility of asset returns so that the risks associated with investment portfolios can be evaluated and insured, and so that financial derivatives can be properly priced.
4. *Economics.* Governments, financial institutions, and policy organizations require forecasts of major economic variables, such as gross domestic product, population growth, unemployment, interest rates, inflation, job growth, production, and consumption. These forecasts are an integral part of the guidance behind monetary and fiscal policy, and budgeting plans and decisions made by governments. They are also instrumental in the strategic planning decisions made by business organizations and financial institutions.

5. *Industrial Process Control*. Forecasts of the future values of critical quality characteristics of a production process can help determine when important controllable variables in the process should be changed, or if the process should be shut down and overhauled. Feedback and feedforward control schemes are widely used in monitoring and adjustment of industrial processes, and predictions of the process output are an integral part of these schemes.
6. *Demography*. Forecasts of population by country and regions are made routinely, often stratified by variables such as gender, age, and race. Demographers also forecast births, deaths, and migration patterns of populations. Governments use these forecasts for planning policy and social service actions, such as spending on health care, retirement programs, and antipoverty programs. Many businesses use forecasts of populations by age groups to make strategic plans regarding developing new product lines or the types of services that will be offered.

These are only a few of the many different situations where forecasts are required in order to make good decisions. Despite the wide range of problem situations that require forecasts, there are only two broad types of forecasting techniques—qualitative methods and quantitative methods.

**Qualitative** forecasting techniques are often subjective in nature and require judgment on the part of experts. Qualitative forecasts are often used in situations where there is little or no historical data on which to base the forecast. An example would be the introduction of a new product, for which there is no relevant history. In this situation, the company might use the expert opinion of sales and marketing personnel to subjectively estimate product sales during the new product introduction phase of its life cycle. Sometimes qualitative forecasting methods make use of marketing tests, surveys of potential customers, and experience with the sales performance of other products (both their own and those of competitors). However, although some data analysis may be performed, the basis of the forecast is subjective judgment.

Perhaps the most formal and widely known qualitative forecasting technique is the **Delphi Method**. This technique was developed by the RAND Corporation (see Dalkey [1967]). It employs a panel of experts who are assumed to be knowledgeable about the problem. The panel members are physically separated to avoid their deliberations being impacted either by social pressures or by a single dominant individual. Each panel member responds to a questionnaire containing a series of questions and returns the information to a coordinator. Following the first questionnaire, subsequent

questions are submitted to the panelists along with information about the opinions of the panel as a group. This allows panelists to review their predictions relative to the opinions of the entire group. After several rounds, it is hoped that the opinions of the panelists converge to a consensus, although achieving a consensus is not required and justified differences of opinion can be included in the outcome. Qualitative forecasting methods are not emphasized in this book.

**Quantitative** forecasting techniques make formal use of historical data and a **forecasting model**. The model formally summarizes patterns in the data and expresses a statistical relationship between previous and current values of the variable. Then the model is used to project the patterns in the data into the future. In other words, the forecasting model is used to extrapolate past and current behavior into the future. There are several types of forecasting models in general use. The three most widely used are regression models, smoothing models, and general time series models. Regression models make use of relationships between the variable of interest and one or more related predictor variables. Sometimes regression models are called **causal forecasting models**, because the predictor variables are assumed to describe the forces that cause or drive the observed values of the variable of interest. An example would be using data on house purchases as a predictor variable to forecast furniture sales. The method of least squares is the formal basis of most regression models. **Smoothing models** typically employ a simple function of previous observations to provide a forecast of the variable of interest. These methods may have a formal statistical basis, but they are often used and justified heuristically on the basis that they are easy to use and produce satisfactory results. General **time series models** employ the statistical properties of the historical data to specify a formal model and then estimate the unknown parameters of this model (usually) by least squares. In subsequent chapters, we will discuss all three types of quantitative forecasting models.

The form of the forecast can be important. We typically think of a forecast as a single number that represents our best estimate of the future value of the variable of interest. Statisticians would call this a **point estimate** or **point forecast**. Now these forecasts are almost always wrong; that is, we experience **forecast error**. Consequently, it is usually a good practice to accompany a forecast with an estimate of how large a forecast error might be experienced. One way to do this is to provide a **prediction interval (PI)** to accompany the point forecast. The PI is a range of values for the future observation, and it is likely to prove far more useful in decision-making than a single number. We will show how to obtain PIs for most of the forecasting methods discussed in the book.

Other important features of the forecasting problem are the **forecast horizon** and the **forecast interval**. The forecast horizon is the number of future periods for which forecasts must be produced. The horizon is often dictated by the nature of the problem. For example, in production planning, forecasts of product demand may be made on a monthly basis. Because of the time required to change or modify a production schedule, ensure that sufficient raw material and component parts are available from the supply chain, and plan the delivery of completed goods to customers or inventory facilities, it would be necessary to forecast up to 3 months ahead. The forecast horizon is also often called the forecast **lead time**. The **forecast interval** is the frequency with which new forecasts are prepared. For example, in production planning, we might forecast demand on a monthly basis, for up to 3 months in the future (the lead time or horizon), and prepare a new forecast each month. Thus the forecast interval is 1 month, the same as the basic period of time for which each forecast is made. If the forecast lead time is always the same length, say,  $T$  periods, and the forecast is revised each time period, then we are employing a **rolling or moving horizon** forecasting approach. This system updates or revises the forecasts for  $T-1$  of the periods in the horizon and computes a forecast for the newest period  $T$ . This rolling horizon approach to forecasting is widely used when the lead time is several periods long.

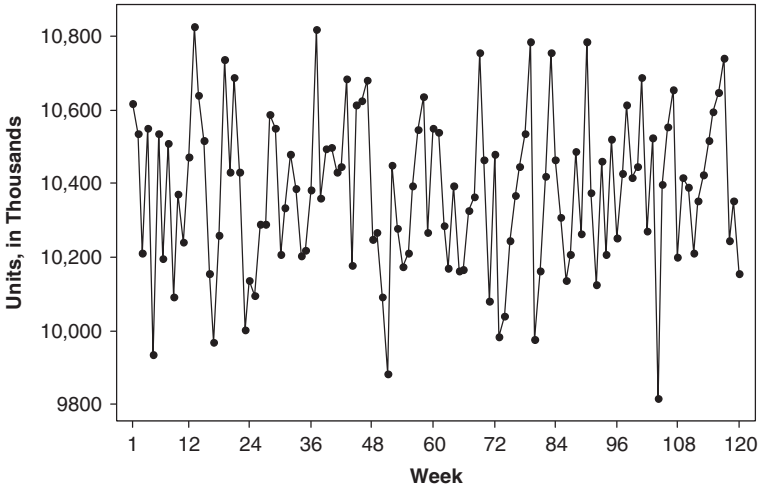
## 1.2 SOME EXAMPLES OF TIME SERIES

Time series plots can reveal **patterns** such as random, trends, level shifts, periods or cycles, unusual observations, or a combination of patterns. Patterns commonly found in time series data are discussed next with examples of situations that drive the patterns.

The sales of a mature pharmaceutical product may remain relatively flat in the absence of unchanged marketing or manufacturing strategies. Weekly sales of a generic pharmaceutical product shown in Figure 1.2 appear to be constant over time, at about  $10,400 \times 10^3$  units, in a random sequence with no obvious patterns (data in Appendix B, Table B.2).

To assure conformance with customer requirements and product specifications, the production of chemicals is monitored by many characteristics. These may be input variables such as temperature and flow rate, and output properties such as viscosity and purity.

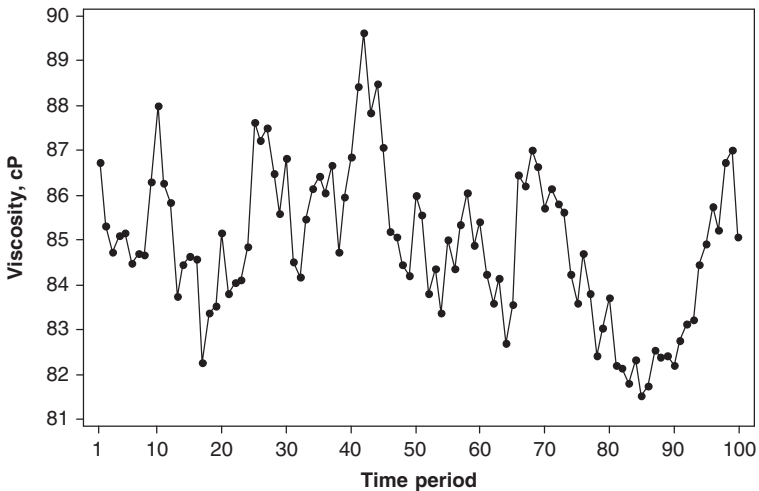
Due to the continuous nature of chemical manufacturing processes, output properties often are **positively autocorrelated**; that is, a value above the long-run average tends to be followed by other values above the



**FIGURE 1.2** Pharmaceutical product sales.

average, while a value below the average tends to be followed by other values below the average.

The viscosity readings plotted in Figure 1.3 exhibit autocorrelated behavior, tending to a long-run average of about 85 centipoises (cP), but with a structured, not completely random, appearance (data in Appendix B, Table B.3). Some methods for describing and analyzing autocorrelated data will be described in Chapter 2.



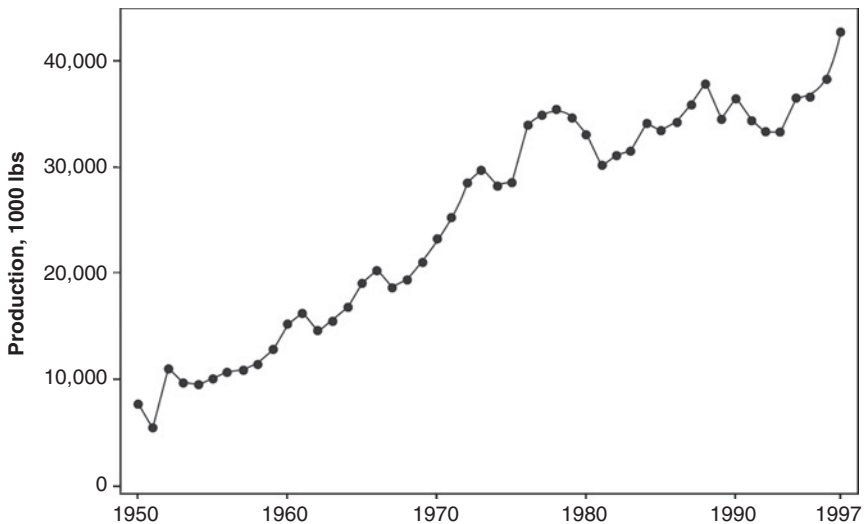
**FIGURE 1.3** Chemical process viscosity readings.

The USDA National Agricultural Statistics Service publishes agricultural statistics for many commodities, including the annual production of dairy products such as butter, cheese, ice cream, milk, yogurt, and whey. These statistics are used for market analysis and intelligence, economic indicators, and identification of emerging issues.

Blue and gorgonzola cheese is one of 32 categories of cheese for which data are published. The annual US production of blue and gorgonzola cheeses (in  $10^3$  lb) is shown in Figure 1.4 (data in Appendix B, Table B.4). Production quadrupled from 1950 to 1997, and the **linear trend** has a constant positive slope with random, year-to-year variation.

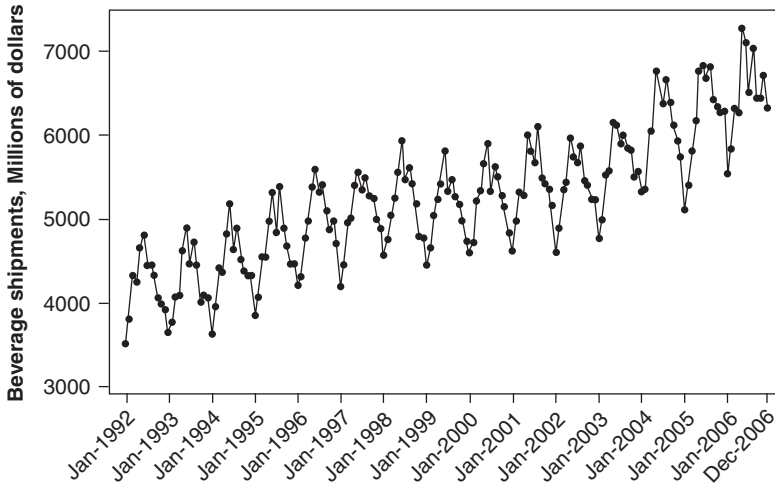
The US Census Bureau publishes historic statistics on manufacturers' shipments, inventories, and orders. The statistics are based on North American Industry Classification System (NAICS) code and are utilized for purposes such as measuring productivity and analyzing relationships between employment and manufacturing output.

The manufacture of beverage and tobacco products is reported as part of the nondurable subsector. The plot of monthly beverage product shipments (Figure 1.5) reveals an overall increasing trend, with a distinct **cyclic pattern** that is repeated within each year. January shipments appear to be the lowest, with highs in May and June (data in Appendix B, Table B.5). This monthly, or **seasonal**, variation may be attributable to some cause



**FIGURE 1.4** The US annual production of blue and gorgonzola cheeses. *Source:* USDA–NASS.





**FIGURE 1.5** The US beverage manufacturer monthly product shipments, unadjusted. *Source:* US Census Bureau.

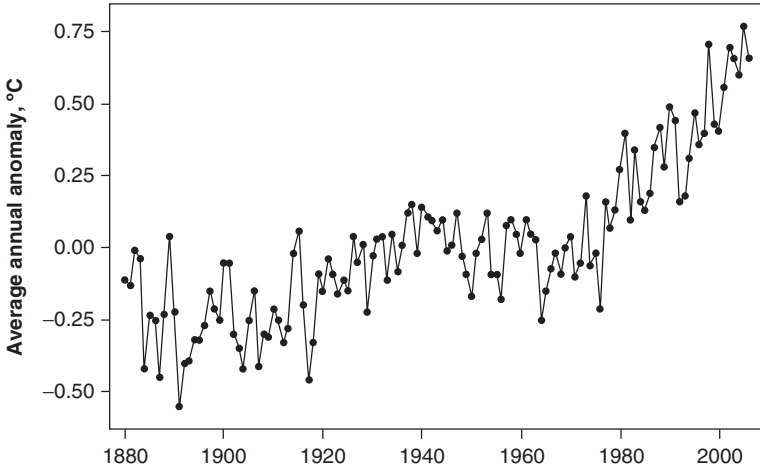
such as the impact of weather on the demand for beverages. Techniques for making seasonal adjustments to data in order to better understand general trends will be discussed in Chapter 2.

To determine whether the Earth is warming or cooling, scientists look at annual mean temperatures. At a single station, the warmest and the coolest temperatures in a day are averaged. Averages are then calculated at stations all over the Earth, over an entire year. The change in global annual mean surface air temperature is calculated from a base established from 1951 to 1980, and the result is reported as an “anomaly.”

The plot of the annual mean anomaly in global surface air temperature (Figure 1.6) shows an increasing trend since 1880; however, the slope, or rate of change, varies with time periods (data in Appendix B, Table B.6). While the slope in earlier time periods appears to be constant, slightly increasing, or slightly decreasing, the slope from about 1975 to the present appears much steeper than the rest of the plot.

Business data such as stock prices and interest rates often exhibit **non-stationary** behavior; that is, the time series has no natural mean. The daily closing price adjusted for stock splits of Whole Foods Market (WFMI) stock in 2001 (Figure 1.7) exhibits a combination of patterns for both mean level and slope (data in Appendix B, Table B.7).

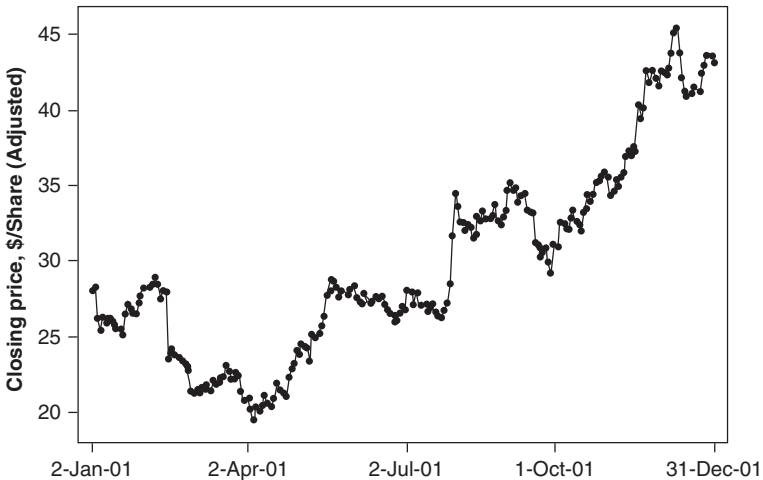
While the price is constant in some short time periods, there is no consistent mean level over time. In other time periods, the price changes



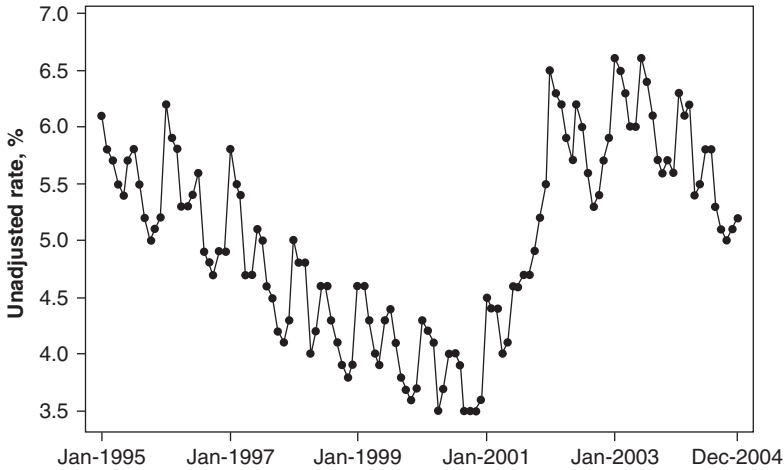
**FIGURE 1.6** Global mean surface air temperature annual anomaly. *Source:* NASA-GISS.

at different rates, including occasional abrupt shifts in level. This is an example of nonstationary behavior, which will be discussed in Chapter 2.

The Current Population Survey (CPS) or “household survey” prepared by the US Department of Labor, Bureau of Labor Statistics, contains national data on employment, unemployment, earnings, and other labor market topics by demographic characteristics. The data are used to report



**FIGURE 1.7** Whole foods market stock price, daily closing adjusted for splits.



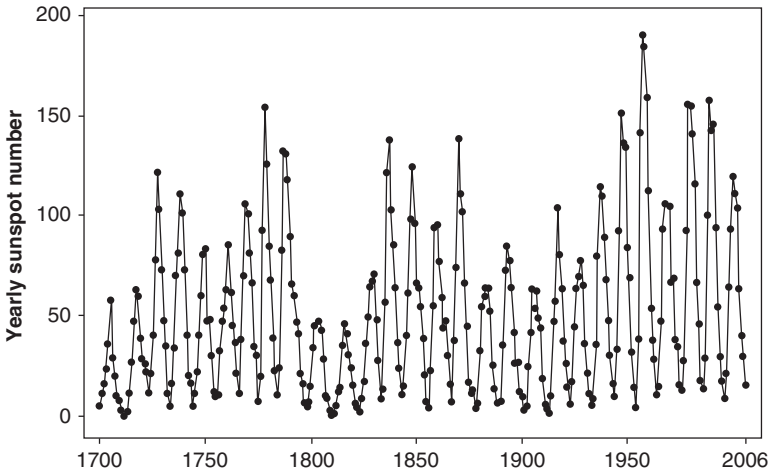
**FIGURE 1.8** Monthly unemployment rate—full-time labor force, unadjusted.  
*Source:* US Department of Labor-BLS.

on the employment situation, for projections with impact on hiring and training, and for a multitude of other business planning activities. The data are reported unadjusted and with seasonal adjustment to remove the effect of regular patterns that occur each year.

The plot of monthly unadjusted unemployment rates (Figure 1.8) exhibits a mixture of patterns, similar to Figure 1.5 (data in Appendix B, Table B.8). There is a distinct cyclic pattern within a year; January, February, and March generally have the highest unemployment rates. The overall level is also changing, from a gradual decrease, to a steep increase, followed by a gradual decrease. The use of seasonal adjustments as described in Chapter 2 makes it easier to observe the nonseasonal movements in time series data.

Solar activity has long been recognized as a significant source of noise impacting consumer and military communications, including satellites, cell phone towers, and electric power grids. The ability to accurately forecast solar activity is critical to a variety of fields. The International Sunspot Number  $R$  is the oldest solar activity index. The number incorporates both the number of observed sunspots and the number of observed sunspot groups. In Figure 1.9, the plot of annual sunspot numbers reveals cyclic patterns of varying magnitudes (data in Appendix B, Table B.9).

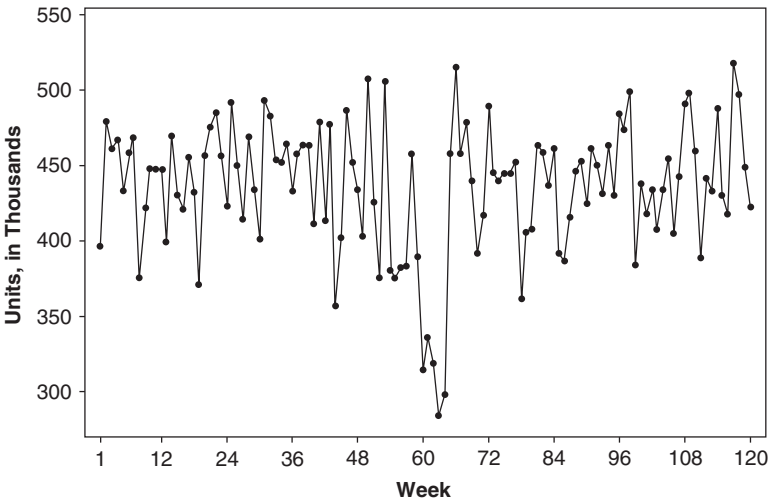
In addition to assisting in the identification of steady-state patterns, time series plots may also draw attention to the occurrence of **atypical events**. Weekly sales of a generic pharmaceutical product dropped due to limited



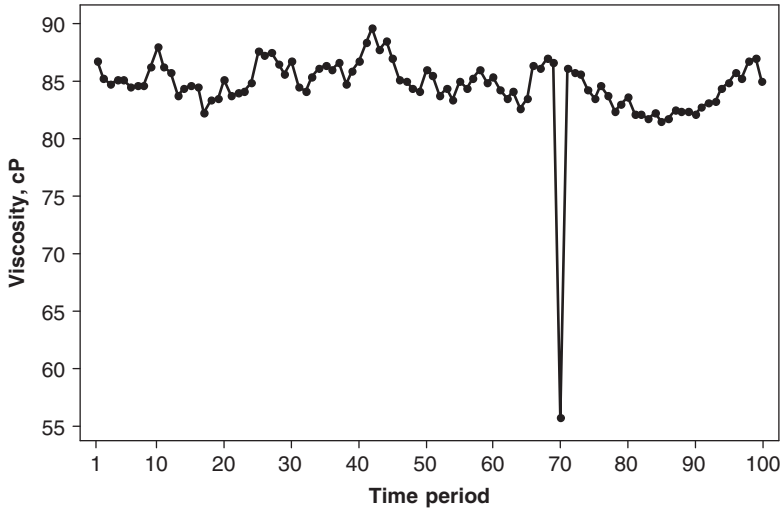
**FIGURE 1.9** The international sunspot number. *Source:* SIDC.

availability resulting from a fire at one of the four production facilities. The 5-week reduction is apparent in the time series plot of weekly sales shown in Figure 1.10.

Another type of unusual event may be the failure of the data measurement or collection system. After recording a vastly different viscosity reading at time period 70 (Figure 1.11), the measurement system was



**FIGURE 1.10** Pharmaceutical product sales.



**FIGURE 1.11** Chemical process viscosity readings, with sensor malfunction.

checked with a standard and determined to be out of calibration. The cause was determined to be a malfunctioning sensor.

### 1.3 THE FORECASTING PROCESS

A process is a series of connected activities that transform one or more inputs into one or more outputs. All work activities are performed in processes, and forecasting is no exception. The activities in the forecasting process are:

1. Problem definition
2. Data collection
3. Data analysis
4. Model selection and fitting
5. Model validation
6. Forecasting model deployment
7. Monitoring forecasting model performance

These activities are shown in Figure 1.12.

**Problem definition** involves developing understanding of how the forecast will be used along with the expectations of the “customer” (the user of

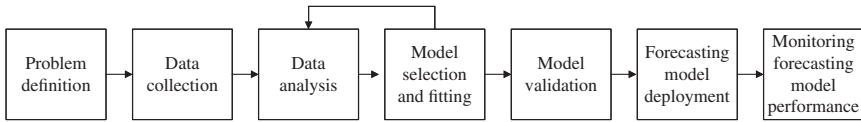


FIGURE 1.12 The forecasting process.

the forecast). Questions that must be addressed during this phase include the desired form of the forecast (e.g., are monthly forecasts required), the forecast horizon or lead time, how often the forecasts need to be revised (the forecast interval), and what level of forecast accuracy is required in order to make good business decisions. This is also an opportunity to introduce the decision makers to the use of prediction intervals as a measure of the risk associated with forecasts, if they are unfamiliar with this approach. Often it is necessary to go deeply into many aspects of the business system that requires the forecast to properly define the forecasting component of the entire problem. For example, in designing a forecasting system for inventory control, information may be required on issues such as product shelf life or other aging considerations, the time required to manufacture or otherwise obtain the products (production lead time), and the economic consequences of having too many or too few units of product available to meet customer demand. When multiple products are involved, the level of aggregation of the forecast (e.g., do we forecast individual products or families consisting of several similar products) can be an important consideration. Much of the ultimate success of the forecasting model in meeting the customer expectations is determined in the problem definition phase.

**Data collection** consists of obtaining the relevant history for the variable(s) that are to be forecast, including historical information on potential predictor variables.

The key here is “relevant”; often information collection and storage methods and systems change over time and not all historical data are useful for the current problem. Often it is necessary to deal with missing values of some variables, potential outliers, or other data-related problems that have occurred in the past. During this phase, it is also useful to begin planning how the data collection and storage issues in the future will be handled so that the reliability and integrity of the data will be preserved.

**Data analysis** is an important preliminary step to the selection of the forecasting model to be used. Time series plots of the data should be constructed and visually inspected for recognizable patterns, such as trends and seasonal or other cyclical components. A trend is evolutionary movement, either upward or downward, in the value of the variable. Trends may